

On Exascalng Experience while dealing with Large Applications

Dieter Kranzlmüller

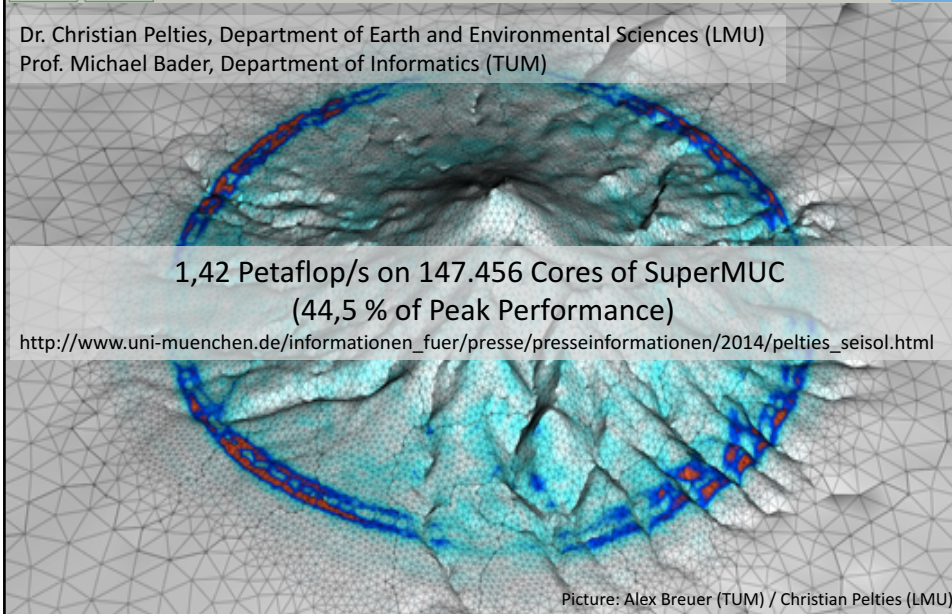
Munich Network Management Team
Ludwig-Maximilians-Universität München (LMU) &
Leibniz Supercomputing Centre (LRZ)
of the Bavarian Academy of Sciences and Humanities



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)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

SuperMUC @ LRZ

lrz



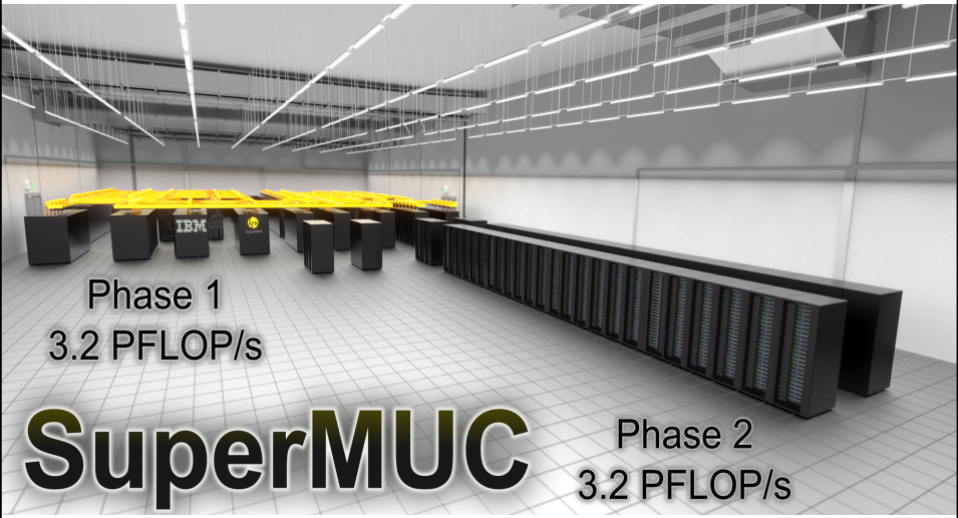
Video: SuperMUC rendered on SuperMUC by LRZ
<http://youtu.be/OIAS6iiqWrQ>

MNM D. Kranzmüller HPC Summit 2017 3

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

SuperMUC Phase 1 + 2

lrz



Phase 1
3.2 PFLOP/s

SuperMUC

Phase 2
3.2 PFLOP/s

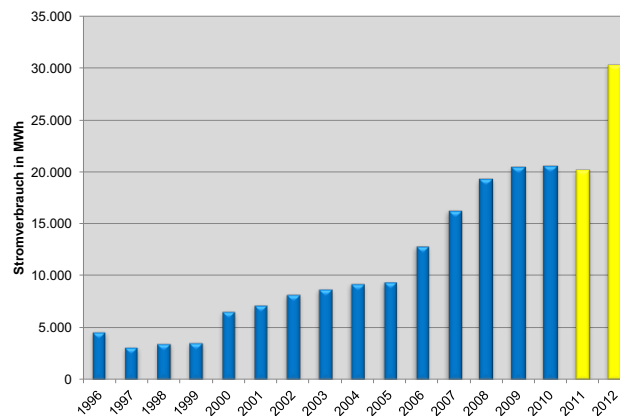
MNM D. Kranzmüller HPC Summit 2017 4

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






Photos: Torsten Bloth, Lenovo




High Energy Efficiency




- ✓ 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
 - Total annual savings of ~2 Mio. € for SuperMUC Phase 1 and 2

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 + 3.6 Pflop/s	229960




1st LRZ Extreme Scale Workshop


- July 2013:
 - 1st LRZ Extreme Scale Workshop**
- Participants:
 - 15 international projects
- Prerequisites:
 - Successful run on 4 islands (32768 cores)
- Participating Groups (Software packages):
 - LAMMPS, VERTEX, GADGET, WaLBerla, BQCD, Gromacs, APES, SeisSol, CIAO
- Successful results (> 64000 Cores):
 - Invited to participate in PARCO Conference (Sept. 2013) including a publication of their approach


 D. Kranzmüller
 HPC Summit 2017 9



1st LRZ Extreme Scale Workshop


- Regular SuperMUC operation
 - 4 Islands maximum
 - Batch scheduling system
- Entire SuperMUC reserved 2,5 days for challenge:
 - 0,5 Days for testing
 - 2 Days for executing
 - 16 (of 19) Islands available
- Consumed computing time for all groups:
 - 1 hour of runtime = 130.000 CPU hours
 - 1 year in total


 D. Kranzmüller
 HPC Summit 2017 10

LMU		LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN		Results (Sustained TFlop/s on 128000 cores)		lrz
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	

MNM D. Kranzmüller HPC Summit 2017 11

- | LMU | | LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | | Extreme Scaling Continued | | lrz |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------|--|---------------------------|--|-----|
| <ul style="list-style-type: none"> ■ Lessons learned → Stability and scalability ■ LRZ Extreme Scale Benchmark Suite (LESS) will be available in two versions: public and internal ■ All teams will have the opportunity to run performance benchmarks after upcoming SuperMUC maintenances ■ 2nd LRZ Extreme Scaling Workshop → 2-5 June 2014 <ul style="list-style-type: none"> – Full system production runs on 18 islands with sustained Pflop/s (4h SeisSol, 7h Gadget) – 4 existing + 6 additional full system applications – High I/O bandwidth in user space possible (66 GB/s of 200 GB/s max) – Important goal: minimize energy*runtime (3-15 W/core) ■ 3rd Extreme Scale-Out with new SuperMUC Phase 2 | | | | | | |
- MNM D. Kranzmüller HPC Summit 2017 12

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Extreme Scale-Out SuperMUC Phase 2

- 12 May – 12 June 2015 (30 days)
- Selected Group of Early Users

- Nightly Operation: general queue max 3 islands
- Daytime Operation: special queue max 6 islands (full system)

- Total available: 63,432,000 core hours
- Total used: 43,758,430 core hours (Utilisation: 68.98%)

Lessons learned (2015):

- Preparation is everything
- Finding Heisenbugs is difficult
- MPI is at its limits
- Hybrid (MPI+OpenMP) is the way to go
- I/O libraries getting even more important

D. Kranzmüller

HPC Summit 2017 13

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

4th Extreme Scale Workshop 2016

- 4 Day Workshop (29 February – 3 March 2016)
- 13 Projects:

	Application	Field	Institution	PI
1	INDEXA	CDF	TU München	M. Kronbichler
2	MPAS	Climate Science	KIT	D. Heinzler
3	Inhouse	Material Science	TU Dresden	F. Ortman
4	HemeLB	Life Science	UC London	P. Coveney
5	KPM	Chemistry	FAU Erlangen	M. Kreutzer
6	SWIFT	Cosmology	U Durham	M. Schaller
7	LISO	CFD	TU Darmstadt	S. Kraheberger
8	ILDBC	Lattice Boltzmann	FAU Erlangen	M. Wittmann
9	Walberla	Lattice Boltzmann	FAU Erlangen	Ch. Godenschwager
10	GASPI	Framework	ITWM Kaiserslautern	M. Kühn
11	GADGET	Cosmology	LMU München	K. Dolag
12	VERTEX	Astrophysics	MPI for Astrophysics	T. Melson
13	PSC	Plasma	LMU München	K. Bamberg

- 147,456 cores in 9216 Nodes
- 14.1 Mio CPUh
- Max Time per Job 6h
- Daily and nightly operation mode

D. Kranzmüller

HPC Summit 2017 14

VERTEX: Simulation Code for Supernova Explosions (plasma + neutrino dynamics)

A. Marek and Team (Max Planck-Institute for Astrophysics, Garching)

Finalists:

- INDEXA
- PSC
- waLBerla
- VERTEX

Leibniz Extreme Scaling Award
Extreme Scale Workshop 2016@LRZ

Quelle
Kein St
Für HR

Motivate your users!

MNM D. Kranzmüller HPC Summit 2017 15

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

LRZ Scaling Workshop: Emergent Applications lrz

May 2017:

- 3 Day hands-on coding workshop to enhance the scalability of users' applications beyond one SuperMUC island
- Goal: Scale up to 8 islands of SuperMUC phase 1
- Teaming up with LRZ, Lenovo, IBM, Intel and Alinea

6 successful applicants:

■ MPAS	Climate research
■ VLASOV6D	Plasma physics
■ ECHO	Astrophysics
■ BFPS	Turbulence research
■ MGLET	CFD
■ TERRA-NEO	Geo Science

MNM D. Kranzmüller HPC Summit 2017 16



Phase 1 (IBM System x iDataPlex):

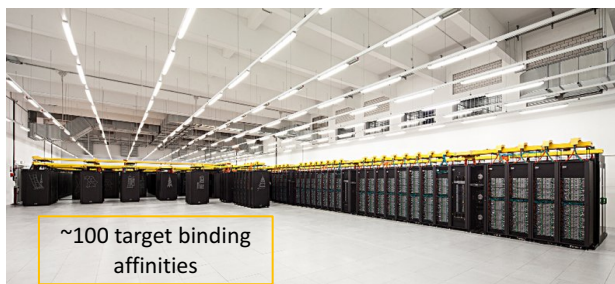
- 3.2 PFlops peak performance
- 9216 IBM iDataPlex dx360M4 nodes in 18 compute node islands
- 2 Intel Xeon E5-2680 processors and 32 GB of memory per compute node
- 147,456 compute cores
- Network Infiniband FDR10 (fat tree)

Phase 2 (Lenovo NeXtScale WCT):

- 3.6 PFlops peak performance
- 3072 Lenovo NeXtScale nx360M5 WCT nodes in 6 compute node islands
- 2 Intel Xeon E5-2697v3 processors and 64 GB of memory per compute node
- 86,016 compute cores
- Network Infiniband FDR14 (fat tree)

Common GPFS file systems with 10 PB and 5 PB usable storage size respectively
 Common programming environment
 Direct warm-water cooled system technology

- Running on all cores of SuperMUC Phase1+2



- 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 Projects COMPAT and MAPPER
<http://www.compat-project.eu>

On Exascaling Experience while dealing with Large Applications

Dieter Kranzmüller
kranzmueller@lrz.de

Photo: Karl Behler

