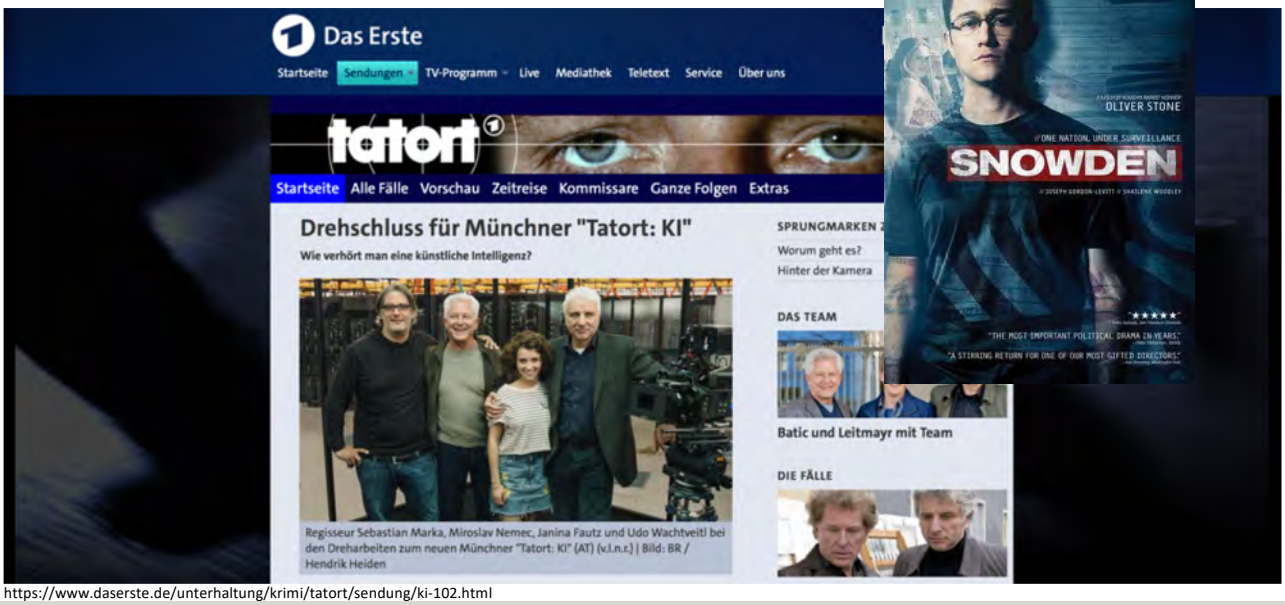
SuperMUC-NG

… und das Rennen um den schnellsten Computer der Welt

Dieter Kranzlmüller

Munich Network Management Team
Ludwig-Maximilians-Universität München (LMU)
&
Leibniz-Rechenzentrum (LRZ)
der Bayerischen Akademie der Wissenschaften

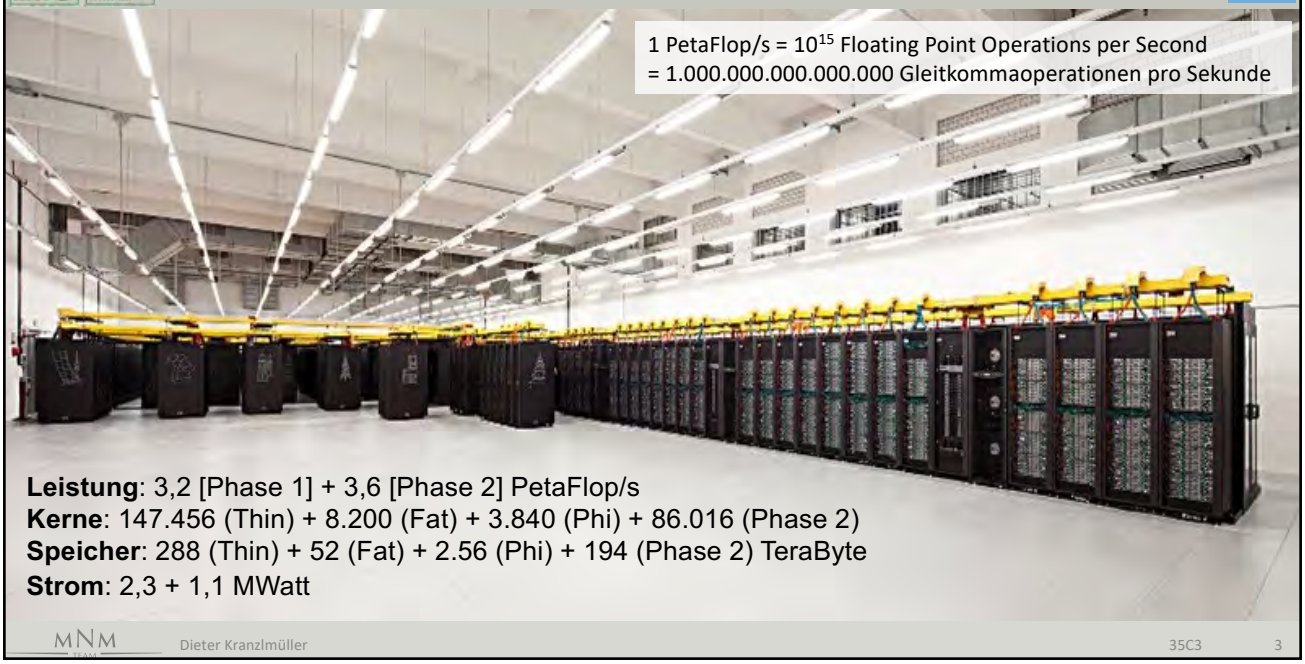
Tatort „Künstliche Intelligenz“



https://www.snowdenfilm.com

https://www.daserste.de/unterhaltung/krimi/tatort/sendung/ki-102.html

SuperMUC Phase 1 (2012) + 2 (2015)



1 PetaFlop/s = 10^{15} Floating Point Operations per Second
= 1.000.000.000.000.000 Gleitkommaoperationen pro Sekunde

Leistung: 3,2 [Phase 1] + 3,6 [Phase 2] PetaFlop/s
Kerne: 147.456 (Thin) + 8.200 (Fat) + 3.840 (Phi) + 86.016 (Phase 2)
Speicher: 288 (Thin) + 52 (Fat) + 2.56 (Phi) + 194 (Phase 2) TeraByte
Strom: 2,3 + 1,1 MWatt

Leistung: 26,7 PetaFlop/s
Kerne: 311.040 Intel Xeon
Speicher: 719 TeraByte

Gauss Centre for Supercomputing e.V. (GCS)

- Kooperation der nationalen Höchstleistungrechenzentren (HLRS, JSC, LRZ)
- Ziel: **Höchstleistungsrechner für die Wissenschaft in Deutschland**
- Finanzierung: **Bundesministerium für Bildung und Forschung** (BMBF) + Baden-Württemberg, **Bayern**, Nordrhein-Westfalen

Projekt: SiVeGCS - *Sicherstellung und erweiterte Verfügbarkeit der Supercomputing-Ressourcen des GCS im Rahmen der nationalen Höchstleistungsrechner-Infrastruktur (2017–2025)*

- Sicherstellung der Verfügbarkeit der leistungsfähigsten Höchstleistungsrechner in Deutschland und Europa,
- Bereitstellung modernster HPC-Technologien der höchsten Leistungsklasse für die Wissenschaft und Forschung
- Offener Zugang im wettbewerblichen Verfahren nach rein wissenschaftlichen Kriterien (Peer-Review Verfahren),
- Sicherstellung einer effizienten Nutzung,
- Effektive Ergebnisverwertung und
- umfassende Öffentlichkeitsarbeit.

LRZ: Hochparalleles System für ein **breites Anwendungsspektrum**, insbes. aus den Anwendungsgebieten Astro-, Geo-, Lebens- und Umweltwissenschaften sowie aus versch. Bereichen der Physik, der Strömungsmechanik und der Chemie

Beschaffung SuperMUC-NG – Wettbewerblicher Dialog

- 8.12.2016: Europäische Vergabeplattform SIMAP: Aufruf zur *Teilnahme am wettbewerblichen Dialog für die Vergabe des europäischen Höchstleistungsrechners SuperMUC-NG am LRZ*
  - Durchführung des Wettbewerblichen Dialogs
  - Bewerbungs- und Vertragsbedingungen
  - Description of Goods and Services
  - Decision Criteria and Benchmark Description
  - NDA-Benchmark Erklärung
  - Raum- und Kühlungspläne
- 23.01.2017: Bewerber erbringen Nachweis ihrer finanziellen und wirtschaftlichen Leistungsfähigkeit, benennen Referenzinstallationen und legen firmeneigene Konzepte zur Erhöhung der Energie- und Kühlungseffizienz dar

1. Runde des Wettbewerblichen Dialogs: 5 Firmen
2. Runde des Wettbewerblichen Dialogs: 2 Firmen

Beschaffung SuperMUC-NG – Description of Goods and Services

M 51: Online measurements of the direct current (DC) **energy** consumption of individual compute nodes must be provided by means of monotonically increasing counters.

The accuracy and measurement frequency must be specified.

Check here that this requirement will be fulfilled: [ ]

Mandatory

I 80: Tools for reading out the node level energy and power counters should deliver energy values in units of Joules and power values in units of Watts. No additional calibration or transformation of values should be needed.

Important

Reuse of system waste heat for the cooling of air or chilled water cooled system components will be positively evaluated.

T 28: Reuse of system waste heat to drive adsorption chillers is desired. If offered, the technical description of the implementation should be delivered and the cooling power that can be extracted should be specified.

[Insert text here]

Beschaffung SuperMUC-NG – Bewerbungs- und Vertragsbedingungen

- Erklärung Wettbewerblicher Dialog gemäß § 119 Abs. (1) und (6) GWB und § 14 Abs. (3) VgV sowie § 18 VgV

…

- **Total Cost of Ownership** (TCO):
  Summe aus Investitions-, Betriebs- und Wartungskosten darf die brutto zur Verfügung stehenden Mittel gemäß Anschreiben nicht überschreiten
- nur ein Hauptangebot je Bieter und keine Nebenangebote
- Installationszeitraum: Q3/2018 – Q4/2019

…

- Leistungsprüfung durch Leistungsmessungen
- Funktionsprüfung im Benutzerbetrieb

…

- Verhalten bei Bieterfragen

https://www.lrz.de/presse/ereignisse/2017-12-14_supermuc-ng_vertrag/

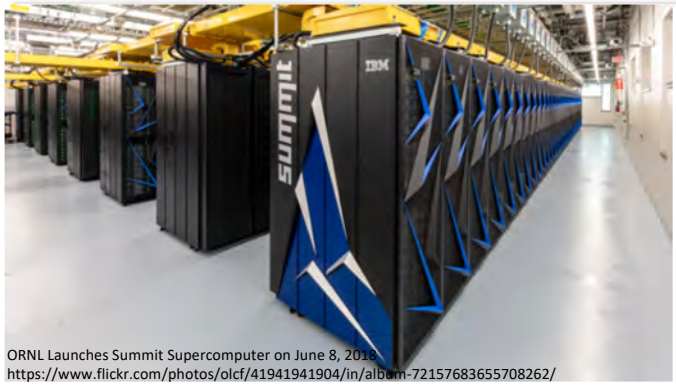
https://www.lrz.de/services/compute/supermuc/supermuc-ng/

ComputeNodes	Thin Nodes	Fat Nodes	Total (Thin + Fat)
Processor	Intel Skylake	Intel Skylake	Intel Skylake
Cores per Node	48	48	48
Memory per node (GByte)	96	768	NA
Number of Nodes	6,336	144	6,480
Number of Cores	304,128	8,912	311,040
PEAK @ nominal (PFlop/s)	26.3	0.6	26.9
Linpack (Pflop/s)	TBD	TBD	19.476
Memory (TByte)	608	111	719
Filesystems			
High Performance Parallel Filesystem	50 PB @ 500 GB/s		
Data Science Storage	20 PB @ 70 GB/s		
Home Filesystem	256 TB		
Infrastructure			
Cooling	Direct warm water cooling		
Waste Heat Reuse	Reuse for producing cold water with adsorption coolers		
Software			
Operating System	Suse Linux (SLES)		
Batch Scheduling System	SLURM		
High Performance Parallel Filesystem	IBM Spectrum Scale (GPFS)		
Programming Environment	Intel Parallel Studio XE GNU compilers OpenHPC Software Stack		
Message Passing	Intel MPI, (OpenMPI).		

Top 500 Supercomputer – www.top500.org – Platz 1 (Nov 2018)

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , **IBM** DOE/SC/Oak Ridge National Laboratory United States	2,397,824	143,500.0	200,794.9	9,783

Rpeak – Theoretische Maximalleistung
Rmax – Erreichte Maximalleistung im Linpack Benchmark



ORNL Launches Summit Supercomputer on June 8, 2018
https://www.flickr.com/photos/olcf/41941941904/in/album-72157683655708262/

Top 500 Supercomputer – www.top500.org – Platz 1-5 (Nov 2018)

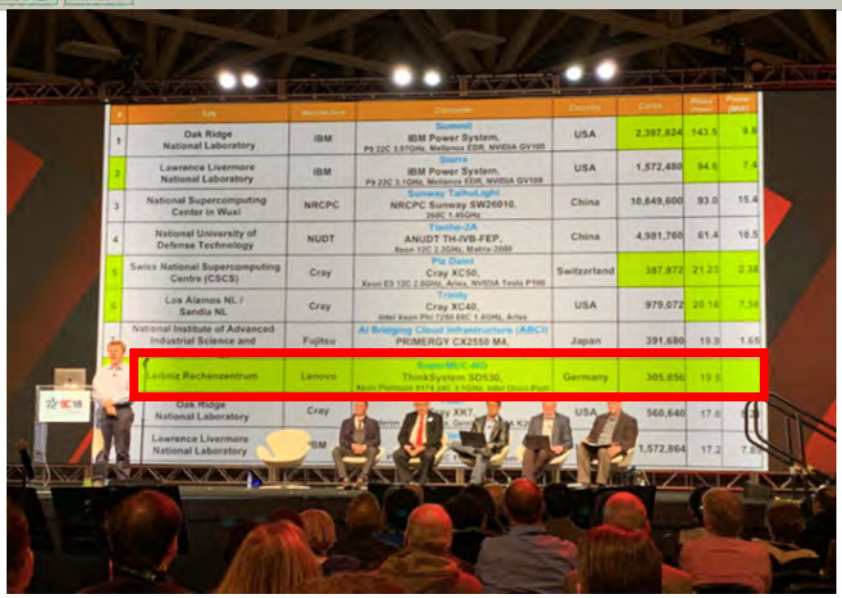
Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , **IBM** DOE/SC/Oak Ridge National Laboratory United States	2,397,824	143,500.0	200,794.9	9,783
2	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , **IBM / NVIDIA / Mellanox** DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
3	**Sunway TaihuLight** - Sunway MPI, Sun NRCPC National Supercomputing Center in Wu China				
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel X Express-2, Matrix-2000 , **NUDT** National Super Computer Center in Gua China				
5	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , **Cray Inc.** Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	27,154.3	2,384

https://insidehpc.com/2018/11/update-piz-daint-fastest-supercomputer-europe/

In this video from SC18 in Dallas, Michele De Lorenzi from CSCS in Switzerland provides an update on Piz Daint, the fastest supercomputer in Europe.

"Recently upgraded with two additional cabinets full of NVIDIA V100 GPUs, the Cray XC50 system comes in at #5 in the world with 21.23 Petaflops of performance on the LINPACK benchmark."

Supercomputing 2018, Dallas, Texas – https://sc18.supercomputing.org



Top 500 Birds of a Feather Session, 13. November 2018

Autoren der TOP500 Liste:

- Erich Strohmaier (NERSC/Lawrence Berkeley National Laboratory),
- Jack Dongarra (University of Tennessee, Knoxville),
- Horst Simon (NERSC/Lawrence Berkeley National Laboratory),
- Martin Meuer (Prometeus),
- Hans Meuer (Universität Mannheim) [1993-2014]

Top 500 Supercomputer – www.top500.org – SuperMUC-NG (Nov 2018)

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
8	SuperMUC-NG - ThinkSystem SD530, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path Lenovo Leibniz Rechenzentrum Germany	305,856	19,476.6	26,873.9	

Der LINPACK Benchmark - https://www.top500.org/project/linpack/

Löser für lineares Gleichungssystem (Jack Dongarra):

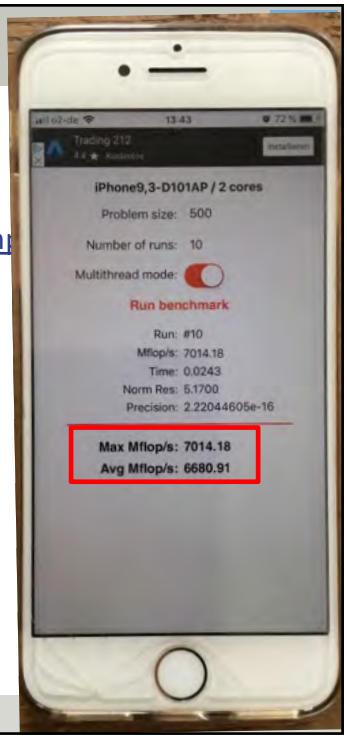
Rank	Pre	First	Firs	Name	Computer	Site	Manufactu	Country	Year	Segment	Total Co	Accelera	Rmax [TFl	Rpeak [TFl	Nmax
1	1	51	1	Summit	IBM Power S	DOE/SC/Oak	IBM	United States	2018	Research	2397824	2196480	143500	200794,88	16693248
2	3	51	3	Sierra	IBM Power S	DOE/NNSA/L	IBM / NVIDIA	United States	2018	Research	1572480	1382400	94640	125712	11902464
3	2	47	1	Sunway Taih	Sunway MPP	National Sup	NRCPC	China	2016	Research	1,1E+07		93014,5939	125435,904	12288000
4	4	41	1	Tianhe-2A	TH-IVB-FEP (	National Sup	NUDT	China	2018	Research	4981760	4554752	61444,5	100678,664	9773000
5	6	40	114	Piz Daint	Cray XC50, X	Swiss Nationa	Cray Inc.	Switzerland	2017	Research	387872	319424	21230	27154,3	3743232
6	9	46	6	Trinity	Cray XC40, X	DOE/NNSA/L	Cray Inc.	United States	2017	Research	979072		20158,7	41461,15	12353536
7	5	51	5	AI Bridging C	PRIMERGY C	National Instit	Fujitsu	Japan	2018	Research	391680	348160	19880	32576,6349	2918016
8		52	8	SuperMUC-N	ThinkSystem	Leibniz Rech	Lenovo	Germany	2018	Academic	305856		19476,6	26873,856	8402688

- 64 Bit Gleitkommaoperation (Addition oder Multiplikation)
- "*Performance of Various Computers Using Standard Linear Equations Software*", Jack Dongarra, University of Tennessee, Knoxville TN, 37996, Computer Science Technical Report Number CS - 89 – 85, today's date, http://www.netlib.org/benchmark/performance.ps
- **Amdahlsches Gesetz**: Geschwindigkeitszuwachs ist durch sequentiellen Teil beschränkt
- **Gustafsons Gesetz**: genügend großes Problem kann effizient parallelisiert werden

Do it yourself – LINPACK Benchmark

App für Ihr Smartphone:

- iOS: https://itunes.apple.com/de/app/linpack/id380883195?mt=8
- Android: https://play.google.com/store/apps/details?id=com.sqi.lin
- Google „Linpack App"



iPhone 7:

- Max MFlop/s: 7014,18
- Avg MFlop/s: 6680,91

Cray Y-MP Supercomputer (1988):

- Max MFlop/s: 2664

By Dave Pape [Public domain], from Wikimedia Commons https://commons.wikimedia.org/wiki/File:Cray_Y-MP_GSFC.jpg

Top 500 Supercomputer – Leistungsentwicklung

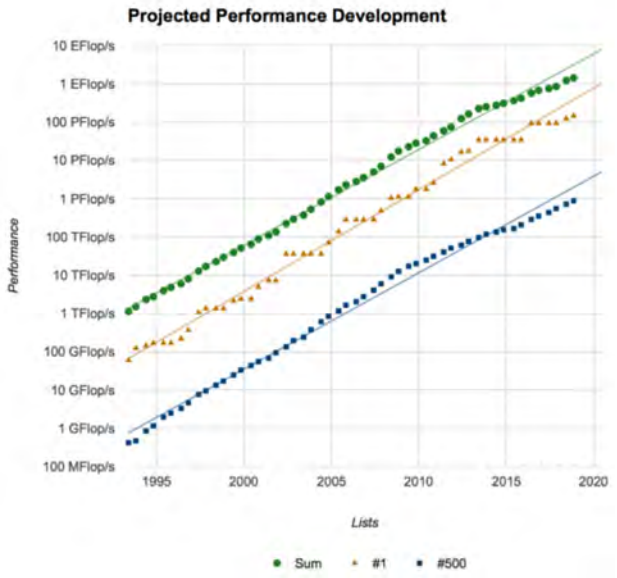
https://www.top500.org/statistics/perfdevel/

Platz 1:

- Juni 1993:
  59,7 GFlop/s (CM-5) = $59{,}7 * 10^9$
- Nov 2018:
  122,3 PFlop/s (Summit) = $122{,}3 * 10^{15}$

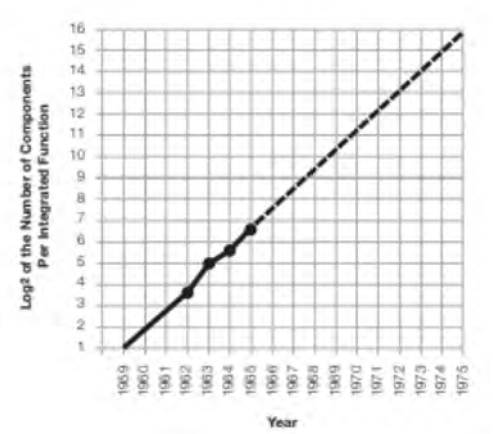
- Leistungssteigerung:
  $59{,}7 * 10^9$ ➔ $122{,}3 * 10^{15}$ = ca. $2 * 10^6$
  Zeitraum: 25 Jahre, 5 Monate

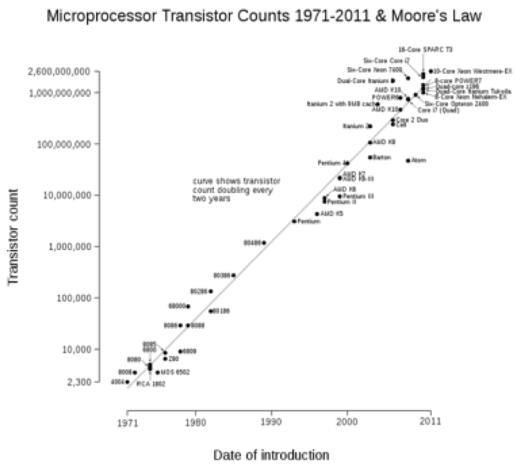
- Exponentielles Wachstum
  ==> Mooresches Gesetz



Mooresches Gesetz

- Gordon E. Moore, „Cramming more components onto integrated circuits“, Electronics Magazine, p.4, 19.04.1965



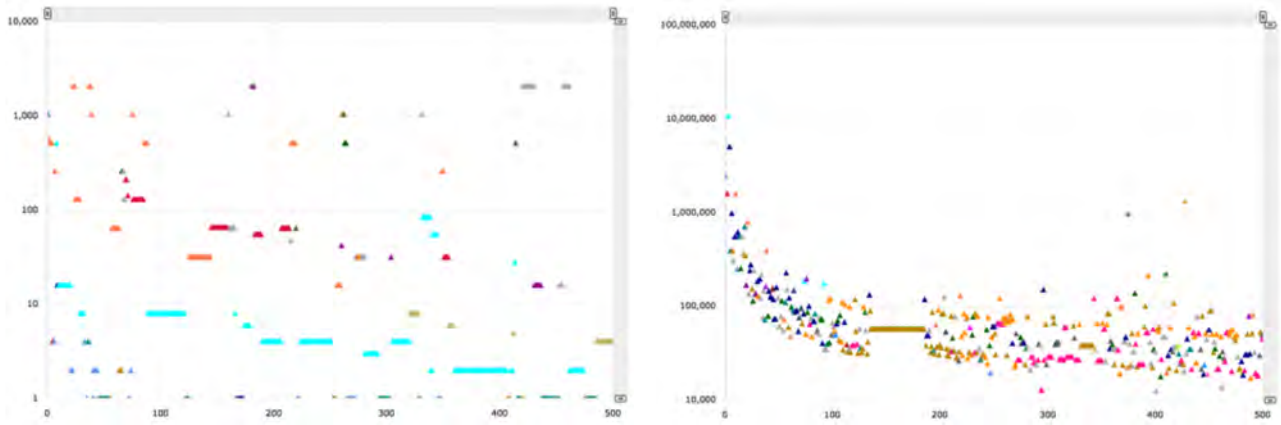


Von Wgsimon - Eigenes Werk, CC BY-SA 3.0
https://commons.wikimedia.org/w/index.php?curid=15193542

Skalierung – Anzahl der Rechenkerne

https://www.top500.org/statistics/efficiency-power-cores/

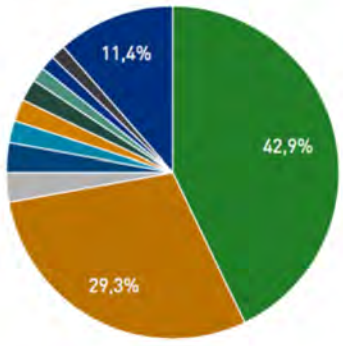
1993: 2018:



Top 500 Sublist Generator

https://www.top500.org/statistics/list/

Accelerator/Co-Processor System Share



- NVIDIA Tesla P100
- NVIDIA Tesla V100
- NVIDIA Tesla K20x
- NVIDIA Tesla K80
- NVIDIA Tesla V100 SXM2
- NVIDIA Tesla K40
- NVIDIA Tesla P100 NVLink
- NVIDIA Volta GV100
- Intel Xeon Phi 5120D
- Intel Xeon Phi 5110P
- Others

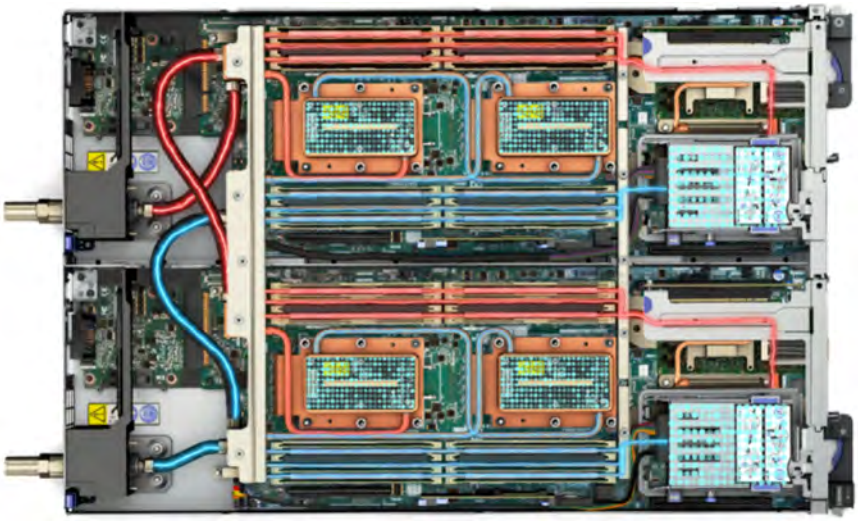
		Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
		12	106,675,088	182,288,507	2,441,572
		8,2	65,671,290	136,467,348	1,645,796
		0,8	24,615,000	38,926,835	731,712
		0,8	8,461,770	13,754,030	271,870
		0,6	31,937,000	52,026,313	609,840
		0,6	8,824,090	14,612,320	201,328
		0,6	10,108,000	15,759,869	176,868
		0,4	238,140,000	326,506,880	3,970,304
		0,4	2,571,200	4,099,204	122,512
		0,4	3,515,893	4,729,128	272,136
NVIDIA 2050	2	0,4	3,837,000	7,685,300	307,008
Matrix-2000	1	0,2	61,444,500	100,678,664	4,981,760
NVIDIA 2070	1	0,2	901,900	1,700,210	78,660
PEZY-SCnp	1	0,2	1,001,010	1,533,460	1,313,280

SuperMUC-NG Prozessoren (Intel Xeon Platinum 8174 Scalable Processors)

CPU	INTEL thin	INTEL fat
Processor type	Intel Skylake	Intel Skylake
Total number of nodes of this type	6336	144
Number of islands with this node type	8	1
Number of processors (sockets) per node	2	2
Number of cores per processor	24	24
Memory per node (GByte)	96	768
Nominal frequency (GHz)	3,10	3,10
Peak frequency (GHz)	3,90	3,90
Floating point operations per core per clock (Fused MulAdd counts as 2)	32	32
Peak floating point performance of one core @ nominal (GFlop/s)	99,20	99,20
Peak AVX-512 floating point performance of one node (all cores active)	3533	3533
Total	INTEL thin	INTEL fat
Total CPU Cores	304128	6912
Total Peak PFlop/s @ maximum AVX-512 frequency (all cores active)	22,38	0,51
SUM PEAK PFlop/s @ maximum AVX-512 frequency (all cores active)	22,89	
Total Nodes	6480	

SuperMUC-NG Knoten – mit Heißwasserkühlung

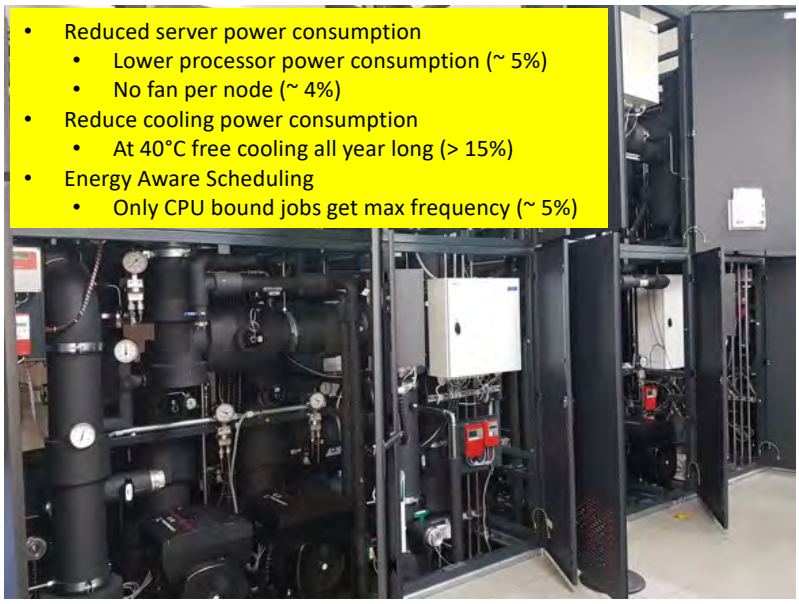
Lenovo Think System SD650:

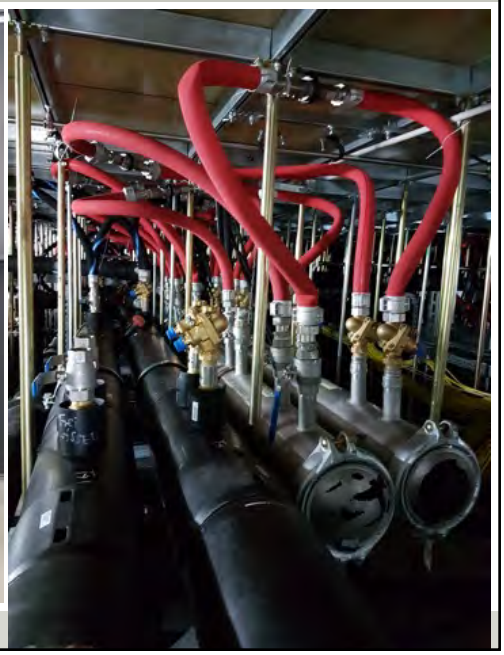


Bilder: Lenovo – Produkt: https://www.lenovo.com/de/de/data-center/servers/high-density/ThinkSystem-SD650/p/77XX7DSSD65

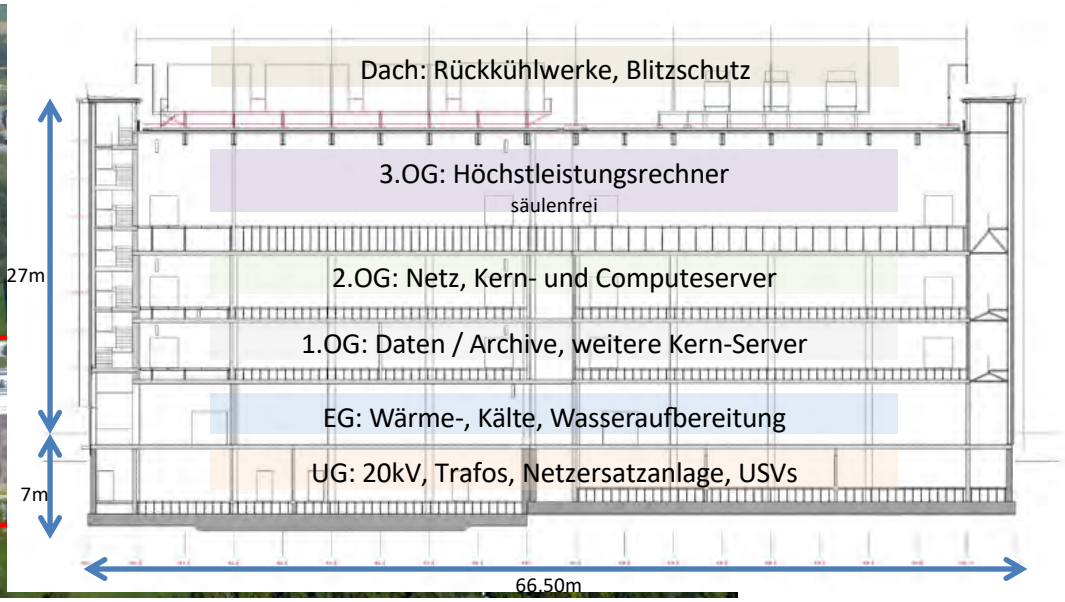
SuperMUC-NG Cool Manager – Kühlwasser 3000 l/h pro Rack

- Reduced server power consumption
  - Lower processor power consumption (~ 5%)
  - No fan per node (~ 4%)
- Reduce cooling power consumption
  - At 40°C free cooling all year long (> 15%)
- Energy Aware Scheduling
  - Only CPU bound jobs get max frequency (~ 5%)



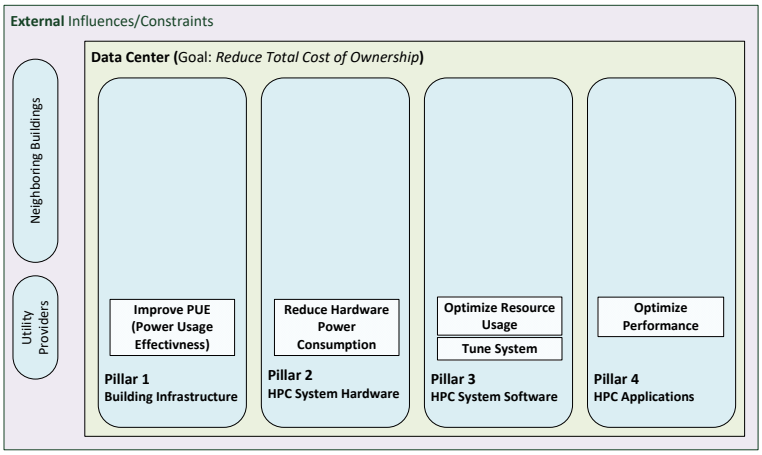


Leibniz-Rechenzentrum am Forschungsgelände Garching b. München

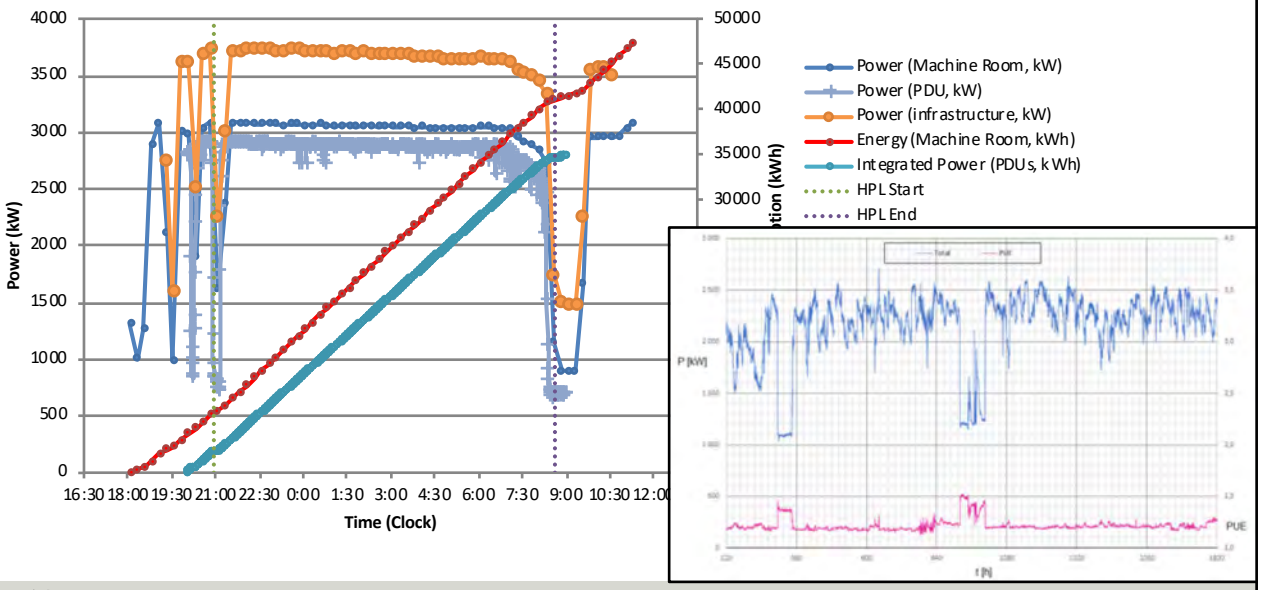


Gesamtheitliche Sicht der Infrastruktur – 4 Säulenmodell

Torsten Wilde, Axel Auweter, Hayk Shoukourian
„The 4 Pillar Framework for energy efficient HPC data centers“
Comput Sci Res Dev (2014) 29: 241.
https://doi.org/10.1007/s00450-013-0244-6



SuperMUC Phase 1 – HPL Lastkurven



LINPACK als Test für die Zuverlässigkeit der Infrastruktur

Stromversorgung: Trafo + Notstromaggregat + USV

Top 500 Supercomputer – www.top500.org – SuperMUC-NG

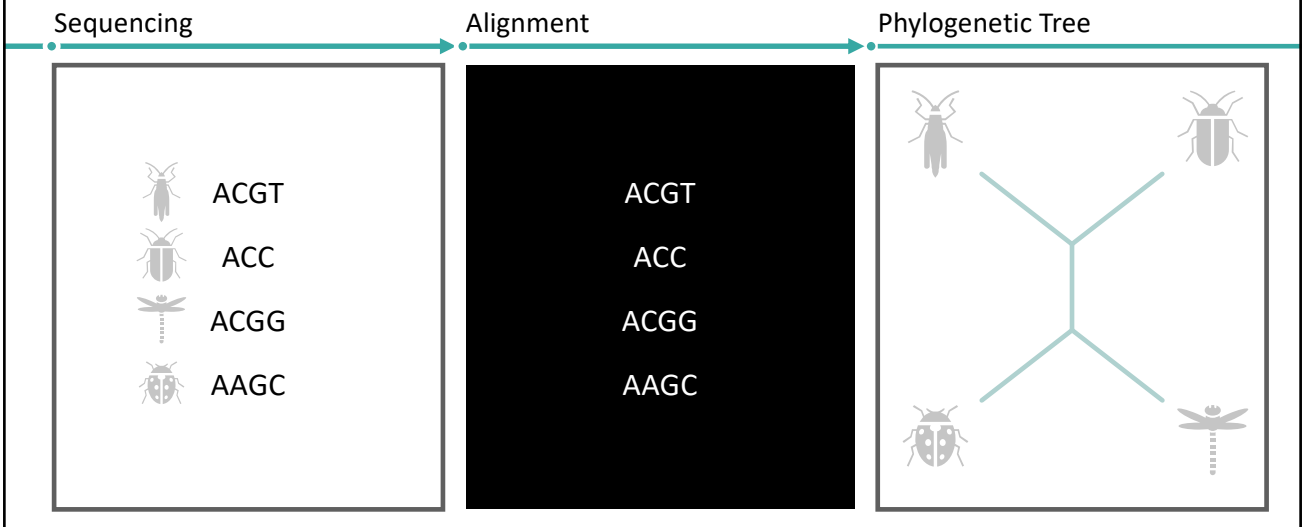
Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
8	**SuperMUC-NG** - ThinkSystem SD530, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , Lenovo Leibniz Rechenzentrum Germany	305,856	19,476.6	26,873.9	

GRAPH 500

#1 Data-intensive applications **SSSP benchmark**

#5 Data-intensive applications **BFS benchmark**

Bestimmung von Phylogenetischen Bäumen (A. Stamatakis, HITS)



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

Anzahl der möglichen Bäume für 150 Arten

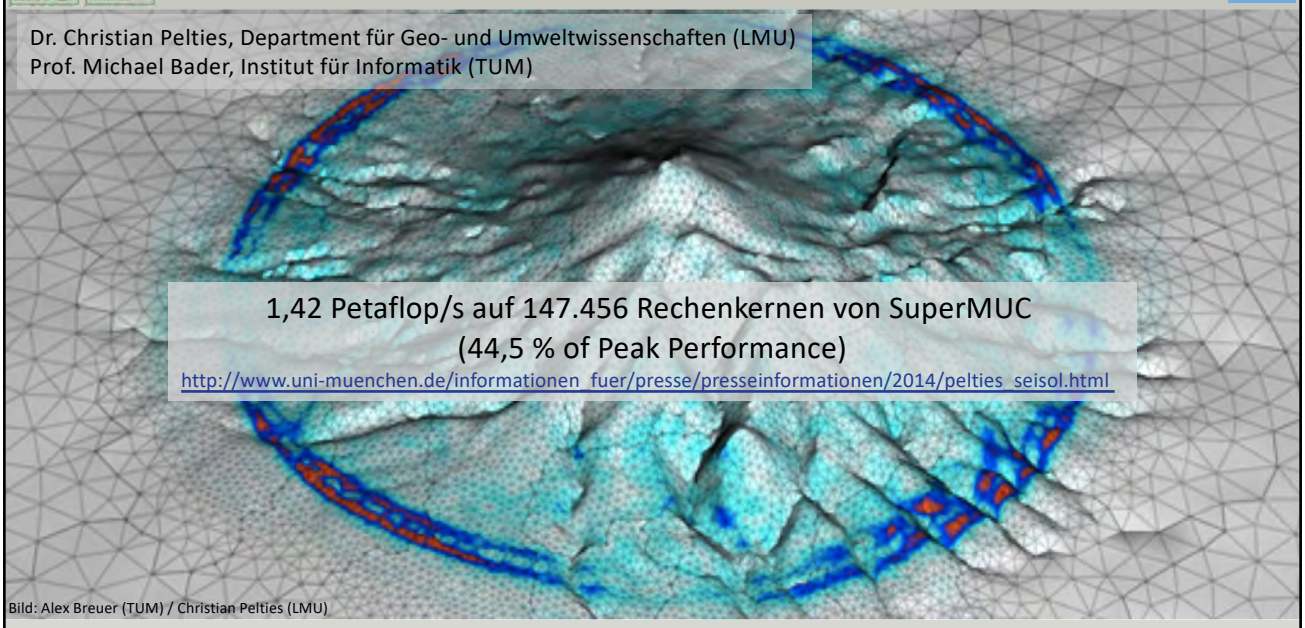
4226517247809112252219618802377042809718932383449882294285747988083143403217875902453679849195116830764946928674148027385702212982924284576878148734552121861861600804474608426626044448936698500560246811618644126422742544072667661492790654064936029763974619174693267509311908892414066940546035766601562 5

$\approx 4.22 \times 10^{301}$

Ergebnisse der Berechnungen auf SuperMUC-NG

SeisSol – Numerische Simulation von seismischen Wellenphänomenen

Dr. Christian Pelties, Department für Geo- und Umweltwissenschaften (LMU)
Prof. Michael Bader, Institut für Informatik (TUM)



1,42 Petaflop/s auf 147.456 Rechenkernen von SuperMUC
(44,5 % of Peak Performance)
http://www.uni-muenchen.de/informationen_fuer/presse/presseinformationen/2014/pelties_seisol.html

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

LRZ Anwendungsspektrum

Computational Fluid Dynamics
Optimisation of turbines/wings, noise reduction

Material Science
Semiconductors

Fusion
Plasma in a future fusion reactor (ITER)

Chemistry
Catalytic reactions

Astrophysics
Origin and evolution of stars and galaxies

Medicine and Medical Engineering
Blood flow, aneurysms, air conditioning

Solid State Physics
Superconductivity, surface properties

Biophysics
Properties of viruses, genome analysis

Geophysics
Earth quake scenarios

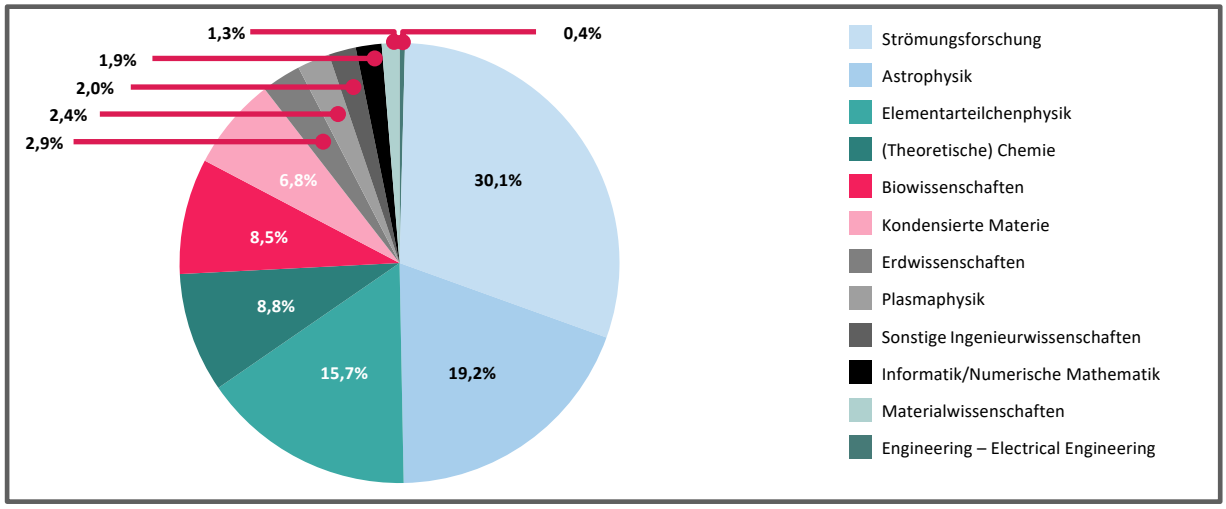
Climate research
Currents in oceans

...

...

Verteilung der Rechenzeit nach Fachgebieten

SuperMUC Phase 1 und Phase 2



LRZ Anwendungsspektrum

Computational Fluid Dynamics
Optimisation of turbines/wings, noise reduction

Fusion
Plasma in a future fusion reactor (ITER)

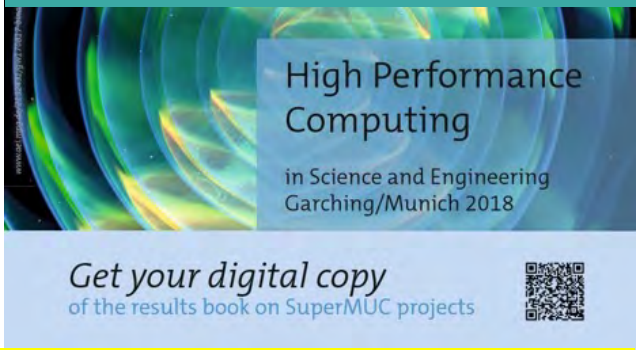
Astrophysics
Origin and evolution of stars and galaxies

Solid State Physics
Superconductivity, surface properties

Geophysics
Earth quake scenarios

...

> 7.6 billion compute hours consumed
> 5.6 million jobs processed
> 750 research projects carried out
> 1,995 researchers as clients



SuperMUC-NG Next Gen Science Symposium
https://www.lrz.de/presse/ereignisse/SuperMUC-NG-Science-Symposium/

Top 500 Supercomputer – www.top500.org – SuperMUC-NG

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
8	**SuperMUC-NG** - ThinkSystem SD530, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , **Lenovo** Leibniz Rechenzentrum Germany	305,856	19,476.6	26,873.9	

Bild: Morio, Nico Rosberg 2013 Catalonia test (19-22 Feb) Day 3.jpg
https://commons.wikimedia.org/wiki/File:Nico_Rosberg_2013_Catalonia_test_(19-22_Feb)_Day_3.jpg

Fotos: Minh Chung, Cong-Hoa-Street, Ho Chi Minh City, Vietnam

Letzte Worte

Worüber ich nicht sprechen konnte, weil die Zeit nicht gereicht hat:

- Der nächste Schritt: Exascale – die Exaflop/s Barriere
  Wer kommt zuerst an: China, USA oder doch Japan?
- Der europäische Höchstleistungsrechner (EuroHPC)
  Wie die EU Kommission in die Top 3 vorstoßen will?
- Die europäische Prozessorintiative
  Warum einen eigenen Prozessor bauen möchte?

…

- Nutzung der Rechnerabwärme – Absorptionskältemaschinen am LRZ
  Wir brauen Bier mit dem heißen Wasser aus SuperMUC?

SuperMUC-NG am Leibniz-Rechenzentrum (LRZ)

Tag der offenen Tür
am Forschungsgelände Garching b. München
➔ Oktober 2019
https://www.lrz.de/wir/tag-der-offenen-tuer/

SuperMUC-NG Next Gen Science Symposium:
https://www.lrz.de/presse/ereignisse/SuperMUC-NG-Science-Symposium/

SuperMUC Berichtsband:
https://www.lrz.de/services/compute/supermuc/magazinesbooks/

Dieter Kranzlmüller
kranzlmueller@lrz.de