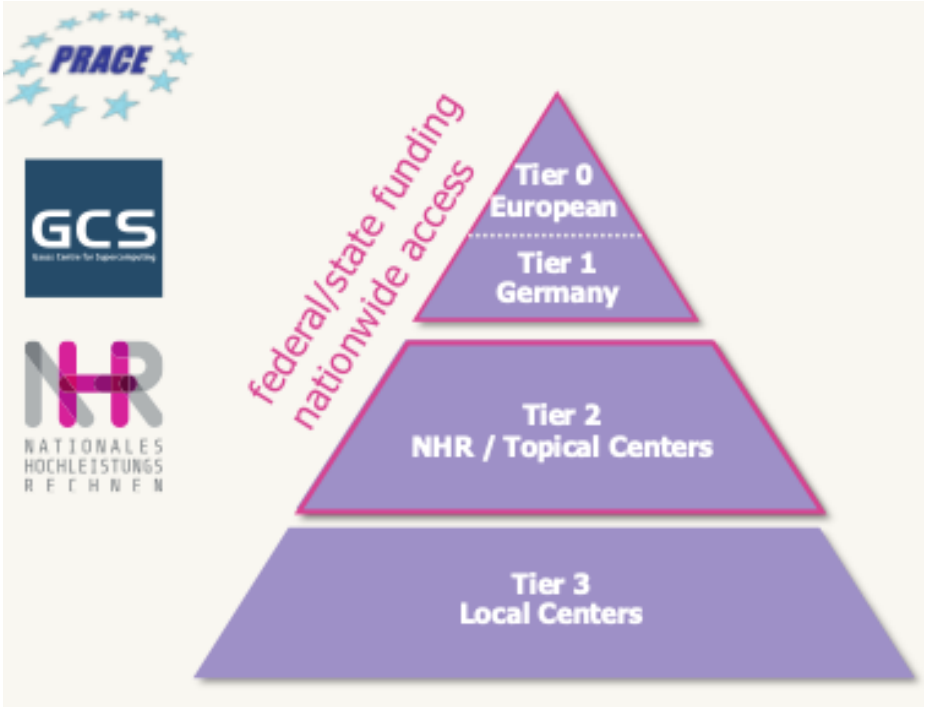
Karlsruher Institut für Technologie

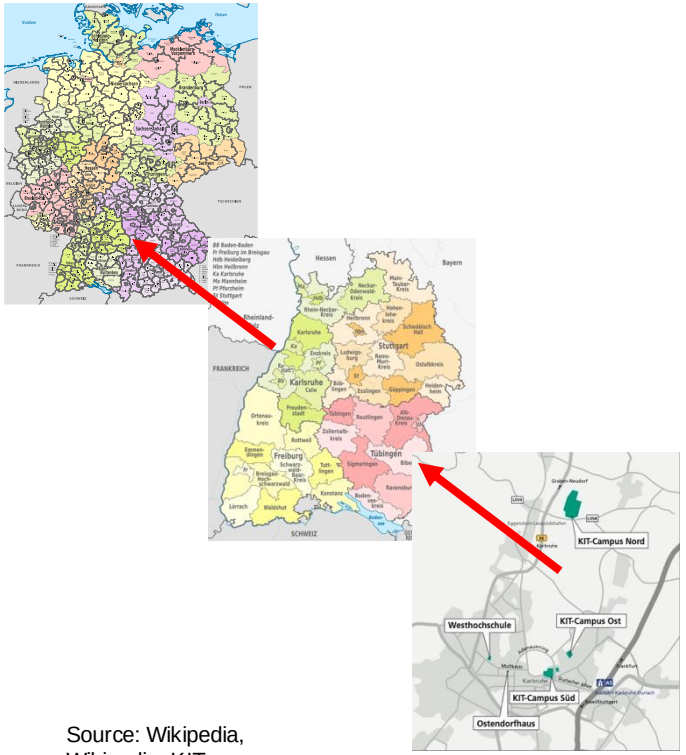
HPC in Germany: Local, state, and federal level

Journées calcul et données 2024

Martin Frank

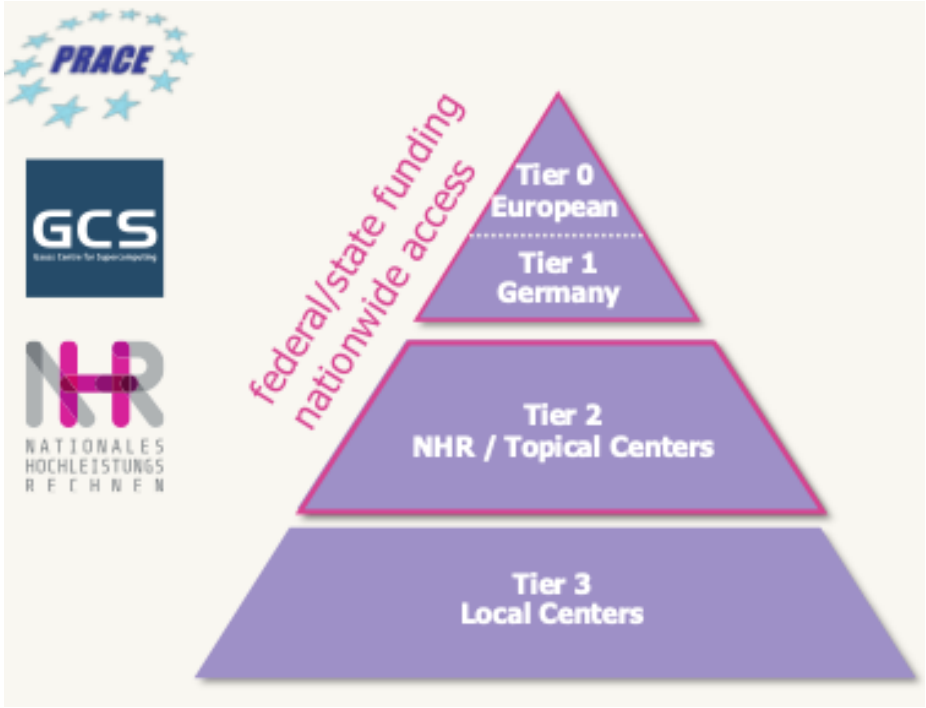
Outline

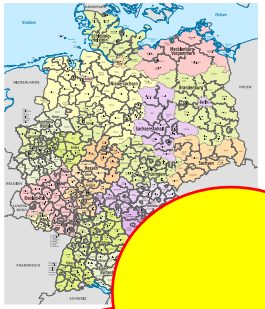




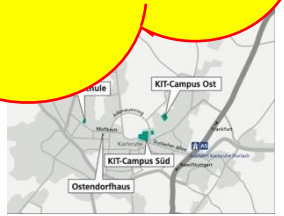
Source: Wikipedia,
Wikipedia, KIT

Outline





Plus Challenges & Future Directions



Source: Wikipedia, Wikipedia, KIT

Karlsruher Institut für Technologie

HPC in Karlsruhe

Big Data Infra. & Services, Data Science

- **GridKa:** the German data and analysis center for LHC (+ HL-LHC) and further particle and astroparticle experiments
  - ~ 70.000 CPU-cores, ~ 47 PB disk, ~ 73 PB tape, 400 Gb/s WAN
  - https://www.scc.kit.edu/en/research/gridka.php
- **LSDF:** multi-disciplinary large-scale data facility
  - About 10 PB online storage, tightly connected to HoreKa
  - https://www.scc.kit.edu/en/services/11228.php
- **Data Science Research**
  - Interdisciplinary: SimDataLabs and Helmholtz.AI
  - Generic: data management/metadata, data analytics/AI/ML
  - Active in large national & EU projects

HPC – Research Infrastructure

- **HoreKa**: Tier-2 flagship system
  - 17 PFLOPS, ~60,000 cores, 668 NVidia A100
  - Access via proposal
  - HoreKa-Blue (CPU), HoreKa-Green (GPU)
  - https://www.scc.kit.edu/dienste/horeka.php
- **bwUniCluster2.0**: Tier-3 local system
  - 902 compute nodes (different generations), 136 NVIDIA V100
  - Access via one signed form, JupyterHub, containers
  - https://www.scc.kit.edu/dienste/bwUniCluster_2.0.php
- SimDataLabs, SSPE Team
  - Joint projects (call for collaboration)
  - Voucher system in development
  - https://www.scc.kit.edu/forschung/sdl.php, https://www.scc.kit.edu/en/research/sspe.php

HoreKa-Teal

- GPU extension „**HoreKa-Teal**“
  - 22 GPU nodes
  - 4 NIVIDIA H100 GPUs each

- **#6 Green500 in 06/24**
- **#2 Germany**

- Efficiency: 63 GFLOP/(s*W)
- Peak power: 6 PFLOP/s

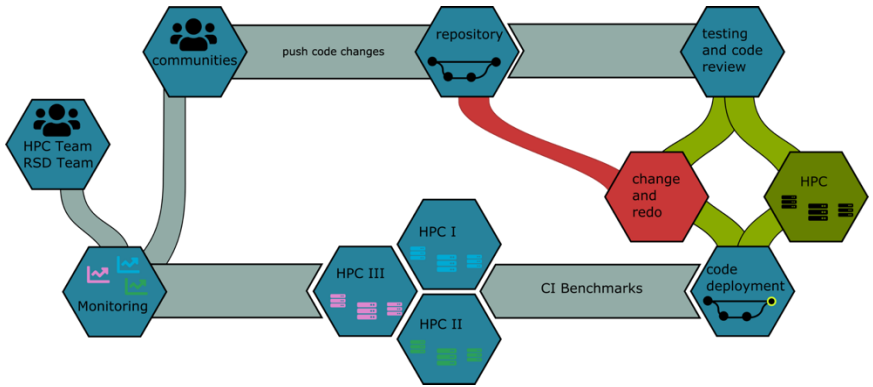
Data Center Building

- Passive cooling with **warm water** (42 Celsius in, 47 Celsius out)
- Utilization of **waste heat** (~80 kW) to heat office building
- German Data Center Award 2017 Category: Newly constructed energy- and resource-efficient data centers

User Support

- Continuous Integration framework

- Team Software Sustainability and Performance Engineering
- 4 Simulation and Data Lifecycle Labs
  - Earth System Science
  - Materials Science
  - Engineering
  - Elementary particle physics

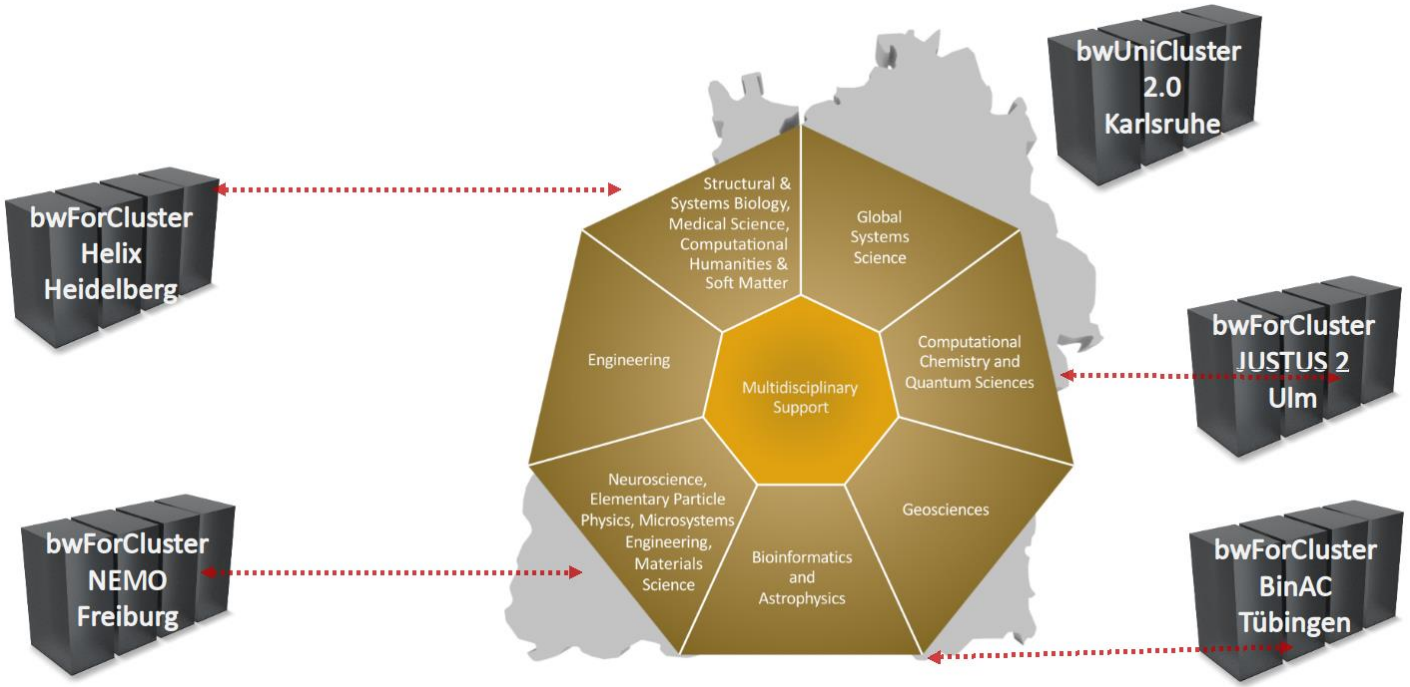


Karlsruher Institut für Technologie

HPC in Baden-Württemberg

Baden-Württemberg
THE LÄND

HPC in Baden-Württemberg



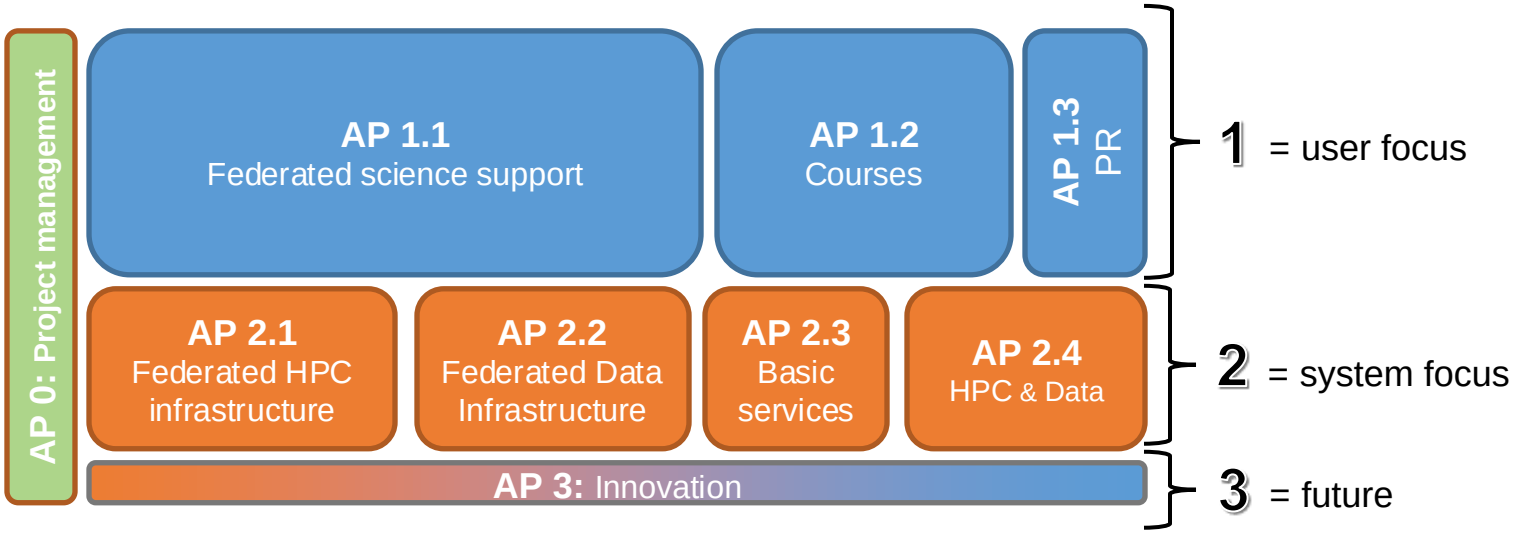
Baden-Württemberg HPC-DIC Strategy

- 3rd Framework 2025-2032
- DFG-reviewed in 2023
- Basis for funding by the state
- Strategy document (in German) published

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/148845

- Federated identity management
- Federated software portfolio (680+ packages)
- Integration of data management with HPC
- bwHPC-Wiki
- Federated resource provision (450+)
- Joint support projects (110+ Tiger Teams)
- (Online) courses
- Annual symposium

State-Funded Services Project (bwHPC-S5)



Karlsruher Institut für Technologie

National High-Performance Computing
Nationales Hochleistungsrechnen (NHR)

NHR Timeline

2012:
German Science and Humanities Council (Wissenschaftsrat): Position paper | Strategic further development of High Performance Computing in Germany

2015:
German Science and Humanities Council (Wissenschaftsrat): Recommendations on the funding of high and ultra-high performance computing in Germany

2018:
GWK resolution on funding of the National High Performance Computing

2019:
Constitution of the NHR Strategy Committee and establishment of an interim NHR office

2020:
Tender and funding decision

2021:
Start of funding

August 2021:
Foundation of the Association for National High Performance Computing – NHR Association

https://www.wissenschaftsrat.de/download/archiv/1838-12

NHR as Part of German HPC Infrastructure

- Recommendation by Research Council to introduce Tier-2 National High Performance Computing (NHR) infrastructure
- Competitive applications in 2020
  - official start: Jan 1, 2021
  - total funding 625M Euro (2021-2030)
  - currently 9 NHR centers
- Key aspects
  - joint federal/state-funding
  - transition from regional to competence-oriented for nationwide use
  - free access for all researchers from German universities
  - strengthen methodological competences through coordinated training, continuing education of users

NHR Members

- Rhein-Westfälische Technische Hochschule Aachen
- Zuse Institut Berlin (ZIB)
- Technische Universität Darmstadt
- Technische Universität Dresden
- Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg
- GWDG/Georg-August-Universität Göttingen
- Karlsruher Institut für Technologie
- Johannes Gutenberg Universität Mainz für das Konsortium Süd-West (Goethe-Universität Frankfurt, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Johannes Gutenberg Universität Mainz, Universität des Saarlandes)
- Universität Paderborn

Executive Board
Christof Schütte (ZIB)
Gerhard Wellein (FAU Erlangen)
Christian Plessl (Uni Paderborn)

NHR Office
Dörte Sternel (Managing Director)

Access to NHR Resources

Computing time allocation

- On a project-by-project basis according to a joint science-guided procedure
- NHR center freely selectable by applicants
- Central application portal (JARDS)

Training program

- Coordinated training program across centers
- Announcement via mailing list NHR announcements and website
- Courses on various HPC topics

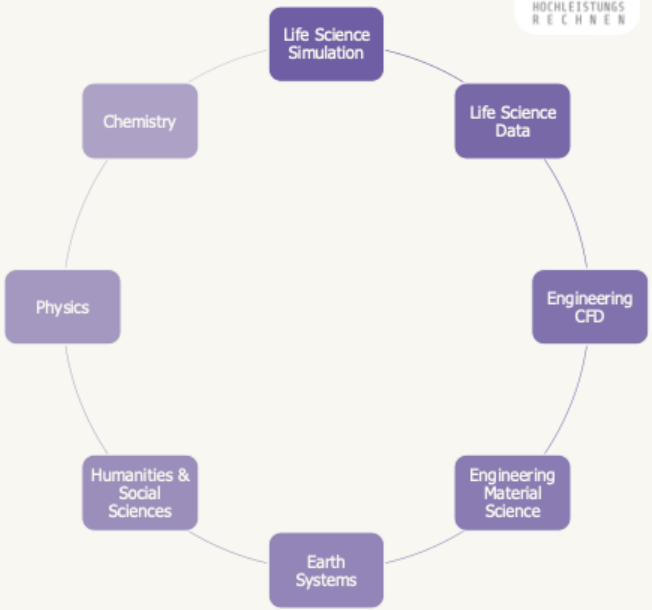
Consulting

- Center-specific focal points

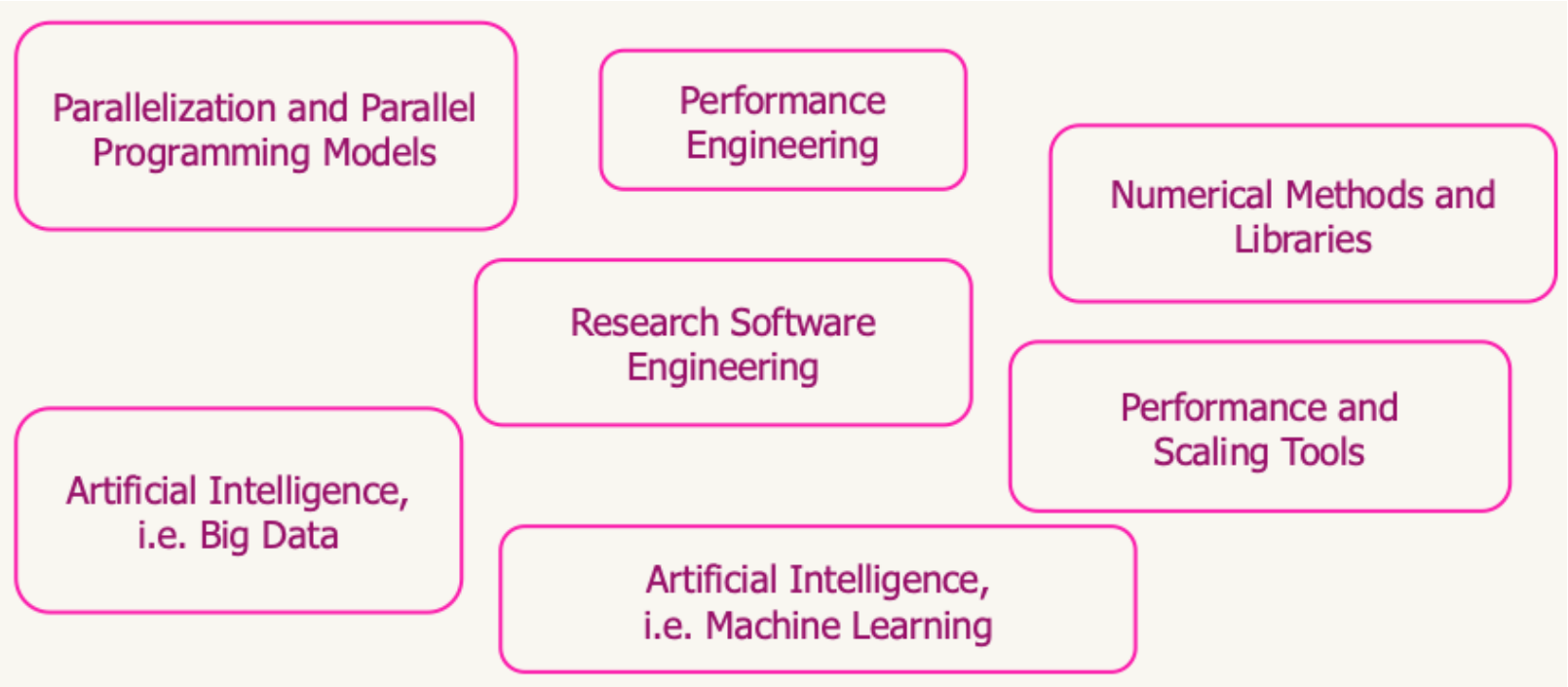
Application Specialization

Topic Specialization

- Centers coordinate to provide broad and complementary coverage of
  - science domains
  - methods
  - technologies
- Goal: provide tailored hardware, software, support, training
- One coordinating center per domain/topic
  - but no sole representation / responsibility
  - users can still apply at any center for their projects



Methodological Focus



Ongoing and Planned Developments

Attracting New Communities

- AI: in October 2023 Joint Special Information Initiative for AI community started
  - Special course programs for new users constantly further developed
  - "AI hardware" is available
  - Overlap between AI service centers and NHR centers

Data: MoU with National Research Data Infrastructure Germany (NFDI) in progress

- Already broad participation of NHR centers in NFDI consortia

Internationalization

- Deepening and expanding cooperation with international partner organizations (MoU with JHPCN)

Karlsruher Institut für Technologie

Challenges & Future Directions

Challenges

- Scientific challenges
- Energy efficiency and sustainability
- System components and architectures
- Software
- Data management
- Sensitive data
- Digital sovereignty
- Structures and organization

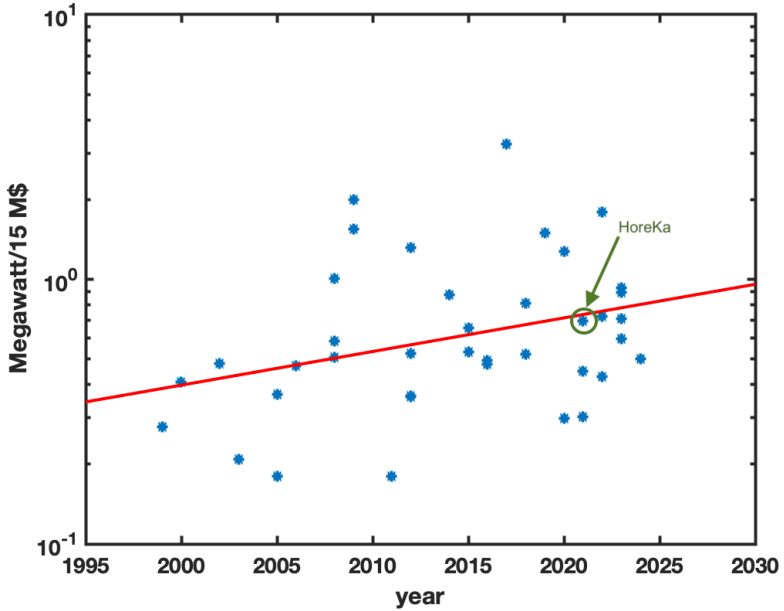
Energy: Cost & Availability

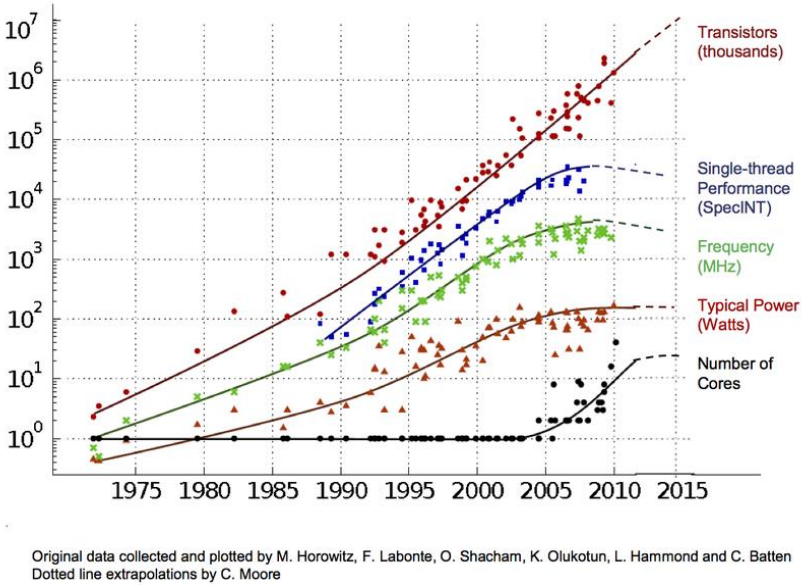
- Electricity costs of HoreKa calculated as
  800 kW x 365 days @ 17 Ct/kWh = 1.2 M€
- Electricity price 2023 ~40 Ct/kWh → annual costs = 2.8 M€

- Not possible to finance in the long run from total budget
  (7 M€/a, out of which 1.2 M€/a electricity)
- New systems planned for 2026 and 2032 @ 15 M€ each

- German Energy Efficiency Act, derived from European Energy Efficiency Directive:
  - Data centers that are currently in operation must have PUE <1.3 by July 1, 2030
  - Data centers that go into operation starting mid 2026 >10 % utilization of waste heat
  - Stricter regulations for new data centers
  - Law regulates air cooling temperatures, mandates establishment of an energy management system, further reporting duties

Outlook: Computing Power per Investment





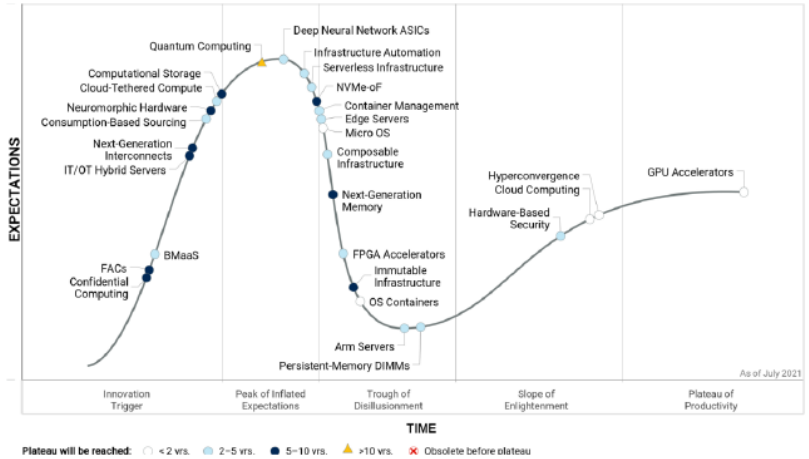
Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

Challenges

Architectures

- End of Moore's Law (doubling of compute power at same cost, area and power consumption)
- Significant growth in computing power only through specialized, heterogeneous architectures (GPUs, TPUs, chiplets)
- Innovation process to explore and test potentially disruptive technologies, including Quantum Computing

Figure 1: Hype Cycle for Compute Infrastructure, 2021



Medium-Term Path Toward GreenHPC

Energy budget

- Make energy consumption transparent
- Energy budget instead of compute budget (Fugaku)

→ **Offer support and joint development**

Operational models for renewable/varying energy supply

- Couple to availability/price
- Scheduling/dumping jobs
- Other budgeting models

→ **Offer support and joint development**

Software

- Key component of scientific work
- Challenge heterogeneity of the hardware

- RSE aiming at energy-optimized research software
- 3 pillars
  - Development
  - Education
  - Dissemination

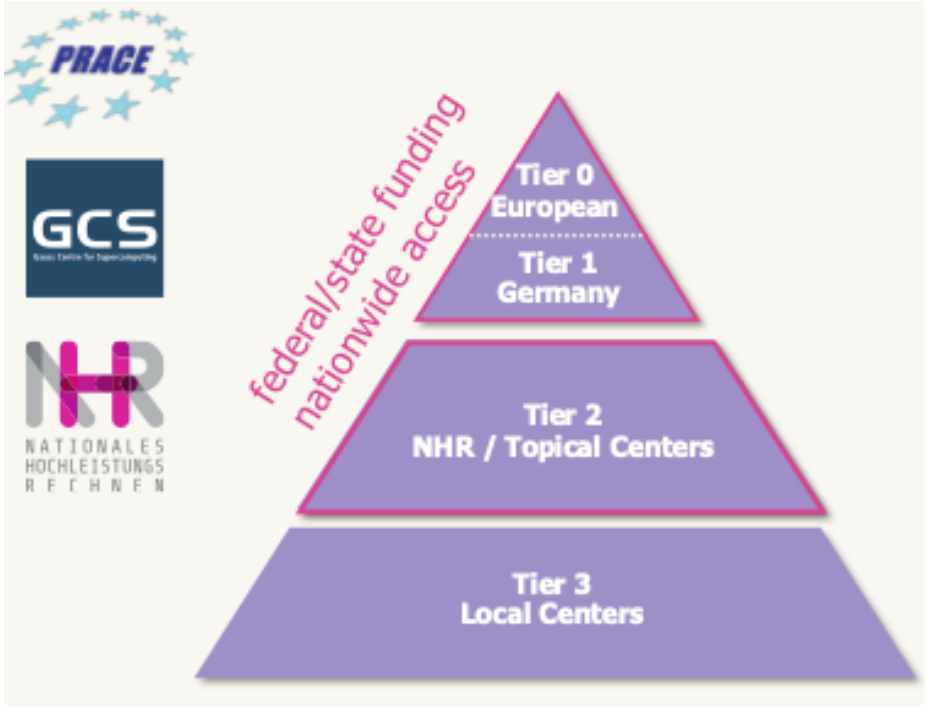
Digital Sovereignty

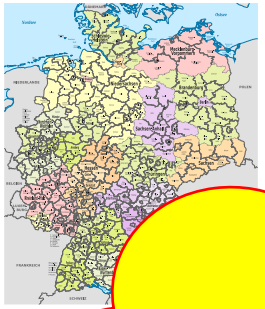
Digital sovereignty means having the skills and opportunities to complete tasks in a digital world independently, autonomously and securely:

- Transparency
- Ability to act
- Unrestricted choice of tools
- Maintaining and developing skills

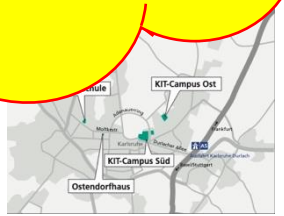
Plenty of possibilities for cooperation

Conclusion





Plus Challenges & Future Directions



Source: Wikipedia, Wikipedia, KIT