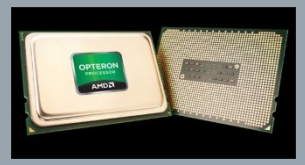
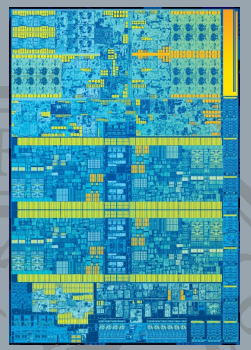
Programming Techniques for Supercomputers

Prof. Dr. G. Wellein[a,b] , Dr. G. Hager[a]
A. Afzal[a], D. Lacey[a], J. Laukemann[a], D. Ernst[a]

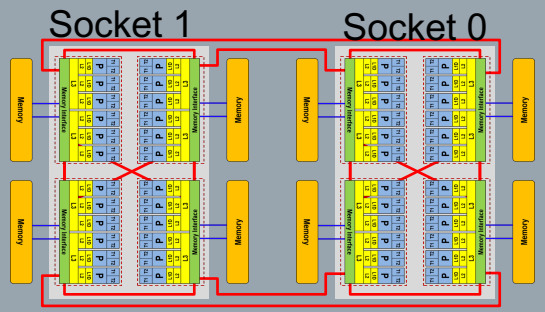
[a] Erlangen National Center for High Performance Computing (NHR@FAU)
[b] Department für Informatik

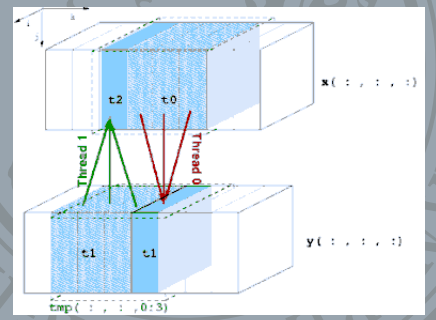


© Intel



FAU Erlangen-Nürnberg
Sommersemester 2023





Audience & Contact

- Audience
  - **Computational Engineering**
  - **Computer Science**
  - **Computational & Applied Mathematics**
  - Physics, Engineering, Materials Science, Chemistry,…

- Contact:
  - Gerhard Wellein: gerhard.wellein@fau.de (Lecture&Tutorial)
  - Georg Hager: georg.hager@fau.de (Lecture&Tutorial)

  - A. Afzal (ayesha.afzal@fau.de) and Dominik Ernst: Tutorial
  - Dane Lacey and Jan Laukemann: Tutorial (**CAM**)

- For questions first use moodle – forum (see next slide)

How to **register** for the **lecture** & **tutorials**

- Sign up for lecture & tutorials at moodle:

https://go-nhr.de/PTfS

- For all students doing the tutorial (to get access to our HPC systems): Please send an email (**using your FAU address**) to georg.hager@fau.de:
  - Subject: PTfS-registration
  - Content:
    - First name, last name
    - PTfS or PTfS-CAM

- Registration for exam (starts later): campo.fau.de

Format of course

- Lecture only: 5 ECTS
  - Material covered in the lecture
  - Register in “campo”
  - Written exam: 60 Minutes
- Lecture & Exercises: (5 + 2.5) ECTS
  - Material covered in lecture AND tutorial
  - Register for lecture AND exercise in “campo”
  - Written exam: 90 Minutes
  - Bonus system based on total credits of returned exercise sheets:
    - 60% - 79% an upgrade in the final mark of 1 stage, e.g. 2.7 → 2.3
    - 80% and higher upgrade in the final mark of 2 stage, e.g. 2.7 → 2.0
    - You must pass the exam before “boost” is applied! 1.0 is still best mark
- No “supporting material” allowed in exam
- **PTFS-CAM** students: Lecture & Excercises **+ Additional programming project**

Organization & Format

- 4 hrs of lecture (2 x 2hrs) / week : !!! In person !!!
  - Monday (16:15 – 17:45): 02.134-113
  - Tuesday (10:15 – 11:45): 02.134-113
  - Wednesday (10:15 – 11:45): 02.134-113 (backup slot – see moodle for lecture dates)

GW needs to travel several times during the semester → backup slot

- **During lectures: DON'T BE SHY AND ASK QUESTIONS!**

- 2 hrs of tutorial (for 7.5 ECTS and CAM students) / week
  - **Monday (14:15 – 15:45) 02.134-113**

  OR

  - **Thursday (14:14 – 15:45) 02.134-113**

Organization & Format: Lecture

- Lecture/Tutorial is completely documented in moodle LMS:

  https://go-nhr.de/PTfS (see also StudOn entry)

- Please enroll into the lecture and specify your matriculation number!
  - Homework assignments, announcements etc. all handled via moodle
  - Use forum for questions or discussion with other students

- Recordings of 2020 and 2021 PTfS lectures are available at https://www.fau.tv
  - https://www.fau.tv/course/id/2351 (2021 - Zoom lecture)
  - https://www.fau.tv/course/id/1233 (2020 – CAMTASIA recording)

Organization & Format: Tutorials and Homework

- New homework assignments are released every Thursday around 10:00 a.m.
- Report submission
  - Deadline: one week later, i.e., Thursday 10:00 a.m. No extensions!
  - Deadline for Assignment 0: Thursday, to be announced
  - Upload a single file in Moodle
    - PDF report or
    - compressed archive including a PDF report and supporting material
  - Grading will be done based on PDF report
- Problems? Questions?
  - Ask during tutorial Q&A
  - Use Moodle forum

Monday / Thursday Tutorial Sessions: Content

- Presentation of solution to previous assignment
- Presentation of current (new) assignment
- !!! Opportunity to ask questions !!!

Organization & Format: PTfS-CAM

- CAM students have to do a programming and performance modeling project:
- Shared memory parallel Preconditioned Conjugate Gradient (PCG) solver
- Extra tutorials:
  - Basic shell introduction: April 22, **after** the normal tutorial
  - Introduction to C: April 23, **instead** of lecture
  - Introduction to project work (OpenMP parallelization)
  - Introduction to project work (performance modelling)
  - Consultation hour every week (Jan / Dane)
  - Dates & times will be announced in moodle → SEE MOODLE

Format of the course

- Prerequisite (for exercises):
  - Basic programming knowledge in C/C++ or FORTRAN
  - Using LINUX / UNIX OS environments (including ssh)

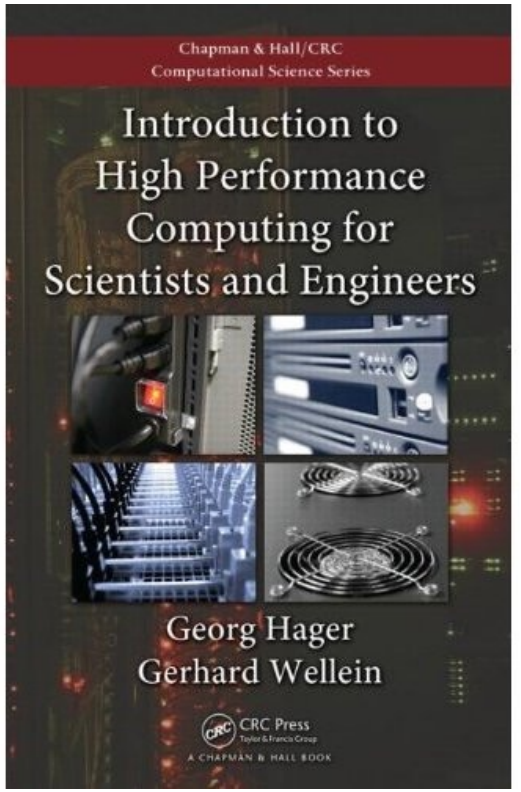
- Recommended
  - First experiences with parallel programming – though we will introduce necessary basics

- Tutorials:
  - You **must** submit your homework assignments for inspection to qualify for bonus!
  - Do the exams and programming exercises even if you do not submit!
  - Some topics will be covered (in more detail) in the tutorials
  - Topics additionally covered in the tutorials are part of the 7.5 ECTS exam
  - Practical parallelization skills will be tested in the 7.5 ECTS exam!

Supporting material

- Books:
  - G. Hager and G. Wellein:
    Introduction to High Performance Computing for Scientists and Engineers.
    CRC Computational Science Series, 2010.
    ISBN 978-1439811924

    - 10 copies are available in the library
    - discounted copies – ask us



  - J. Hennessy and D. Patterson: *Computer Architecture. A Quantitative Approach.* Morgan Kaufmann Publishers, 2017. ISBN 978-0128119051
  - W. Schönauer: *Scientific Supercomputing*.
    (cf. http://www.rz.uni-karlsruhe.de/~rx03/book/)

Supporting material

- Documentation:
  - http://www.openmp.org
  - http://www.mpi-forum.org
  - http://developer.intel.com/products/processor/manuals
  - http://developer.amd.com/documentation/guides

- The big ones and more useful HPC related information:
  - http://www.top500.org

Related teaching activities

- Regular seminar on

  “Efficient numerical simulation on multicore processors” (**MuCoSim**)
  - 5 ECTS
  - 2 hrs per week
  - 2 talks + written summary

  - Topics from code optimization, code parallelization and code benchmarking on latest multicore / manycore CPUs and GPUs

  - This semester: Wednesday 16:00 – 17:30

  - See moodle: http://go-nhr.de/MuCoSim

  - Also offered during winter term

Related teaching activities

- Lecture on

  Practical parallel algorithms with MPI (**PAMPI**)

- Winter term
- 5 ECTS
- 2 hrs of lecture
- 2 hrs of tutorials

- Lecturer: Dr. Jan Eitzinger

Scope of the lecture

Scope of the lecture

Understand relevant hardware features/concepts of modern HPC systems and derive efficient parallelization & implementation strategies

→Identify **basic hardware concepts** (CPU & GPU)
single core, chip/device, node-level

→Learn strategies how to **efficiently use (program) the hardware**
(code transformations/optimizations, OpenMP/CUDA)

→Establish appropriate **performance expectations/models** to

→assess the attainable performance and

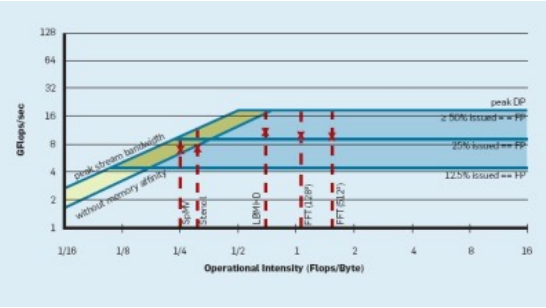
→identify the “hardware bottleneck”

Scope of the lecture

- Hardware coverage:
  - Single-core + Multi-Core — Intel Core/Xeon & AMD Rome & ARM-based
  - Many-core / GPU: — NVIDIA V100/A100
  - Shared memory nodes — (multiple multi-/many-core & GPUs)
  - Distributed memory computers — (multiple/many nodes)

- Programming models (mostly basic introduction; performance issues):
  - OpenMP — (shared memory nodes)
  - CUDA — (GPUs)
  - MPI — (distributed memory) → **PAMPI lecture**

- Performance Modelling
  - Expectations
  - Roofline Model [Williams&Patterson, 2009]
  - (ECM Model)



Structure of the lecture

- Introduction
  - Performance: Basics, Measuring & Reporting, Benchmarks
- Modern (multicore) processors
  - Single core: Basics, Pipelining, Superscalarity, SIMD
  - Memory Hierarchy: Caches & Main Memory
  - Multicore: Technology & Basics
  - GPU (*)
- Parallel computers: Shared Memory
  - Shared-memory system architectures: UMA, ccNUMA
  - OpenMP basics
- Performance Modelling / Engineering:
  - Roofline Model
  - Case Studies: Dense&Sparse Matrix-Vector-Multiplication / Stencils / …
- Shared Memory in depth
  - Advanced OpenMP, Pitfalls, Data Placement
- Hardware performance monitoring and model validation (*) → LIKWID

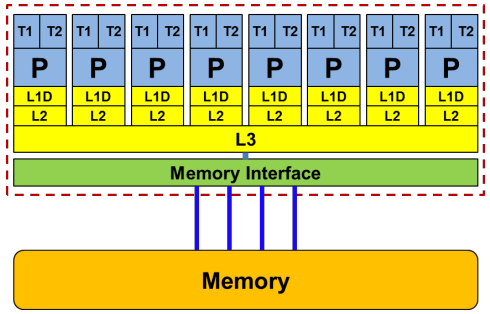
Performance Analysis and Modeling

Scope of the lecture – a typical example

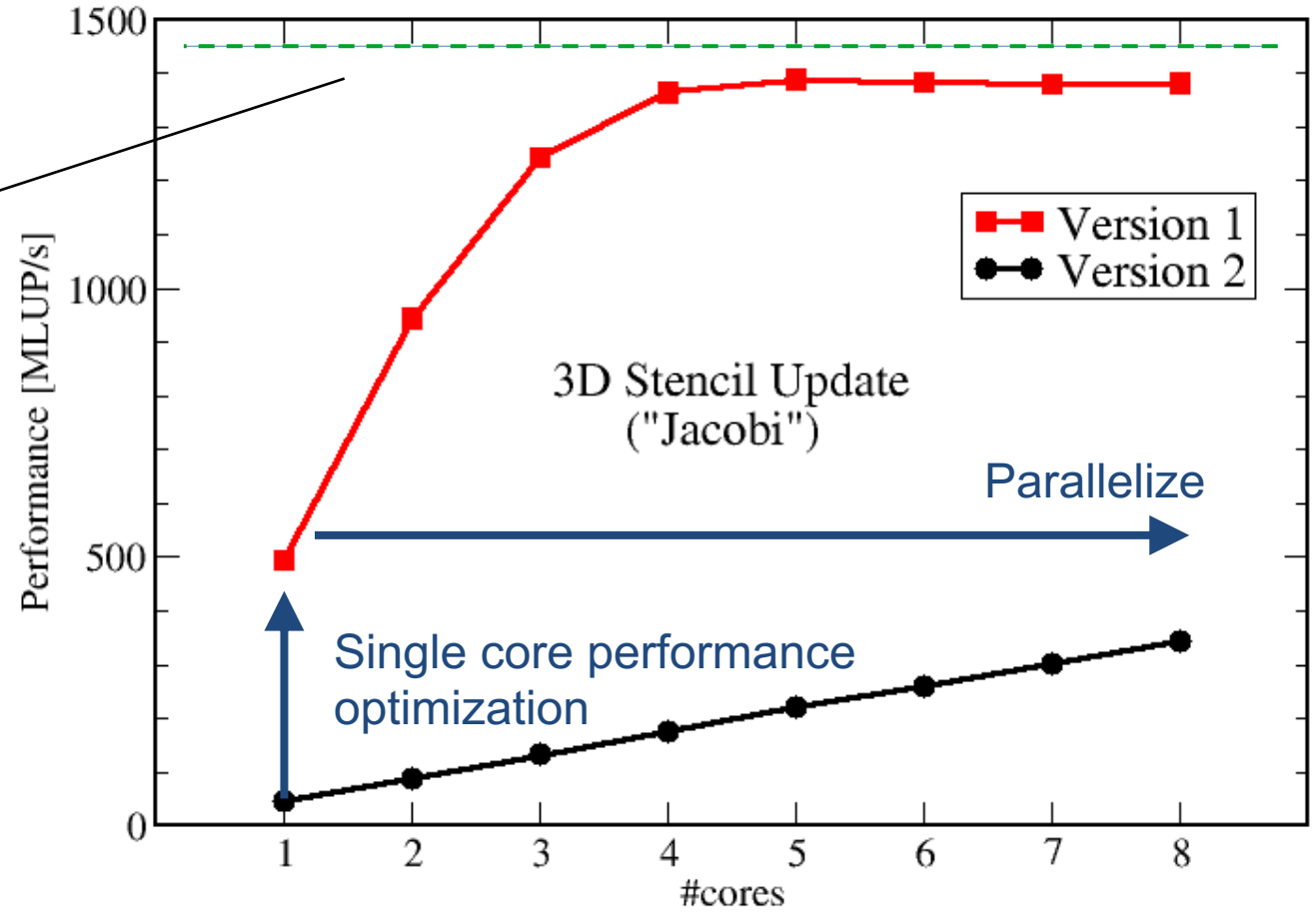
```
!$OMP PARALLEL DO
do k = 1 , 400
  do j = 1 , 400; do i = 1 , 400
     y(i,j,k)= b*(  x(i-1,j,k)+ x(i+1,j,k)+ x(i,j-1,k)+
                    x(i,j+1,k)+ x(i,j,k-1)+ x(i,j,k+1))
    enddo; enddo
enddo
!$OMP END PARALLEL DO
```

Parallelize

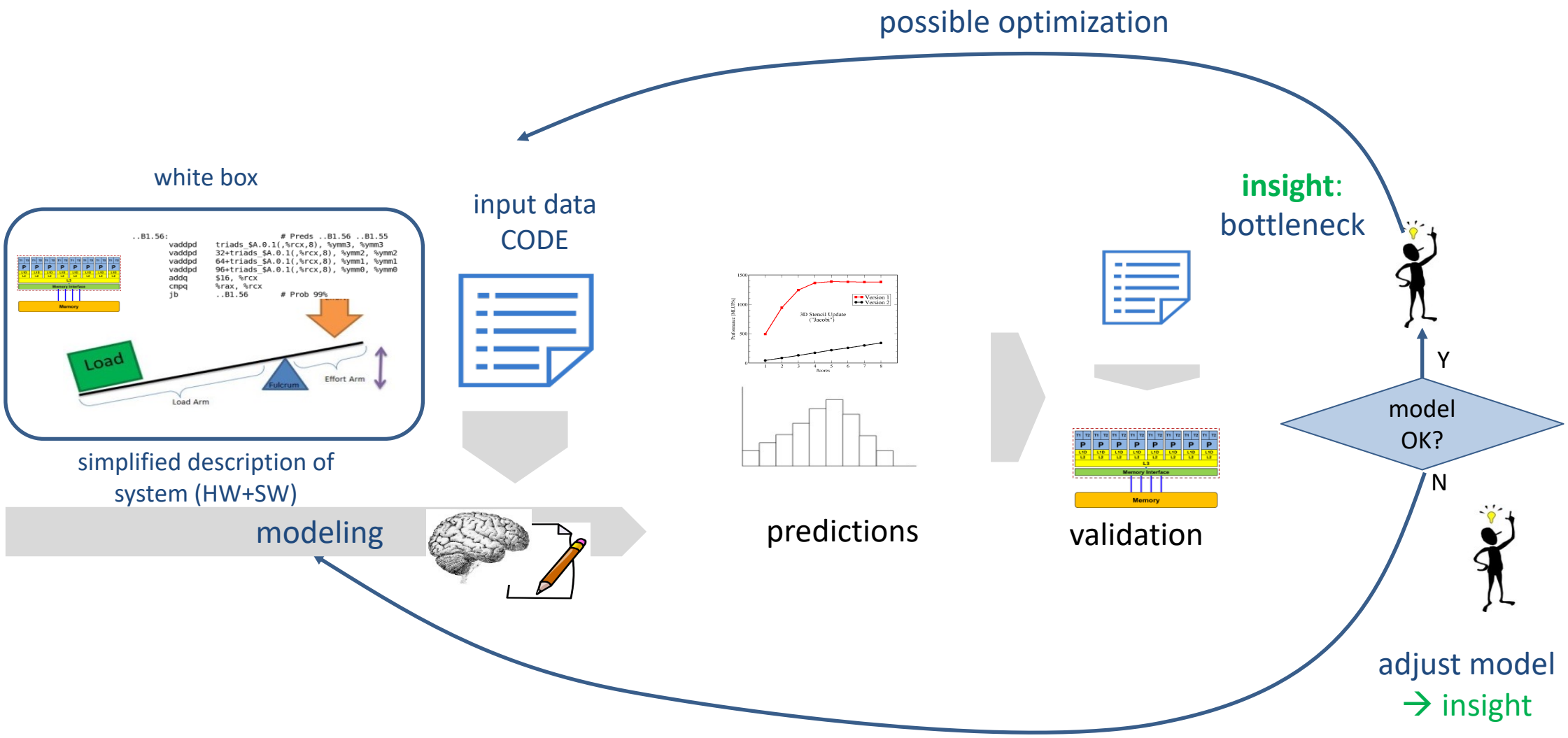
Establish limit simple performance model (Here: Roofline)



Intel® Xeon® Prozessor E5-2670



Code optimization/parallelization – no black boxes!



„Performance Engineering“

NHR FAU

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Introduction

Supercomputers:
The Big Ones and the working horses

Most powerful computers in the world: TOP500

- Top 500: Survey of the 500 most powerful supercomputers
  - http://www.top500.org

  - How to rate the performance?

  → Solve large dense system of equations: **`A x = b`** („LINPACK“)

  - Max. performance achieved with 64-Bit Floating Point Numbers: R_{max}

  - Published twice a year (ISC in Germany, SC in USA)

  - First: 1993 (CM5/1,024 procs.): 60 GFlop/s (TOP1)
  - Nov. 2023 (HPE – AMD GPUs): 1,194,000,000 GFlop/s (TOP1)

  - Performance increase: 75% p.a. from 1993 – 2023

Most powerful computers in the world: TOP500

- Performance measures: MFlop/s, GFlop/s, TFlop/s, PFlop/s, EFlop/s
  - Number of FLOATING POINT operations per second
  - FLOATING POINT operations: Typically Add & Multiply operations (→Chapter 3 of the lecture)
  - Performance may depend on accuracy (of input operands):
    - double precision, double: 64 Bit
    - single precision, float: 32 Bit
    - half precision: 16 Bit
    - default: 64 Bit
    - See Chapter 3 for details

10^6: MFlop/s
10^9: GFlop/s

10^{12}: **T**Flop/s ← Single node
10^{15}: **P**Flop/s ← TOP500 systems
10^{18}: **E**Flop/s ← TOP1 in 2023

TOP3 as of November 2022

Number of processors

LINPACK Performance

Peak Performance

Power@ LINPACK

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016

Investment for such systems: 150 – 200 M€

Power bill @ 30ct/kWhr: 1 MW ←→ € 2,500,000 p.a.

Source: www.top500.org

TOP10 as of November 2022

Most systems use accelerators, e.g. GPUs: 1,3,4,5,6,8,9

Leading systems: US / Japan / China

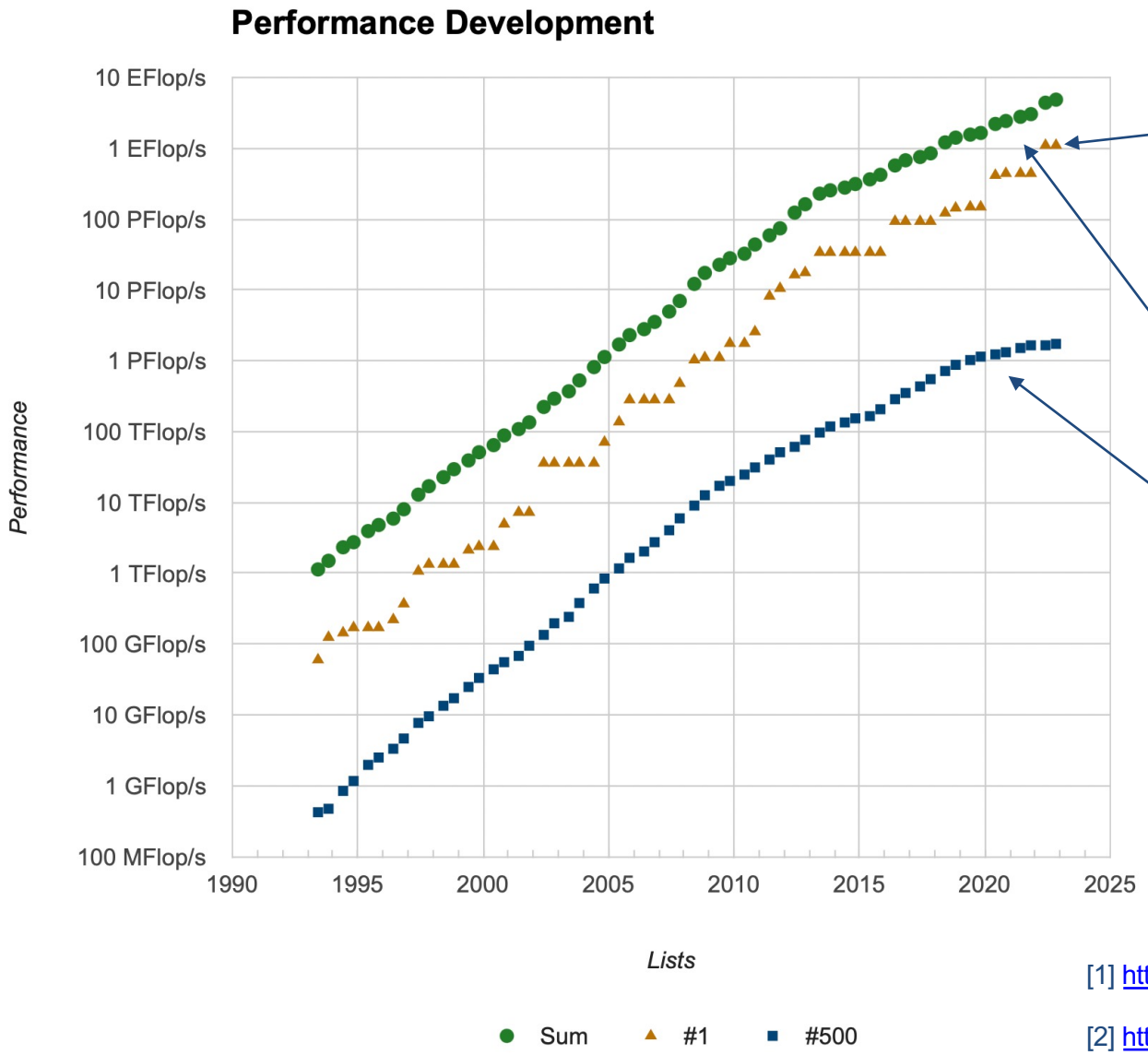
Two systems from Europe in Top 10: 3,4

Germany:
JUWELS Booster Module@FZJ Jülich: #12
SuperMUC-NG@LRZ Garching : #29

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,463,616	174.70	255.75	5,610
5	**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	2,414,592	148.60	200.79	10,096
6	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438
7	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93.01	125.44	15,371
8	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70.87	93.75	2,589
9	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63.46	79.22	2,646
10	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, NUDT National Super Computer Center in Guangzhou China	4,981,760	61.44	100.68	18,482

Source: www.top500.org

Performance Trend & Projection (2022)



Source: www.top500.org

Frontier supercomputer at Oak Ridge Nat. Lab breaks exascale barrier ExaFlop/s in 2022 [1]

600 million USD investment [2]

Basic trend: Slope changes
→ performance increase slows down

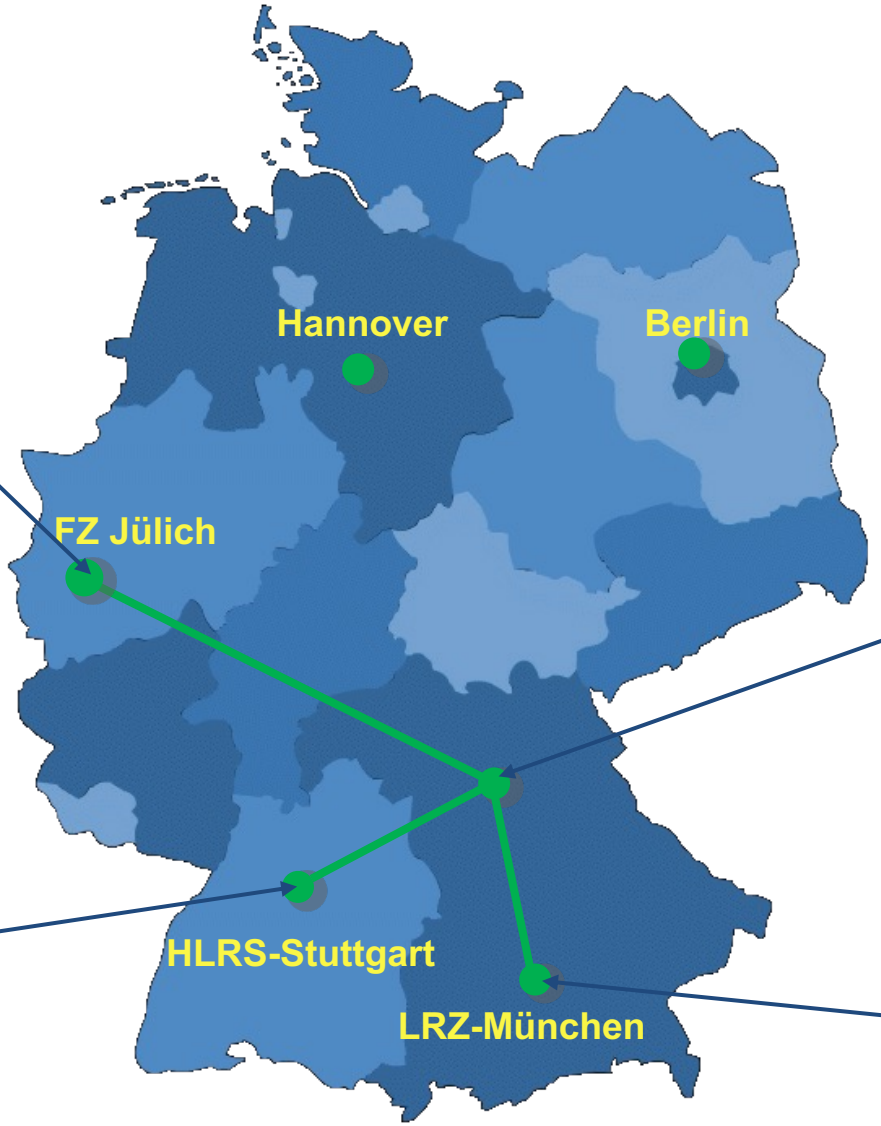
[1] https://www.ornl.gov/news/frontier-supercomputer-debuts-worlds-fastest-breaking-exascale-barrier

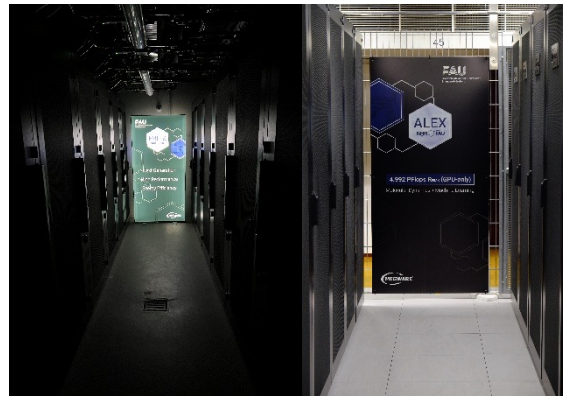
[2] https://www.olcf.ornl.gov/2019/05/07/no-scaling-back-doe-cray-amd-to-bring-exascale-to-ornl/

HPC Centers in Germany: A view from Erlangen

Forschungszentrum Jülich
in der Helmholtz-Gemeinschaft

Jülich Supercomputing Center
JUWELS (Booster)
71 PF/s





NHR FAU

FAU Erlangen/-Nürnberg
Fritz (5 PF/s) & Alex (5 PF/s)

H L R S

HLR Stuttgart
Hawk: 26 PF/s

SuperMUC-NG: 26.8 PF/s

Erlangen National Center for High Performance Computing

www.nhr.fau.de

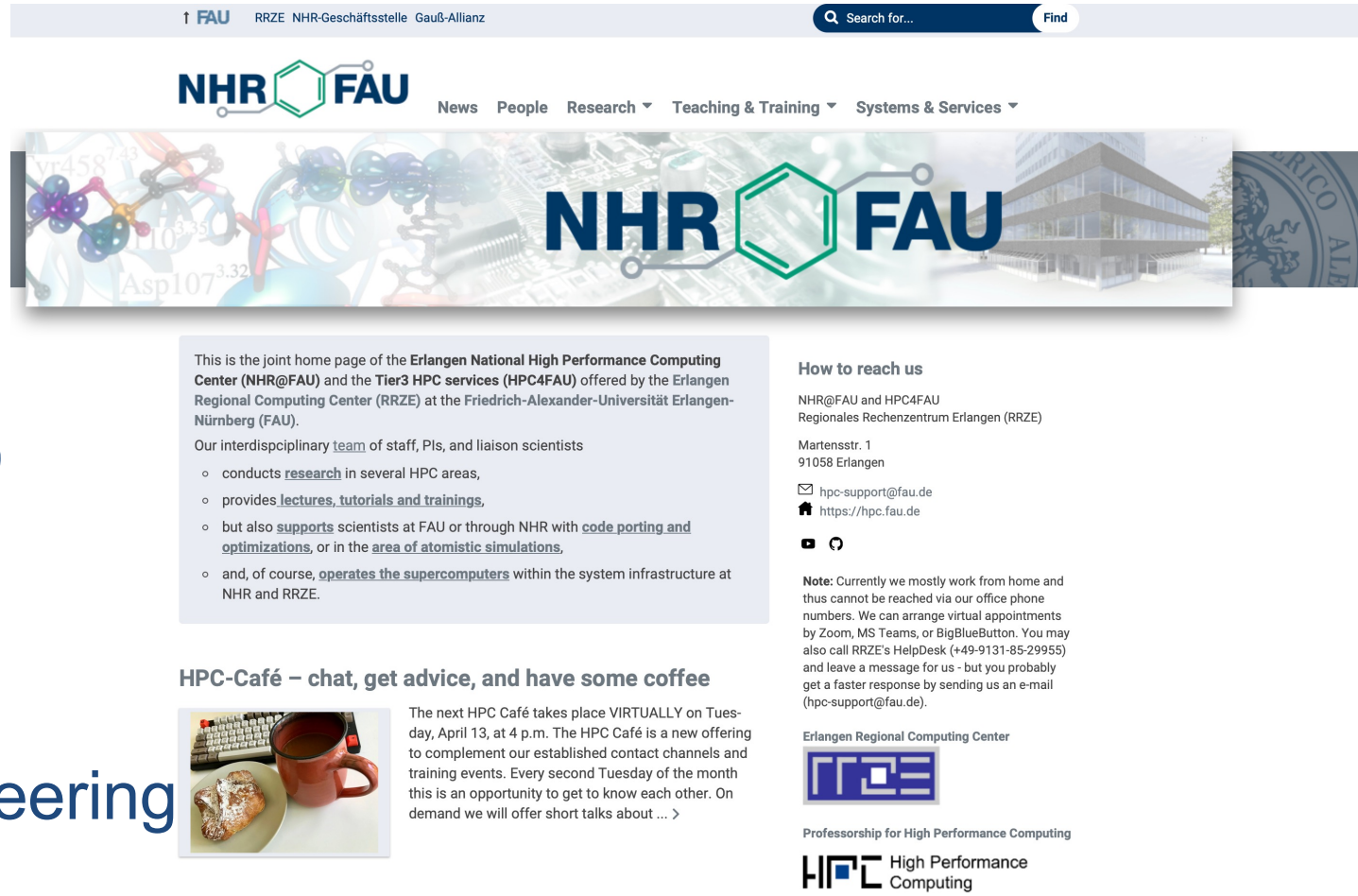
One out of nine national HPC centers at German universities

Ongoing:

- New system installed (10 M€+)
- New HPC center - 2030

Focus topics:

- Node-level Performance engineering
- Atomistics simulations



Student positions, master theses, doctoral positions available

Fritz & Alex: Fact Sheet

	#nodes	Node conf.	Storage	Typical job sizes	Peak (FP64)
Fritz	992 Intel ICL (71,424 cores) **(+120,000 c in 2025)**	2 * 36 c (8360Y) 256 GB 1 x HDR100	Shared PFS • 3 PB • >20 GB/s	1 – 64 nodes	5.9 PF/s (4.1 PF/s)
	64 Intel SPR (6,656 cores)	2 * 52 c (8470) **1 TB / 2 TB** 1 x HDR100		1 – 4 nodes	n.y.a.

178	**Fritz** - Megware D50TNP, Xeon Platinum 8360Y 36C 2.4GHz, InfiniBand HDR100, MEGWARE Universitaet Erlangen - Regionales Rechenzentrum Erlangen Germany	71,424	3.58	5.45 672

Power consumption (kW) for LINPACK

Fritz: https://doc.nhr.fau.de/clusters/fritz/

Fritz & Alex: Fact Sheet

NHR FAU

	#nodes	Node conf.	Storage	Typical job sizes	Peak (FP64)
Alex	38 AI/ML (**304 NVIDIA A100**) (+320 H100 in 2024)	8 * NVIDIA A100 2 * 64 c (AMD) 1 TB 2 x HDR200	Node local 14 TB NVMe	1 – 8 GPUs	6.1 PF/s
	44 MD (**352 NVIDIA A40**)	8 * NVIDIA A40 2 * 64 c (AMD) 0.5 TB	Node local 7 TB NVMe	1 – 8 GPUs	---

157	**Alex** - MEGWARE NF5488A5, AMD EPYC 7713 64C 2GHz, NVIDIA A100 SXM4 80 GB, Infiniband HDR, MEGWARE Universitaet Erlangen - Regionales Rechenzentrum Erlangen Germany	37,696	4.03	6.08	179

Power consumption (KW) for LINPACK (A100)

Alex: https://doc.nhr.fau.de/clusters/alex/

HPC compute infrastructure – TOP500

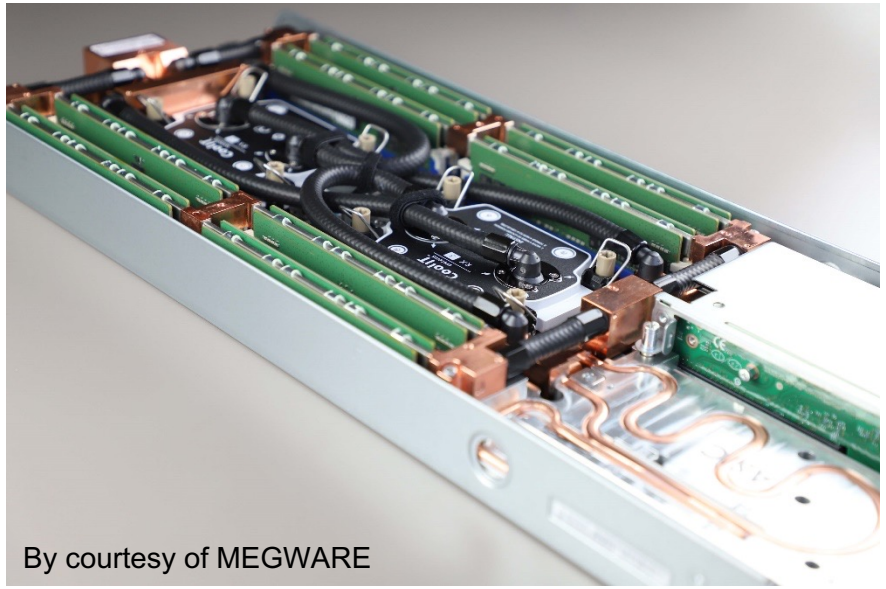
Alex	List	Fritz
#184 2.9 PF/s	06/22	**#323** 2.2 PF/s
#174 3.2 PF/s	11/22	**#151** 3.6 PF/s
#157 4.0 PF/s	06/23	**#178** 3.6 PF/s
#187 4.0 PF/s	11/23	**#213** 3.6 PF/s

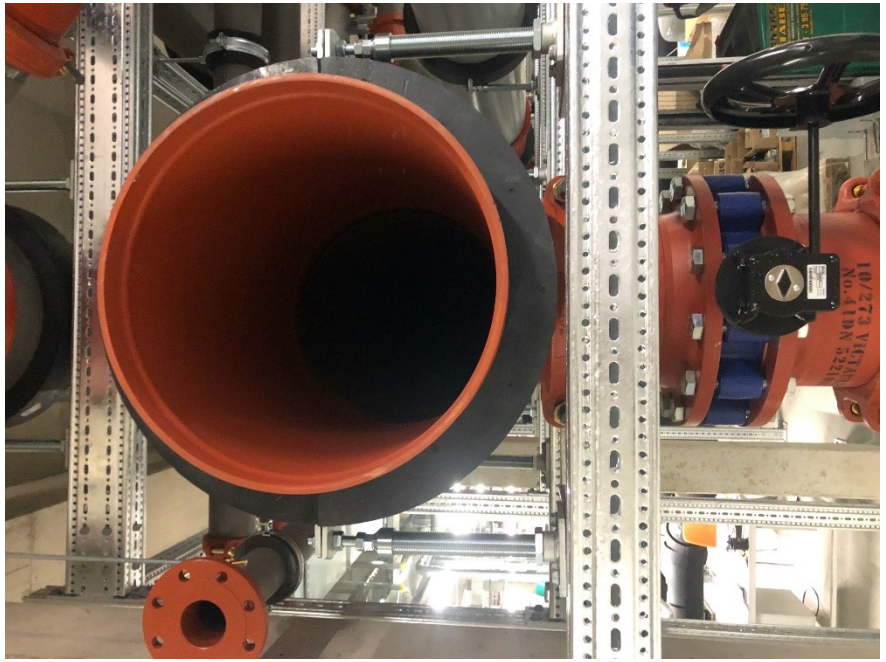
RRZE-Serverraum

NatFak Kältezentrale

NF-Technikzentrale: “data center” for Fritz (750 KW)

By courtesy of MEGWARE


By courtesy of MEGWARE



Lecture plan until May

- **15.4.2024: Lecture**
- **16.4.2024: Lecture**
- **17.4.2024: Lecture**

- **22.4.2024: Lecture**
- 23.4.2024: No Lecture, instead: **Introduction to C**
- 24.4.2024: No Lecture

- **29.4.2024: Lecture**
- **30.4.2024: Lecture**
- 01.5.2024: PUBLIC HOLIDAY