Interactive Visualization of Large Simulation Datasets

Andreas Gerndt, Rolf Hempel, Robin Wolff
DLR Simulation and Software Technology

EuroMPI 2010 Conference, Stuttgart, Sept. 13-15, 2010

Introduction

DLR
German Aerospace Center

- Research on
  - Aeronautics
  - Space
  - Transport
  - Energy
- Space Agency
- Project Management Agency

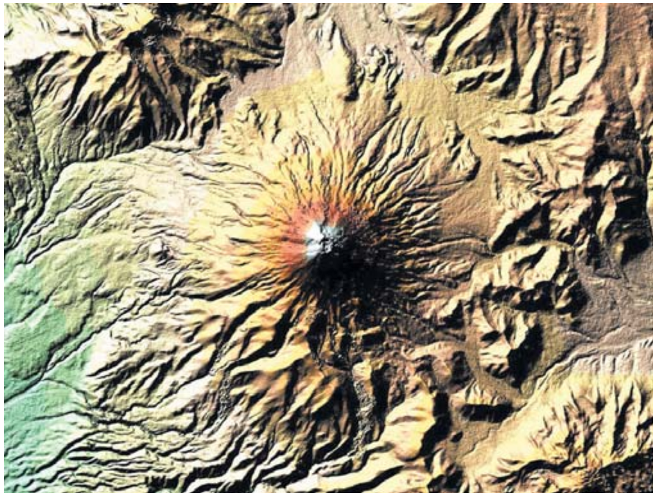
Locations and employees

6500 employees across
29 research institutes and
facilities at

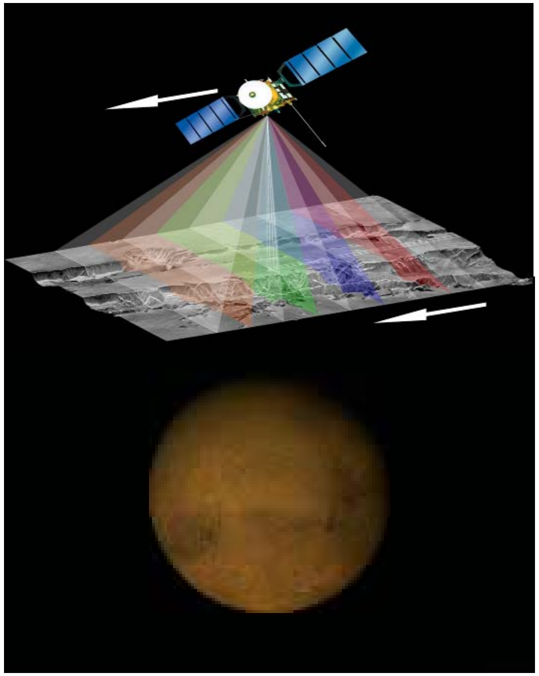
- 13 sites.

Offices in Brussels,
Paris and Washington.

Visualization at DLR
Applications in Experiment and Simulation



- Earth observation
  - National data repository
  - Technology development
    - Left: SAR image of Cotopaxi volcano
  - Atmospheric and climate research



- Planetary science
  - Several space missions, e.g. Mars Express
  - High-resolution Stereo Camera
  - Multi-spectral data
  - 10 – 100 meters / pixel

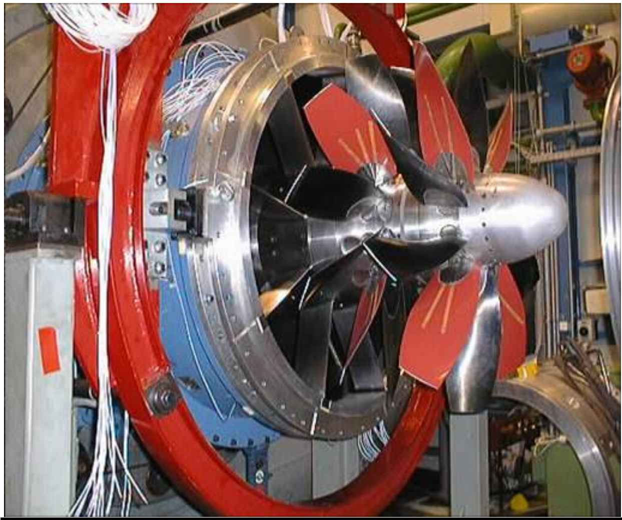
Visualization at DLR
Applications in Experiment and Simulation

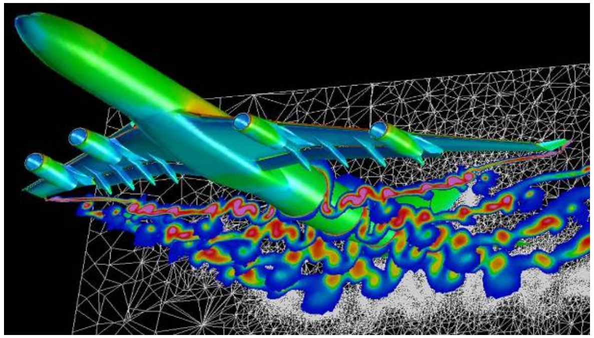
- Simulators for cars, aircraft, helicopters, …
  - Left: Dynamic car simulator
  - Research into driver assistance systems
  - Visualization of external view

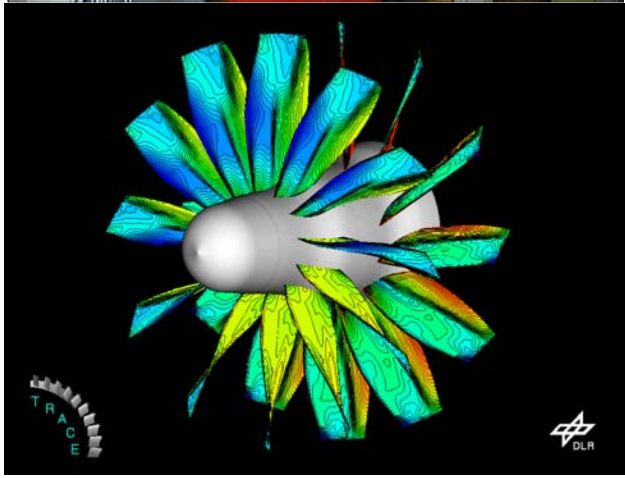
- Robotics / Mechatronics
  - Spacecraft design
  - On-orbit satellite servicing
  - Astronaut training

Visualization at DLR
Applications in Experiment and Simulation

- Aerodynamics
  - CFD and experiment
  - Turbines, airplanes, helicopters, spacecraft
  - Development of CFD codes (TAU, TRACE)
  - Experiments used to validate simulations
- Prominent DLR institutes specialized in CFD
  - Inst. for aerodynamics and flow technology
  - Institute for propulsion technology

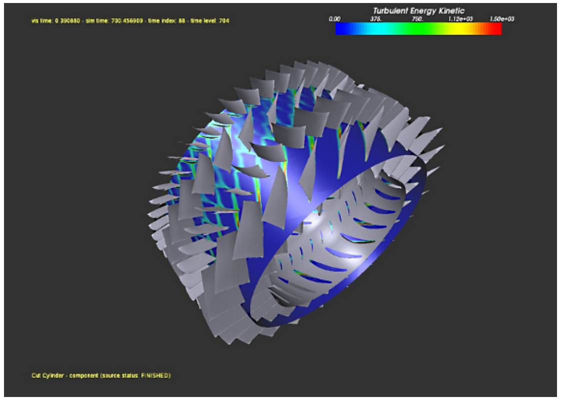


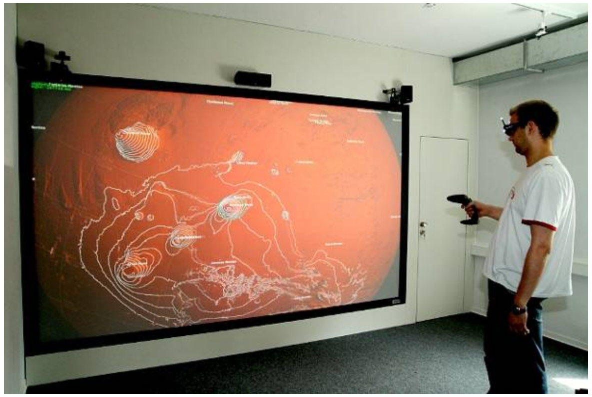


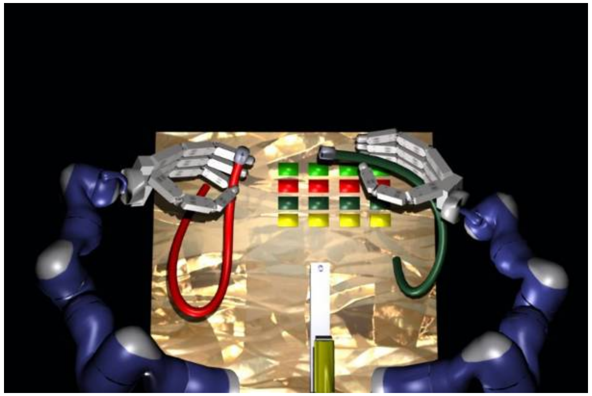


Visualization at DLR
DLR Simulation and Software Technology

- Research on software / computer science topics
- Co-operations with engineering institutes in advanced software projects
- Main topics include HPC and Visualization
- Visualization projects with DLR institutes on
  - CFD around airplanes / in turbomachines
  - planetary surface data
  - on-orbit servicing of satellites

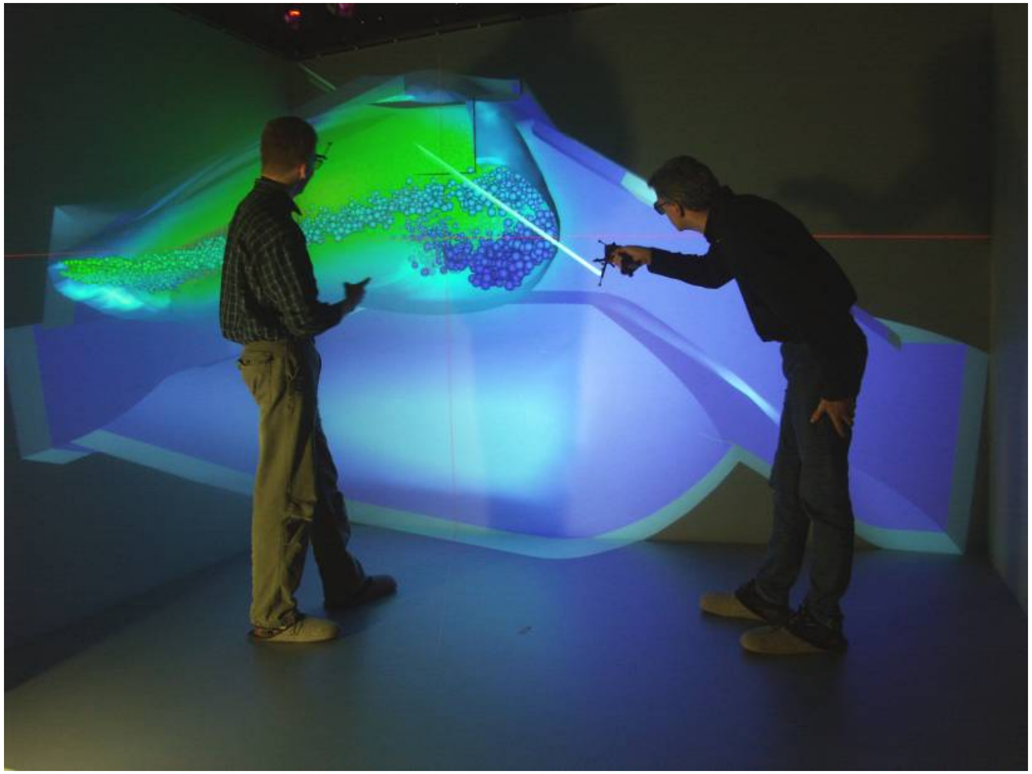






Objectives of Virtual Reality (VR) Research

- Important aspects
  - Immersion
  - Interactivity
  - 3D Visualization
  - Multi-modality



Interactive CFD data exploration in a virtual environment

Overview

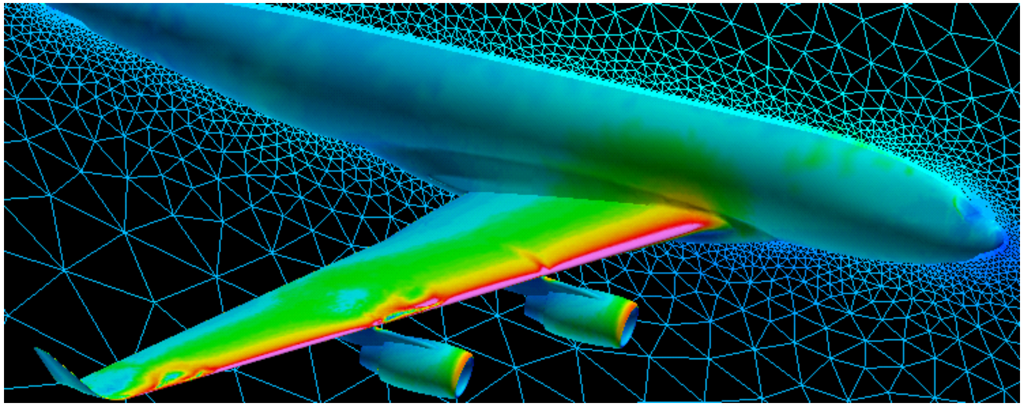
- **Problem Description**
- **Pipeline Distribution / Parallelization Framework**
- **Large-scale Dataset I/O**
- **Multi-Resolution and Data Streaming**
- **Co-Execution of Simulation and Post-Processing**

Problem Description

Problem Description

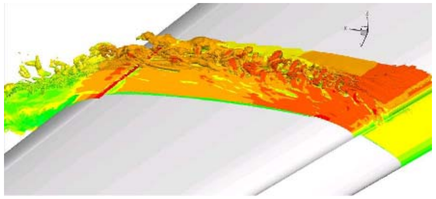
What is a „large“ simulation dataset?

- CFD calculations at the Institute for aerodynamics and flow technology with the TAU code
  - Solution of the Raynolds-averaged Navier-Stokes equations
  - HPC system: $C^2A^2S^2E$ (see next slide)



Simulation of the air flow around a complete transport airplane configuration

  - 300-400 / 2000 mio. grid points (without / with acoustics)
  - Test case: Slat track noise (LES)
    - 80 mio. pnts., 6.7 GB, 10,000 time steps
    - 19 minutes / time step on 2048 cores



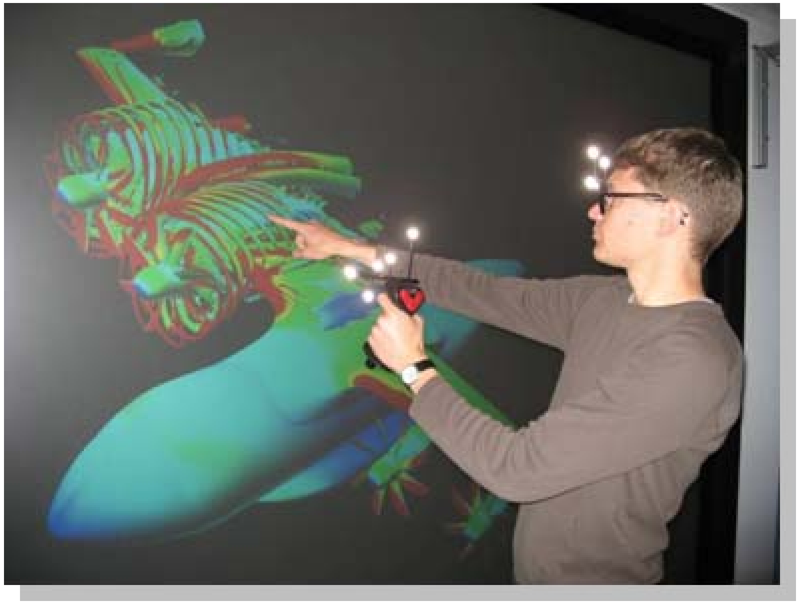
Problem Description
The $C^2A^2S^2E$ Supercomputer

- $C^2A^2S^2E$ Cluster (SUN)
  - 16 compute racks, 48 blades each
  - 2 AMD Opteron quad core processors, 1.9 GHz (Barcelona)
- 768 nodes, 6144 cores
  - 46.6 TFlop/s Peak performance
  - TAU 1 core: 1 GFlop/s
    all cores: 3 TFlop/s
- 12288 GB Main memory
  - 16 GB per node (758 nodes)
  - 32 GB per node (10 nodes)
- SUN Mega Switch
- Parallel file system (GPFS)
- High speed link to Airbus Bremen (100 Mbit/s)
- Dedicated to aircraft research
- 2010 performance update by factor 3

DLR AIRBUS Niedersachsen

Problem Description
First Experiences with Virtual Reality

- 2007: VR system set up at Institute for aerodynamics and flow technology
  - Powerwall 2.7m x 2.0m
  - 2 projectors
  - IR tracking system
  - 2 dual-core Opteron processors
  - Visualization software: Ensight Gold 8.2
- Problems
  - System cannot handle full-resolution datasets
  - Visualization is too slow
  - Missing interaction metaphors
- System mostly used for „Colorful pictures for visitors“

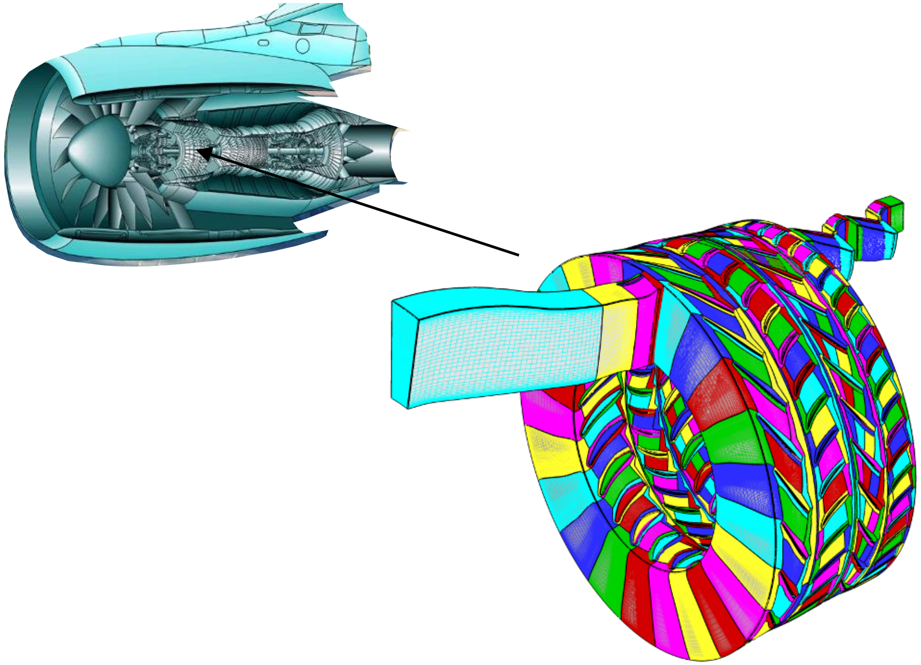


Problem Description

CFD Simulation of a Turbo-Engine Compressor Section

- Four-stage compressor section
- Simulation of the transition point between laminar and turbulent flow → instationary problem
- Computational domain split into blocks (colored)
- 123 Million mesh points
- National Supercomputer HLRB-II (SGI Altix 4700) at the Leibnitz-Rechenzentrum in Munich

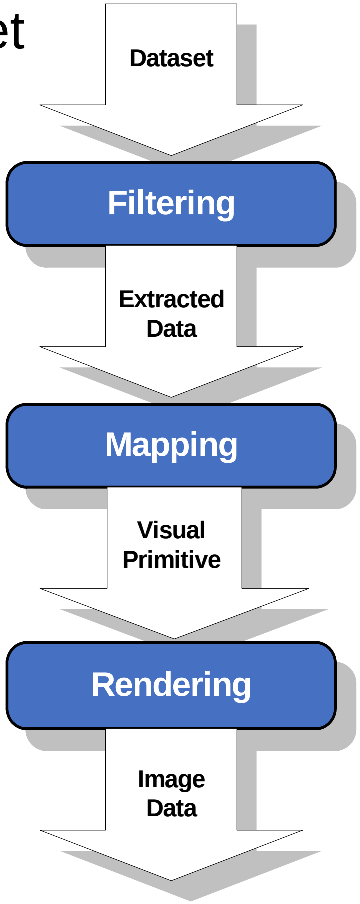
- In total 15 Terabytes generated
- Compute time: 210,000 CPUh
- Example is two years old
- Today: Intention to run a problem three times as large



Pipeline Distribution

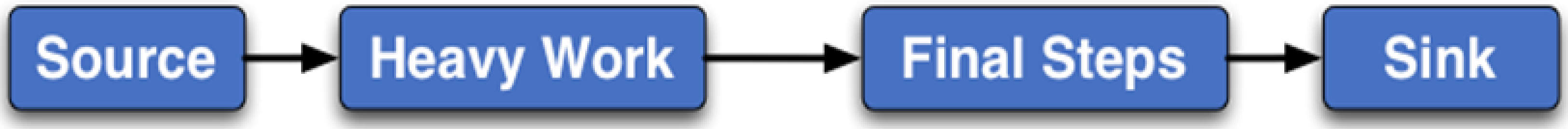
Pipeline Distribution
Visualization Pipeline

- Visual analysis of simulation data is organized as a post-processing pipeline
- Pipeline input: large-scale grid-based CFD dataset
- Filtering stage processes the data
  - Produces more manageable data sizes (cutting, clipping, converting, merging, re-sampling, …)
  - Extracts features of interest (isosurfaces, particle tracing, vortices, topology information, …)
  - May contain multiple processing steps and can be organized as a data flow network (multiple inputs and outputs)
- Mapping
  - Creates visualization objects
- Rendering



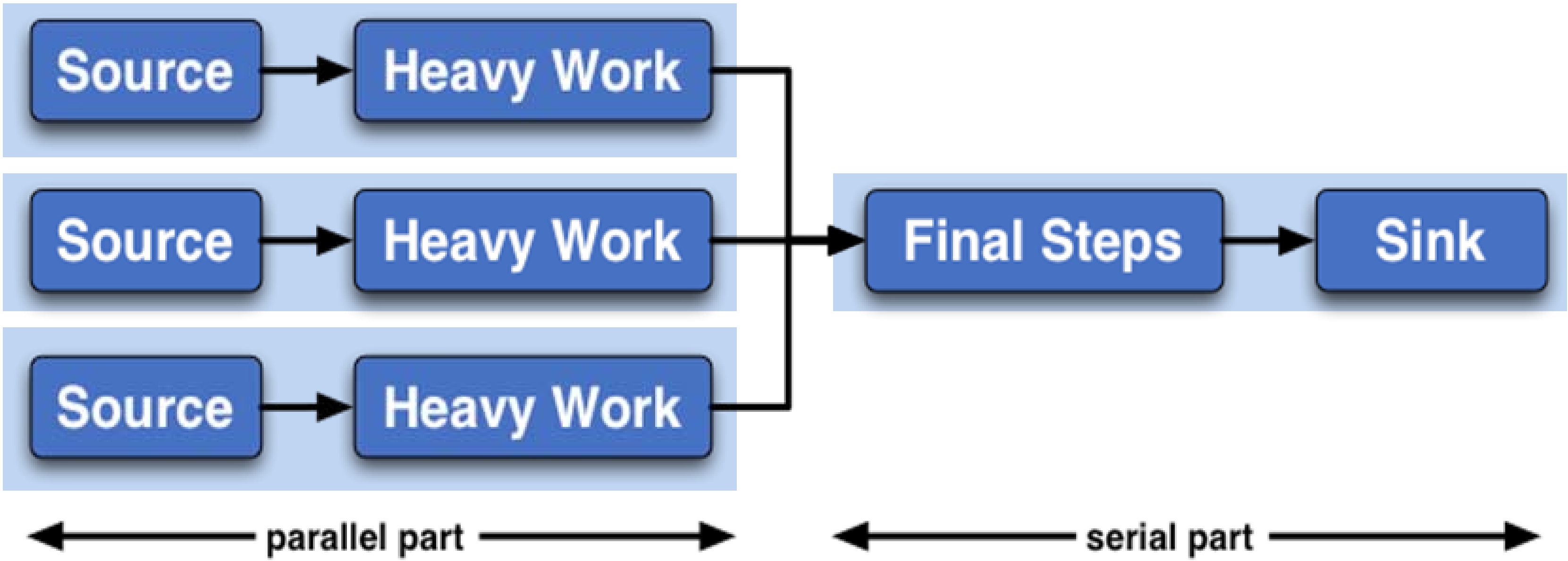
Pipeline Distribution
Data Parallelization

- Several parallelization approaches possible
- By nature of the post-processing pipeline, heavy work is usually located at the beginning of the pipeline
  - Data parallelization is the most efficient approach in most cases
  - Communication costs are reduced to a minimum



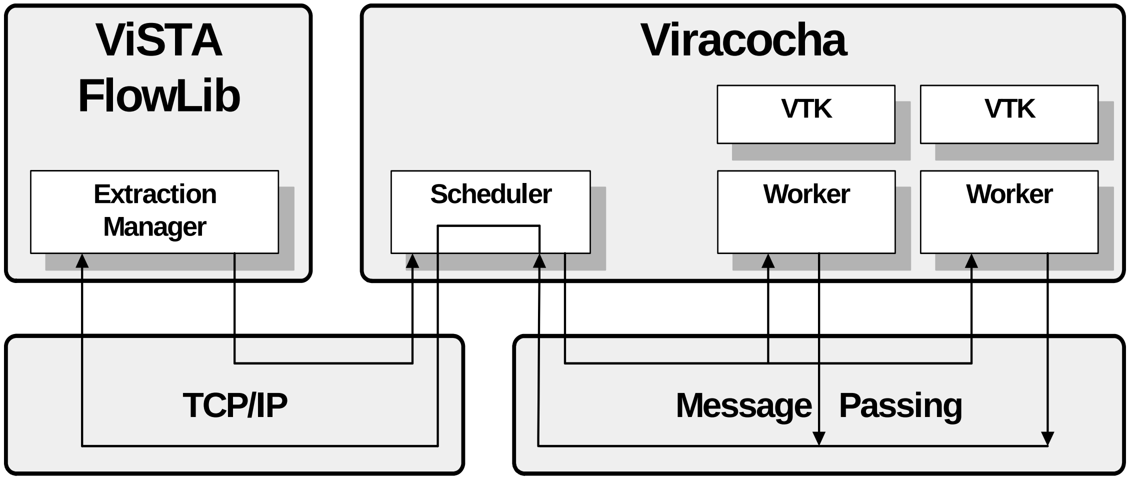
Pipeline Distribution
Data Parallelization

- Determine heavy work and parallelize these pipeline sections
- Split data and allocate sections to parallel processes
- Combine partial results and execute remaining steps



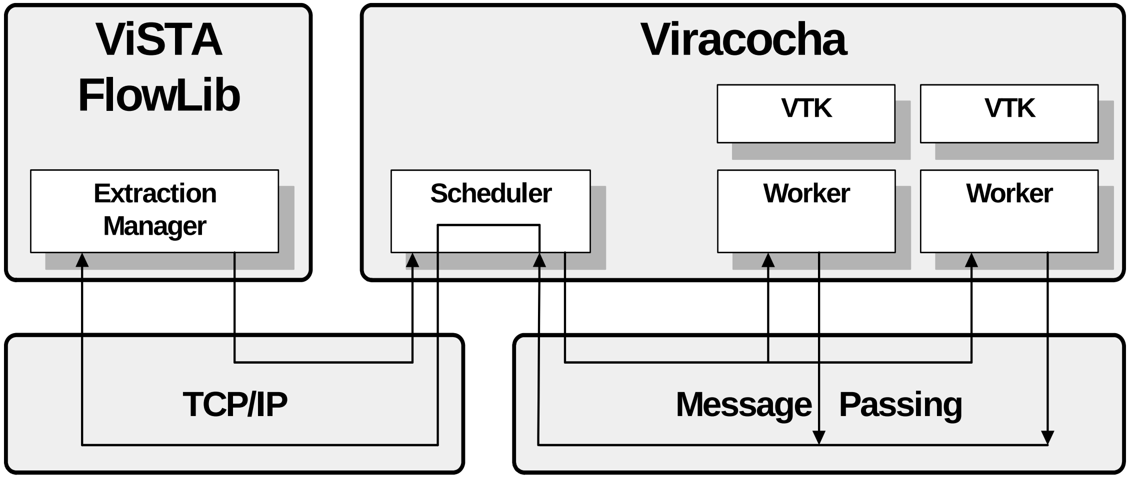
Pipeline Distribution
Distributed Post-Processing Framework

- Distributed post-processing framework Viracocha
  - Uses data parallelization approach
  - Has been developed in co-operation with the VR Group at Aachen University (RWTH)
  - Message-passing best choice for parallel post-processing
  - TCP/IP for communication in heterogeneous hardware systems



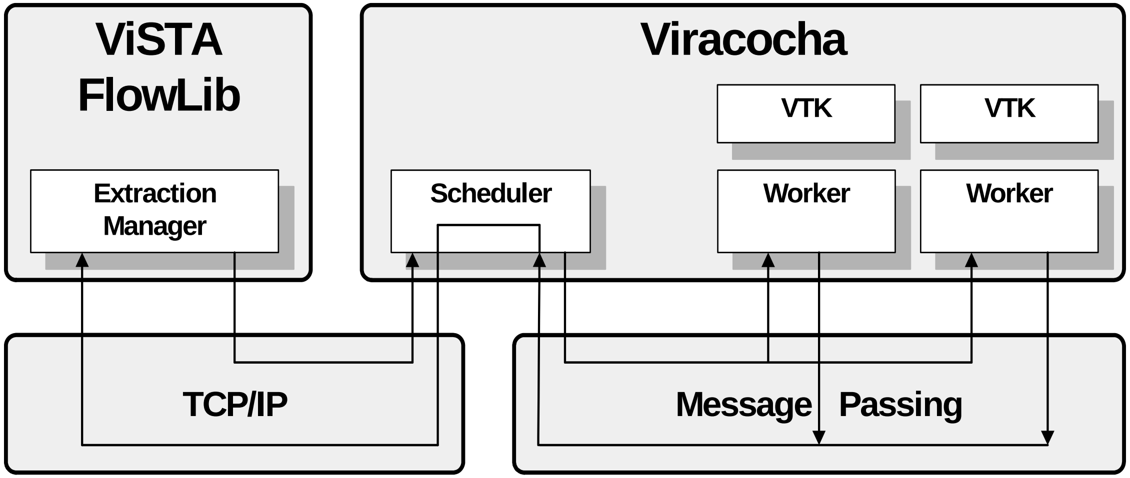
Pipeline Distribution
Distributed Post-Processing Framework

- Generic Viracocha layer architecture
  - Application-independent
  - Easy integration of new functions
  - Parallel concept (scheduler, workers) stays untouched
  - Visualization Toolkit (VTK) is used for most of the basic post-processing algorithms
  - Algorithms from own research added to Toolkits



Pipeline Distribution
Distributed Post-Processing Framework

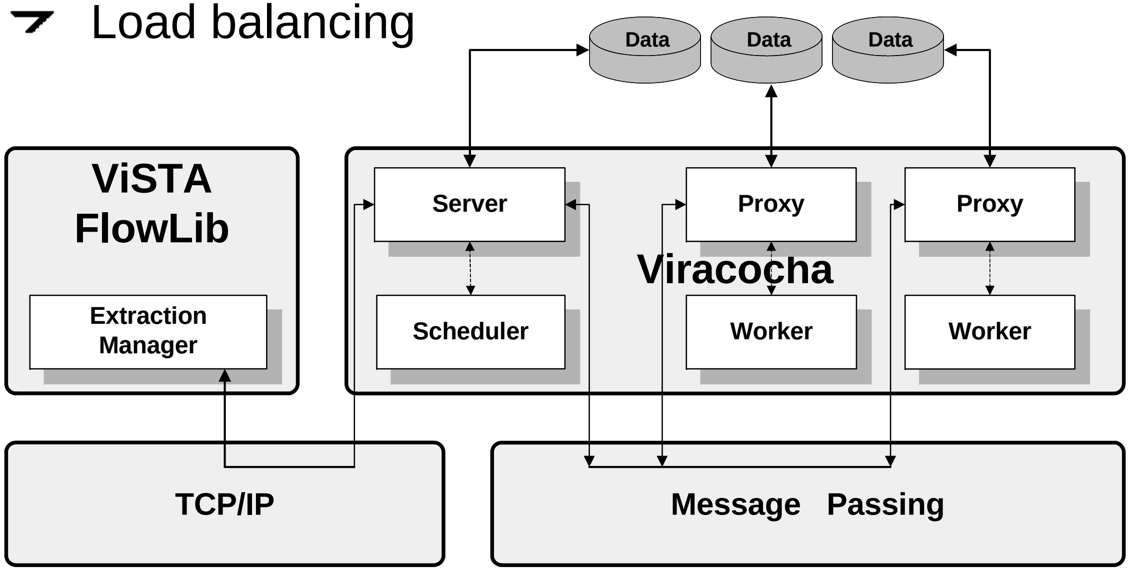
- ViSTA FlowLib is used as frontend
  - Optimized for visualization of unsteady simulation data
  - Interactive visualization using Virtual Reality technologies
  - Real-time interaction by decoupling heavy data handling and post-processing work
  - More features can be requested from backend without disrupting interactive exploration



Large-scale Dataset I/O

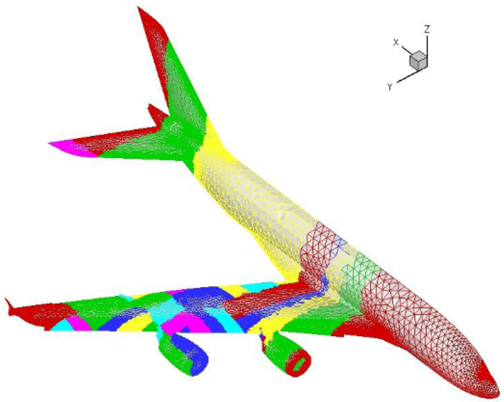
Large-scale Dataset I/O
Viracocha Data Management System

- Concurrent data management handles I/O load
  - Caching (primary / secondary)
  - Prefetching
  - Optimized loading strategies
  - Data services
    - Ask scheduler for next data
    - Load balancing

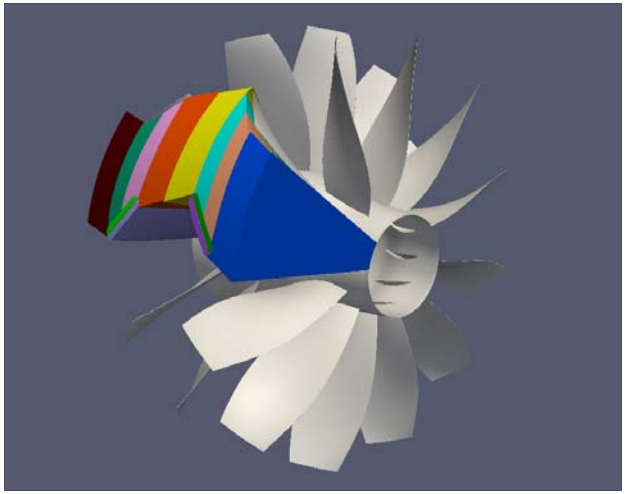


Large-scale Dataset I/O
Domain Decomposition

- Multi-block datasets
  - Can easily be distributed among several processes
  - One block per time level = one file on disk
  - Sufficient for cell-based (local) algorithms
- Unstructured datasets
  - Make use of TAU's partitioner
  - Online Monitoring: no I/O required



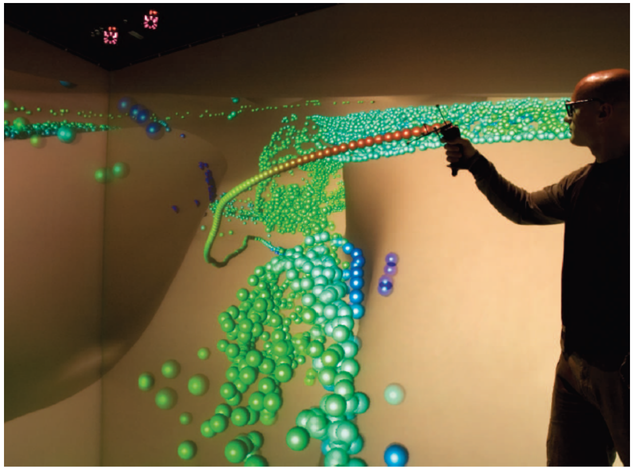
Domain decomposition for a simulation using TAU



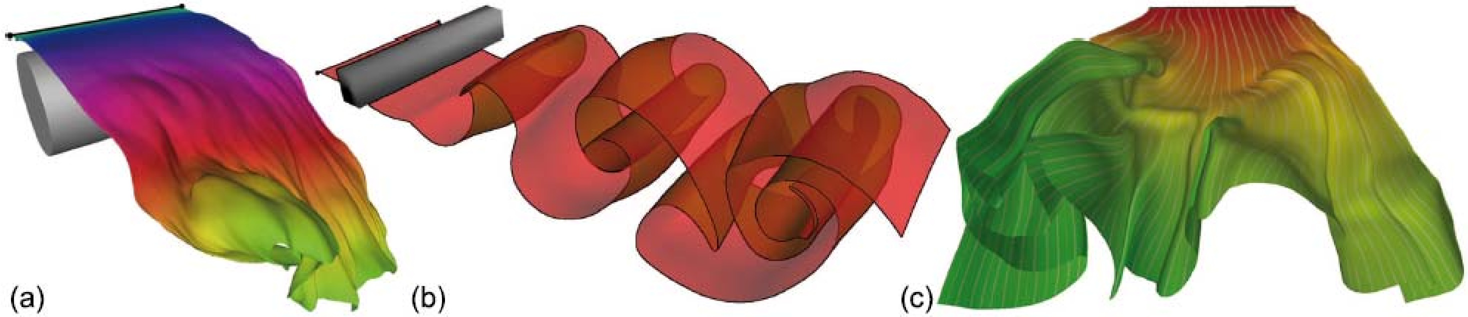
Blocks of a *Propfan* dataset, block-wise colored

Large-scale Dataset I/O
Loading on Demand

- Global post-processing algorithms
  - E.g. particle integration
  - Only block containing current particle position needs to be loaded into main memory
  - Metadata about block connectivity helps to identify next block when particle leaves current block
  - Load on demand considerably reduces I/O load



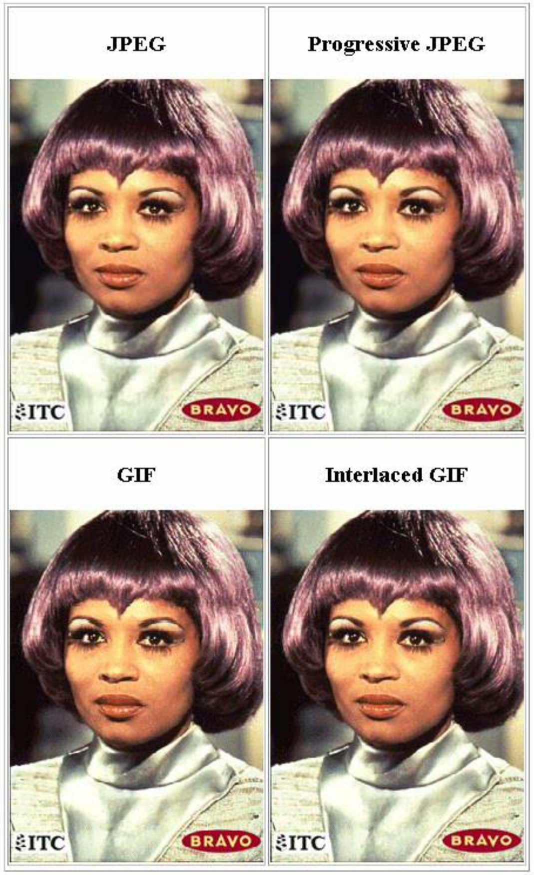
Interactive particle seeding in an immersive environment



Multi-Resolution and Data Streaming

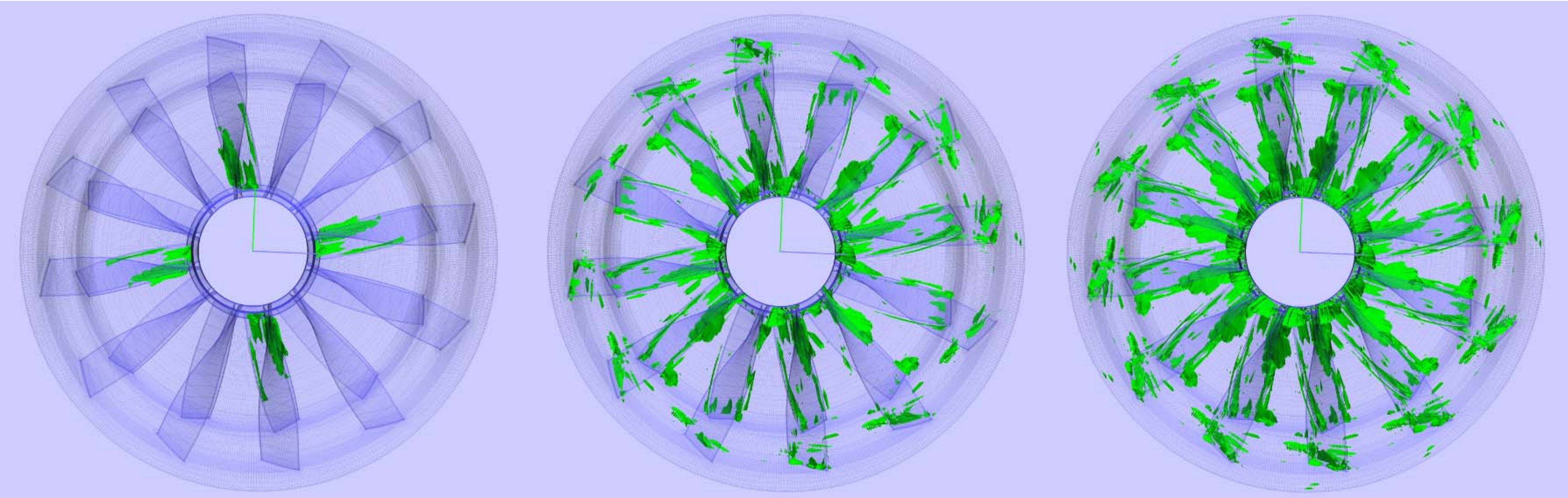
Multi-Resolution and Data Streaming
Data Streaming

- Decouple frontend and backend
  - Improvement of interactivity within virtual environments
- However: Extraction still needs time
- Solution: Data streaming
  - Motivation: Progressive JPG >>>
    - Immediate feedback
    - Fast impression of the final result
  - Exploit first approximate results
    - Quick start of data exploration
    - Possible to abort running comp. and restart it with new parameters
- Problem
  - Additional communication and more computation time
  - Appropriate CFD algorithms
  - Progressive approaches hardly available



Source: http://wp.netscape.com/eng/mozilla/2.0/relnotes/demo/pjpegdemo.html

Multi-Resolution and Data Streaming
Data Streaming

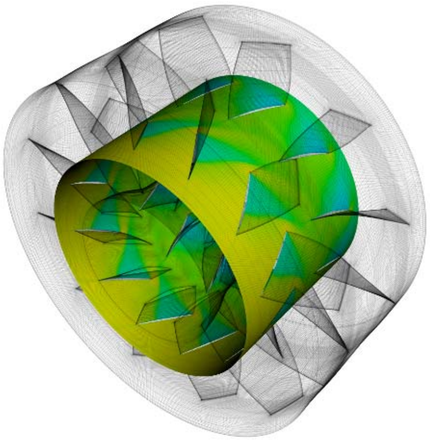


Multiple streaming steps of Lambda-2 vortex extraction computed by 4 workers and then streamed independently to the visualization frontend

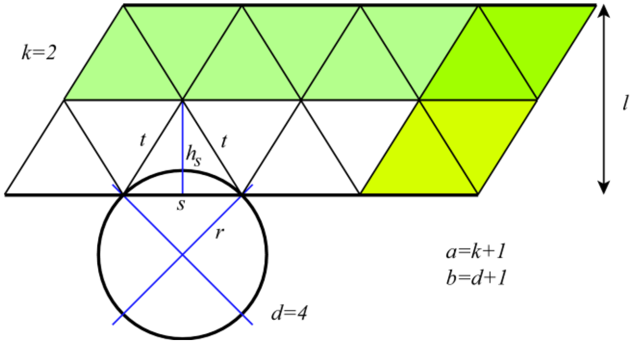
Multi-Resolution and Data Streaming

Progressive Streaming

- Analytical cut function
  - Intersections on cell edges have to be found
  - Geometry of cut mesh unknown in advance
  - Fast



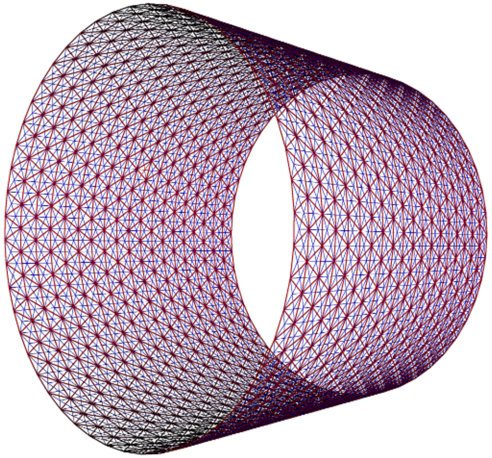
- Sampling
  - Geometry is already known everywhere in advance
  - Cut surface can be visualized immediately
    ➔ Interactivity!
  - Only scalars sampled at vertices are transmitted
    - Computed in parallel on backend



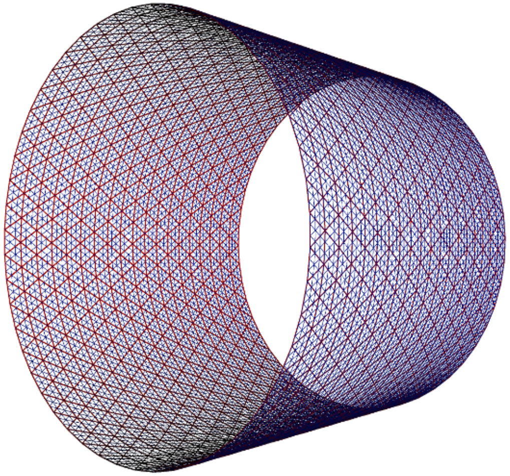
Triangulation parameters of a cut cylinder (right), resulting cut through the propfan (top)

Multi-Resolution and Data Streaming
Progressive Streaming

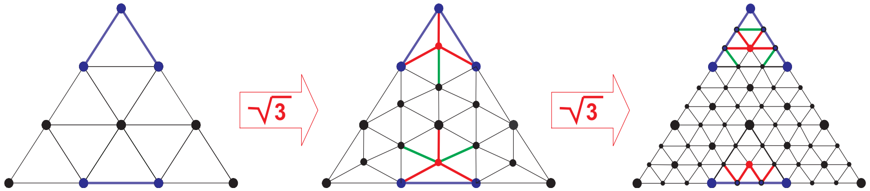
- Progressive approach
  - Realized as surface sub-division
  - Dyadic split: classical approach for triangle meshes
  - Sqrt-3 refinement
    - Fewer points to add
    - Slower refinement
    - Smoother streaming



Sqrt-3 refinement (Level 1)



Sqrt-3 refinement (Level 2), initial grid (red)

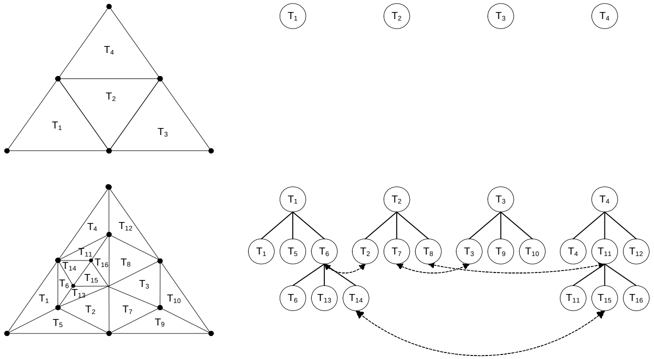


Sqrt-3 refinement steps with special boundary treatment

Multi-Resolution and Data Streaming
Multi-Resolution Geometry

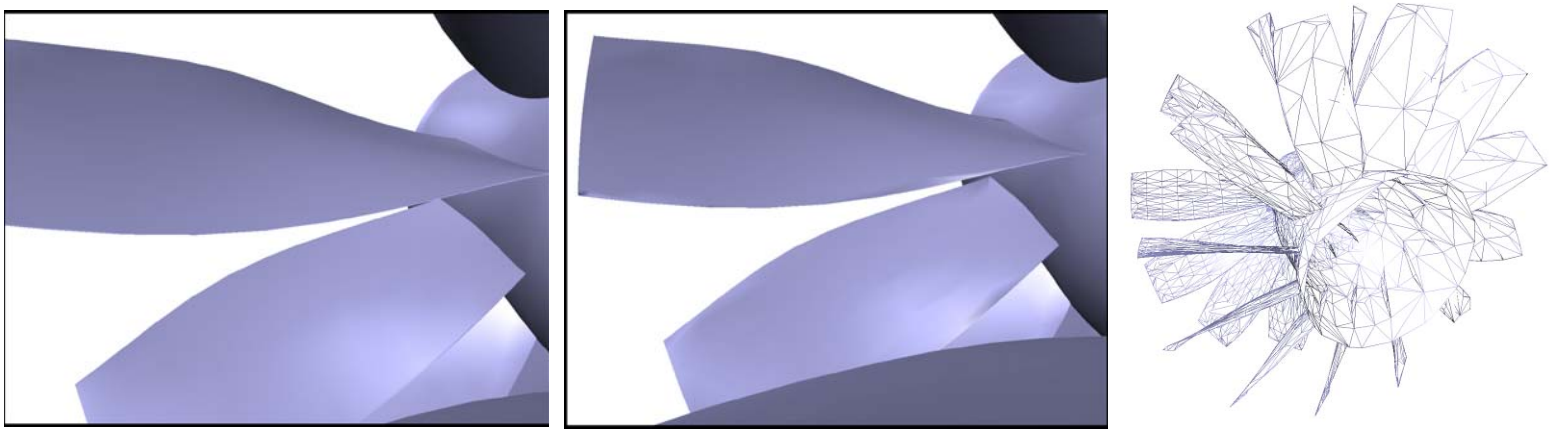
- Adaptive rendering
  - Cache multi-level streaming data for Level-of-Detail switch
  - Increase frame rate by using different LODs selected by priorities
    - E.g. context geometry vs. extraction data
  - Handled by multi-resolution manager

- Selective refinement
  - Get viewpoint of the user
  - Close parts of objects are rendered with higher resolution
  - Requires adaptive meshes
    - Main advantage of Sqrt-3 data structure
  - Approach can also be applied for polylines (e.g. particle trajectories)



Sqrt-3 selective refinement

Multi-Resolution and Data Streaming
Multi-Resolution Geometry



View-dependent optimized view of propfan blades, original with 340,000 triangles (left), view-dependent optimized with 8,532 triangles (mid), and wireframe presentation of the entire view-dependent optimized model (right)

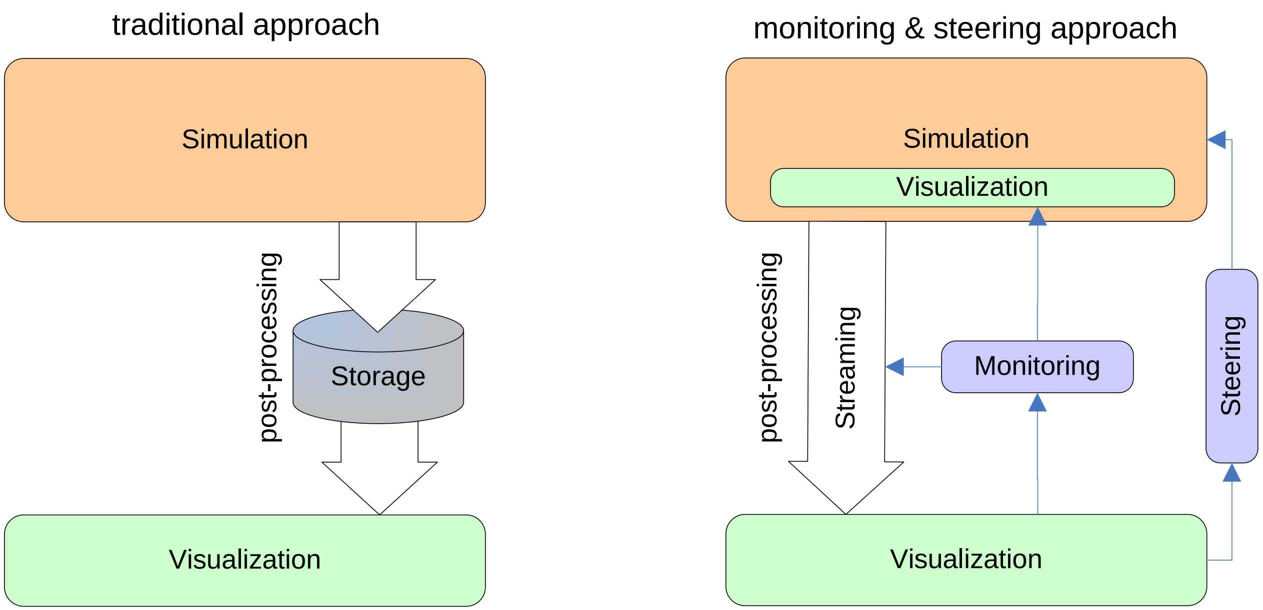
Co-Execution of Simulation and Post-Processing

Co-Execution of Simulation and Post-Processing
Online Monitoring and Computational Steering

- Goal: Interact with the simulation during the run

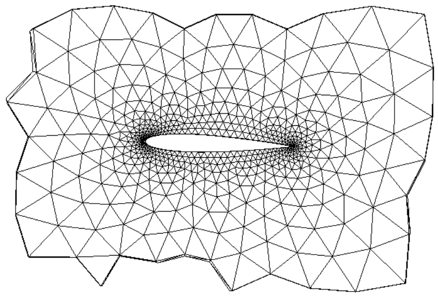
- Approach:
  - Interactive post-processing
  - Integrate steering module
  - Attach streaming visualization pipeline

- Challenges:
  - Steer simulation and maintain data coherence
  - Integration of interactive post-processing

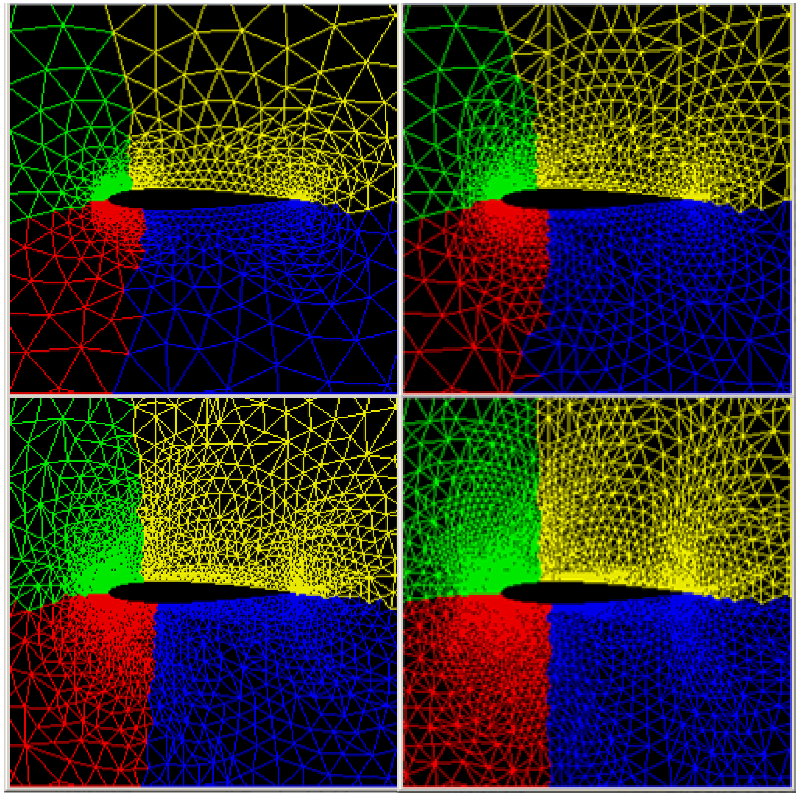


Co-Execution of Simulation and Post-Processing
Online Monitoring and Computational Steering

- First results of prototype
  - Mesh adaption while simulation runs
  - Processing done in parallel (shown by color)



original mesh



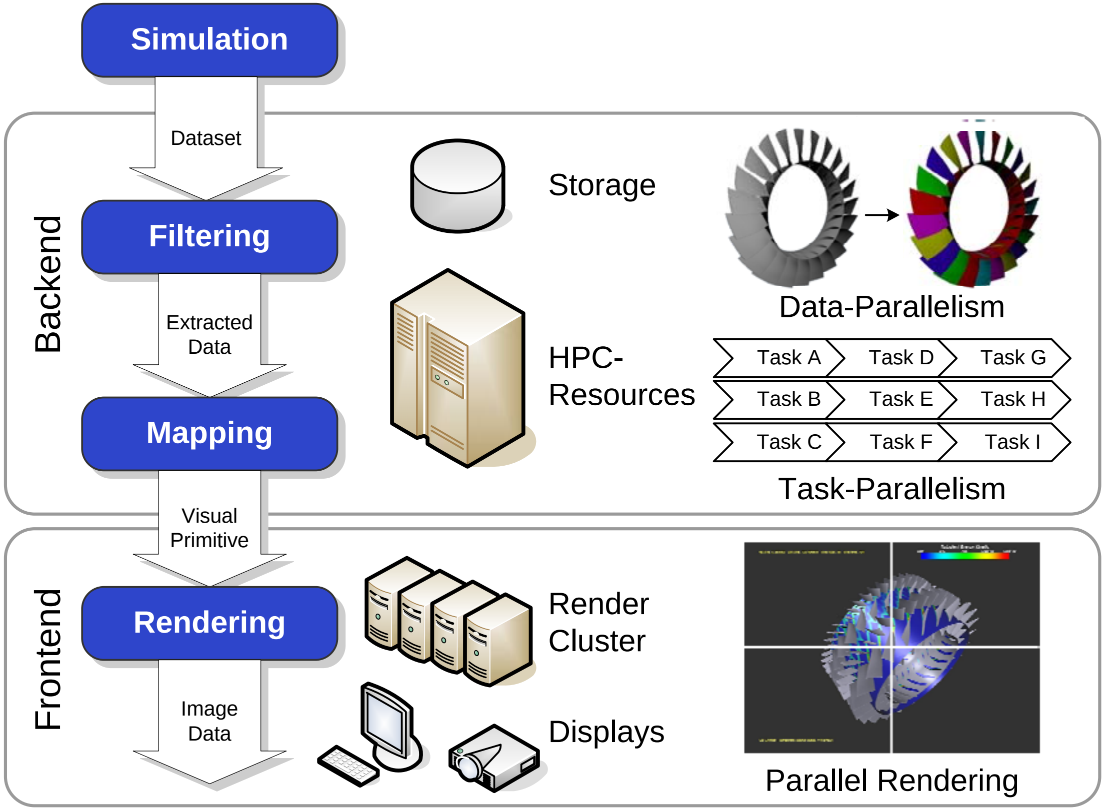
refined mesh

Co-Execution of Simulation and Post-Processing

Distributed Post-Processing Infrastructure

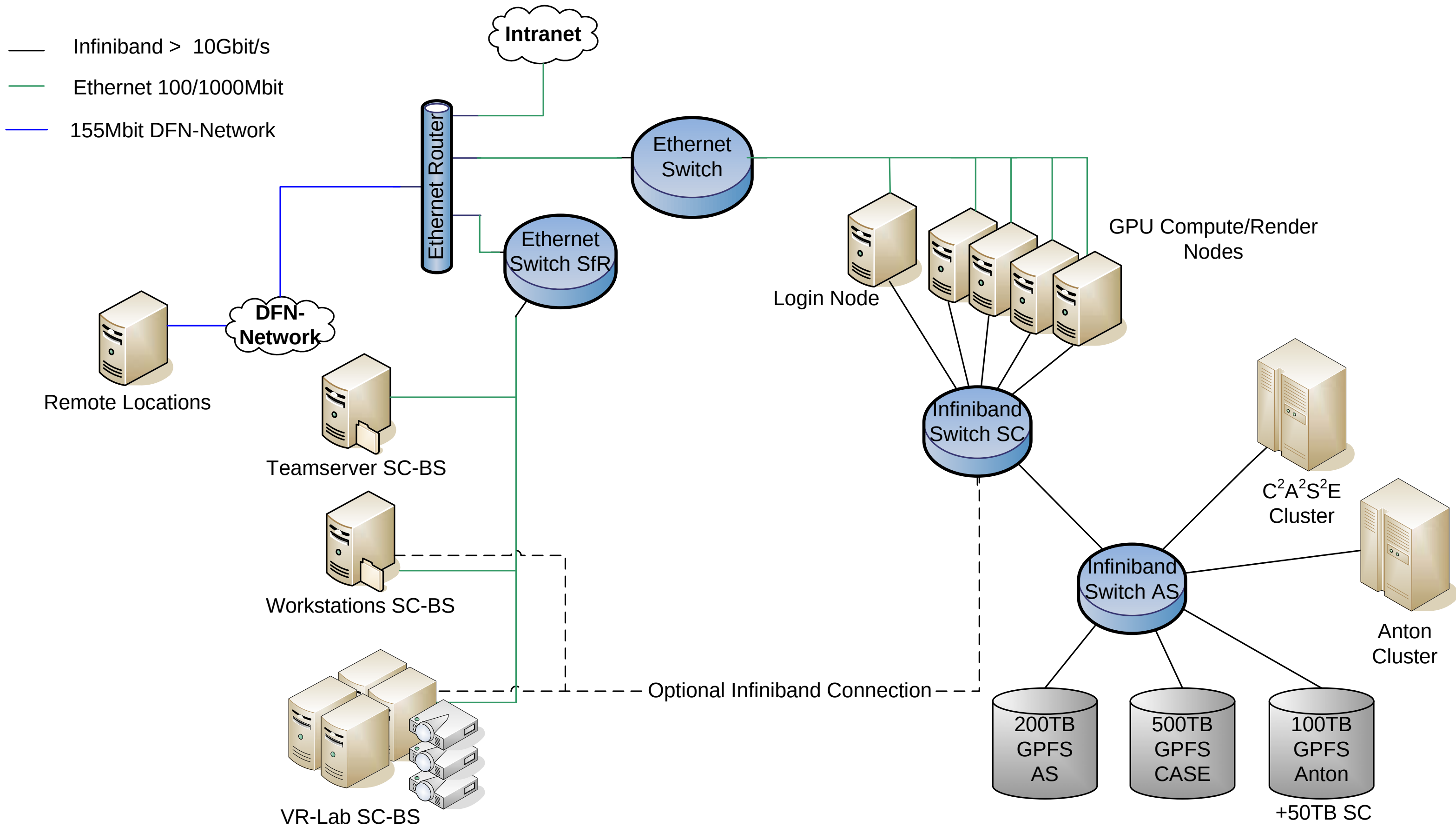
Goals:

- Exploit parallelism using various techniques and resources
- Reduce amount of data communicated across pipeline
- Rendering in parallel on GPGPU cluster (integrated in HPC environment)
- “Light-weight” frontend sufficient



Co-Execution of Simulation and Post-Processing
Distributed Post-Processing Infrastructure

- Planned Infrastructure at DLR:



Summary

- Growing importance of visualization in large-scale numerical simulation
- Today VR techniques mainly used for „pretty pictures“
- Fast processing of large datasets required for interactive visualization
  ➡ HPC (parallel execution) for first stages in visualization pipeline
- Close integration of application and postprocessing requires dedicated hardware / software infrastructure
- Pure batch execution model for HPC no longer feasible