Introduction to HPC

Lecture 1: INTRODUCTION TO HPC

“High Performance Computing module ”

DATA SCIENCE & ARTIFICIAL INTELLIGENCE

SCIENTIFIC & DATA-INTENSIVE COMPUTING

First Section

- 03.03 Introduction to HPC
- 03.10 HPC Hardware and software

Agenda

- Prologue: why and where HPC ?
- What is HPC ?
- Performance and metrics
- Supercomputers and TOP500
- Parallel computers

Before starting: HPC prefix..

Factor	Name	Symbol
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k

- How large is your HD on your laptop ?
- How large is your RAM?
- How powerful is your CPU in your laptop ?
- How large is the L1 cache of your CPU ?
- What is the CPU frequency of your laptop?

A first citation

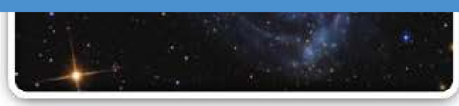
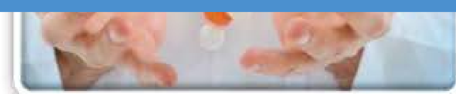
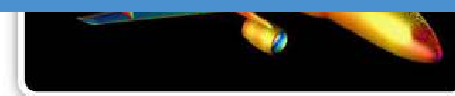
- "CRUCIAL PROBLEMS that we can only hope to address computationally REQUIRE US TO DELIVER EFFECTIVE COMPUTING POWER ORDERS-OF-MAGNITUDE GREATER THAN WE CAN DEPLOY TODAY."

DOE's *Office of Science*, *2012*

..even more valid nowadays, though

What do they have in common ?

THEY ALL NEED COMPUTATIONAL POWER AND USE HIGH PERFORMANCE COMPUTING TO DELIVER BETTER RESULTS FASTER

Pricing shares and winning trading by millisecond arbitrage

Forecasting an hurricane and its impact with increased precision

Finding oil under a salt crust saving billions in exploration and drilling



HPC not easy to define..

High performance computing (HPC), also known as supercomputing, refers to computing systems with extremely high computational power that are able to solve hugely complex and demanding problems.

[Taken from https://ec.europa.eu/digital-single-market/en/high-performance-computing]

Complex problem 1:

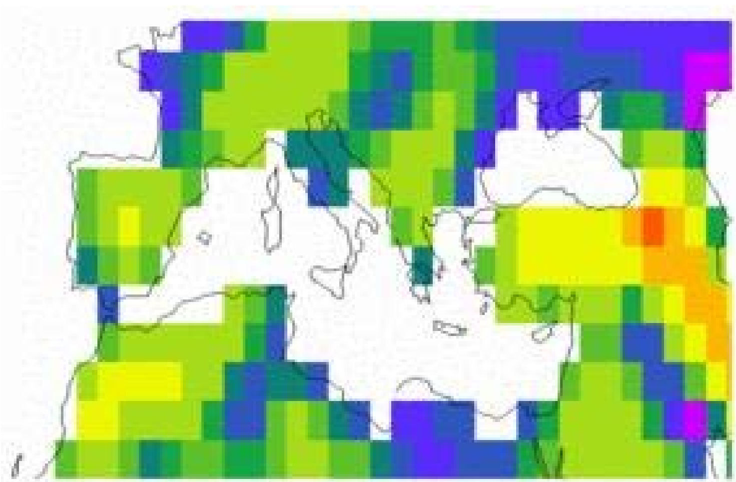
Weather forecast..

Recipe:

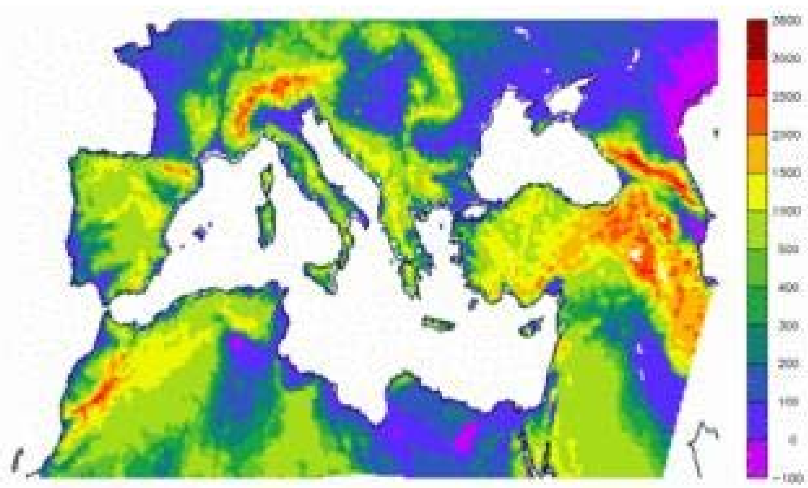
- Define a mathematical model to describe the problem
- Solve it computationally
  - Discretization over a 3d grid
  - Integrate equations
  - Check results..

Complex problem 1 : climate change over the Mediterranean sea

- What are the requirements in term of RAM to have decent results ?



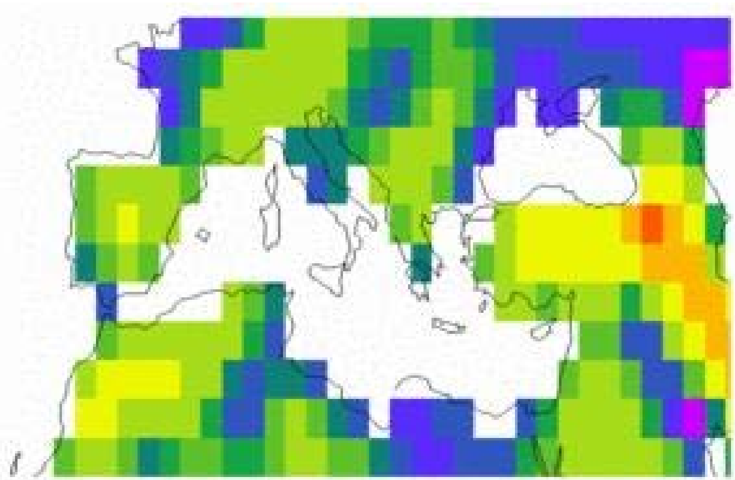
200 km



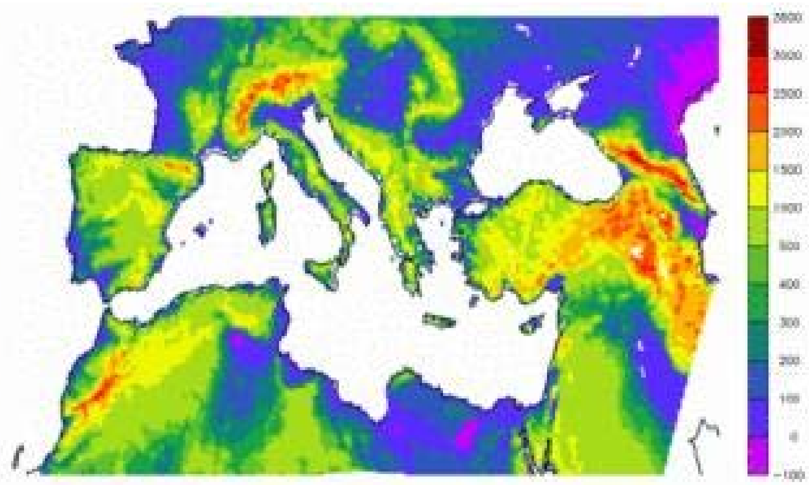
25 km

Complex problem: climate change over the Mediterranean sea

- Resolution:
  - 200km -> ~ 1GB 2km -> ? GB



200 km



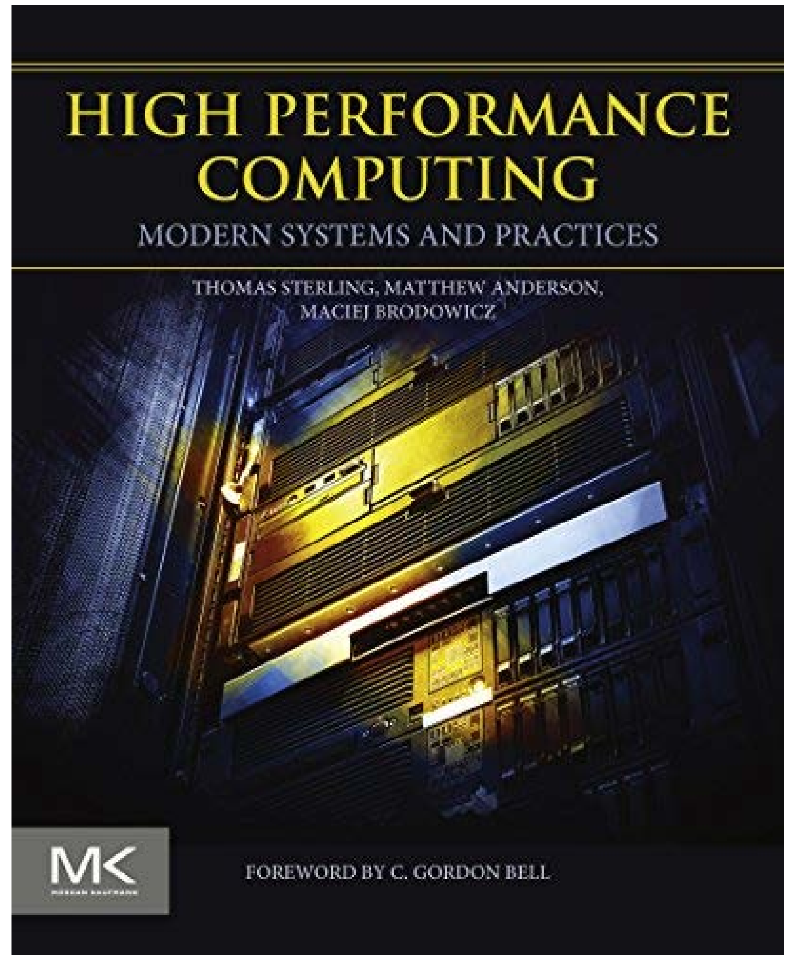
25 km

Complex problems solved by simulations

- Simulation has become the way to research and develop new scientific and engineering solutions.
- Used nowadays in leading science domains like aerospace industry, astrophysics, etc.
- Challenges related to the complexity, scalability and data production of the simulators arise.
- Impact on the relaying IT infrastructure.

Interested in more example ?

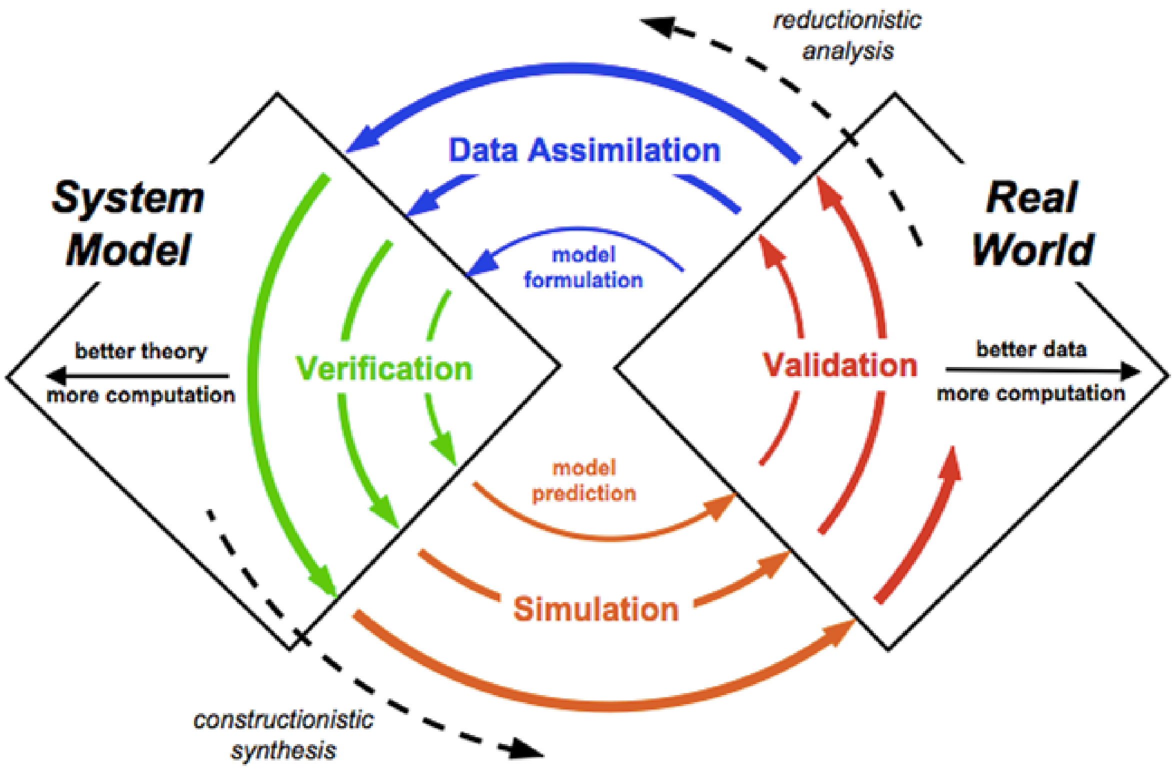
- See chapter one section 1.2 of reference 4
- Look around on the internet..



Research is changing..

- Inference Spiral of System Science

  As models become more complex and new data bring in more information, we require ever increasing computational power



Data are flooding us..

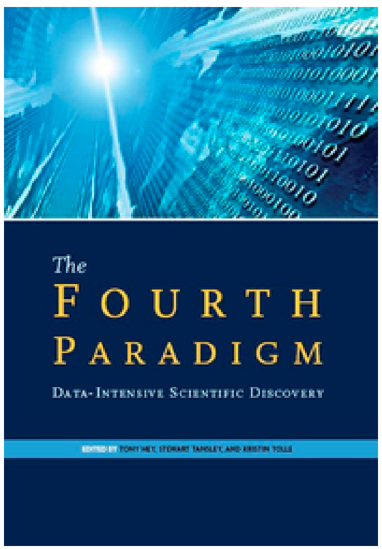
In today's world, larger and larger amounts of data are constantly being generated, from 33 zettabytes globally in 2018 to an expected 181 zettabytes in 2025). As a result, the nature of computing is changing, with an increasing number of data-intensive critical applications. HPC is key to processing and analysing this growing volume of data, and to making the most of it for the benefit of citizens, businesses, researchers and public administrations.

[Taken again from https://ec.europa.eu/digital-single-market/en/high-performance-computing]

Data intensive science

The Fourth Paradigm: Data-Intensive Scientific Discovery

Presenting the first broad look at the rapidly emerging field of data-intensive science



Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets.

The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies.

In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, the collection of essays expands on the vision of pioneering computer scientist Jim Gray for a new, fourth paradigm of discovery based on data-intensive science and offers insights into how it can be fully realized.

Critical praise for *The Fourth Paradigm*

Download

- Full text, low resolution (6 MB)
- Full text, high resolution (93 MB)
- By chapter and essay

Purchase from Amazon.com

- Paperback
- Kindle version

In the news

- Sailing on an Ocean of 0s and 1s *(Science* Magazine)
- A Deluge of Data Shapes a New Era in Computing *(New York Times)*
- A Guide to the Day of Big Data *(Nature)*

Big data challenge: from HPC to HPDA through AI

- Organizations are expanding their definitions of high-performance computing (HPC) to include workloads such as artificial intelligence (AI) and high-performance data analytics (HPDA) in addition to traditional HPC simulation and modeling workloads.

From https://insidebigdata.com/2019/07/22/converged-hpc-clusters/

Complex problem 2: ChatGPT…

TECH

ChatGPT and generative AI are booming, but the costs can be extraordinary

PUBLISHED MON, MAR 13 2023•8:58 AM EDT | UPDATED MON, APR 17 2023•2:09 AM EDT

Jonathan Vanian
@JONATHANVANIAN

Kif Leswing
@KIFLESWING

WATCH LIVE

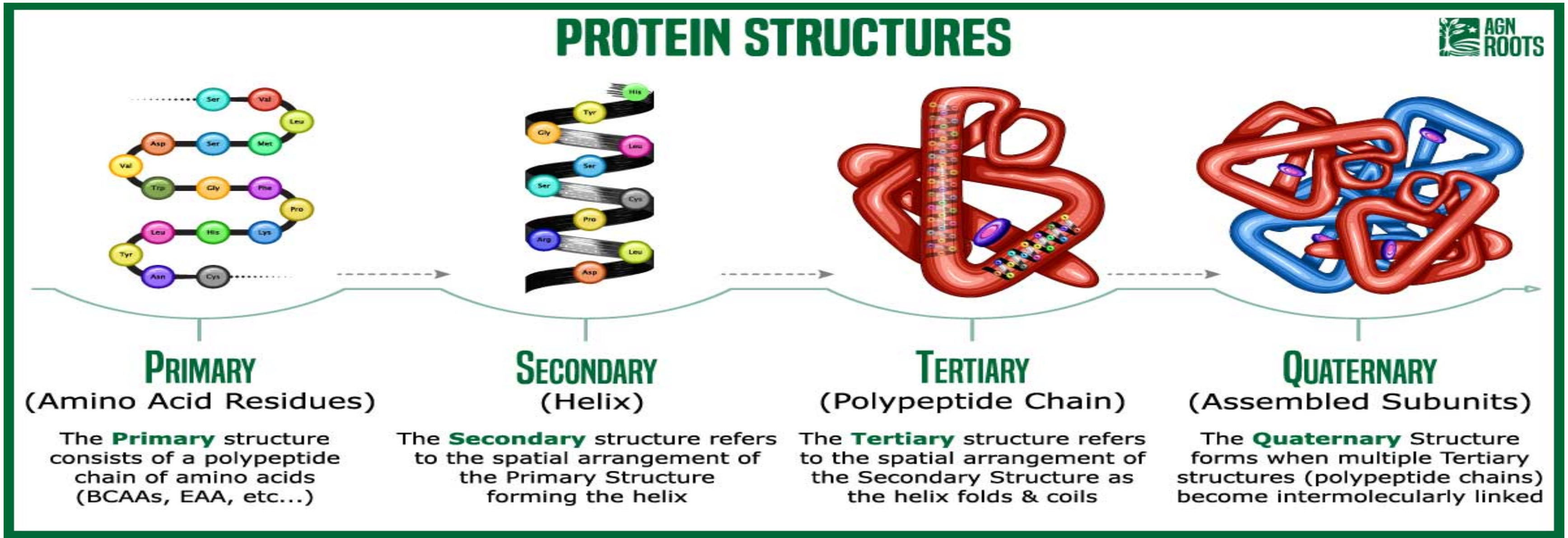
KEY POINTS

- The cost to develop and maintain the software can be extraordinarily high.
- Nvidia makes most of the GPUs for the AI industry, and its primary data center workhorse chip costs $10,000.
- Analysts and technologists estimate that the critical process of training a large language model such as GPT-3 could cost over $4 million.

Complex problem 2
Protein folding..

Determine the e structure of the protein from its aminoacid sequence

- Solve it computationally..
    A. Run very long MD simulation on each sequence and wait..
    B. Predict new structures training an AI algorithm..



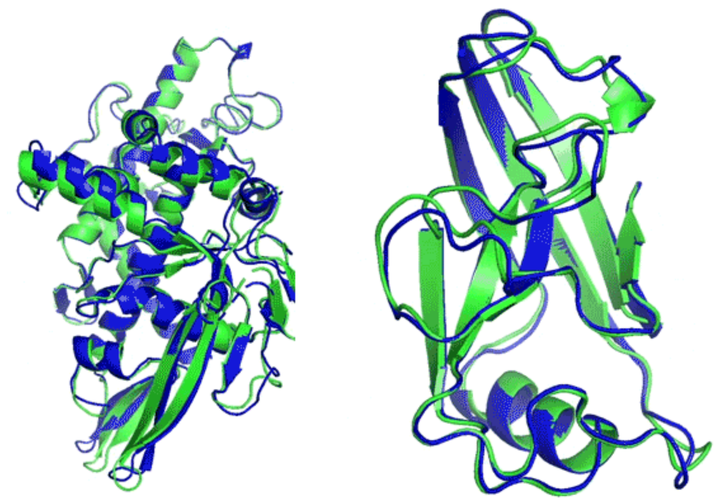
Approach A: Folding@home

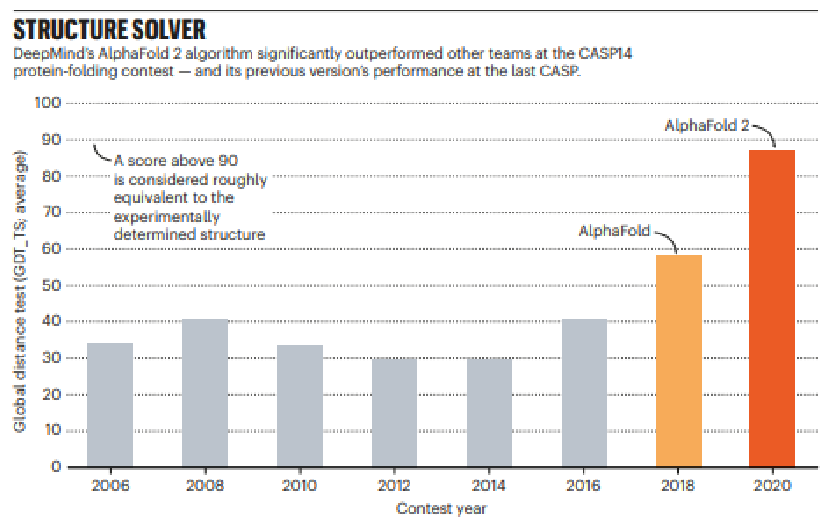
- A more than 20 years project
- It allows everybody to run an MD simulation contributing to the problem..
- Impressive distributed computational power..
- From statistics page:

OS	GPU	CPU	Tflops GPU	Tflops x86
Windows	2440+ 12864	31417	27504	55243
Linux	98 + 19257	34794	36561	72456
Total	2538+ 32121	73018	64429	128063

Approach B: AlphaFold

- [AlphaFold](#) is an AI system developed by [DeepMind](#) that predicts a protein's 3D structure from its amino acid sequence.
- Presented in 2018 and 2020 at CASP outperformed all other approaches
- Scores obtained are roughly equivalent to experimentally determined structured





AlphaFold DB

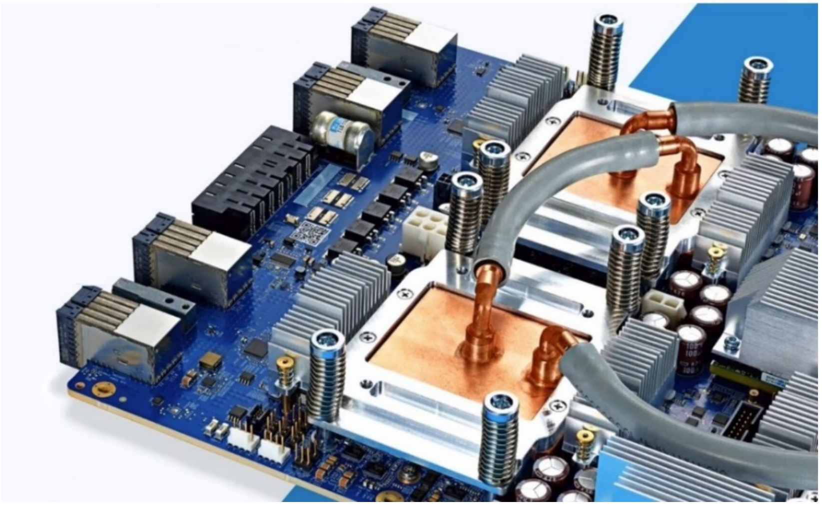
AlphaFold DB release in July 2022 open access to over 200 million protein structure predictions to accelerate scientific research.

Full dataset download for AlphaFold Database - UniProt (214M):

- The full dataset of all predictions is available at no cost and under a CC-BY-4.0 licence .. from Google Cloud Public Datasets. We've grouped We suggest that you only download the full dataset if you need to process all the data with local computing resources (the size of the dataset is 23 TiB, ~1M tar files).

How much computing power does AlphaFold use ?

“We train the model on Tensor Processing Unit (TPU)* v3 with a batch size of 1 per TPU core, hence the model uses 128 TPU v3 cores…
The initial training stage takes approximately 1 week, and the fine-tuning stage takes approximately 4 additional days.”



* A Tensor Processing Unit (TPU) is an application specific integrated circuit (ASIC) developed by Google to accelerate machine learning. Google offers TPUs on demand, as a cloud deep learning service called Cloud TPU.

FastFold vs. AlphaFold

FastFold: Reducing AlphaFold Training Time from 11 Days to 67 Hours

We successfully scaled the AlphaFold model training to 512 NVIDIA A100 GPUs and obtained aggregate 6.02 PetaFLOPs at the training stage. The overall training time is reduced to 67 hours from 11 days with significant economic cost savings.

TABLE IV
RESOURCE AND TIME COST COMPARE.

Implementation	Framework	Training Process	Hardward	Step Time (s)	Training Time (days)	Resource
AlphaFold	JAX [18]	Initial training Fine-tuning	128 × TPUv3	/ /	11	33792 TPU hours
OpenFold	PyTorch	Initial training Fine-tuning	128 × A100	6.186 20.657	8.39	25774 GPU hours
FastFold	PyTorch	Initial training Fine-tuning	256 × A100 512 × A100	2.487 4.153	**2.81**	**20738 GPU hours**

The need of a different approach..

- *However, as AI development outpaces hardware development, leading to skyrocketing training costs, adopting cost-saving solutions has become imperative.*

[taken from https://arxiv.org/pdf/2408.14158v2](https://arxiv.org/pdf/2408.14158v2)

Fire-Flyer AI-HPC: A Cost-Effective
Software-Hardware Co-Design for Deep Learning

Wei An, Xiao Bi, Guanting Chen, Shanhuang Chen,Chengqi Deng, Honghui Ding, Kai Dong, Qiushi Du, Wenjun Gao, Kang Guan, Jianzhong Guo, Yongqiang Guo, Zhe Fu, Ying He, Panpan Huang, Jiashi Li, Wenfeng Liang, Xiaodong Liu, Xin Liu, Yiyuan Liu, Yuxuan Liu, Shanghao Lu, Xuan Lu, Xiaotao Nie, Tian Pei, Junjie Qiu, Hui Qu, Zehui Ren, Zhangli Sha, Xuecheng Su, Xiaowen Sun, Yixuan Tan, Minghui Tang, Shiyu Wang, Yaohui Wang, Yongji Wang, Ziwei Xie, Yiliang Xiong, Yanhong Xu, Shengfeng Ye, Shuiping Yu, Yukun Zha, Liyue Zhang*, Haowei Zhang, Mingchuan Zhang, Wentao Zhang, Yichao Zhang, Chenggang Zhao, Yao Zhao, Shangyan Zhou, Shunfeng Zhou, Yuheng Zou

DeepSeek-AI, Beijing, China
research@deepseek.com

Agenda

- ~~Prologue: why and where HPC ?~~
- What is HPC ?
- Performance and metrics ?
- Supercomputers and TOP500
- Parallel Computers

HPC: a ~~first~~ second definition

High Performance Computing (HPC) is the use of servers, clusters, and supercomputers – plus associated software, tools, components, storage, and services – for scientific, engineering, or AI tasks that are particularly intensive in computation, memory usage, or data management

HPC is used by scientists and engineers both in research and in production across industry, government and academia.

[to be continued]

Elements of the HPC ecosystem..

- use of servers, clusters, and supercomputers

  → HARDWARE
- associated software, tools, components, storage, and services

  → SOFTWARE
- scientific, engineering, AI tasks

  → PROBLEMS TO BE SOLVED..

A list of HPC items

COMPUTATIONAL SERVERS

ACCELERATORS

HIGH SPEED NETWORKS

HIGH END PARALLEL STORAGE

IS ALL THIS ENOUGH ?

MIDDLEWARE

SCIENTIFIC/TECHNICAL/ DATA ANALYSIS SOFTWARE

RESEARCH/TECHNICAL DATA

PROBLEMS TO BE SOLVED

Last but not least: people

- Human capital is by far the most important aspect
- Two important roles:
  - HPC providers
    - plan/install/manage HPC resources
  - HPC user :
    - use at best HPC resource

MIXING/INTERPLAYING ROLES INCREASES COMPETENCE LEVELS

Agenda

- ~~Prologue: why and where HPC ?~~ ✓
- ~~What is HPC ?~~ ✓
- Performance and metrics
- Supercomputers and TOP500
- Parallel Computers

It is all about Performance

- It is difficult to define Performance properly

  “speed” / “how fast” are vague terms
- Performance as a measure again ambiguous and not clearly defined and in its interpretation
- In any case performance it is at core to HPC as a discipline
- Let discuss it in some details

Does P stand just for Performance ?

- Performance is not always what matters..

to reflect a greater focus on the productivity, rather than just the performance, of large-scale computing systems, many believe that HPC should now stand for High Productivity Computing. [from wikipedia]

Performance vs Productivity

- A possible definition:
  - Productivity = (application performance) / (application programming effort)
- Example:
  - To speed up a code by a factor of two it takes 6 months work
  - does this deserve to be done ?
- people in HPC arena have different goals in mind thus different expectations and different definitions of productivity.

How do measure (basic) performance of HPC systems

- How fast can I crunch numbers on my CPUs ?
- How fast can I move data around ?
  - from CPUs to memory
  - from CPUs to disk
  - from CPUs on different machines
- How much data can I store ?

Number crunching on CPU: what do we count ?

- Rate of [million/billions of] floating point operations per second ([M|G]flops) FLOPs/S
- Theoretical peak performance:
  - determined by counting the number of floating-point additions and multiplications that can be completed during a period of time, usually the cycle time of the machine

FLOPS=clock_rate*Number_of_FP_operation*Number_of_cores

Sustained (peak) performance

- Real (sustained) performance: a measure

FLOPS= (total number of floating point operations done by a program) / (time the program takes to run in second)

- Number_of_floating_point_operations not easy to be defined for real application
- benchmarks are available for that..
- Top500 list uses HPL Linpack:
  - Sustained peak performance is what's matter in TOP500

Agenda

- ~~Prologue: why and where HPC ?~~ ✓
- ~~What is HPC ?~~ ✓
- ~~Performance and metrics~~ ✓
- Supercomputers and TOP500
- Parallel Computers

TOP 500 List

- The TOP500 list www.top500.org
- published twice a year from 1993
  - ISC conference in Europe (June)
  - Supercomputing conference in USA (November)
- List the most powerful computers in the world
- yardstick: Linpack benchmark (HPL)

HPL: some details

- From http://icl.cs.utk.edu/hpl/index.html:
  - The code solves a uniformly random system of linear equations and reports time and floating-point execution rate using a standard formula for operation count.
  - Number_of_floating_point_operations = $2/3n^3 + 2n^2$ (n=size of the system)

```
================================================================================
T/V                N    NB     P     Q               Time                 Gflops
--------------------------------------------------------------------------------
WR03R2L2       86000  1024     2     1             191.06              2.219e+03
--------------------------------------------------------------------------------
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)=        0.0043644 ...... PASSED
```

HPL&TOP 500 List

- For each machine the following numbers are reported using HPL:
  - Rmax: the performance in GFLOPS for the largest problem run on a machine.
  - Rpeak: the theoretical peak performance GFLOPS for the machine.
  - The measure of the power required to run the benchmark

And the winner is..

EL CAPITAN - HPE CRAY EX255A, AMD 4TH GEN EPYC 24C 1.8GHZ, AMD INSTINCT MI300A, SLINGSHOT-11, TOSS

Site:	DOE/NNSA/LLNL
Manufacturer:	HPE
Cores:	11,039,616
Processor:	AMD 4th Gen EPYC 24C 1.8GHz
Interconnect:	Slingshot-11
Installation Year:	2024

And the winner is..

Performance	
Linpack Performance (Rmax)	1,742.00 PFlop/s
Theoretical Peak (Rpeak)	2,746.38 PFlop/s
Nmax	25,446,528
Power Consumption	
Power:	29,580.98 kW
Power Measurement Level:	2
Software	
Operating System:	TOSS
Compiler:	g++ 12.2.1 and hipcc 6.2.0
Math Library:	AMD rocBLAS 6.0.2 and Intel MKL 2016
MPI:	HPE Cray MPI

11/2024 Highlights (from www.top500.org)

- **11/2024 Highlights**

- The 64th edition of the TOP500 reveals that **El Capitan** has achieved the top spot and is officially the third system to reach exascale computing after Frontier and Aurora
- The new **El Capitan** system at the **Lawrence Livermore National Laboratory** in California, U.S.A., has debuted as the most powerful system on the list with an HPL score of 1.742 EFlop/s. It has 11,039,616 combined CPU and GPU cores and is based on AMD 4th generation EPYC processors with 24 cores at 1.8GHz and AMD Instinct MI300A accelerators. El Capitan relies on a Cray Slingshot 11 network for data transfer and achieves an energy efficiency of 58.89 Gigaflops/watt. This power efficiency rating helped El Capitan achieve No. 18 on the GREEN500 list as well.
- The **Frontier** system at **Oak Ridge National Laboratory** in Tennessee, U.S.A, has moved down to the No. 2 spot. It has increased its HPL score from 1.206 Eflop/s on the last list to 1.353 Eflop/s on this list.
- The **Aurora** system at **Argonne Leadership Computing Facility** in Illinois, U.S.A, has claimed the No. 3 spot on this TOP500 listThe **Eagle** system installed on the **Microsoft Azure Cloud** in the U.S.A. claimed the No. 4 spot and remains the highest-ranked cloud-based system on the TOP500. It has an HPL score of 561.2 PFlop/s
- The only other new system in the TOP 5 is the **HPC6** system at No. 5. This machine is installed at **Eni S.p.A** center in Ferrera Erbognone, Italy and has the same architecture as the No. 2 system Frontier. The HPC6 system at Eni achieved an HPL benchmark of 477.90 PFlop/s and is now the fastest system in Europe.

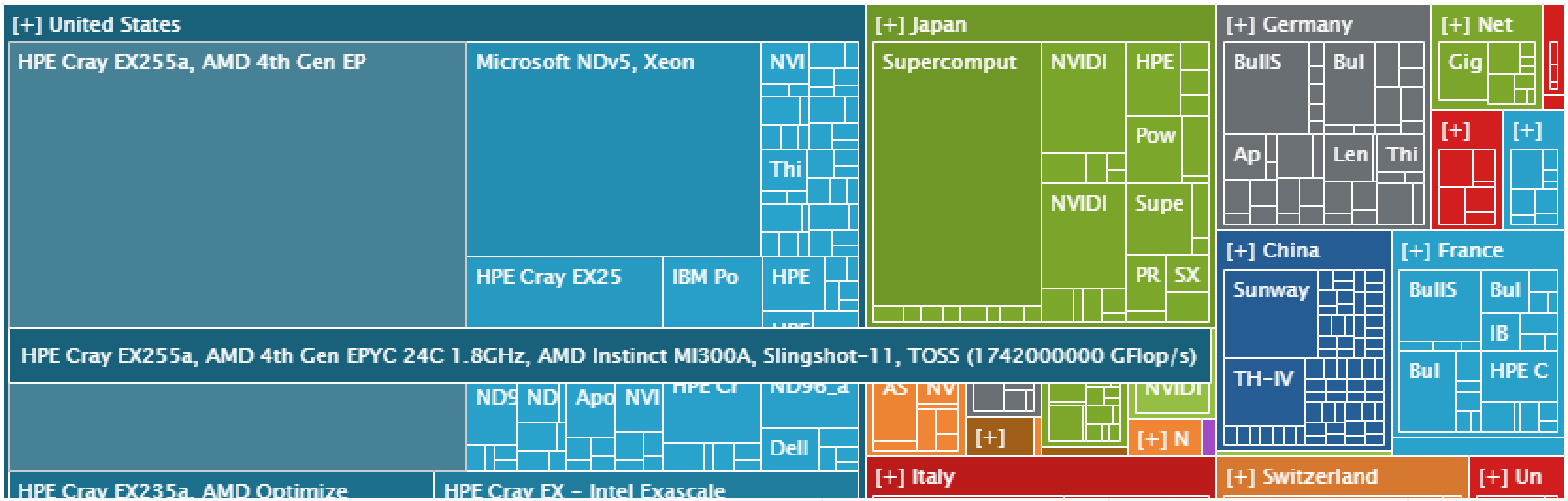
06/2023 Highlights (from www.top500.org)

- **06/2023 Highlights**
- Frontier is the No. 1 system in the TOP500.
- Fugaku, the No. 2 system, is installed at the RIKEN Center for Computational Science (R-CCS) in Kobe, Japan. It has 7,630,848 cores which allowed it to achieve an HPL benchmark score of 442 Pflop/s.
- The LUMI system, another HPE Cray EX system installed at EuroHPC center at CSC in Finland is the No. 3 with a performance of 309.1 Pflop/s. The European High-Performance Computing Joint Undertaking (EuroHPC JU) is pooling European resources to develop top-of-the-range Exascale supercomputers for processing big data. One of the pan-European pre-Exascale supercomputers, LUMI, is located in CSC’s data center in Kajaani, Finland.
- The No. 4 system Leonardo is installed at a different EuroHPC site in CINECA, Italy. It is an Atos BullSequana XH2000 system with Xeon Platinum 8358 32C 2.6GHz as main processors, NVIDIA A100 SXM4 40 GB as accelerators, and Quad-rail NVIDIA HDR100 Infiniband as interconnect. It achieved a Linpack performance of 238.7 Pflop/s.

EU systems from 5th to 9th places..

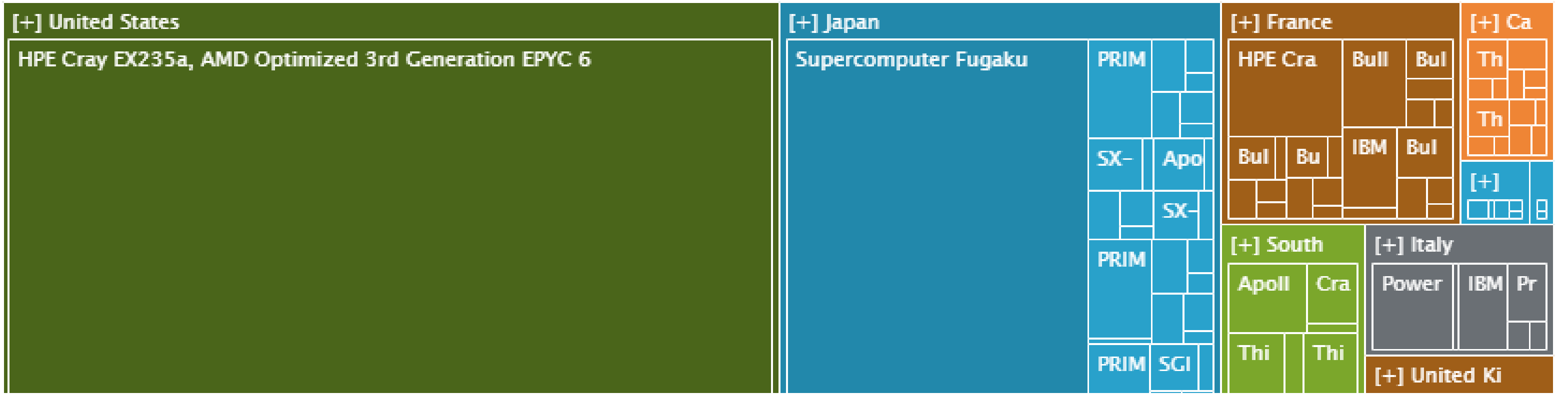
5	Eni S.p.A. Italy	**HPC6** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9 HPE	3,143,520	477.90	606.97	8,461
6	RIKEN Center for Computational Science Japan	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D Fujitsu	7,630,848	442.01	537.21	29,899
7	Swiss National Supercomputing Centre (CSCS) Switzerland	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS HPE	2,121,600	434.90	574.84	7,124
8	EuroHPC/CSC Finland	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11 HPE	2,752,704	379.70	531.51	7,107
9	EuroHPC/CINECA Italy	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband EVIDEN	1,824,768	241.20	306.31	7,494

By country: 11/24



	Countries	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
1	United States	173	34.6	6,477,869,130	10,267,226,568	56,856,656
2	China	63	12.6	319,062,154	548,507,445	20,121,848
3	Germany	40	8	399,727,830	577,160,600	4,736,708
4	Japan	34	6.8	940,710,340	1,249,016,118	11,736,944
5	France	24	4.8	298,085,500	431,753,642	4,500,728
6	Italy	14	2.8	849,088,330	1,097,238,408	7,314,432
7	United Kingdom	14	2.8	84,813,684	142,257,324	1,970,096

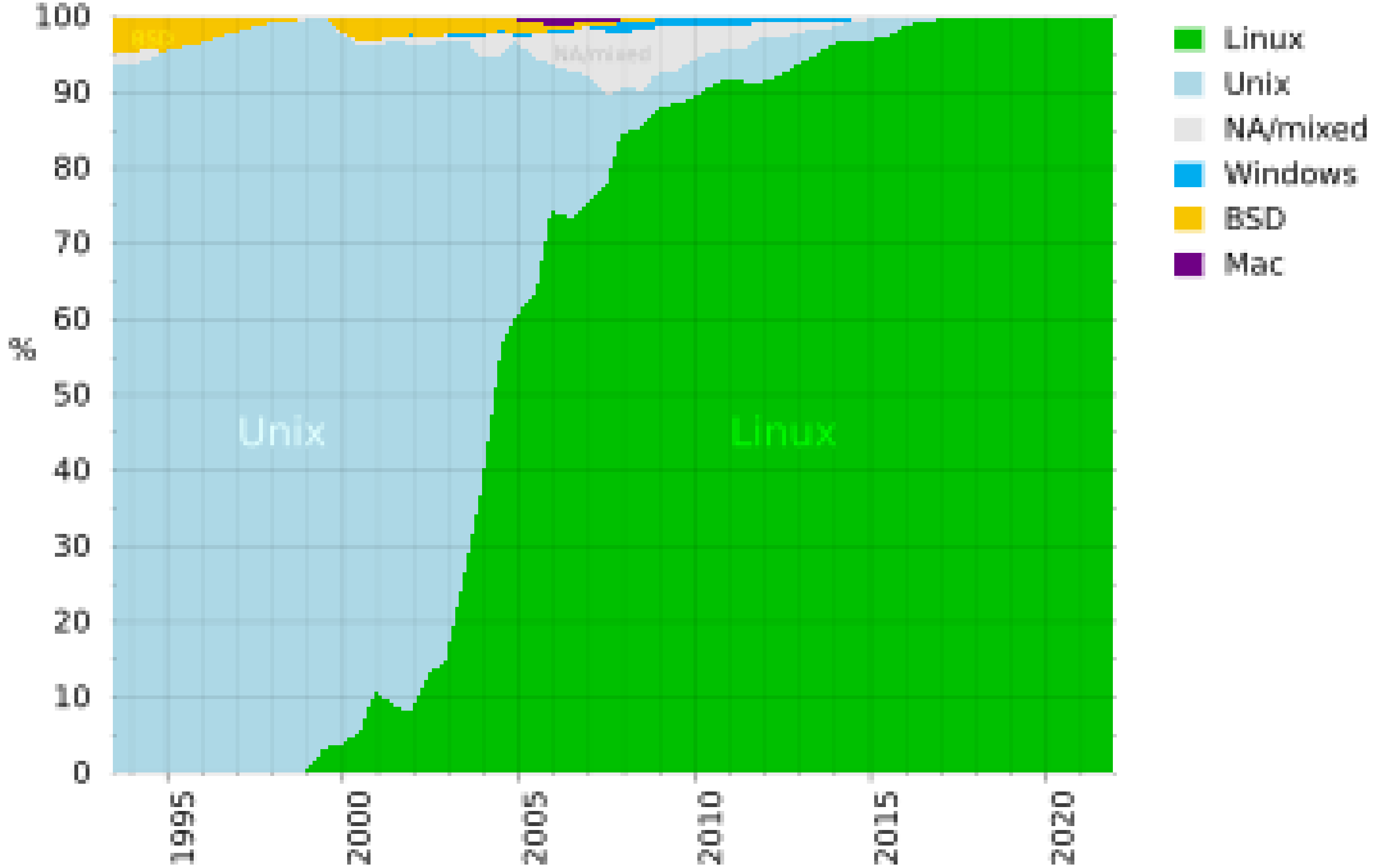
By country: 6/23

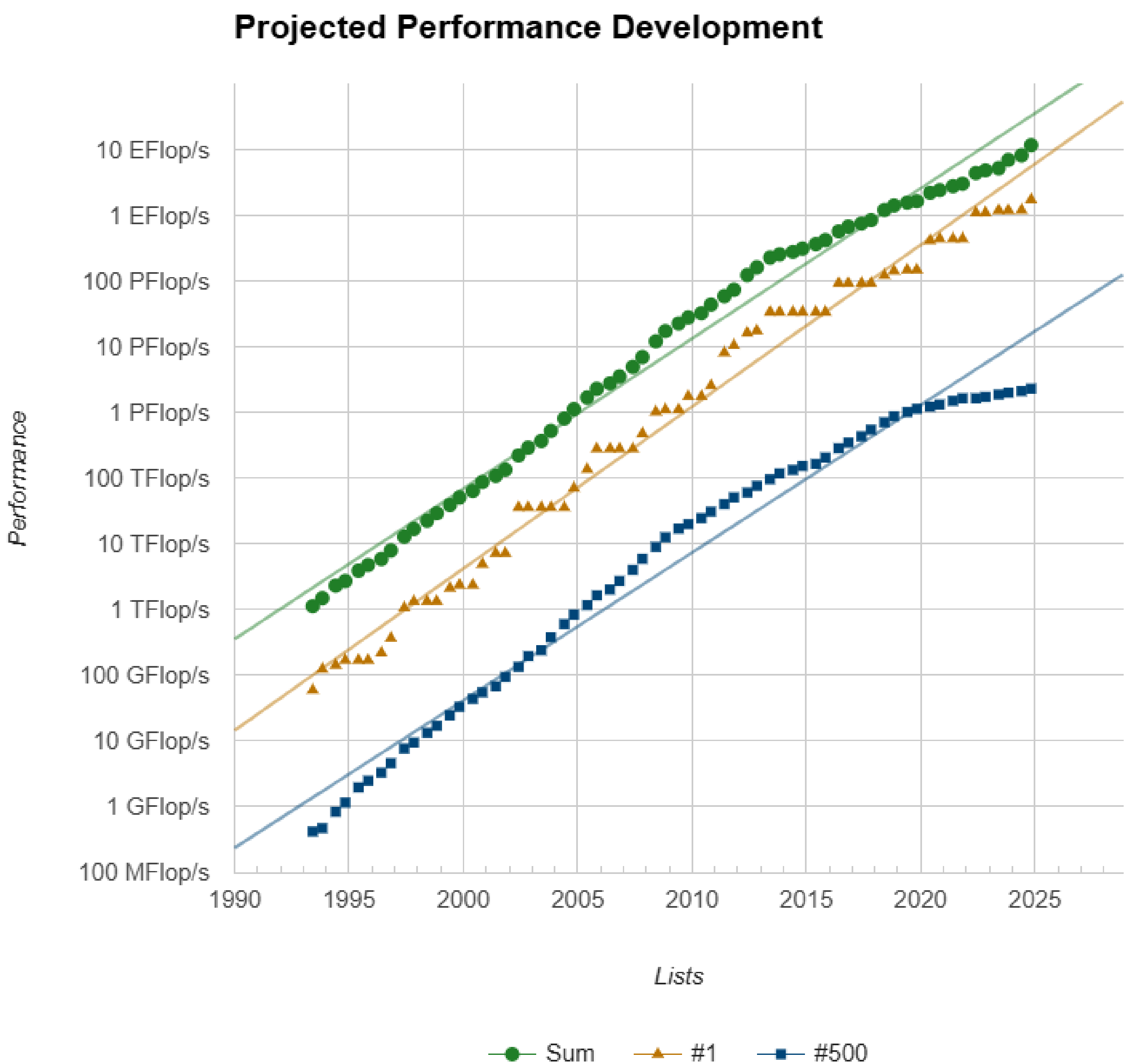


Installations by countries/regions:

		Count	System Share (%)	Rmax (TFlops)	Rpeak (TFlops)	Cores
1	China	173	34.6	530,240	1,158,771	29,413,676
2	United States	128	25.6	2,085,045	3,150,398	27,715,304
3	Japan	33	6.6	626,506	817,353	11,984,068
4	Germany	31	6.2	200,537	306,054	3,896,660
5	France	22	4.4	168,660	242,484	3,874,520
6	Canada	14	2.8	47,805	80,390	1,076,384
7	United Kingdom	12	2.4	57,018	78,629	1,779,888
8	Russia	7	1.4	73,715	101,737	741,328
9	Italy	6	1.2	78,529	114,512	1,447,536

By operating system





Projected Performance Development
10 EFlop/s
1 EFlop/s
100 PFlop/s
10 PFlop/s
1 PFlop/s
100 TFlop/s
10 TFlop/s
1 TFlop/s
100 GFlop/s
10 GFlop/s
1 GFlop/s
100 MFlop/s
Performance
1990
1995
2000
2005
2010
2015
2020
2025
Lists
Sum
#1
#500

Is Top500 ok for AI workload ?

The answer is not really

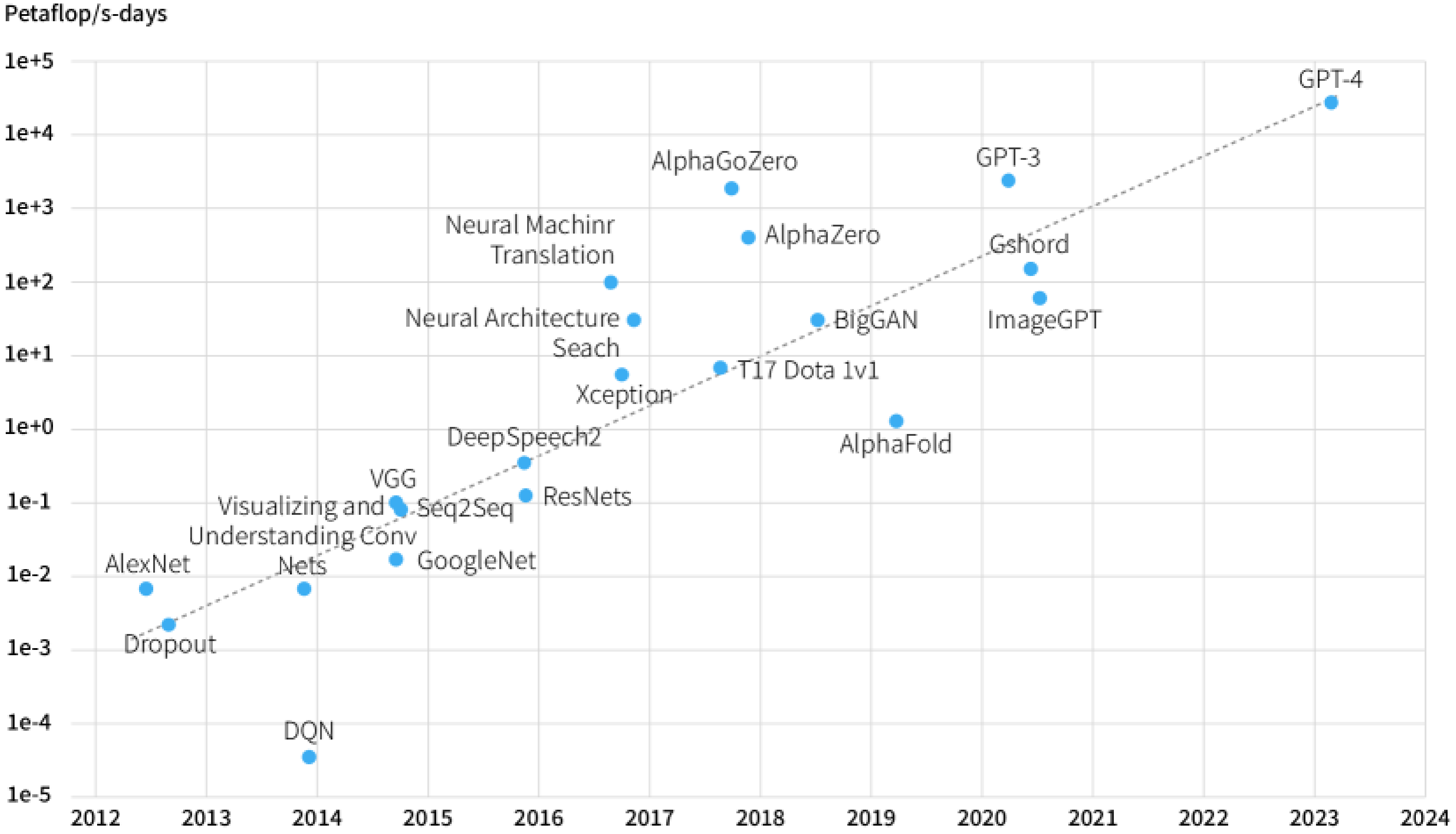
- Floating point in TOP500 is double precision
  - But AI works is single and/or half precision
- Data movement is not consider in TOP500
  - RAM access
  - Disk access

One word on DeepSeek

DeepSeek says that it spent 2.66 million GPU-hours on H800 accelerators to do the pretraining, 119,000 GPU-hours on context extension, and a mere 5,000 GPU-hours for supervised fine-tuning and reinforcement learning on the base V3 model, for a total of 2.79 million GPU-hours. At the cost of $2 per GPU hour – we have no idea if that is actually the prevailing price in China – then it cost a mere $5.58 million to train V3.

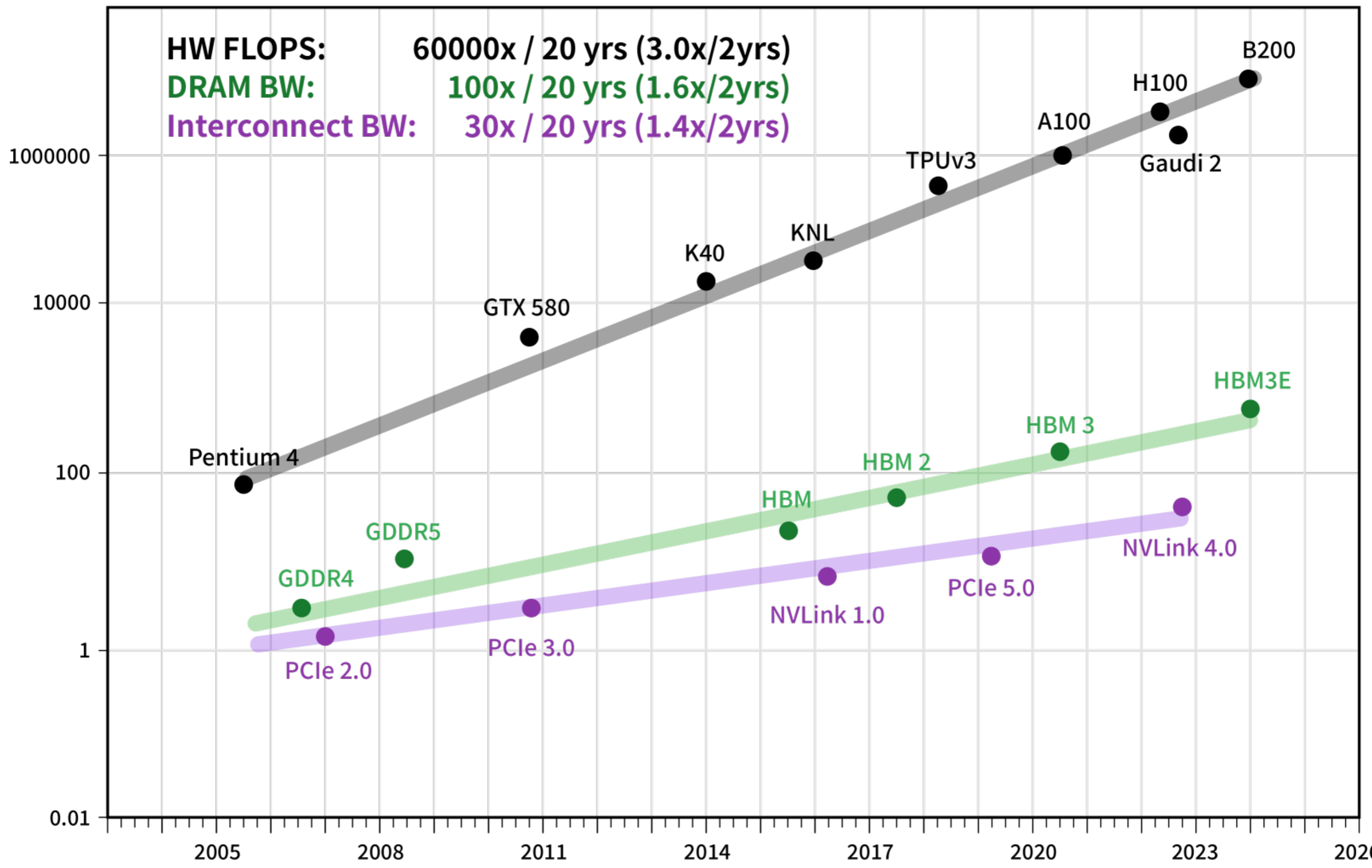
From: How Did DeepSeek Train Its AI Model On A Lot Less – And Crippled – Hardware?

HPC needs for AI..

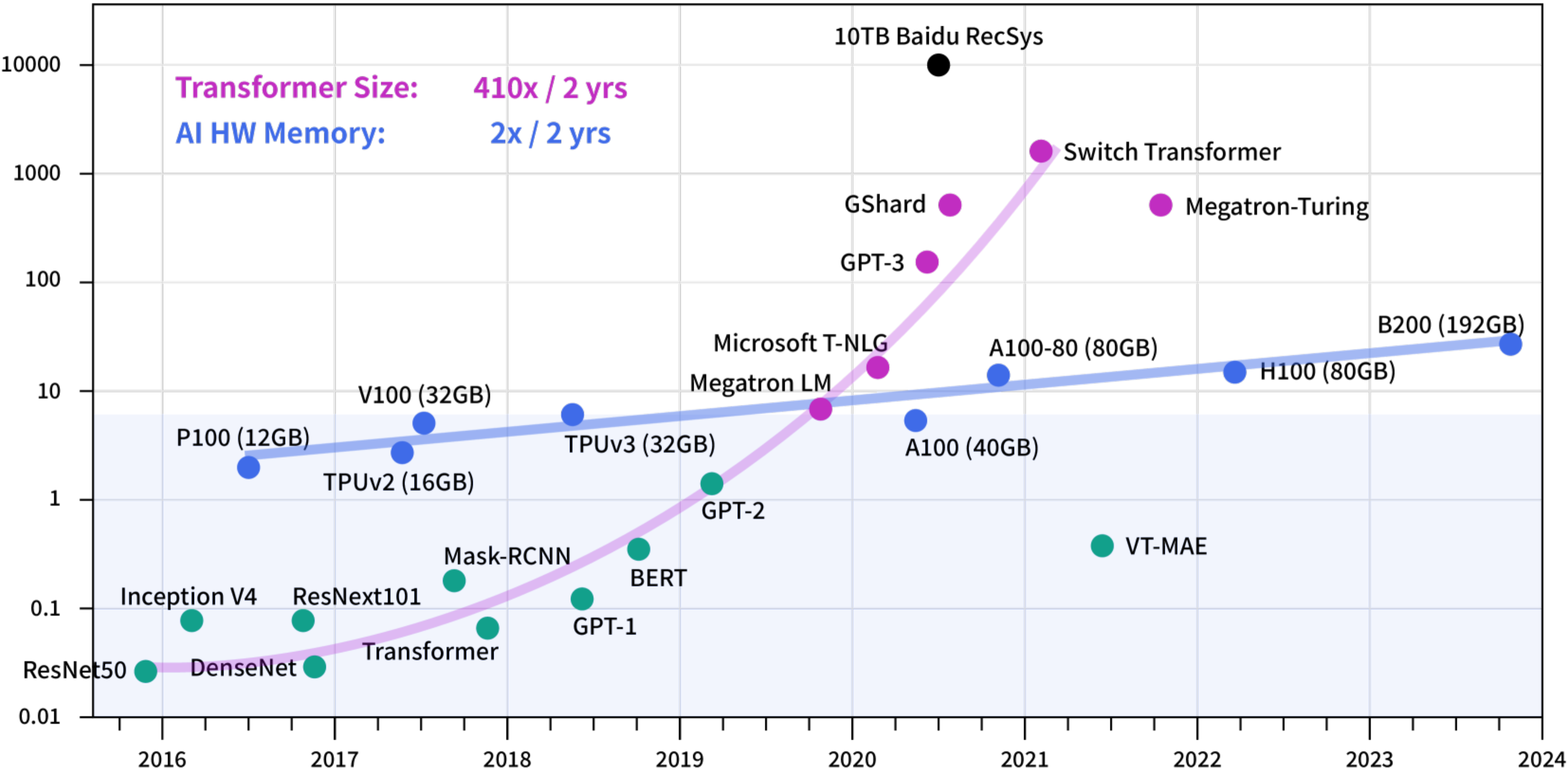


[Taken from https://arxiv.org/pdf/2408.14158v2](https://arxiv.org/pdf/2408.14158v2)

Scaling of Peak Hardware FLOPS, and Memory/Interconnect Bandwidth.



AI needs



10TB Baidu RecSys
Transformer Size: 410x / 2 yrs
AI HW Memory: 2x / 2 yrs
Switch Transformer
GShard
Megatron-Turing
GPT-3
B200 (192GB)
Microsoft T-NLG
A100-80 (80GB)
H100 (80GB)
Megatron LM
V100 (32GB)
P100 (12GB)
TPUv3 (32GB)
A100 (40GB)
TPUv2 (16GB)
GPT-2
VT-MAE
Mask-RCNN
BERT
Inception V4
ResNext101
GPT-1
Transformer
ResNet50
DenseNet
10000
1000
100
10
1
0.1
0.01
2016
2017
2018
2019
2020
2021
2022
2023
2024

Agenda

- ~~Prologue: why and where HPC ?~~ ✓
- ~~What is HPC ?~~ ✓
- ~~Performance and metrics~~ ✓
- ~~Supercomputers and TOP500~~ ✓
- Parallel Computers

- To be continued