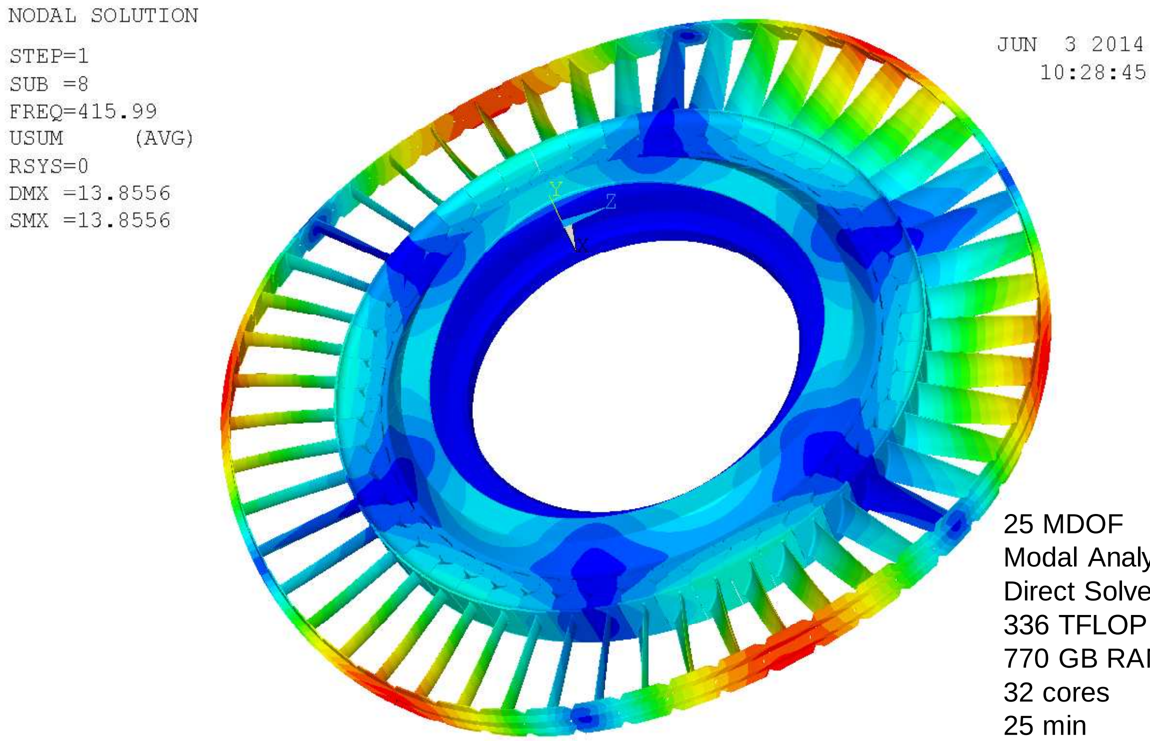
TFLOP Performance for ANSYS Mechanical 2.0



25 MDOF
Modal Analysis
Direct Solver
336 TFLOP
770 GB RAM
32 cores
25 min

Dr. Herbert Güttler

MicroConsult Engineering GmbH
Holunderweg 8
89182 Bernstadt
www.microconsult-engineering.de

Contents

ANSYS Mechanical performance requirements

What can be achieved using CPUs and GPUs

Workstation & Cluster, multiple GPUs

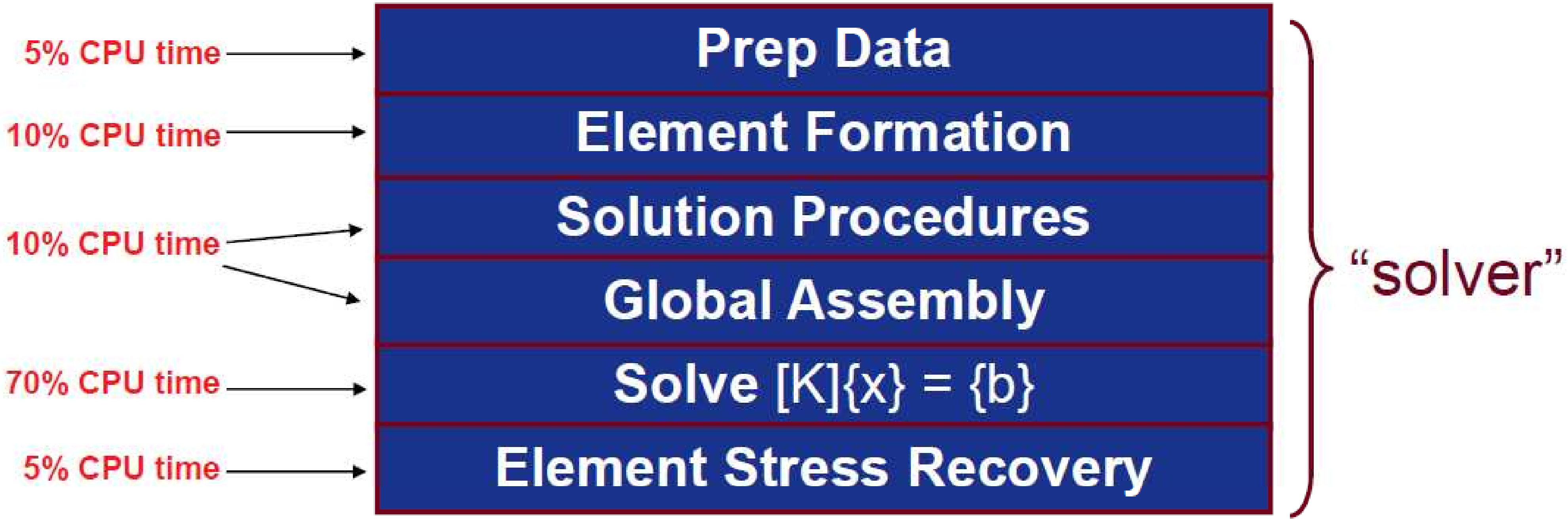
Cluster performance & scaling

Application examples

Sandy Bridge and Ivy Bridge results

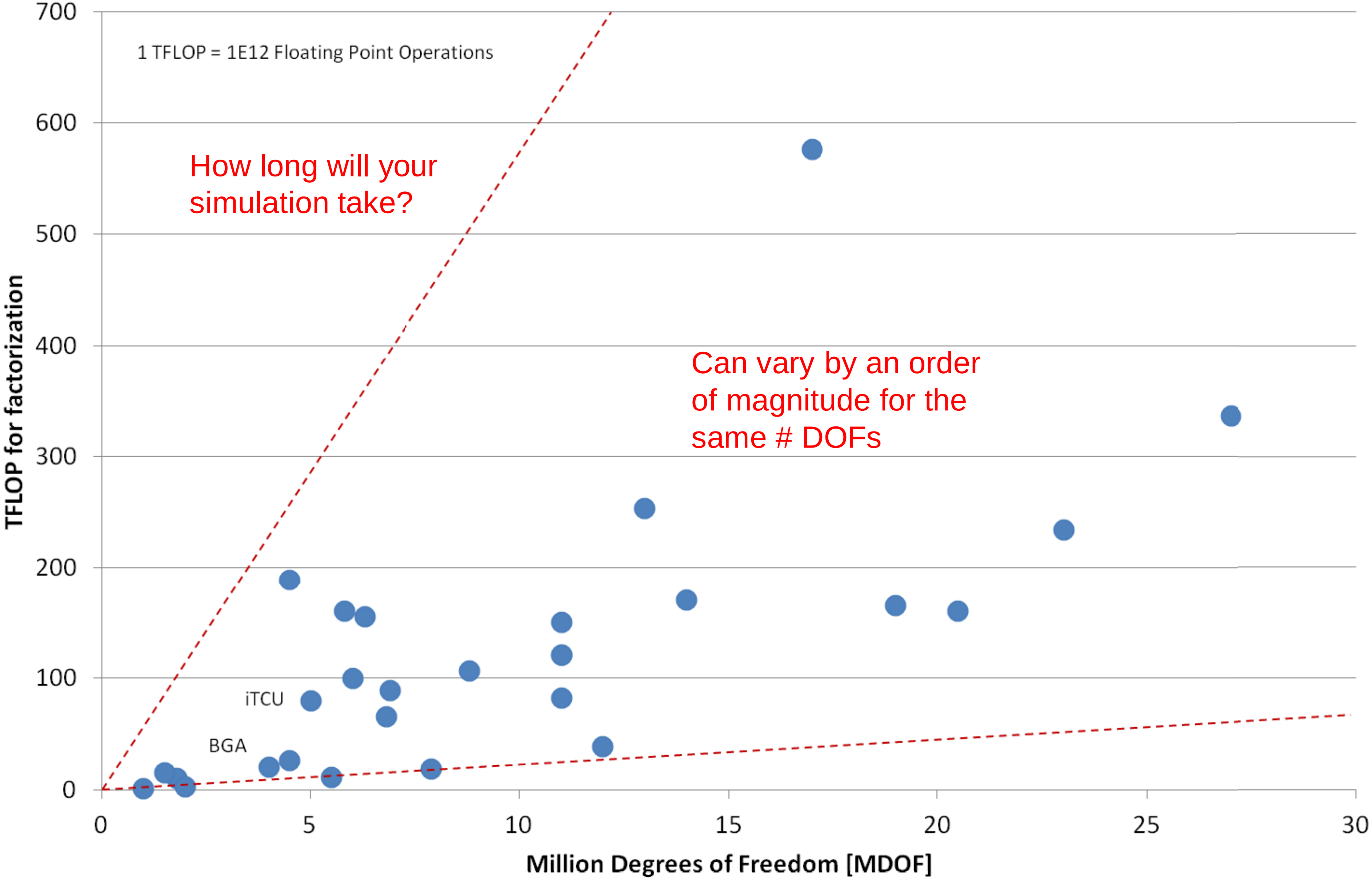
HPC for Optimization

ANSYS Solver Basics: Solution Overview



- Equation solver dominates solution CPU time! Need to pay attention to equation solver
- Equation solver also consumes the most system resources (memory and I/O)

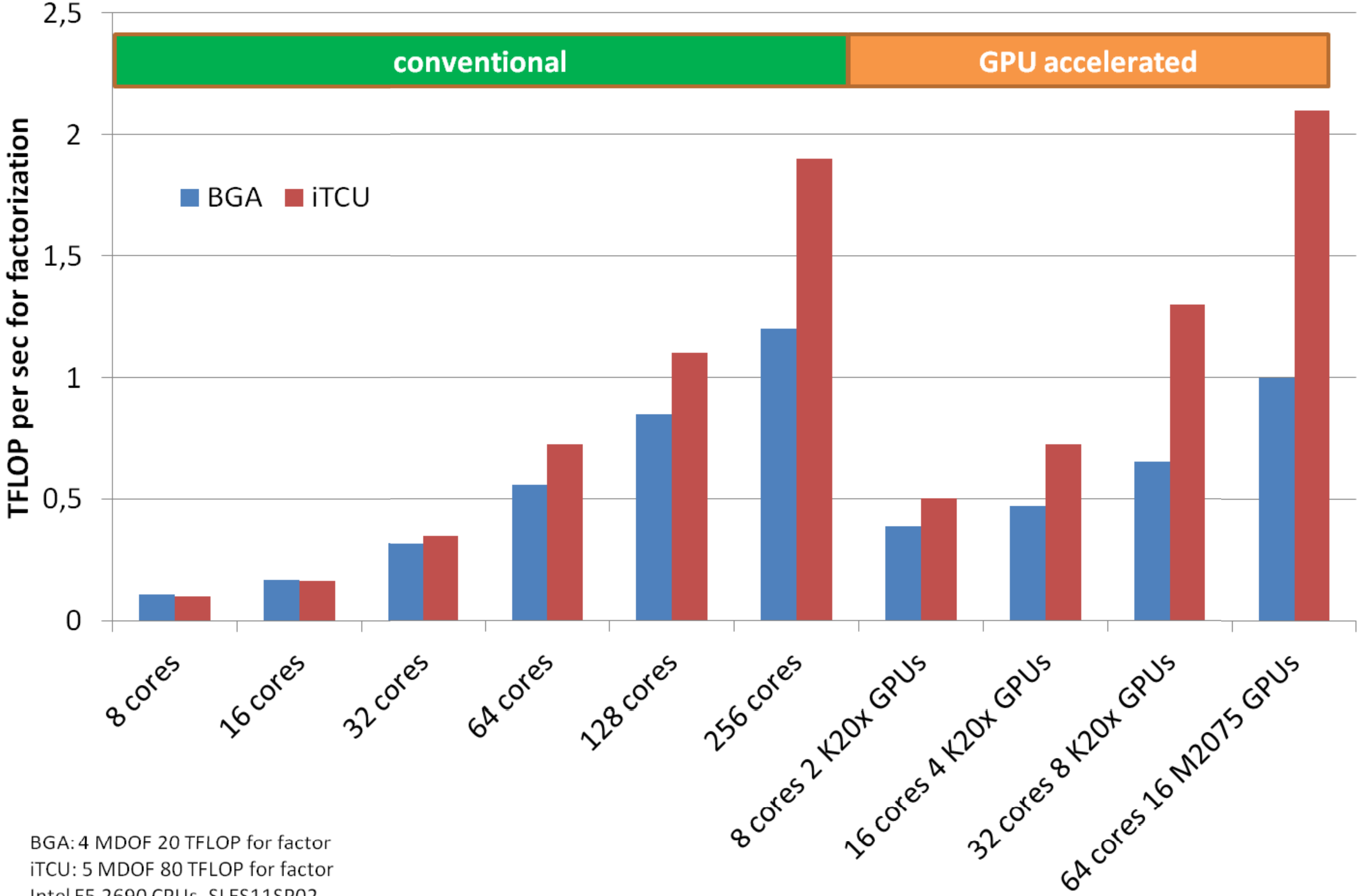
Numerical Effort for a random selection of MCE Projects ANSYS MAPDL, sparse solver



Stats data can be found in .DSP file

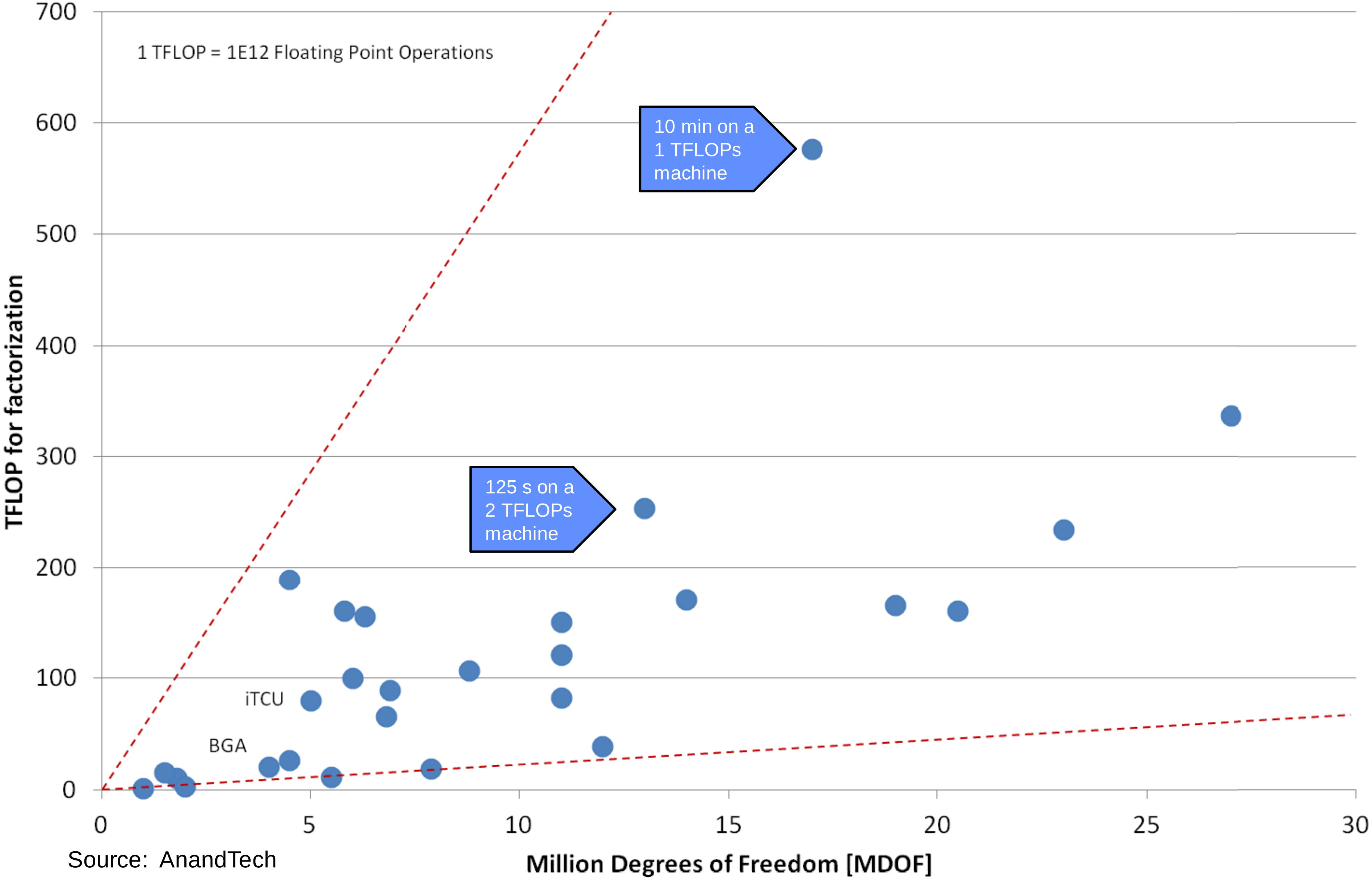
```
==========================
 = multifrontal statistics =
 ==========================

     number of equations                    =        5162979  = 5.1 MDOF
     no. of nonzeroes in lower triangle of a =      406156585
     no. of nonzeroes in the factor l        =    11942613850
     ratio of nonzeroes in factor (min/max)  =         0.0088
     number of super nodes                   =          64204
     maximum order of a front matrix         =          13857
     maximum size of a front matrix          =       96015153
     maximum size of a front trapezoid       =       60763830
     no. of floating point ops for factor    =     7.7694D+13  = 78 TFLOP
     no. of floating point ops for solve     =     4.6605D+10
     ratio of flops for factor (min/max)     =         0.0048
     near zero pivot monitoring activated
     number of pivots adjusted               =              0
     negative pivot monitoring activated
     number of negative pivots encountered   =              0
     factorization panel size                =            128
     number of cores used                    =            512
     time (cpu & wall) for structure input   =       1.060000        1.079678
     time (cpu & wall) for ordering          =       6.287842        6.287842
     time (cpu & wall) for other matrix prep =      13.262158       14.075546
     time (cpu & wall) for value input       =       1.120000        1.138784
     time (cpu & wall) for matrix distrib.   =       5.430000        5.780427
     time (cpu & wall) for numeric factor    =      19.650000       25.445240
     computational rate (mflops) for factor  =  3953900.312372  3053386.012269  = 4 TFLOPs
     time (cpu & wall) for numeric solve     =       0.340000        0.682836
     computational rate (mflops) for solve   =  137074.487818   68252.599236
     effective I/O rate (MB/sec) for solve   =  522253.790742  260042.399186


  Solver Memory allocated on core    0       =      725.389 MB
  Solver Memory allocated on core    1       =      711.322 MB
  …
  Solver Memory allocated on core  510       =      665.193 MB
  Solver Memory allocated on core  511       =      434.874 MB
  Total Solver Memory allocated by all cores =   346653.300 MB

  DSP Matrix Solver         CPU Time (sec) =         52.770
  DSP Matrix Solver     ELAPSED Time (sec) =         60.485
  DSP Matrix Solver      Memory Used ( MB) =        725.389
```

Performance Results



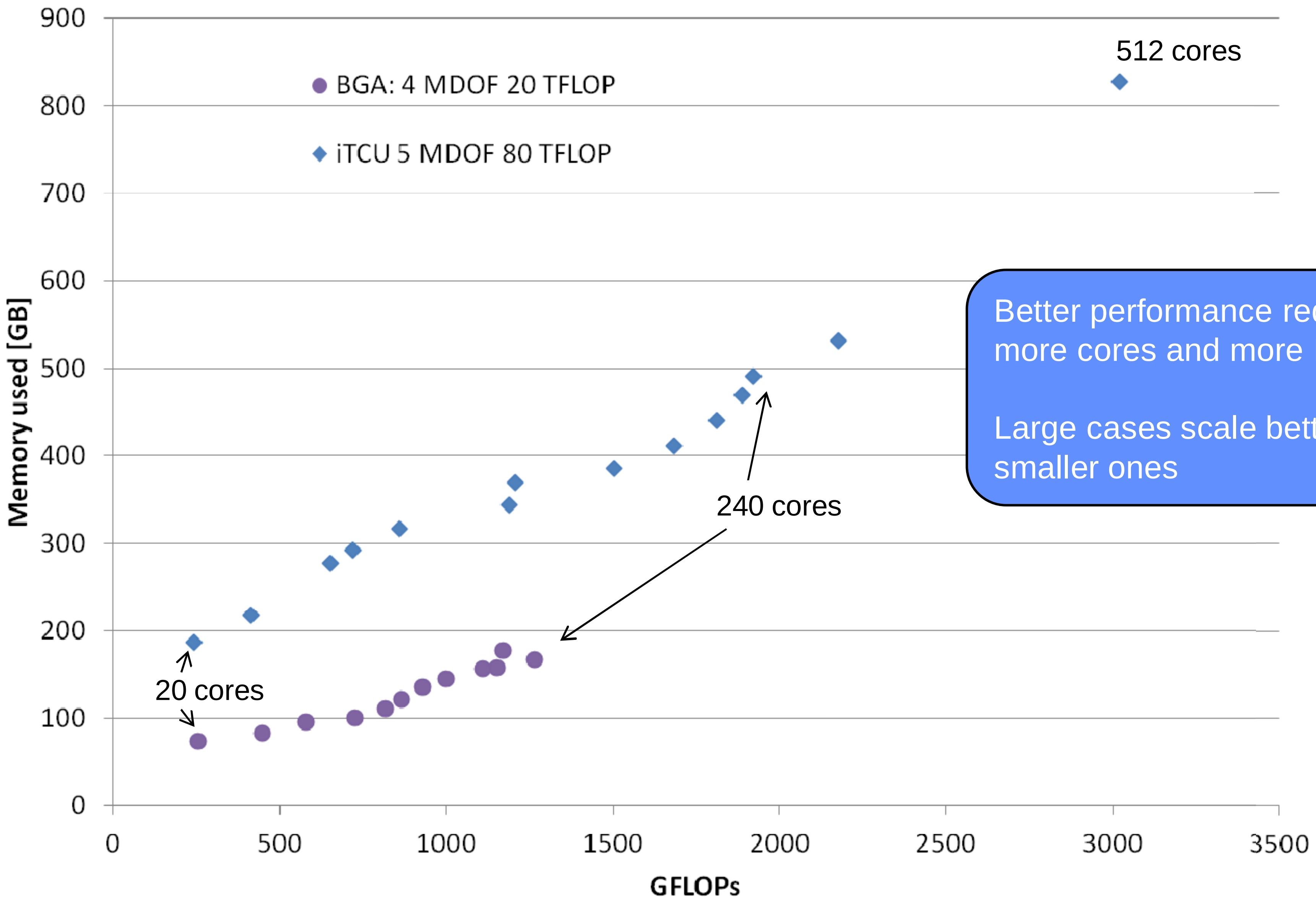
BGA: 4 MDOF 20 TFLOP for factor
iTCU: 5 MDOF 80 TFLOP for factor
Intel E5 2690 CPUs, SLES11SP02

Numerical Effort for a random selection of MCE Projects ANSYS MAPDL, sparse solver



Source: AnandTech

Scaling & Ressources



Better performance requires more cores and more RAM

Large cases scale better than smaller ones

Hardware: March 2014

240 E5 2690 V2 Ivy Bridge cores @ 3.0 GHz
64 E5 4627 V2 Ivy Bridge cores @ 3.3 GHz
320 E5 2690 Sandy Bridge cores @ 2.9 GHz

6..32 GB / core RAM (6 TB total)

Accelerators:
22 Fermi M207x ,
10 Kepler K20x
2 Kepler K40
2 Xeon Phi 7210P

Peak Performance ANSYS
single job: 4 TFLOPs
single node: 0.4 TFLOP (4S)
single node: 0.7 TFLOP (2S+2K40)
accumulated / 24 Jobs: 10 TFLOPs

Infiniband interconnect (FDR/QDR)
Compute servers SSD only
Remote Access: 3x HP-RGS

6 Fileservers , 200TB

SLES 11 SP02 for compute nodes

Closed loop aircooled rack (20kW)

Interconnect: FDR Performance

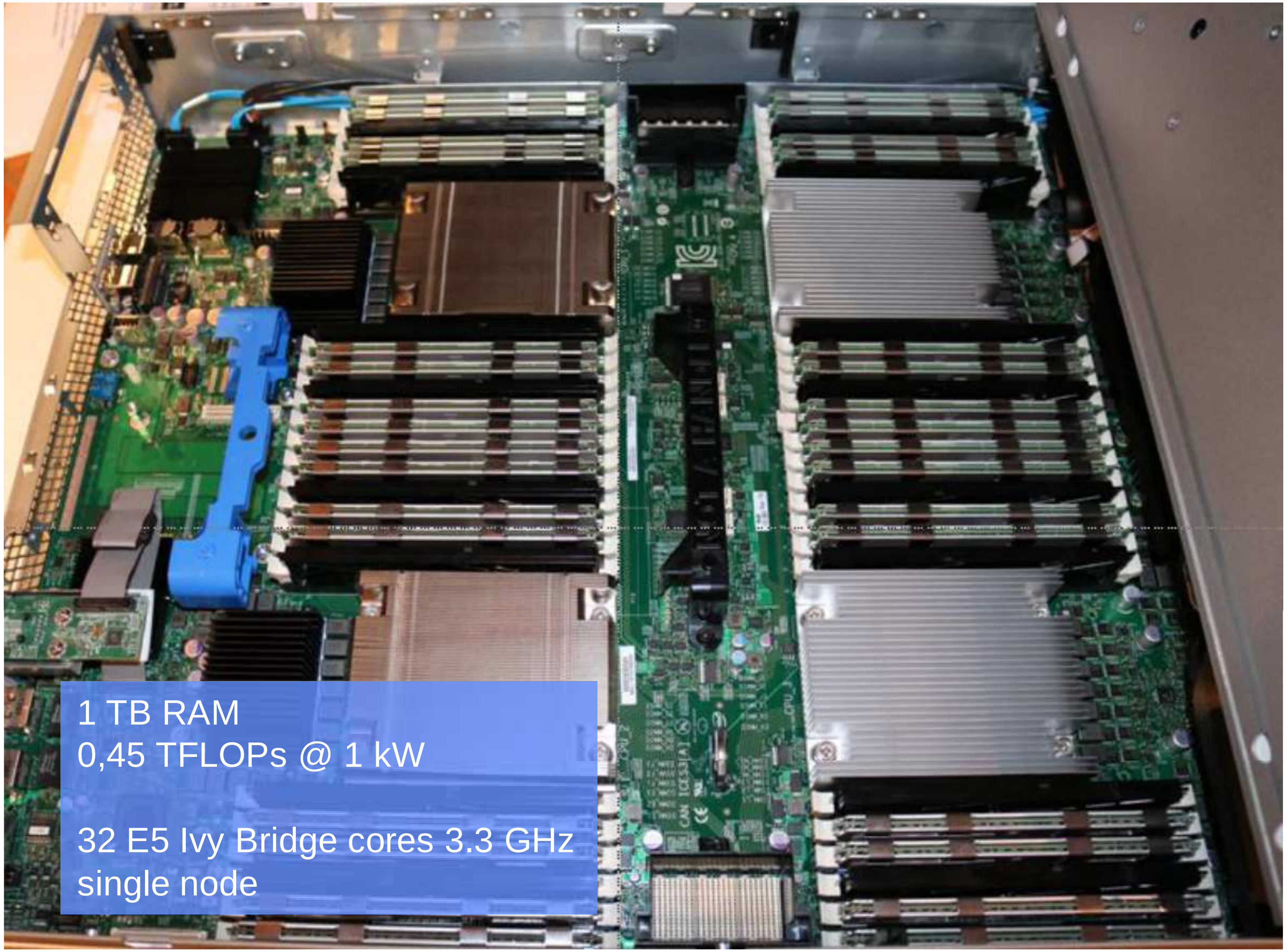
Latencies	Bandwidth
Latency time from master to core 1 = 1.259 µs	Communication speed from master to core 1 = 8077.06 MB/sec
Latency time from master to core 2 = 1.175 µs	Communication speed from master to core 2 = 8857.00 MB/sec
Latency time from master to core 3 = 1.235 µs	Communication speed from master to core 3 = 9372.93 MB/sec
…	
Latency time from master to core 9 = 2.183 µs	Communication speed from master to core 9 = 5312.38 MB/sec
Latency time from master to core 10 = 2.393 µs	Communication speed from master to core 10 = 5377.34 MB/sec
Latency time from master to core 11 = 1.836 µs	Communication speed from master to core 11 = 5081.82 MB/sec
…	
Latency time from master to core 16 = 1.979 µs	Communication speed from master to core 16 = 5121.90 MB/sec
Latency time from master to core 17 = 2.012 µs	Communication speed from master to core 17 = 5313.56 MB/sec
Latency time from master to core 18 = 2.008 µs	Communication speed from master to core 18 = 5249.56 MB/sec
…	…
Latency time from master to core 28 = 1.993 µs	Communication speed from master to core 28 = 4939.63 MB/sec
Latency time from master to core 29 = 2.366 µs	Communication speed from master to core 29 = 4939.24 MB/sec
Latency time from master to core 30 = 2.333 µs	Communication speed from master to core 30 = 4765.06 MB/sec
Latency time from master to core 31 = 2.119 µs	Communication speed from master to core 31 = 4925.74 MB/sec

core – core on die

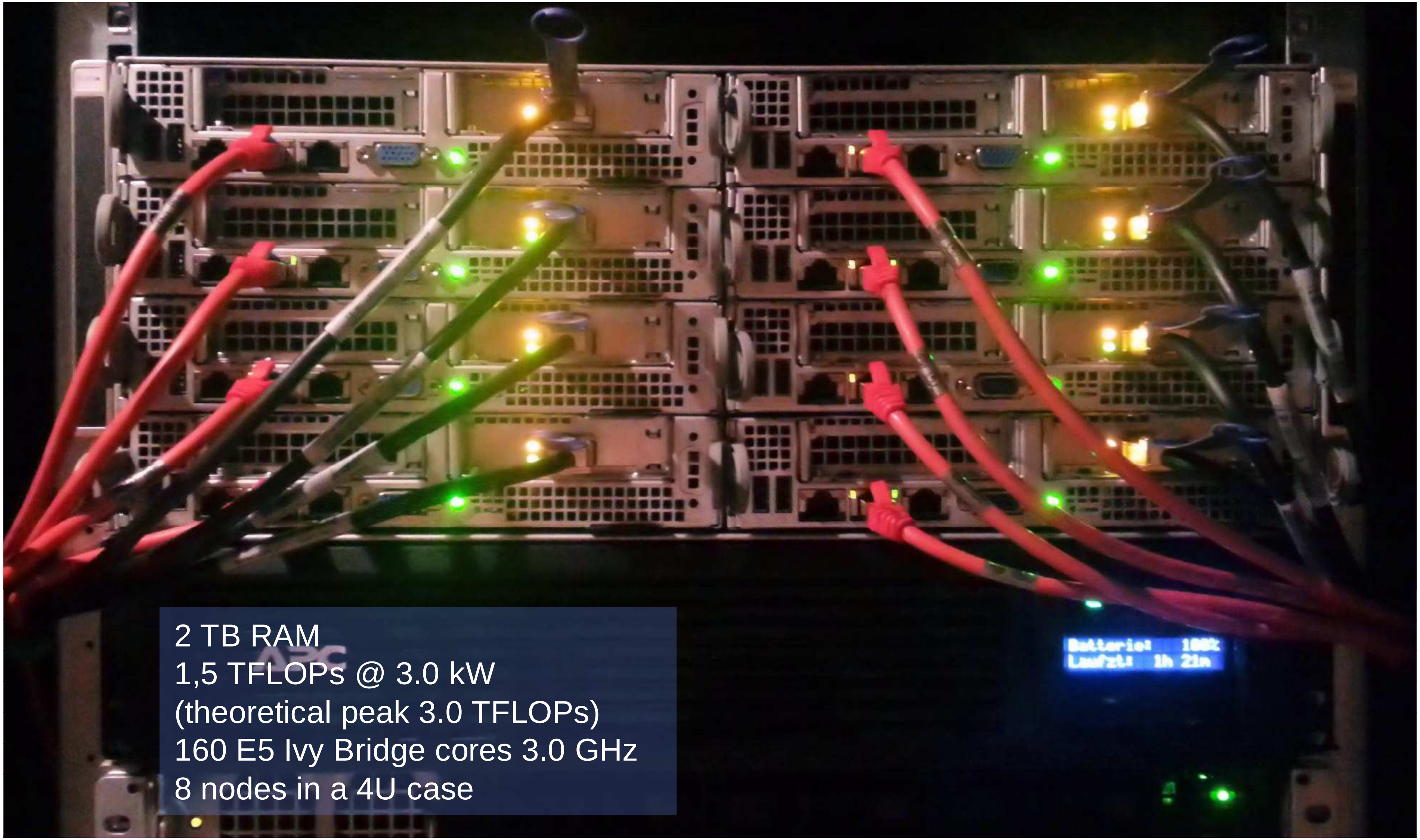
socket - socket

node - node

Tools (Hardware: March 2014)



Tools (Hardware: Sept 2013)



Tools (Hardware: June 2013)

GPU Performance Data

SELECT THE RIGHT TESLA GPU

Features	Tesla K20X	Tesla K20	Tesla K10	Tesla M2090	Tesla M2075
Number and Type of GPU	1 Kepler GK110		2 Kepler GK104s	1 Fermi GPU	1 Fermi GPU
GPU Computing Applications	Seismic processing, CFD, CAE, Financial computing, Computational chemistry and Physics, Data analytics, Satellite imaging, Weather modeling		Seismic processing, signal and image processing, video analytics	Seismic processing, CFD, CAE, Financial computing, Computational chemistry and Physics, Data analytics, Satellite imaging, Weather modeling	
Peak double precision floating point performance	1.31 Tflops	1.17 Tflops	190 Gigaflops (95 Gflops per GPU)	665 Gigaflops	515 Gigaflops
Peak single precision floating point performance	3.95 Tflops	3.52 Tflops	4577 Gigaflops (2288 Gflops per GPU)	1331 Gigaflops	1030 Gigaflops
Memory bandwidth (ECC off)	250 GB/sec	208 GB/sec	320 GB/sec (160 GB/sec per GPU)	177 GB/sec	
Memory size (GDDR5)	6 GB	5 GB	8GB (4 GB per GPU)	6 GigaBytes	
CUDA cores	2688	2496	3072 (1536 per GPU)	512	

K20X	K40 (Atlas)
1x GK110	1x GK180
3.95 TF 2.90 TF	>4.0 TF
1.32 TF 1.22 TF	>1.4 TF
6 GB	12 GB
250 GB/s	288 GB/s
-	AMBER, ANSYS
Gen 2	Gen 3
2688	2880
235W	235W (245W SXM)
PCIe Passive, SXM	PCIe Passive, Active & TTP, SXM

GPU Acceleration

Real life test with customer case @ MicroConsult:
Hardware: E5 2690, 2x Tesla K40m Accelerator, DSPARSE

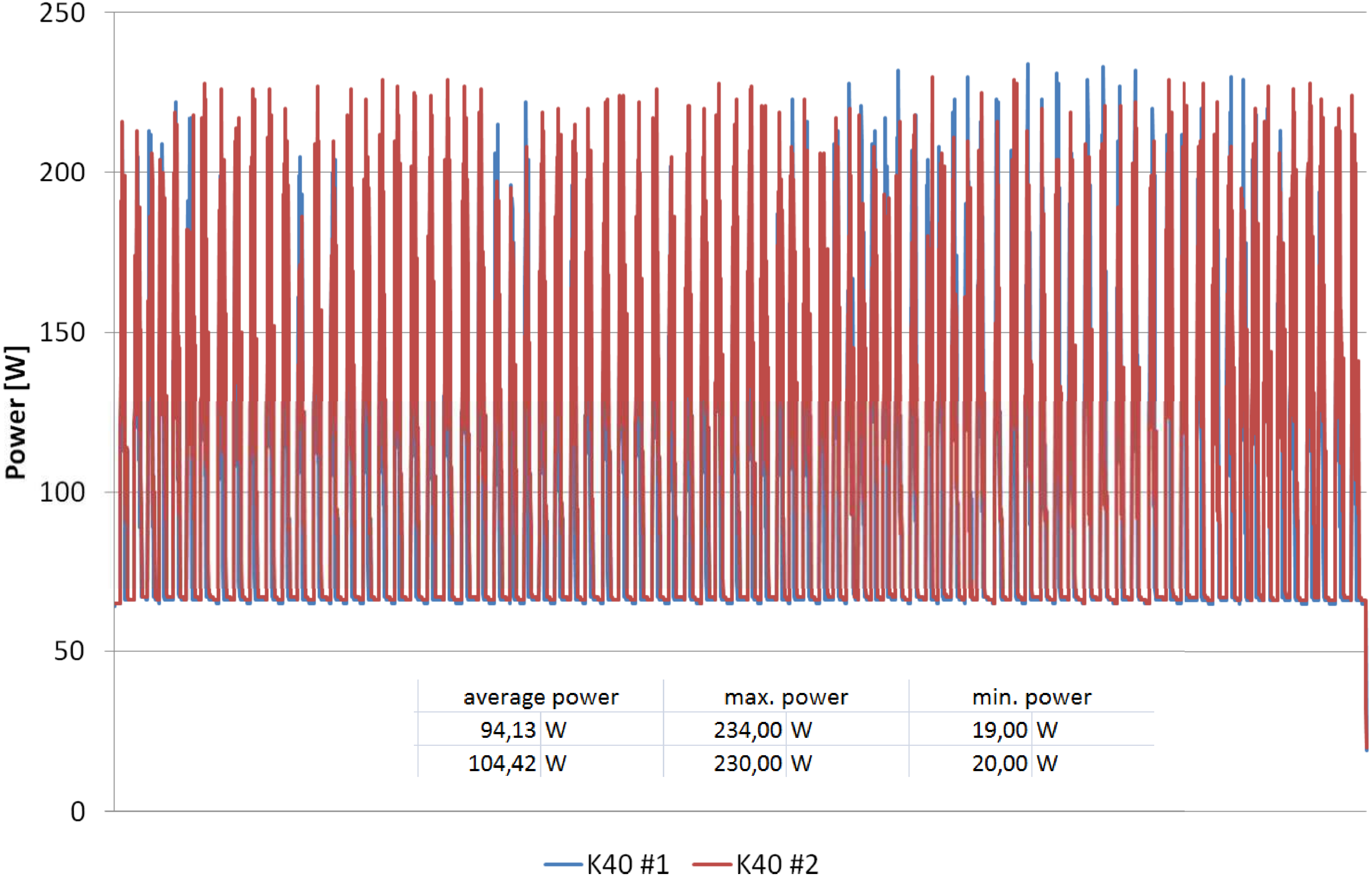
66s later →

```
Sun Nov 24 20:02:36 2013
+------------------------------------------------------+
| NVIDIA-SMI 5.319.60     Driver Version: 319.60       |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla K40m          On   | 0000:03:00.0     Off |                    0 |
| N/A   40C    P0   180W / 235W |   3078MB / 11519MB |     99%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla K40m          On   | 0000:83:00.0     Off |                    0 |
| N/A   44C    P0   167W / 235W |   1679MB / 11519MB |     99%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Compute processes:                                               GPU Memory |
|  GPU       PID  Process name                                     Usage      |
|=============================================================================|
|    0      6167  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      108MB  |
|    0      6169  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    0      6170  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150     1446MB  |
|    0      6164  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150     1333MB  |
|    1      6171  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150     1276MB  |
|    1      6172  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    1      6168  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      105MB  |
|    1      6173  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
+-----------------------------------------------------------------------------+
```

```
Sun Nov 24 20:03:42 2013
+------------------------------------------------------+
| NVIDIA-SMI 5.319.60     Driver Version: 319.60       |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla K40m          On   | 0000:03:00.0     Off |                    0 |
| N/A   37C    P0    72W / 235W |    507MB / 11519MB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla K40m          On   | 0000:83:00.0     Off |                    0 |
| N/A   45C    P0    79W / 235W |    507MB / 11519MB |      0%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Compute processes:                                               GPU Memory |
|  GPU       PID  Process name                                     Usage      |
|=============================================================================|
|    0      6167  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    0      6169  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    0      6170  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    0      6164  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    1      6171  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    1      6172  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    1      6168  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
|    1      6173  /usr/ansys_inc/v150/ansys/bin/linx64/ansys.e150      104MB  |
+-----------------------------------------------------------------------------+
```

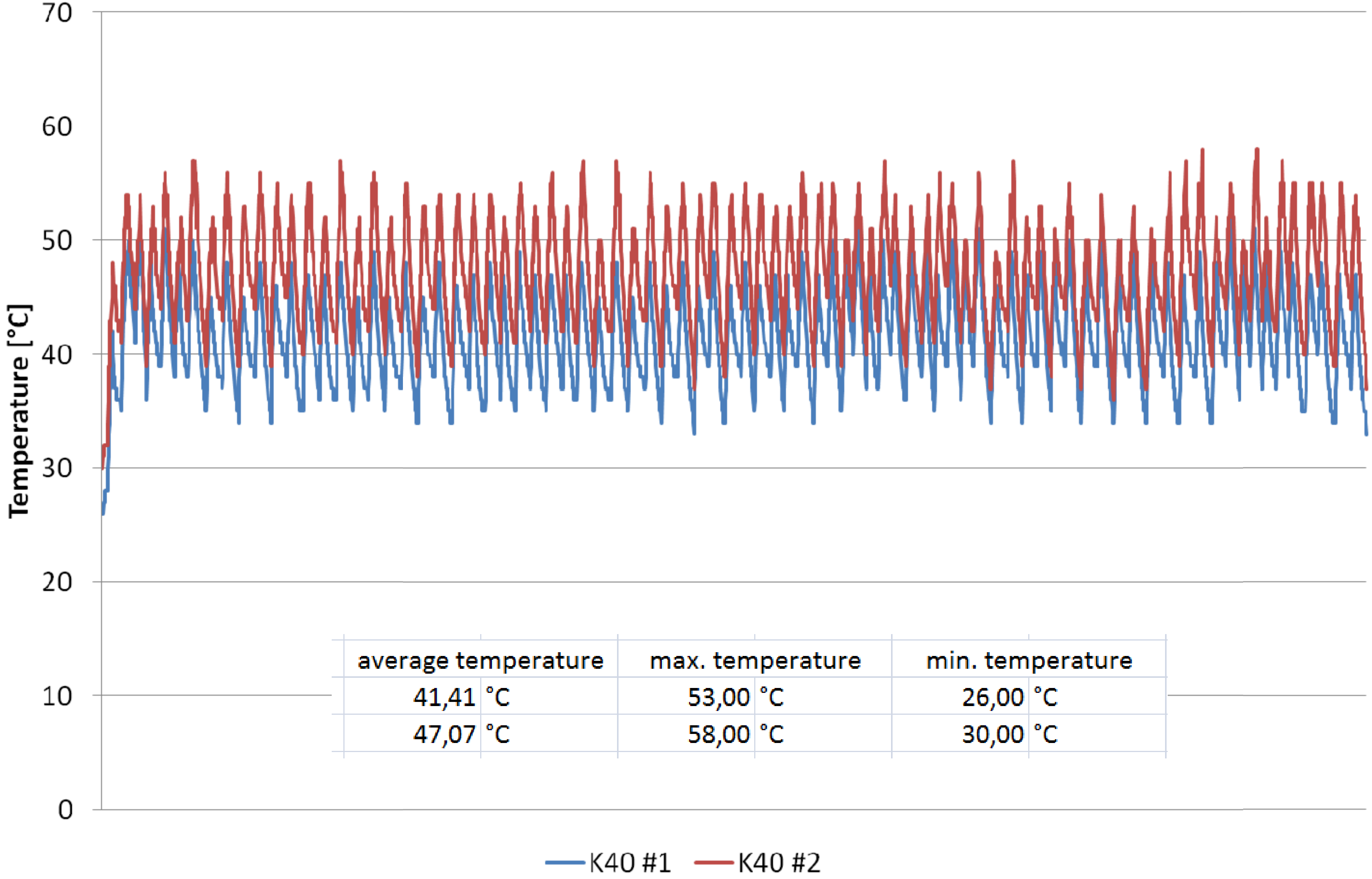
Duty Cycle ca. 30%

GPU Insights: GPU Power



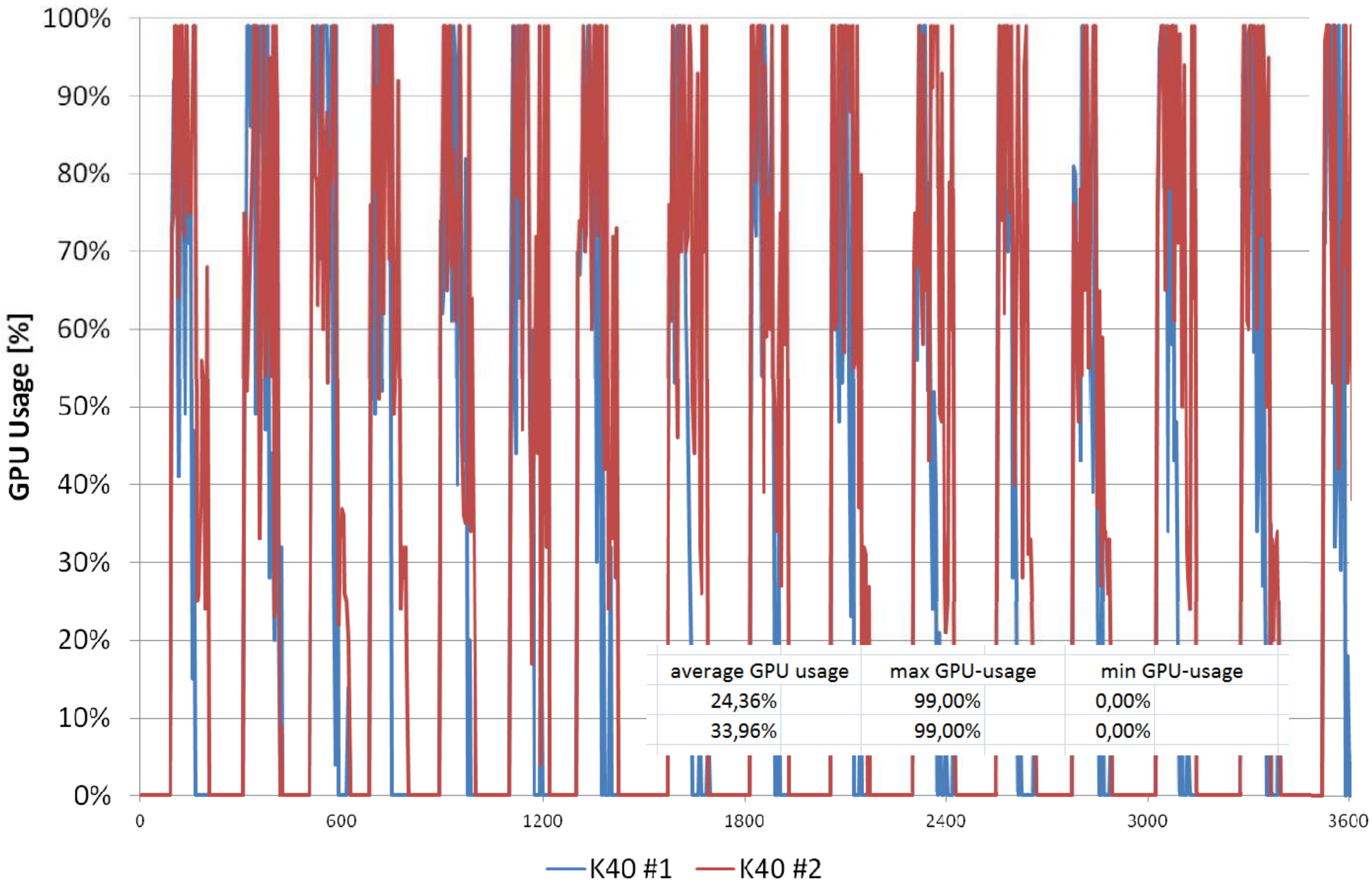
average power	max. power	min. power
94,13 W	234,00 W	19,00 W
104,42 W	230,00 W	20,00 W

GPU Insights: GPU Temperature

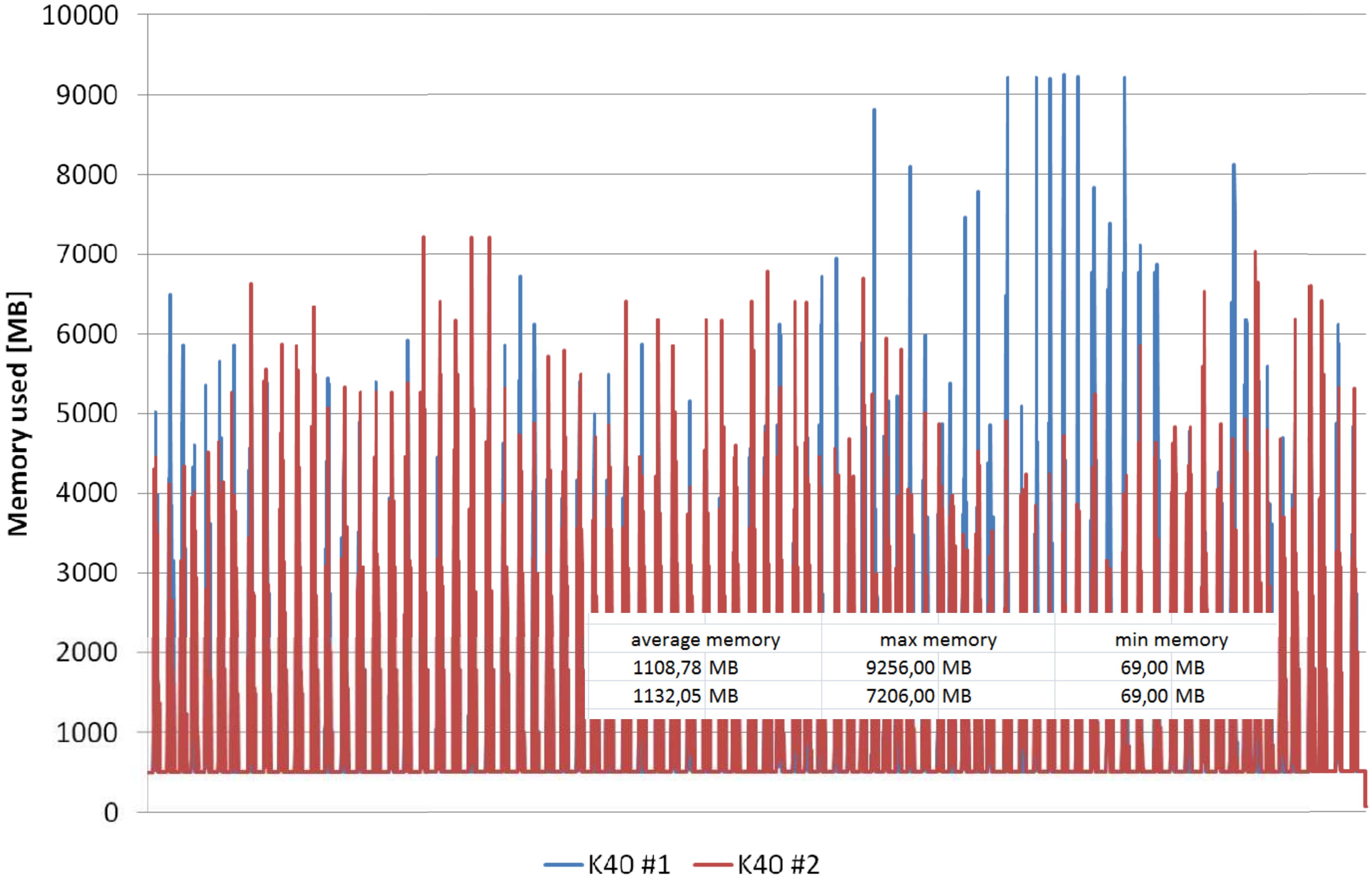


average temperature		max. temperature		min. temperature	
41,41	°C	53,00	°C	26,00	°C
47,07	°C	58,00	°C	30,00	°C

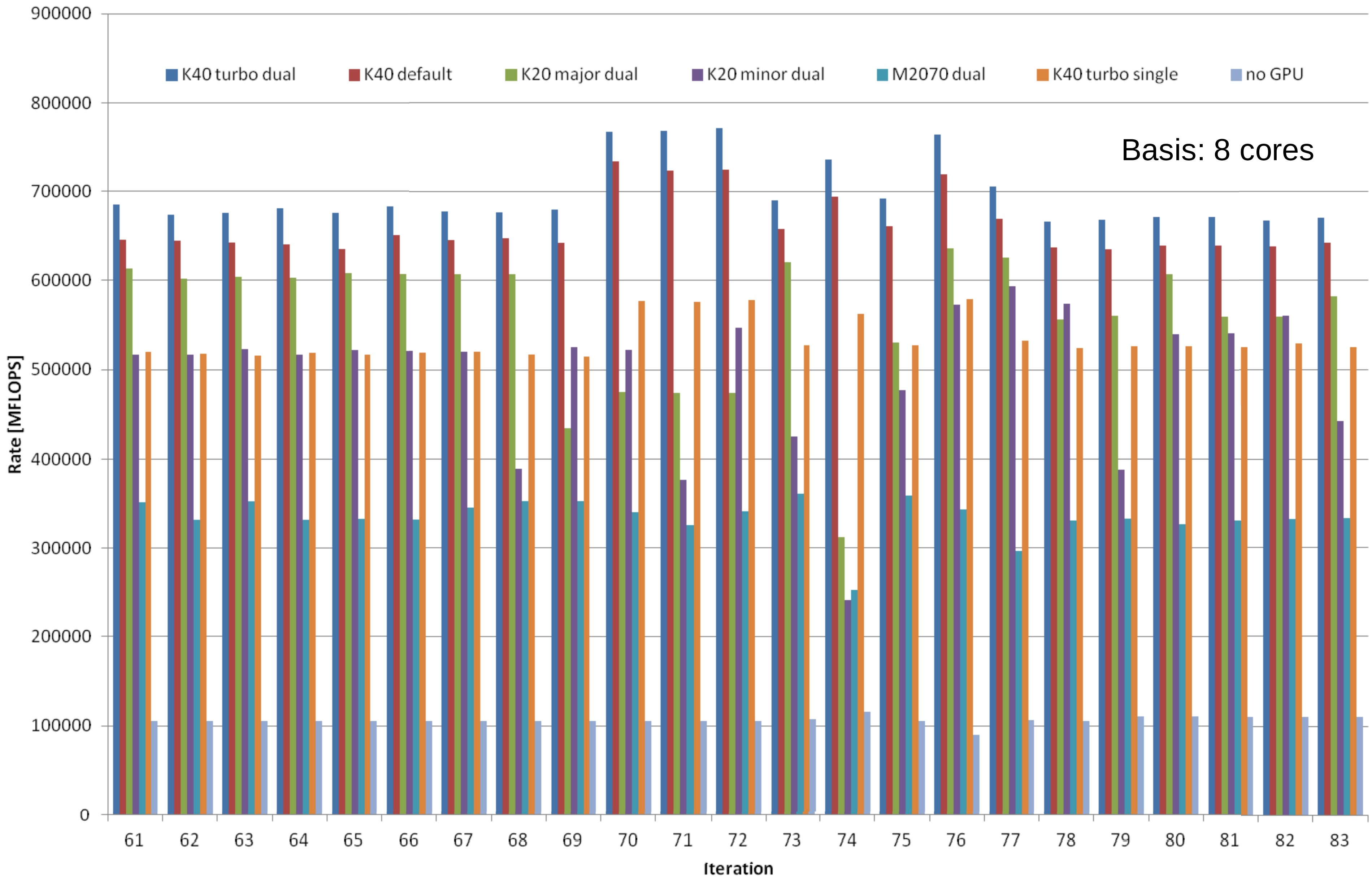
GPU Insights: GPU usage



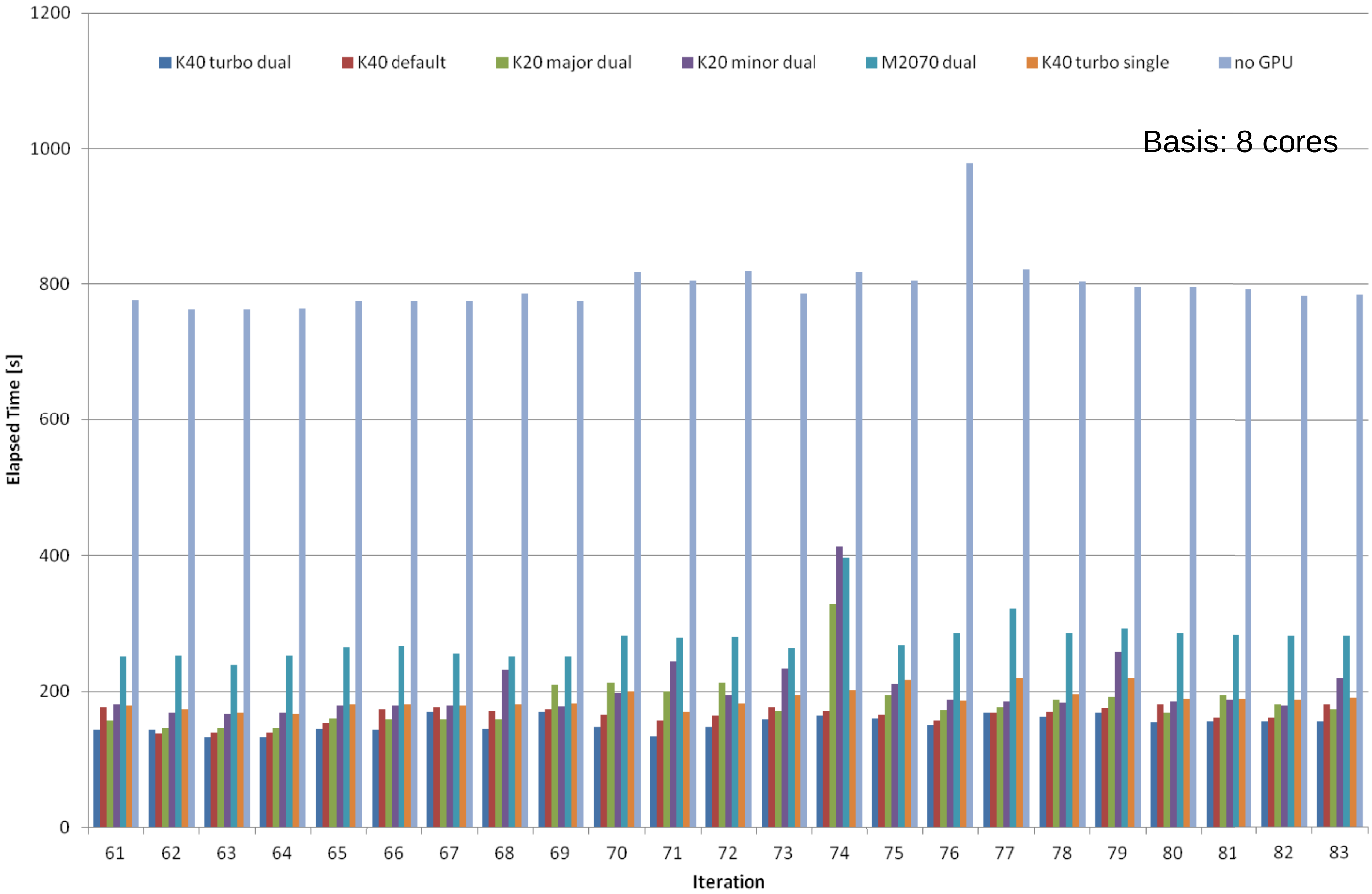
GPU Insights: Memory Usage



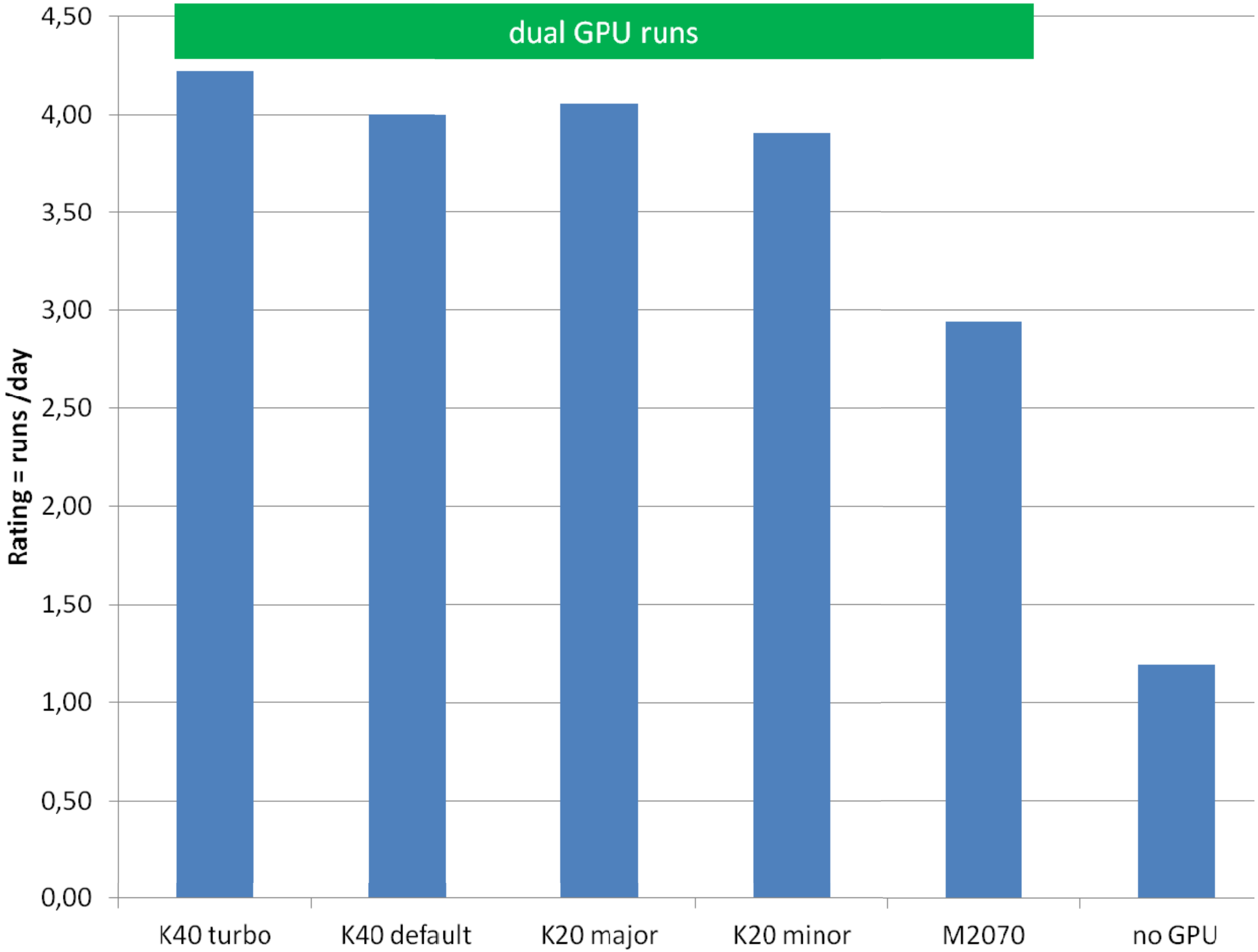
GPU Insights: Factorization rate



GPU Insights: Elapsed Time per Iteration



GPU Acceleration

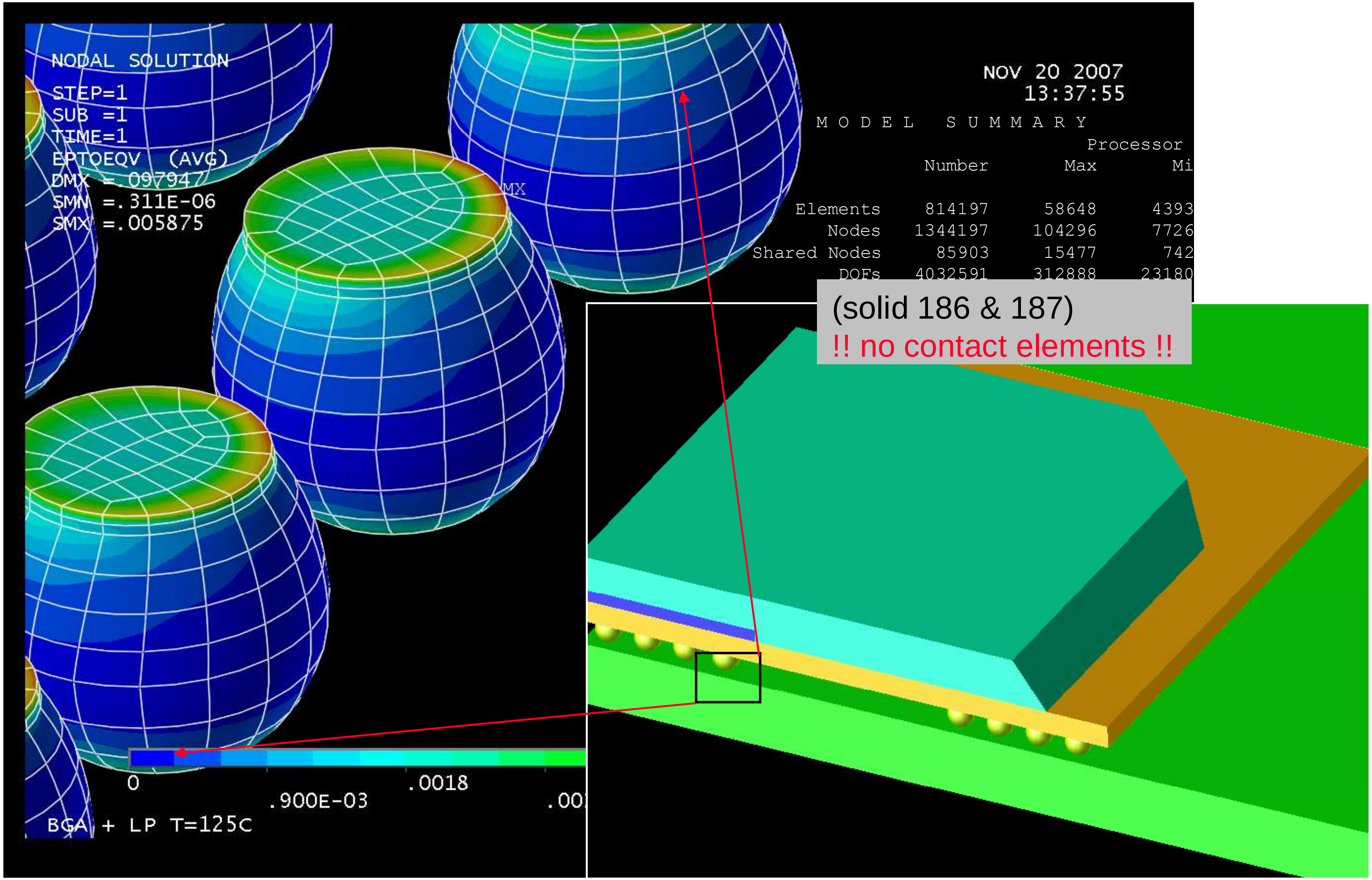


GPUs: Pros and Cons

- GPUs are the only way to boost performance of an existing workstation (after all memory slots are filled and you are running on solid state disks / SSDs)
- The price of a GPU is small compared to a complete workstation or the license costs to run ANSYS on it.
- GPUs are only acceleration the equation solving part
- For large models /large matrix fronts the local memory on the GPU board imposes some limits. It helps to combine multiple GPUs
- Finally, the licensing costs determine the attractiviy of using GPUs.
(Became more attractive with Release 15)

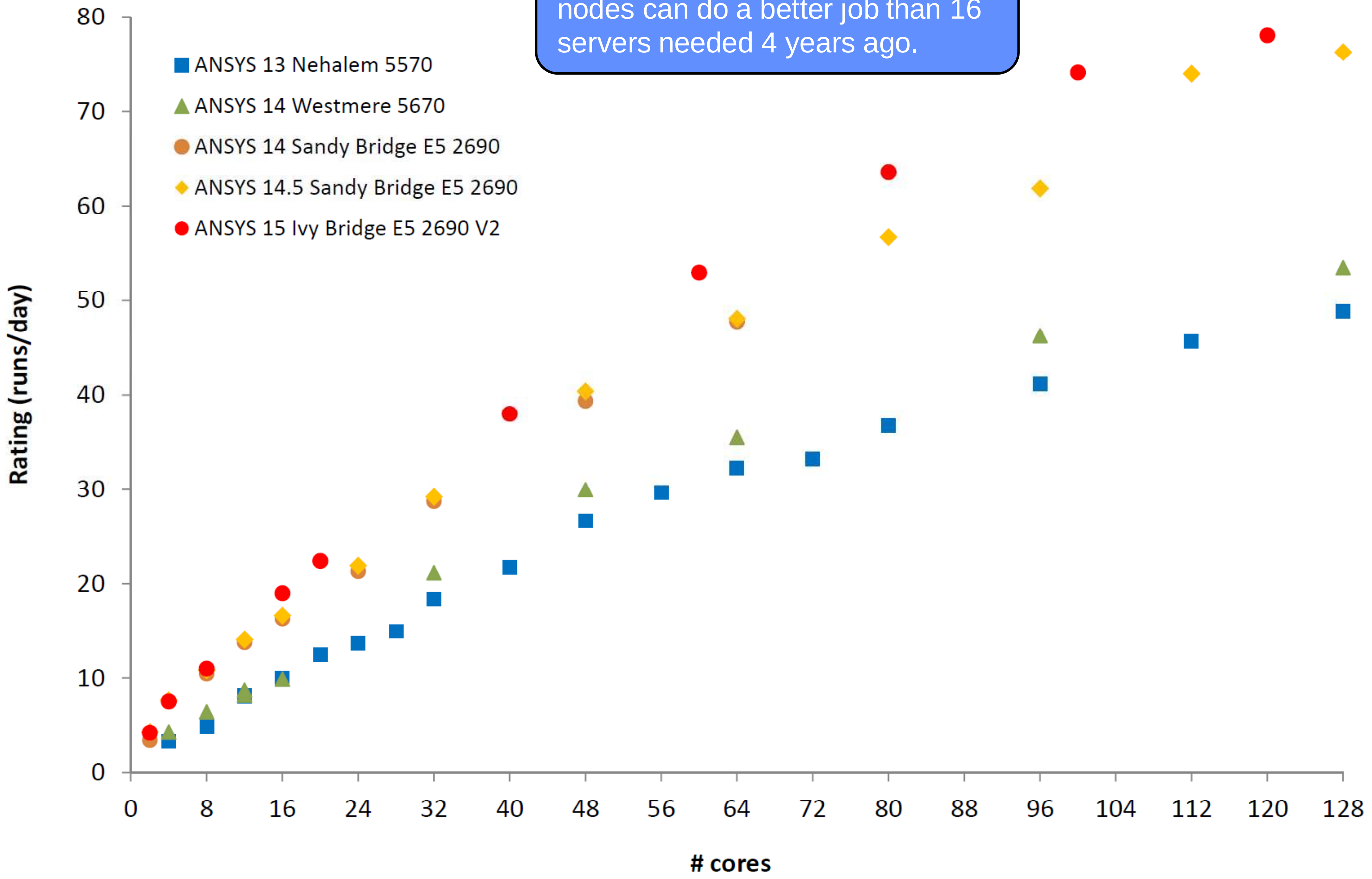
Applications

Example: Ball grid array



HPC with ANSYS 15: BGA

With current technology 6 compute nodes can do a better job than 16 servers needed 4 years ago.



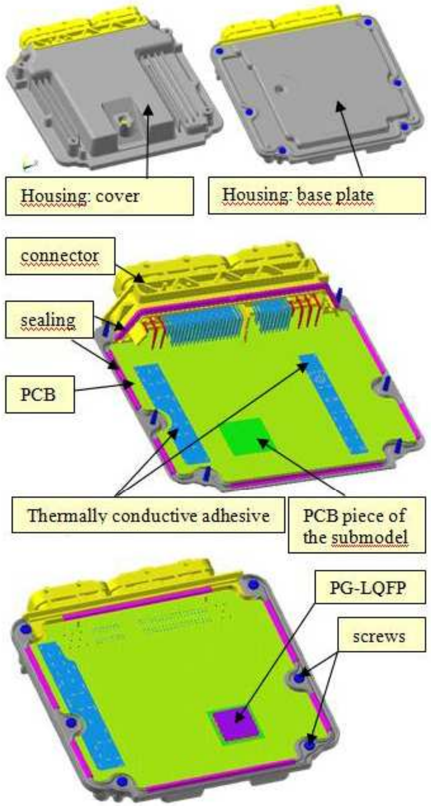
Applications:

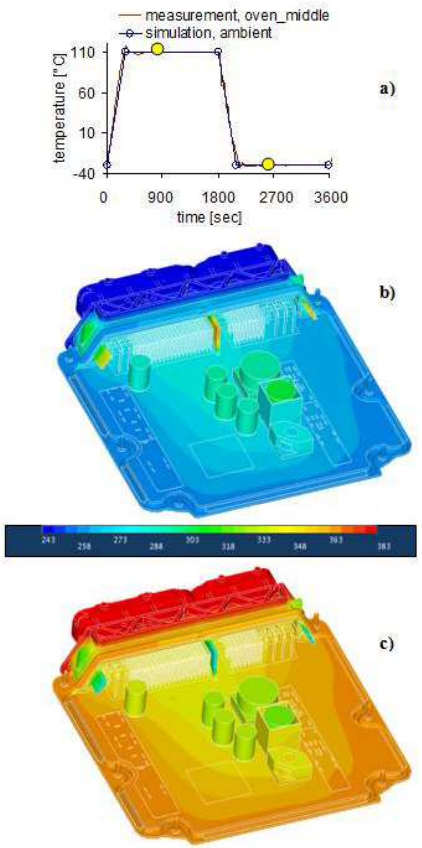
- BGA, LQFP
- Einzelbauteile & Systembetrachtung
- Schwerpunkt Lotkriechen

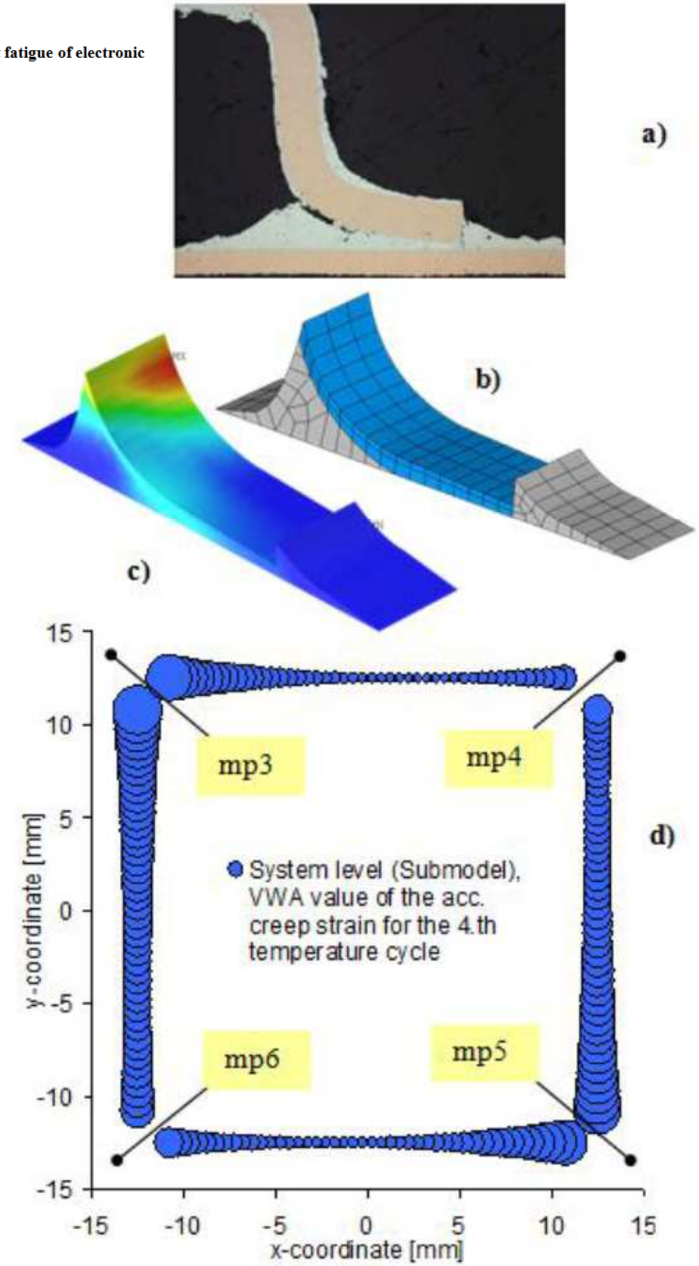
Development of a submodel technique for the simulation of solder joint fatigue of electronic devices mounted within an assembled ECU.

Natalja Schafet, Christian Lemm, Ulrich Becker
Robert Bosch GmbH, Schwieberdingen, Germany

Herbert Güttler, Philipp Schmid
MicroConsult GmbH, Bernstadt,Germany



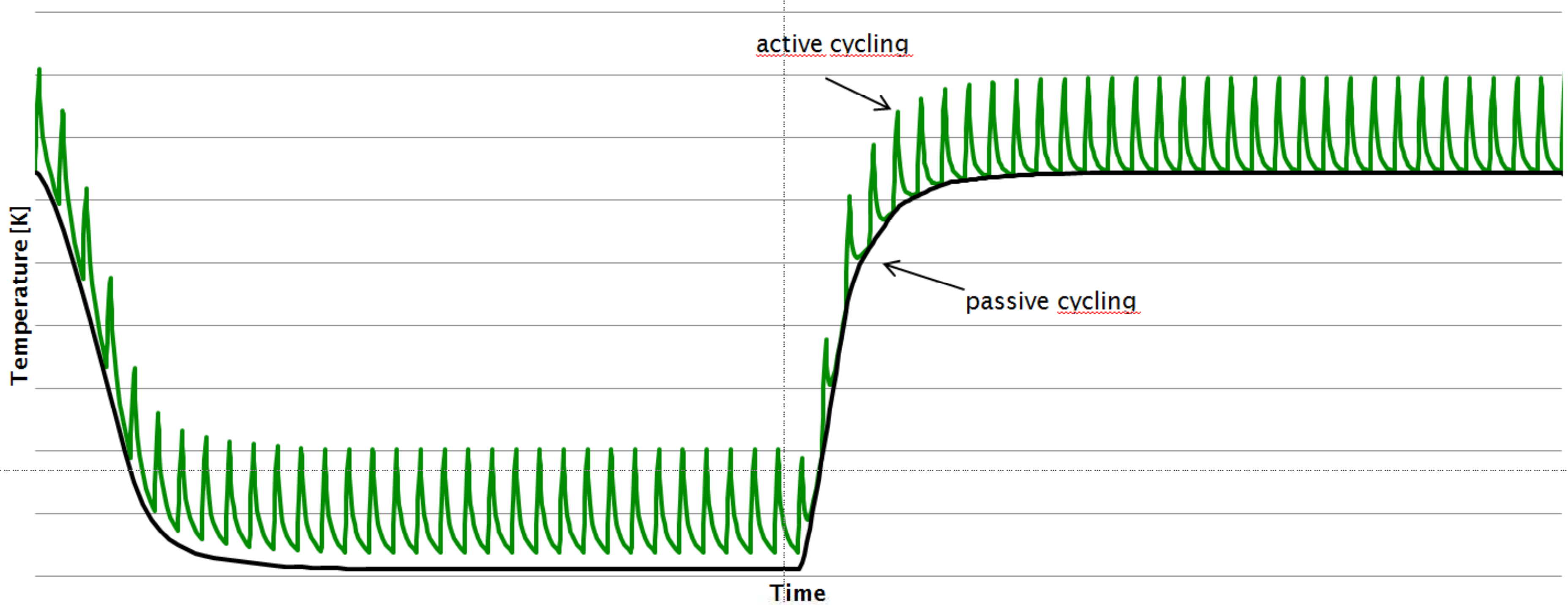




Benchmark Results: Leda Benchmark

Procedure	ANSYS 11	ANSYS12	ANSYS12.1	ANSYS13 SP02	ANSYS 14 UP20110901	ANSYS 14.5	ANSYS 15 UP20130403	ANSYS 15 UP20130906
Thermal (full model) 3 MDOF	4h (8 cores)				***1h (8 cores + 1 GPU) 0.8h (32 cores)***			
Thermo-mechanical Simulation (full model) 7.8 MDOF	~ 5.5 days for 163 iterations (8 cores)	34.3h for 164 iterations (20 cores)	12.5h for 195 iterations (64 cores)	***9.9h for 195 iterations (64 cores)***	***7.5h for 195 iterations (128 cores)***	***6.4h for 196 iterations (128 E5 cores) 6.3h (96 E5 cores + 16 GPUs)***	***7.3h for 196 iterations (128 E5 cores) PCMPI, imbalance= 4860***	***5.1h for 196 iterations (128 E5 cores) 5.2h (64 E5 cores + 16 GPUs) imbal: 2.95 8.6h (24 E5 cores + 8 K20x GPUs) imbal: 2.3***
Interpolation of boundary conditions	37h for 16 Loadsteps	Identical to ANSYS 11	Identical to ANSYS 11	***0.2h (improved algorithm)***	***0.2h***			
Submodell: Creep Strain Analysis 5.5 MDOF	~ 5.5 days for 492 iterations (16 cores)	38.5h for 492 iterations (16 cores)	8.5h for 492 iterations (76 cores)	***6.1h for 488 iterations (128 cores)***	***5.9h for 498 iterations (64 cores + 8GPUs) 4.2h (256 cores)***	***4.0h for 498 iterations (128 E5 cores)***		***4.1h for 488 iterations (128 E5 cores)***
	2 weeks	5 days	2 days	1 day	½ day	½ day		

Done in half a day? Comparison between 2008 and 2013

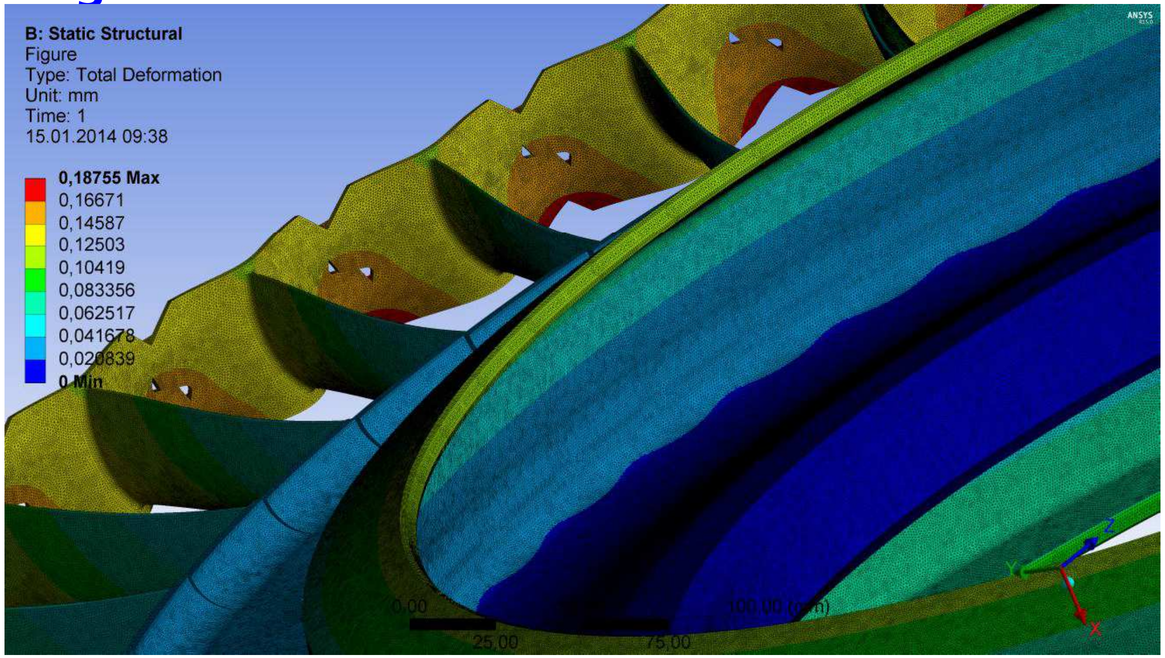


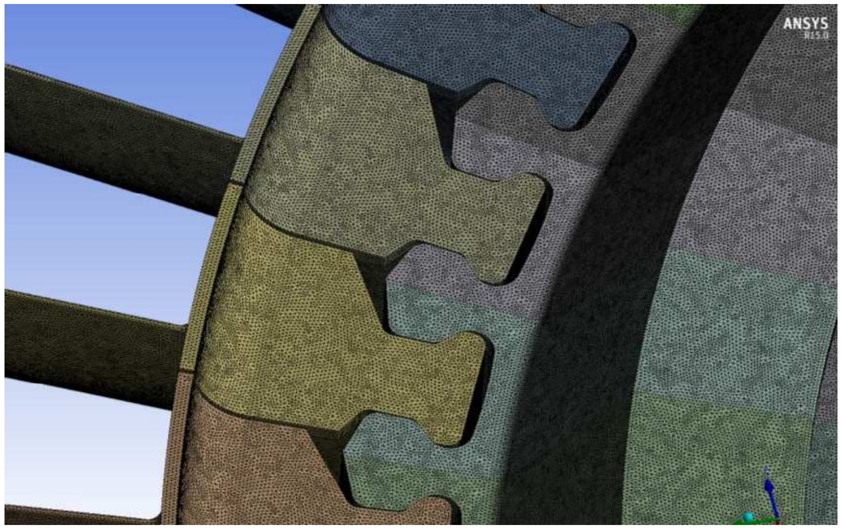
	nodes	DOF	TFLOP/iteration
thermal model	949k	949k	1,3
mechanical model	949k	2,8M	39
cluster @ 128 cores	**hours**	**iterations**	
passiveTTA:	3	487	
passiveTMA	34	916	
activeTTA	35	6051	
activeTMA	235	8965	

New possibilities because of more compute power

‚active cycling' imposes10x increase in numerical effort

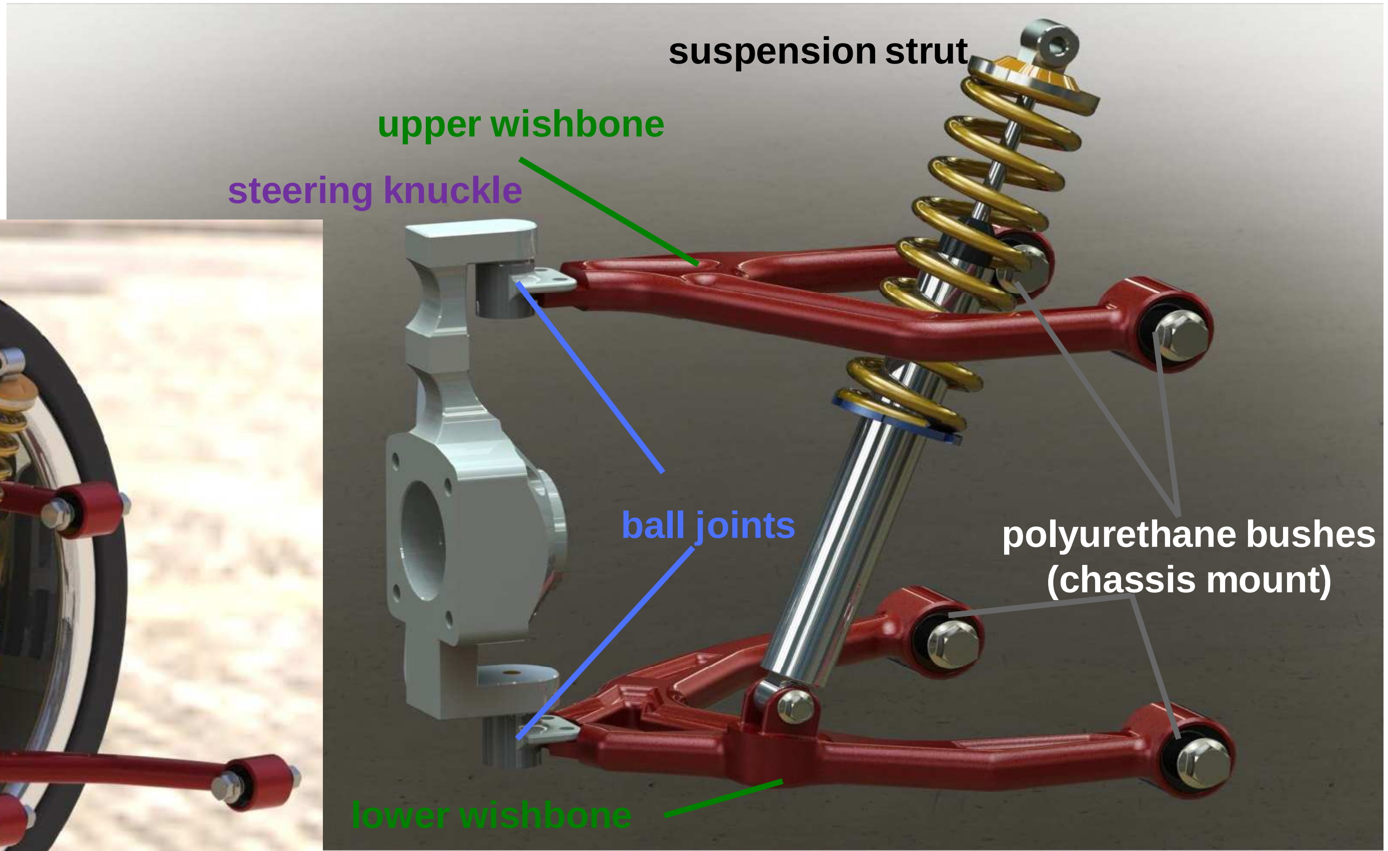
PCG Solver: Pushing the limits





Model	memory used	solve	wall	# cores	# nodes
25M	61098	69,3	279	64	16
50M	99401	123	526	64	16
100M	196039	200	1011	64	16
300M	454470	978	13698	34	3
500M	736546	1529	6714	32	1
600M	925332	1416	7273	34	5
750M	1148959	1865	10743	34	5
850M	1557951	1140	9490	64	16
860M	1693157	1158	10024	64	16
900M	1777981	1288	10576	64	16
950M	1942394	1446	12153	64	16
1000M	2108354	1612	34575	64	16
1250M	3547185	2541	20842	128	17

HPC & Optimization

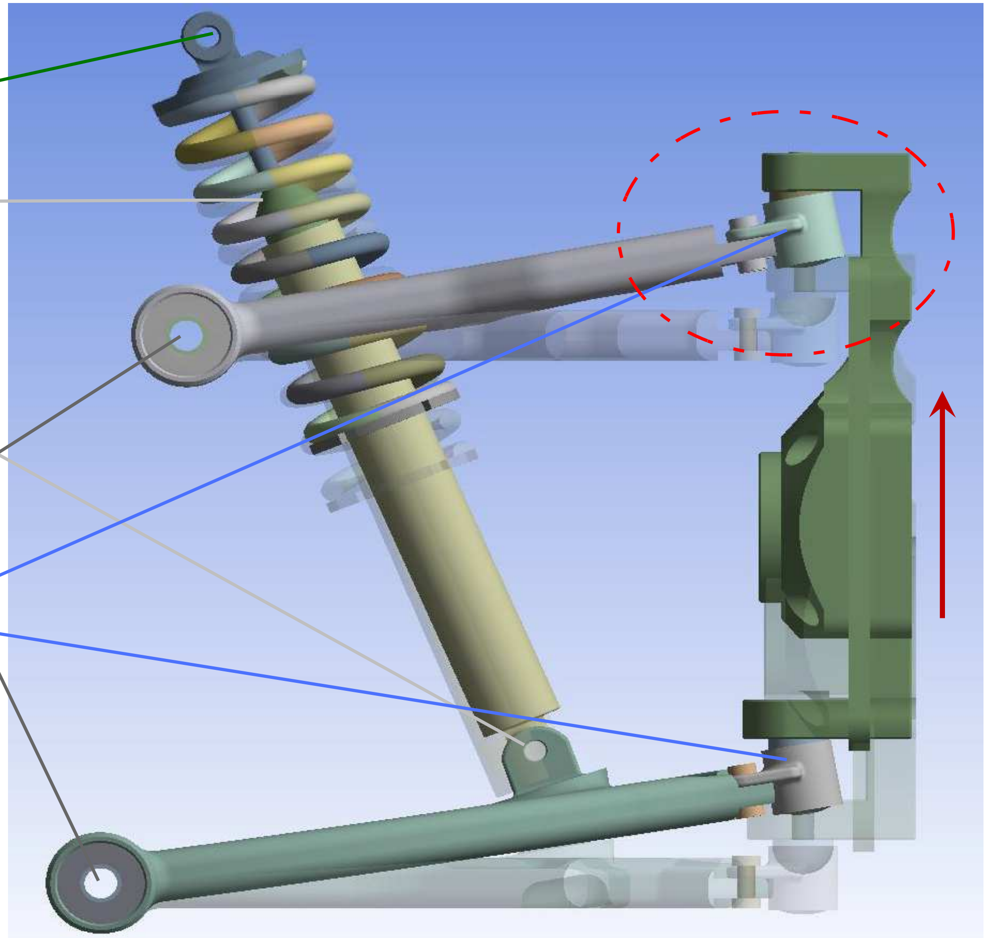


Simulation – Overview

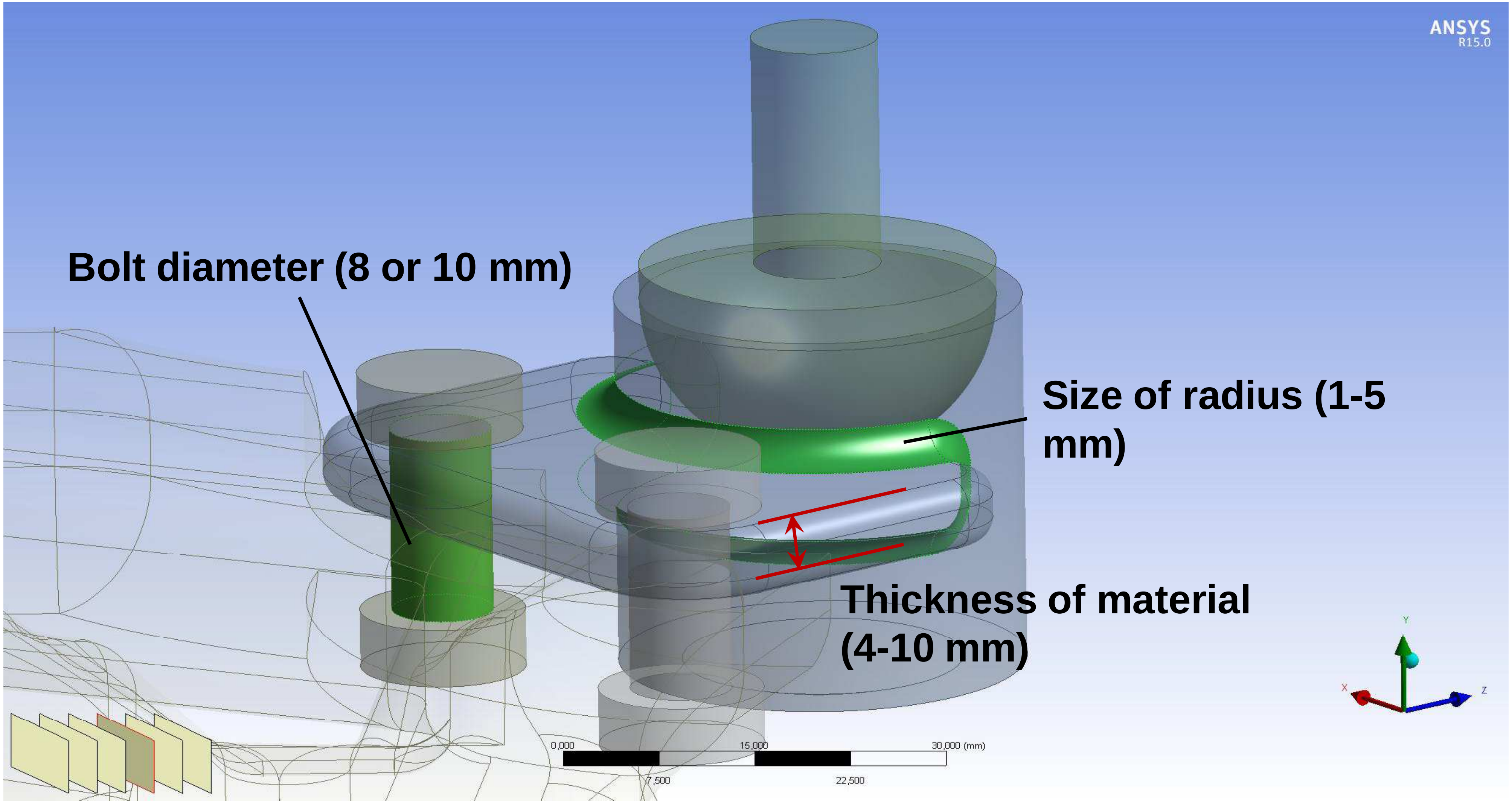
- The purpose of the simulation is to show the movement of the suspension parts when the wheel is lifted by 50 mm e.g. when the car is parked on a curbstone.

Required degrees of freedom:

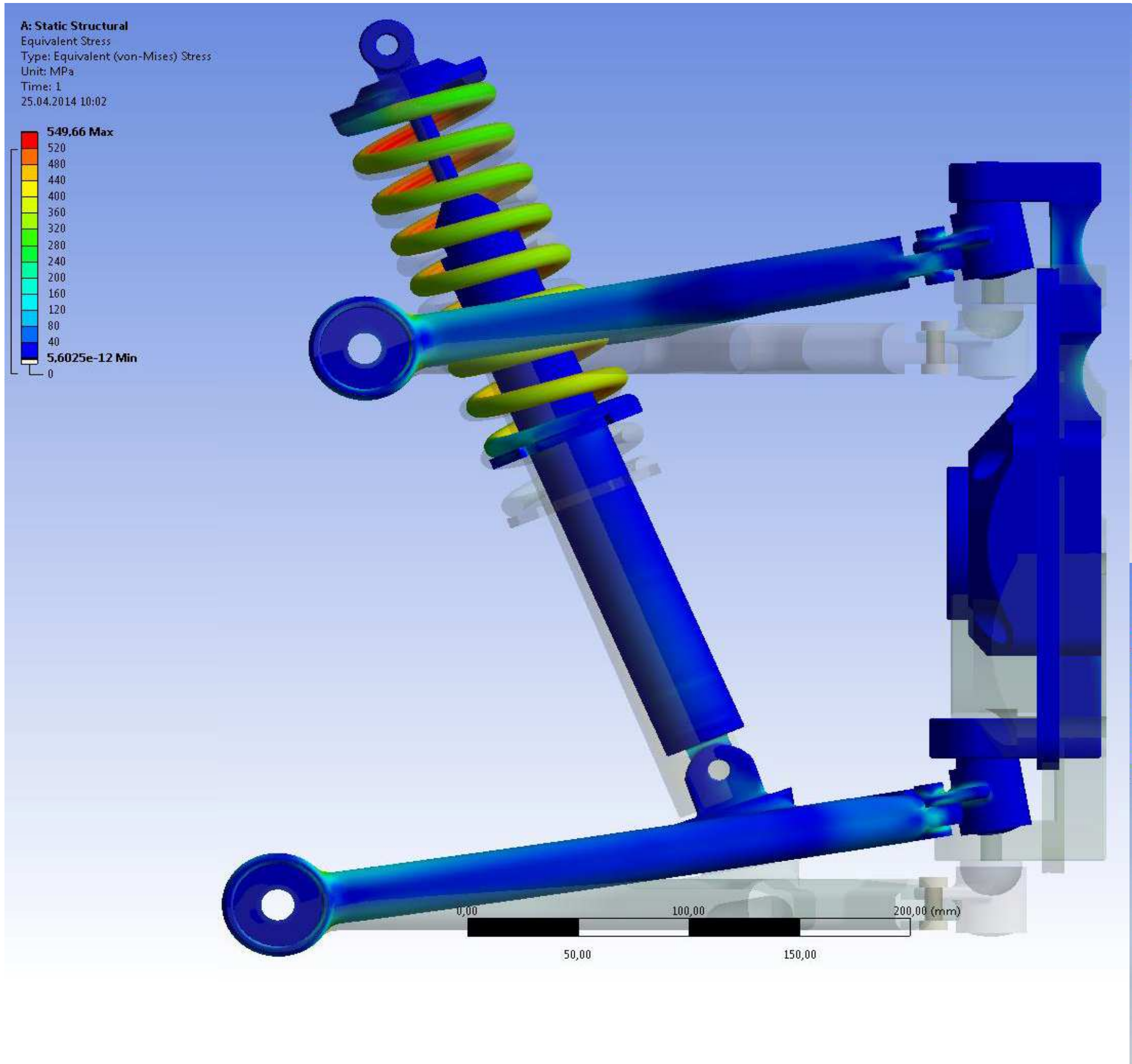
- upper chassis mount may rotate
- cylindrical joint between piston rod and damper housing
- revolute joint between strut and wishbone
- bushes are fixed via their center sleeves and flex due to movement
- spherical joints to represent the ball joints

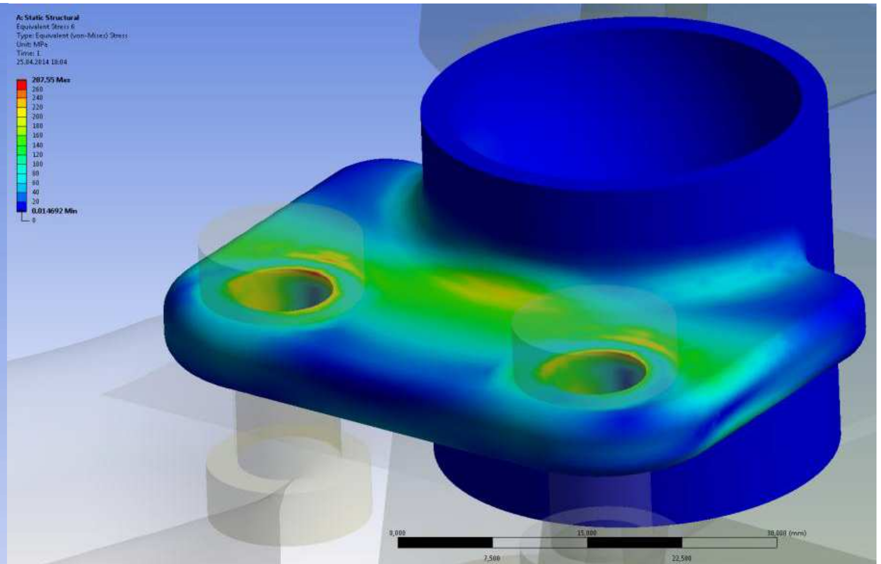


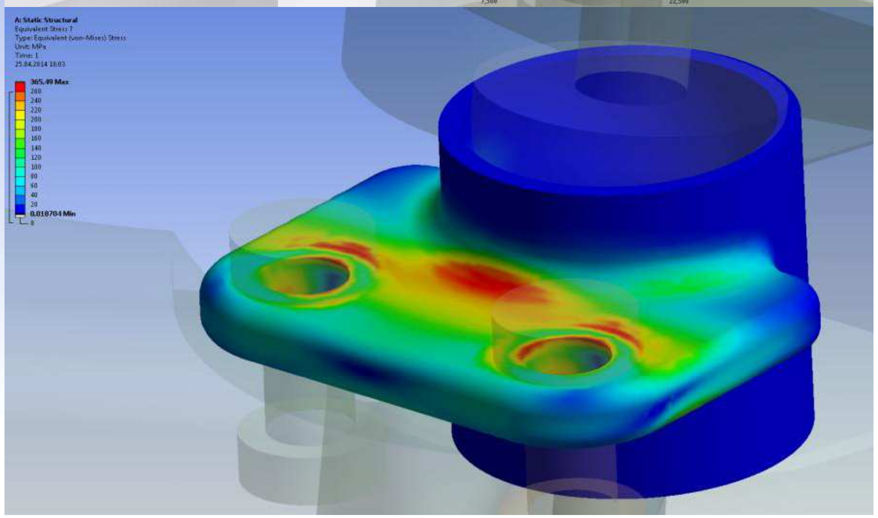
Simulation – Parametrization



Simulation – Basic results (von Mises Stress)

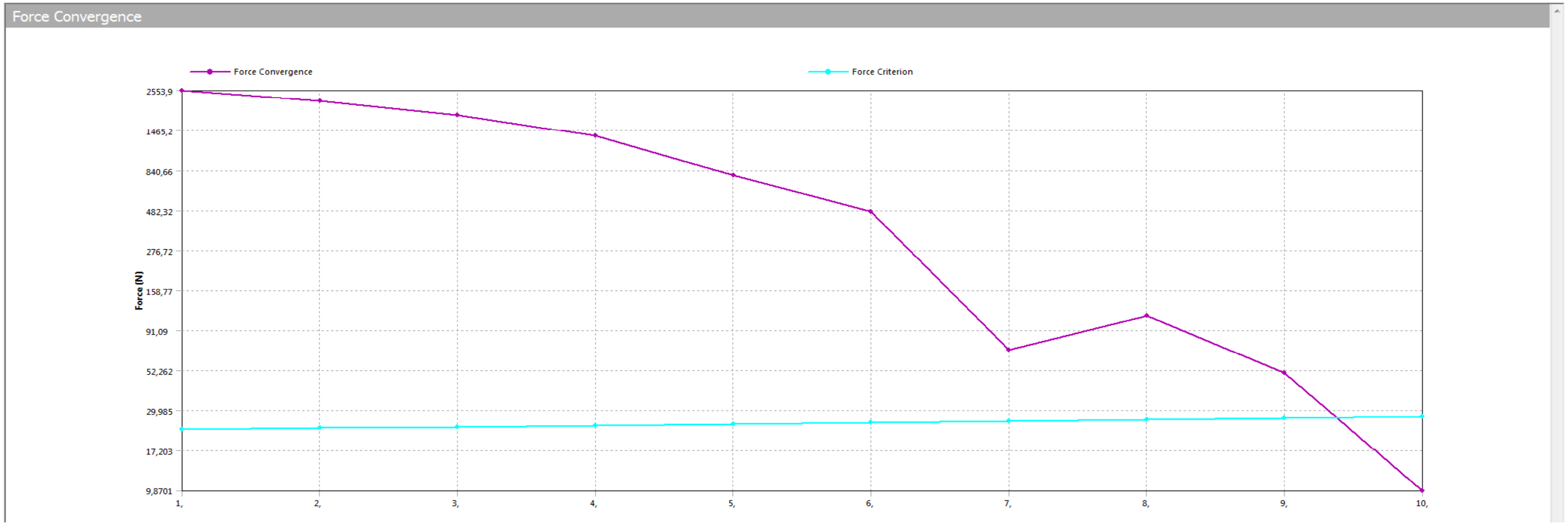




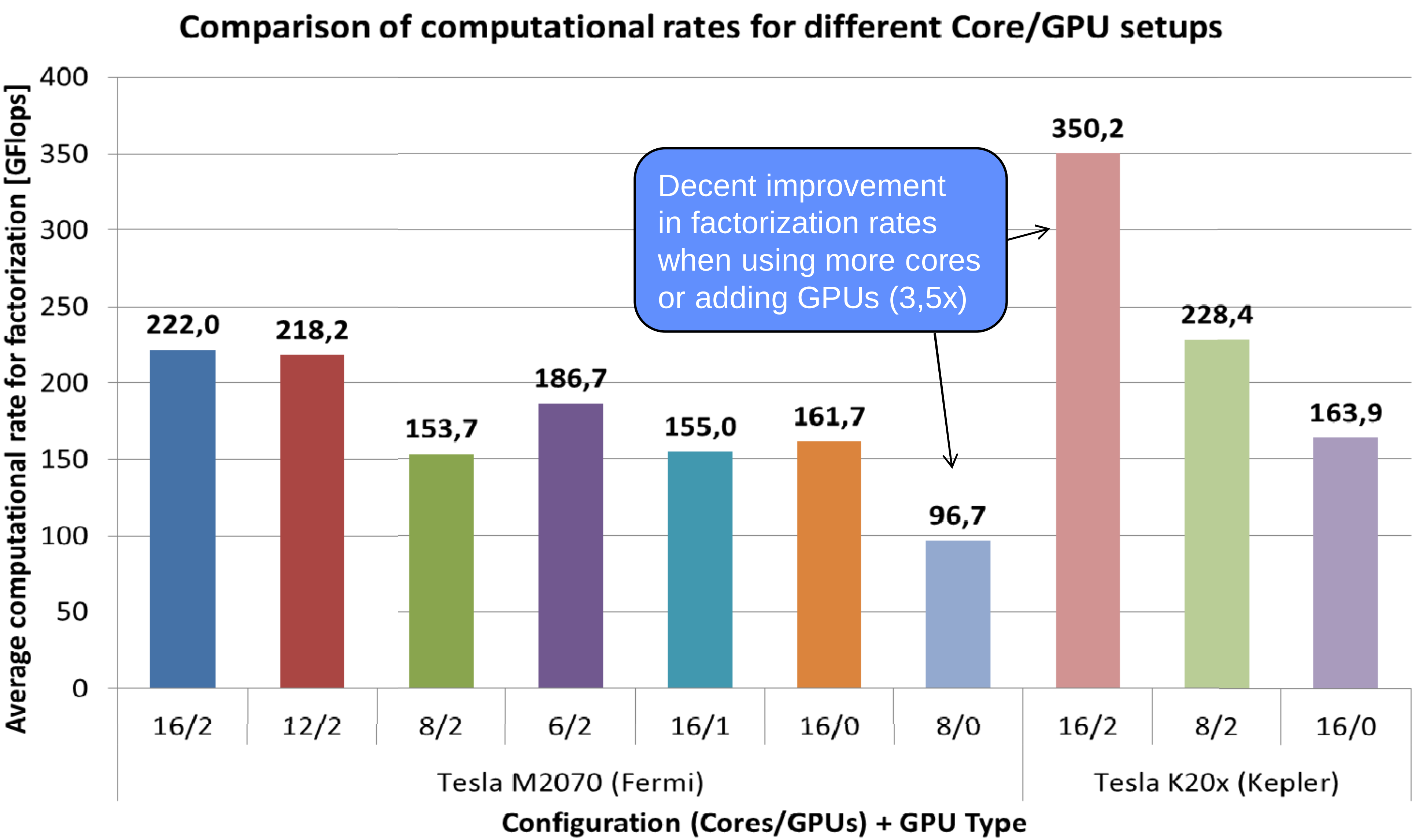


Simulation – Statistics

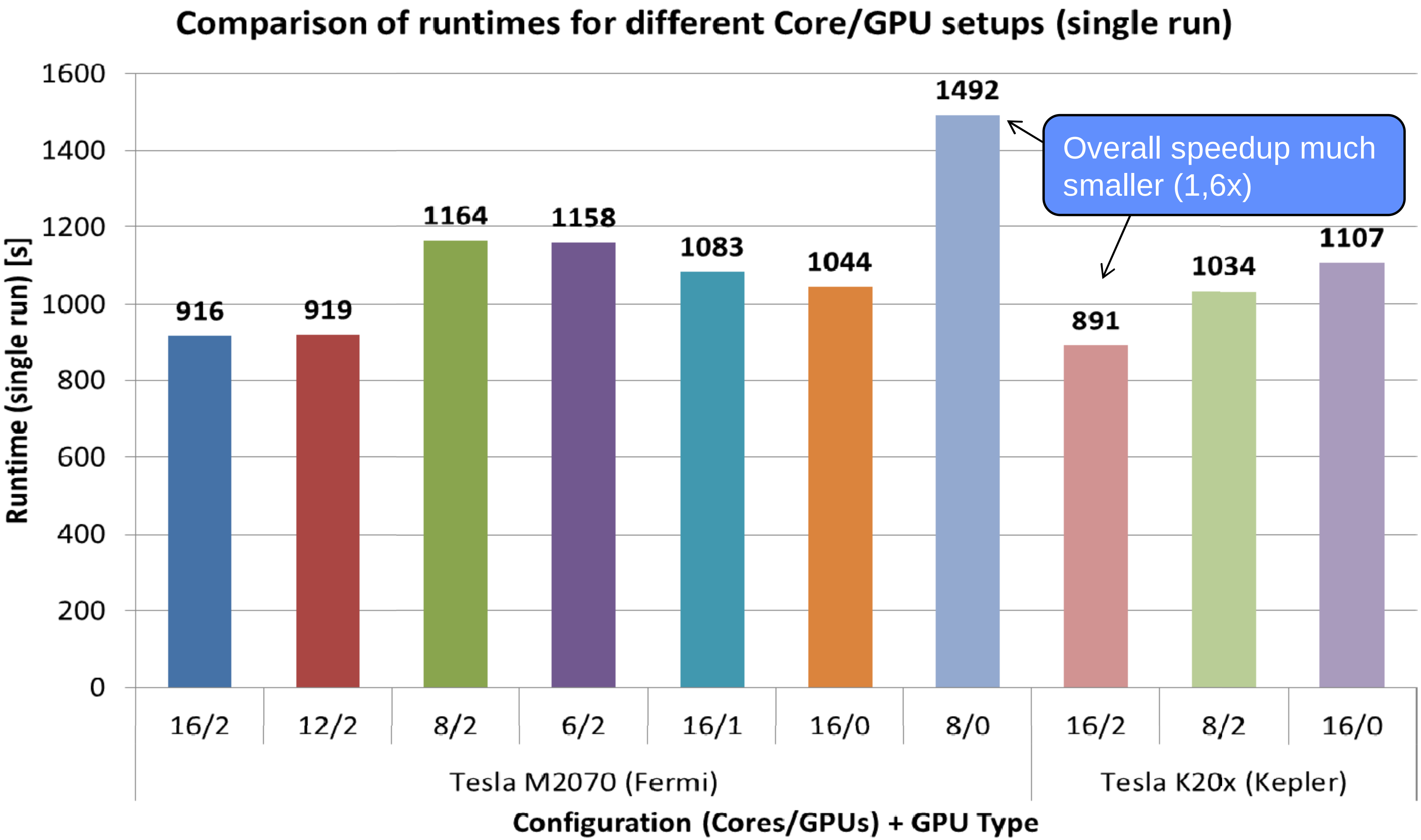
- Number of active bodies: **42** (Spring is defined as multi body part due to meshing problems)
- Number of nodes: **1.112.726**
- Number of Elements: **621.586**
- Number of contacts: **28**
- Solver: **Distributed Sparse**
- Total memory required for in-core solution: **40 GB**
- Total Number of Iterations: **10**



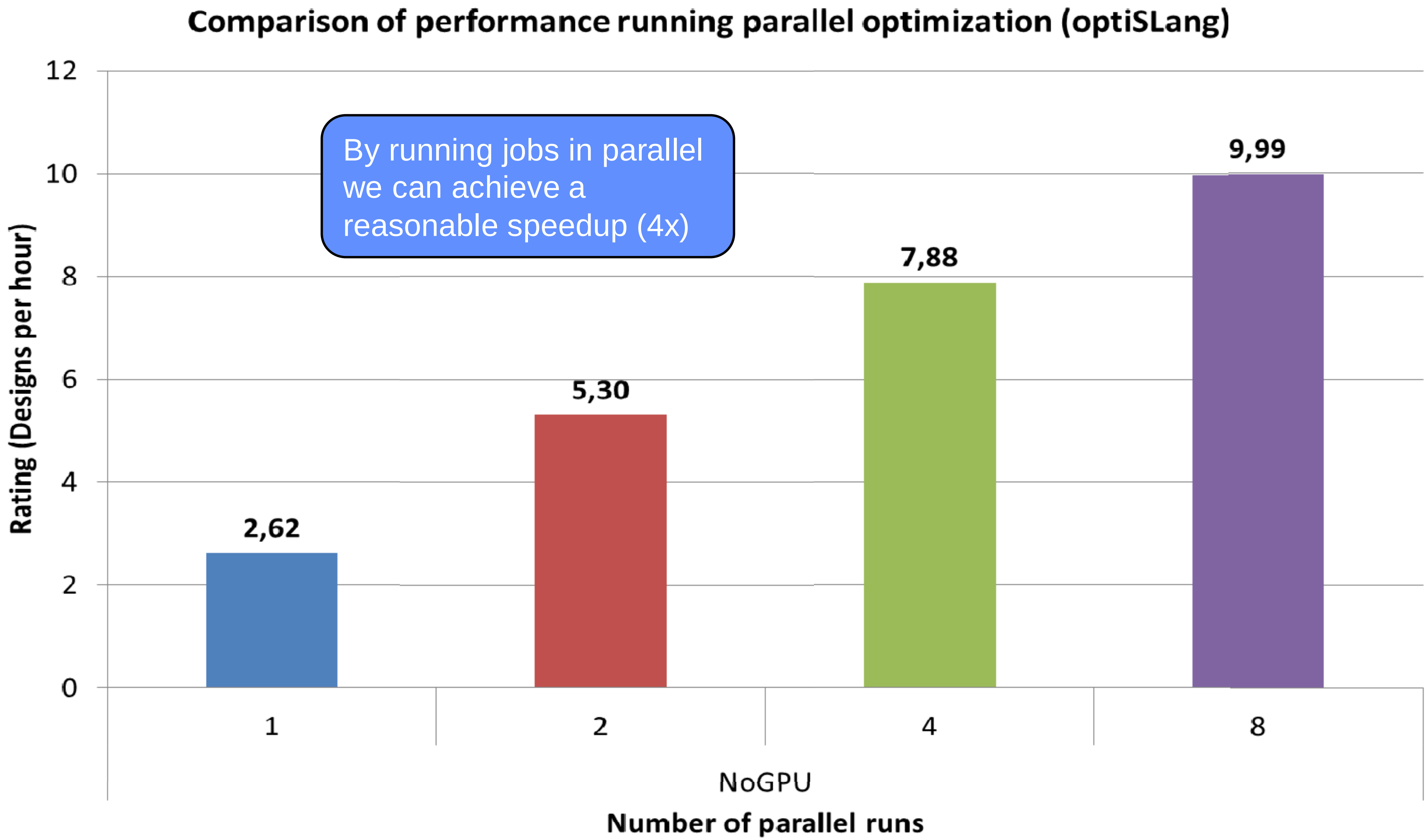
Simulation – Performance stats (single run)



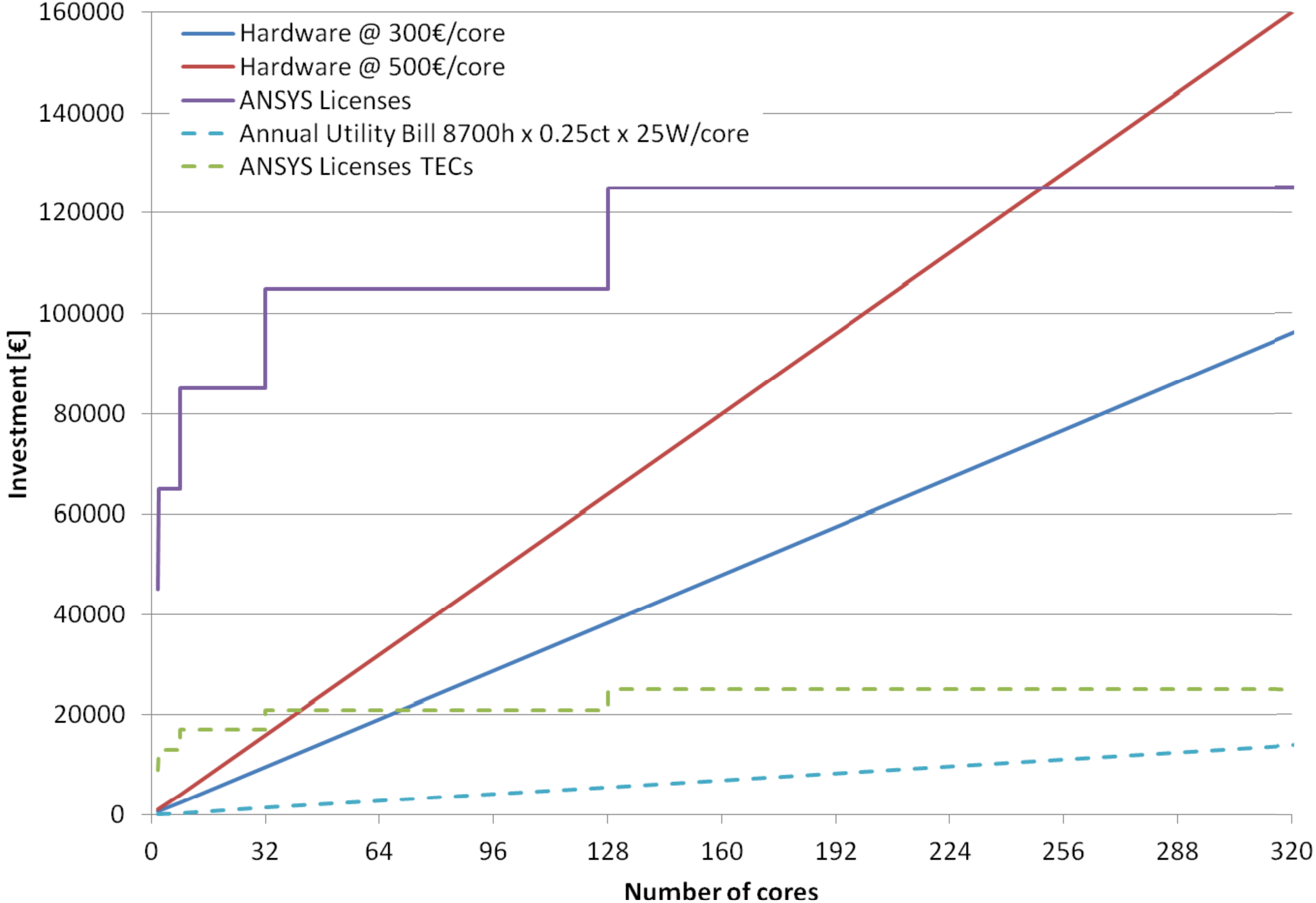
Simulation – Performance stats (single run)



Simulation – Performance stats (optimization)



Some thoughts about value:



Conclusions

- ANSYS Mechanical routinely delivers TFLOP per second performance in a HPC environment!
  - Works for a wide range of cases and setups
  - Validated with up to 512 cores (Sparse solver)
  - Models with up to 1250 MDOF could be solved (PCG)
  - Licensing cost is dominating factor (up to 256 cores)
  - Go for HPC to get the best return out of your licensing investment

Acknoweldgements

- Jeff Beisheim, ANSYS Inc.
- Erke Wang, Peter Tiefenthaler, CADFEM GmbH
- Natalja Schafet, Robert Bosch GmbH
- Philipp Schmid, Holger Mai, MicroConsult Engineering GmbH

http://www.cadfem.de/seminare/ansys/strukturmechanik/s/11979.html

29.4.2014 : 9:49 : +0200 Home → Seminare → ANSYS → Konstruktionbegleitend

Seminare

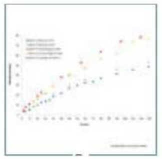
- Überblick
- ANSYS
  - Geometrie und Vernetzung
  - Konstruktionbegleitend
  - Strukturmechanik
  - Optimierung
  - Akustik
  - Elektromechanik
  - Temperaturfelder
  - Strömungsmechanik
  - Multiphysik
  - Update ANSYS 15.0
- LS-DYNA
- FTI FORMING SUITE
- Optimierung: optiSLang
- Medical / AnyBody
- Qualität und Recht
- CADFEM Customization
- Experten für Experten
- International Seminars
- Info-Veranstaltungen
- CADFEM esocaet
- Seminarsuche
- Individualtraining

1. Seminar auswählen! 2. Termin auswählen! 3. Daten eingeben! 4. Anmeldung Absenden!

Präsenz-Seminar:
Potentiale von High Performance Computing erfolgreich ausschöpfen

Wie setze ich HPC mit ANSYS erfolgreich ein und was habe ich davon? Im Kurs wird dieses Zusammenspiel anhand von vielen praktischen Beispielen beantwortet. Die Teilnehmer verstehen mit Abschluss des Seminars die geschwindigkeitssteigernden Faktoren einer Berechnung und können eigene Modelle im Hinblick auf die Rechenzeit optimal in ANSYS aufbauen.

- **Einführung**
  - Eignung der verschiedenen Solver zur Performance-Steigerung
  - Werkzeuge zur Performanceanalyse
  - Einflüsse des Modellaufbaus auf die Skalierungseigenschaften
  - Randbedingungen durch die Lizensierung
- **Praktischer Teil I**
  - Ausgewählte Beispiele aus der Praxis
- **Bedeutung der Hardware**
  - CPUs (Multicore vs. Taktfrequenz)
  - GPUs als Hardware Beschleuniger
  - Clustering von Hardware zur Leistungssteigerung
  - Optimierung von Ressourcen (SSDs, RAM)
  - Interconnects (Ethernet vs. Infiniband)
- **Modell-Tuning**
  - Design Rules für gut skalierende Modelle
  - Einfluss von Kontakten
- **Praktischer Teil II**
  - Übungen mit Beispielen



Termine und Anmeldung

Bitte wählen Sie hier Ihren Termin:

→ Seminarnummer: 13079
27.06.2014 Lein.-Echt. (DE)

→ Seminarnummer: 13080
12.09.2014 Grafing (DE)

→ Seminarnummer: 13081
05.12.2014 Lein.-Echt. (DE)

Uhrzeit
Tag 1: 09:00 - 17:00 Uhr

Allgemein

Voraussetzungen
Grundwissen in ANSYS

Vortragssprache
Deutsch

Sprache Unterlagen
Englisch

Referent
Herbert Güttler

Preis
595,-- EUR zzgl. MwSt.

Dauer
1 Tag(e)

Hinweis

Das Seminar widmet sich strukturmechanischen Anwendungen unter Verwendung von ANSYS Mechanical APDL und ANSYS Workbench Mechanical.

May 2009, Ansys12, 512 cores, 1 TFLOP per second

ANSYS 12.0 High-Performance Computing Capabilities Mean More Productive Use of Engineering Simulation

Major Advances in New Release Deliver Teraflop Performance

SOUTHPOINTE, Pa.--(BUSINESS WIRE)--May. 11, 2009-- ANSYS, Inc. (NASDAQ: ANSS), a global innovator of simulation software and technologies designed to optimize product development processes, today announced significant high-performance computing (HPC) milestones achieved with ANSYS® 12.0. The new release delivers impressive performance gains that enable product development teams to increase the value of simulation by considering large, high-fidelity models in shorter turnaround times. Key HPC achievements in the recent release include optimized parallel computing performance on multi-core processors, expanded support for large simulations, scaling breakthroughs, and support for parallel file systems. These product enhancements deliver best-in-class HPC capability for multiphysics simulations, engineered to scale from multi-core desktop workstations to departmental clusters and large enterprise supercomputers.

"Engineers today are including more geometric details in their analyses and looking for a more-realistic treatment of physical phenomena. Both of these factors drive the need for outstanding performance on the latest computing platforms," said Jim Cashman, president and CEO of ANSYS, Inc. "ANSYS 12.0 is our most HPC-capable release to date, and it delivers tremendous value to customers who need faster turnaround and the ability to consider high-fidelity multiphysics simulations. HPC is an important part of Simulation Driven Product Development™, since it delivers efficiency and productivity improvements to engineering organizations seeking to gain a competitive advantage in today's business climate."

ANSYS 12.0 technology incorporates optimization for the latest multi-core processors and benefits greatly from recent improvements in processor architecture, resulting in highly efficient use of parallel processing to reduce the turnaround time for simulation. Improved algorithms for model partitioning, combined with optimized communications and load balancing between processors, have yielded additional parallel scaling breakthroughs. For structural mechanics analyses, dramatically improved scaling is observed on desktop parallel systems, and a major performance milestone of over 1 teraflop has been achieved via cluster computing on 512 cores. Teraflop performance is well over 100 times faster than the fastest single-core performance currently observed, reducing run times from days to minutes on the most challenging simulations. For fluid dynamics simulations, nearly ideal linear speedup has been demonstrated out to 1,024 cores, roughly doubling the core count for ideal scaling compared to previous releases. At 2,048 cores, scaling remained at approximately 80 percent of ideal linear performance. ANSYS FLUENT® 12.0 software introduces parallel input and output (I/O) of files, dramatically reducing turnaround time for large simulations that involve extensive I/O and removing file handling as a bottleneck for scaling on large clusters. In addition, ANSYS 12.0 technology includes important enhancements that enable larger simulations than ever before, setting the stage for customers to consider highly detailed physical phenomena and full-assembly models in their fluids or structural simulations.