Programming Techniques for Supercomputers

Prof. Dr. G. Wellein[a,b]
Dr. G. Hager[a]
A. Afzal[b], C.L. Alappat[b], D. Ernst[b]

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

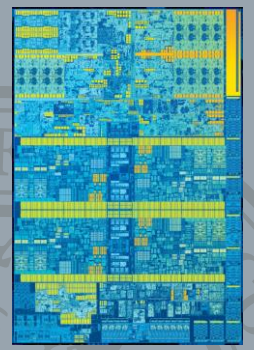
FAU Erlangen-Nürnberg
Sommersemester 2022

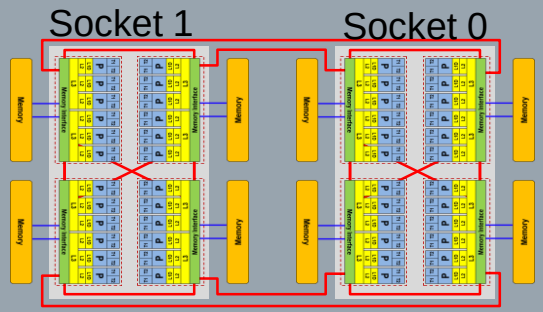




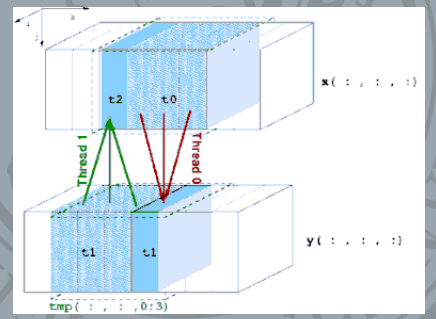
© Intel











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 (dominik.ernst@fau.de): Tutorial
  - Christie Louis Alappat: **christie.alappat@fau.de** (Tutorial & **CAM**)

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

Format of course

- Lecture only: 5 ECTS
  - Material covered in the lecture
  - Register in “meincampus”
  - Written exam: 60 Minutes
- Lecture & Exercises: (5 + 2.5) ECTS
  - Material covered in lecture AND tutorial
  - Register for lecture AND exercise in “meincampus”
  - 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: Please contact me / **Christie** via email

Organization & Format

- 4 hrs of lecture (2 x 2hrs) / week : !!! In person !!!
  - Monday (16:15 – 17:45): H10
  - Thursday (8:30 – 10:00): H6
  - (Tuesday: backup slot – if some lectures need to be moved due to business travel)
- **During lectures: DON'T BE SHY AND ASK QUESTIONS!**

- 2 hrs of tutorial (for 7.5 ECTS and CAM students) / week
  - Tuesday (8:15 – 9:45) 02.133-113

OR

  - Wednesday (10:15 – 11:45) 02.134-113

!You need to wear a mask if minimum distance of 1,5 meter cannot be maintained!

Organization & Format: Lecture

- Lecture/Tutorial is completely documented in moodle LMS:

  http://tiny.cc/PTfS (see also PTfS univis 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 Friday around 12:00 p.m.
- Report submission
  - Deadline: one week later, i.e., Friday 12:00 p.m. No extensions!
  - Deadline for Assignment 0: Friday, May 6 at 12:00 p.m.
  - 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

Tuesday/Wednesday Tutorial Sessions

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

Organization & Format: PTfS-CAM

- In addition to lecture & exercise, CAM students have to do a programming and performance modeling project (Conjugate Gradient solver)
- PTfS-CAM students have to register via e-mail
- If you did not, please send an e-mail to gerhard.wellein@fau.de and christie.alappat@fau.de
- All registered students will receive an e-mail next week regarding the extra tutorials

Format of the course

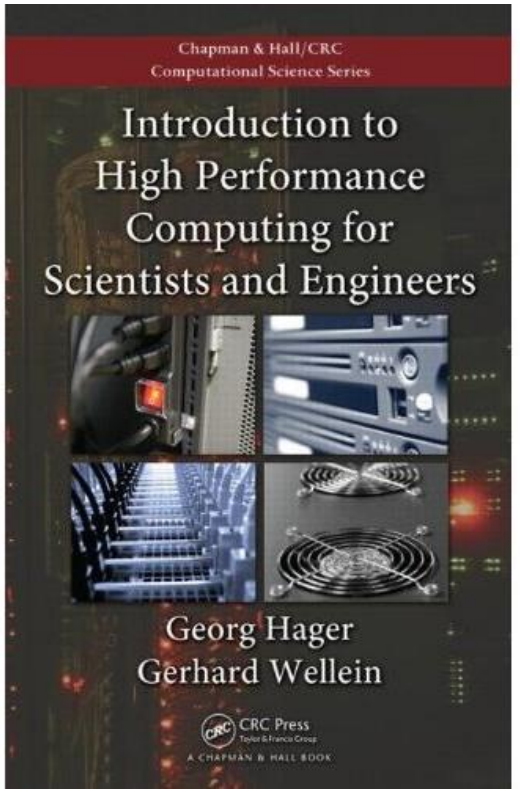
- Prerequisite (for exercises):
  - Basic programming knowledge in C/C++ or FORTRAN
  - Using LINUX / UNIX OS environments (including ssh)

- 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: Monday 16:00 – 17:30 – RRZE (zoom)

  - See moodle: http://tiny.cc/MuCoSim

  - Also offered during winter term

Scope of the lecture



Scope of the lecture

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

→Identify basic hardware concepts
(single core, chip/device, node-level, massively parallel)

→Learn strategies how to efficiently use (program) the hardware
(code modifications, OpenMP/CUDA, MPI)

→Establish appropriate performance expectations/models to

→assess the attainable performance and

→identify the “hardware bottleneck”

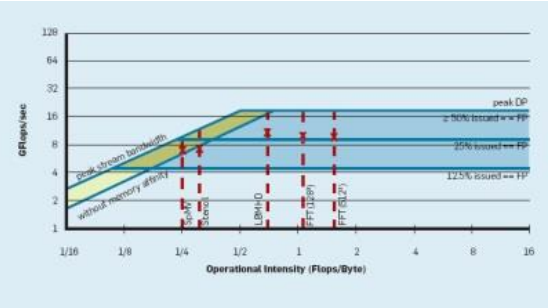
Scope of the lecture

- 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)

Performance Models – how fast may a code run on a given architecture?

- 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
  - Manycore / 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
- Parallel computers: Distributed Memory
  - Architecture & Communication networks
  - MPI in a nutshell
- Hardware performance monitoring and model validation (*)

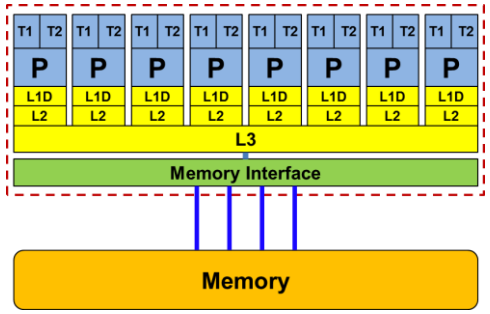
Performance Analysis and Modeling

Scope of the lecture – a typical example

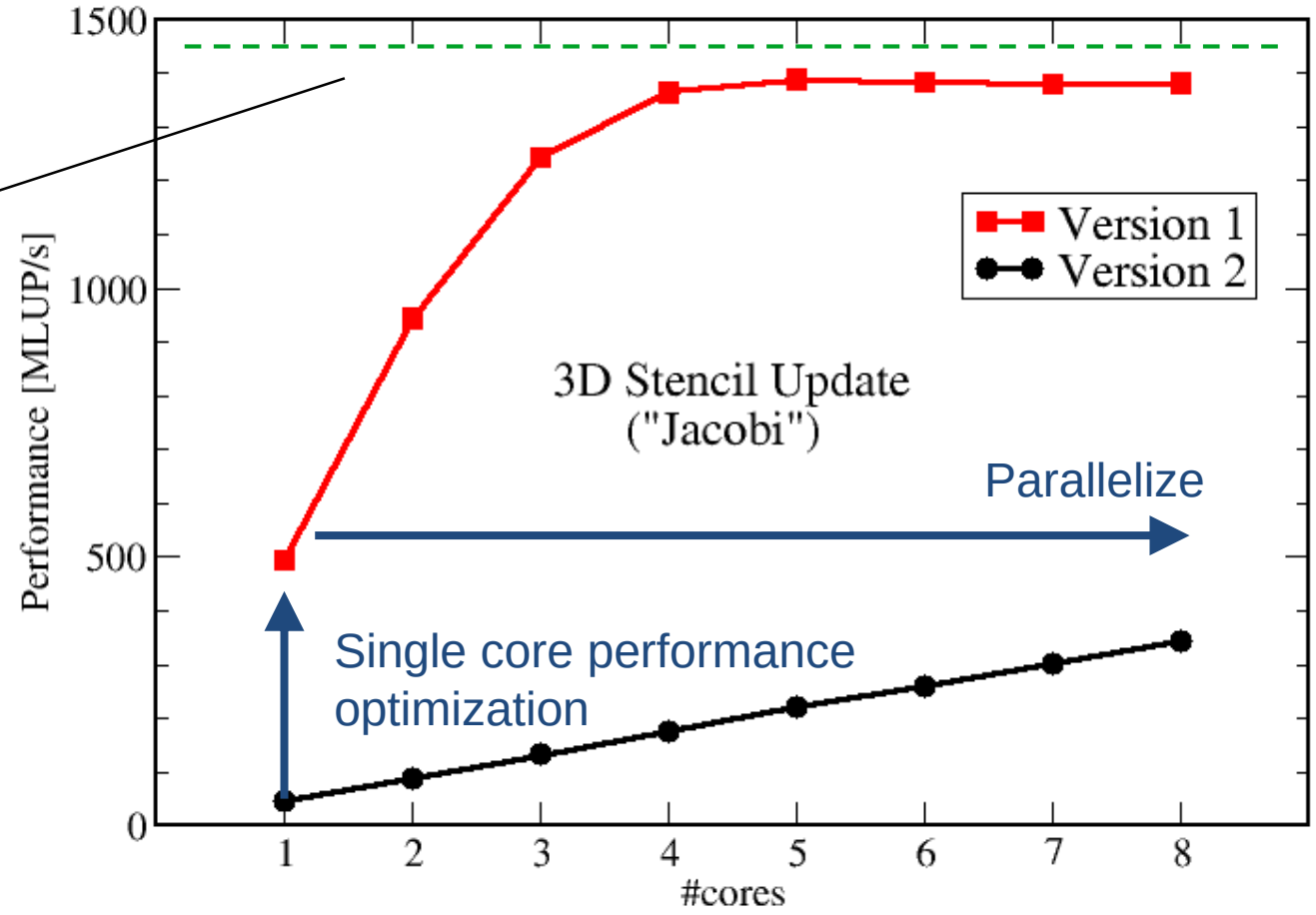
```
!$OMP PARALLEL DO
do k = 1 , 400
  do j = 1 , 400; do i = 1 , 400
     y(i,j,k)= b*(  x(i-1,j,k)+ x(i+1,j,k)+ x(i,j-1,k)+
                    x(i,j+1,k)+ x(i,j,k-1)+ x(i,j,k+1))
    enddo; enddo
enddo
!$OMP END PARALLEL DO
```

Parallelize

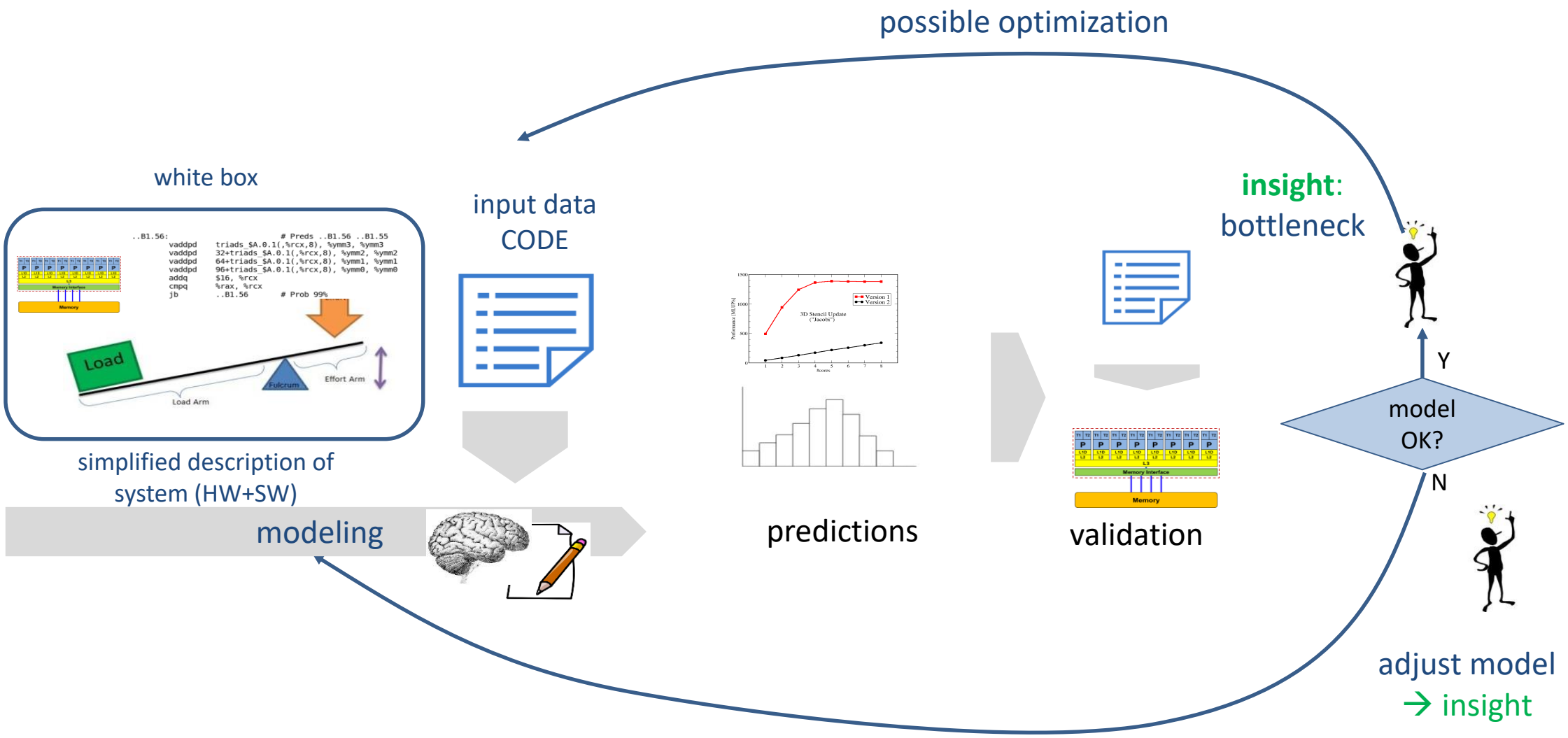
Establish limit simple performance model (Here: Roofline)



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. 2020 (Fujitsu A64FX / $7.6*10^6$ cores): 442,010,000 GFlop/s (TOP1)

  - Performance increase: 80% p.a. from 1993 – 2020

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 2021/2 (?)

TOP3 as of November 2020 and 2021

Number of processors

LINPACK Performance

Peak Performance

Power@ LINPACK

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**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	148,600.0	200,794.9	10,096
3	**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	94,640.0	125,712.0	7,438

Investment for such systems: 150 – 200 M€

Power bill @ 15ct/kWhr: 1 MW ←→ € 1,300,000 p.a.

Source: www.top500.org

TOP10 as of November 2020*

Most systems use accelerators, e.g. GPUs: 2,3,5,6,7,10

Leading systems: Japan / US / China

Two systems from Europe (2020): 7,8

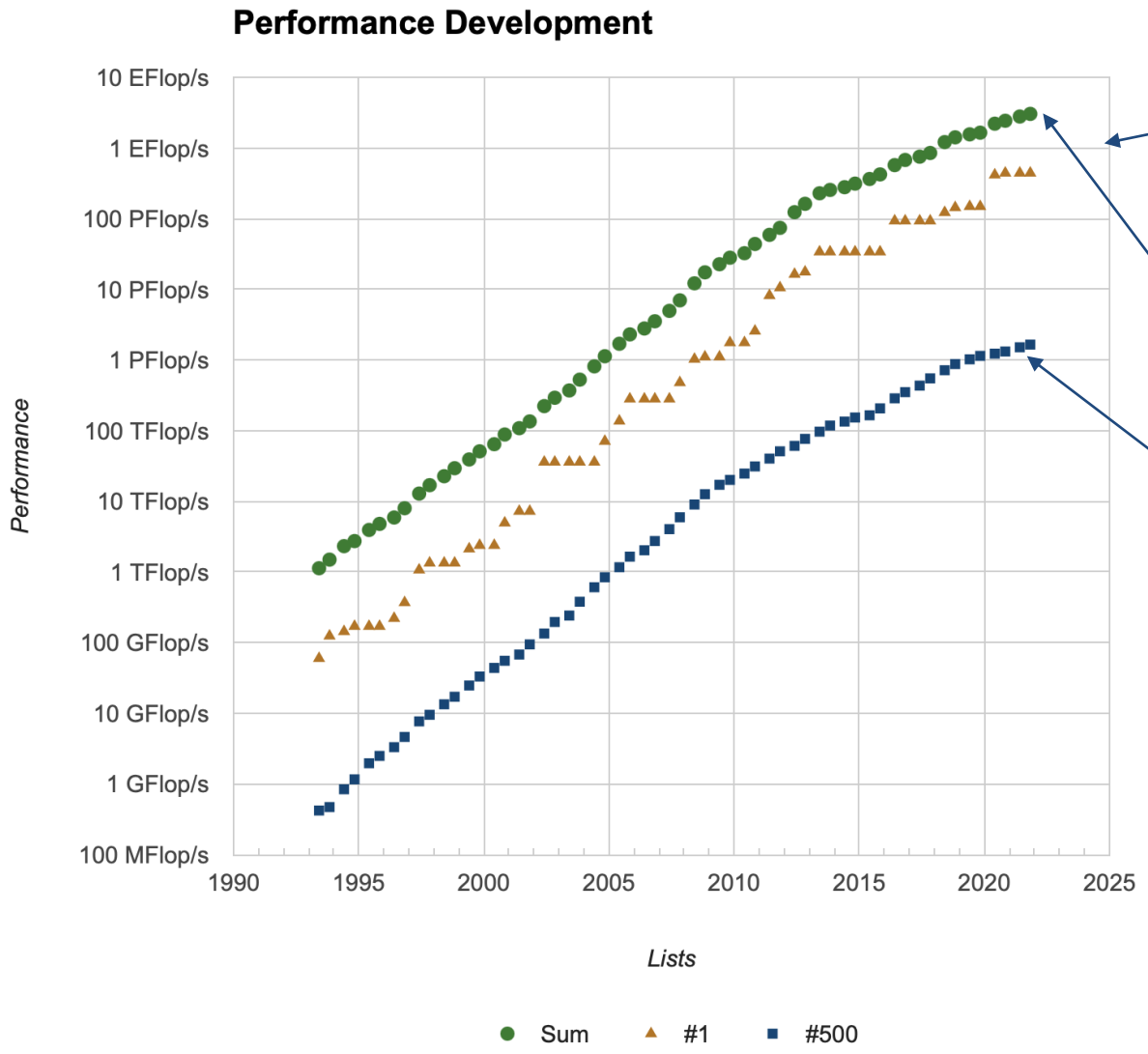
SuperMUC-NG@LRZ Garching (2020/21): #15/23

* Only 2 new TOP10 entries in 2021….

Source: www.top500.org

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**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	148,600.0	200,794.9	10,096
3	**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	94,640.0	125,712.0	7,438
4	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
5	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63,460.0	79,215.0	2,646
6	**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	61,444.5	100,678.7	18,482
7	**JUWELS Booster Module** - Bull Sequana XH2000 , AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite, Atos Forschungszentrum Juelich (FZJ) Germany	449,280	44,120.0	70,980.0	1,764
8	**HPC5** - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, Dell EMC Eni S.p.A. Italy	669,760	35,450.0	51,720.8	2,252
9	**Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR, Dell EMC Texas Advanced Computing Center/Univ. of Texas United States	448,448	23,516.4	38,745.9	
10	**Dammam-7** - Cray CS-Storm, Xeon Gold 6248 20C 2.5GHz, NVIDIA Tesla V100 SXM2, InfiniBand HDR 100, HPE Saudi Aramco Saudi Arabia	672,520	22,400.0	55,423.6	

Performance Trend & Projection (2021)



Source: www.top500.org

ExaFlop/s machine in 2021 ?!

600 million USD investment at Oak Ridge Nat. Lab. [1]

User operation in 2023 at Oak Ridge Nat. Lab. [2]

Basic trend: Slope changes
→ performance increase slows down

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

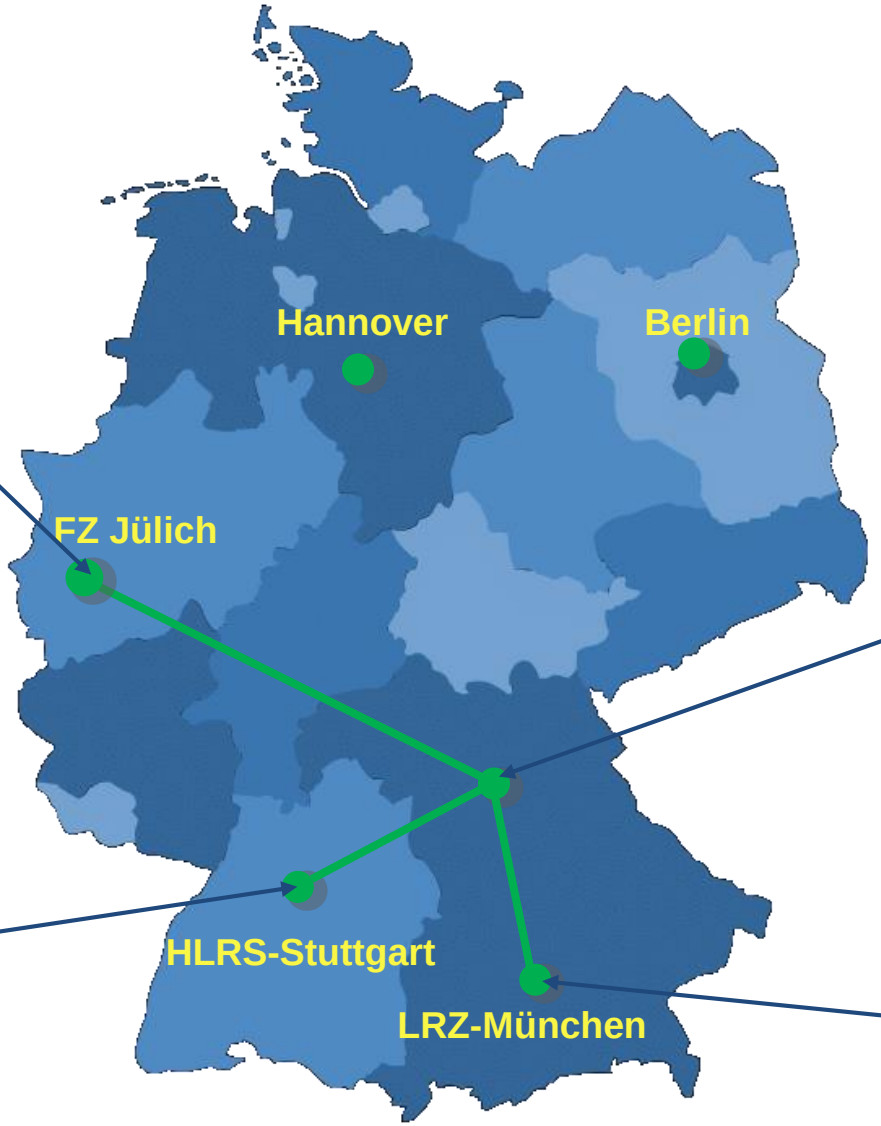
[2] https://www.ornl.gov/news/forging-ahead-frontier-ready-crush-science

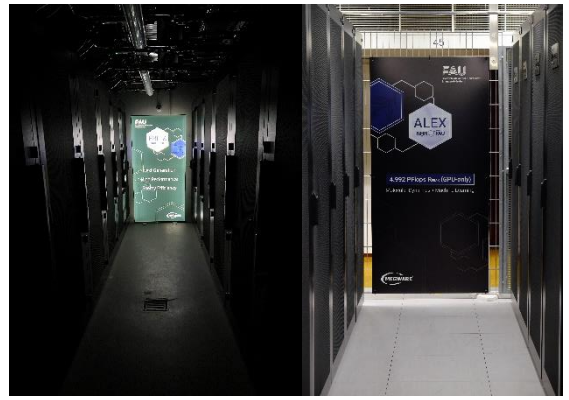
HPC Centers in Germany: A view from Erlangen

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)

HLRS



HLR Stuttgart
Hawk: 26 PF/s



SuperMUC-NG: 26.8 PF/s

Erlangen National Center for High Performance Computing

www.nhr.fau.de

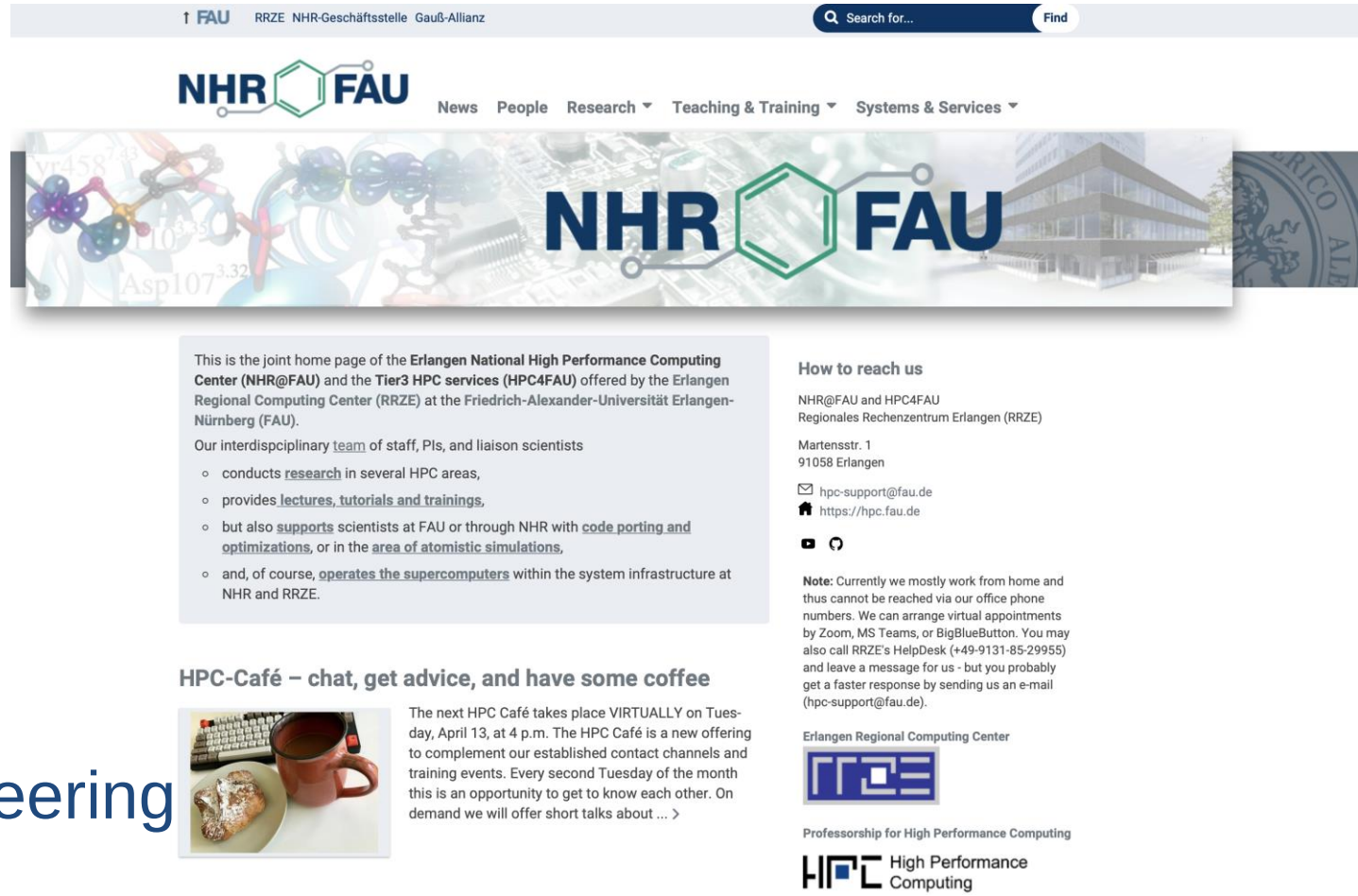
One out of nine national HPC centers at German universities

Ongoing:

- New system installed (10 M€+)
- New HPC center: 2027 (?)

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	944 Intel ICL (67,968 cores)	2 * 36 c (Xeon 8360Y) 256 GB 1 x HDR100	Shared PFS • 3 PB • 80 GB/s	1 – 64 nodes	**5.2 PF/s**
Alex	32 NVIDIA A100 (256 A100)	8 * A100 2 * 64 c (AMD) 1 TB 2 x HDR200	Node local 14 TB NVMe	1 – 8 GPUs	**5.0 PF/s**
	38 NVIDIA A40 (304 A40)	8 * A40 2 * 64 c (AMD) 0.5 TB	Node local 7 TB NVMe	1 – 8 GPUs	---

Fritz: https://hpc.fau.de/systems-services/systems-documentation-instructions/clusters/fritz-cluster/
Alex: https://hpc.fau.de/systems-services/systems-documentation-instructions/clusters/alex-cluster/

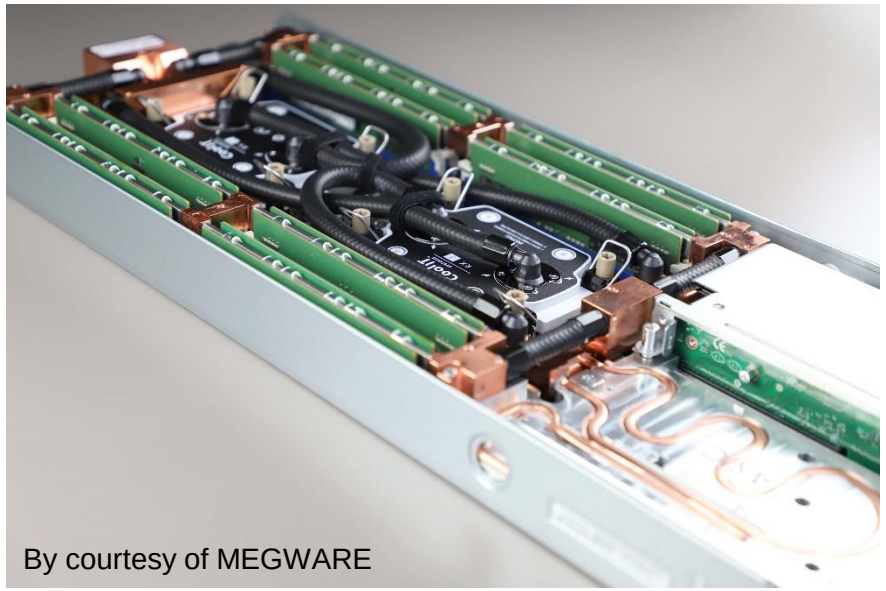
NF-Technikzentrale: “data center” for Fritz (750 KW)



By courtesy of MEGWARE






By courtesy of MEGWARE

