

NEXT-GENERATION SUPERCOMPUTING IN EUROPE: A COMMUNITY IN ACTION ADVANCING HARDWARE, SOFTWARE, AND APPLICATION PERFORMANCE FOR THE EXASCALE ERA

What's next for Infrastructure? | ESA-NASA International Workshop on AI Foundation Model for EO | 5-7 May 2025 | ESA-ESRIN | Frascati, Italy

GABRIELE CAVALLARO (WWW.GABRIELE-CAVALLARO.COM)

HEAD OF SIMULATION AND DATA LAB "AI AND ML FOR REMOTE SENSING", JÜLICH SUPERCOMPUTING CENTRE (FORSCHUNGSZENTRUM JÜLICH)

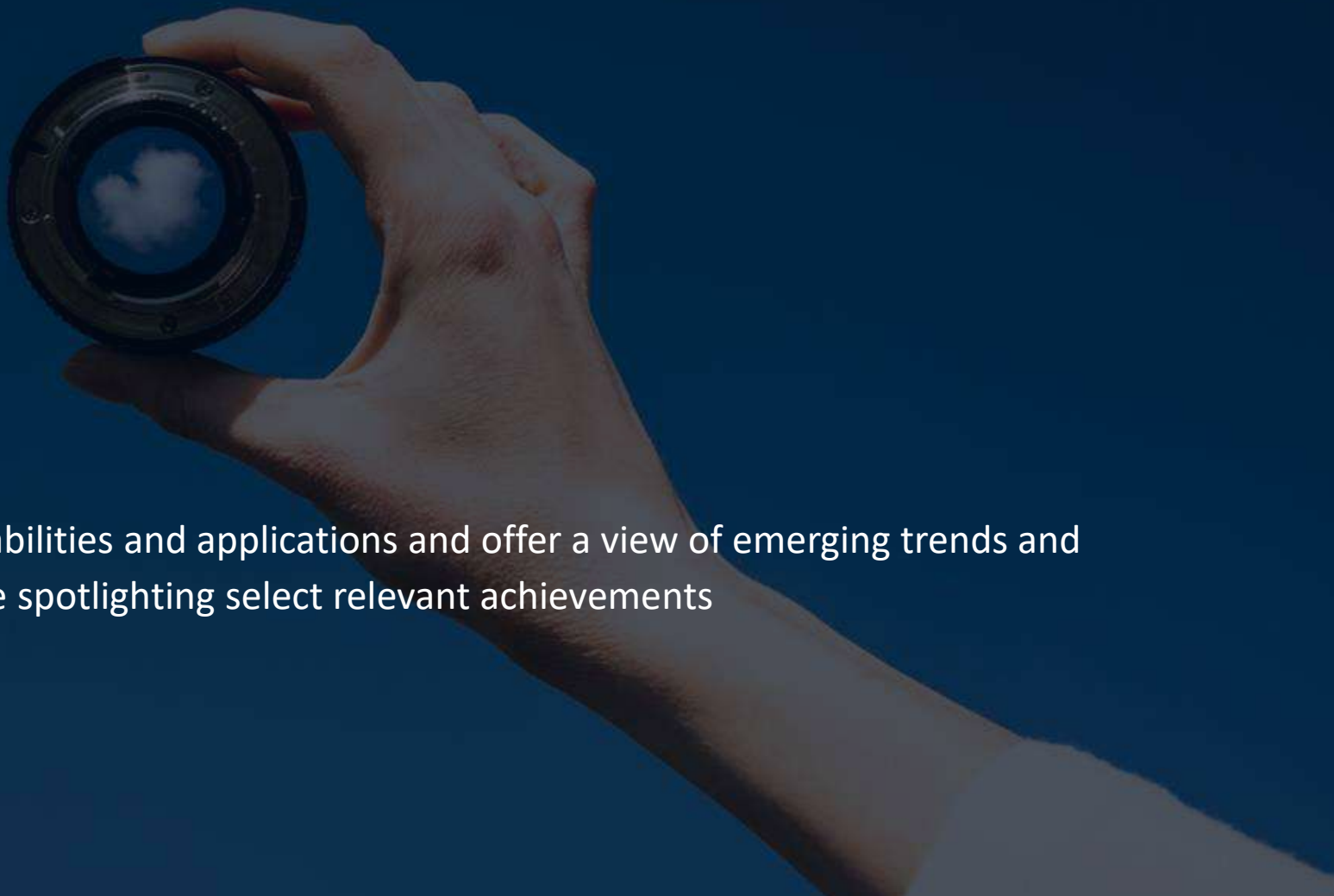
ASSOCIATE PROFESSOR, FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING (UNIVERSITY OF ICELAND)



JÜLICH
SUPERCOMPUTING
CENTRE



OBJECTIVE OF THIS TALK



Highlight today's supercomputing capabilities and applications and offer a view of emerging trends and challenges, while spotlighting select relevant achievements

OUTLINE

- 1 Background and Achievements
- 2 Evolution of Computing Technologies
- 3 Sustainability and Future Trends

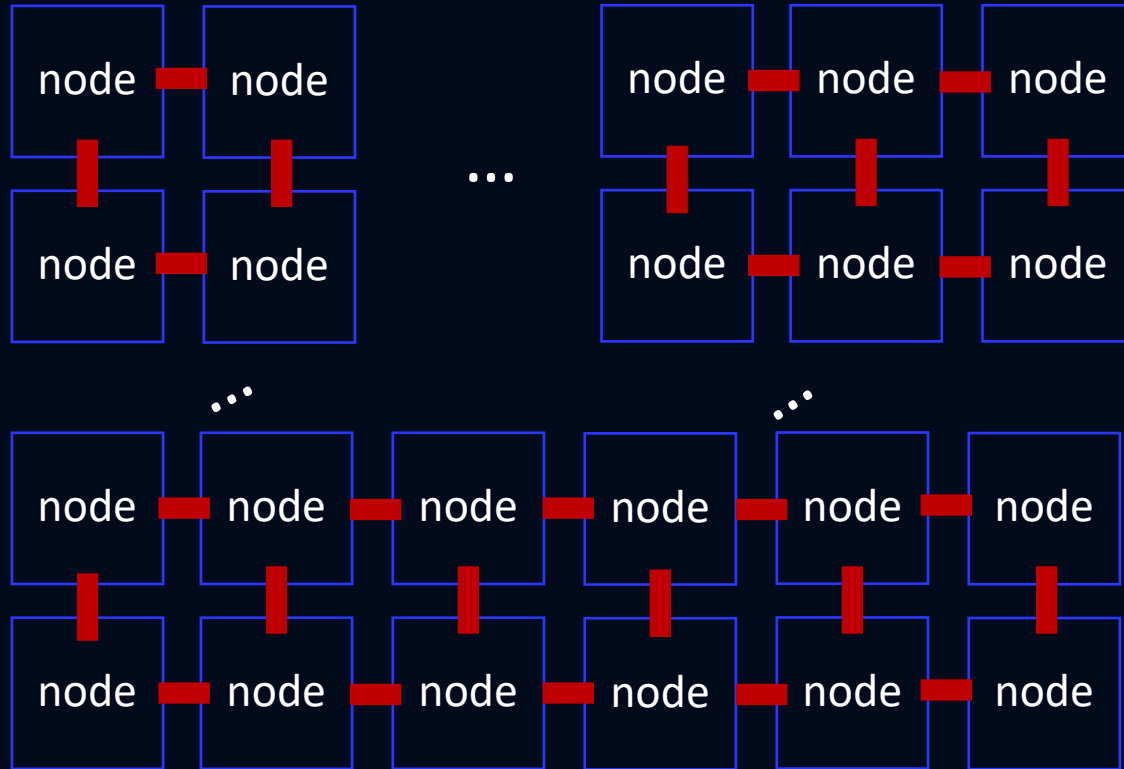
1

BACKGROUND AND ACHIEVEMENTS

WHAT ACTUALLY IS A SUPERCOMPUTER?

A large supercomputer system in a data center. The image shows several tall, dark, rectangular server racks with perforated doors, arranged in a long row. The racks are situated in a large, industrial-style room with a high ceiling and exposed metal trusses. The lighting is dim, with some blue light emanating from the racks. The overall atmosphere is technical and high-tech.

HIGH-PERFORMANCE COMPUTING SYSTEMS

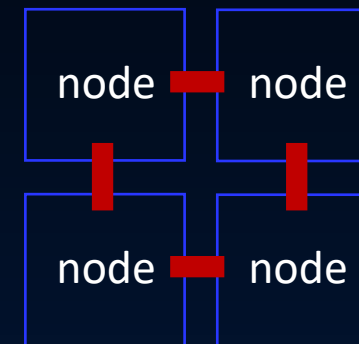


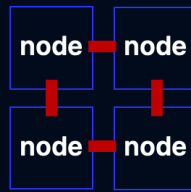
- High number of **compute nodes**
- Vast amounts of memory
- **High-speed interconnects**

HPC == tightly coupled parallel workloads



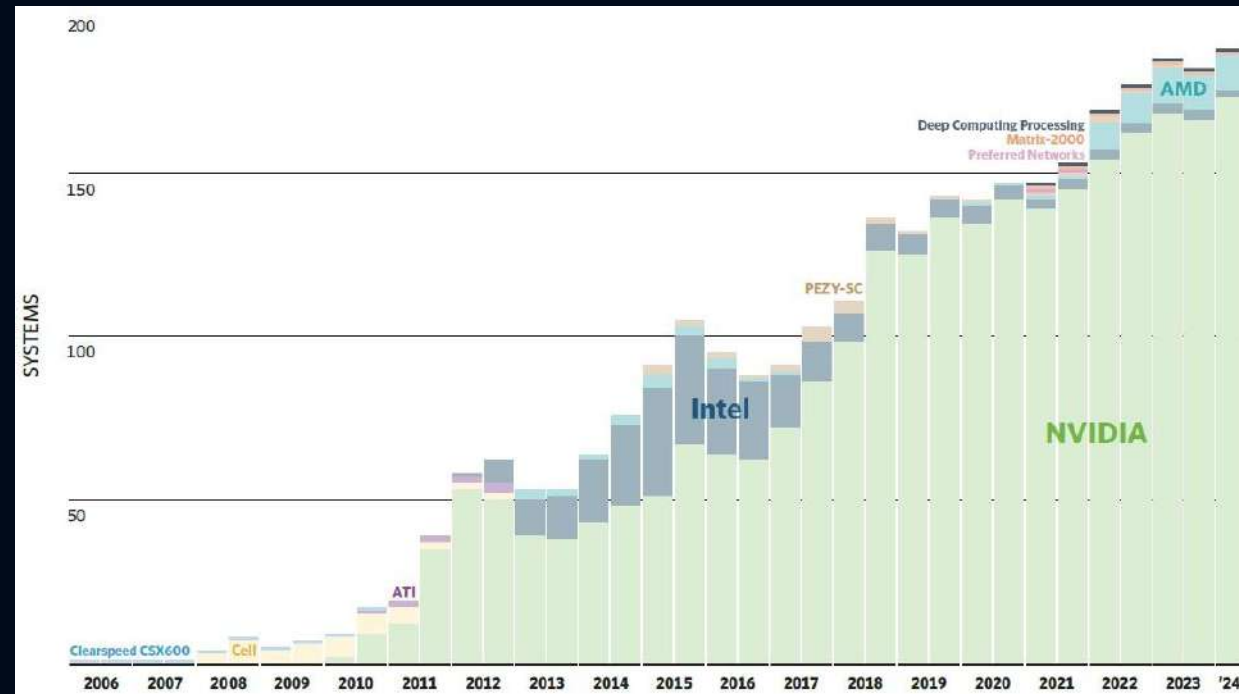
WHAT IS INSIDE COMPUTE NODES?





HPC SYSTEMS ARE ACCELERATED

- Last decade
 - More and more GPUs installed in HPC machines
 - More and more HPC machines with GPUs
 - More and more GPUs in each system
- Future
 - GPUs selected as technology for enabling Exascale
 - Even larger GPU machines with larger GPUs





FROM PETASCALE TO EXASCALE COMPUTING

≥ one quintillion (“1” followed by 18 zeros) calculations per second

Frontier supercomputer debuts as world's fastest, breaking exascale barrier



May 30, 2022

EXASCALE ERA (SINCE 2022)

HPE/Cray/AMD EPYC/Radeon FRONTIER

JUPITER: FIRST EUROPEAN EXASCALE SYSTEM



JU Pioneer for Innovative and Transformative Exascale Research (JUPITER)

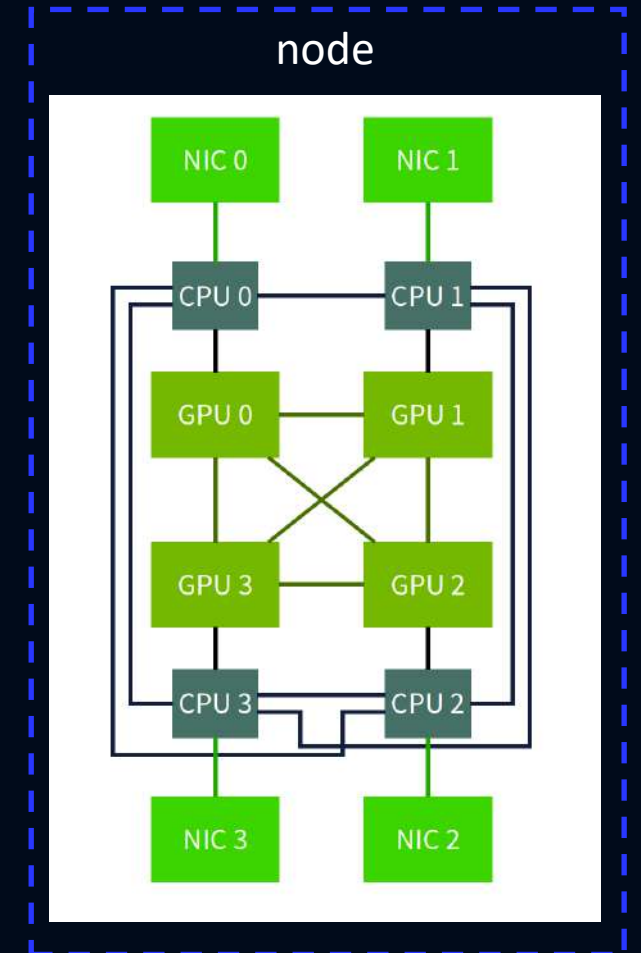
Performance of 1 million smartphones (a stack as tall as Mount Everest)

≈ 6000 nodes of JUPITER

JUPITER BOOSTER

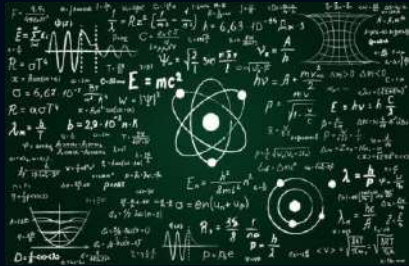
~6000 nodes, ~24 000 GPUs, 224 000 network devices

- **GPU** : 4 × NVIDIA H100 Grace-Hopper flavor 96 GB memory per GPU
- **CPU**: 4 × NVIDIA Grace, 4 × 72 cores; 4 × 120 GB LPDDR5X memory
- **Network** : 4 × NVIDIA Mellanox InfiniBand NDR200, 4 × 25GB/s



WHAT ARE SUPERCOMPUTERS USED FOR*?

ENGINE OF SCIENTIFIC PROGRESS. TACKLE PRESSING SOCIETAL PROBLEMS



Physics



Chemistry



Medicine



Climate

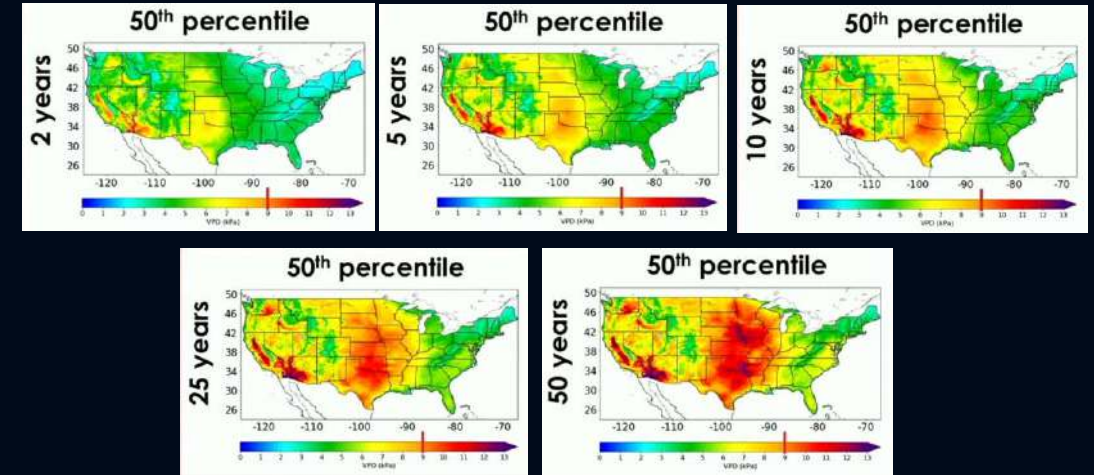


Society

Used for data-intensive and computation-heavy scientific and engineering applications

IMPROVE PREDICTION OF FLASH DROUGHTS WITH HPC

- New method to predict flash droughts
 - Vapor Pressure Deficit (VPD): Threshold 9kPa
 - Humidity=10%, T = 46°C
- Predictions, drought days/year above threshold:
 - +10 days in California(CA) coastline
 - +30-40 days in NW and MW
 - +100 days in CA central valley



HPC usage: Intrepid, Mira (BG/Q), Theta (KNL)

- Estimated compute time: ~1M node-hours
- Used 10 years of simulation data from other facility work (from multiple climate models)



COMMON TRENDS IN SCIENTIFIC COMMUNITIES

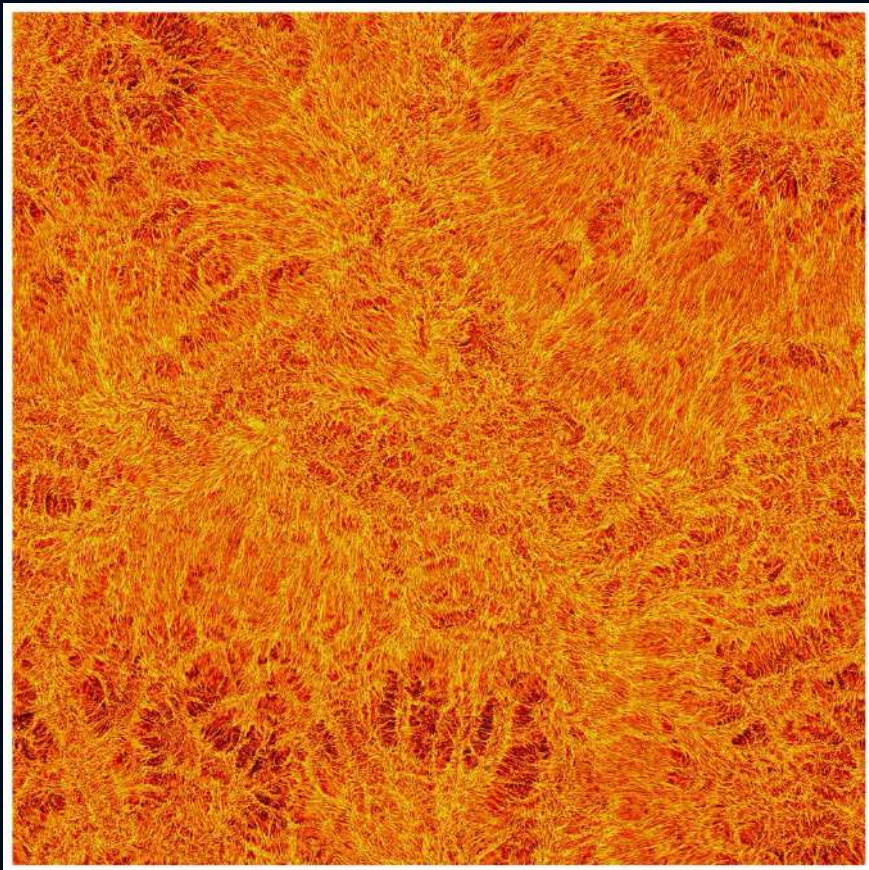
- **More Digital Twins**
- **More AI for science**
 - Foundation models
 - Data assimilation and interpretation
- **Tighter link to experiments and data sources**
 - Streaming data
 - Jointly analysing data from different sources
- **Challenges around dealing with and moving large data volumes**
 - Rather than about compute power

A photograph of a moving scene. In the background, a white moving truck is parked on a paved driveway, with its rear door open. Two movers in blue uniforms and caps are loading the truck with cardboard boxes and furniture. In the foreground, a man and a woman are standing with their backs to the camera, watching the movers. The man is wearing a red and blue striped shirt, and the woman is wearing a light blue shirt and dark shorts. To the right of the truck, there is a stack of three cardboard boxes. The scene is set in front of a house with a stone wall and a garden.

COMPUTE IS MUCH EASIER
TO MOVE THAN DATA

HOW TO ANALYSE EXTREME SCALE SIMULATIONS?

Example: in situ visualization to access more data



In Situ Workflow using ASCENT

- Application: Thermal convection
 - For understanding complex phenomena in stellar and planetary interiors and atmospheres driven by several factors in combination
 - Researchers often study the idealized model of convection – Rayleigh–Bénard convection (RBC)
- Large HPC setup with 3360 GPUs:
 - $46,7 \times 10^9$ unique points
 - Computed with nekRS 23.0



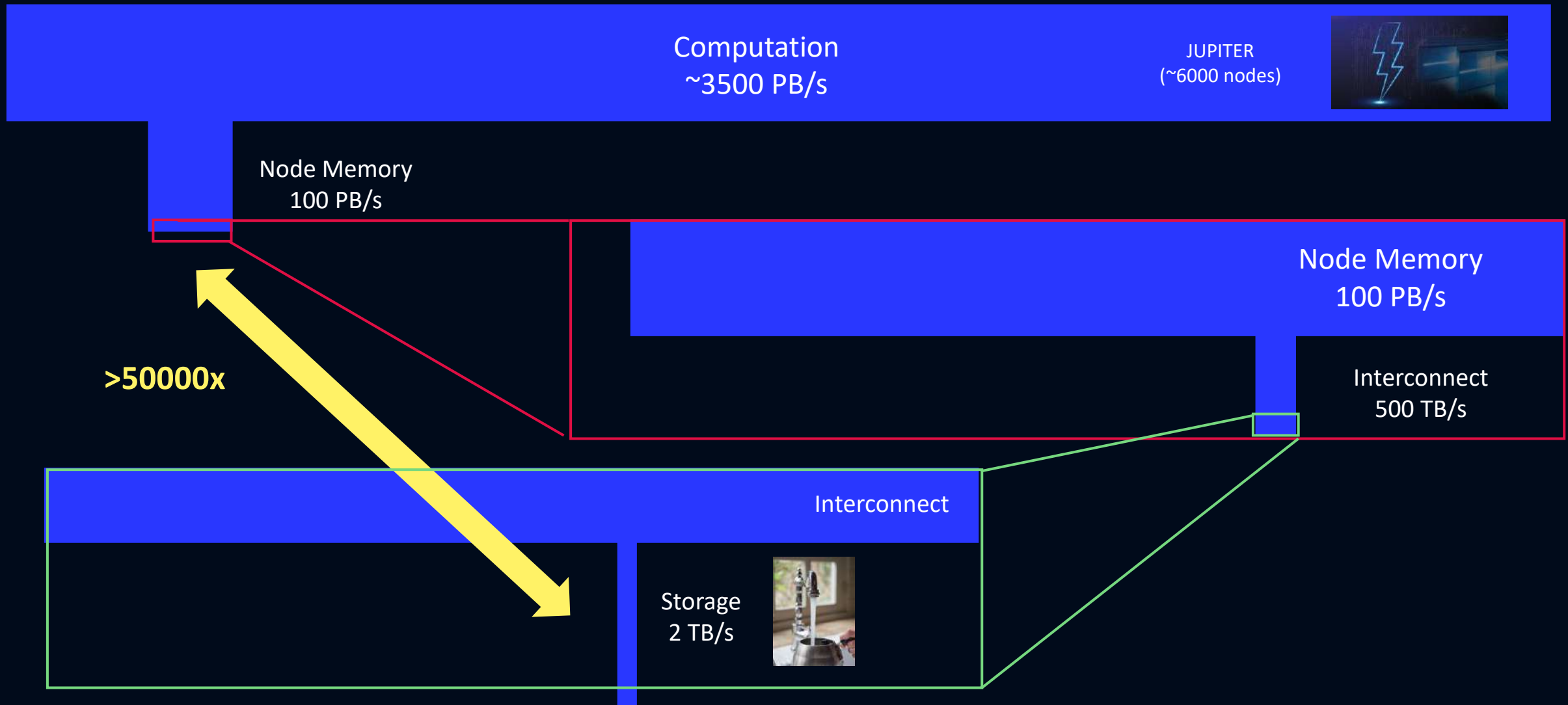
Convective mesoscale turbulence at very low Prandtl numbers
Ambrish Pandey, Dmitry Krasnov, Katepalli R. Sreenivasan and Jörg Schumacher

PR "NekRS + Ascent for In-Situ Visualization" <https://github.com/Nek5000/nekRS/pull/547>

Credits: Victor A. Mateevitsi, Mathis Bode, Nicola Ferrier, Paul Fischer, Jens Henrik Göbbert, Joseph A. Insley, Yu-Hsiang Lan, Misun Min, Michael E. Papka, Saumil Patel, Silvio Rizzi, ROSHAN Samuel, Jörg Schumacher and Jonathan Windgassen

Jülich Supercomputing Centre, "Jülich Wizard for European Leadership Science (JUWELS)", <https://www.fz-juelich.de/en/ias/jsc/systems/supercomputers/juwels>

HOW TO ACCESS MORE DATA?



Credits: Jens Henrik Göbbert, (Jülich Supercomputing Centre, Forschungszentrum Jülich)

Jülich Supercomputing Centre (Forschungszentrum Jülich), "JUPITER | The Arrival of Exascale in Europe", <https://www.fz-juelich.de/en/ias/jsc/jupiter>

A person is shown working on a dense network of cables in a server room. The cables are bundled and organized, with some colors like yellow and blue visible. The person is wearing a dark jacket and is focused on the task. The background shows more server racks and cables, creating a complex and technical environment.

IN HPC, HIGH FRACTION OF HARDWARE COST
GOES INTO NETWORK

HPC SPENDS WHATEVER IT TAKES TO ADDRESS THE CHALLENGES IN DISTRIBUTED COMPUTING

1. The network is reliable
2. Latency is low and fixed
3. Bandwidth is high and fixed
4. The network is secure
5. Topology doesn't change
6. There is one administrator
7. The network is homogeneous
8. Transport cost is negligible



Eight* fallacies of distributed computing

*Peter Deutsch of Sun Microsystems, 1994 ... item 8 added in 1997 by James Gosling

Source: https://en.wikipedia.org/wiki/Fallacies_of_distributed_computing

Credits and many thanks to Tim Mattson (University of Bristol and Merly.ai), "The Future and what to do about it: People, processors, and programming"

EUROPE IS INVESTING IN AI LEADERSHIP

To support scientific advances, infrastructure development, and wider technology adoption.

HPC POWER IN THE EU IS PUBLICLY ACCESSIBLE

Europe hosts 30% of the world's top ten supercomputers

- European network of cutting-edge supercomputers deployed by the European High-Performance Computing Joint Undertaking (**EuroHPC JU**)
 - EuroHPC was launched in 2018 and co-funded by the EU, Member States, and private actors.
- TOP500 (November 2024, global ranking)
 - LUMI (#8)
 - Leonardo (#9)
 - MareNostrum 5 ACC (#11)



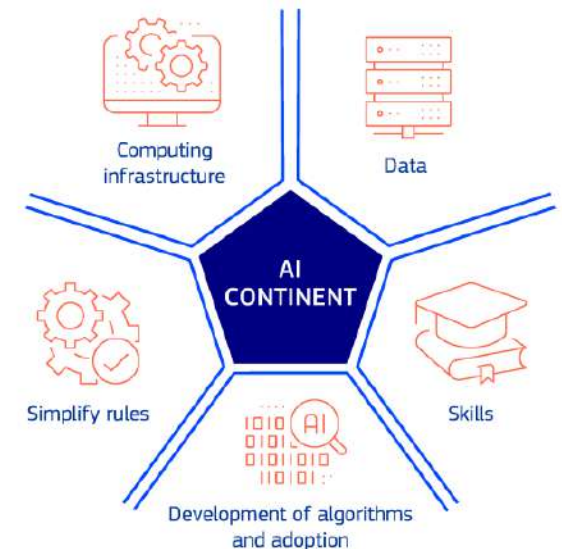
AI CONTINENT ACTION PLAN

5 Pillars for Europe to become the AI Continent

1. Building a large-scale AI computing infrastructure
2. Increasing access to high-quality data
3. Promoting AI in strategic sectors
4. Strengthening AI skills and talents
5. Simplifying the implementation of the AI act

PRESS RELEASE | Feb 11, 2025 | Paris | 3 min read

EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence*



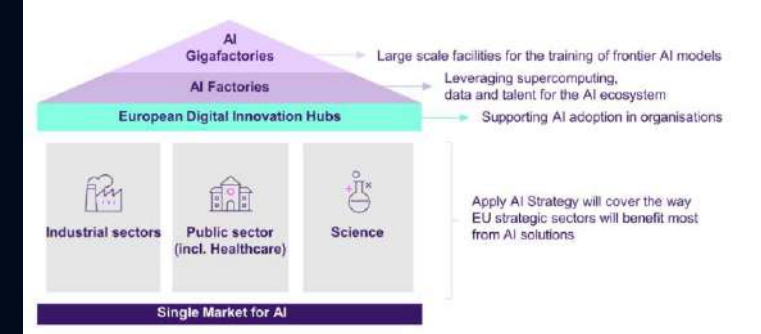
European Commission, "Shaping Europe's leadership in artificial intelligence with the AI continent action plan", https://commission.europa.eu/topics/eu-competitiveness/ai-continent_en

European Commission, "AI Continent Action Plan COM(2025)165", 2025, https://commission.europa.eu/topics/eu-competitiveness/ai-continent_en

AI Continent Action Plan, <https://digital-strategy.ec.europa.eu/en/factpages/ai-continent-action-plan>

AI CONTINENT ACTION PLAN

Area: Computing infrastructure



AI Factories

- Objective: train and finetune AI models
- Budget: €10 billion from 2021 to 2027
- At least 13 operational AI factories by 2026

AI Gigafactories

- Objective: train and develop complex AI models
- 4x more powerful than AI Factories
- €20 billion mobilised by InvestA!
- Deploy up to 5 Gigafactories

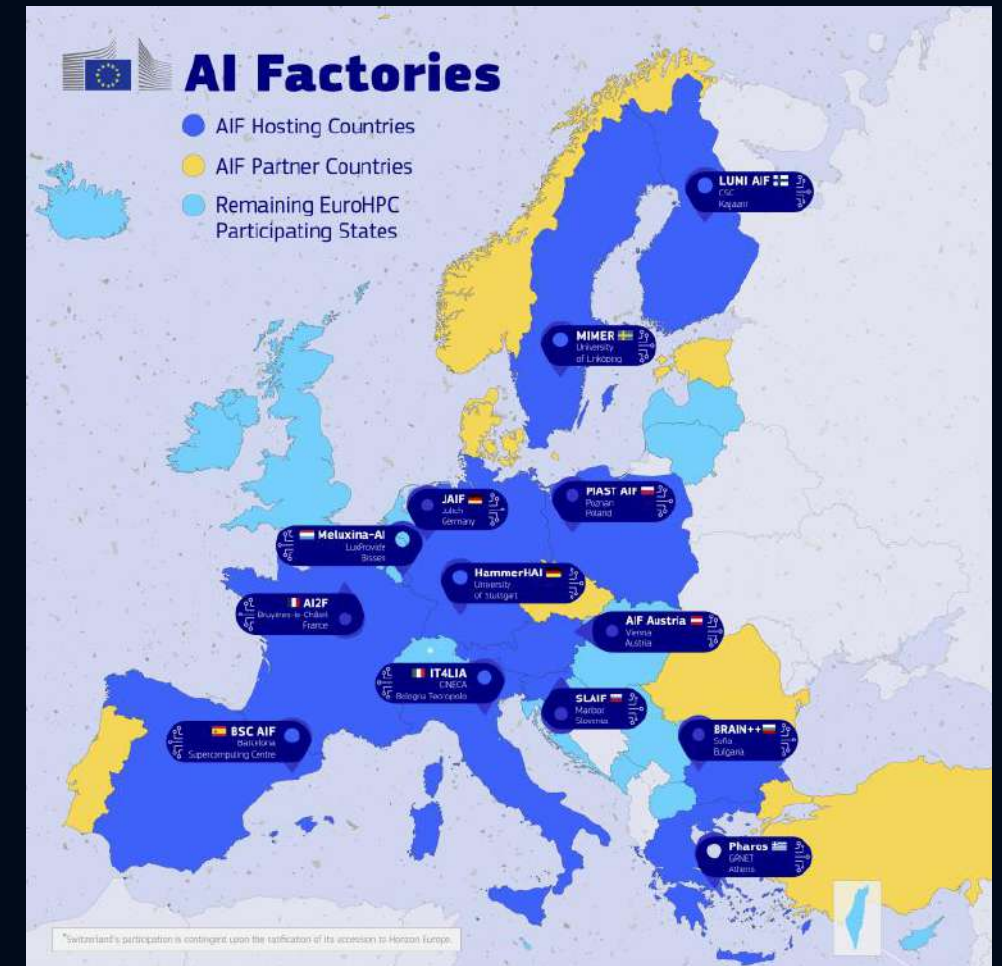
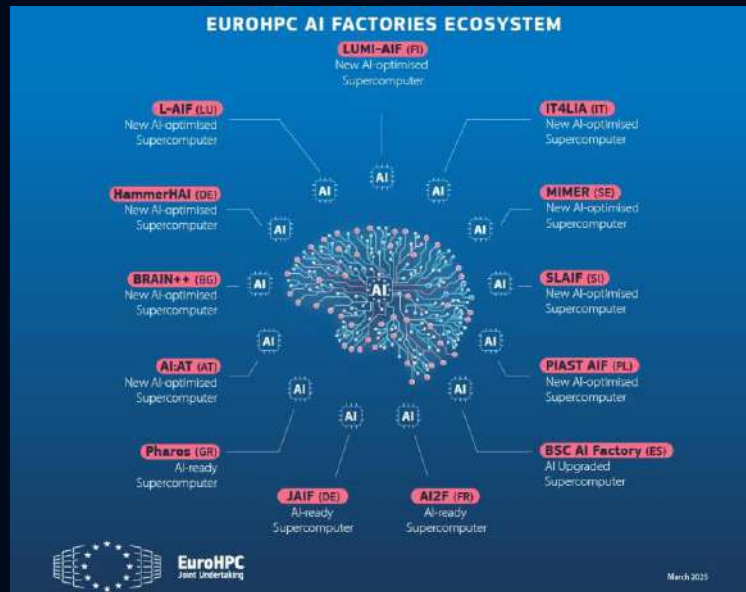
Cloud and AI development Act

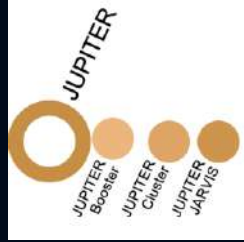
- Objective: boost research in highly sustainable infrastructure
- Encourage investments
- Triple the EU's data centre capacity in the next 5-7 years

EUROHPC AI FACTORIES

To triple the current EuroHPC AI computing capacity

- Facilitate access to AI provided by HPC facilities
- Dynamic ecosystems that foster innovation, collaboration, and development in the field of AI
- Support startups, industry, and researchers to develop cutting-edge AI models and applications.





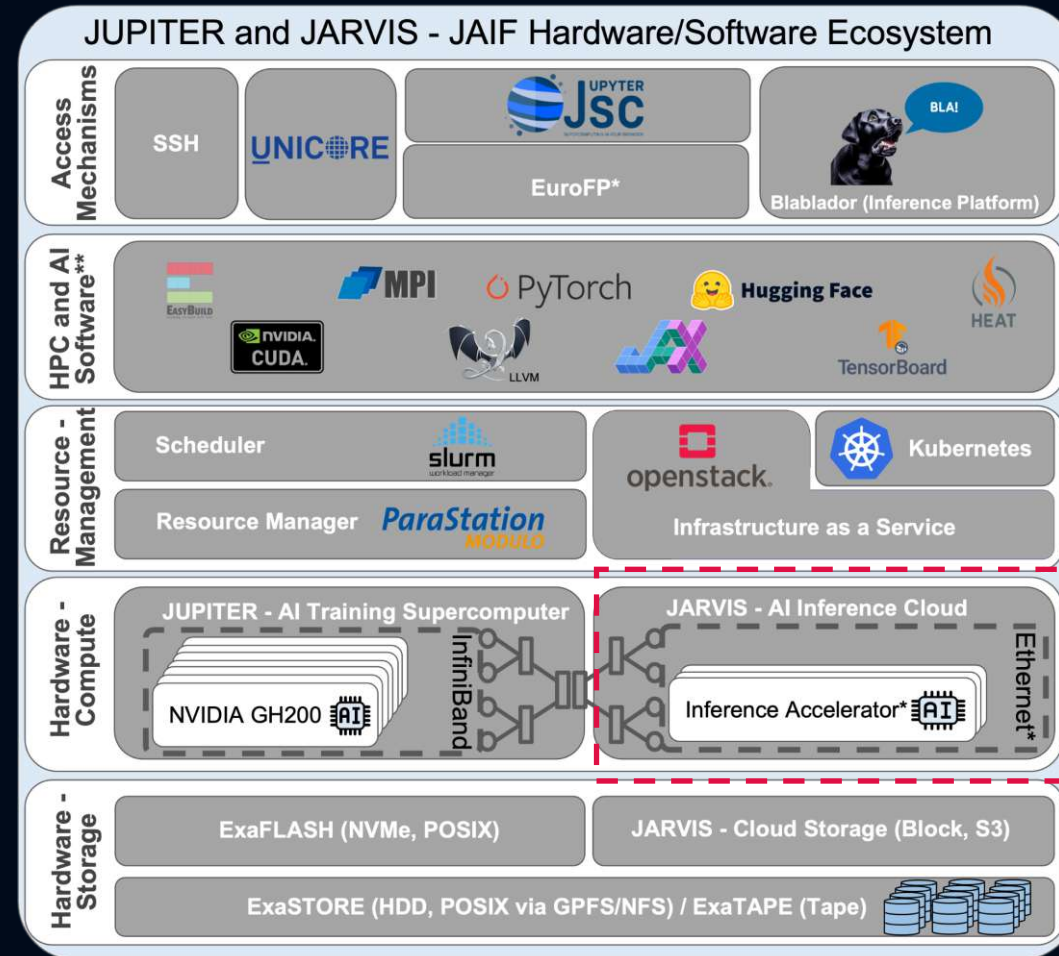
THE JUPITER AI FACTORY (JAIF)

Modular JUPITER - Hybrid Training/Inference AI System

Consortium



Contact: jaif@fz-juelich.de



JUPITER inference system **JARVIS** (JUPITER Advanced Research Vehicle for Inference Services), a cloud-based AI inference platform

*Depending on procurements and available functionality

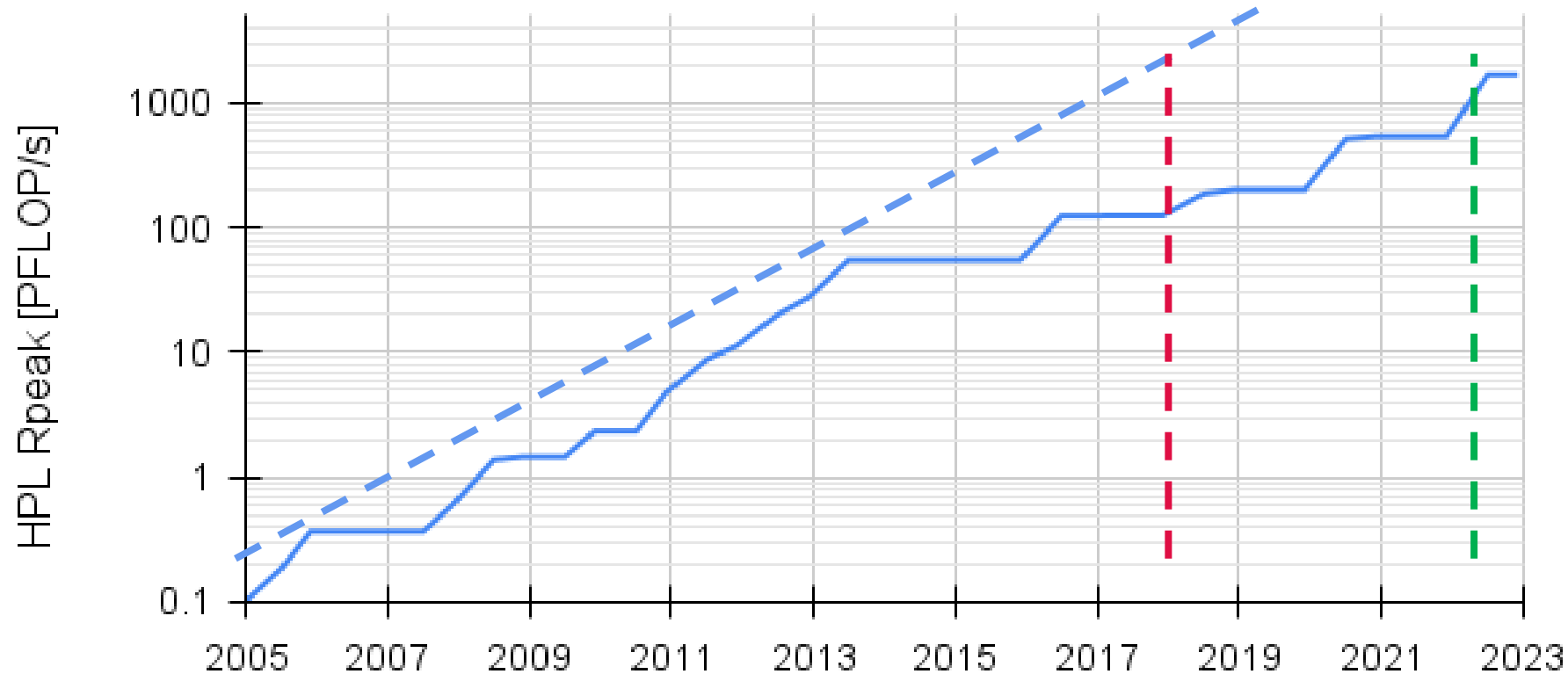
**This is a subset of the available software

2

EVOLUTION OF COMPUTING TECHNOLOGIES

EXASCALE ERA

Top #1: HPL Rpeak [PFLOP/s]



- **1997:** First 1 TFLOP/s computer:
(*Intel ASCI Red/9152*)
- **2008:** First 1 PFLOP/s computer: (*Roadrunner*)
- So.... First 1 EFLOP/s computer: **2018 !!**
 - Well... not really
- It took 4 more years... **2022**

HPE/Cray/AMD
EPYC/Radeon
FRONTIER



<https://www.top500.org/>

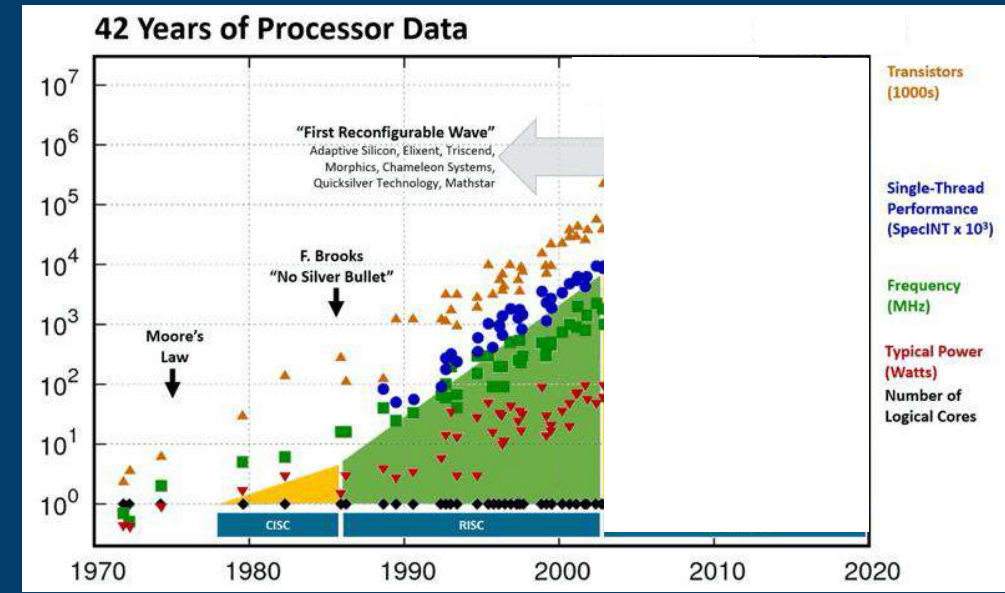
<https://www.ornl.gov/news/ornl-celebrates-launch-frontier-worlds-fastest-supercomputer>



UNTIL ~2005, PERFORMANCE CAME FROM THE BOTTOM

Moore's Law

Drove the semiconductor industry to cram more and more transistors and logic into the same volume



Semiconductor Technology

Up until ~2005,

performance came

from the bottom

Credit and many thanks: Tim Mattson (University of Bristol and Merly.ai), "Parallel Programming with Python"

Charles E. Leiserson et al., "There's plenty of room at the Top: What will drive computer performance after Moore's law?", in Science, vol. 368, 2020, <http://doi.org/10.1126/science.aam9744>

J. L. Hennessy, D. A. Patterson, "A New Golden Age for Computer Architecture", in Communications of the ACM, vol. 62 no. 2, pp. 48-60, 2019, <https://doi.org/10.1145/3282307>

Hennessy and Patterson, Turing Lecture 2018, overlaid over "42 Years of Processors Data" <https://www.karirupp.net/2018/02/42-years-of-microprocessor-trend-data/>; "First Wave" added by Les Wilson, Frank Schirmer Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten New plot and data collected for 2010-2017 by K. Rupp

SINCE ~2005 PERFORMANCE COMES FROM "THE TOP"

End of Dennard's Scaling

Limits in how much it is possible to shrink voltage and current without losing predictability

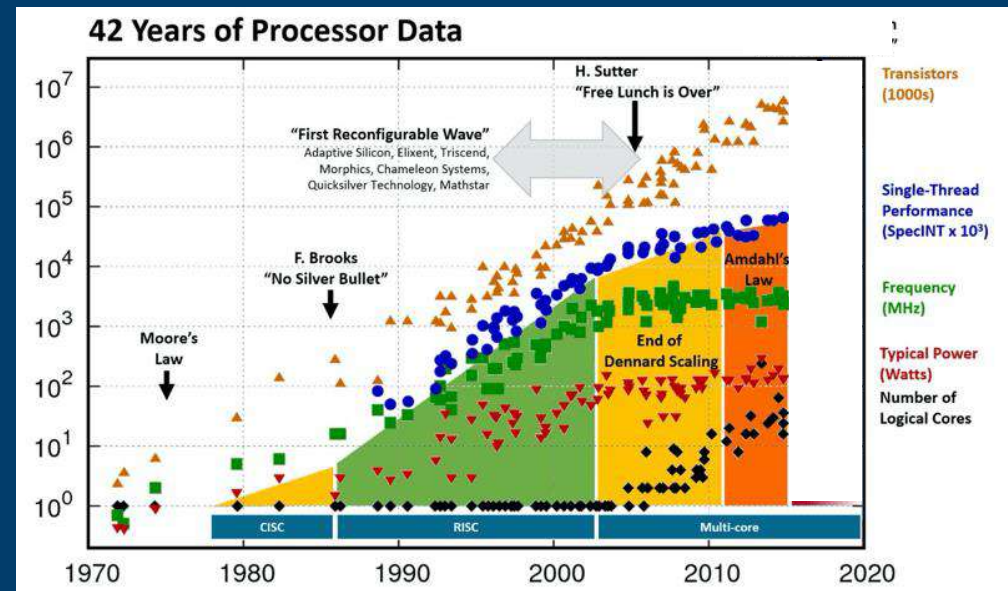
Multi-Core Era

Triggered by the Instruction Level Parallelism (ILP) wall

Amdahl's Law

Challenges in terms of energy efficiency, thermal management and parallelizability

Better software technology and algorithms



Hennessy and Patterson, Turing Lecture 2018, overlaid over "42 Years of Processors Data" <https://www.karup.net/2018/02/42-years-of-microprocessor-trend-data/>; "First Wave" added by Les Wilson, Frank Schirmer; Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, D. Shacham, K. Olukotun, L. Hammond, and C. Batten. New plot and data collected for 2010-2017 by K. Rupp

Still better HW architecture (HW architecture matters,
but much less than software and algorithms)

Semiconductor Technology

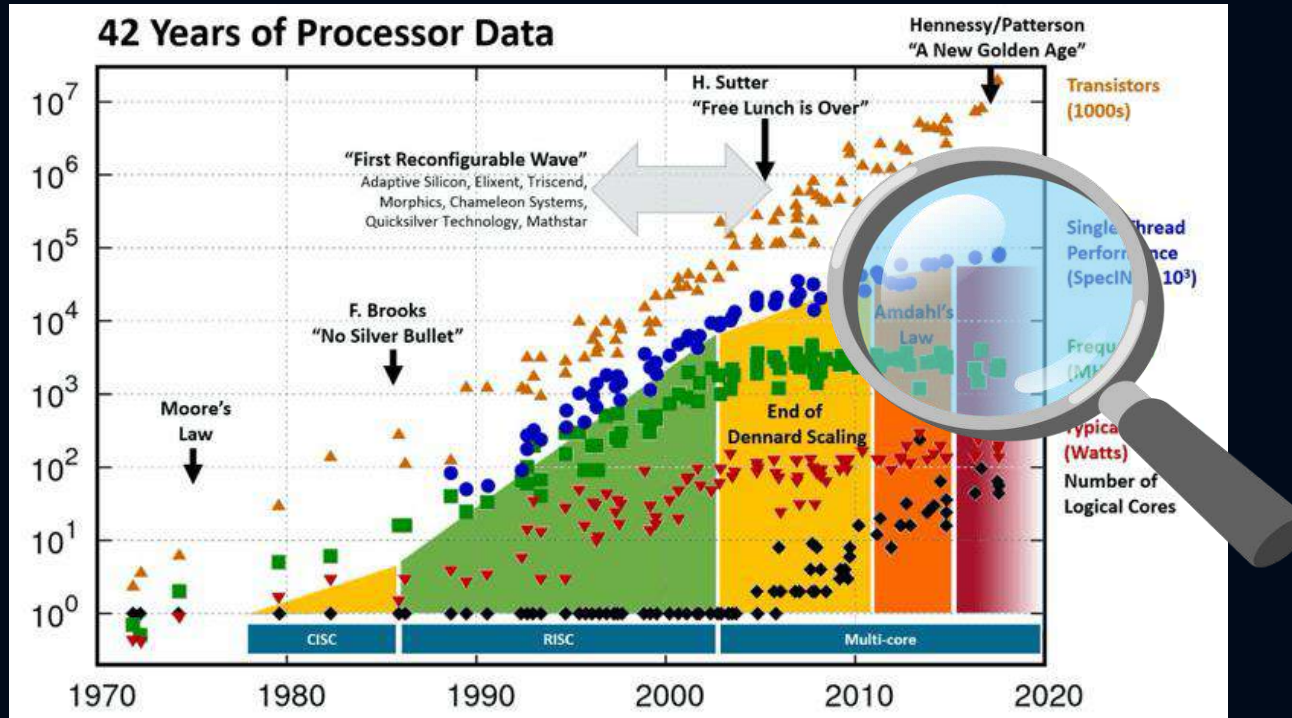
A conceptual image featuring a globe at the center, surrounded by a large number of small, diverse human figures. The figures are arranged in a circular pattern around the globe, suggesting a global community or the impact of human actions on the planet. The background is a dark, muted blue.

3

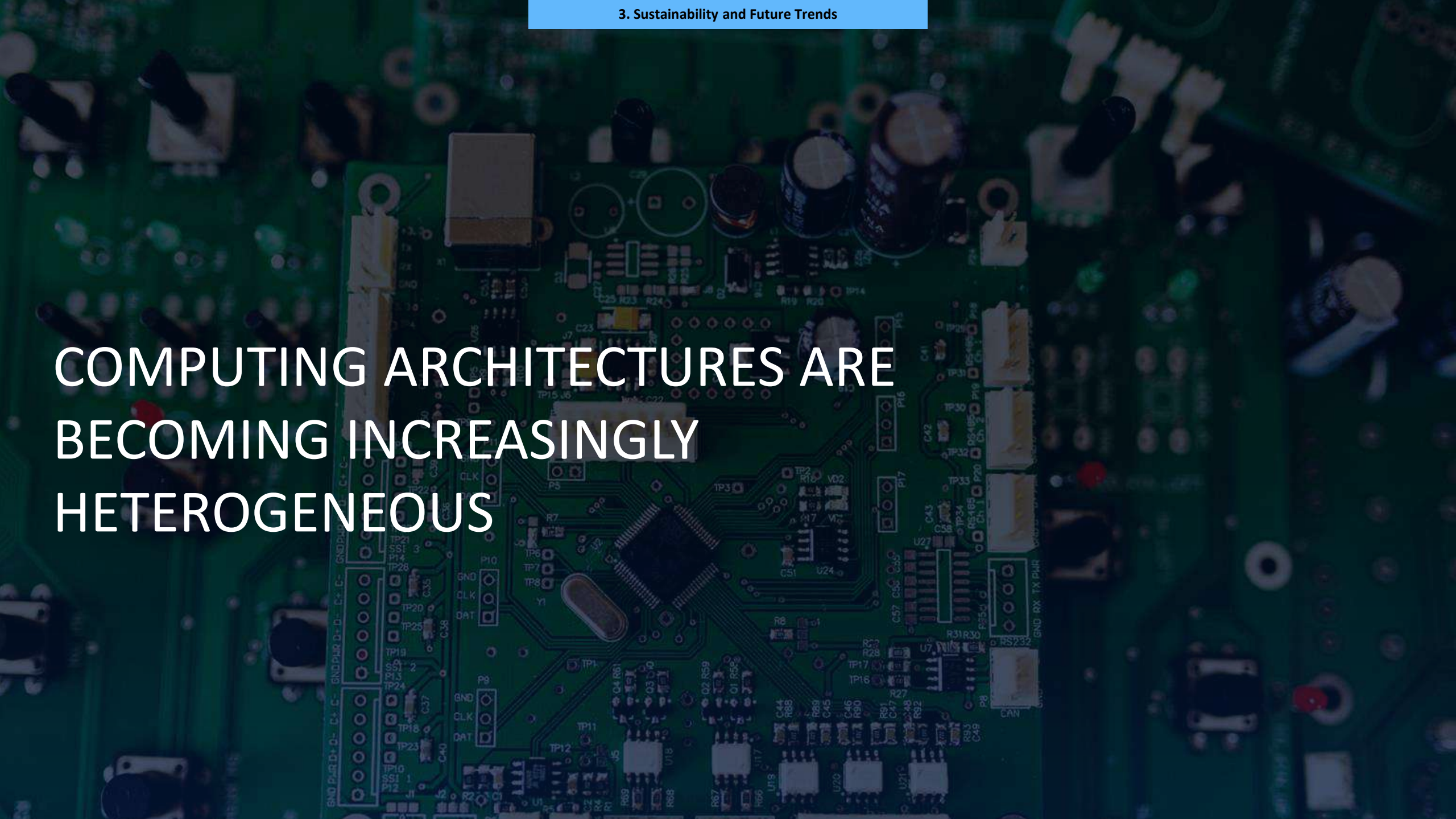
SUSTAINABILITY AND FUTURE TRENDS

Challenges for Exascale and Post Exascale

WAVES OF EFFICIENCY WILL COME FROM A TIGHTER INTEGRATION OF TECHNOLOGIES FROM ALGORITHMS TO HOUSING INFRASTRUCTURE



- Stagnation of general-purpose computing
- Increase in domain-specific computing optimizations to counteract the end of Dennard scaling
- Objective: obtain the best performance-cost-energy tradeoffs for defined computing tasks

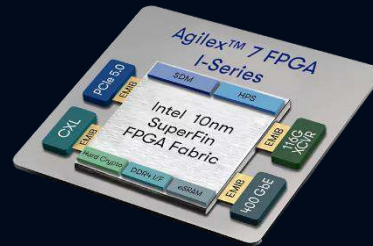


COMPUTING ARCHITECTURES ARE
BECOMING INCREASINGLY
HETEROGENEOUS

HETEROGENEITY IN HPC



Intel Xe GPU



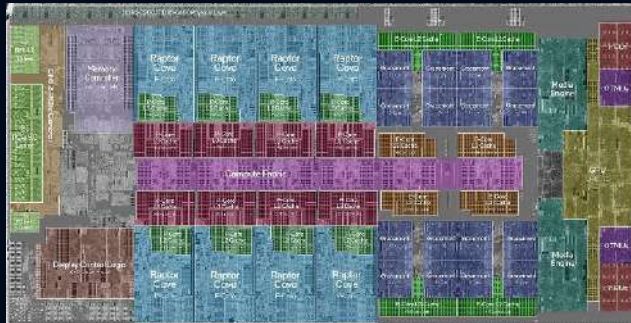
Intel Agilex FPGA



RISC-V



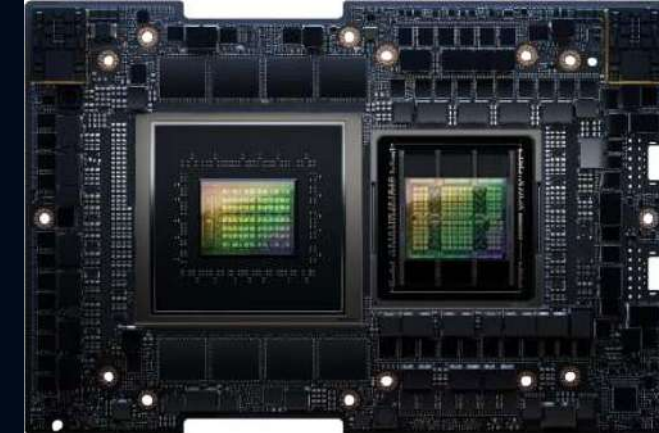
ARM Cortex CPU



Intel Raptor Lake CPU Hybrid Architecture with 16 efficiency cores and 8 performance cores + integrated GPU



AMD EPYC with 192 Zen5 cores



NVIDIA Grace Hopper

Microprocessors: the heart of computing

COMMON PATTERN EVOLUTION

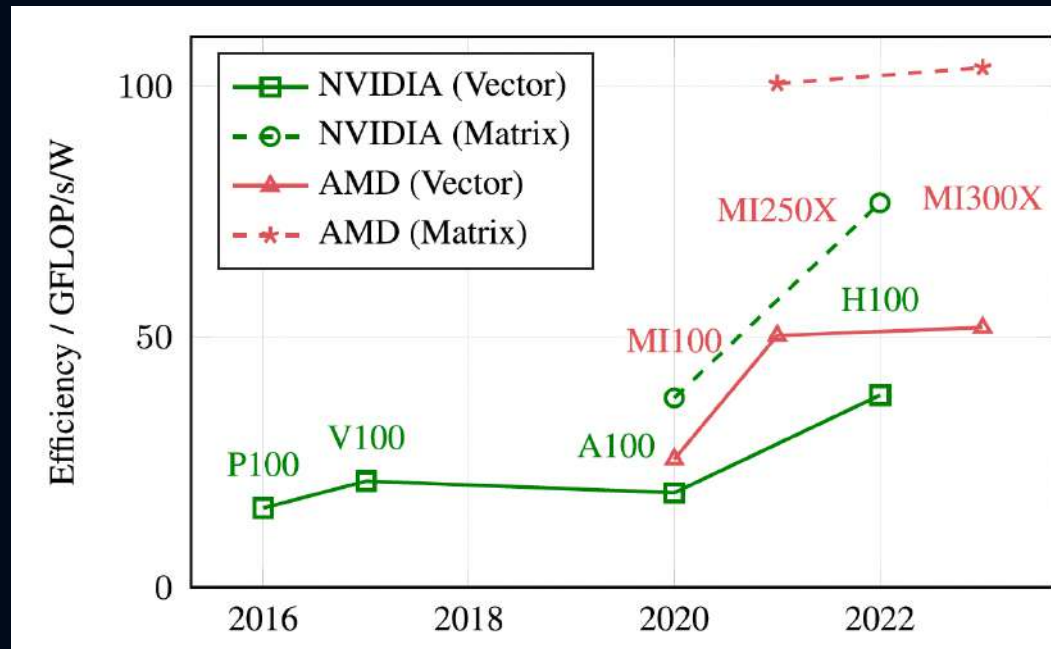
Specialist → Generalist → Specializing Generalist

1. Specialist technology working in limited conditions
2. Generalist technology leading to widespread adoption
3. Further specialization for various problem domains

Future systems will be increasingly heterogeneous, integrating GPUs, CPUs, FPGAs, and a wide range of domain-specific accelerators.

Specialized processors performing tasks suited to their architecture are more efficient than general-purpose ones

SPECIALIZATION ENABLES INCREASES IN ENERGY EFFICIENCY



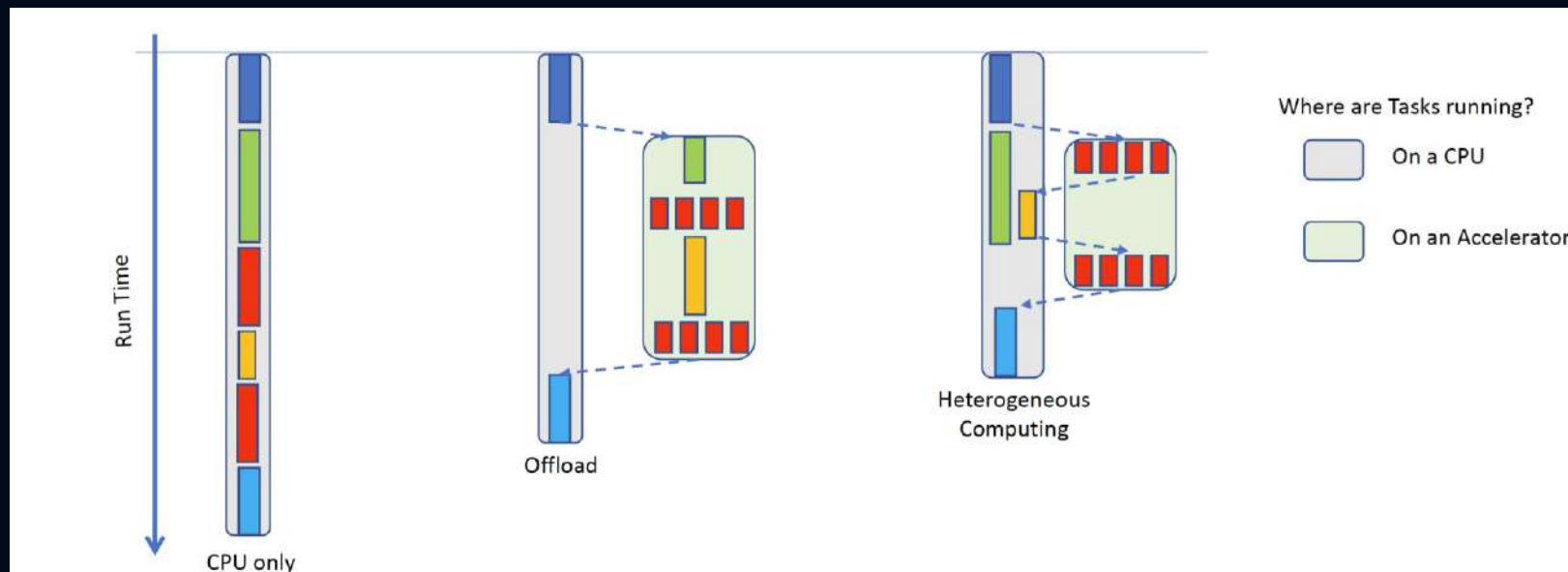
High-Performance Linpack (HPL) energy efficiency

(Matrix) → tensor cores for dedicated acceleration of matrix multiplication

- **Efficiency plateau:** Over last decade, raw GPU energy efficiency barely improved once basic data-parallelism was exploited
- **Power of specialization:** Adding tensor cores yields the largest jumps in FLOPS per watt (more than CPU or general-GPU tweaks)

OFFLOAD VS. HETEROGENEOUS COMPUTING

- Offload: The CPU moves work to an accelerator and waits for the answer.
- Heterogeneous Computing: Run sub-problems in parallel on the hardware best suited to them.





GROWING HETEROGENEITY OF BOTH
COMPUTE HARDWARE AND ENVIRONMENTS

COMPUTE IS MUCH EASIER TO MOVE THAN DATA

ORCHESTRATION OF NEW COMPLEX WORKFLOWS

With heterogenous computing and data processing environments

Live Orchestration



Currently in conflict with current resource management practices on supercomputers.

Computing Continuum

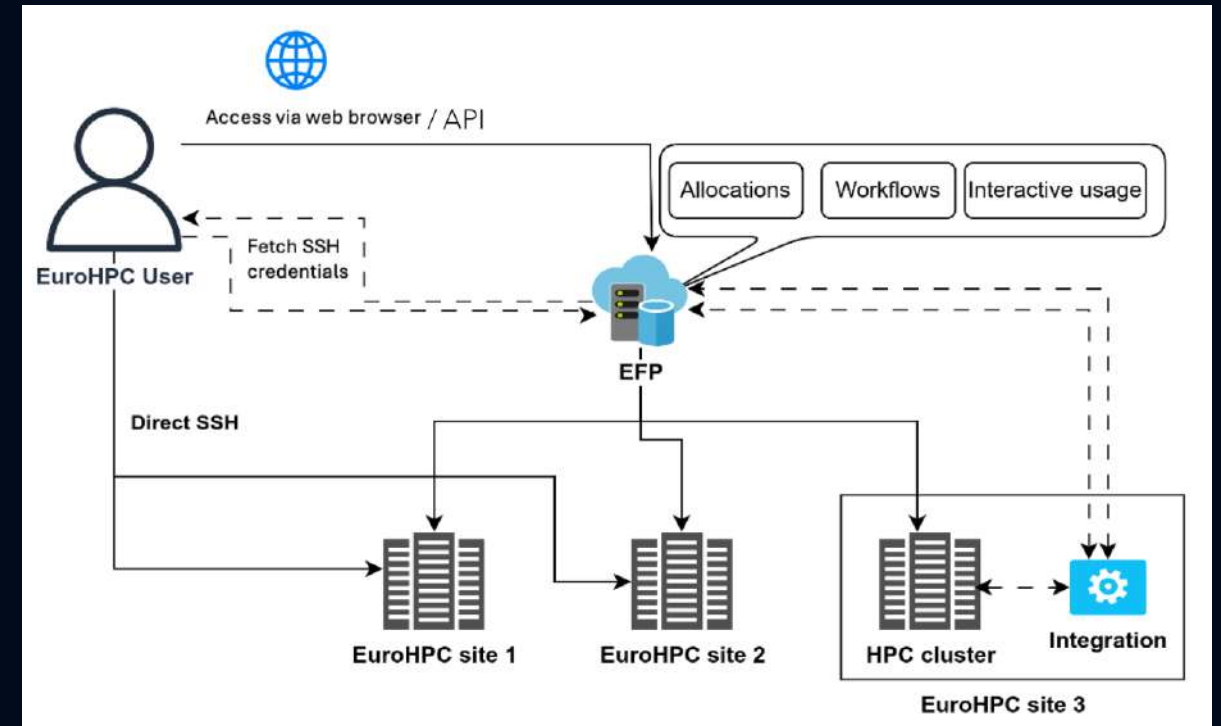
- HPC: one element in a continuum of computing, in charge of simulating very large tasks (e.g., digital twin of the Earth) but fed with **data coming from the real world**
- Many implications for HPC technologies:
 - e.g. computing loads with **hard or soft real-time requirements** for obtaining results or streaming processing “on the fly”,
 - communication of data in and out with all the **security** related aspects,
 - authentication of the systems connected to the HPC engine, etc.

Computing and data processing environment where edge devices, heterogeneous nodes and both cloud and HPC resources are seamlessly integrated

FEDERATION

EuroHPC Federation Platform (EFP)

- Federated identity and Single-Sign-On (SSO)
- Resource allocation, management and monitoring across systems
- Direct access utilizing SSH certificates.
- Interactive web-based usage
- Federated software catalogue
- Advanced workflows and data transfer



IS THE FUTURE OF LARGE-SCALE SYSTEMS IN THE CLOUD?

WILL LOW-LATENCY NETWORKS + OPTIMIZED SOFTWARE BRIDGE THE GAP?

HPC

Execution agent: Processes
Memory: Distributed memory, local memory owned by individual processes
Typical Execution Pattern: SPMD

Chip-to-chip optical networks
reduce latency and increase
bandwidth

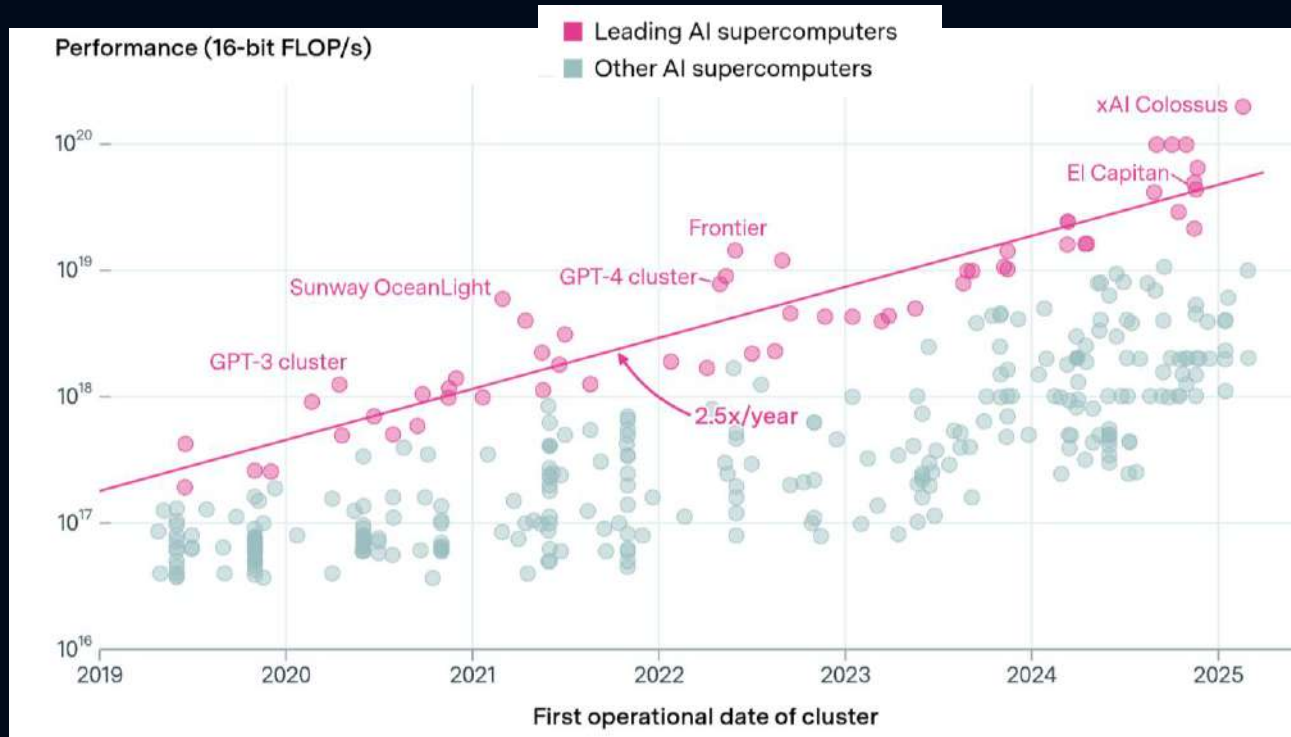
CLOUD

Execution agent: Microservices
Memory: Distributed object store (in memory) backed by a persistent storage system
Typical Execution Pattern: Event driven tasks, Faas, and Actors

Data Streaming Accelerators and IPUs reduce **latency** and jitter (faster, more predictable cloud performance for real-time and AI workloads)

TRENDS IN AI SUPERCOMPUTERS

THE PERFORMANCE OF LEADING AI SUPERCOMPUTERS



- Double in performance “every 9 months”
- Cost billions of dollars
- Require as much power as mid-sized cities
- Companies now own 80% of all AI supercomputers, while governments’ share has declined.

EPOCH AI, "Trends in AI Supercomputers AI supercomputers double in performance every 9 months, cost billions", 2025, <https://epoch.ai/blog/trends-in-ai-supercomputers>

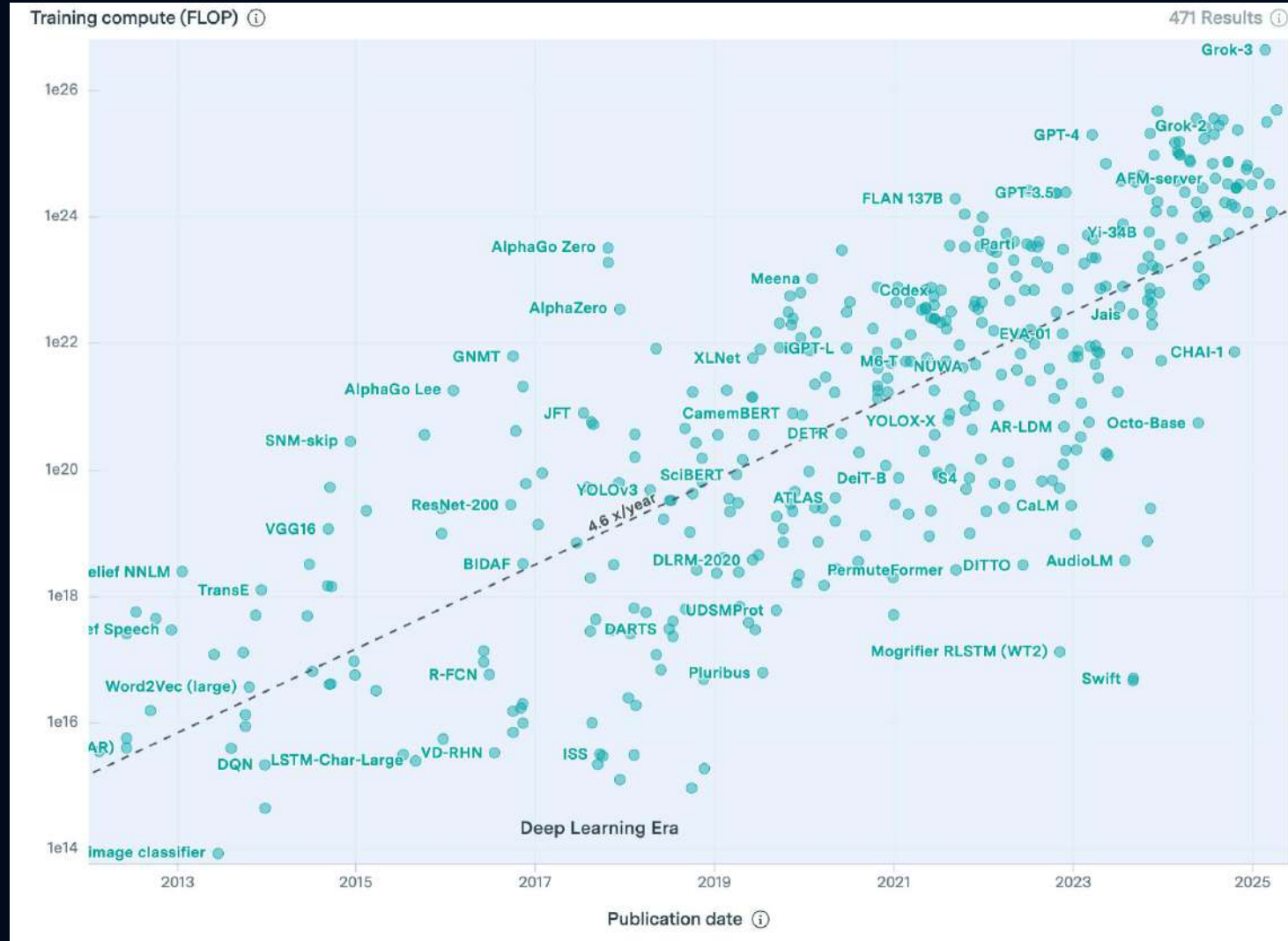
XAI Brings Colossus Online, https://www.perplexity.ai/page/xai-brings-colossus-online-f6EQHsQ_S.egPXbSN17CmQ



NOTABLE AI MODELS

Scaling “Spending” Law

- 2012: AlexNet
 - \$500 GPU x 2 = \$1K
 - 5 days of training
- 2022: GPT-4 (speculation)
 - \$8K GPU x 25,000 = \$200M
 - 90 days of training



IN THE CURRENT SETUP ANY GROWTH BEYOND MOORE'S LAW IS ACHIEVED ONLY BY SPENDING MORE

Rising energy consumption

Year	OOMs	# of H100s-equivalent	Cost	Power	Power reference class
2022	~GPT-4 cluster	~10k	~\$500M	~10 MW	~10,000 average homes
~2024	+1 OOM	~100k	\$billions	~100MW	~100,000 homes
~2026	+2 OOMs	~1M	\$10s of billions	~1 GW	The Hoover Dam, or a large nuclear reactor
~2028	+3 OOMs	~10M	\$100s of billions	~10 GW	A small/medium US state
~2030	+4 OOMs	~100M	\$1T+	~100GW	>20% of US electricity production

REDUCE ENERGY CONSUMPTION AND CONSTRAINS

Justify use of energy and natural resources

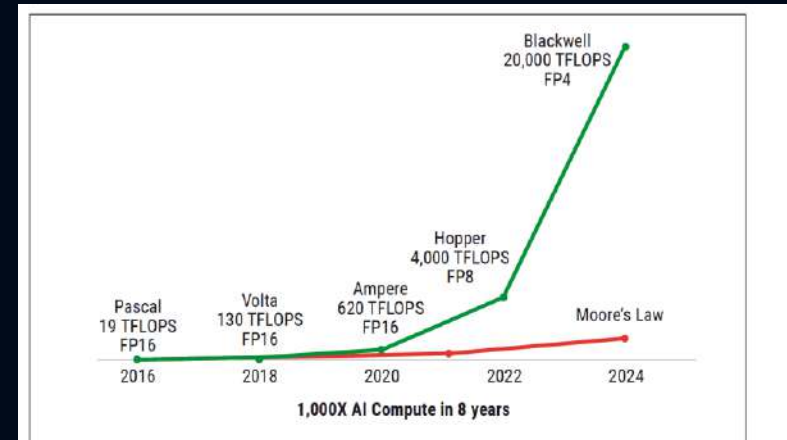
Decentralized training



To overcome power constraints, companies may increasingly use decentralized training approaches, which would allow them to distribute a training run across AI supercomputers in several locations.

...

Lower precision arithmetic



Performance improvement of NVIDIA GPUs on AI workloads
(source NVIDIA, J. Huang keynote at Computex 2024)

Twin advantages of higher floating point operation rates and reduced data movement and energy consumption

And many more solutions ...

TECHNOLOGY DEPENDENCY

TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY (TSMC) FABRICATES MOST OF THE HIGH-END SEMICONDUCTORS ON WHICH THE WORLD DEPENDS.

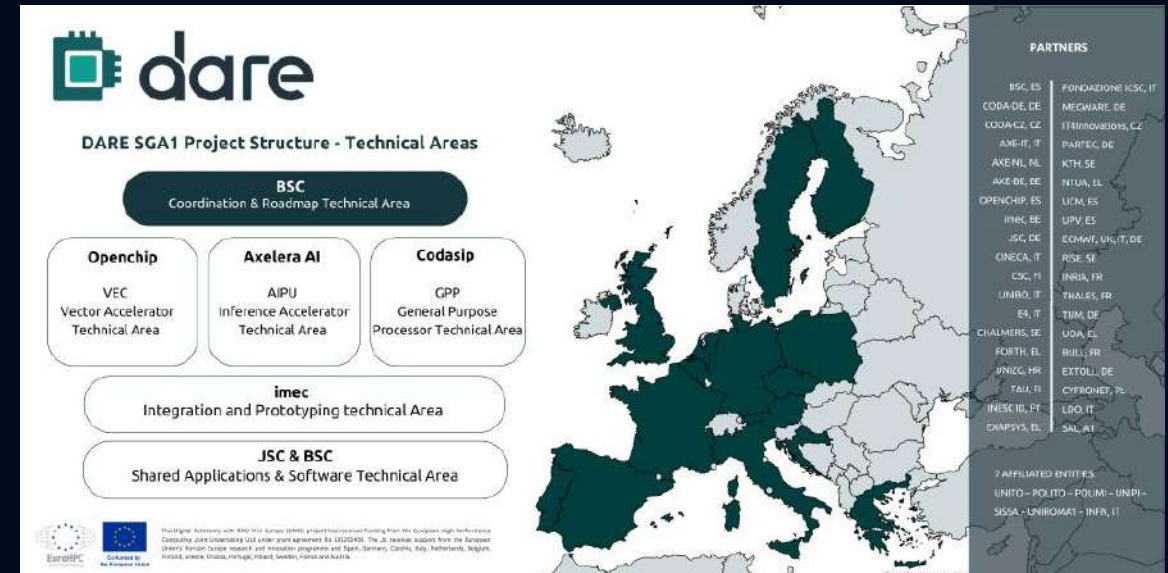
- Today's computing innovation is increasingly driven by
 - the needs of large-scale cloud and AI providers,
 - balanced against the technical and financial challenges of a post-Moore's Law semiconductor era,
 - all unfolding within a shifting geopolitical landscape.

- Amazon, Apple, Google, Microsoft, Meta, and OpenAI are flying high
 - the Chinese DeepSeek AI shock showed there is no monopoly on AI innovation though.

BUILDING A SELF-SUFFICIENT EUROPEAN CHIP MANUFACTURING AND PROCESSOR-DESIGN ECOSYSTEM

Plan for the upcoming years

- In 2018, the European Processor Initiative was launched with the aim of developing a designed-in-Europe Arm-based central processing unit (CPU) and RISC-V-based accelerators.
- DARE: major international follow-up project
 - Budget of EUR 240 million to develop a full HPC ecosystem based on open RISC-V processors (general purpose and accelerators, including AI-specific chips) and their integration into exascale and post-exascale European supercomputers
- Objective: strengthen the EU's strategic technological sovereignty, producing competitive HPC technology to power future European supercomputers





CONCLUSION

CONCLUSION

- **Supercomputing is interdisciplinary** (spanning several research domains)
 - System architectures, hardware components, system software & management, communication and network, programming environment, IO & storage, mathematics and algorithms, ecosystem technologies
- **Dennard scaling has ended**, and the returns from Moore's-Law transistor scaling are diminishing
 - Chip-level performance gains are slowing
 - Shrinking semiconductor feature sizes and increasing transistor complexity (planar → FinFET → GAA) are driving up the cost of leading-edge foundry construction
- To **achieve higher performance**, it is necessary to **scale up the entire system**
 - Major cloud providers are spending tens of billions of dollars annually on enormous data centers packed with AI accelerators that consume gigawatts of power,
 - Creating construction and operating costs that traditional HPC cannot match

CONCLUSION

- Main **thematic trends in supercomputing**
 - Federation, quantum+HPC, energy efficiency & sustainability, AI & foundational models
- **Hardware complexity is growing**: how do we increase performance?
 - Programs must expose greater parallelism and locality for the hardware to exploit
 - More **co-design between users and HPC** is needed: software engineers should collaborate with hardware architects so that new processors provide the abstractions required to use the hardware easily
- Continued algorithmic **research in mixed- and low-precision arithmetic** is essential
 - The gap between the sizes of the computational-science and AI markets is widening
 - Future hardware may prioritize lower precision (detriment of traditional modelling and simulation applications)

CONCLUSION

- **The line between HPC and AI is fading**
 - AI offers new opportunities in computational modelling via efficient, learned surrogate models, while HPC insights increasingly guide AI
- **Deeper overlaps between cloud and HPC are likely in the future**
 - Networking technology (optical interconnects and new topologies) is advancing rapidly
 - With low latency, high bandwidth, and stable performance, cloud platforms can now support loosely synchronous and even synchronous applications
- **HPC applications must evolve** to handle reliability issues and network inhomogeneities

CONCLUSION

- Cloud hyperscalers and the economics of generative AI are becoming dominant

DATE	LEADING AI SUPERCOMPUTER	PERFORMANCE (16-BIT FLOP/s)	H100-EQ [†]	NUMBER OF AI CHIPS	POWER	HARDWARE COST (2025 USD)
MARCH 2025	XAI COLOSSUS	1.98×10^{20}	200K	200K	300MW	\$7B
JUNE 2026	Extrapolated	5×10^{20}	500K	300K	600MW	\$14B
JUNE 2027	Extrapolated	1×10^{21}	1M	500K	1GW	\$25B
JUNE 2028	Extrapolated	3×10^{21}	3M	800K	2GW	\$50B
JUNE 2029	Extrapolated	8×10^{21}	8M	1.3M	5GW	\$100B
JUNE 2030	Extrapolated	2×10^{22}	20M	2M	9GW	\$200B

“...9 GW equals the output of nine nuclear reactors (a scale beyond any existing industrial facility)...”

K, F. Pilz, J. Sanders, R. Rahman, L. Heim, "Trends in AI Supercomputers", in arXiv:2504.16026, 2025, <https://doi.org/10.48550/arXiv.2504.16026>

- Quantum and biologically inspired computing (e.g., neuromorphic), along with semi-custom chiplet-based architectures, include new possibilities and may better match scientific and engineering workloads
 - At present, however, these technologies are immature and unlikely to rival today’s silicon-based semiconductors in the near term
- Nature demonstrates that **human-level intelligence** need not consume megawatts; a **few dozen watts** suffices

SOURCES AND AKNOWLEDGEMENT

- Sources:

- European Technology Platform for High-Performance Computing (ETP4HPC), “Strategic Research Agenda 6” , 2025, <https://etp4hpc.eu/strategic-research-agenda/>
- High Performance, Edge And Cloud computing (HiPEAC), “HiPEAC Vision 2025”, <https://vision.hipeac.net/>
- All other references reported in the presentation

- Acknowledgments:

- Colleagues at the Jülich Supercomputing Centre
- Insights drawn from previous presentations by Tim Mattson (University of Bristol / Merly.ai) and Daniel A. Reed (University of Utah)

- Disclaimer:

- All data and materials have been selected, reworked, and personally interpreted by me.
- Any errors or omissions are my sole responsibility.

THANK YOU FOR YOUR ATTENTION

