Perspective of HPC

challenges in exascale era and beyond

Dr. Gilles Fourestey
Executive Director

EUROfusion Advanced Computing Hub
EPFL

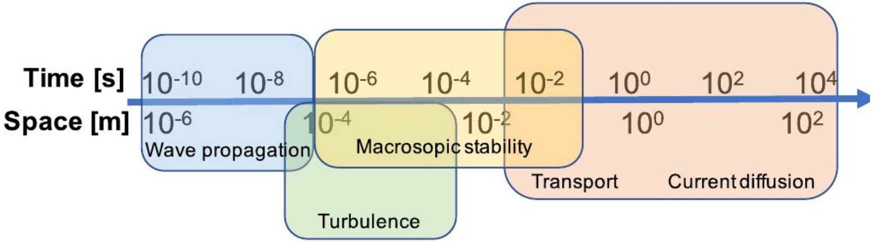
Simulations/HPC to model the hot plasma

- Fusion reactors are **extremely complex to build**
- **Numerical simulations** are an ***essential* tool** to help their design
- But are also **extremely demanding** both in terms of models and resources

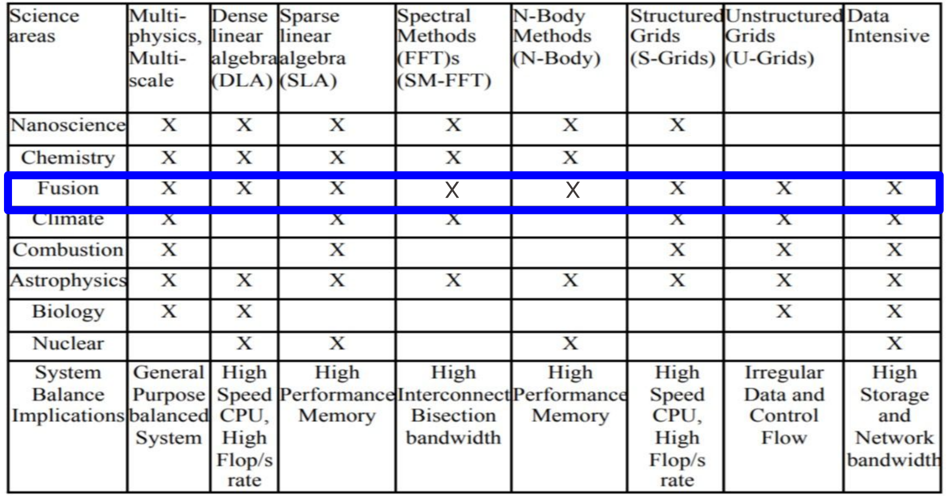
Coupling between different fields…

- Electromagnetic
- Plasma physics kinetic, gyrokinetic,
- Two-fluids,
- MHD models
- Material science plasma-wall interactions
- Wave physics heating systems
- Engineering not included!

with different space/time scales…



… and HPC motifs and their hardware implementations



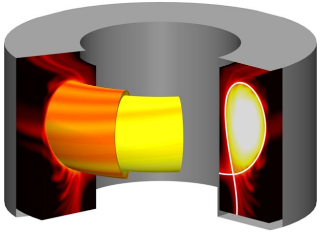
Science areas	Multi-physics, Multi-scale	Dense linear algebra (DLA)	Sparse linear algebra (SLA)	Spectral Methods (FFT)s (SM-FFT)	N-Body Methods (N-Body)	Structured Grids (S-Grids)	Unstructured Grids (U-Grids)	Data Intensive
Nanoscience	X	X	X	X	X	X		
Chemistry	X	X	X	X	X			
Fusion	X	X	X	X	X	X	X	X
Climate	X		X	X		X	X	X
Combustion	X		X			X	X	X
Astrophysics	X	X	X	X	X	X	X	X
Biology	X	X					X	X
Nuclear		X	X		X			X
System Balance Implications	General Purpose balanced System	High Speed CPU, High Flop/s rate	High Performance Memory	High Interconnect Bisection bandwidth	High Performance Memory	High Speed CPU, High Flop/s rate	Irregular Data and Control Flow	High Storage and Network bandwidth

Exascale: more is more...

High-Fidelity Simulations:

- Better capture spatial and temporal scales
- Enhanced resolution
- Improved accuracy
- Multi-Physics simulations

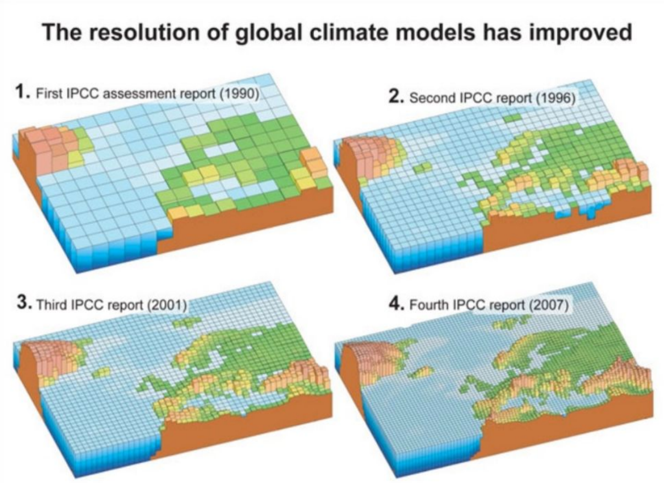
TCV@EPFL
100 Teraflops

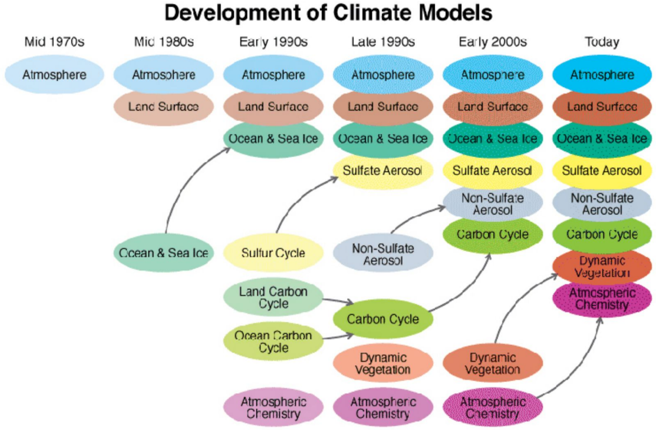


JT-60SA: **100x TCV**
10 Petaflops

ITER: **500x TCV**
100 Petaflops

DEMO: **5000x TCV**
1 Exaflops





Unprecedented level of heterogeneity

- **GPUs** are dominating the **Top500**
- but the **CPU/GPU combo is rarely the same** vendor-wise
- And there is more to come:

intel intel

arm

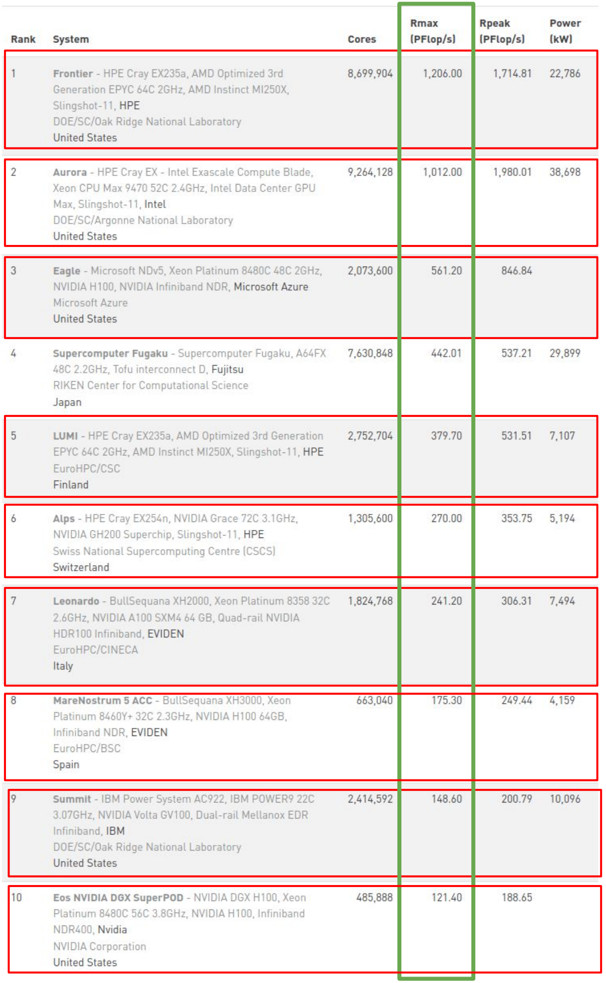
LUMI

AMD AMD

arm NVIDIA

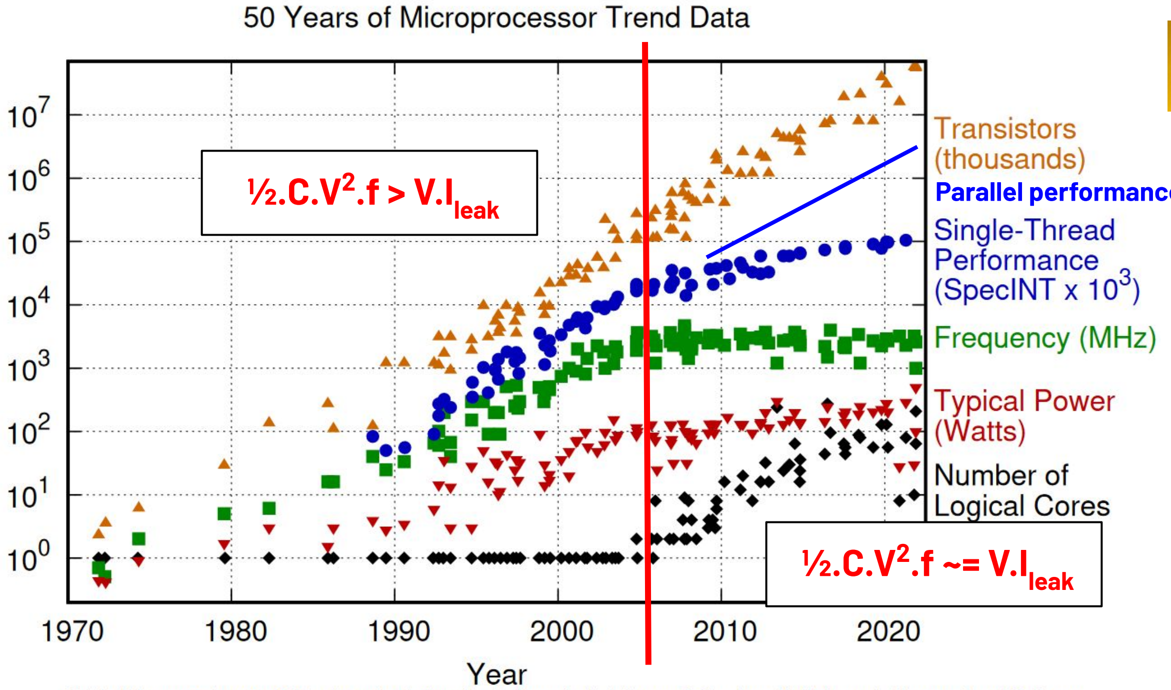
LEONARDO CINECA

intel NVIDIA



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	8,699,904	1,206.00	1,714.81	22,786
2	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
3	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure Microsoft Azure United States	2,073,600	561.20	846.84	
4	**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
5	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107
6	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	1,305,600	270.00	353.75	5,194
7	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, EVIDEN EuroHPC/CINECA Italy	1,824,768	241.20	306.31	7,494
8	**MareNostrum 5 ACC** - BullSequana XH3000, Xeon Platinum 8460Y+ 32C 2.3GHz, NVIDIA H100 64GB, Infiniband NDR, EVIDEN EuroHPC/BSC Spain	663,040	175.30	249.44	4,159
9	**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
10	**Eos NVIDIA DGX SuperPOD** - NVIDIA DGX H100, Xeon Platinum 8480C 56C 3.8GHz, NVIDIA H100, Infiniband NDR400, Nvidia NVIDIA Corporation United States	485,888	121.40	188.65	

Where We Are: The End of Dennard's Scaling



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 2010-2021 by K. Rupp

Frequency Scaling Era, the "free lunch"

- Perf/J increasing
- Memory throughput = arithmetic throughput
- ILP is exposed to programmers

Frequency scaling is dead, Energy Efficiency is king

- Power wall: Perf/J is constant
- Memory wall (mem t.p. < arithmetic t.p.)
- ILP wall, now TLD, DLP is taking over with highly specialised accelerators (GPUs…)

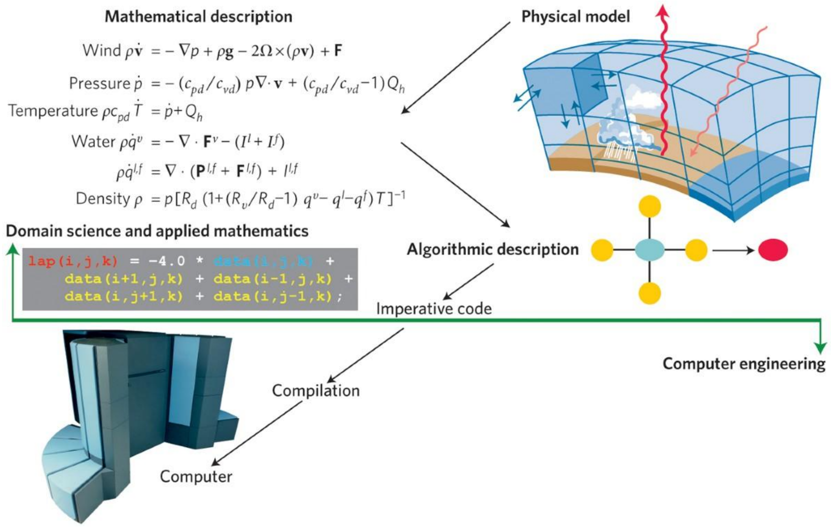
Unprecedented Heterogeneity Means Unprecedented Software Challenges

Rapid evolution of computing hardware leads to:

- **Frequent rewrite** of software
- **Unsustainable** development efforts
- A **suboptimal** use of the hardware

Either:

- Focus on a **specific hardware target** to get **maximum performance**
- Or go **portability/**
- A good trade-off: **separation of concern** between front-end (science) and back-end (software/hardware)



$$\begin{aligned}
\text{Wind } \rho\dot{\mathbf{v}} &= -\nabla p + \rho\mathbf{g} - 2\Omega\times(\rho\mathbf{v}) + \mathbf{F}\\
\text{Pressure } \dot{p} &= -(c_{pd}/c_{vd})\, p\nabla\cdot\mathbf{v} + (c_{pd}/c_{vd}-1)Q_h\\
\text{Temperature } \rho c_{pd}\dot{T} &= \dot{p}+Q_h\\
\text{Water } \rho\dot{q}^v &= -\nabla\cdot\mathbf{F}^v - (I^l + I^f)\\
\rho\dot{q}^{l,f} &= \nabla\cdot(\mathbf{P}^{l,f} + \mathbf{F}^{l,f}) + I^{l,f}\\
\text{Density } \rho &= p[R_d\,(1+(R_v/R_d-1)\,q^v - q^l - q^f)T]^{-1}
\end{aligned}$$

```
lap(i,j,k) = -4.0 * data(i,j,k) +
    data(i+1,j,k) + data(i-1,j,k) +
    data(i,j+1,k) + data(i,j-1,k);
```

Programming revisited, Thomas C. Schulthess (CSCS)

Nature Physics - 2015

Bridging the Productivity Gap

Adopt High-Level Programming Languages and Frameworks:

- **Use of High-Level Languages**: Languages like Python, Julia, and modern C++ offer more abstraction and ease of use compared to traditional HPC languages like Fortran (which is lagging behind) or C.
- **Parallel Programming Libraries**: Utilize libraries such as MPI for distributed computing, OpenMP for shared-memory parallelism, and CUDA or SYCL for GPU programming to manage complexity.

Leverage Domain-Specific Languages (DSLs) or Reusable Libraries:

- **DSLs** can simplify programming by providing constructs tailored to specific domains, hiding low-level implementation details.
- **Mathematical Libraries**: Leverage optimized libraries like BLAS, LAPACK, and PETSc for common computational tasks.

Enhance Collaboration and Code Sharing/Dissemination:

- **Version Control**: Use Git and platforms like GitLab dedicated for collaborative development.
- **Open-Source Contributions**: Share improvements and adaptations with the community to foster collective progress.

Bridging the Productivity Gap

Implement Verification and Validation (V&V):

- **Code Verification**: Regularly test code against analytical solutions or benchmarks to ensure correctness.
- **Model Validation**: Compare simulation results with experimental data to validate models.

Apply Uncertainty Quantification (UQ):

- **UQ Tools**: Integrate UQ methods to assess the impact of input uncertainties on simulation outcomes.
- **Statistical Analysis**: Use probabilistic approaches to quantify confidence levels in results

The EUROfusion Standard Software is a great framework to bridge the gap!

Bridging the Productivity Gap

Co-design:

- Optimization of software and hardware simultaneously to meet requirements
- Foster close collaborations between hardware architects, software developers and domain scientists

More attention to compiler technologies:

- Automated code tuning
- Energy-efficient algorithms (specialized LLVM IR/backends…)

Massive energy consumption:

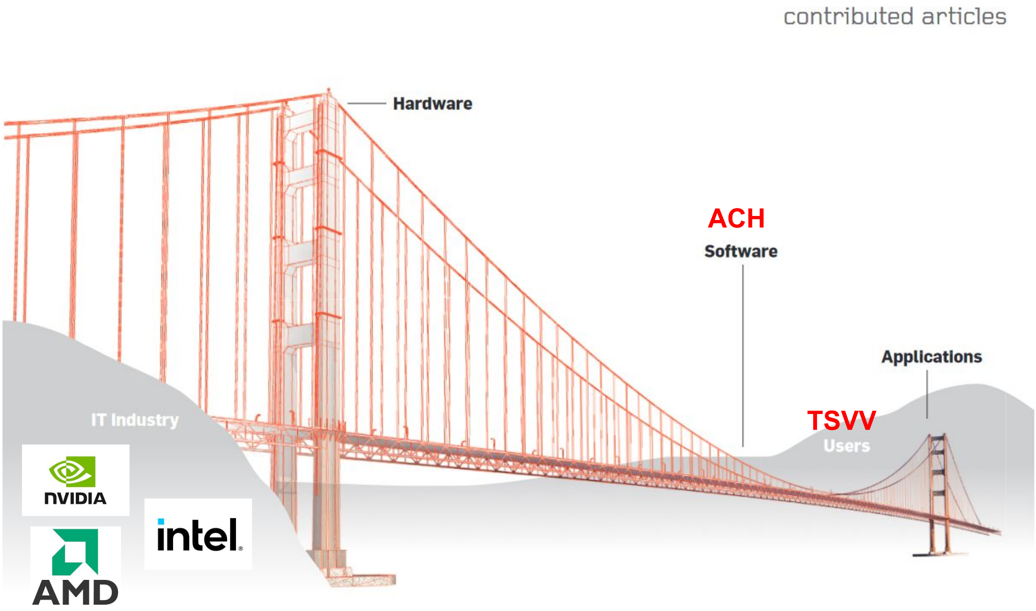
- Energy consumption will go up as performance will increase
- Energy-efficient algorithms must be developed to circumvent the end of Dennard’s scaling (e.g. data locality to avoid data movement)

I/O bottlenecks:

- Massive data generation
- Imbalance and data read/write bottlenecks
- Filesystem scalability

It's Dangerous to Go Alone...

...but the ACHs are here to help!



A View of the Parallel Computing Landscape
"Writing programs that scale with increasing numbers of cores should be as easy as writing programs for sequential computers."

Thank you!