The new SuperMUC petascale system and applications

Dieter Kranzlmüller

Munich Network Management Team
Ludwig-Maximilians-Universität München (LMU) &
Leibniz Supercomputing Centre (LRZ)
of the Bavarian Academy of Sciences and Humanities

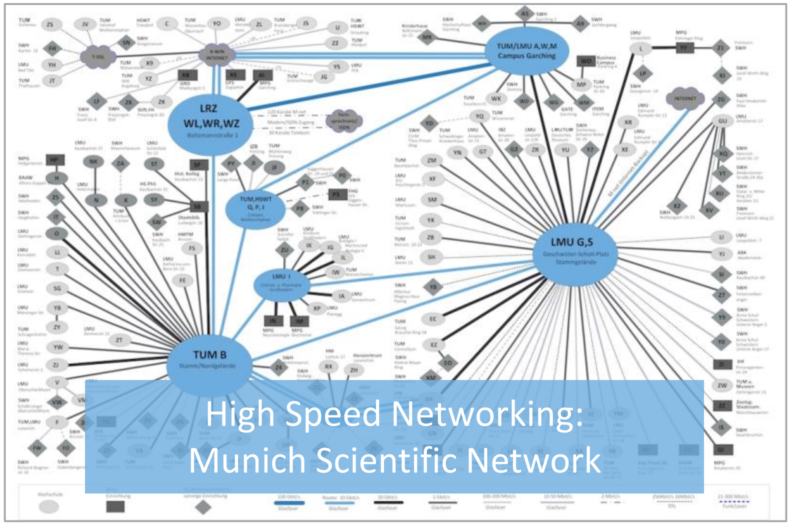
Thank you to HCMUT Team

With approx. 250 employees
for more than 100.000 students and
for more than 30.000 employees
including 8.500 scientists

- European Supercomputing Centre
- National Supercomputing Centre
- Regional Computer Centre for all Bavarian Universities
- Computer Centre for all Munich Universities

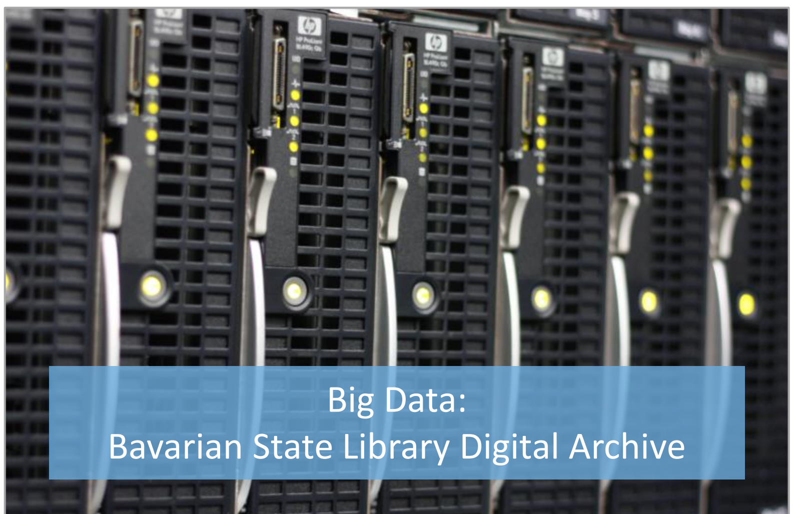
Photo: Ernst Graf

LRZ as IT Competence Centre:
Operating Cutting-edge IT Infrastructure

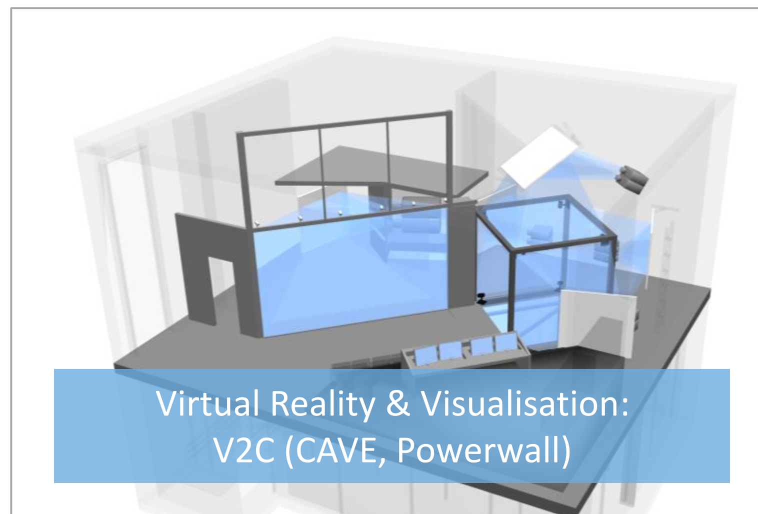


High Speed Networking:
Munich Scientific Network

High Performance Computing:
SuperMUC, LinuxCluster

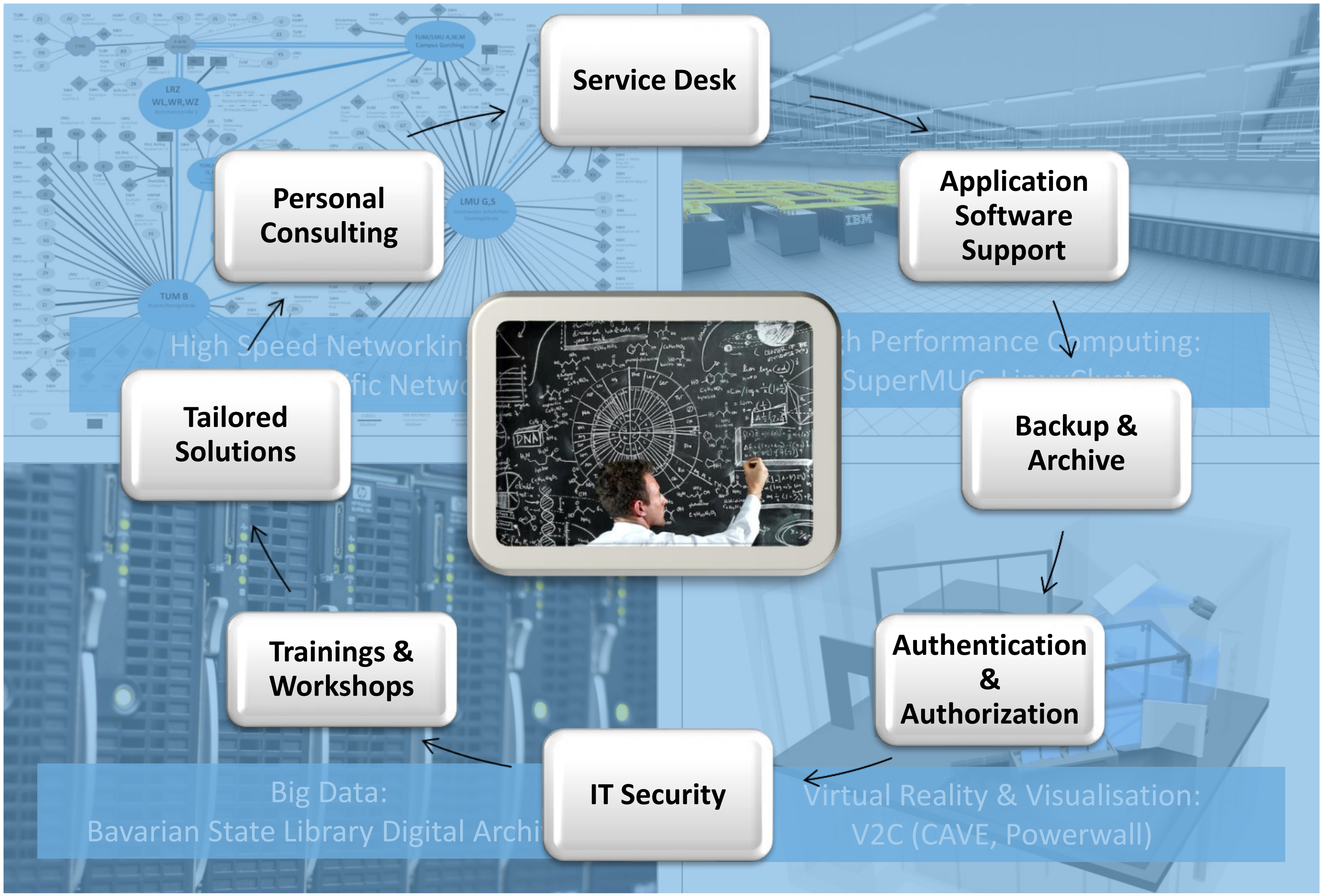


Big Data:
Bavarian State Library Digital Archive



Virtual Reality & Visualisation:
V2C (CAVE, Powerwall)

LRZ as IT Competence Centre:
Providing Comprehensive IT Services for Science



SuperMUC @ LRZ

Video: **SuperMUC rendered on SuperMUC by LRZ**

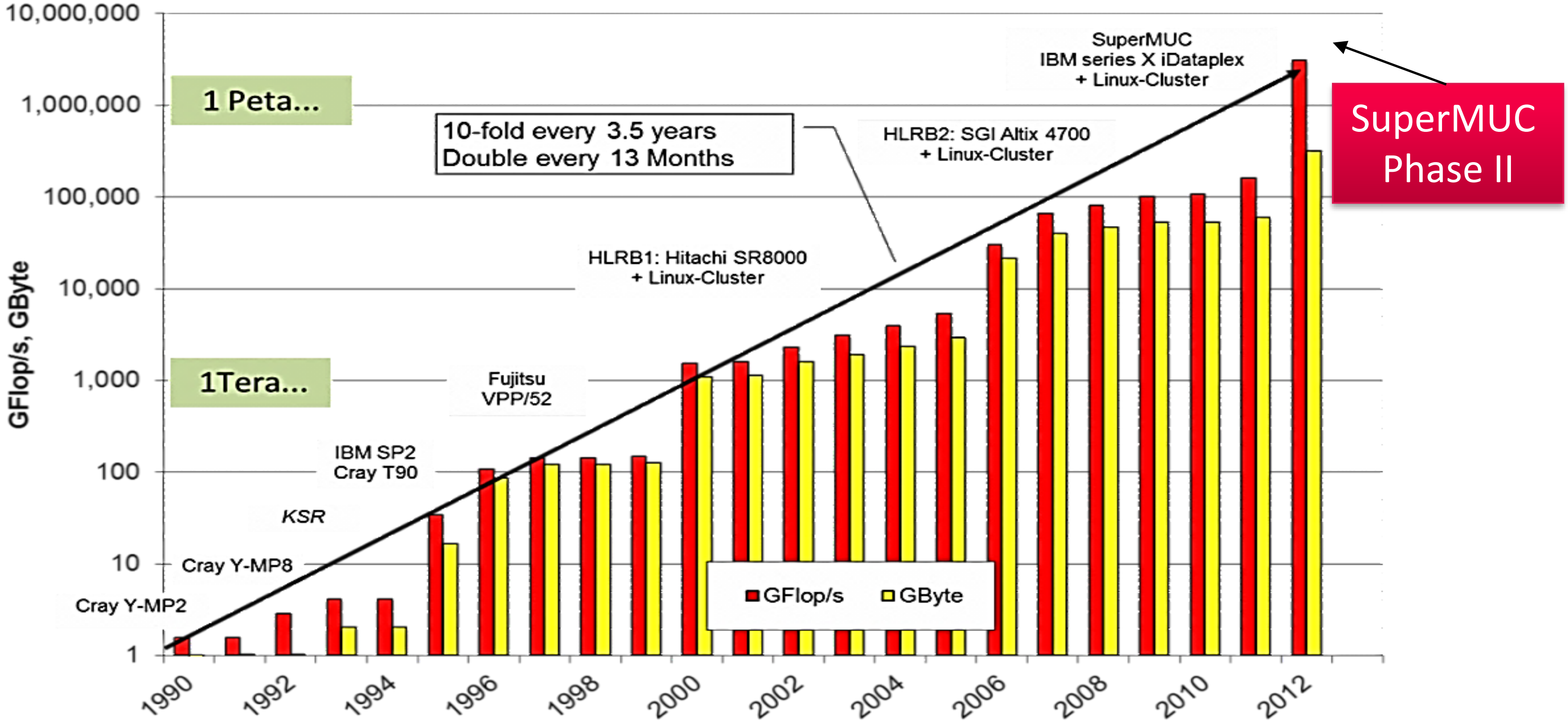
http://youtu.be/OlAS6iiqWrQ

Top 500 Supercomputer List (June 2012)

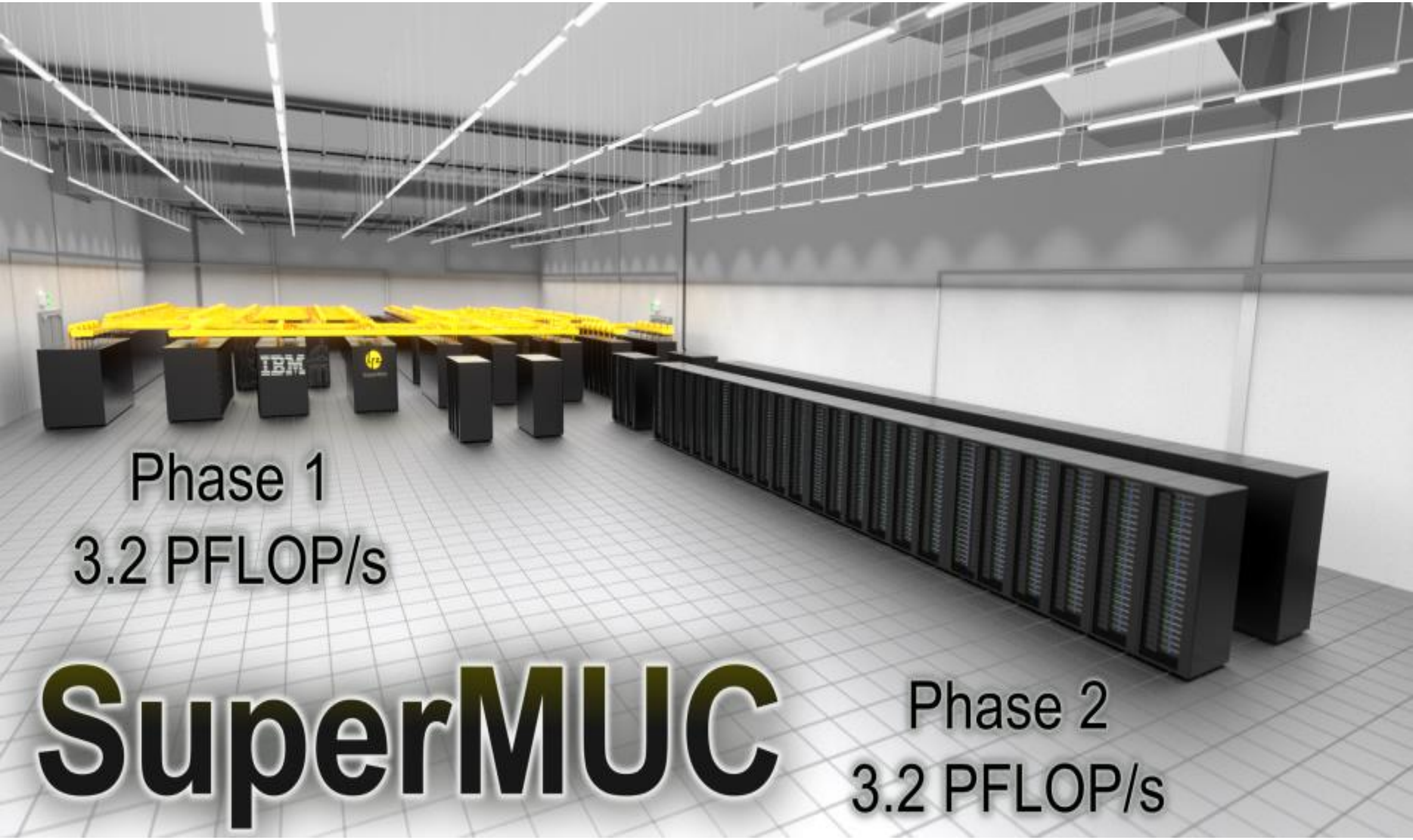
Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}	Power
1	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom / 2011 IBM	1572864	16324.75	20132.66	7890.0
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect / 2011 Fujitsu	705024	10510.00	11280.38	12659.9
3	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	786432	8162.38	10066.33	3945.0
4	Leibniz Rechenzentrum Germany	**SuperMUC** - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR / 2012 IBM	147456	2897.00	3185.05	3422.7
5	National Supercomputing Center in Tianjin China	**Tianhe-1A** - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 / 2010 NUDT	186368	2566.00	4701.00	4040.0
6	DOE/SC/Oak Ridge National Laboratory United States	**Jaguar** - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 / 2009 Cray Inc.	298592	1941.00	2627.61	5142.0
7	CINECA Italy	**Fermi** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	163840	1725.49	2097.15	821.9
8	Forschungszentrum Juelich (FZJ) Germany	**JuQUEEN** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	131072	1380.39	1677.72	657.5
9	CEA/TGCC-GENCI France	**Curie thin nodes** - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR / 2012 Bull	77184	1359.00	1667.17	2251.0
10	National Supercomputing Centre in Shenzhen (NSCS) China	**Nebulae** - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 / 2010 Dawning	120640	1271.00	2984.30	2580.0

www.top500.org

LRZ Supercomputers – Chronology



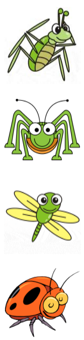
SuperMUC Phase 1 + 2



LRZ Application Mix

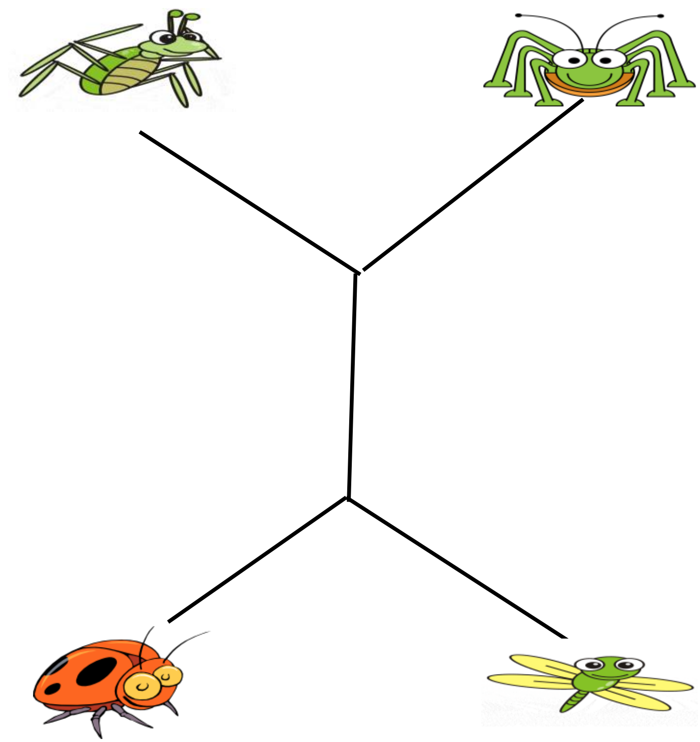
- Computational Fluid Dynamics: Optimisation of turbines/wings, noise reduction
- Fusion: Plasma in a future fusion reactor (ITER)
- Astrophysics: Origin and evolution of stars and galaxies
- Solid State Physics: Superconductivity, surface properties
- Geophysics: Earth quake scenarios
- Material Science: Semiconductors
- Chemistry: Catalytic reactions
- Medicine and Medical Engineering: Blood flow, aneurysms, air conditioning
- Biophysics: Properties of viruses, genome analysis
- Climate research: Currents in oceans
- ...

Phylogenetic Tree Computation



ACGT
ACC
ACGG
AAGC

ACGT
ACC-
ACGG
AAGC



Sequencing → Alignment → Phylogenetic Tree

Alexandros Stamatakis, H-ITS

of possible trees for 150 species

- 4226517247809112252219618802377042809718932383449
  8822942857479880831434032178759024536798491951168
  3076494692867414802738570221298292428457687814873
  4552121861861600804474608426626044448936698500560
  2468116186441264227425440726676614927906540649360
  2976397461917469326750931190889241406694054603576
  66015625

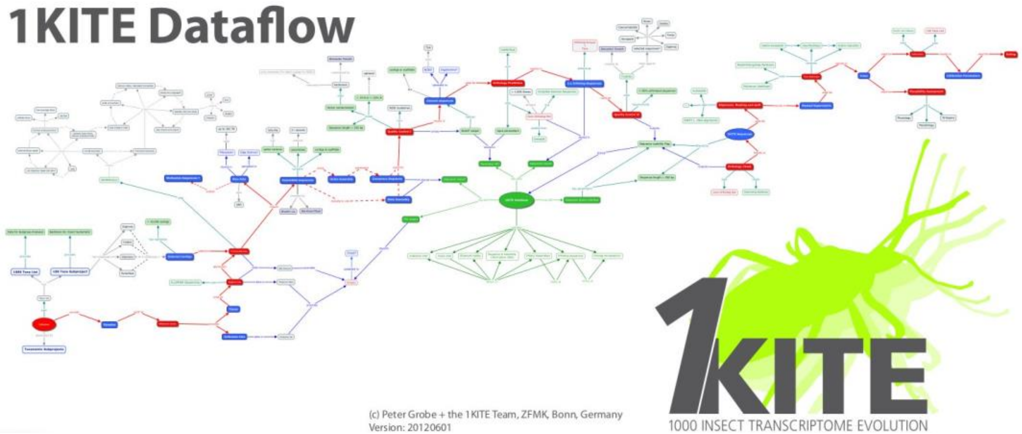
- ≈ 4.22×10^{301}

Alexandros Stamatakis, H-ITS

1KITE Project
(1,000 Insect Transcriptome Evolution)

- Alexandros Stamatakis
  Scientific Computing Group,
  Heidelberg Institute for Theoretical Studies (HITS) /
  Exelixis Lab

- „Big Data“ and High Performance Computing
- Novel software and applications needed
- Reading the data: only 1 minute (instead of 15 minutes)
- 1000 Processors: 17 hours (instead of 10 days)
- Load balancing



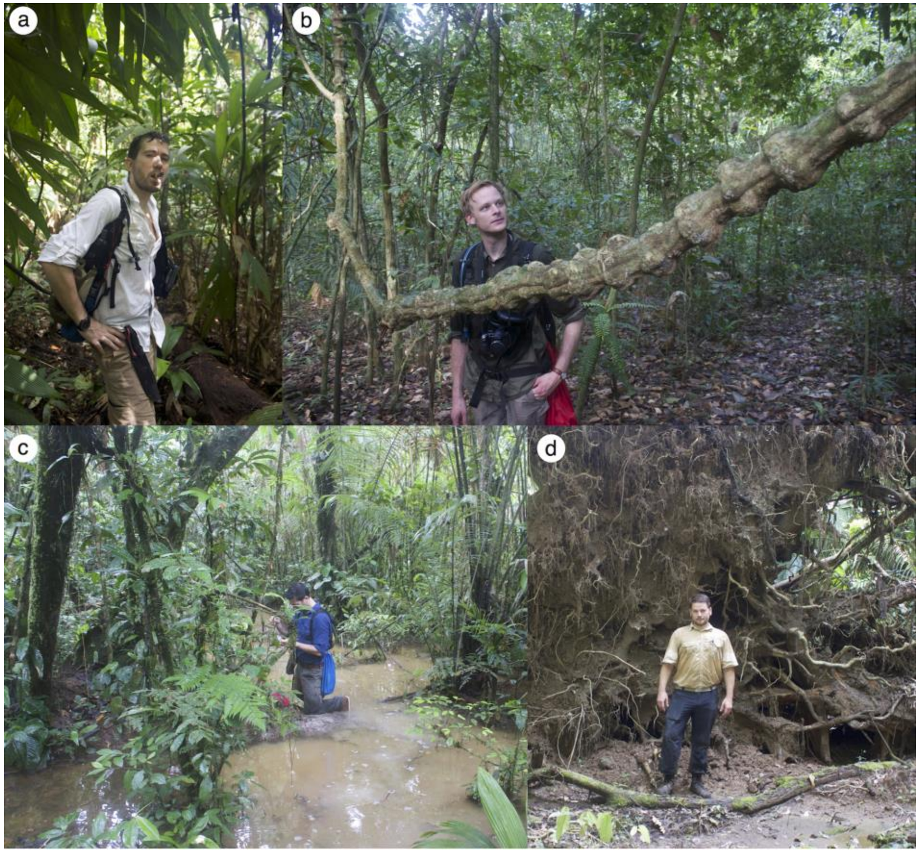
Scientific Results - Publications

Alexandros Stamatakis, H-ITS

Biodiversity

- Neotropical Rainforests are hyperdiverse ecosystems
- Since Humboldt and Bonpland, we know about the high animal and plant richness
- New study now finds that unicellular eukaryotes are even more diverse
- Particularly the parasitic Apicomplexa dominate these forests
- Their presence might drive the
- diversity of macro-organisms

Micah Dunthorn/TU Kaiserslautern

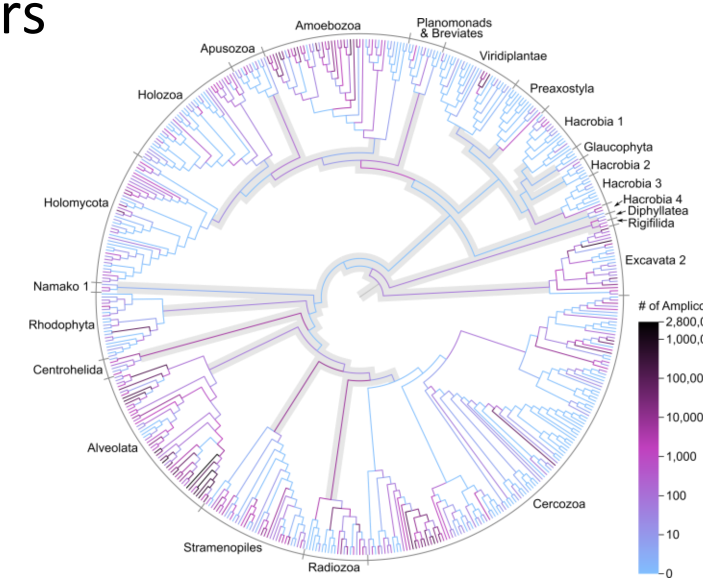


https://natureecoevocommunity.nature.com/channels/521-behind-the-paper/posts/15402-a-larger-microbial-perspective-of-tropical-rainforests

Data Analysis (TU Kaiserslautern, HITS, KIT)

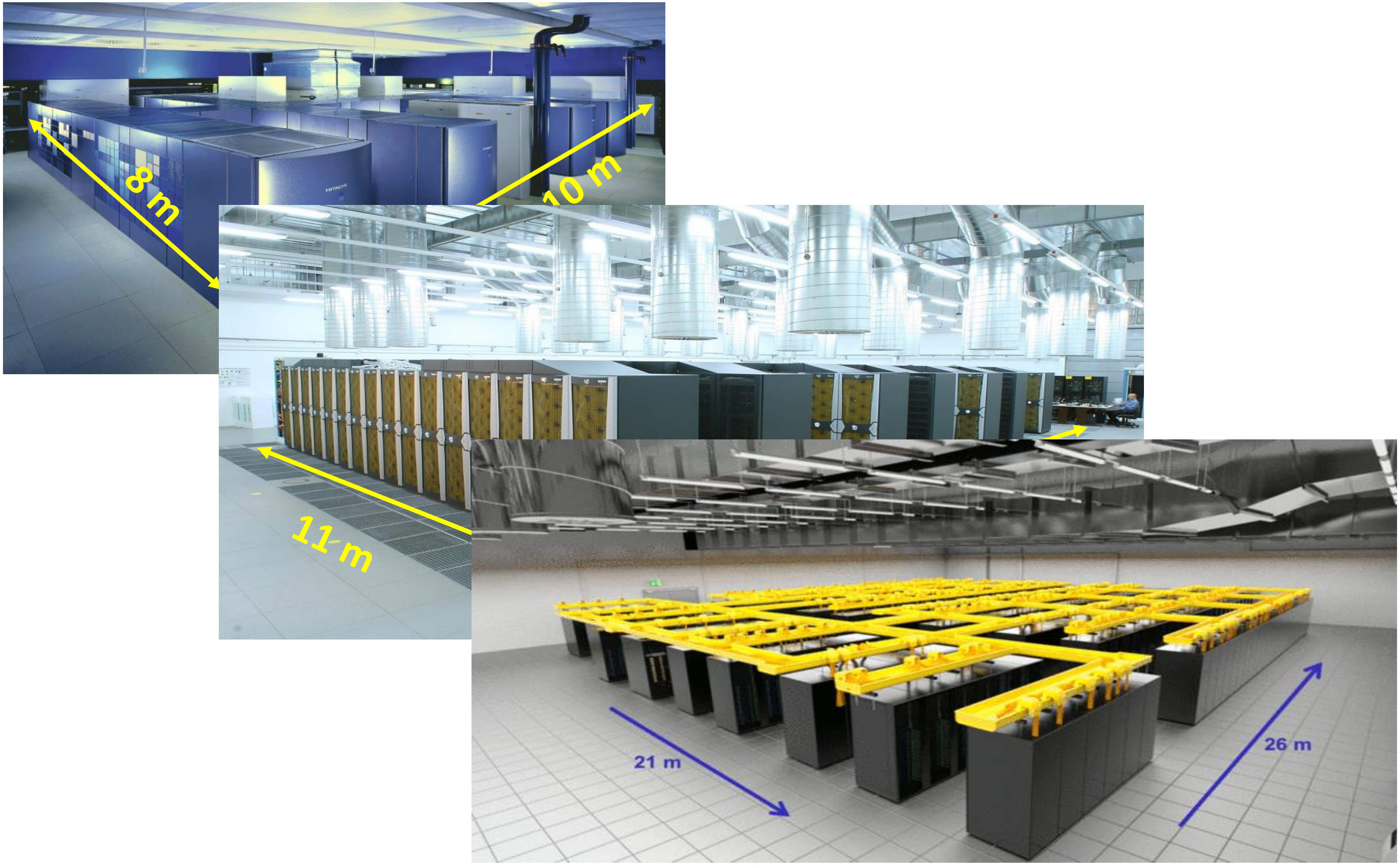
- More than 130 million DNA sequences were analysed
- Most of them belong to yet unknown microbial species
- Thus, a thorough method was necessary for classifying those sequences
- The method takes the evolutionary history of known species into account
- But this comes at the cost of increased computational needs
- Approximately 1 million computation hours on **SuperMUC** were necessary

Mahé et al. (2017). Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests.
Nature Ecology and Evolution 1:09.
DOI: 10.1038/s41559-017-0091



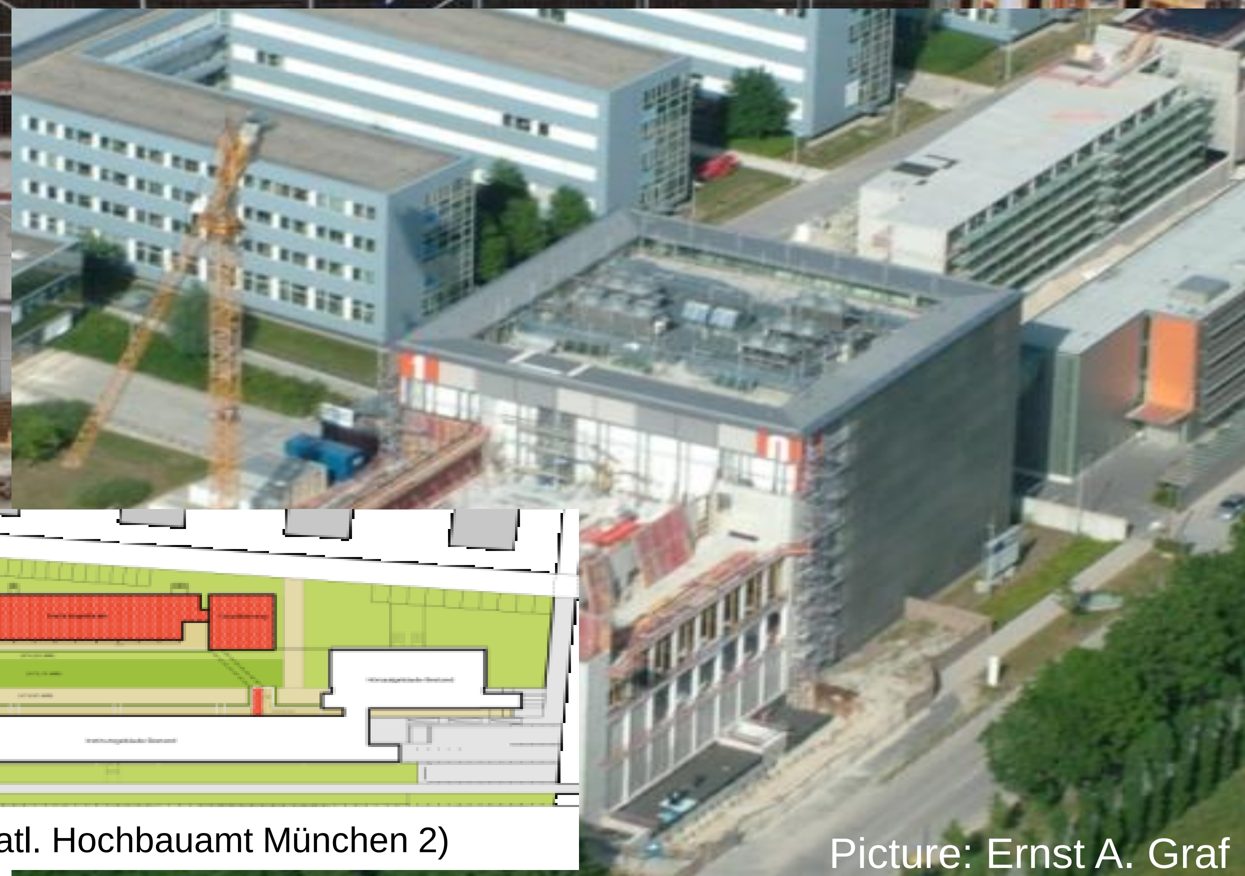
https://www.uni-kl.de/aktuelles/news/news/detail/News/neue-arten-entdeckt-mikroparasiten-tragen-zur-stabilitaet-des-oekosystems-im-regenwald-bei/

SuperMUC and its predecessors



LRZ Building Extension

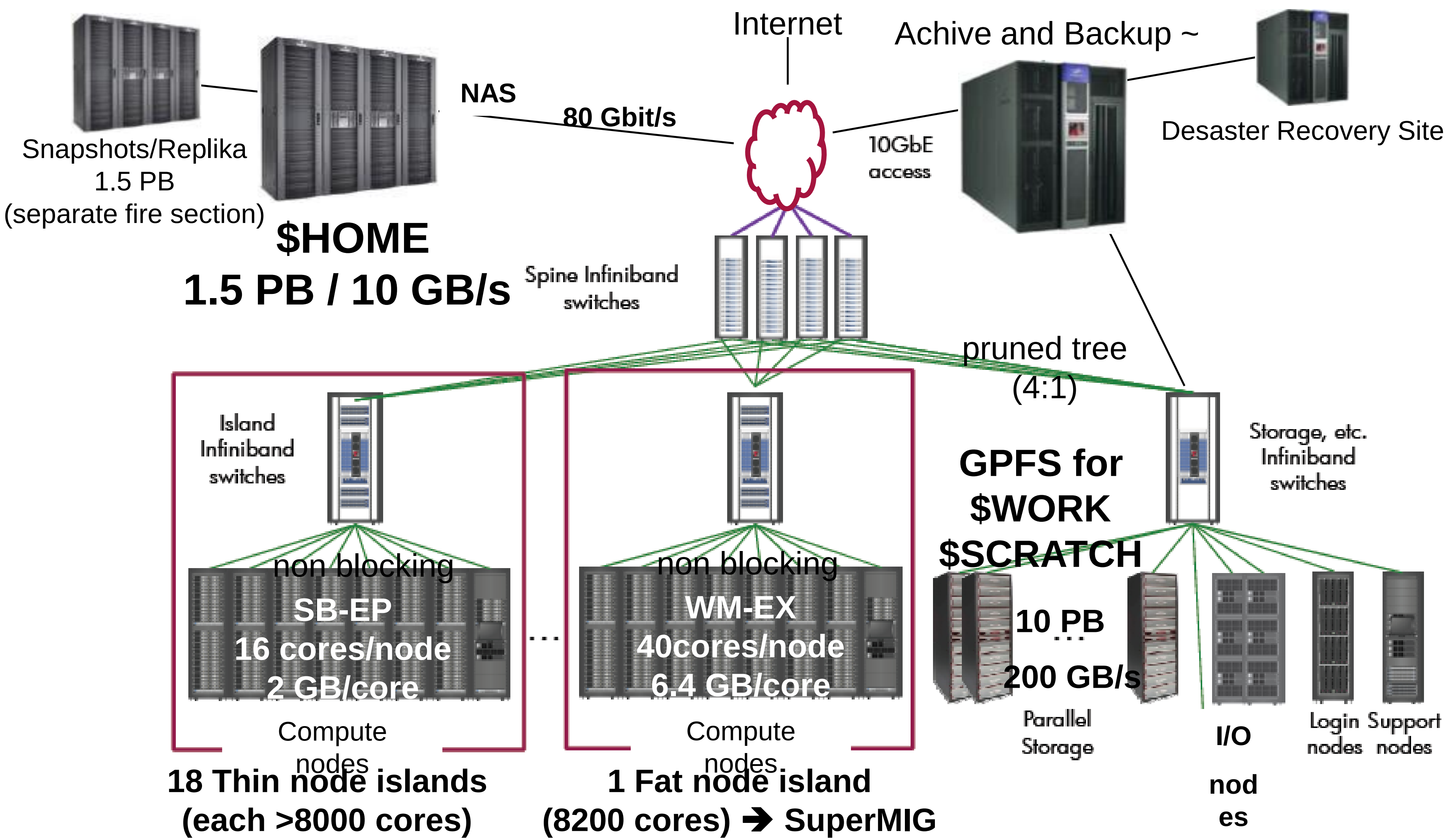
Picture: Horst-Dieter Steinhöfer



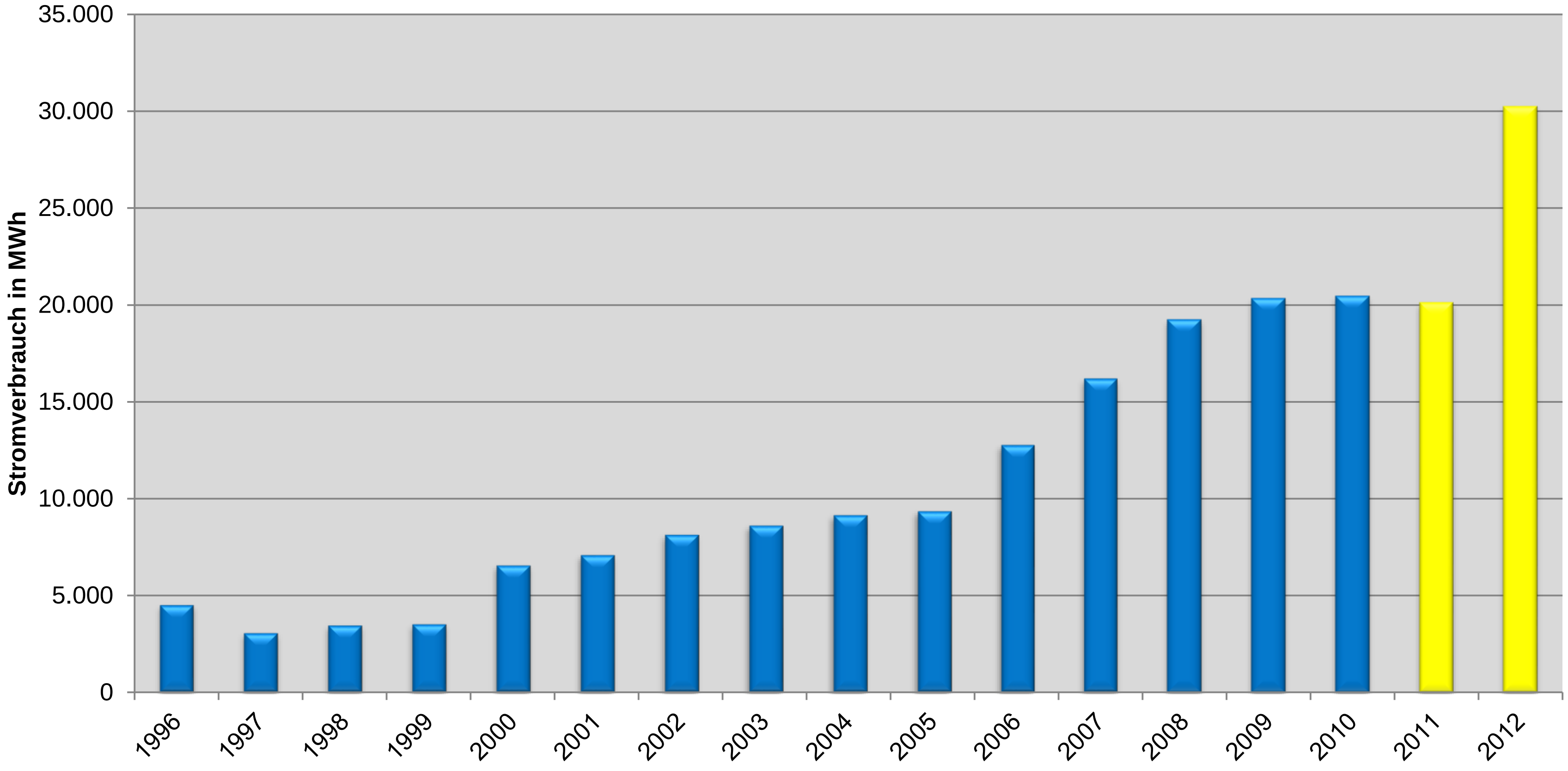
Picture: Ernst A. Graf

Figure: Herzog+Partner für StBAM2 (staatl. Hochbauamt München 2)

SuperMUC Architecture



Power Consumption at LRZ



Cooling SuperMUC

Energy Efficiency on SuperMUC @ LRZ

Photos: Torsten Bloth, Lenovo

- ✓ Usage of Intel Xeon E5 2697v3 processors
- ✓ Direct liquid cooling
  - 10% power advantage over air cooled system
  - 25% power advantage due to chiller-less cooling
- ✓ Energy-aware scheduling
  - 6% power advantage
  - ~40% power advantage
  - Annual savings: ~2 Mio. € for SuperMUC Phase 1 and 2

Increasing Numbers of Cores

Date	System	Flop/s	Cores
2000	HLRB-I	2 Tflop/s	1512
2006	HLRB-II	62 Tflop/s	9728
2012	SuperMUC	3200 Tflop/s	155656
2015	SuperMUC Phase II	3.2 + 3.2 Pflop/s	229960

1st LRZ Extreme Scaling Workshop

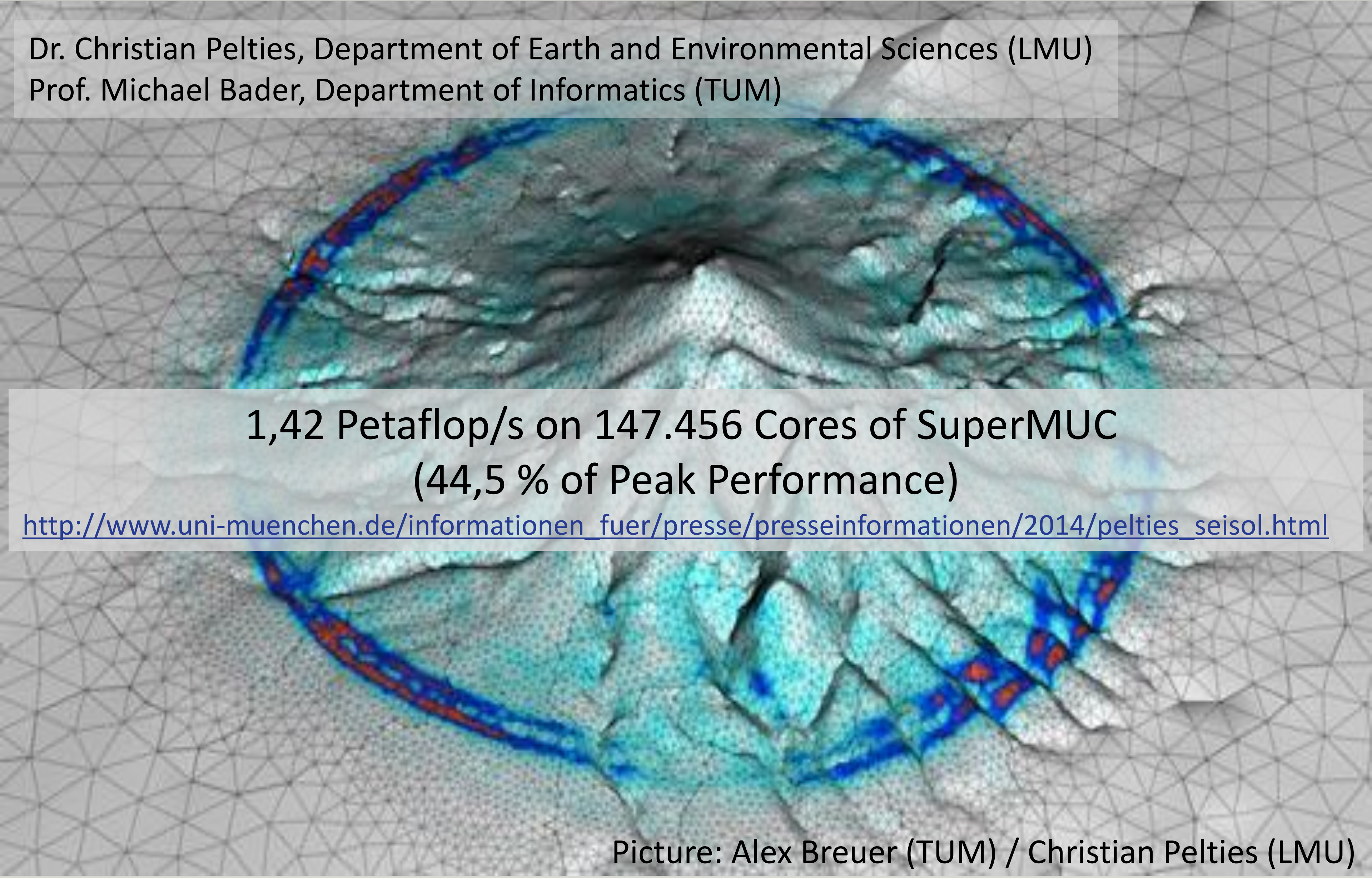
- Results:

Name	MPI	# cores	Description	TFlop/s/island	TFlop/s max
Linpack	IBM	128000	TOP500	161	2560
Vertex	IBM	128000	Plasma Physics	15	245
GROMACS	IBM, Intel	64000	Molecular Modelling	40	110
Seissol	IBM	64000	Geophysics	31	95
waLBerla	IBM	128000	Lattice Boltzmann	5.6	90
LAMMPS	IBM	128000	Molecular Modelling	5.6	90
APES	IBM	64000	CFD	6	47
BQCD	Intel	128000	Quantum Physics	10	27

- Sustained TFlop/s on 64000/128000 cores

SeisSol - Numerical Simulation of Seismic Wave Phenomena

Dr. Christian Pelties, Department of Earth and Environmental Sciences (LMU)
Prof. Michael Bader, Department of Informatics (TUM)



1,42 Petaflop/s on 147.456 Cores of SuperMUC
(44,5 % of Peak Performance)

http://www.uni-muenchen.de/informationen_fuer/presse/presseinformationen/2014/pelties_seisol.html

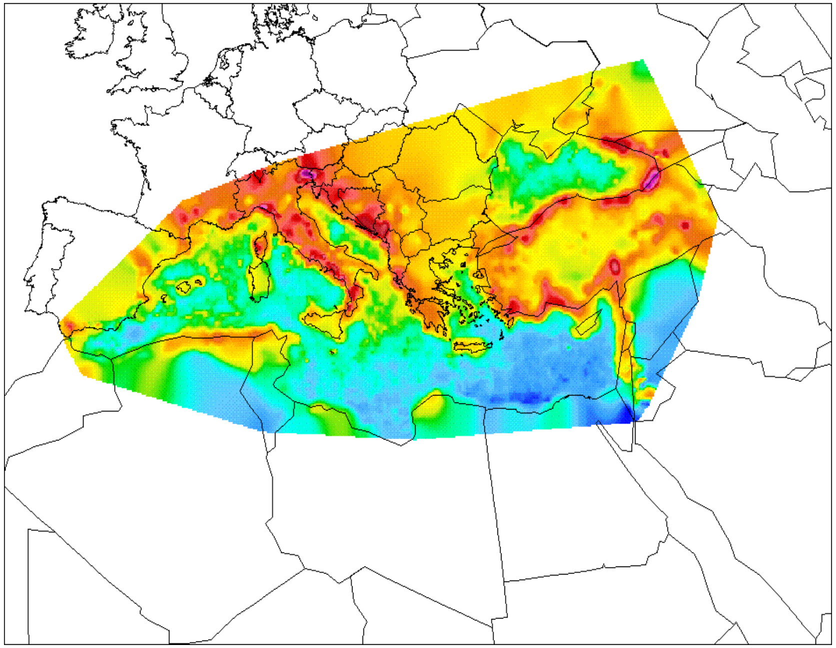
Picture: Alex Breuer (TUM) / Christian Pelties (LMU)

LRZ benefits

- Understanding the (current and future) needs and requirements of the respective scientific domain
- Developing future services for all user groups
- Thematic focusing: **Environmental Computing**

EU Project Series DRIHM*

- Flash Project estimates for 1990-2006
- > 29 billion euros in damages produced by floods
- > 4,500 total number of casualties



SSMI and raingauge observations (1978-1994)

Professor Peter V. Coveney

- holds a chair in Physical Chemistry
- is an Honorary Professor in Computer Science at University College London (UCL)
- is Professor Adjunct at Yale University School of Medicine (USA).

- is Director of the Centre for Computational Science (CCS) and of the Computational Life and Medical Sciences Network (CLMS) at UCL.
- https://www.ucl.ac.uk/chemistry/people/peter-coveney

- leads CompBioMed, A Centre of Excellence in Computational Biomedicine
- http://www.compbiomed.eu

- Goal: *advance the role of computationally based modelling and simulation within biomedicine*.

- Three related user communities:
  - academic,
  - industrial and
  - clinical researchers
- All wish to build, develop and extend such capabilities in line with the increasing power of high performance computers.

- Three distinct exemplar research areas:
  - cardiovascular,
  - molecularly-based and
  - neuro-musculoskeletal medicine.

HORIZON 2020

1 of 9 European Centres of Excellence in HPC
Official start on 1 October 2016; 3 years

How to Utilize SuperMUC as a Tool?

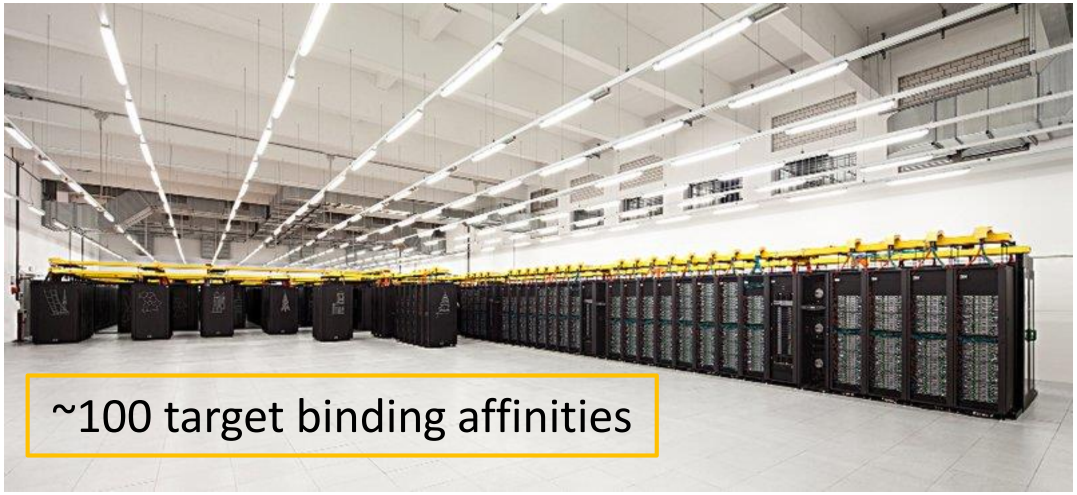
- Target question:
  Can we use the genomic data from an individual candidate and predict whether a standard drug for the treatment of breast cancer will help or not?

- Goal:
  A demonstration of feasibility
  with the power of high performance computing

- Key questions:
  - Provide an answer to the question above
  - Determine how to use IT-Infrastructures for this question
  - Detect insufficiencies of using IT-Infrastructures for this question
  - Derive a workflow for utilizing HPC in daily operation

Molecular Simulation

- Running on all cores of SuperMUC Phase1+2



~100 target binding affinities

- Docking simulation of potentials drugs for breast cancer
- 37 hours total run time
- 241,672 cores
- 8.900.000 CPU hours
- 5 Terabytes of data produced

EU CoE CompBioMed
http://www.compbiomed.eu
EU Projects COMPAT and MAPPER
http://www.compat-project.eu

What‘s next? A View into the Future

Until today:

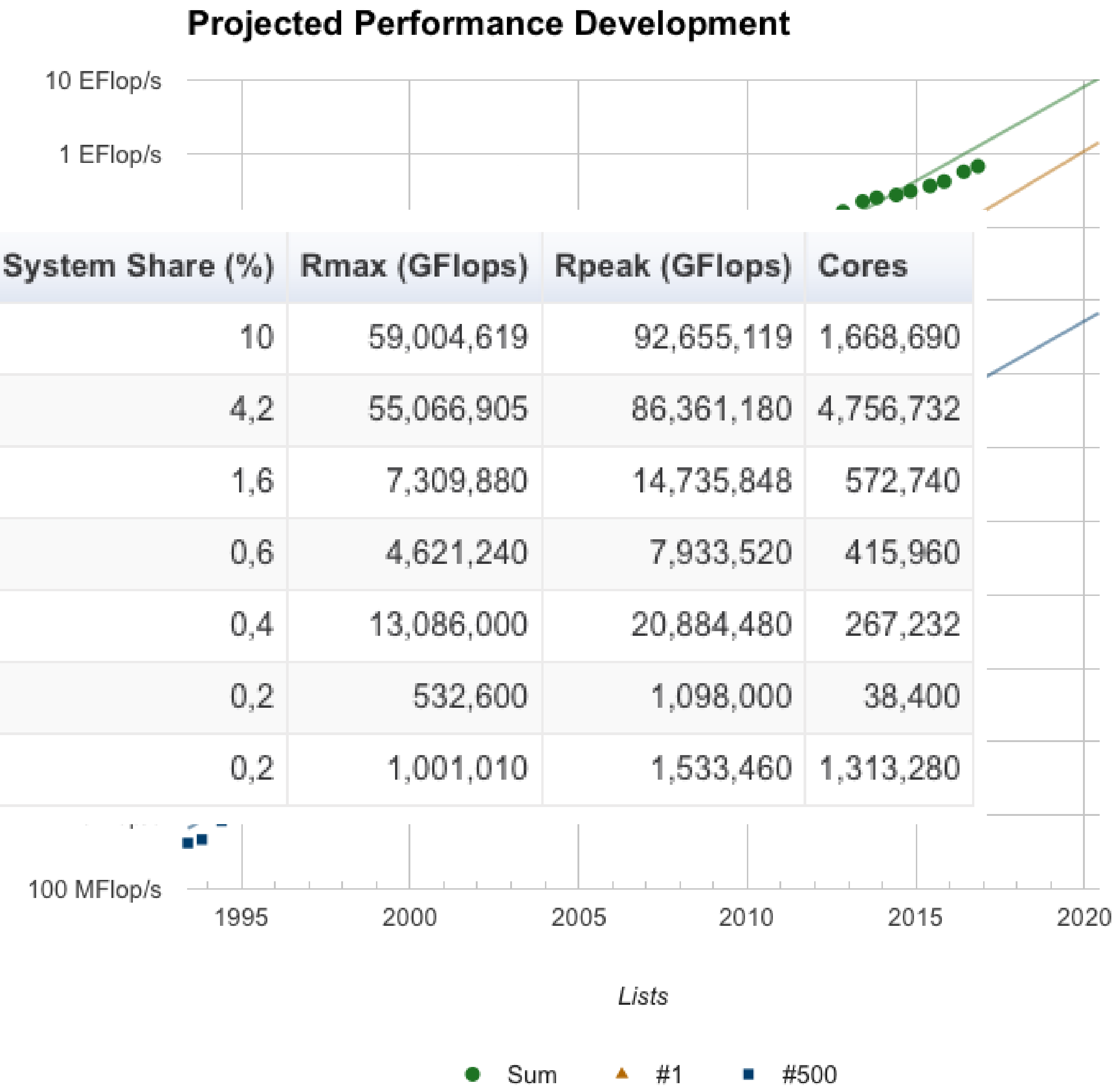
- HLRB-II (pre-SuperMUC): Top 500 06/2007: 56,5 Tflop/s
- SuperMUC Phase 1: Top 500 06/2012: 2897 Tflop/s

Coming up:

- **SuperMUC NG** (Next Generation) – Procurement on-going

Consulting the Top 500 List - www.top500.org

Rank	Site	System
1	National Supercomputing Center in Wuxi China	**Sunway TaihuLight** - Su MPP, Sunway SW26010 1.45GHz, Sunway NRCPC
2	National Super Computer Center in Guangzh China	**Tianhe-2 (MilkyWay-2)**
3	DOE/SC/ Laborato United St	
4	DOE/NN United St	
5	DOE/SC/ United St	
6	Joint Cer Perform Japan	
7	RIKEN A Computa Japan	
8	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint** - Cray XC50, X 2690v3 12C 2.6GHz, Arie interconnect , NVIDIA Te Cray Inc.
9	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Pov 16C 1.60GHz, Custom IBM
10	DOE/NNSA/LANL/SNL United States	**Trinity** - Cray XC40, Xeo 2698v3 16C 2.3GHz, Arie interconnect Cray Inc.



Accelerator/CP Family	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
Nvidia Kepler	50	10	59,004,619	92,655,119	1,668,690
Intel Xeon Phi	21	4,2	55,066,905	86,361,180	4,756,732
Nvidia Fermi	8	1,6	7,309,880	14,735,848	572,740
Hybrid	3	0,6	4,621,240	7,933,520	415,960
Nvidia Pascal	2	0,4	13,086,000	20,884,480	267,232
ATI Radeon	1	0,2	532,600	1,098,000	38,400
PEZY-SC	1	0,2	1,001,010	1,533,460	1,313,280

What‘s next? A View into the Future

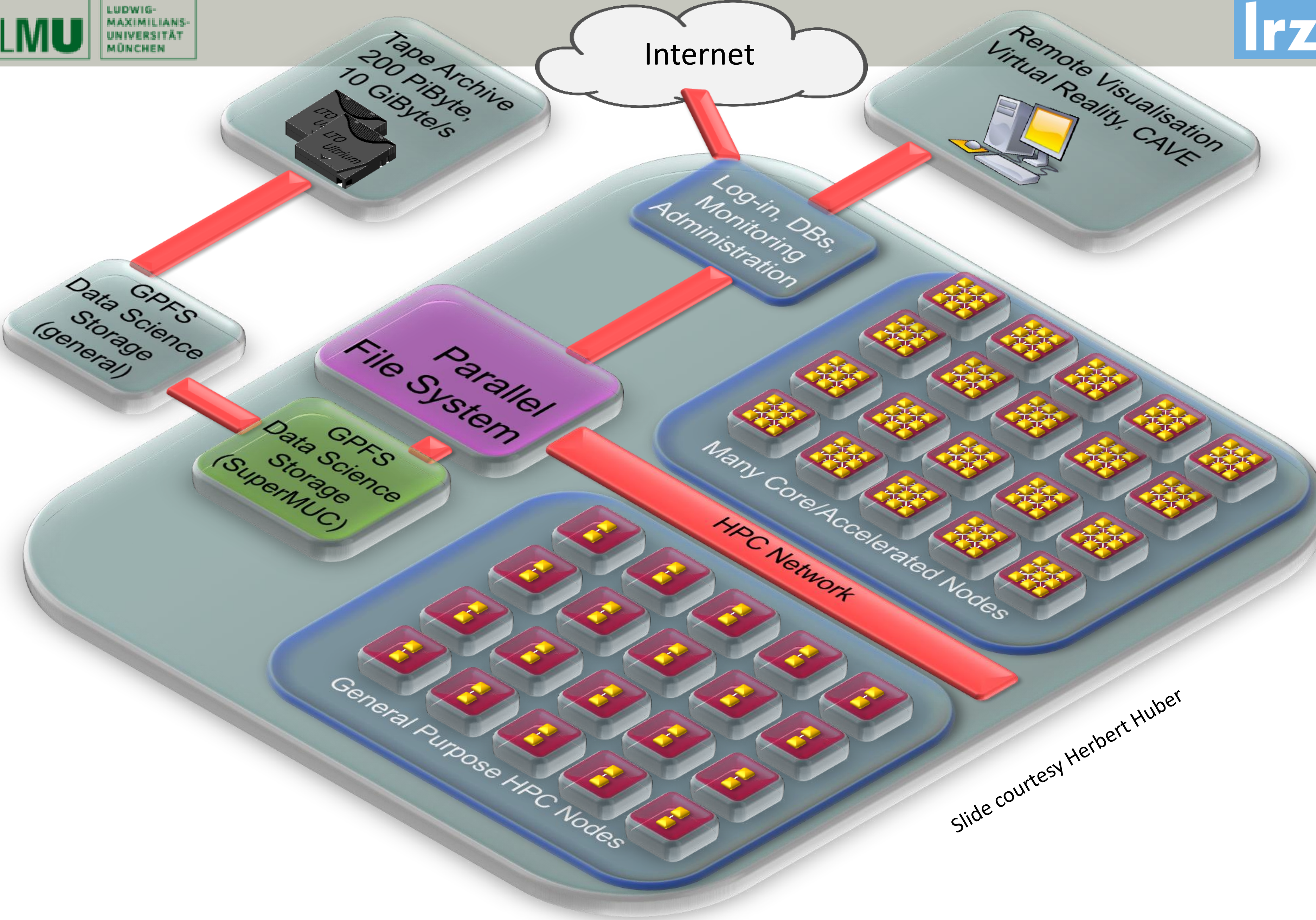
Until today:

- HLRB-II (pre-SuperMUC): Top 500 06/2007: 56,5 Tflop/s
- SuperMUC Phase 1: Top 500 06/2012: 2897 Tflop/s

Coming up:

- **SuperMUC NG** (Next Generation) – Procurement on-going

- **Selection criteria:**
  - LRZ application mix (compute,memory,bandwidth characteristics)
    - Number of cores
    - Memory per core
    - Interconnect
  - Accelerators (Manycore, GPGPU, ...)
  - Virtualization (Docker, Cloud, ...)
  - Workflow engines, HTC applications, ...
  - Power consumption (in total, over time, ...)


Internet
Tape Archive
200 PiByte,
10 GiByte/s
Remote Visualisation
Virtual Reality, CAVE
Log-in, DBs,
Monitoring
Administration
GPFS
Data Science
Storage
(general)
Parallel
File System
GPFS
Data Science
Storage
(SuperMUC)
Many Core/Accelerated Nodes
HPC Network
General Purpose HPC Nodes
Slide courtesy Herbert Huber

Conclusions

- Excellent research needs excellent tools
- Supercomputers provide the highest possible computational performance, interconnectivity and memory capacity

- The complexity of (super-)computers (such as SuperMUC NG) is steadily increasing (not only on the extreme scale)
- Demand of domain science drives computer science research to new frontiers

- Users need the possibility to execute (and optimize) their codes on the full size machines
- The **LRZ Partnership Initiative Computational Science (piCS)** tries to improve user support

http://www.sciencedirect.com/science/article/pii/S1877050914003433

piCS Cookbook

1. Choose focus topics to serve as lighthouse
   - National agreement within GCS: LRZ focuses on Environment (& Energy)
2. Choose user communities
   - Already active at LRZ?
   - Not active at LRZ?
3. Invite them for introductory piCS Workshops
   - Show faces & tour
   - Discussion on joint topics, requirements, interests, ...
4. Establish links between communities and specific points-of-contact
   - Whom to talk to, if there are questions?
   - When to talk to them? In general, as early as possible
   - Maybe, place people into the research groups (weekly, for a certain period)
5. Run joint lectures (e.g. hydrometeorology and computer science)
6. Apply for joint projects
7. Use **HPC Machines** efficiently to do science

The new SuperMUC petascale system and applications

Dieter Kranzlmüller
kranzlmueller@lrz.de

Contributions from: A. Bode, A. Stamatakis, L. Czech, A. Frank, M. Brehm, H.Huber, M. Bader, F. Jamitzky, A. Parodi, ...