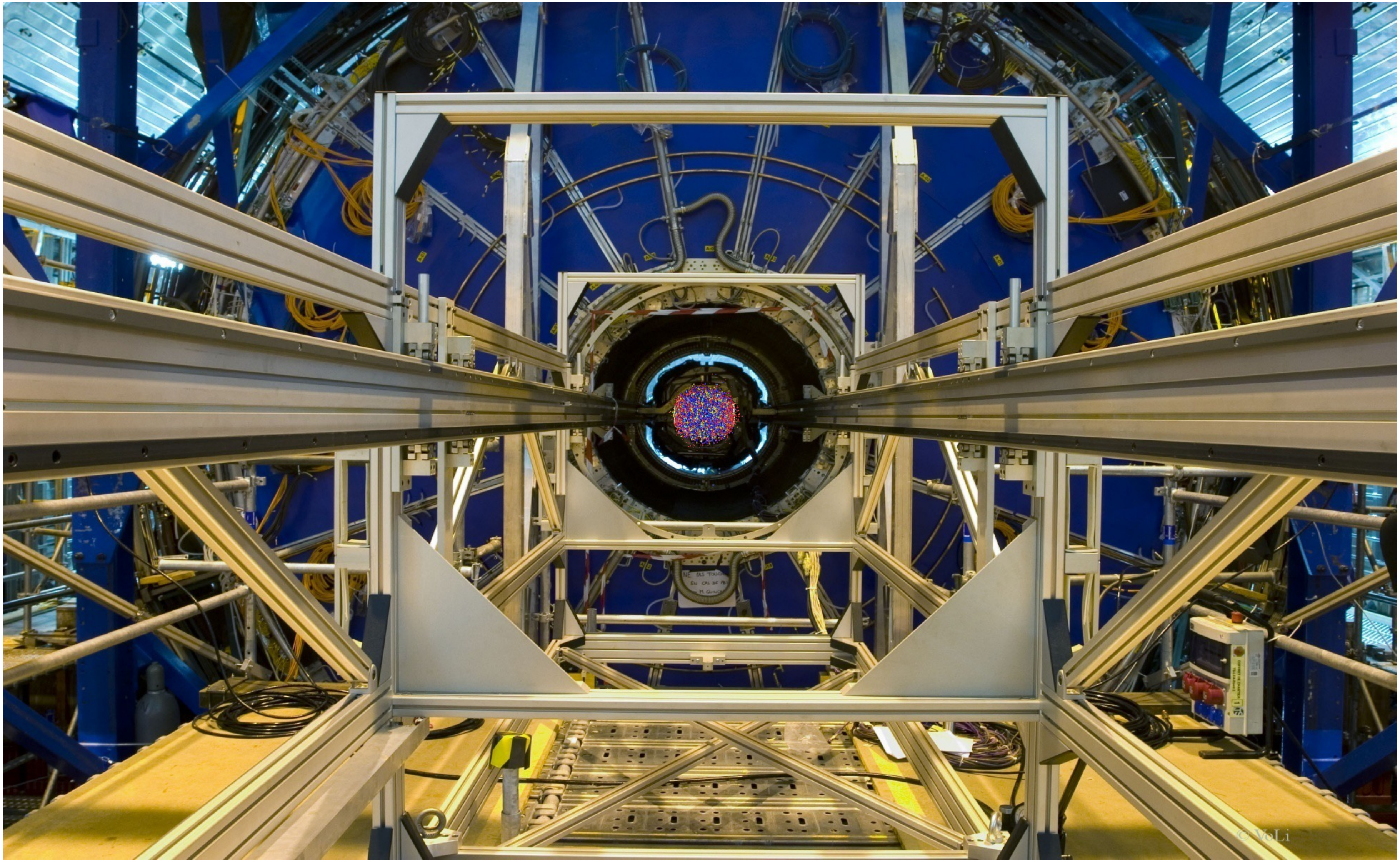
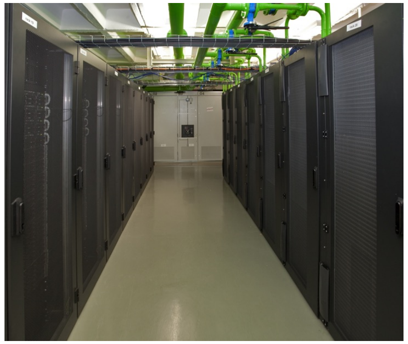


Energy Efficient Computing

Prof. Dr. Volker Lindenstruth
FIAS, IfI, LOEWE Professur
Chair of HPC Architecture
University Frankfurt, Germany
GSI Helmholtzcenter
Phone: +49 69 798 44101
Fax: +49 69 798 44109
Email: voli@compeng.de
WWW: www.compeng.de

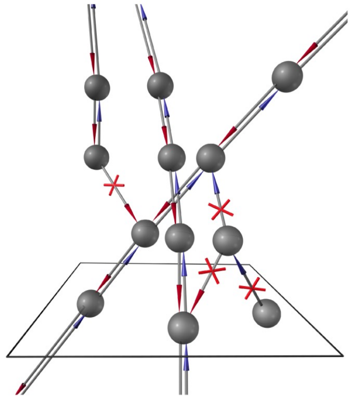
Three main areas for optimization

- **Data Center Architecture**

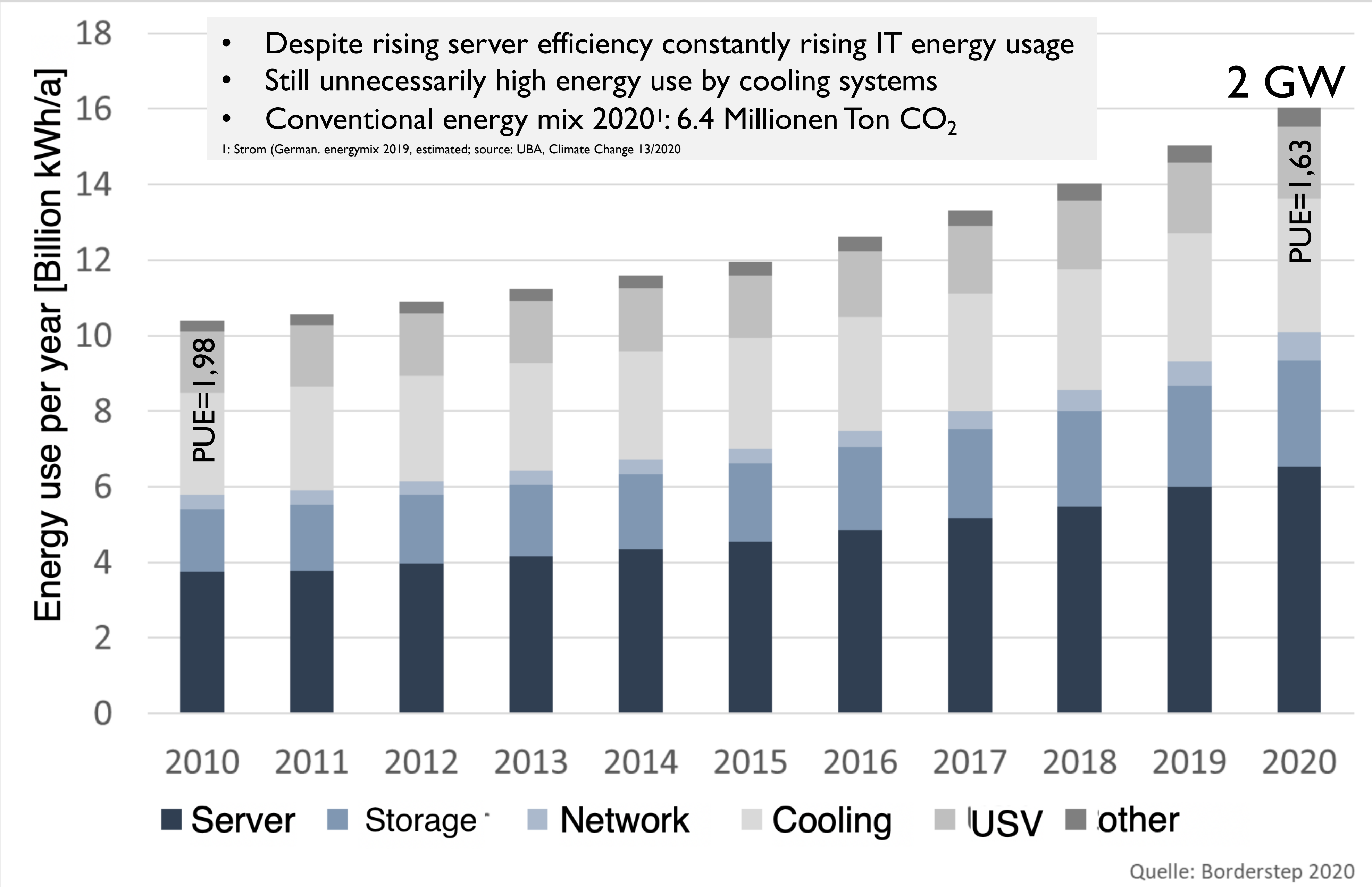


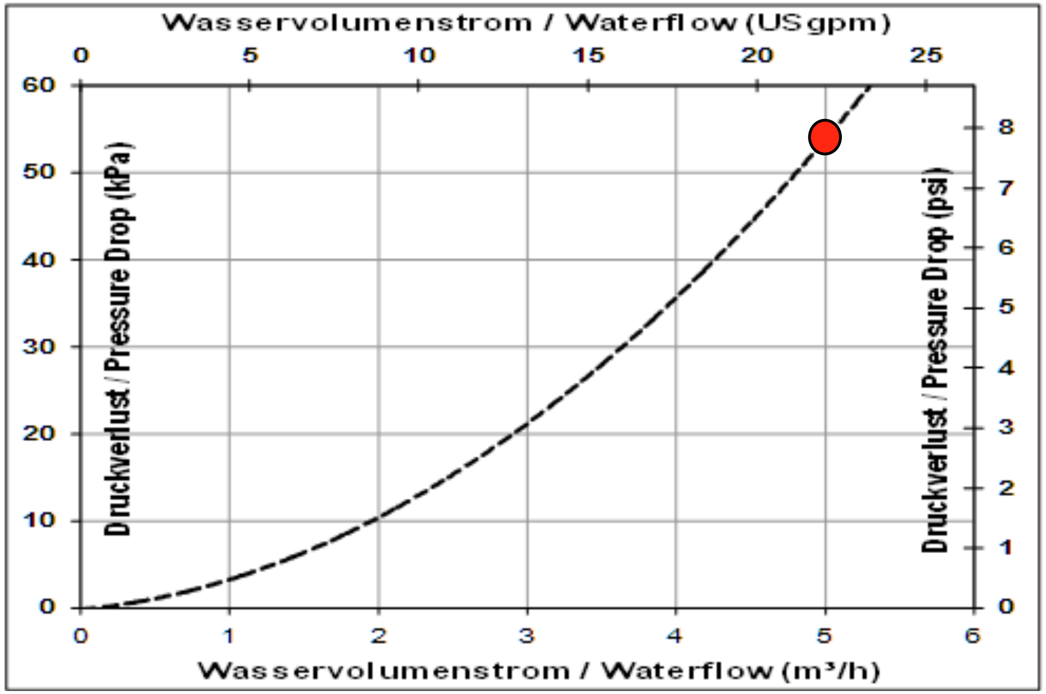
- **Computer Architecture**

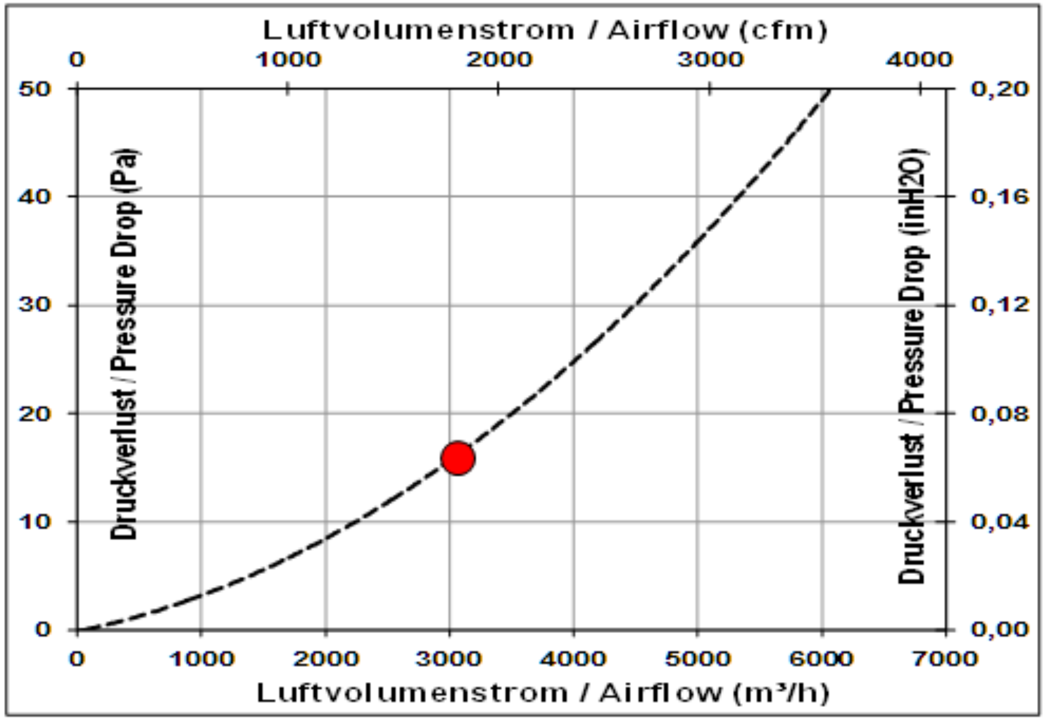
- **Algorithm**



IT Energy use Germany



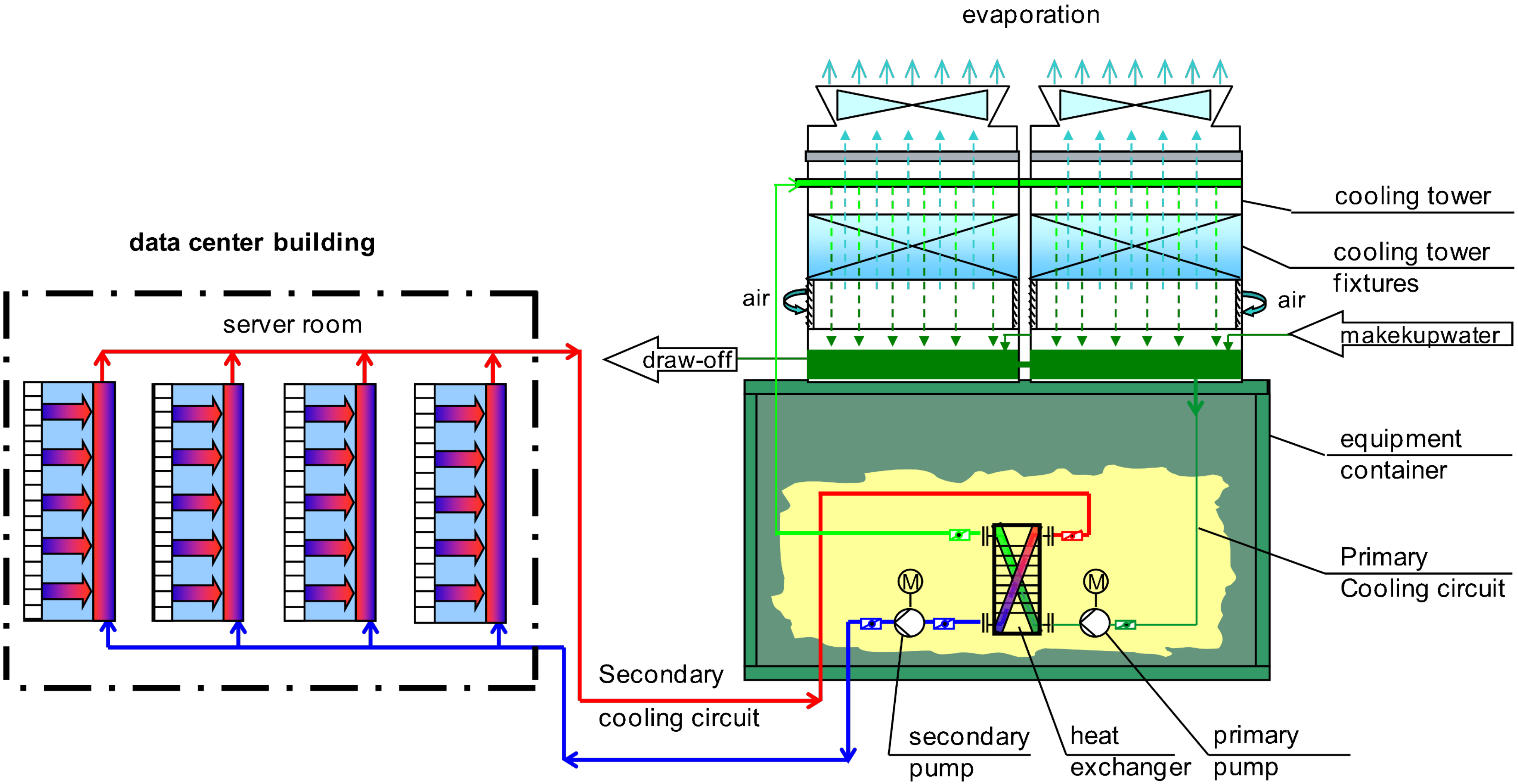


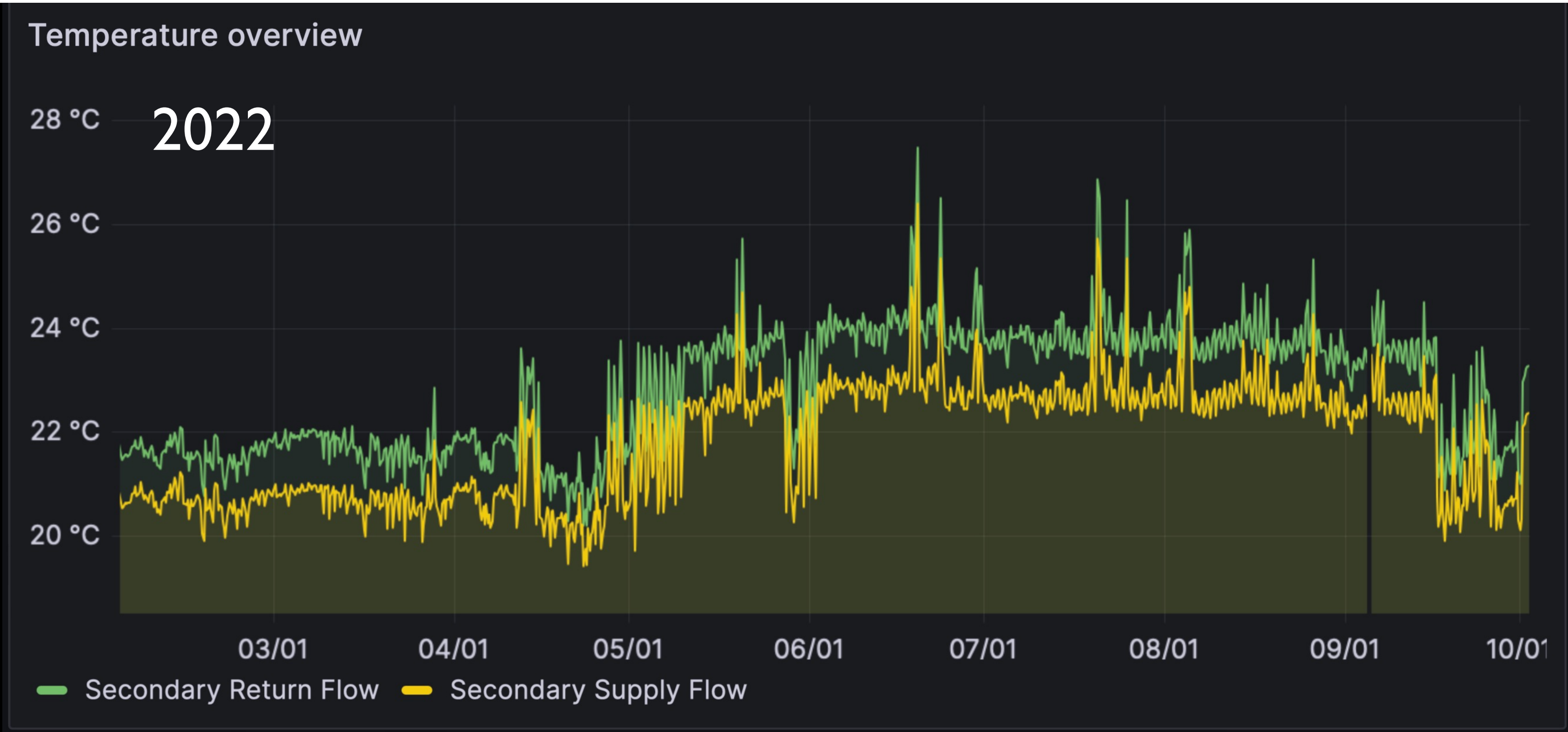


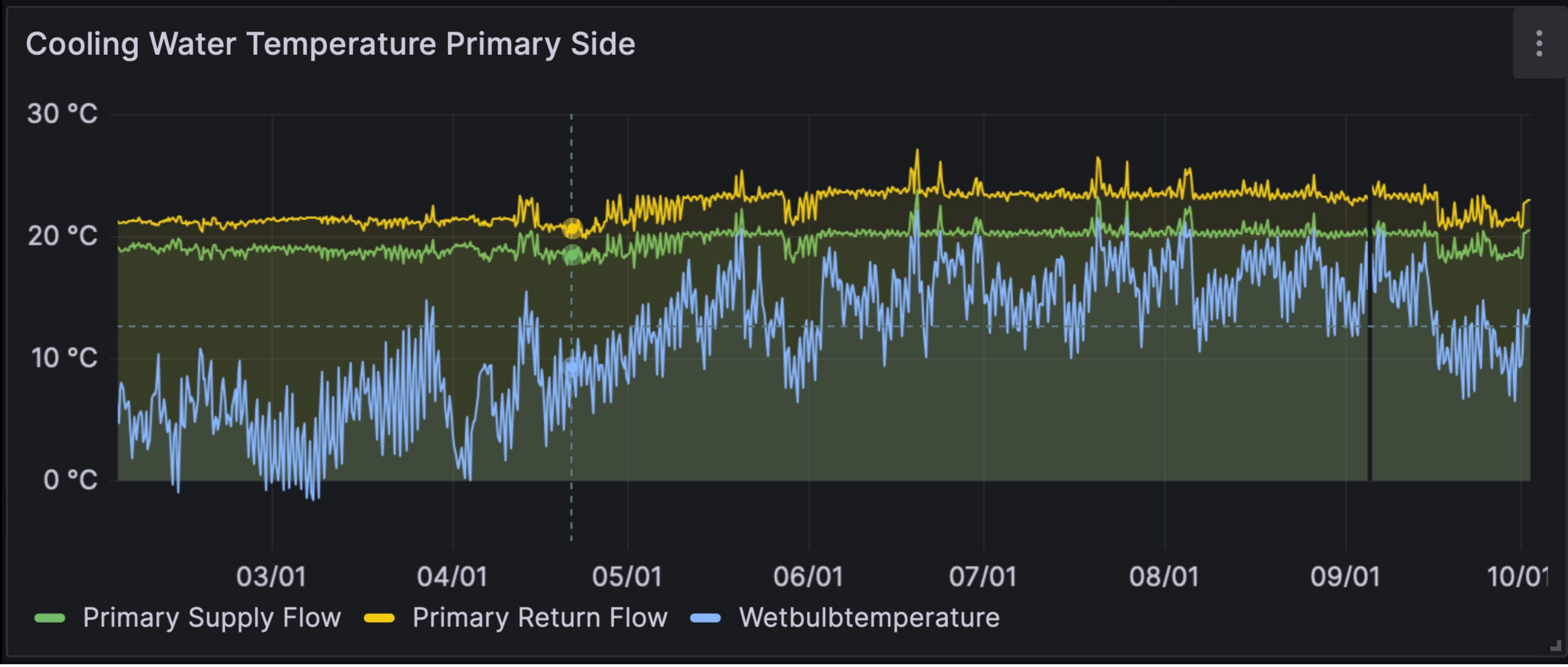
Up to 75 kW/rack

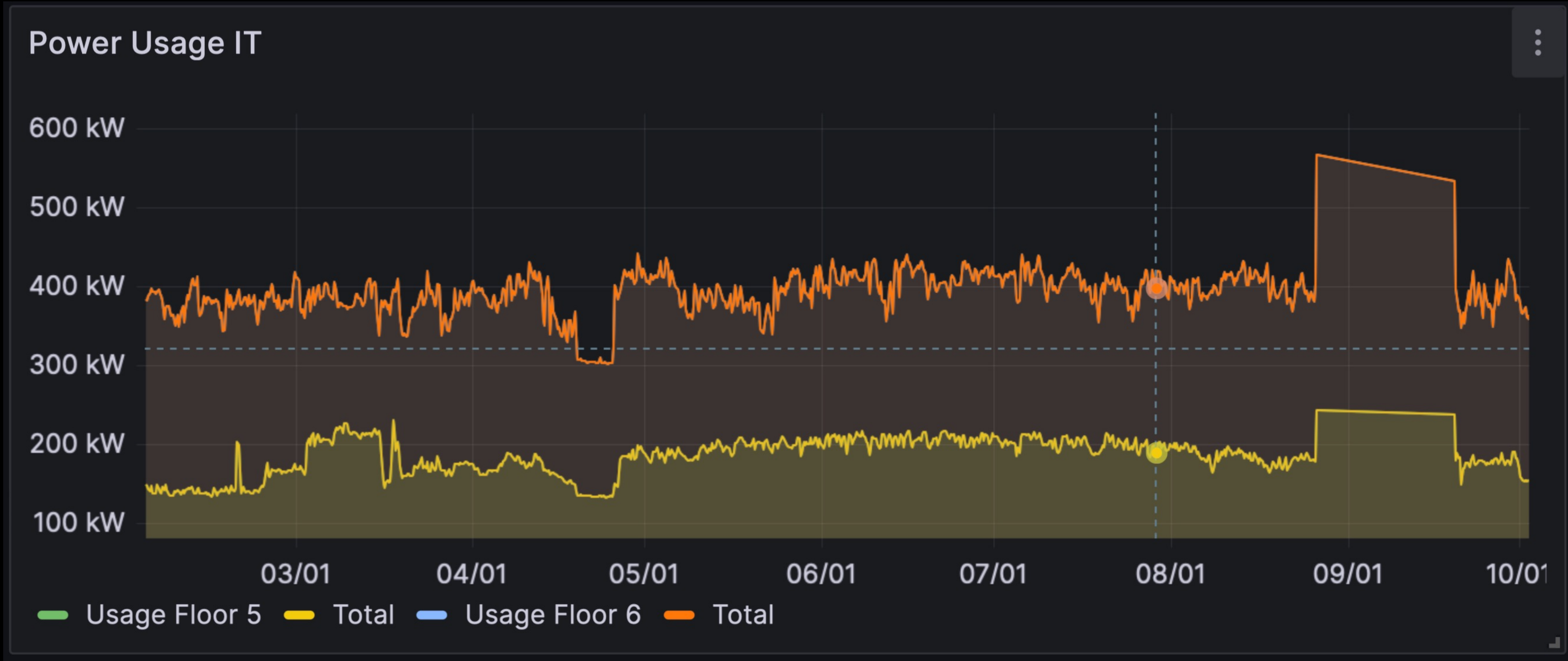
knürr environments for electronics

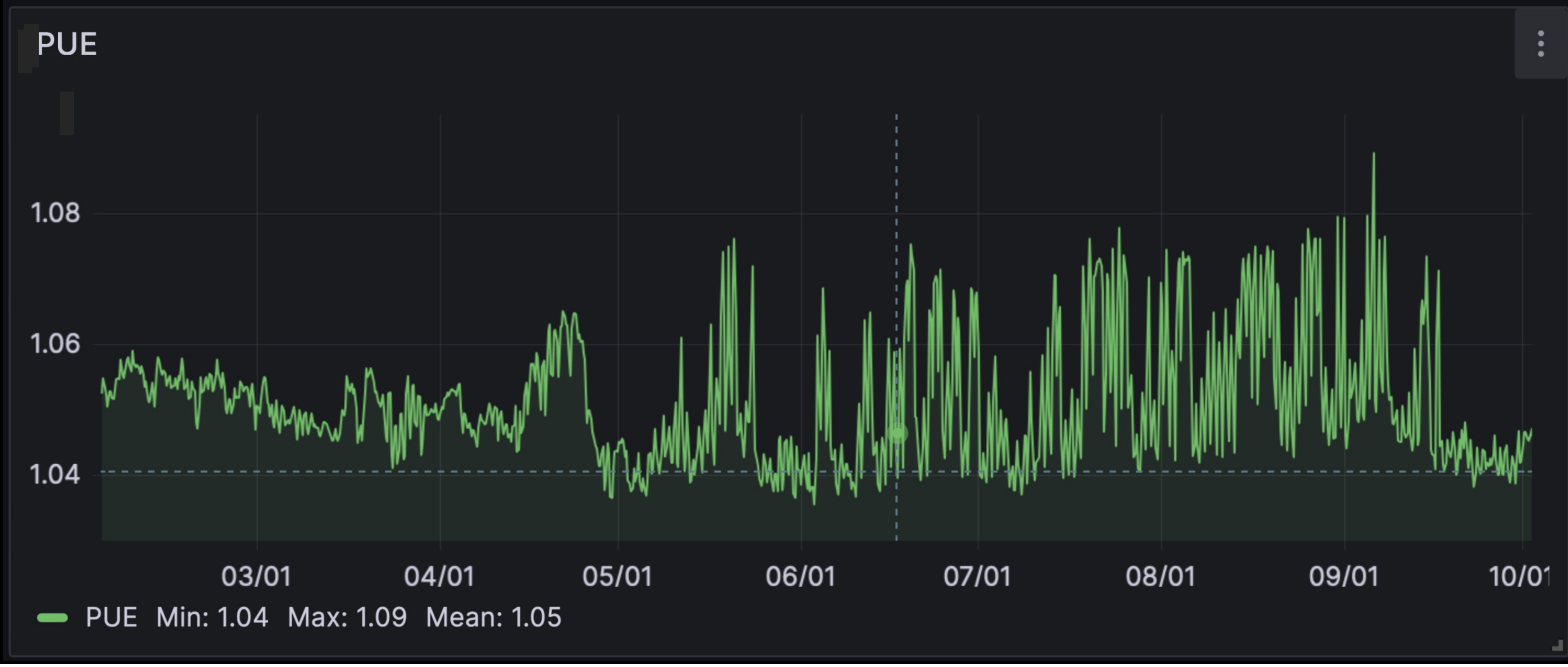
Heiko Ebermann


evaporation
data center building
server room
cooling tower
cooling tower
fixtures
air
air
draw-off
makekupwater
equipment
container
Primary
Cooling circuit
Secondary
cooling circuit
secondary
pump
heat
exchanger
primary
pump


Temperature overview
2022
28 °C
26 °C
24 °C
22 °C
20 °C
03/01
04/01
05/01
06/01
07/01
08/01
09/01
10/01
Secondary Return Flow
Secondary Supply Flow


Cooling Water Temperature Primary Side
30 °C
20 °C
10 °C
0 °C
03/01
04/01
05/01
06/01
07/01
08/01
09/01
10/01
Primary Supply Flow
Primary Return Flow
Wetbulbtemperature


Power Usage IT
600 kW
500 kW
400 kW
300 kW
200 kW
100 kW
03/01
04/01
05/01
06/01
07/01
08/01
09/01
Usage Floor 5
Total
Usage Floor 6
Total


PUE
1.08
1.06
1.04
03/01
04/01
05/01
06/01
07/01
08/01
09/01
PUE Min: 1.04 Max: 1.09 Mean: 1.05

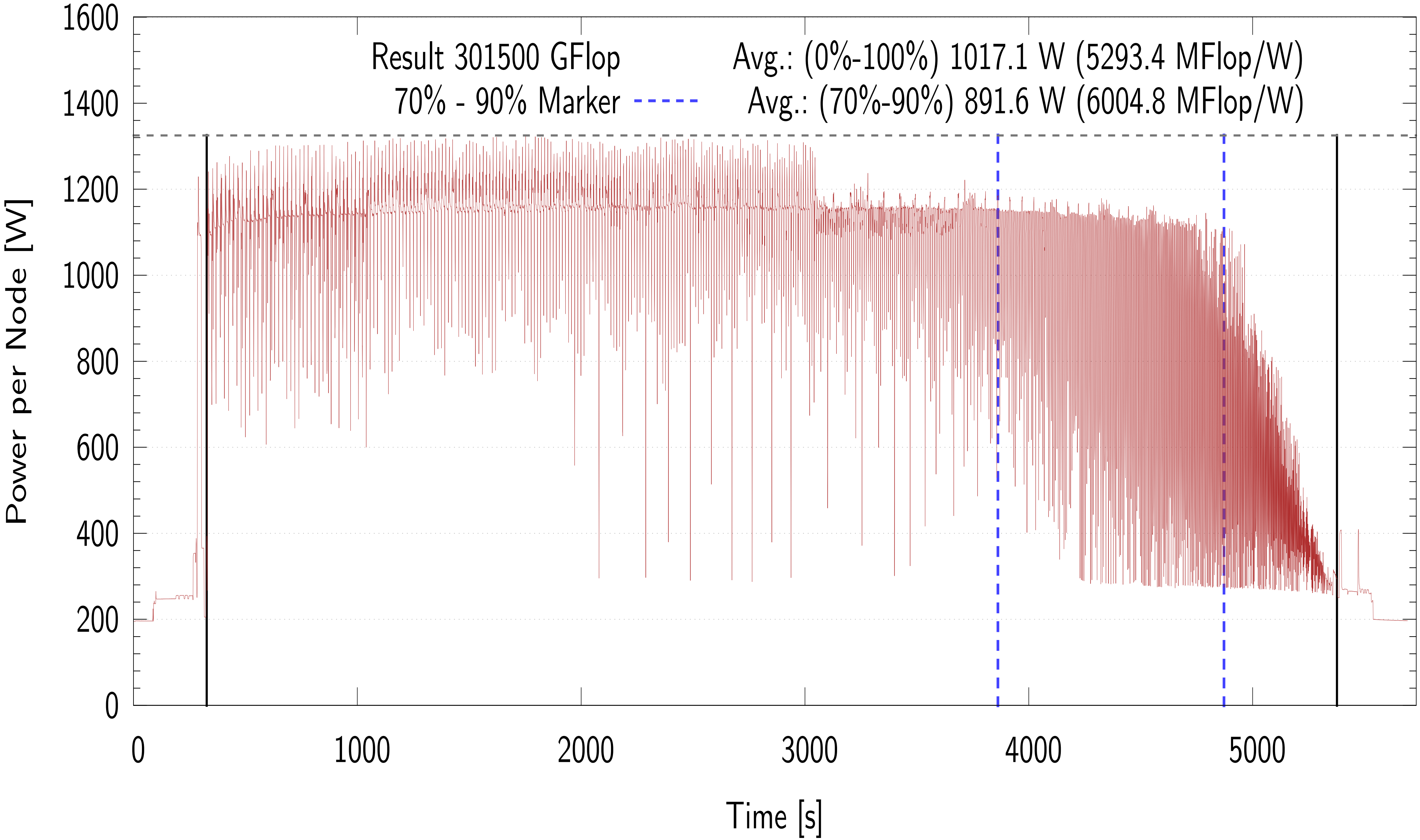
Awards

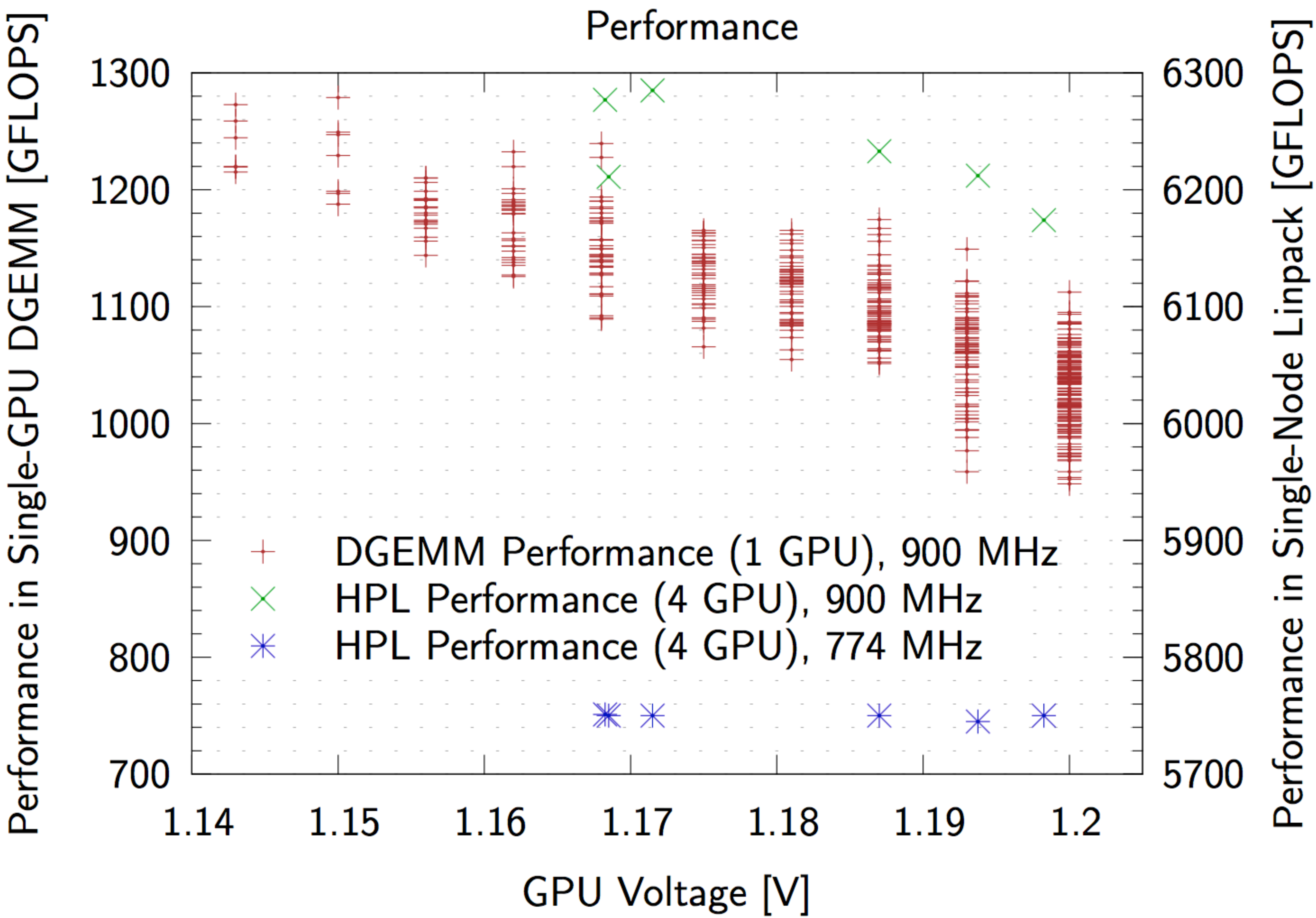
- Green-IT Award Bundesregierung „Visionäre Gesamtkonzepte“
- German data center price 2012 – energy efficiency
- German data center price 2013 – Visionary data center architecture
- Nominated for German data center price 2014 – energy efficiency
- DataCenterDynamics EMEA Award 2013 – Data Center Blueprint
- BroadGroup EMEA Awards Special Commendation – Energy Efficiency
- „Land of Ideas“2012 for LOEWE-CSC
- *Green Cube* Project the Month, BMBF
- 5 nominations with 4 second positions for Data Center Dynamics EMEA Awards – 2011, 2012, 2013
- 2. rank at German Internet award 2012
- 1. rank DataCloud Awards 2015, Monaco
- Blauer Engel
- …

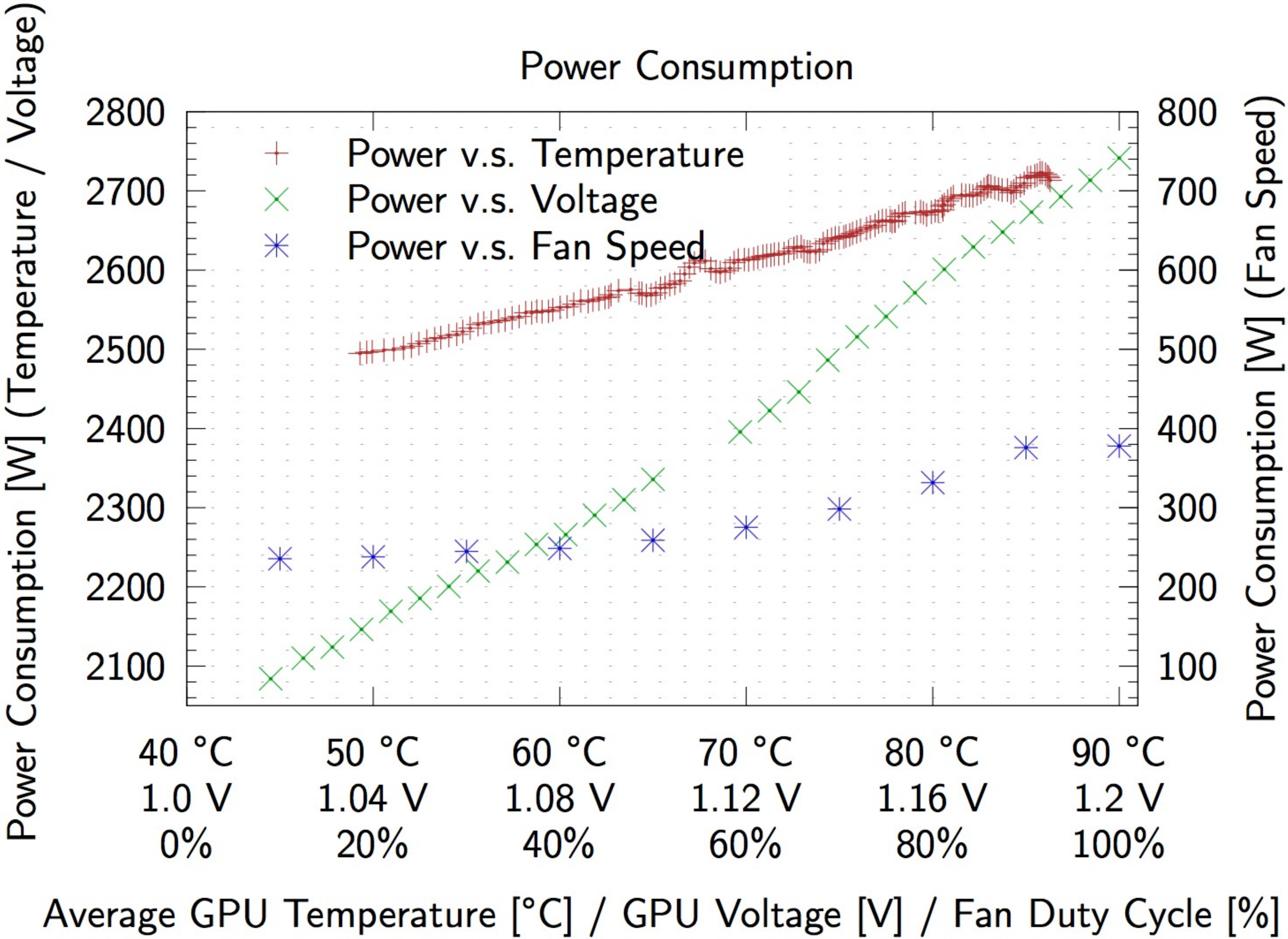
COMPUTER ARCHITECTURE


L-CSC
Green500 Rankings:
#8: LOEWE-CSC 2010
#2: SANAM 2012
#1: L-CSC 2014
#9: GOETHE-HLR 2023
The Green500 Top 1
November 2014
#1 L-CSC-GSI Helmholtz Center
#2 Suiren-High Energy Accelerator Research Organization /KEK
5200 MF/W
@ 1000 TF

HPL power profile




Performance
Performance in Single-GPU DGEMM [GFLOPS]
Performance in Single-Node Linpack [GFLOPS]
GPU Voltage [V]
1300
1200
1100
1000
900
800
700
6300
6200
6100
6000
5900
5800
5700
1.14
1.15
1.16
1.17
1.18
1.19
1.2
DGEMM Performance (1 GPU), 900 MHz
HPL Performance (4 GPU), 900 MHz
HPL Performance (4 GPU), 774 MHz



Power Consumption
Power v.s. Temperature
Power v.s. Voltage
Power v.s. Fan Speed
2800
2700
2600
2500
2400
2300
2200
2100
800
700
600
500
400
300
200
100
Power Consumption [W] (Temperature / Voltage)
Power Consumption [W] (Fan Speed)
40 °C
1.0 V
0%
50 °C
1.04 V
20%
60 °C
1.08 V
40%
70 °C
1.12 V
60%
80 °C
1.16 V
80%
90 °C
1.2 V
100%
Average GPU Temperature [°C] / GPU Voltage [V] / Fan Duty Cycle [%]

Computer Architecture

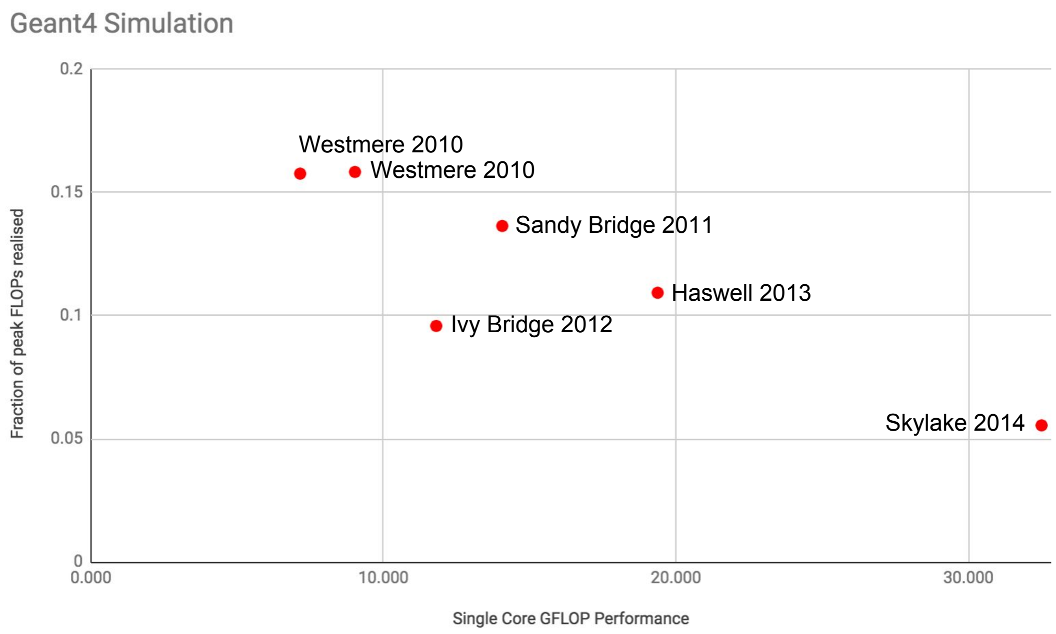
- Single core CPU performance scales only slowly
- Good CPU performance requires vectorization
- In general GPU power effectiveness much larger than CPU power effectiveness
- CPU cost for real applications >7x of GPU cost
- Well vectorized code will run also very well on GPU
- GPU optimized code will also run much faster on CPU
- GPU power depends quadratically on the supply voltage → optimize supply voltage and clock of all GPUs and find optimum
- Server optimizations:
  - » watch power of auxiliary devices (switch off)
  - » watch fan power (efficiency drops with square of RPM)
  - » Avoid 1U server and counter rotating fans
  - » watch airflow inside the server
  - » Optimize similar speed of servers for parallel code

ALGORITHMIC OPTIMIZATIONS

Moores Law ...

How is our Code Doing? Simulation on 5 years of Intel CPUs

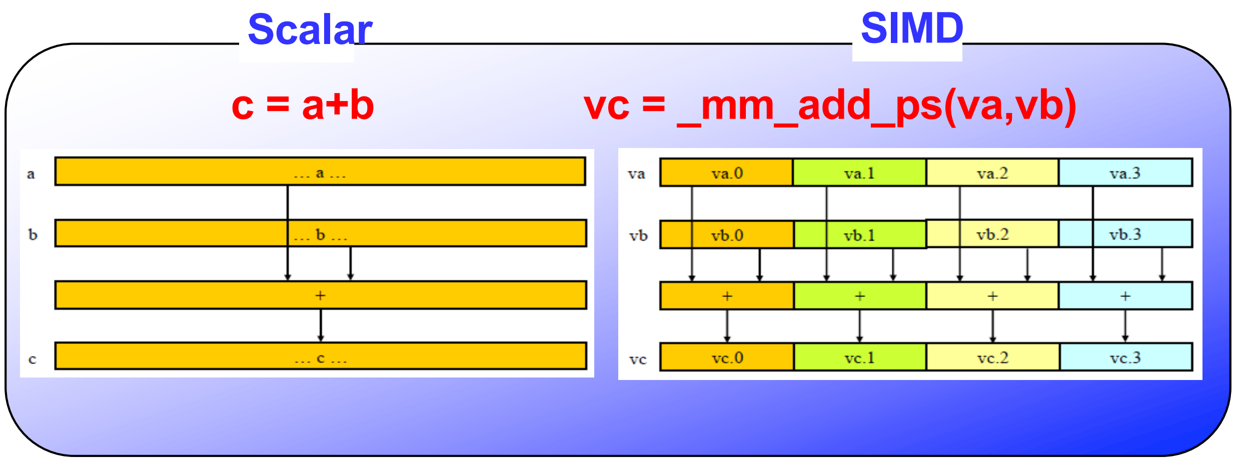
- Fraction of the potential floating point performance we use has been dropping over time
- CPU manufacturers add wider vectors that we do not take advantage of, or deep pipelines where cache misses are very costly
- Confirms what we have long suspected about the growing performance gap on modern architectures



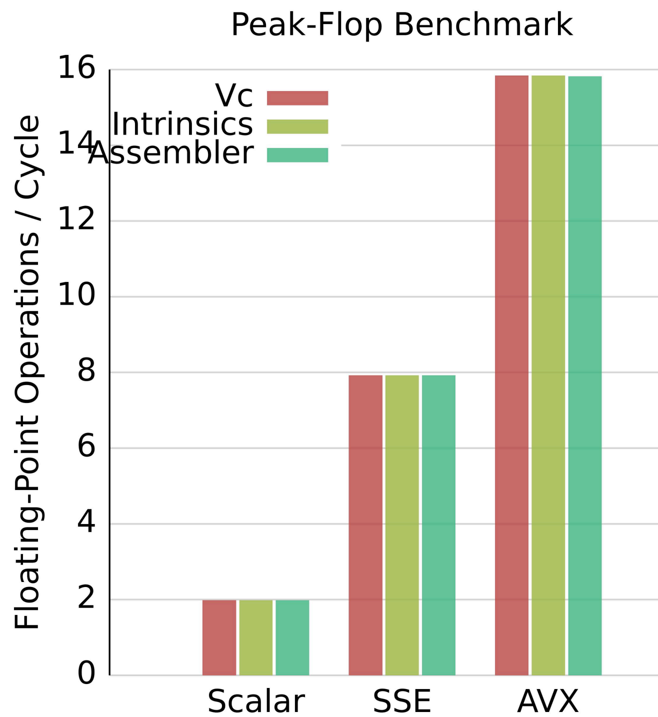
Private ttbar event test in the ATLAS detector with Geant4 10.1

4

Vector Classes (Vc)



Vector classes overload scalar C operators with SIMD/SIMT extensions



Vector classes:

- **provide full functionality for all platforms**
- **support the conditional operators**

Vc increase the speed by the factor:

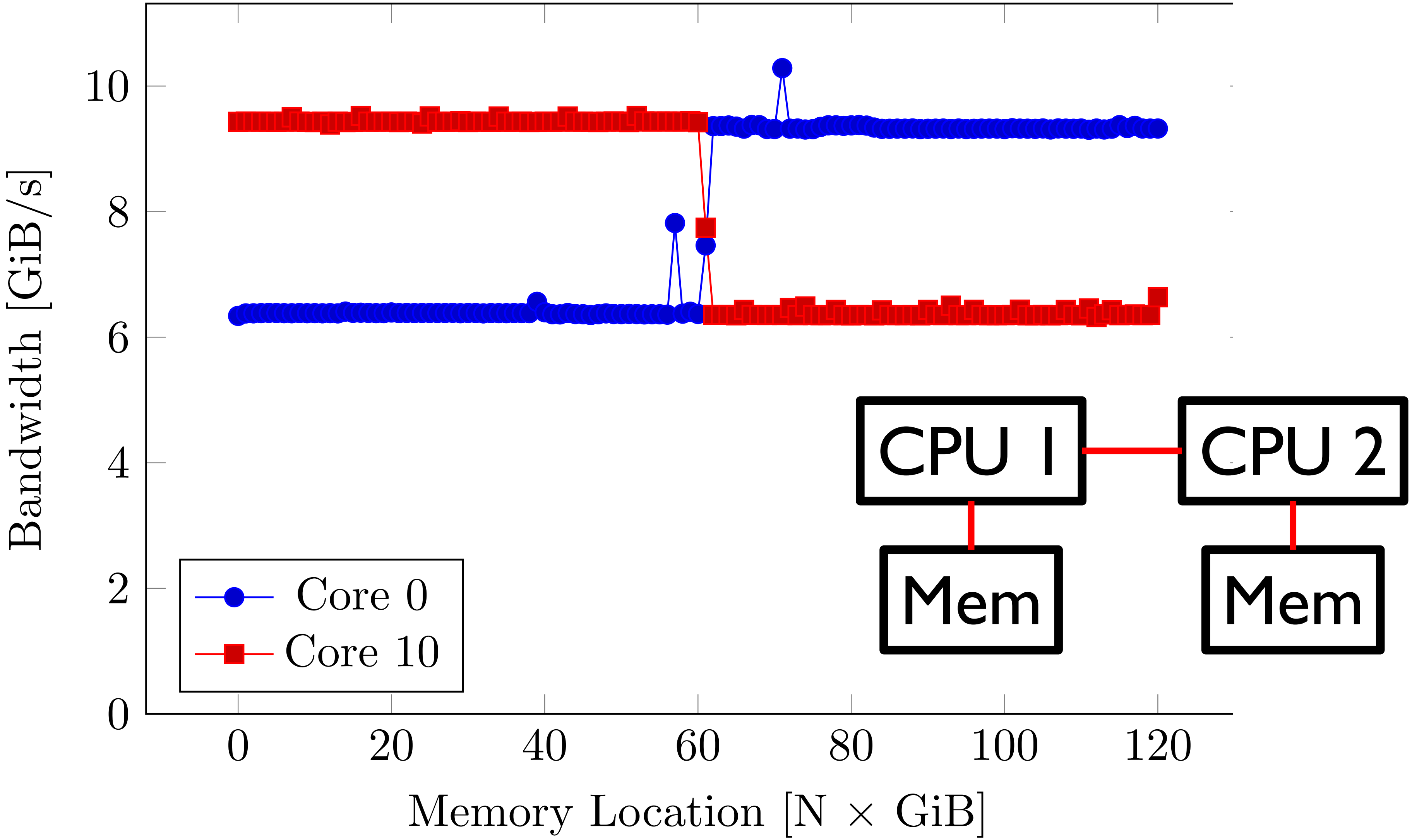
- **SSE2 – SSE4** **4x**
- **CPUs** **8x, 16x**
- **GPGPUs** **>16x**

phi(phi<0) += 360;

Vector classes enable easy vectorization of complex algorithms

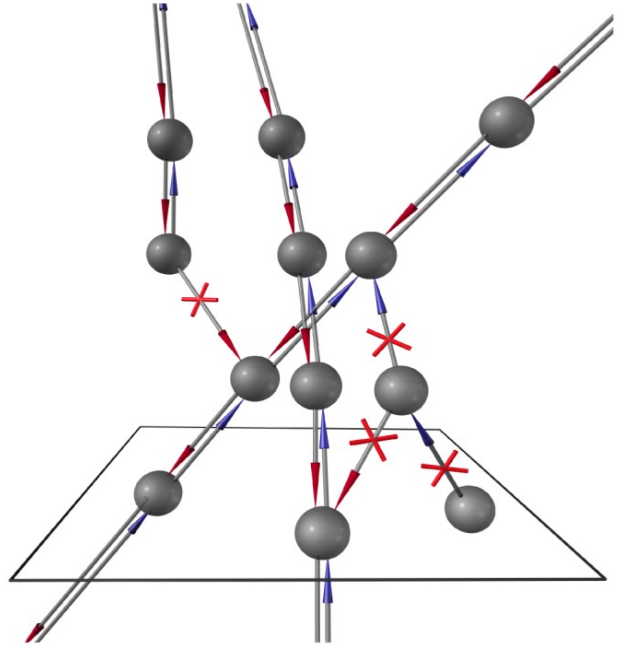
- Standard ISO/IEC 19570:2018 is now based on Vc
- Extending {C++} for explicit data-parallel programming via {SIMD} vector types
  M. Kretz, Dissertation Goethe-University

NUMA in real life



Example Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz, 1333 MHz DDR3

Modern Programming



CBM

Stage	Description	Time/track	Speedup
	Initial scalar version	12 ms	–
1	Approximation of the magnetic field	240 μs	50
2	Optimization of the algorithm	7.2 μs	35
3	Vectorization	1.6 μs	4.5
4	Porting to SPE	1.1 μs	1.5
5	Parallelization on 16 SPEs (2 Cells)	0.1 μs	10
	Final SIMDized version	0.1 μs	120,000

Summarized stages of the porting procedure

Fast simdized Kalman filter based track fit
U. Kebschull, **I. Kisel**, V. Lindenstruth, W.F.J. Müller
Computer Physics Communication **2007**

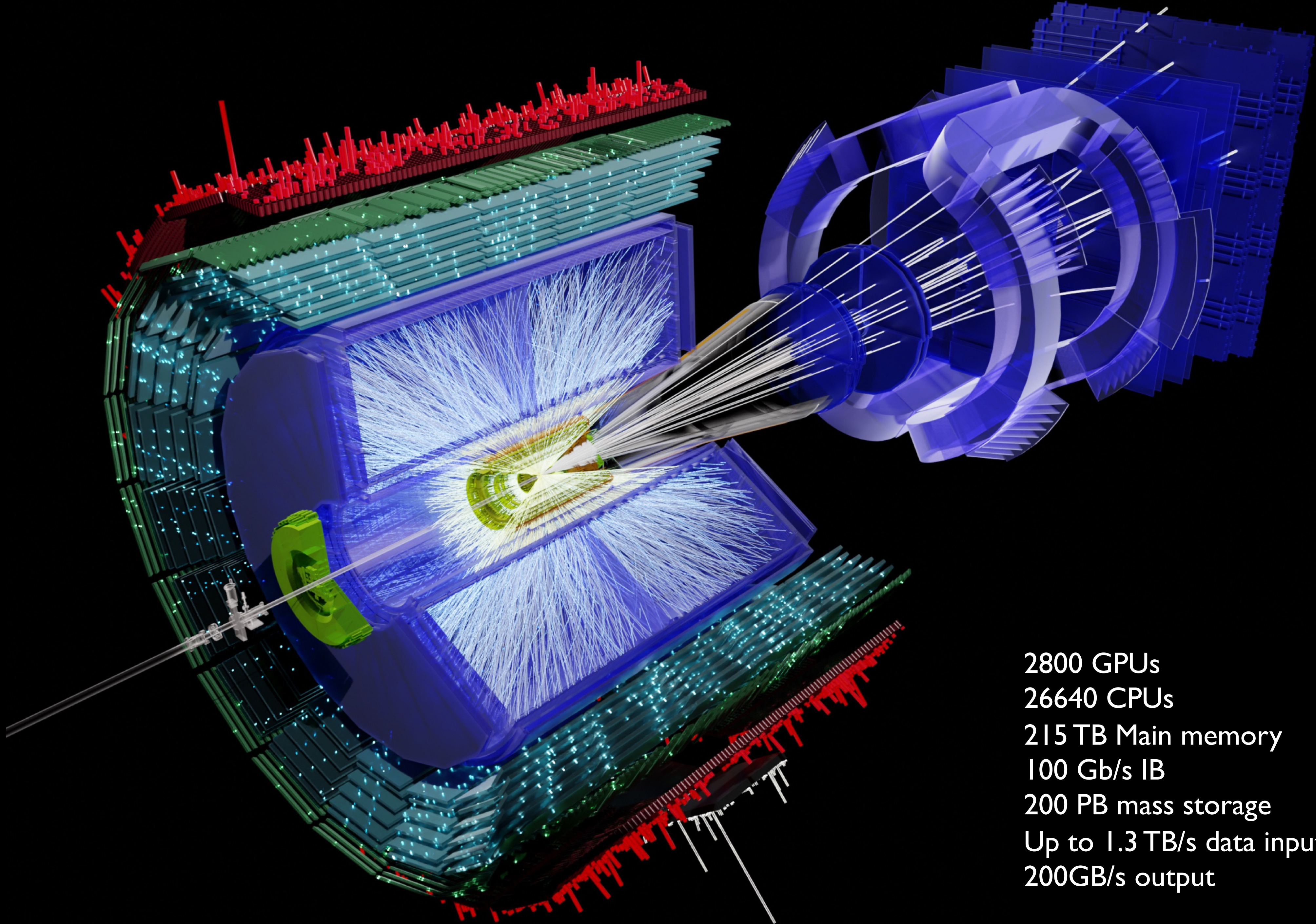
ALICE 2023

O2 system:

- 2800 AMD MI50/MI100 GPUs
- 26440 AMD CPUs (32Core/48Core)
- Real-time on-line processing of ALICE data (up to 1.3 TB/s)

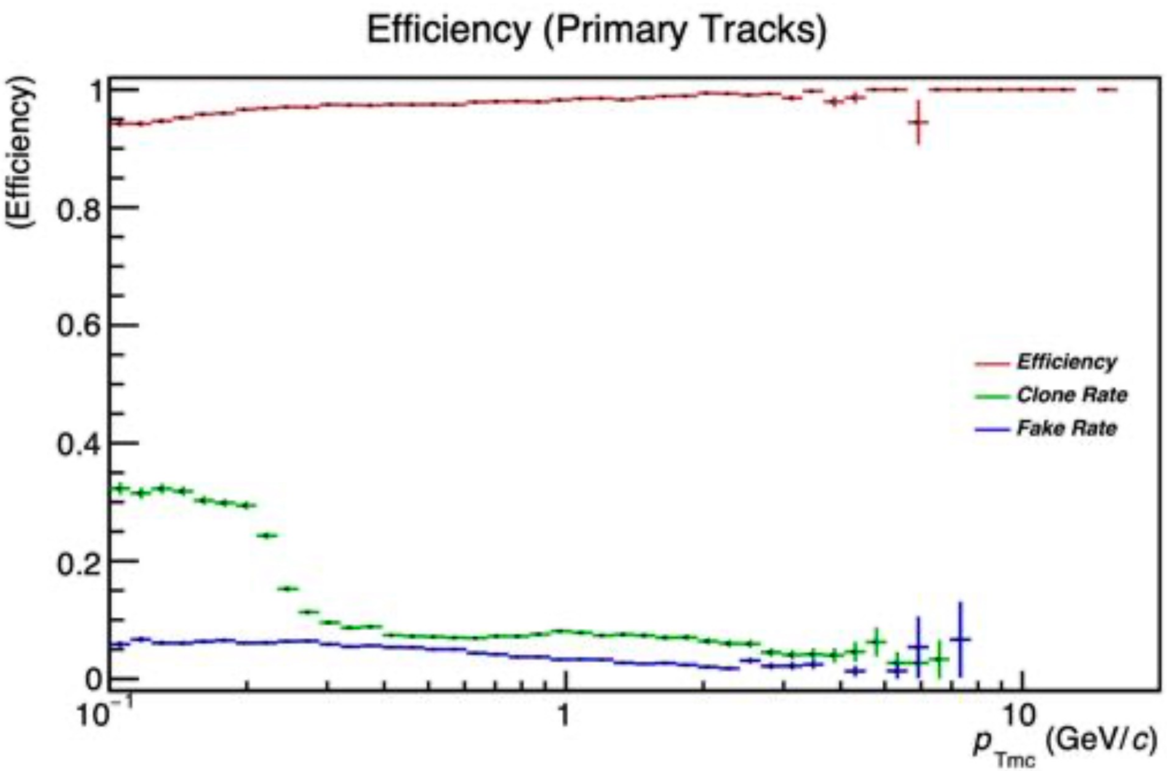
Hardware options for the parallel computation:

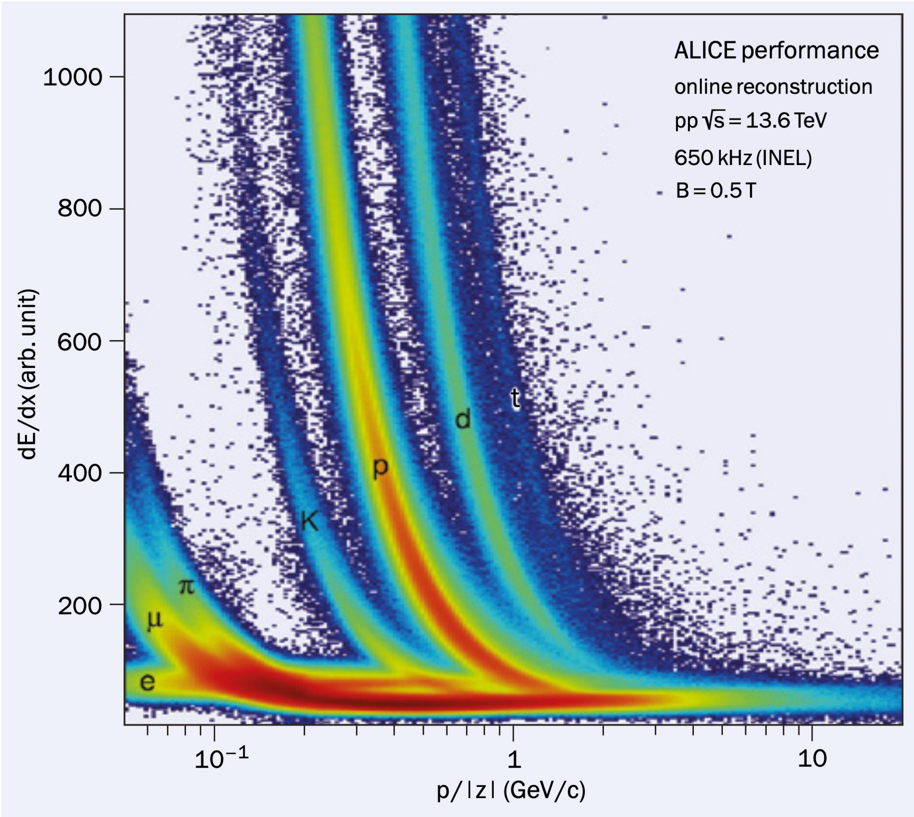
- SIMD CPU instructions (Vectorization)
- multi-threading
- multi-core CPU
- many-core hardware (Graphics cards, …)



2800 GPUs
26640 CPUs
215 TB Main memory
100 Gb/s IB
200 PB mass storage
Up to 1.3 TB/s data input
200GB/s output

ALICE on-line event reconstruction





- 2800 GPUs, 32 GB each
- 24640 CPU cores
- 200 TB main memory
- 200 PB mass storage
- Up to 1.3 TB/s ingres
- 200 GB/s egres
- 95% of processing on GPU

- Data distribution framework developed at FIAS

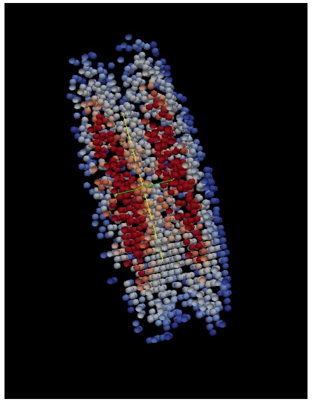
ALICE collab.

New nodes *The event processing node racks in the ALICE computing farm, part of a completely new computing model for Run 3 and beyond.*

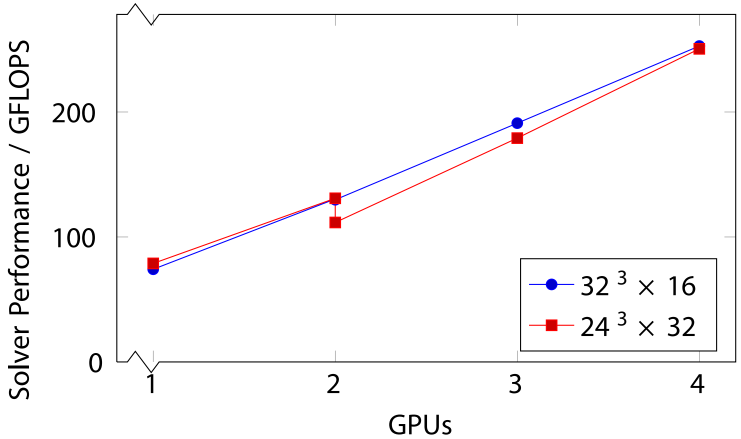
ALICE UPS ITS GAME FOR SUSTAINABLE COMPUTING

The design and deployment of a completely new computing model – the O^2 project – allows the ALICE collaboration to merge online and offline data processing into a single software framework to cope with the demands of Run 3 and beyond. Volker Lindenstruth goes behind the scenes.

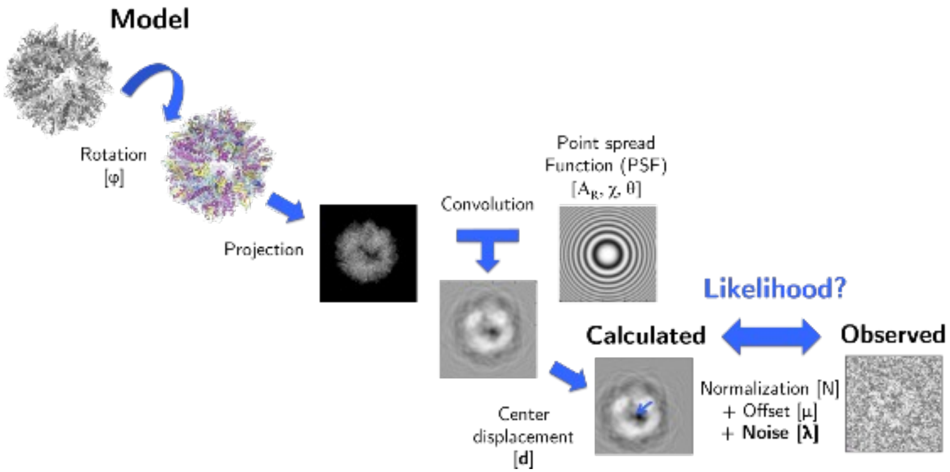
Algorithmic Optimization Examples



- Hybrid Quantum Molecular Dynamics: Speedup 160x, Memory 0.2x



- Lattice QCD: Speedup 44, support for multi GPU usage



- GPU-accelerated computing of Bayesian inference of electron microscopy images: Speedup >100

Algorithmic Optimizations

- The largest potential in Energy and CO_2 saving is algorithmic optimization
- Design algorithm for vectorization and SIMT
- Pay attention to data structure layout and cache efficiency (prefer structures of arrays over arrays of structures)
- Avoid load/store by recompute wherever reasonable
- Pay attention to NUMA domains
- Bring as much as possible computing to GPU

Summary

- 3D Green Cube Data Center
  - Data center architecture allows cost significant savings:
    - » CAPEX: 1.5 €/W for Tier-3 like data center
    - » OPEX: PUE < 1.1
    - » Very small foot print, Green Cube: >30 kW/m^2
  - No assumptions about computer hardware required
  - Indirect free cooling most cost effective
  - Unique and unprecedented cost (CAPEX/OPEX)
  - World wide potential savings 15 GW or 65 MT CO_2/a
- Computer Architecture another area of energy and CO_2 saving
  - Use vectorization of CPU code
  - Use of GPUs wherever possible
- Algorithmic optimization has huge saving potential
  - Demonstrated GPU cost/energy effectiveness >7x of CPU
  - GPU optimized code also performs better on CPU
  - Performance improvements often >100, demonstrated 10000 in particle physics


clustervision