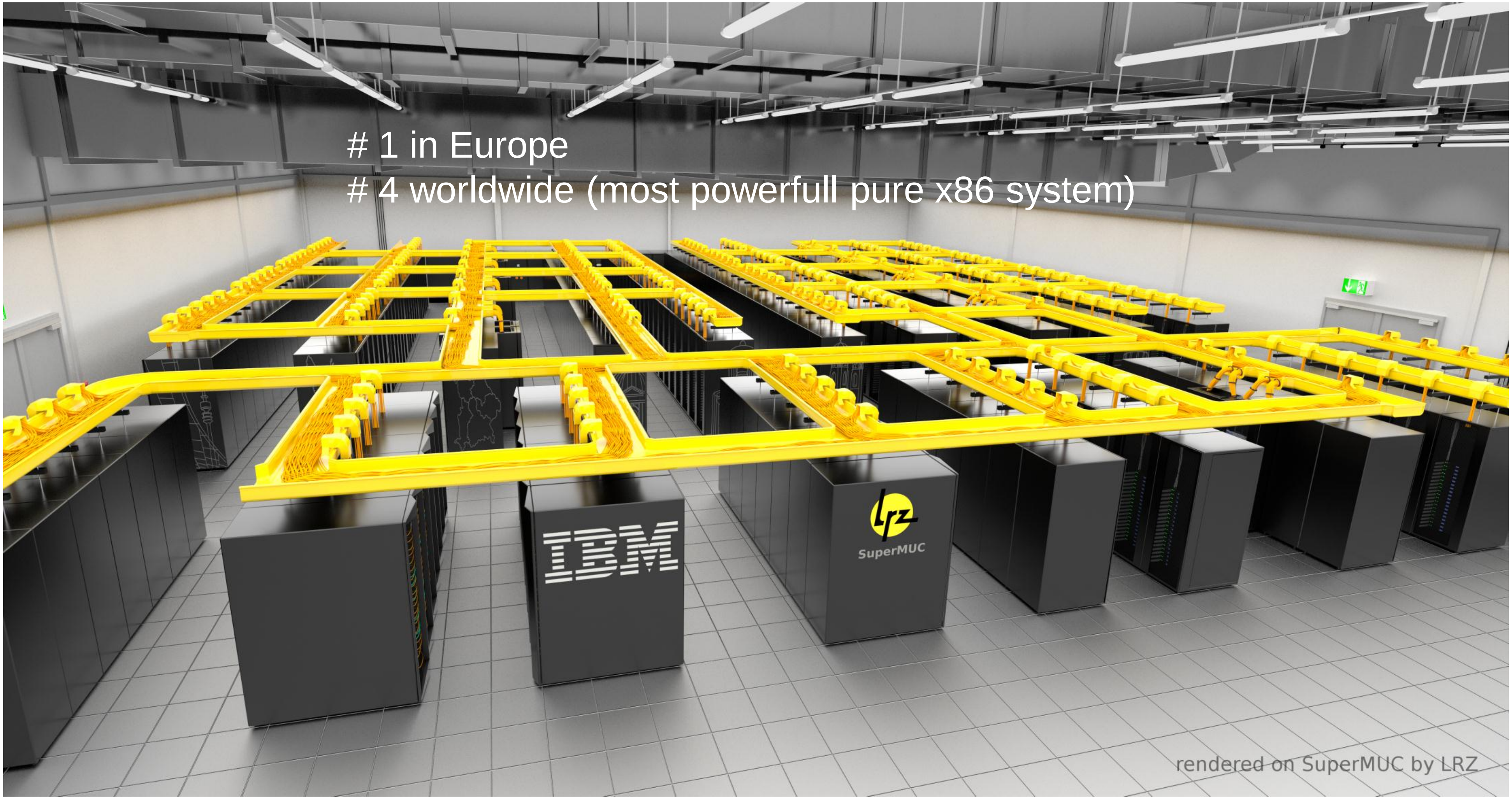
Energy-efficient High Performance Computing with SuperMUC

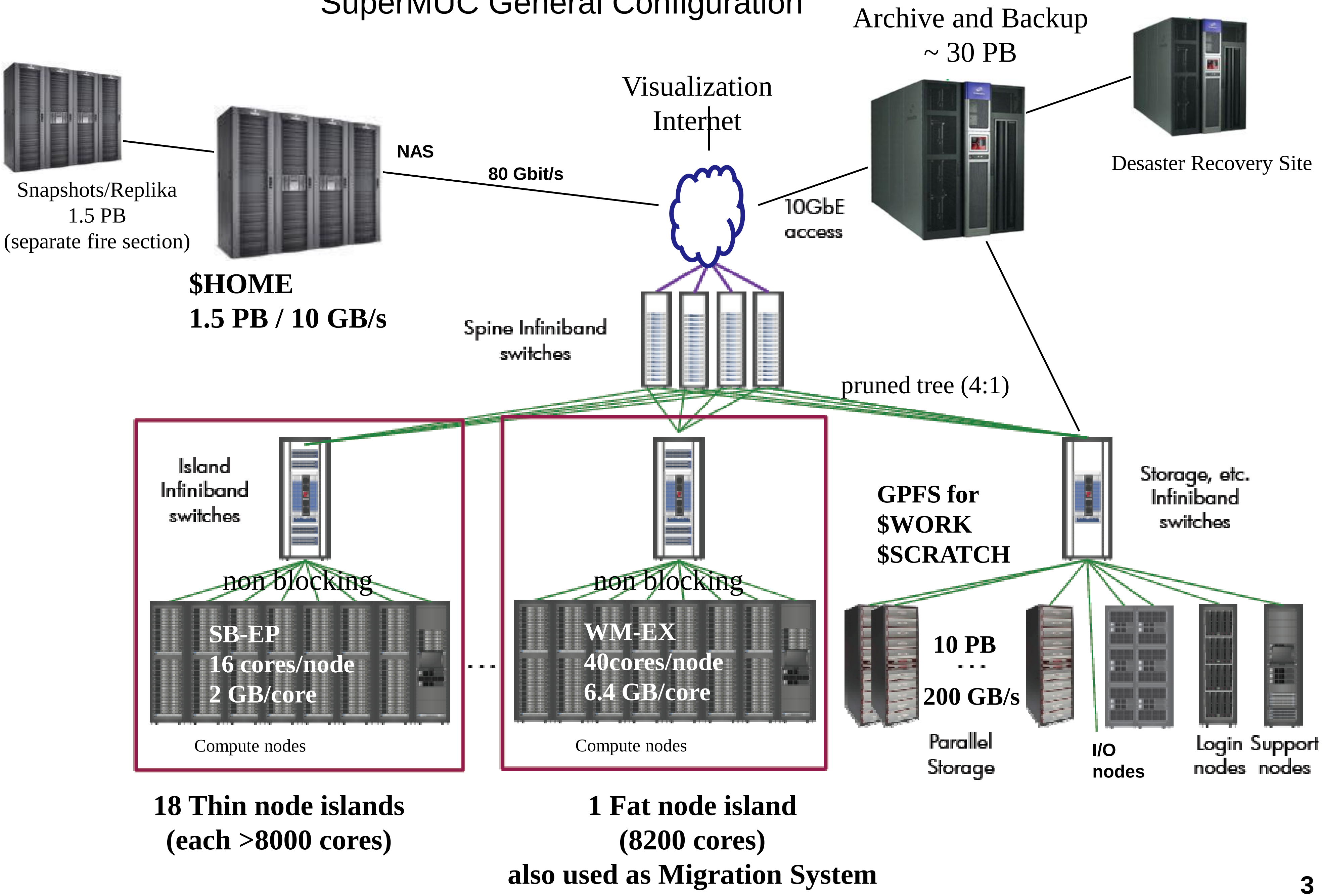
By Ernst A. Graf

Arndt Bode
Chairman of the Board of the Leibniz Supercomputing Centre
and Technische Universität München

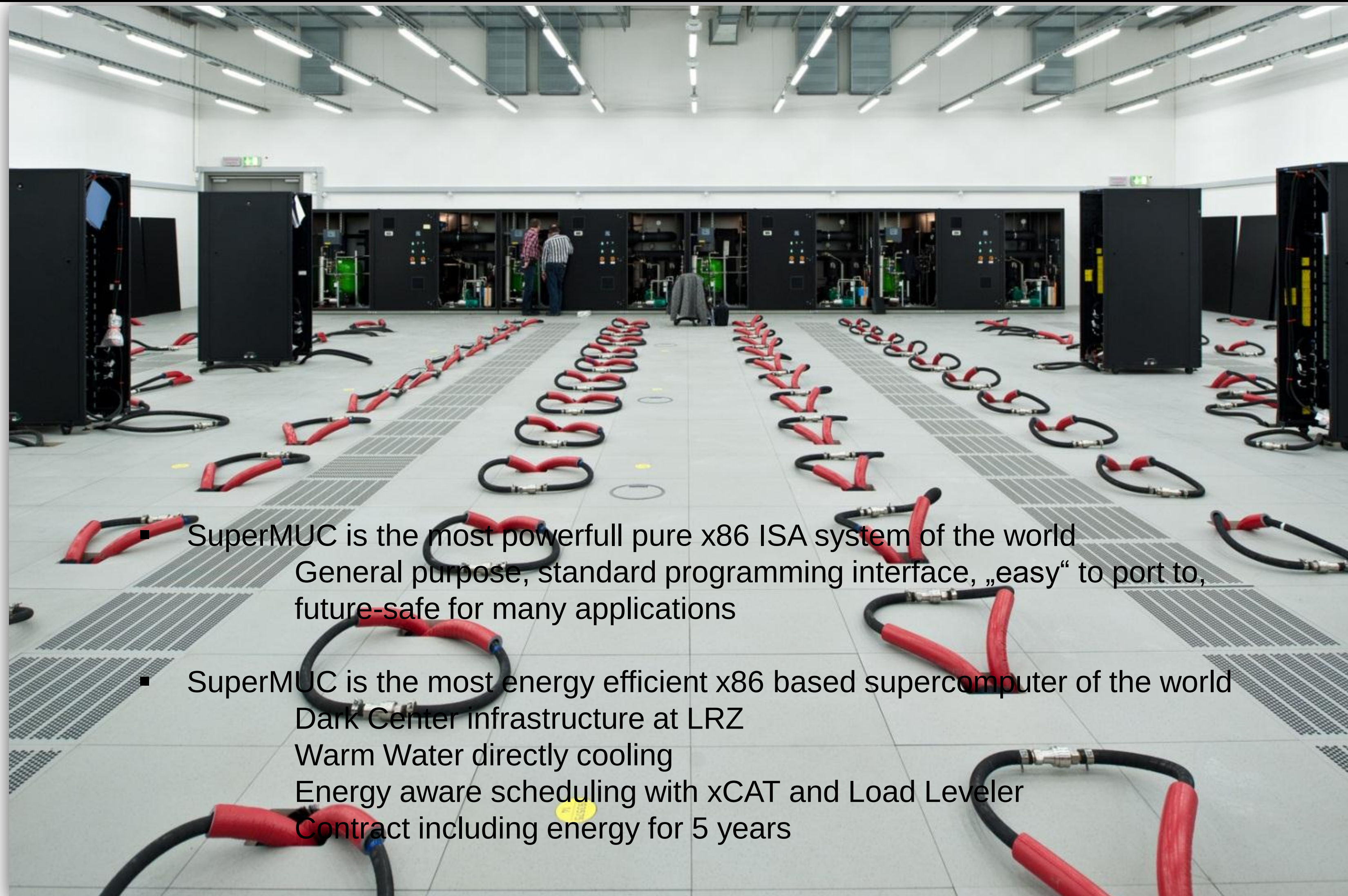
SuperMUC, TOP 500 39th Release, Hamburg June 2012



SuperMUC General Configuration



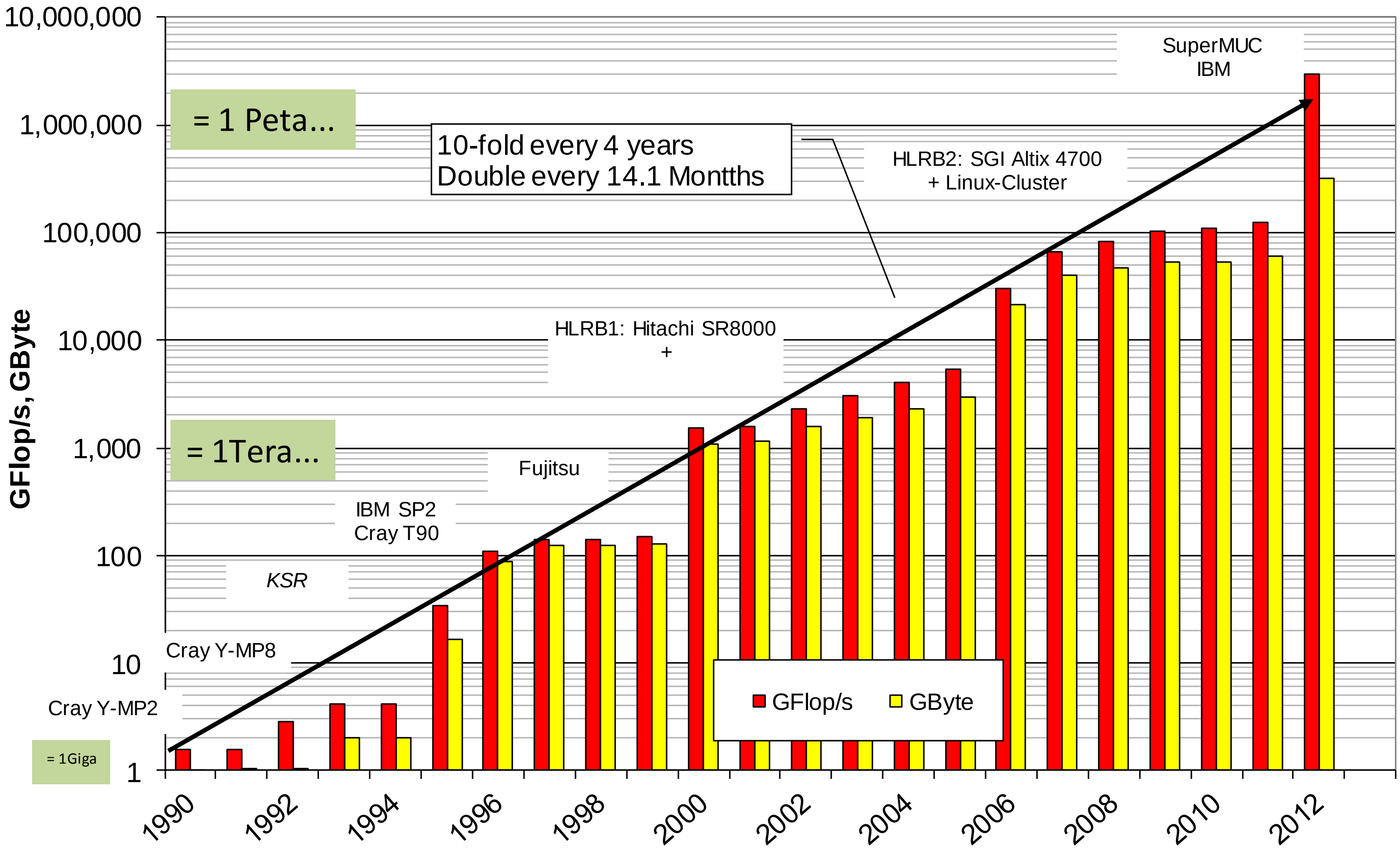
What‘s special about SuperMUC



- SuperMUC is the most powerfull pure x86 ISA system of the world
  - General purpose, standard programming interface, „easy“ to port to, future-safe for many applications

- SuperMUC is the most energy efficient x86 based supercomputer of the world
  - Dark Center infrastructure at LRZ
  - Warm Water directly cooling
  - Energy aware scheduling with xCAT and Load Leveler
  - Contract including energy for 5 years

Supercomputer Architectures at LRZ since 1990



Energy Efficiency and SuperMUC

Motivation: we pay 15.8 Cents per Kwh (regular contract, expect increase)

Energy Efficient HPC

- Reduce the power losses in the power supply chain
- Exploit your possibilities for using compressor-less cooling und use energy-efficient cooling technologies (e.g. direct liquid cooling)
- Re-use waste heat of IT systems

Energy efficient hardware

- Use newest semiconductor technology
- Use of energy saving processor and memory technologies
- Consider using special hardware or accelerators tailored for solving specific scientific problems or numerical algorithms

Energy efficient infrastructure

- Monitor the energy consumption of the compute systems and the cooling infrastructure
- Use energy aware system software to exploit the energy saving features of your target platform
- Monitor and optimize the performance of your scientific applications

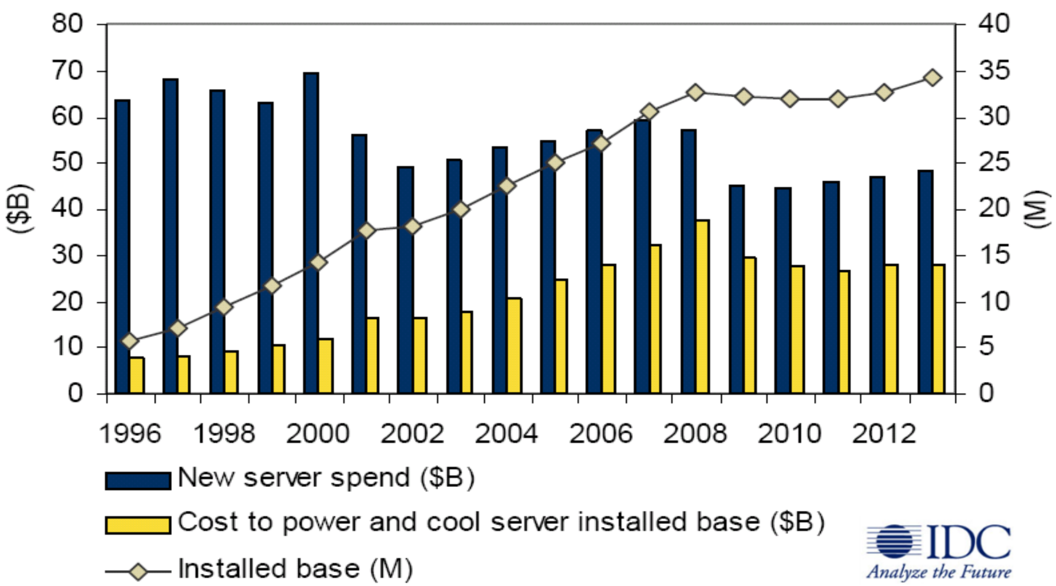
Energy aware software environment

- Use most efficient algorithms
- Use best libraries
- Use most efficient programming paradigm

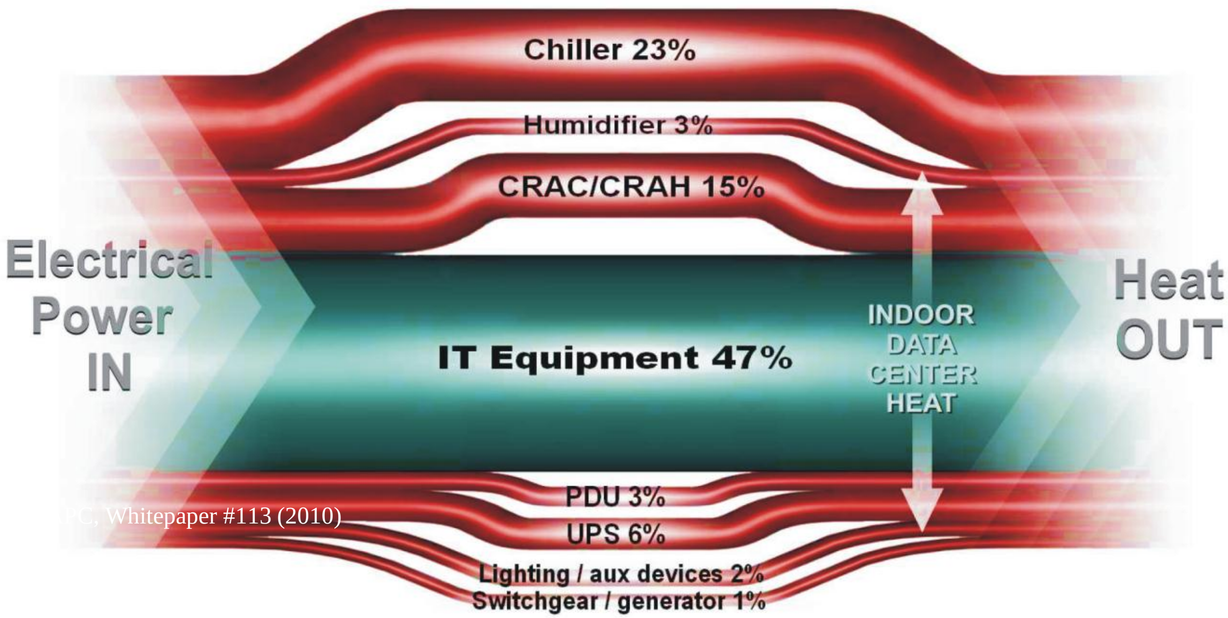
Energy efficient applications

Data Center Market Drivers and Trends

- **Total cost of ownership and environmental footprint**
- **Servers used 330 TWh of electrical energy: $25bn or 2% of the electricity production (2009).**



Energy Consumption of Datacenter



- **Air-cooled datacenters are inefficient.**
  Typical cooling needs as much energy as IT equipment and both are thrown-away.
- **Provocative: datacenter is a huge "heater with integrated logic."**

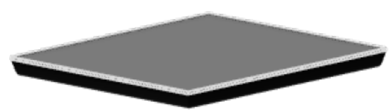
Comparison: Water Cooling vs. Air Cooling

Water

1. High heat capacity

$$c_V \approx 1 \text{ Wh/(L·K)}$$

2. Low thermal resistance



$$\dot{q}'' = R_{th} \cdot \Delta T$$
$$R_{th} = 0.1 \text{ K cm}^2/\text{W}$$
$$\dot{q}'' = 50-100 \text{ W/cm}^2$$
$$\Delta T = 5-10^{\circ}\text{C}$$

Air

1. Low heat capacity

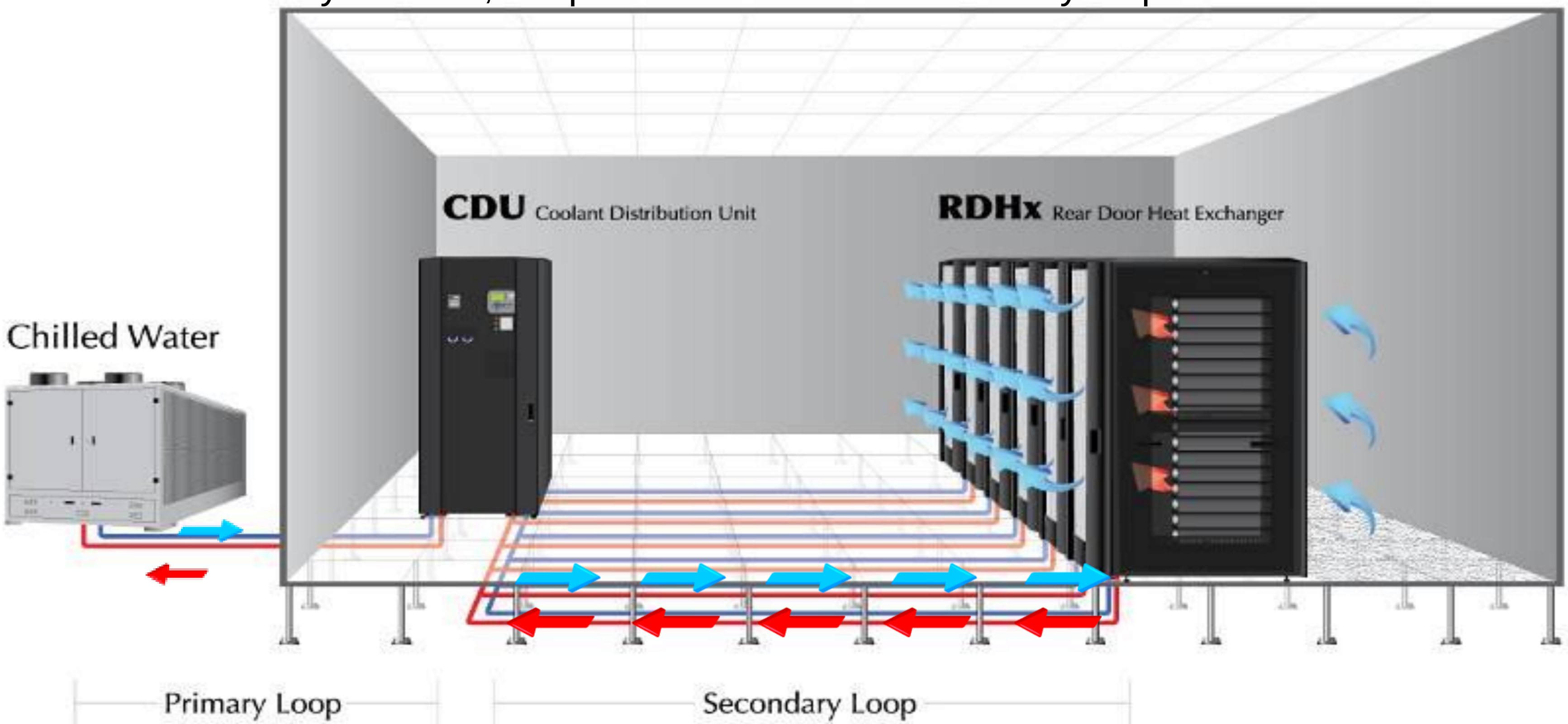
$$c_V \approx 0.0003 \text{ Wh/(L·K)}$$

2. High thermal resistance

$$\dot{q}'' = R_{th} \cdot \Delta T$$
$$R_{th} = 1 \text{ K cm}^2/\text{W}$$
$$\dot{q}'' = 50-100 \text{ W/cm}^2$$
$$\Delta T = 50-100^{\circ}\text{C}$$

IBM Rear Door Heat Exchangers

- Passive Rear Door Heat Exchanger (RDHx) provides up to 100% cooling
  - No condensation, no need for reheat or humidification
  - No moving parts
- CDU creates fully isolated, temperature controlled secondary loop

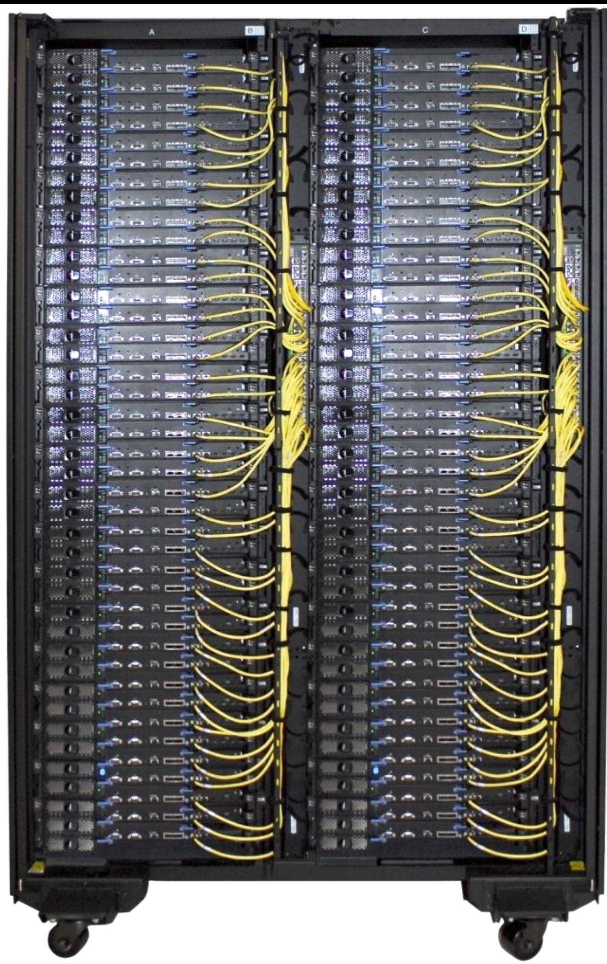


Direct Water Cooling

IBM iDataplex dx360 M4

- **Heat flux > 90% to water; very low chilled water requirement**
- **Power advantage over air-cooled node:**
  - Warm water cooled ~10%
    (cold water cooled ~15%)
  - due to lower $T_{components}$ and no fans
- **Typical operating conditions: T_{air} = 25 – 35°C, T_{water} = 18 – 45°C**

IBM System x iDataPlex Direct Water Cooled Rack



iDataplex DWC Rack w/ water cooled nodes
(front view)

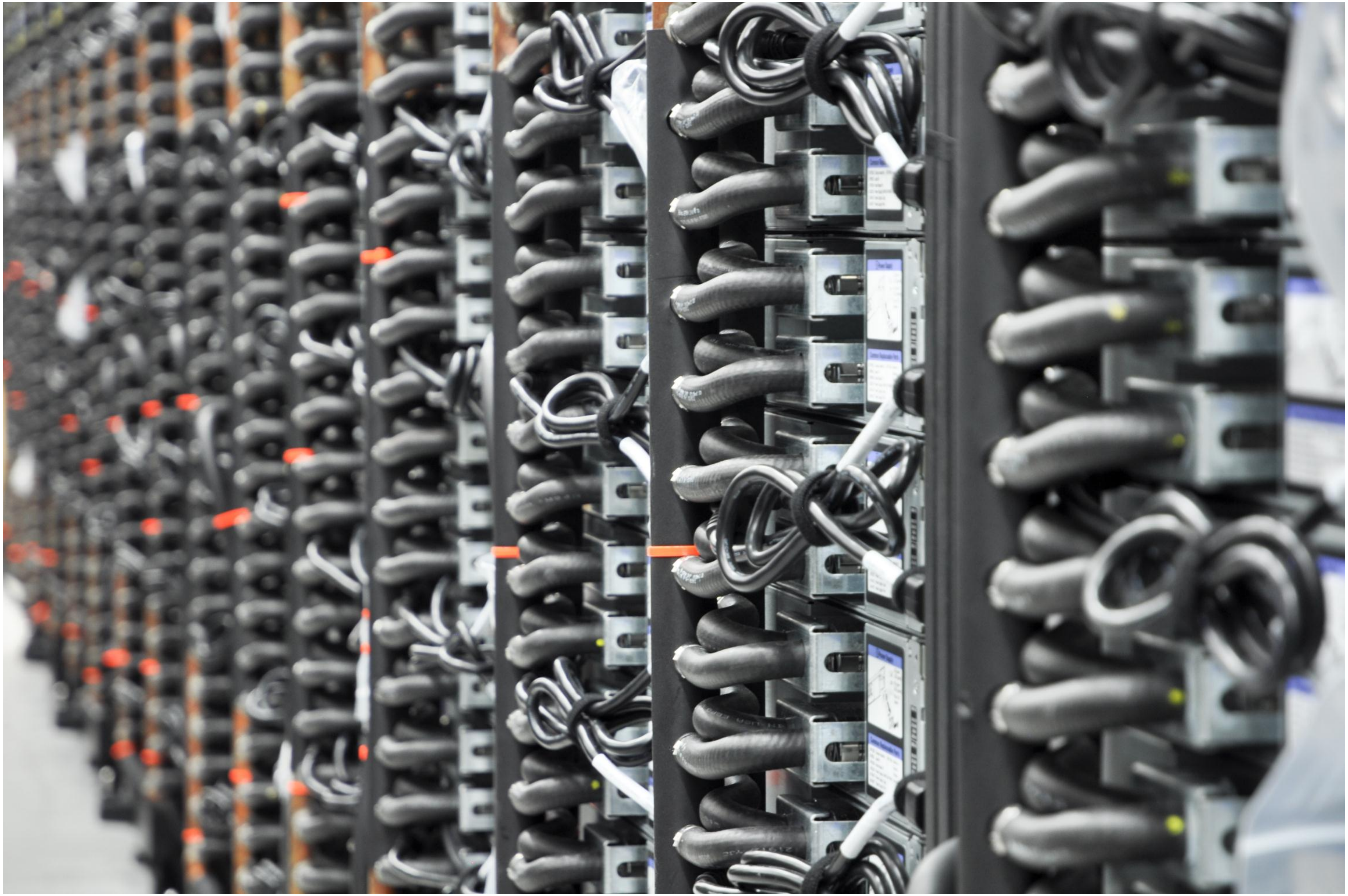
iDataplex DWC Rack w/ water cooled nodes
(rear view of water manifolds)

Hybrid Datacenter w/ Direct Water Cooled Nodes

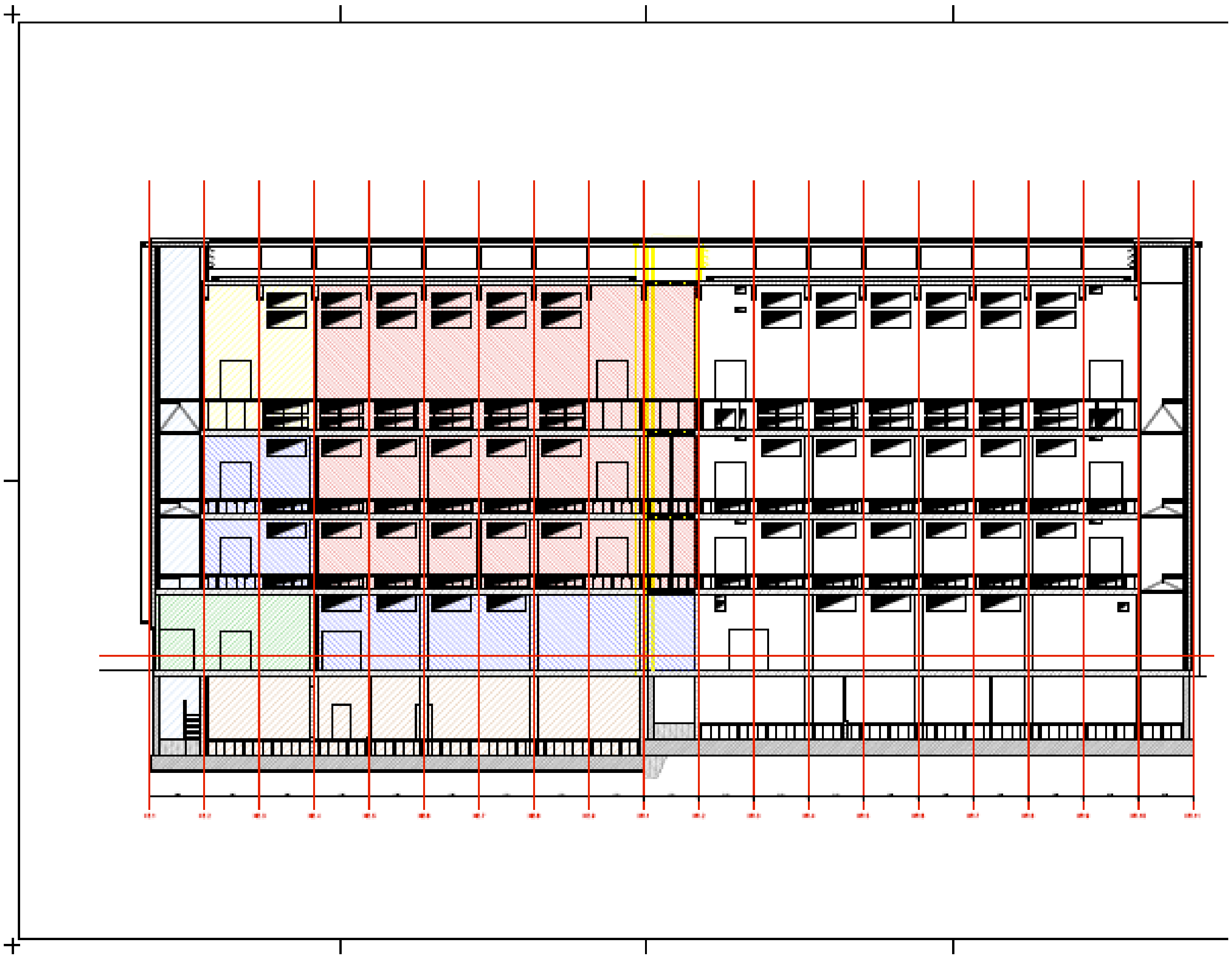
- **Highly energy-efficient hybrid-cooling solution:**
  - Compute racks
    - 90% Heat flux to warm water
    - 10% Heat flux to CRAH
  - Switch / Storage racks
    - Rear door heat exchangers

- **Compute node power consumption reduced ~ 10% due to lower component temperatures and no fans.**

- **Power Usage Effectiveness P_{Total} / P_{IT}: PUE ~ 1.1**



From Computer Cube to Computer Cuboid
Nearly double floor space



Supercomputer

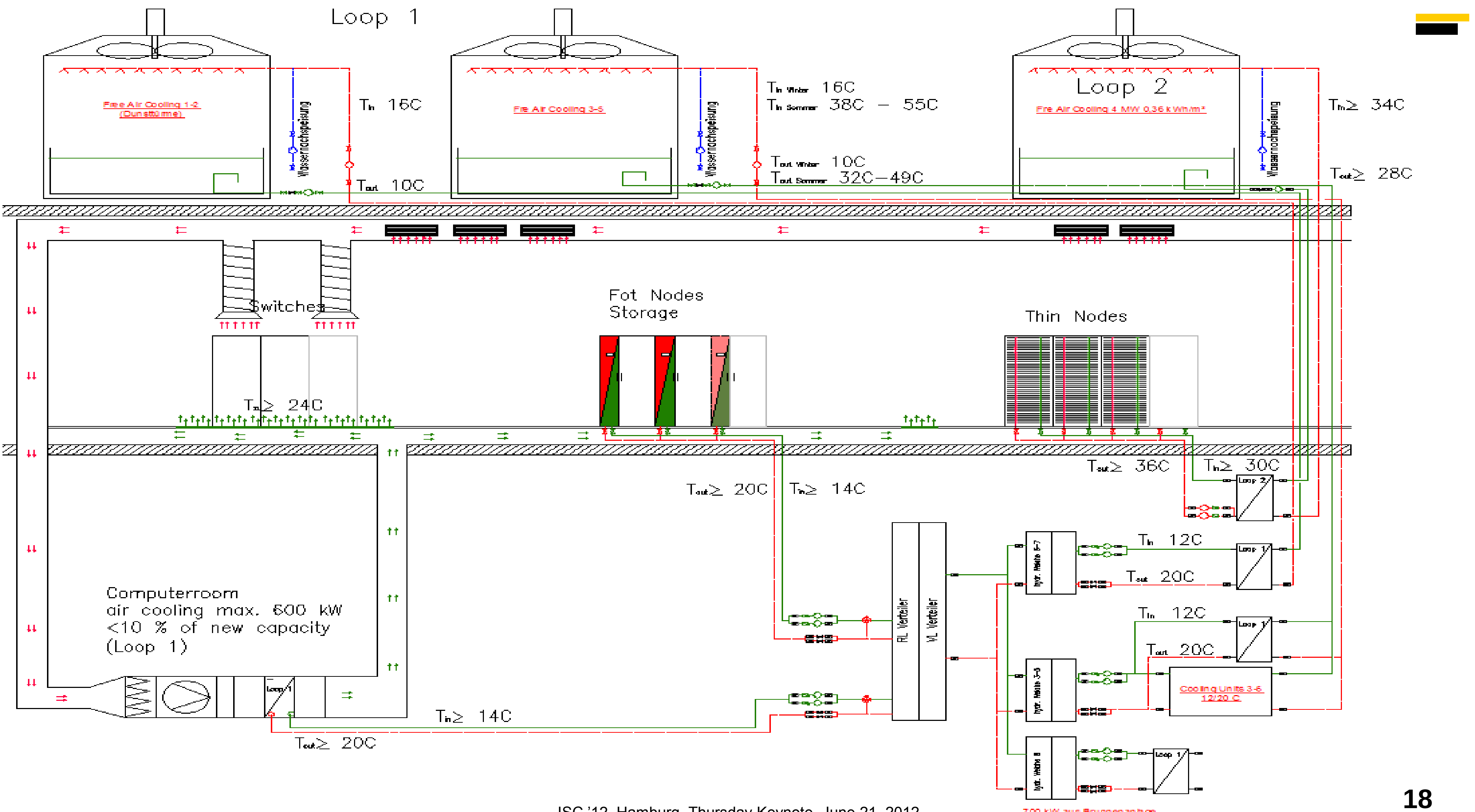
Linux-Cluster,
Hosting/Housing, Servers

(Tape) Archive/Backup, Disks

Cooling, Water Air Processing

Power, Transformers, UPS,

New Cooling Concept: Dedicated Free Cooling Loop



Infrastructure 3rd floor

Infrastructure 3rd floor

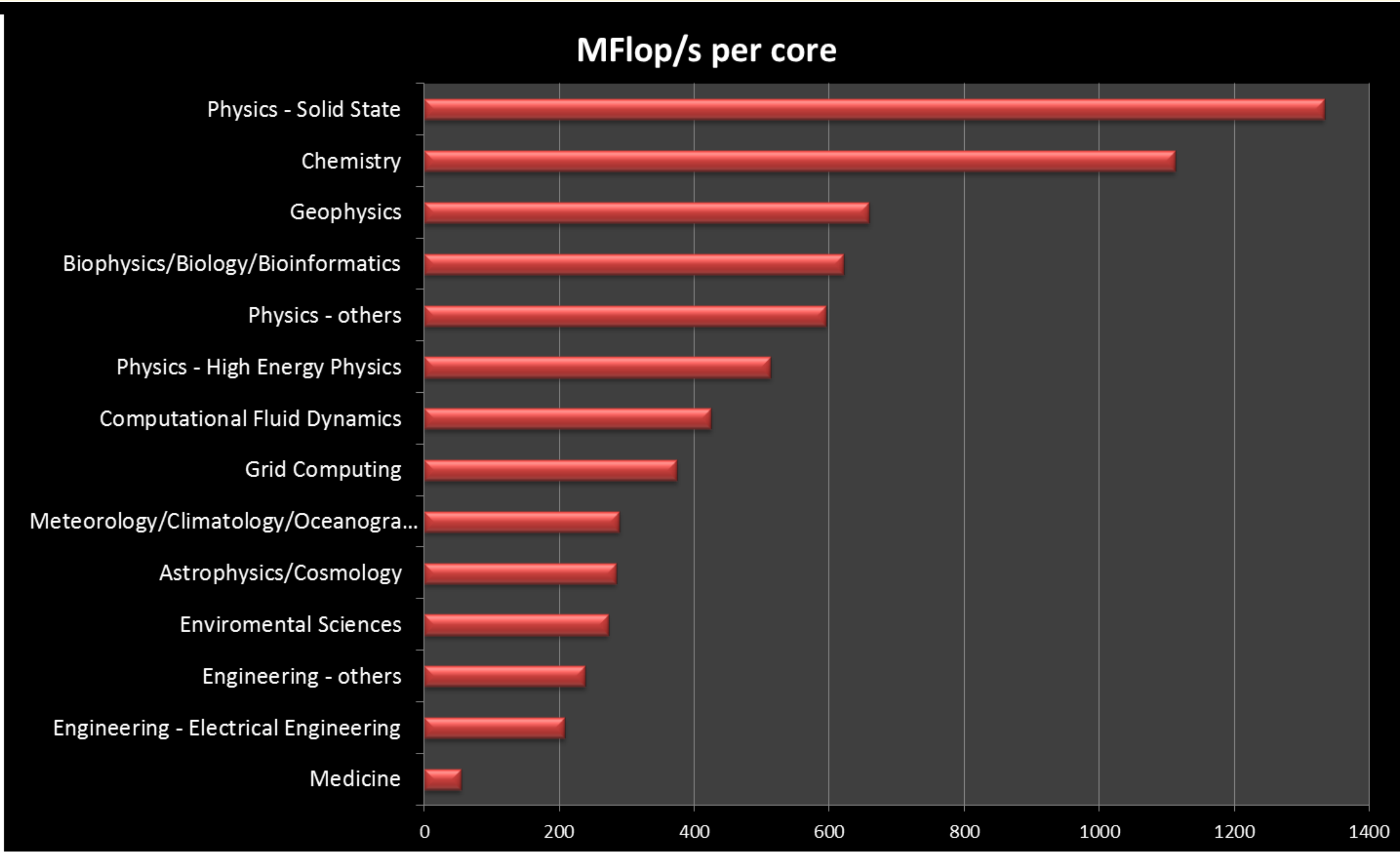
Infrastructure 4th floor

Infrastructure roof

Reconsider TCO

- **GPGPUs for Dusty Decks?**
- **Manycore for Scientific Applications**
  - Statistics at LRZ: Percentage of max performance for different application areas
- **Minimizing Energy Maximizes Programming Cost and Minimizes System Performance Yield**
- **We need tools to find a global optimum!**

Performance per core by Research area

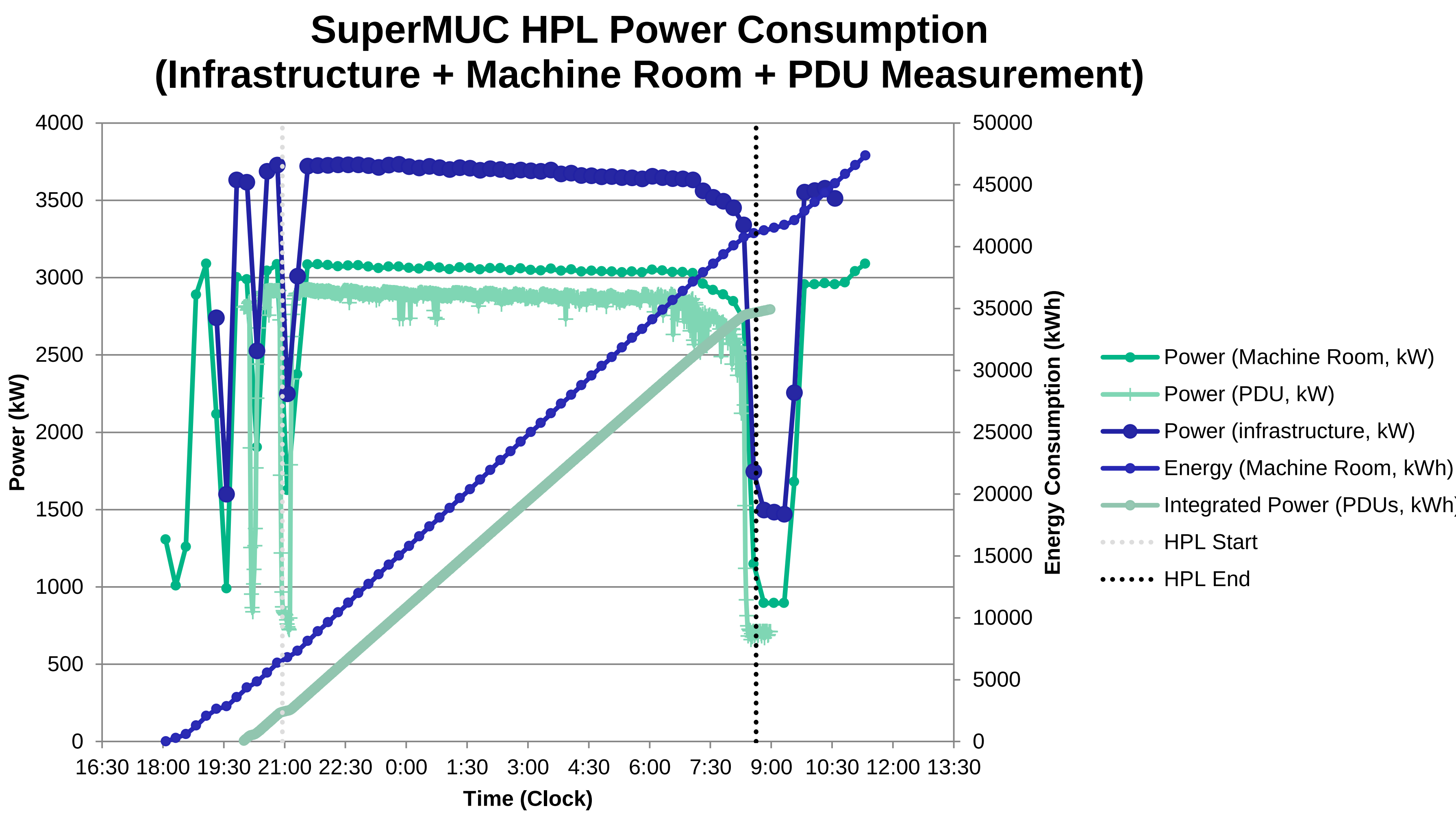


Some Results

- **Node Power Consumption and Cooling Infrastructure**

- **Using the characteristics of new processor products (SB EP)**
  - Power Efficiency depends on clock speed (parabolic curve)
  - Power Efficiency depends on voltage (lower is better)
  - Turbo mode is good for TOP500 - not for Power Efficiency
  - Energy loss with increased chip temperature
    Cool down as much as possible (cost for cooling?)
    Use liquid instead of air
  - Disable unnecessary units (virtualization, …)
  - Need to adjust processor, memory - and interconnect speed

Measurement data




Deutscher Rechenzentrumspreis 2012

SuperMUC users and network

LRZ / BAdW

MCSC, BGCE, ENB

KONWIHR

GCS,GA

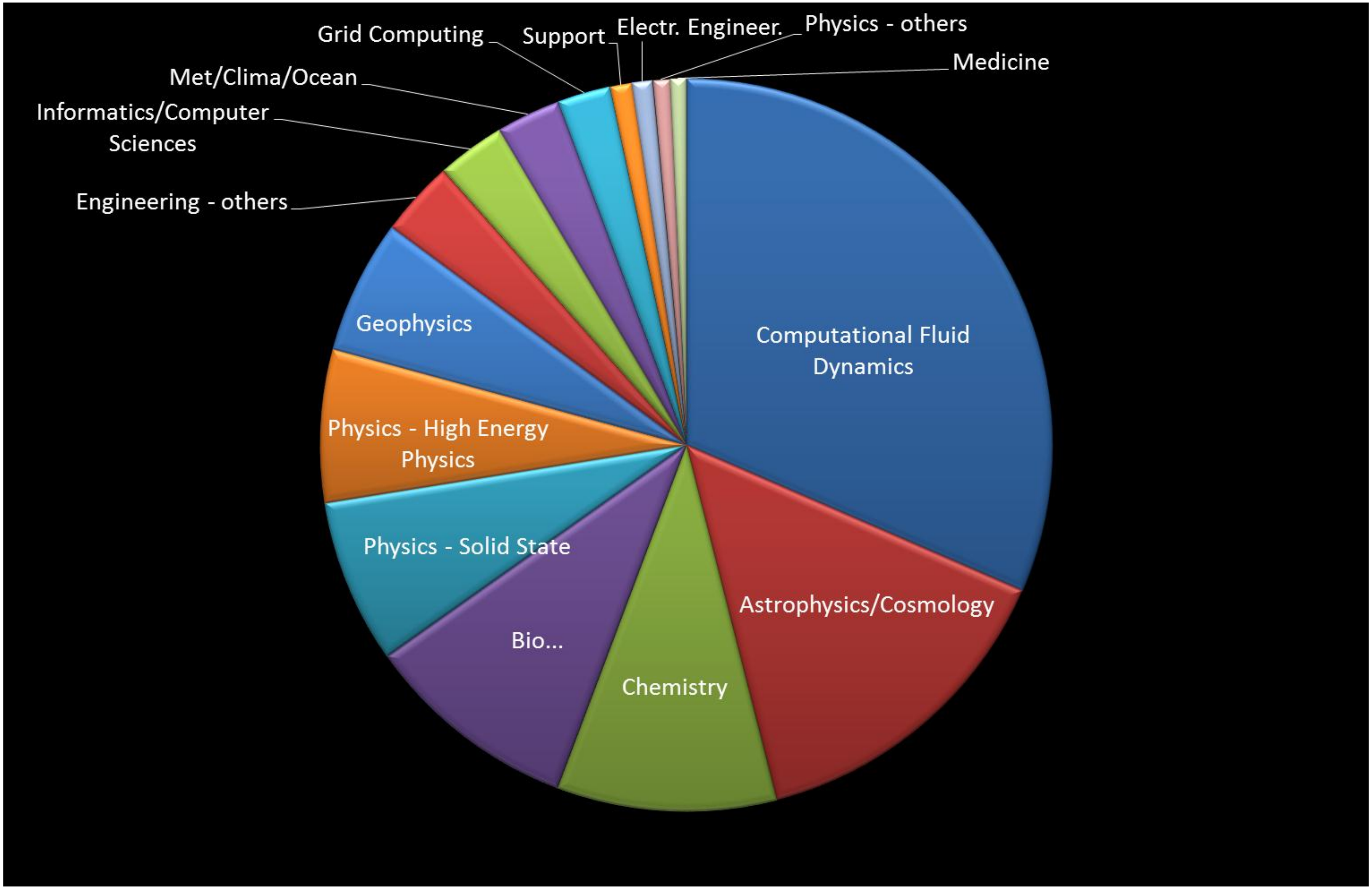
PRACE

PROSPECT

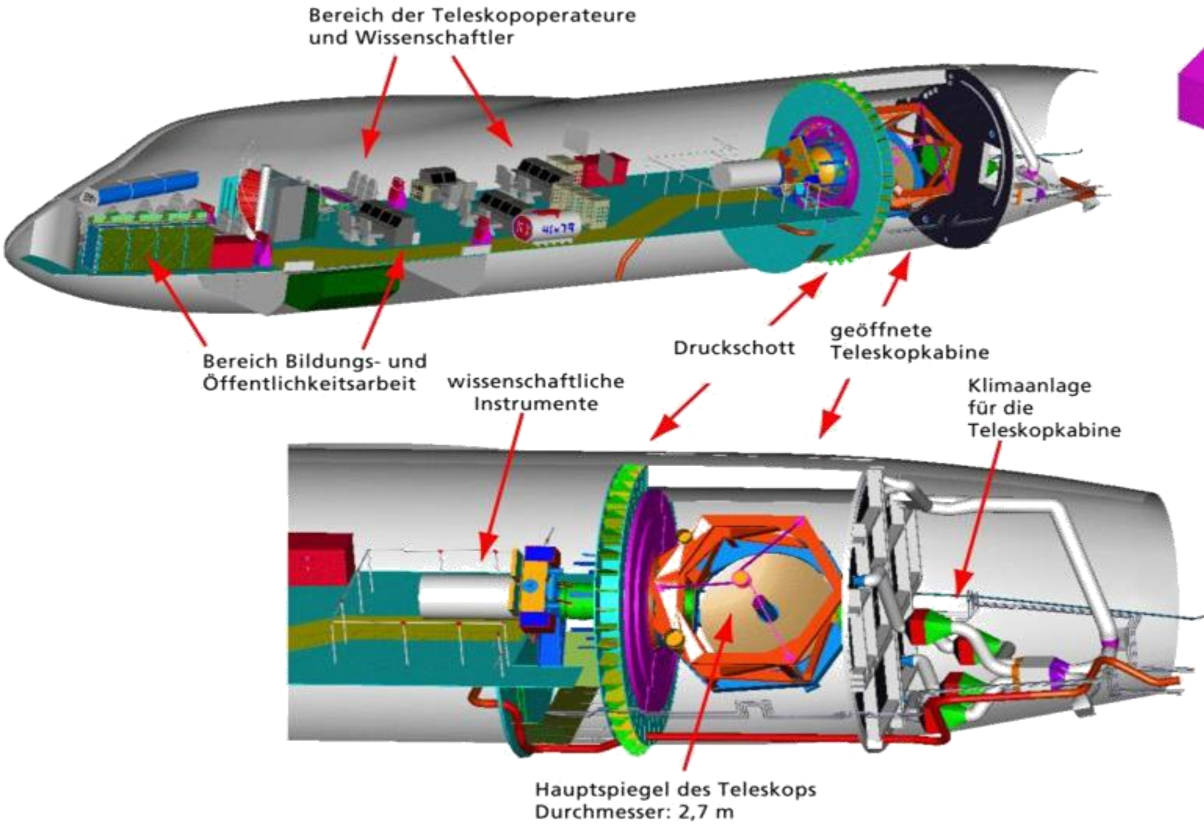
ETP4HPC

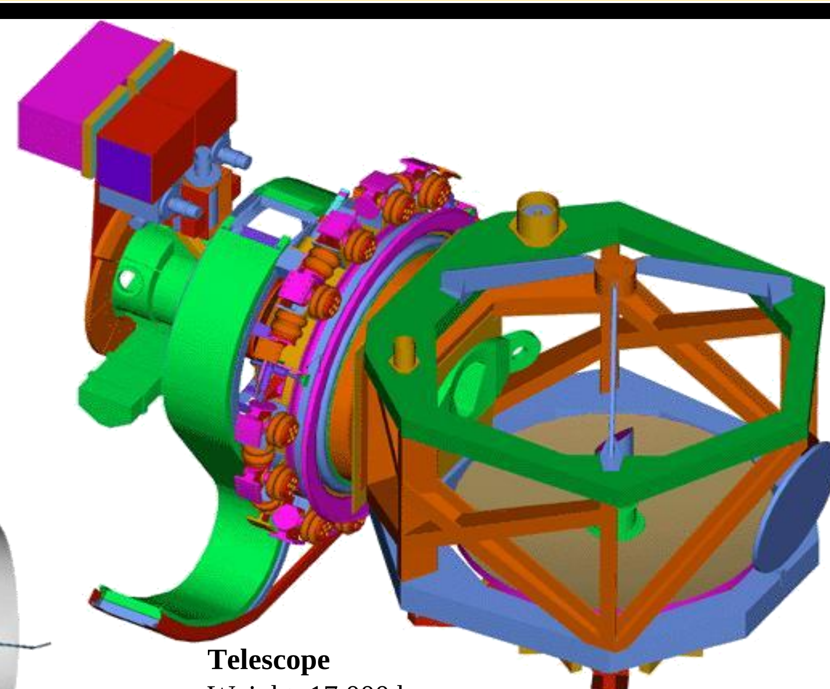
- HPC for science
- HPC for industry
- Support and education
- Future System and Application research and development

Usage 2010
by Research Area



Stratospheric Observatory For Infrared Astronomy - SOFIA

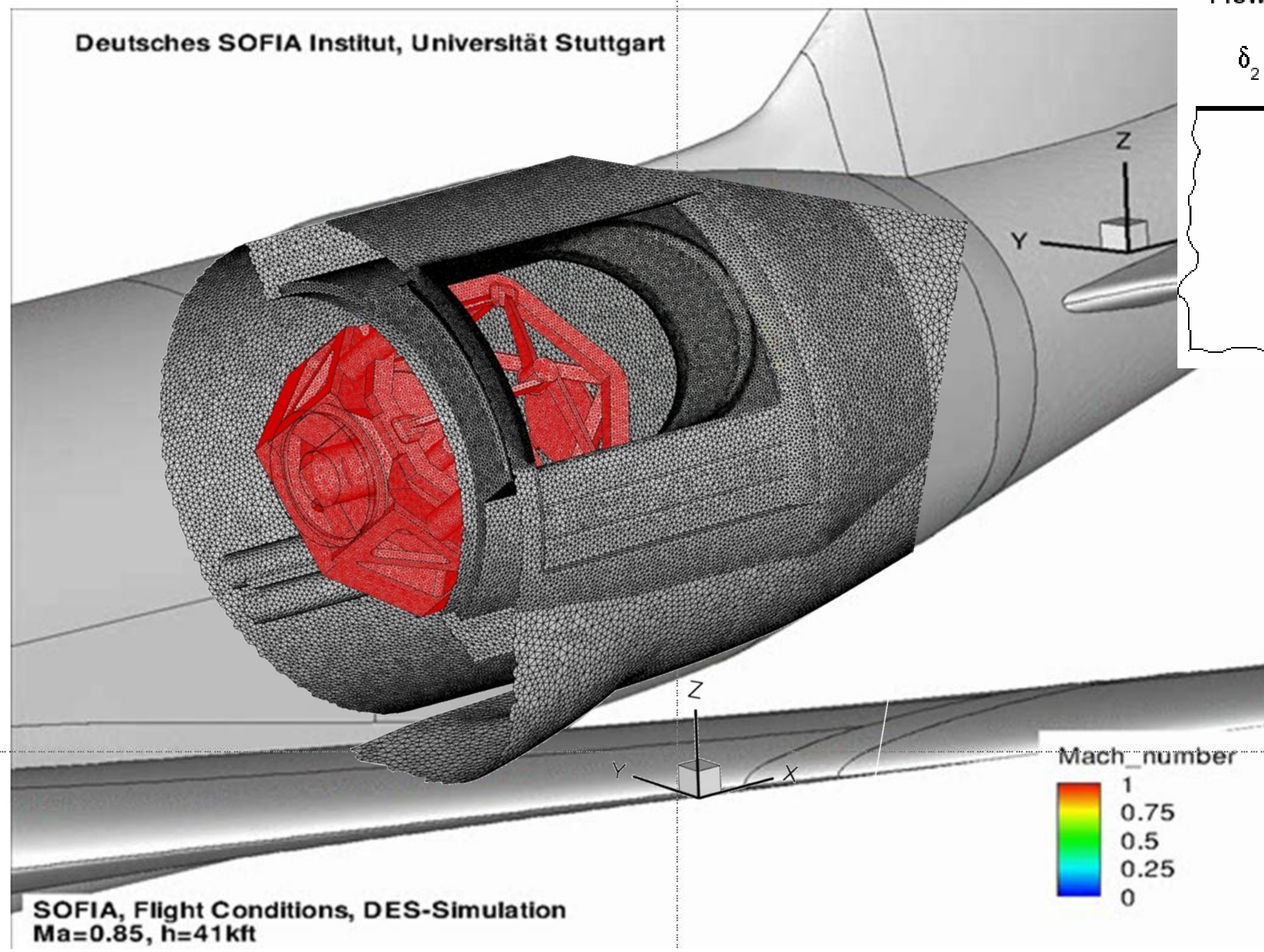


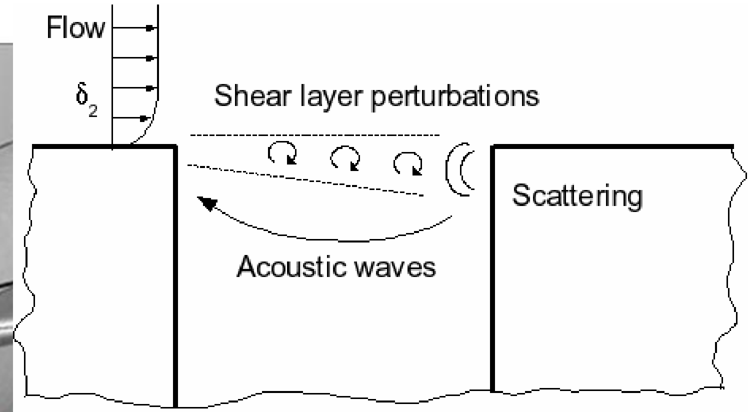


Telescope
Weight: 17.000 kg
Diameter: 2.7m
Elevation range: 20° - 60°
Developed and built in Germany

Platform: Boeing 747 SP
Max. Gross Weight: 300.000 kg
Wingspan: 60m
Max. Range: 15.000km

Shear Layer Visualization

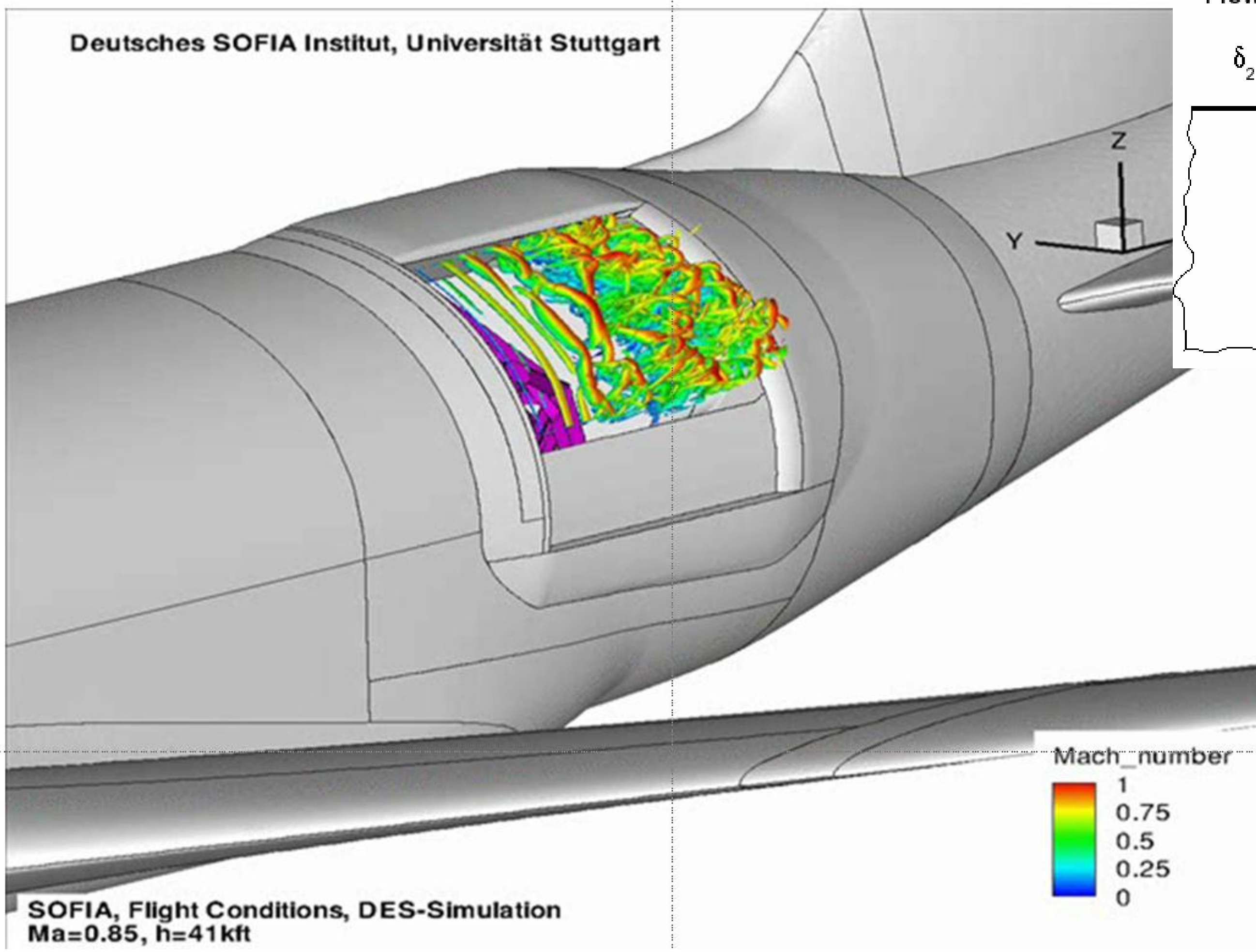


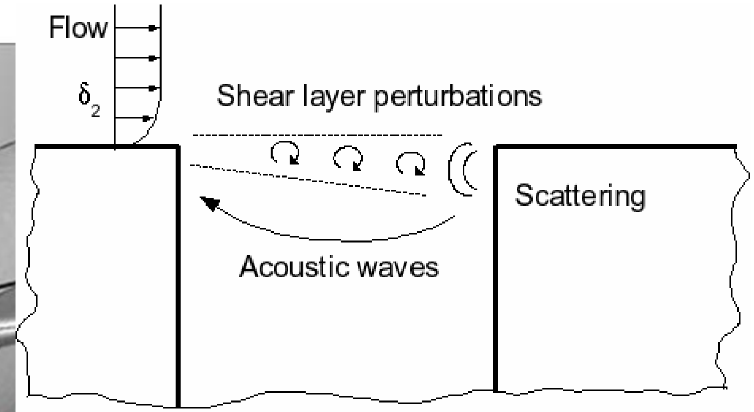


IAG

Sven Schmid
Institut für Aerodynamik und Gasdynamik

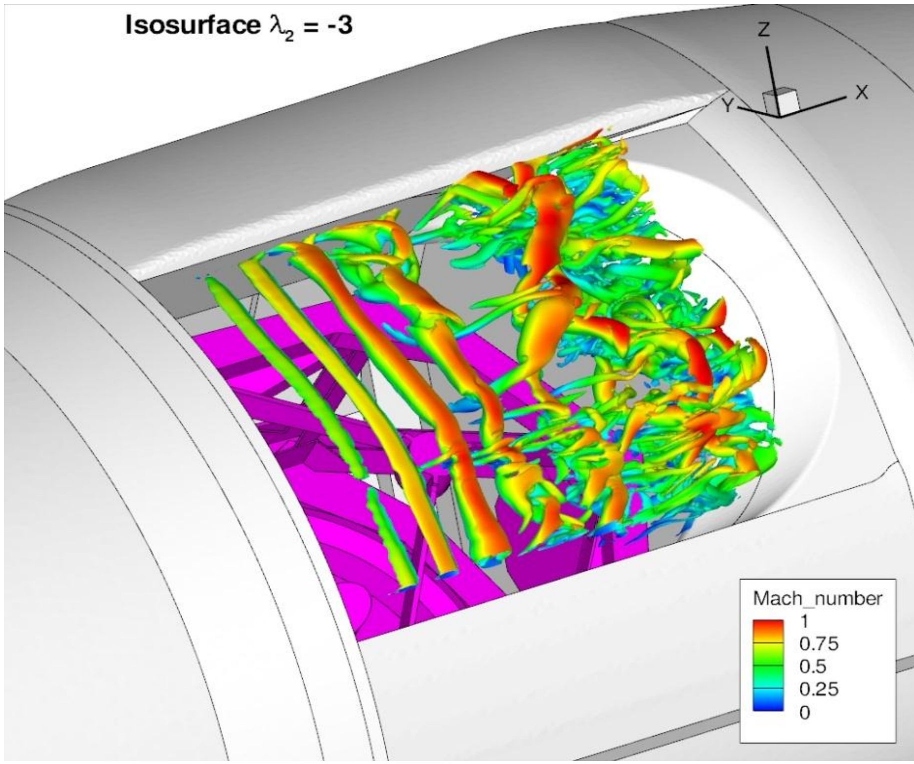
Shear Layer Visualization



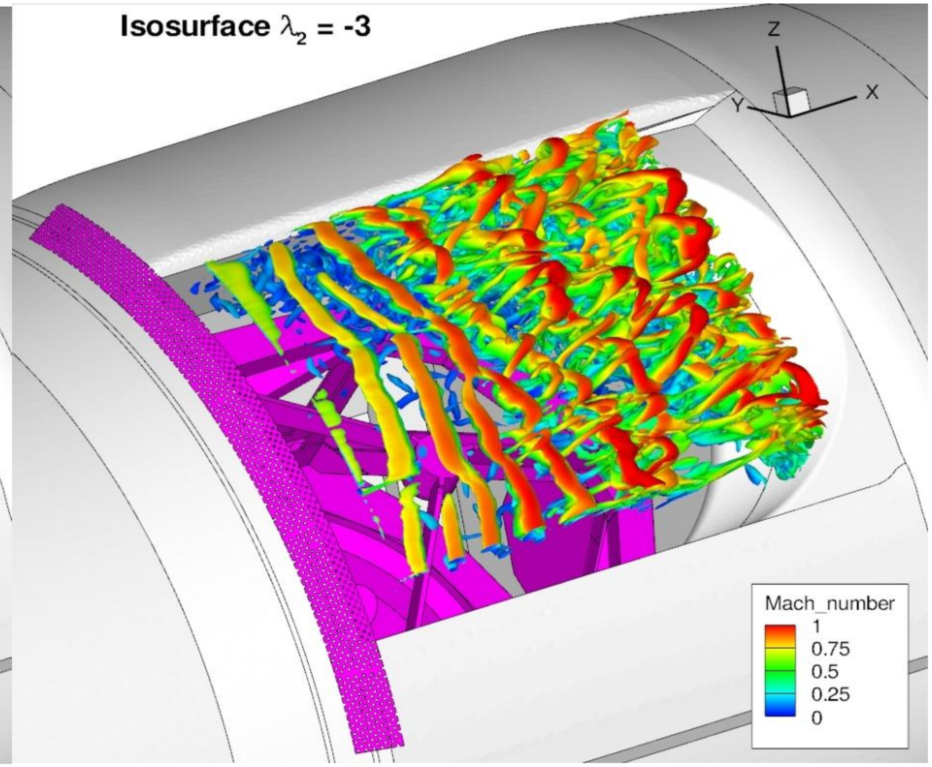


IAG
Sven Schmid
Institut für Aerodynamik und Gasdynamik

SOFIA Configuration with Fence

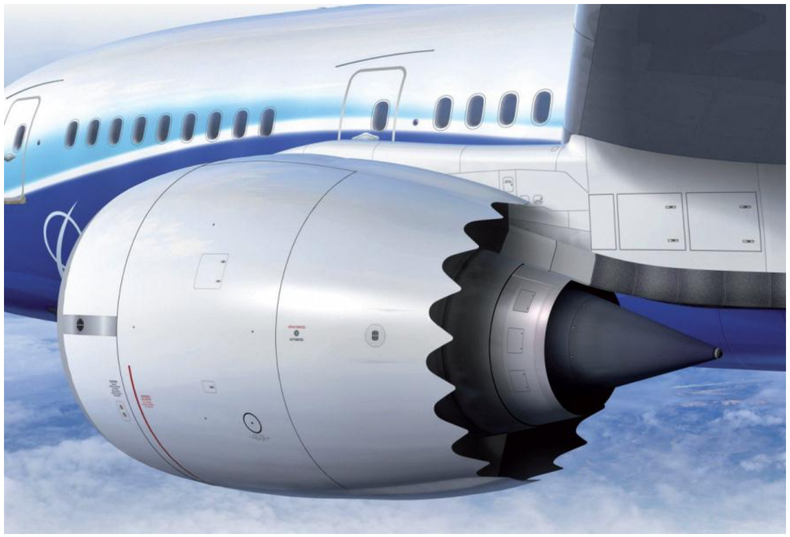


SOFIA configuration without fence

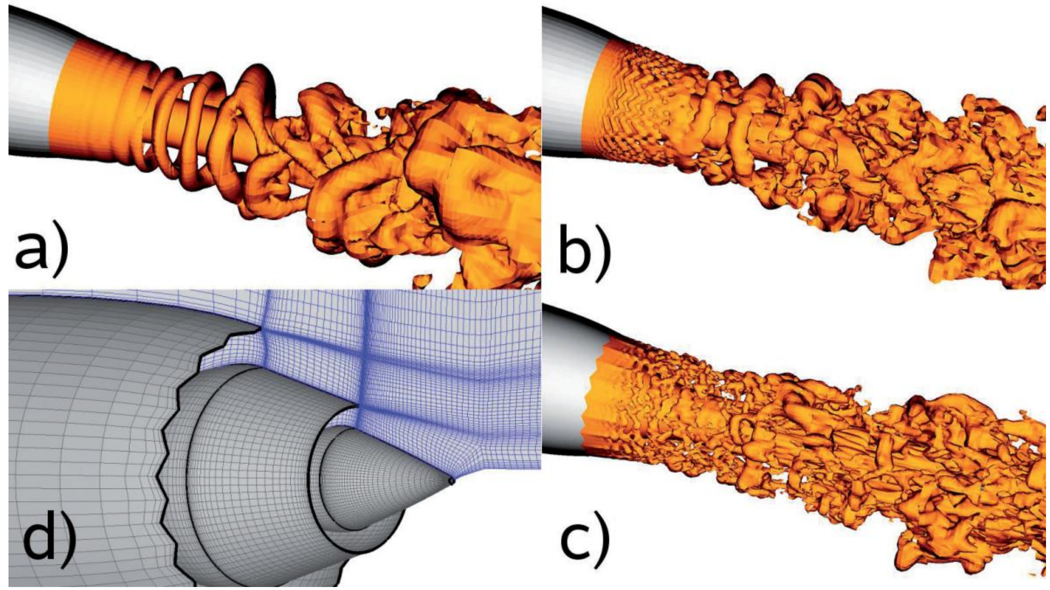


SOFIA configuration with fence

Numerical investigation of the noise from modified nozzle geometries



Rolls-Royce Trent 1000 engine on a Boing 787 Dreamliner (copyright Rolls-Royce plc 2010)



Iso-surfaces of the density showing resolved turbulent flow structures

a) standard DES;
b) modified DES, unserrated nozzle;
c) modified DES, serrated nozzle;
d) serrated short-cowl nozzle surface mesh with every second grid line shown

Institute of Fluid Mechanics and Engineering Acoustics, University of Technology Berlin

Principal Investigator

F. Thiele

Researchers

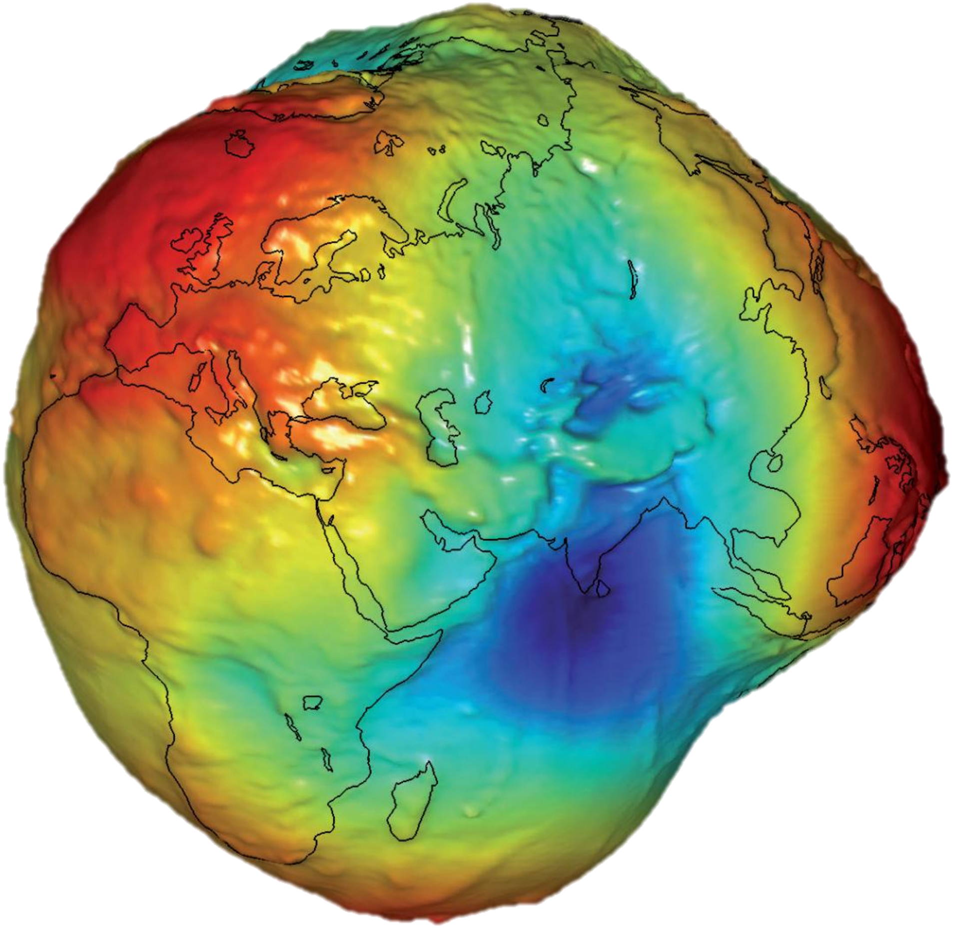
D. Eschricht, J. Yan, L. Panek and K. Tawackolian

Project Partners

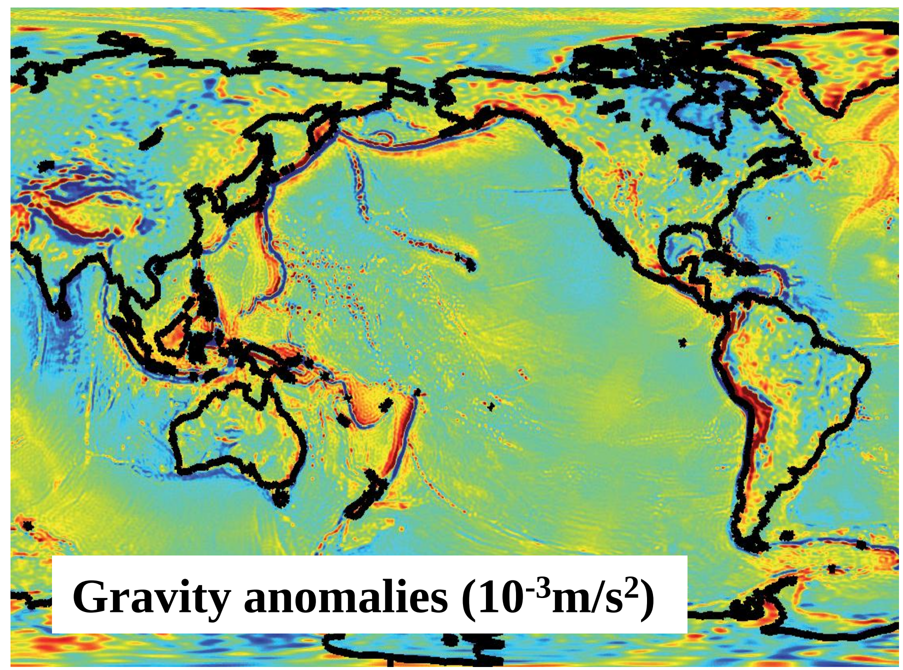
Rolls-Royce Germany

Global high-resolution gravity field determination

Research Institution
Institute for Astronomical and Physical Geodesy
Principal Investigator
Thomas Gruber
Researchers
Thomas Fecher, Prof Roland Pail



The geoid, the physical shape of the Earth



Gravity anomalies ($10^{-3} m/s^2$)

SuperMUC and after

- SuperMUC Phase 2 in 2014: manycore technology
  better energy efficiency
- LRZ in Exascale projects: DEEP (Intel MIC and Xtoll)
  Mont-Blanc (ARM technology)
  EESI
- Successor to SuperMUC needs strong support for users:
  Scalability issues for the „Mega-core-system“
- New „HPC styles“: Big Data
  Realtime HPC
  Integrated Visualization
  Steering