ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

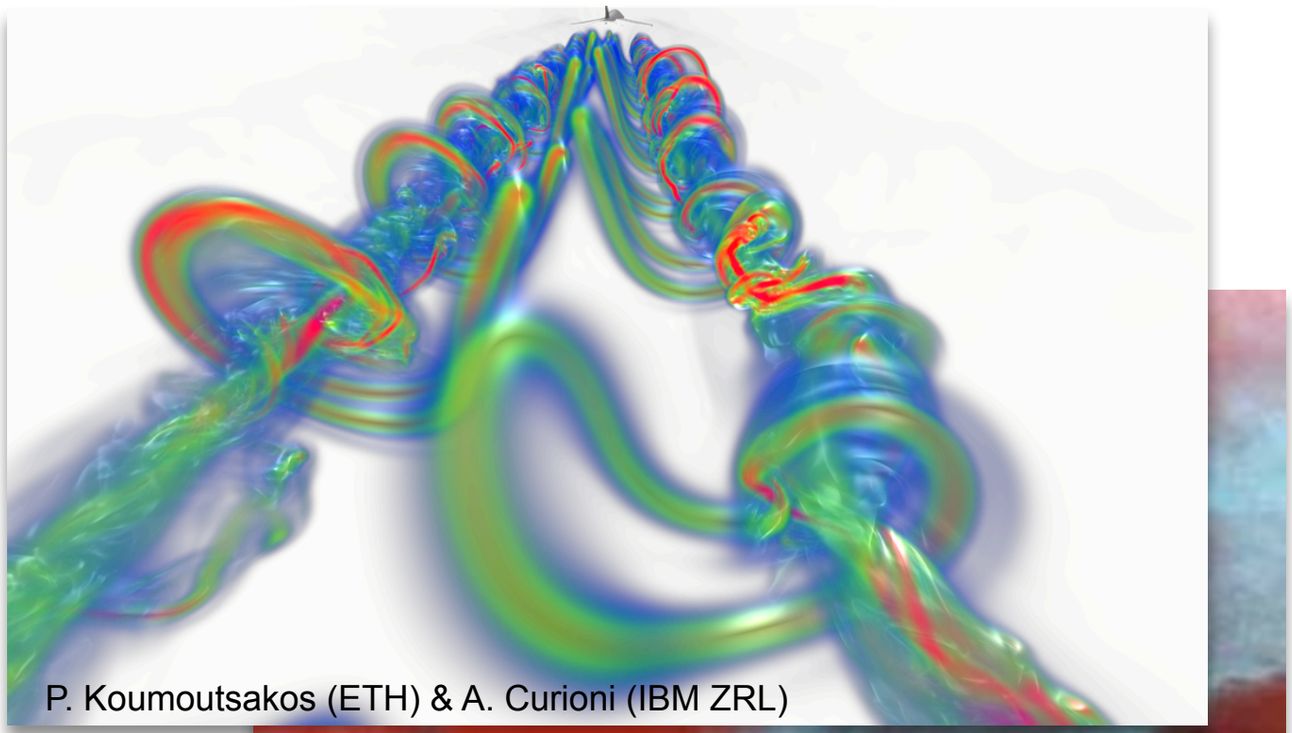
CSCS
Swiss National Supercomputing Centre

Das Unsichtbare sichtbar machen – wenn Supercomputer Prozesse simulieren

Thomas C. Schulthess

Optimized winglets reduce environmental impact of aircraft

- Computational simulation of vortex formation in wake of an aircraft
- Optimized winglets impact
  - fuel consumption
  - reduce noise level / environmental impact
- RUAG develops optimized winglets for Airbus aircraft

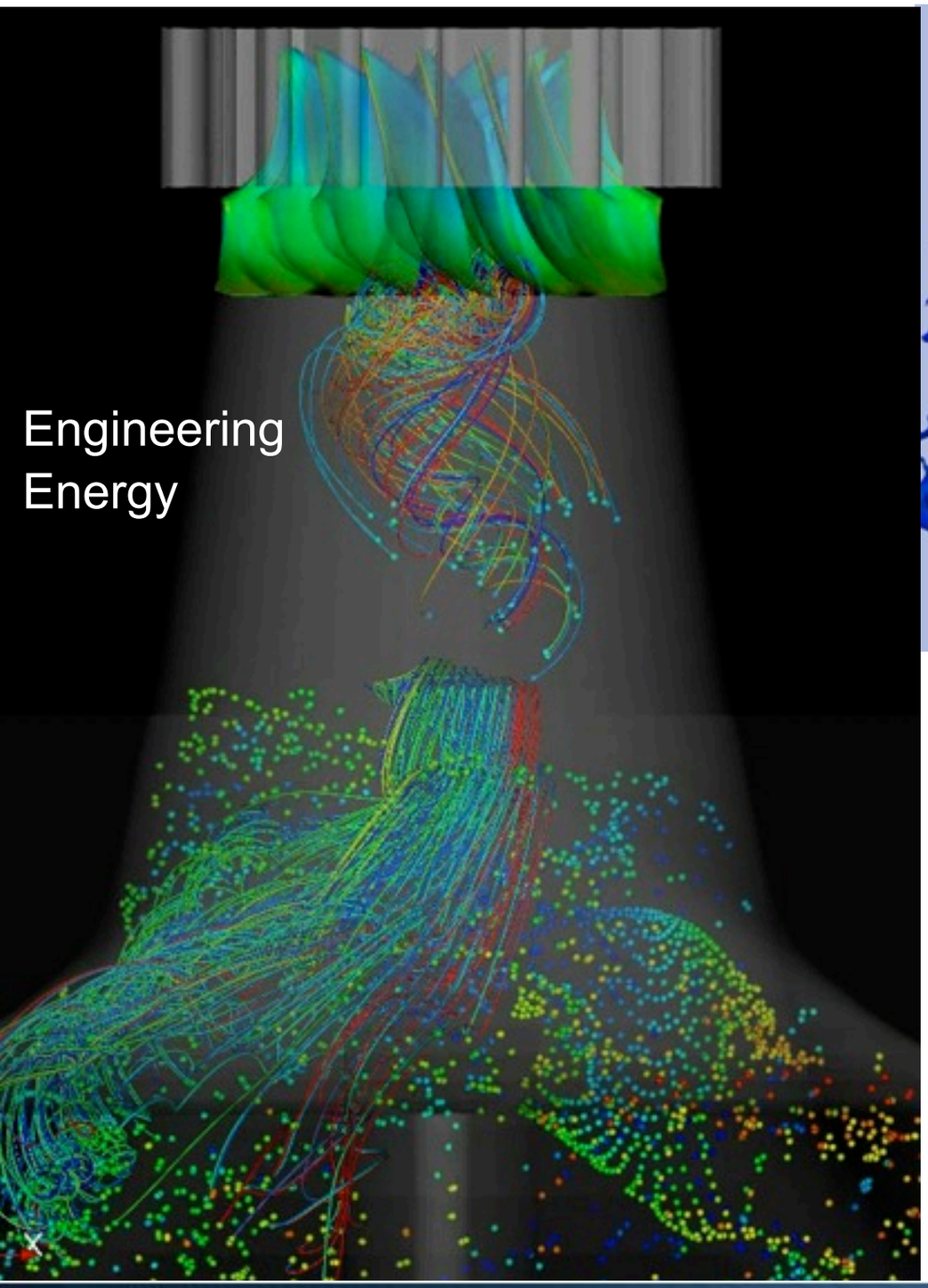


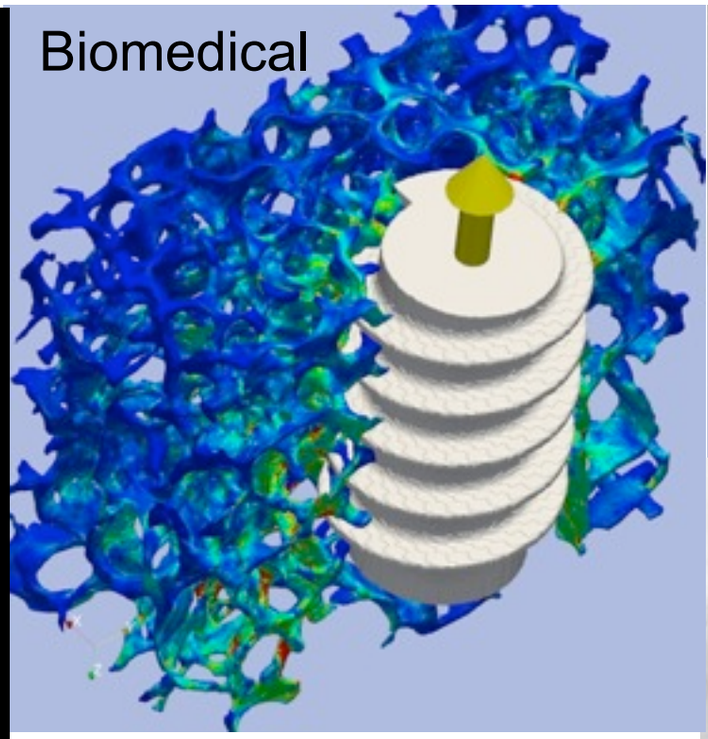
P. Koumoutsakos (ETH) & A. Curioni (IBM ZRL)

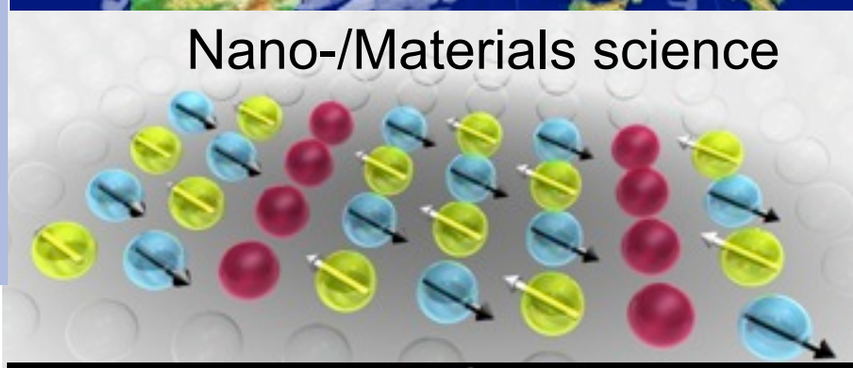
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

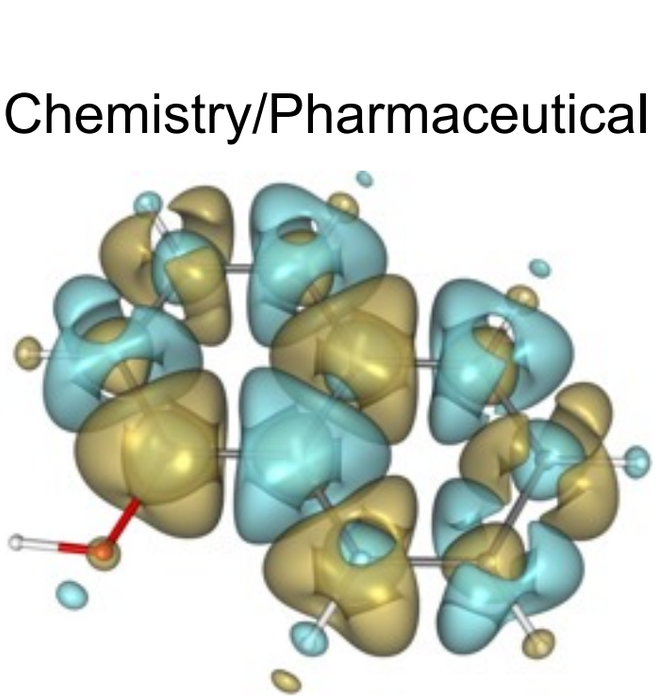
CSCS
Swiss National Supercomputing Centre

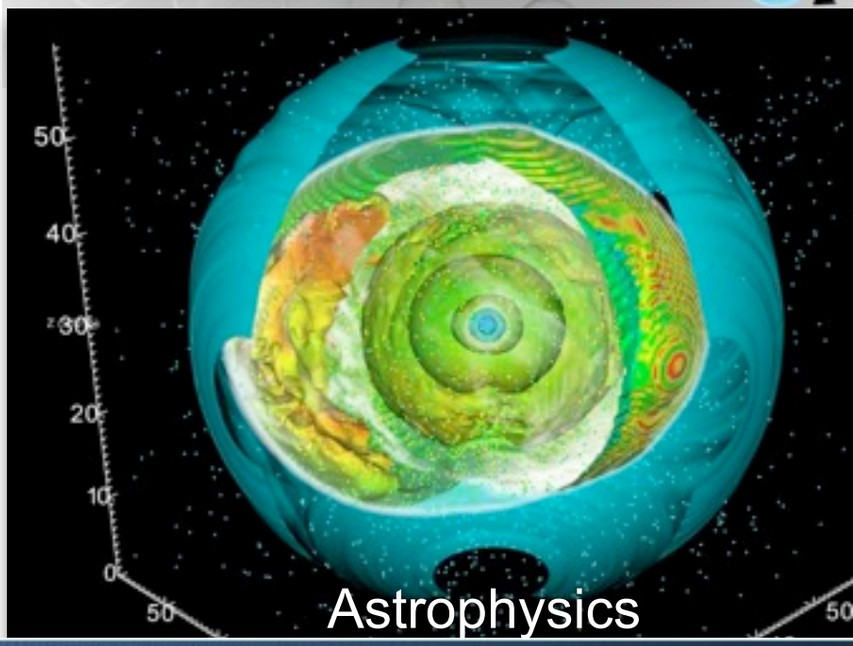
Selected application areas for simulation based science and engineering in Switzerland



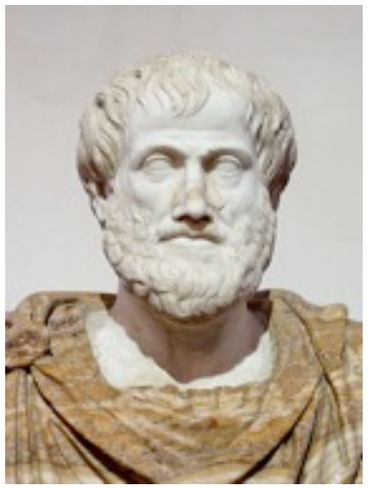




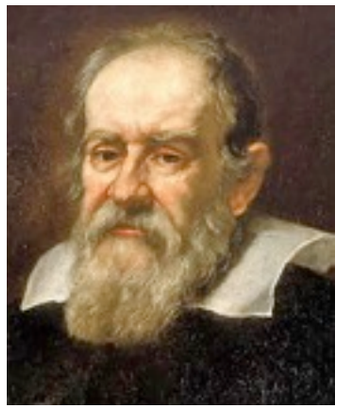


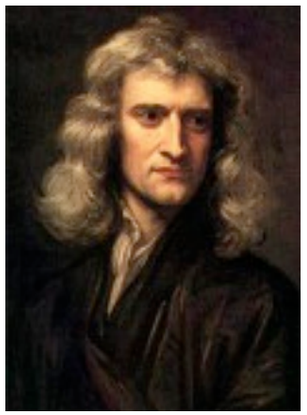


Premise: 3 pillars of 21. century scientific method

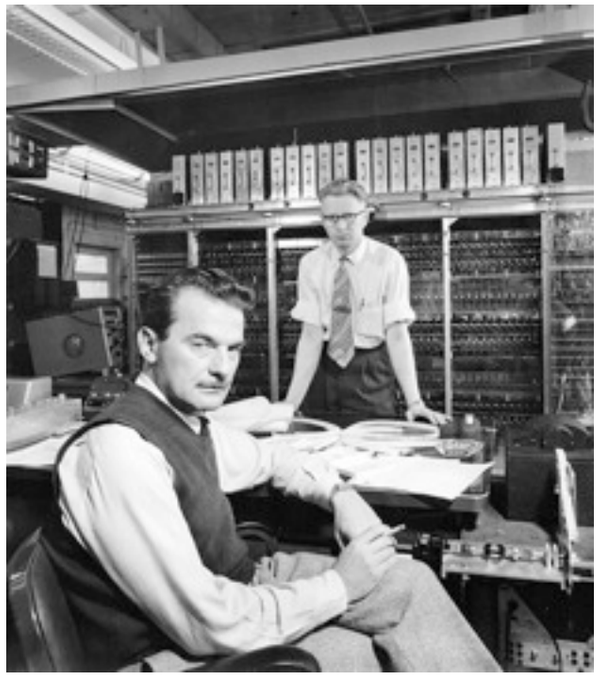


Theory (since antiquity)





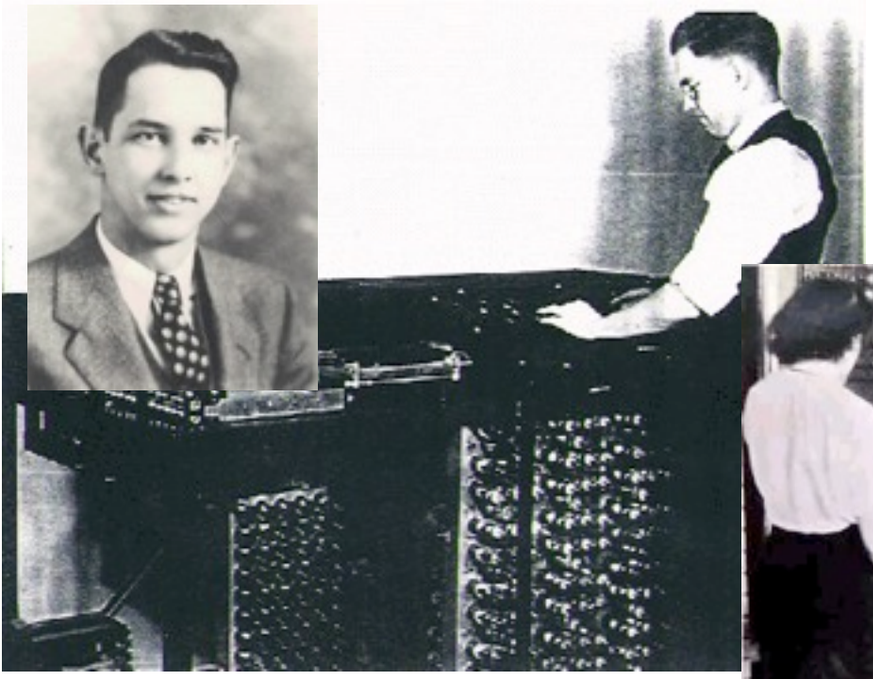
combined with experiment (since Galilei & Newton)



and simulation
(since Metropolis, Teller,
von Neumann, Fermi, ... 1940s)

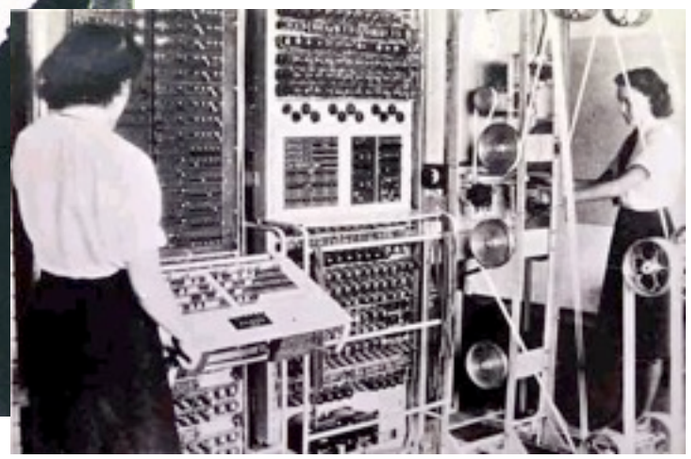
Excellence in Science requires excellence in all three areas: theory, experiment, and simulations

Electronic computing: the beginnings



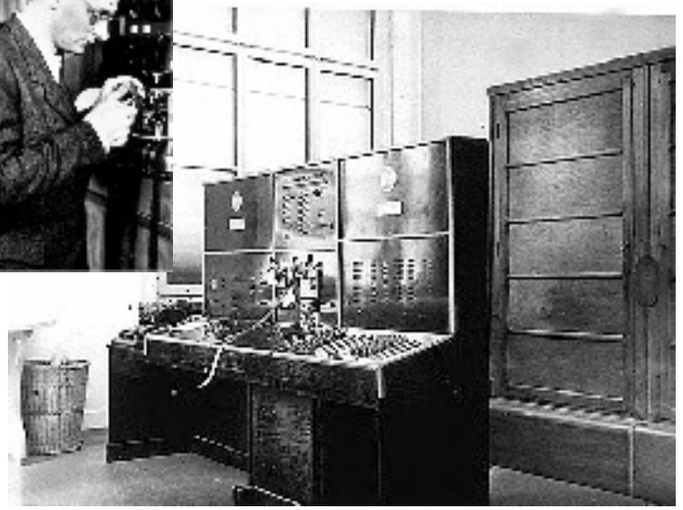
1939-42: Atanasoff-Berry Computer - Iowa State Univ.

1938: Konrad Zuse's Z1 - Germany



1943/44: Colossus Mark 1&2 - Britain

Zuse and Z3 (**1941**)



Z4 @ ETH (1950-54)

1945-51: UNIVAC I
Eckert & Mauchly - "first commercial computer"

1945: John von Neumann report that defines the "von Neuman" architecture

Since the dawn of High-performance computing:
Supercomputing at Los Alamos National Laboratory

1946: ENIAC

1952: MANIAC I

1957: MANIAC II

...

1974: Cray 1 - vector architecture

...

1987: nCUBE 10 (SNL) - MPP architecture

1993: Intel Paragon (SNL)

1993: Cray T_3D

...

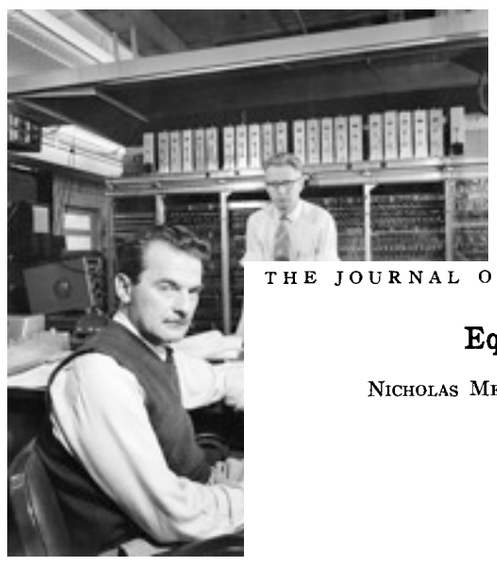
2004: IBM BG/L (LLNL)

2005: Cray Redstorm/XT3 (SNL)

2007: IBM BG/P (ANL)

2008: IBM "Roadrunner"

2008: Cray XT5 (ORNL)



Nicholas Metropolis: group leader in LANL's **T Division** that designed MANIAC I & II

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

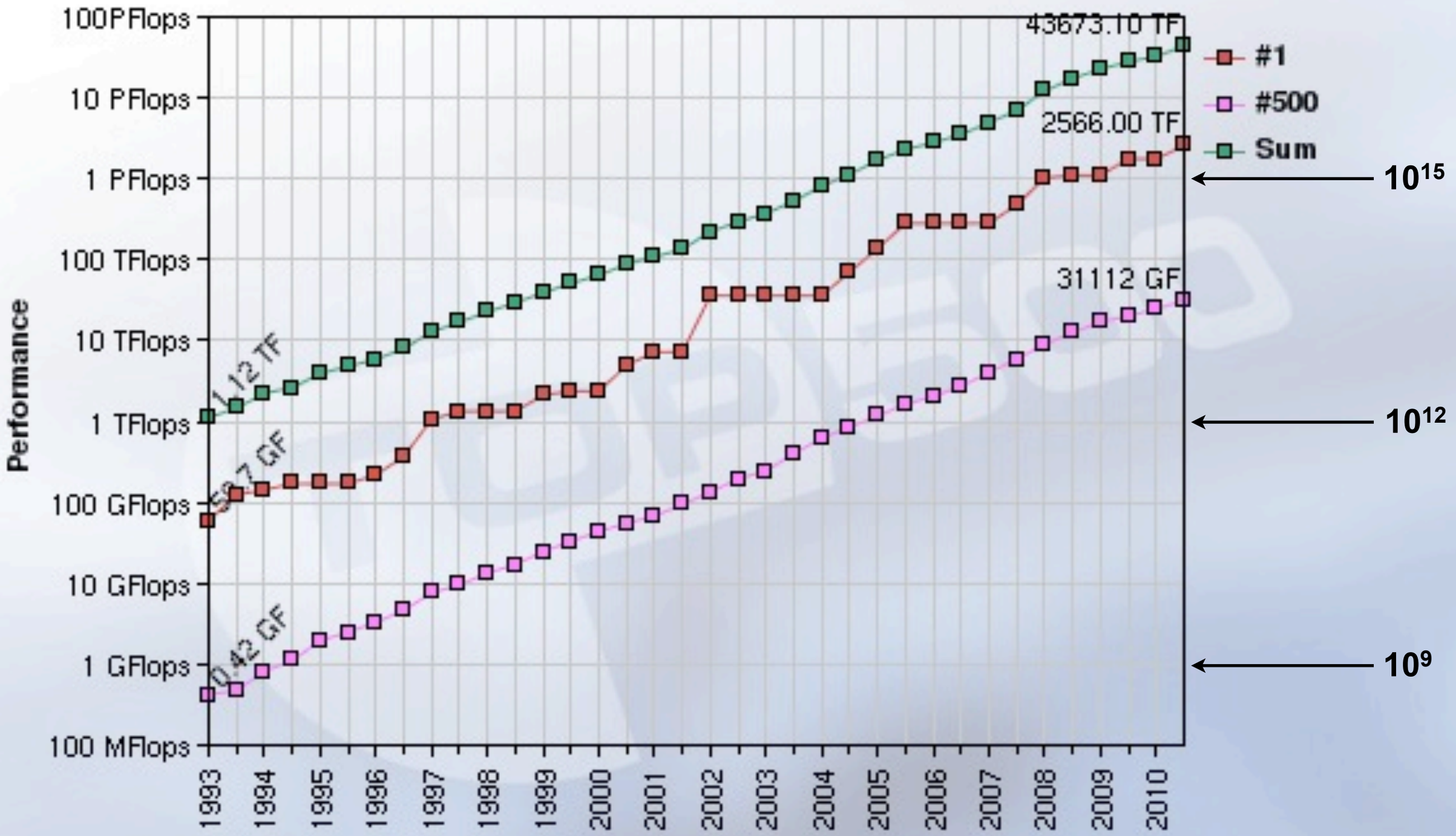
A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

2002:
Japanese Earth Simulator - Sputnik shock of HPC

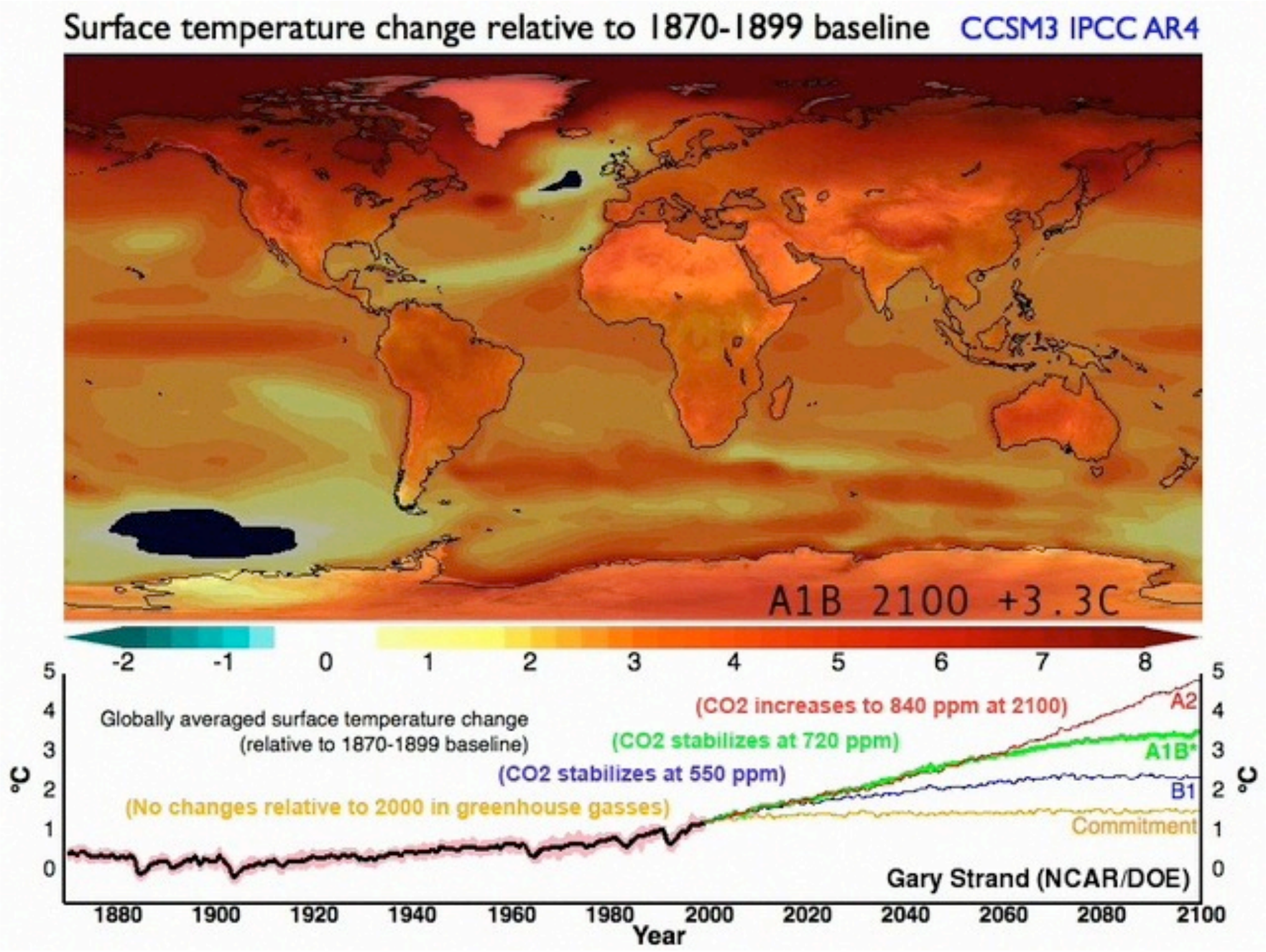
Peak: 1.382 TF/s
Quad-Core AMD Freq.: 2.3 GHz
150,176 compute cores
Memory: 300 TB

Performance Development

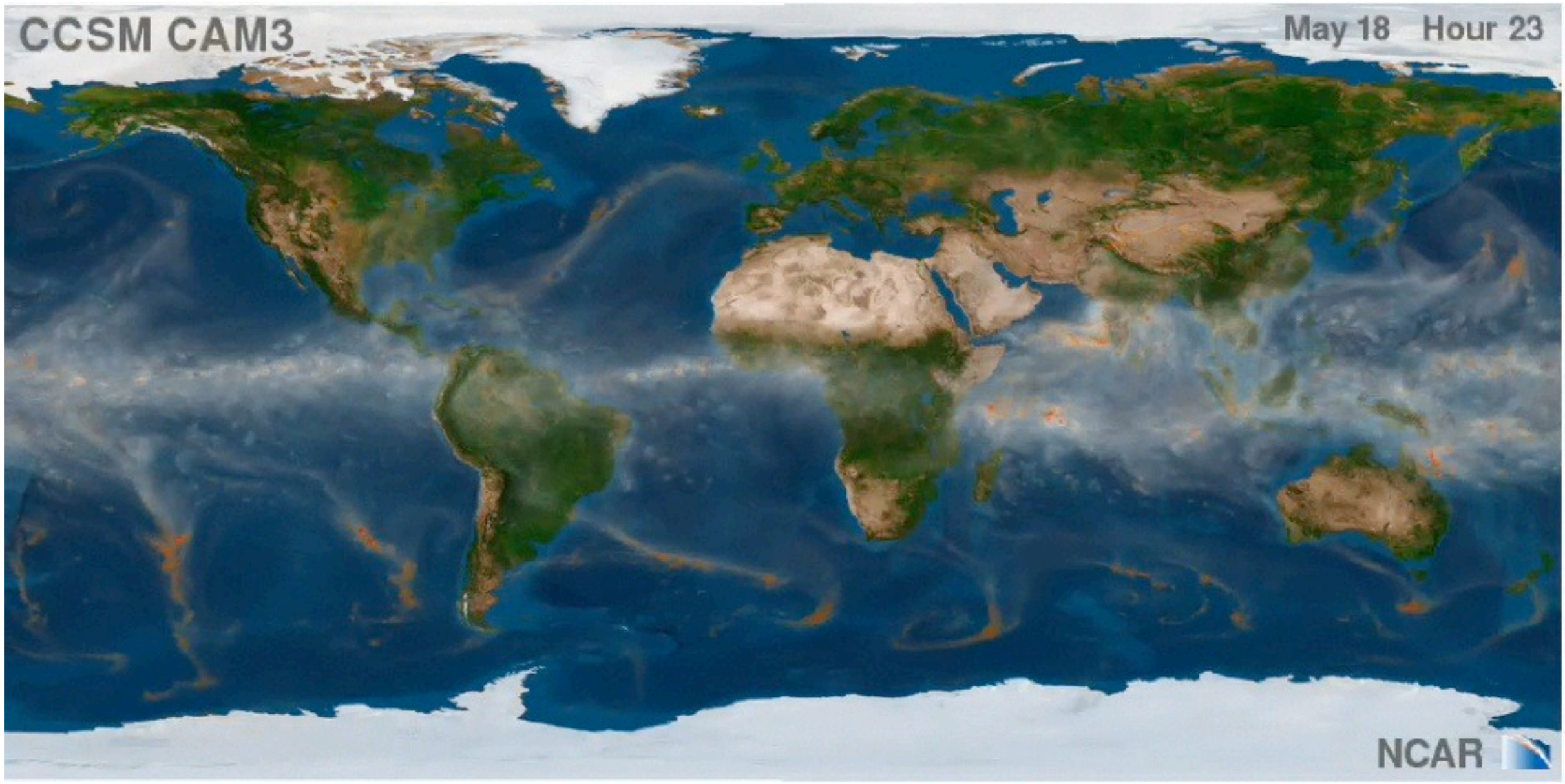
Flops = floating point operation per second **Peta (P) = 10^{15}**



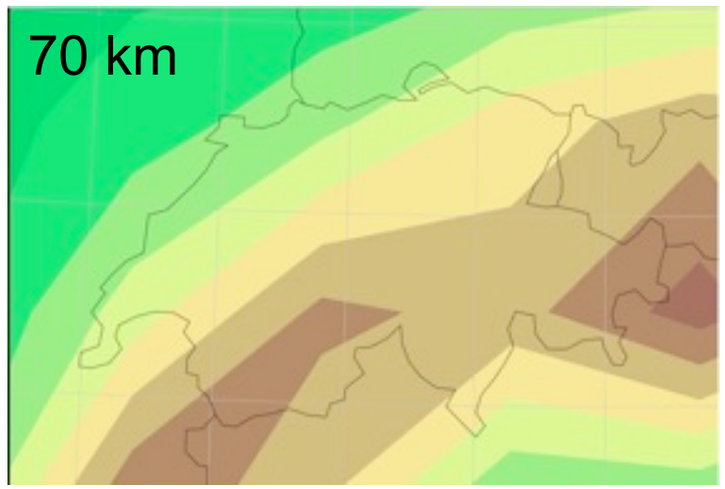
Today's state of the art climate simulation (resolution T85 ~ 148 km)

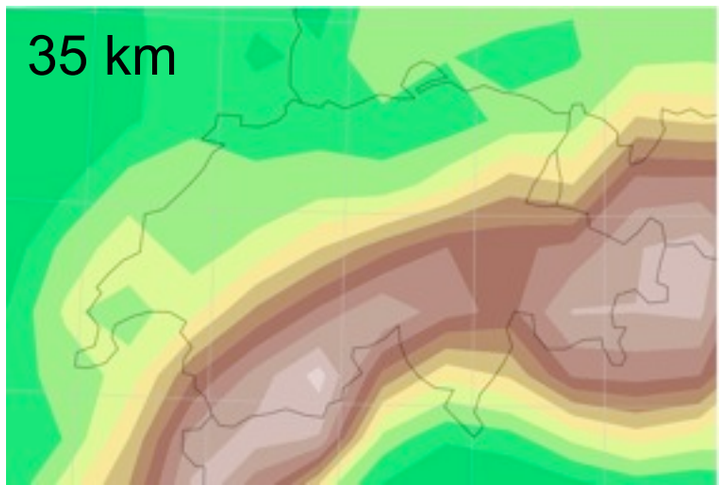


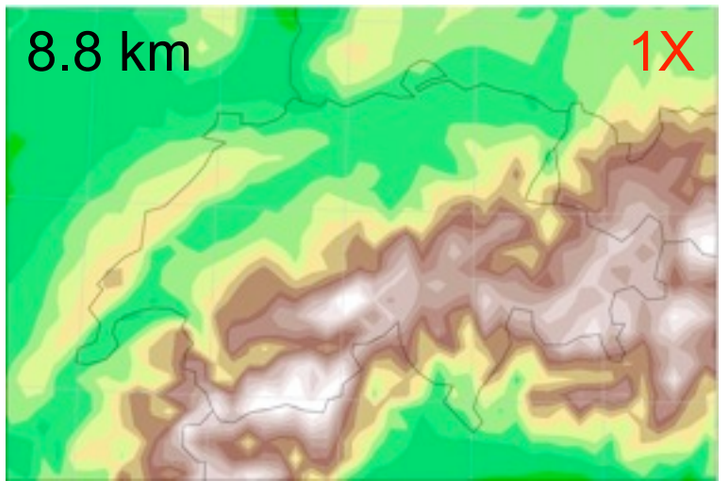
Experimental climate running at higher resolution (resolution T341 ~ 37 km)

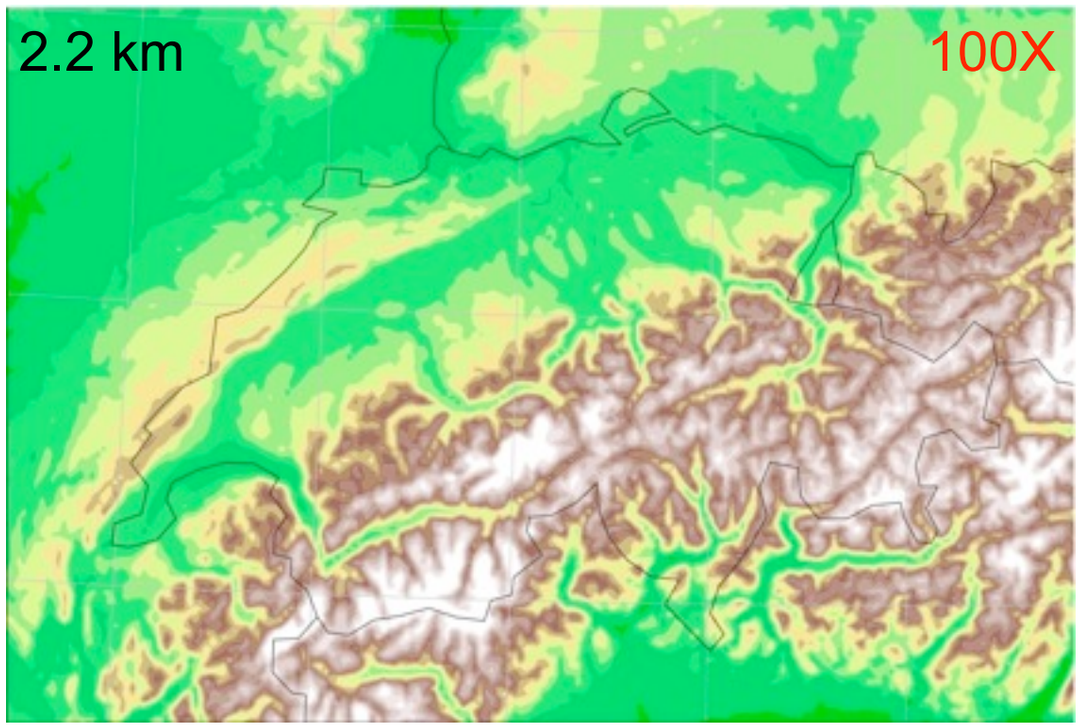


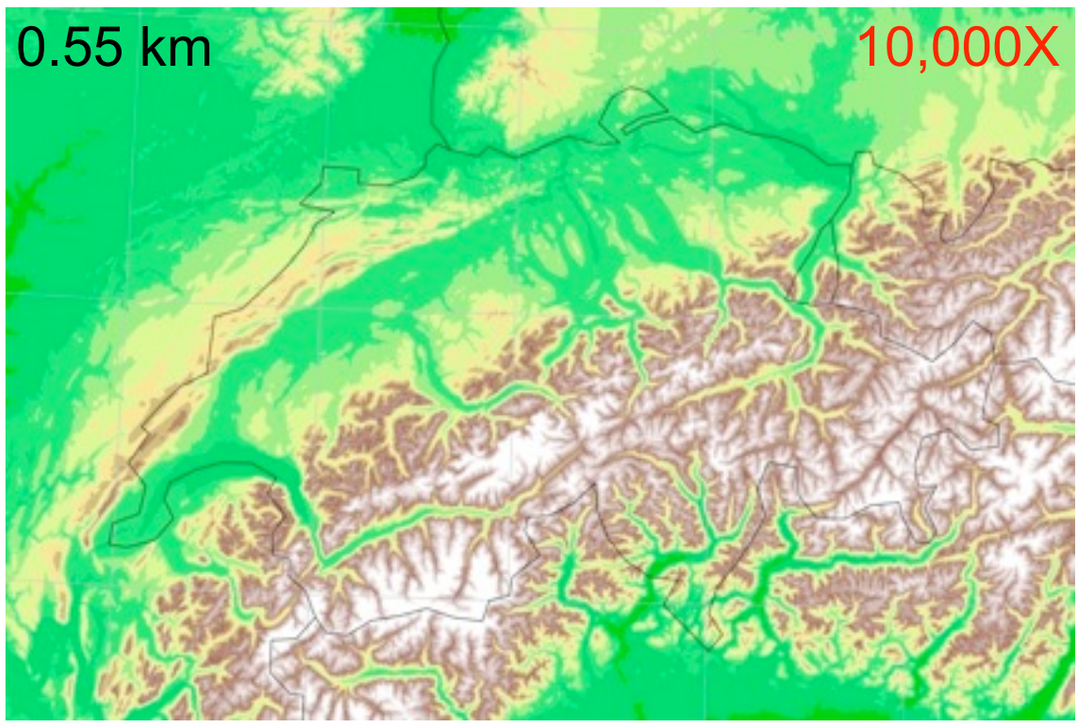
Why resolution is such an issue for Switzerland





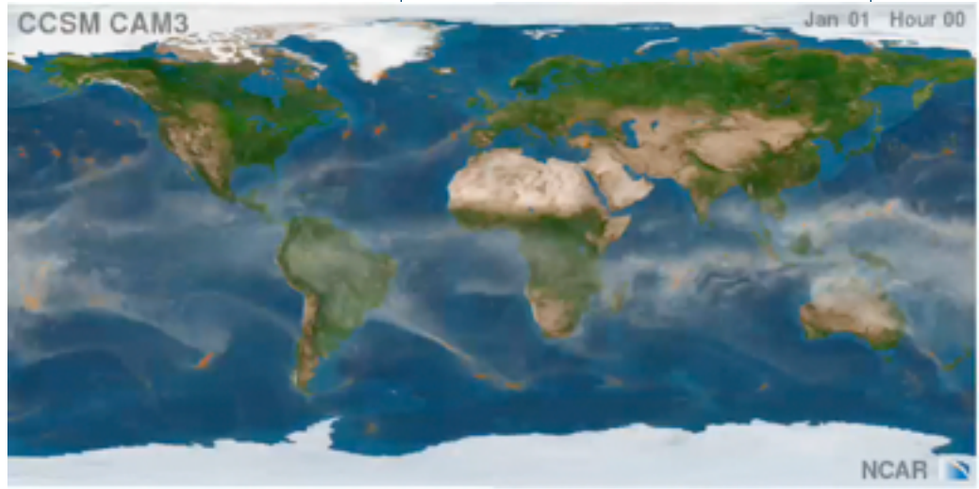




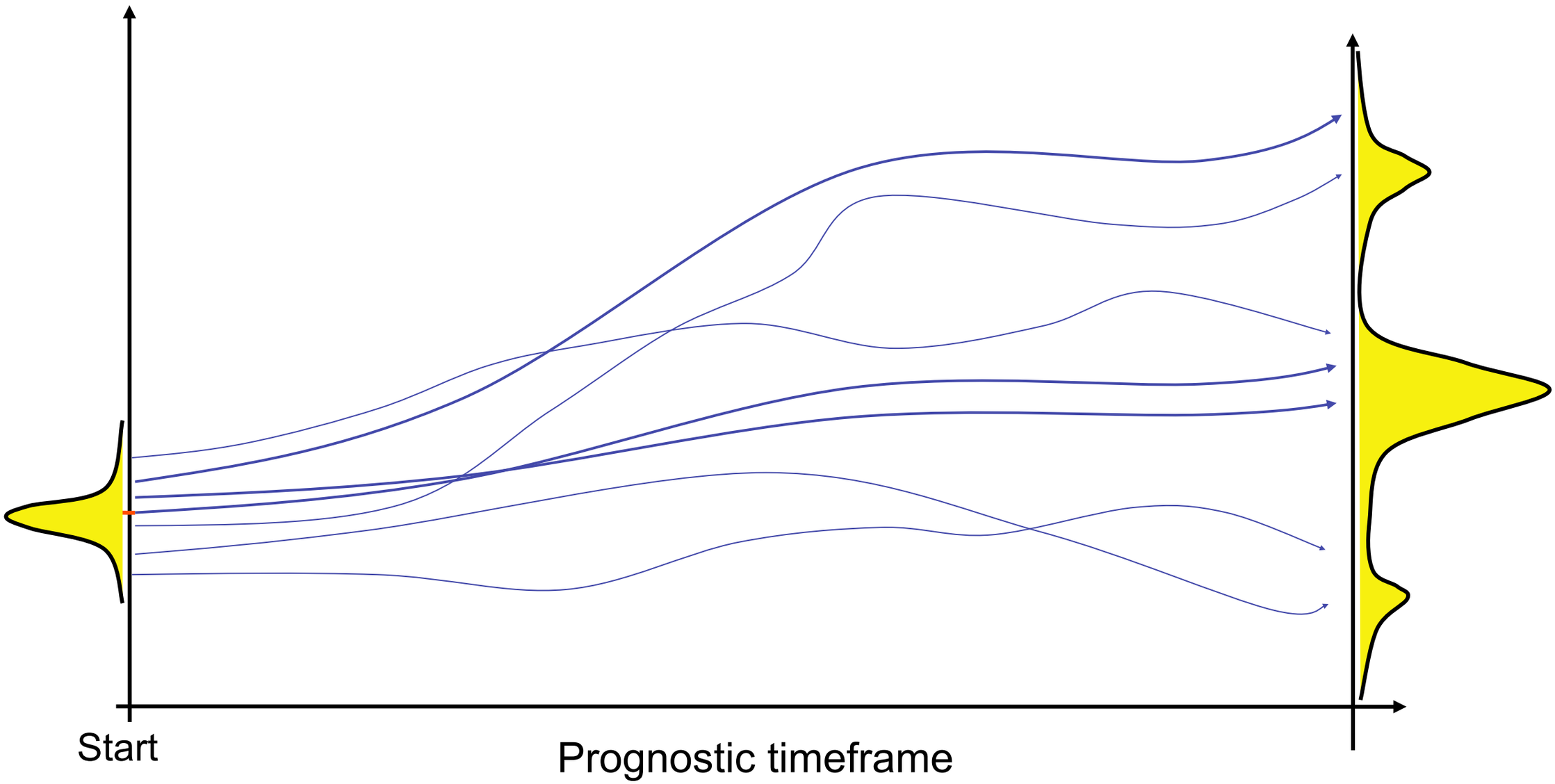


Source: Oliver Fuhrer, MeteoSwiss

Prognostic uncertainty



The weather system is chaotic
→ rapid growth of small perturbations (butterfly effect)



Ensemble method: compute distribution over many simulations

Source: Oliver Fuhrer, MeteoSwiss

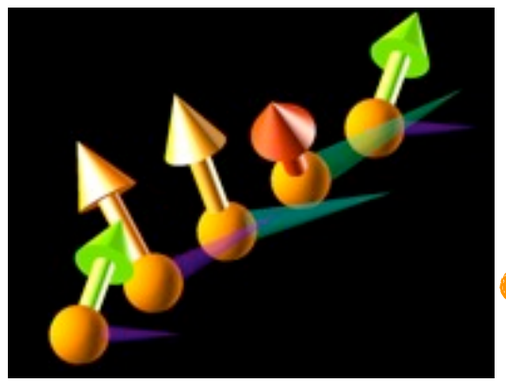
Computer performance and application performance increase ~10^3 every decade

~100 Kilowatts → ~5 Megawatts → 20-30 MW →

~1 Exaflop/s

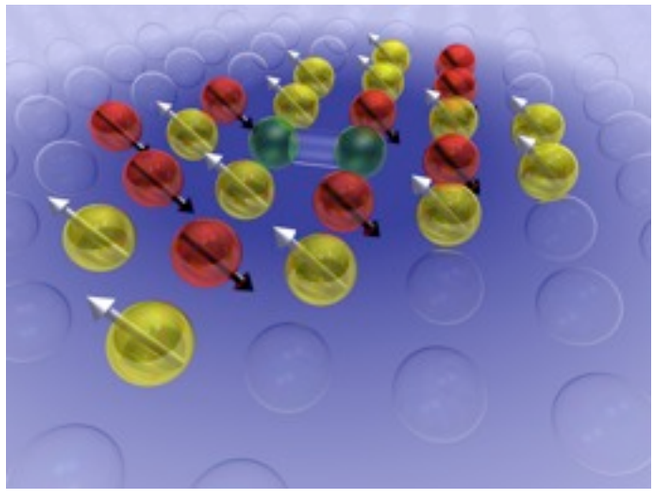
100 million or billion processing cores (!)

1 Gigaflop/s
Cray YMP
8 processors



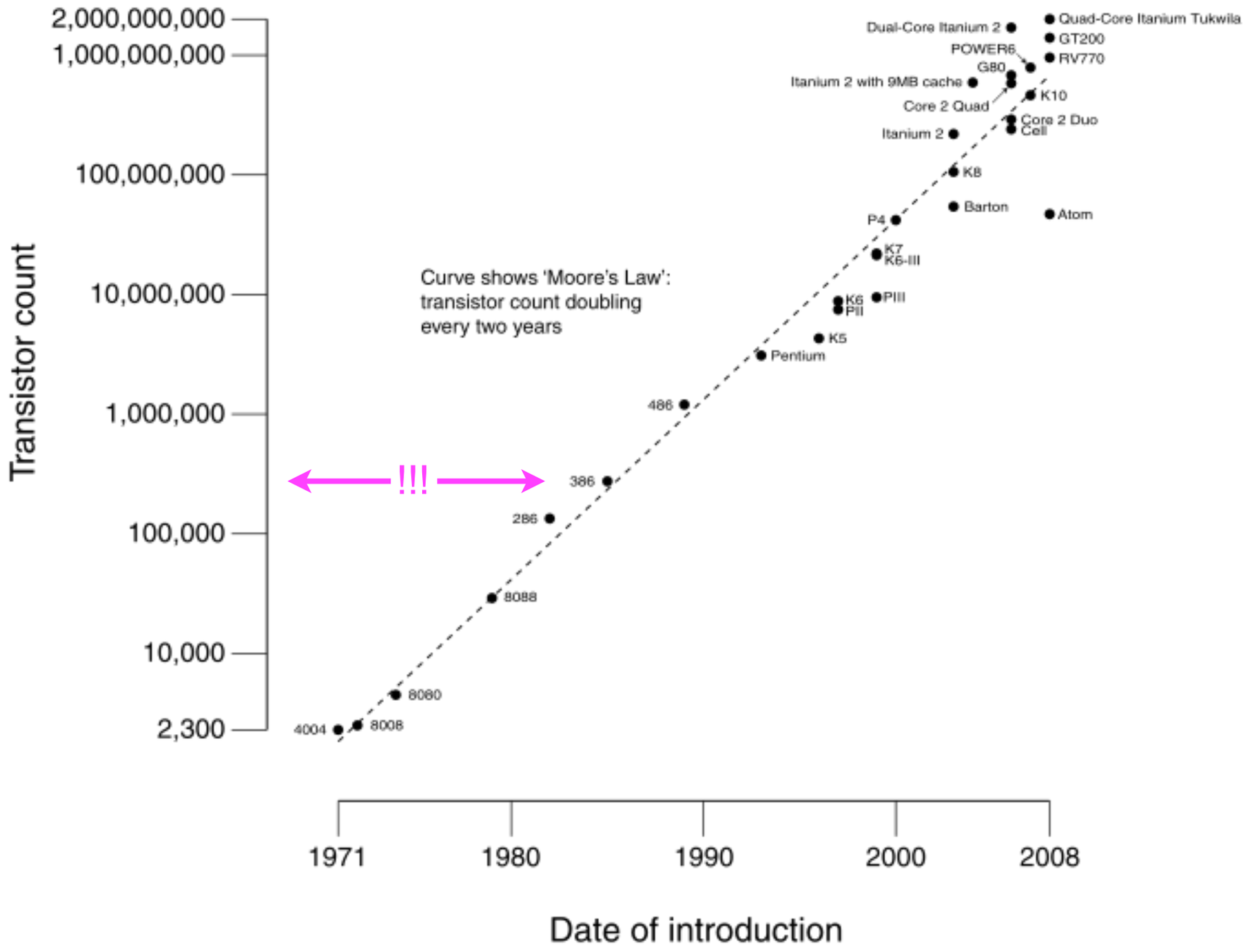
1.02 Teraflop/s
Cray T_{3E}
1'500 processors

1.35 Petaflop/s
Cray XT5
150'000 processors



1988	1998	2008	2018
First sustained GFlop/s Gordon Bell Prize 1988	First sustained TFlop/s Gordon Bell Prize 1998	First sustained PFlop/s Gordon Bell Prize 2008	Another 1,000x increase in sustained performance

CPU Transistor Counts 1971-2008 & Moore's Law



Source: Wikipedia, the free encyclopedia

Moore's Law is still alive and well

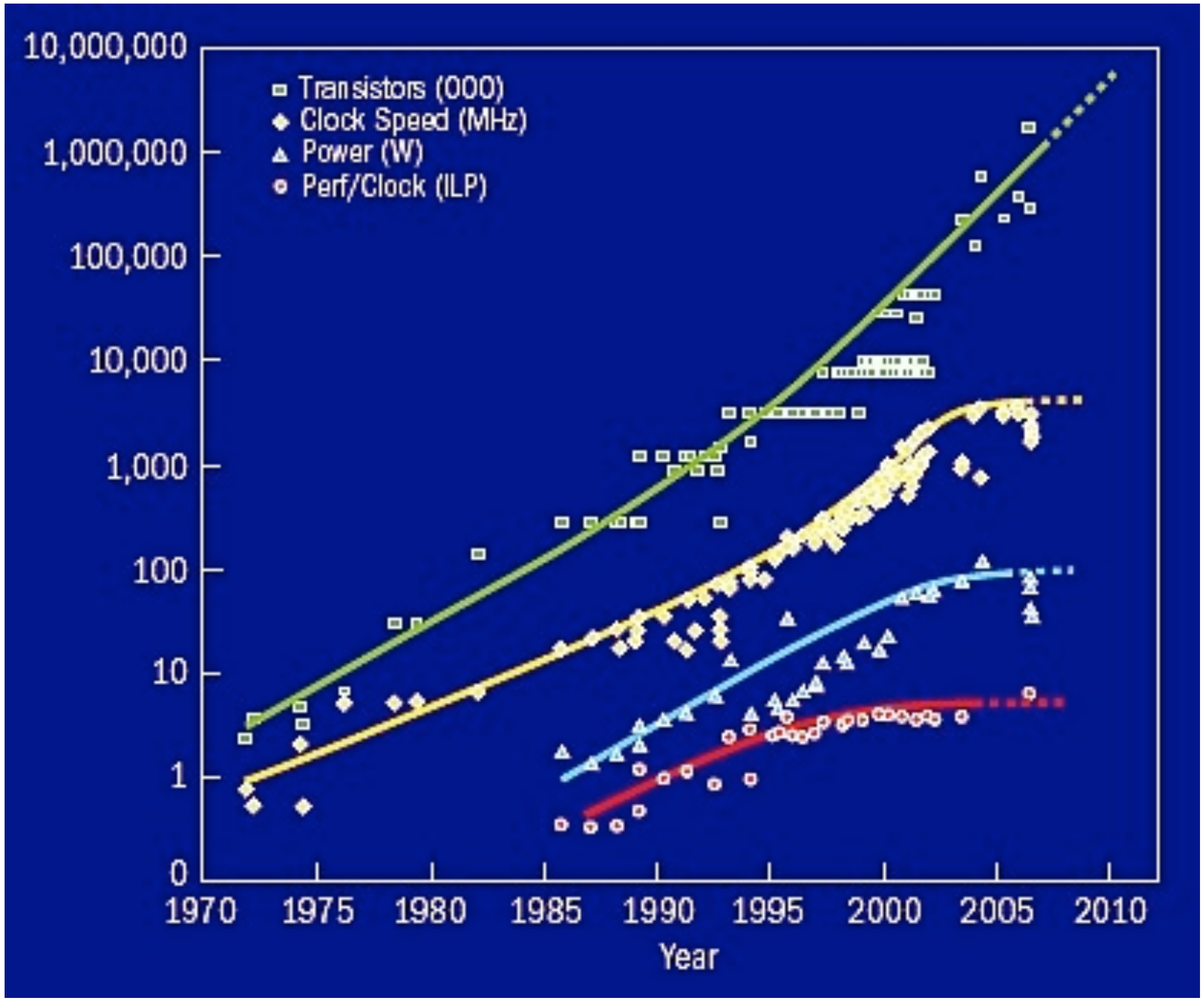
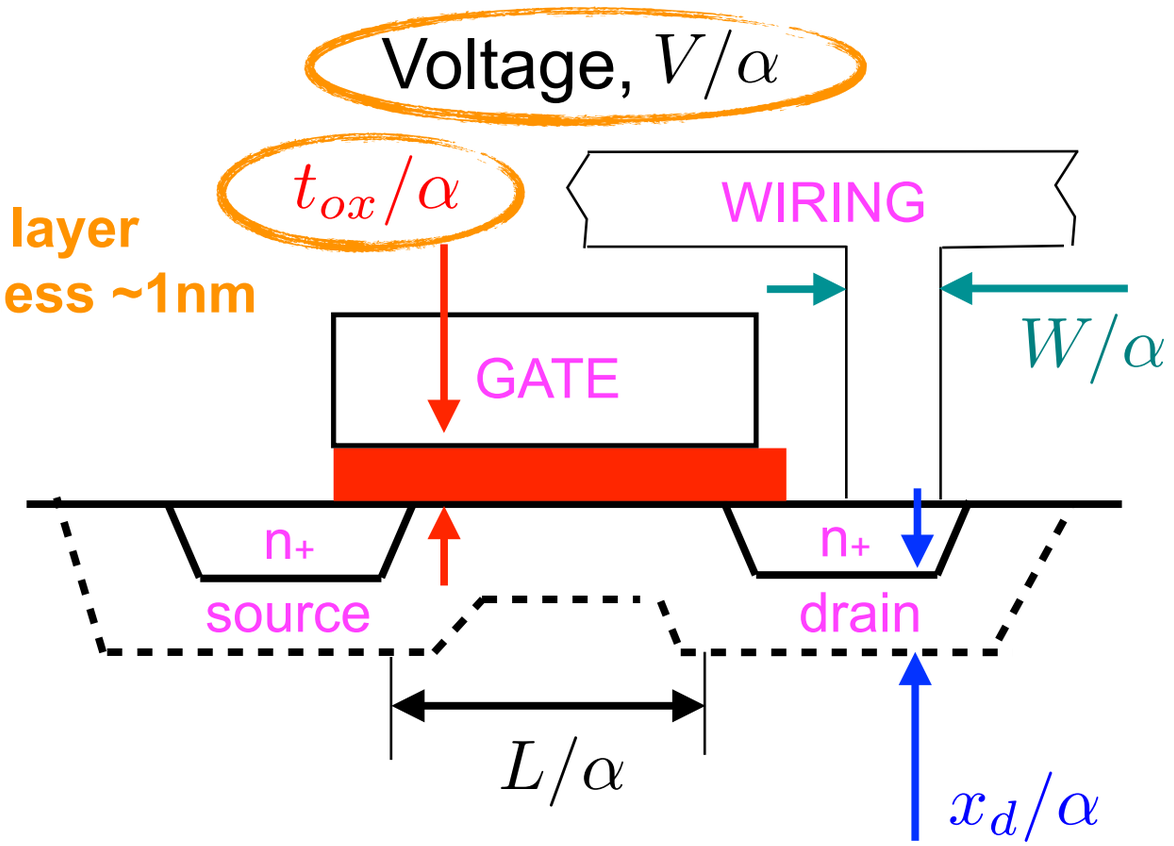


illustration: A. Tovey, source: D. Patterson, UC Berkeley

Limits of CMOS scaling



Oxide layer thickness ~1nm

p substrate, doping $\alpha^* N_A$

SCALING

~~Voltage: V/α~~

~~Oxide: t_{ox}/α~~

Wire width: W/α

Gate Width: L/α

Diffusion: x_d/α

Substrate: $\alpha^* N_A$

CONSEQUENCE:

Higher density: $\sim \alpha^2$

Higher speed: $\sim \alpha$

Power/ckt: $\sim 1/\alpha^2$

~~Power density: ~ constant~~

Source: Ronald Luijten, IBM-ZRL

The power challenge today is a precursor of more physical limitations in scaling – atomic limit!

Moore's Law 2008-2020

Semiconductor Device Scaling Factors

Technology (High Volume)	45nm (2008)	32nm (2010)	22nm (2012)	16nm (2014)	11nm (2016)	8nm (2018)	5nm (2020)
Transistor density	1.75	1.75	1.75	1.75	1.75	1.75	1.75
Frequency scaling	15%	10%	8%	5%	4%	3%	2%
Voltage (Vdd) scaling	-10%	-7.5%	-5%	-2.5%	-1.5%	-1%	-0.5%
Dimension & Capacitance	0.75	0.75	0.75	0.75	0.75	0.75	0.75
SD Leakage scaling/micron	1X Optimistic to 1.43X Pessimistic						

1000 fold increase in performance in 10 years:

> previously: double transistor density every 18 months = 100X in 10 years
frequency increased

> now: "only" 1.75X transistor density every 2 years = 16X in 10 years
frequency almost the same

Need to make up a factor 60 somewhere else

Source: Rajeeb Hazra's (HPC@Intel) talk at SOS14, March 2010

Moore's Law 2008-2020

Semiconductor Device Scaling Factors

Technology (High Volume)	45nm (2008)	32nm (2010)	22nm (2012)	16nm (2014)	11nm (2016)	8nm (2018)	5nm (2020)
Transistor density	1.75	1.75	1.75	1.75	1.75	1.75	1.75
Frequency scaling	15%	10%	8%	5%	4%	3%	2%
Voltage (Vdd) scaling	-10%	-7.5%	-5%	-2.5%	-1.5%	-1%	-0.5%
Dimension & Capacitance	0.75	0.75	0.75	0.75	0.75	0.75	0.75
SD Leakage scaling/micron	1X Optimistic to 1.43X Pessimistic						

Sources: International Technology Roadmap for Semiconductors and Intel

Moore's Law Takes Miracles ... But It Isn't The Miracle That Will Carry The Day

8

Source: Rajeeb Hazra's (HPC@Intel) talk at SOS14, March 2010

Petaflop/s = 10^{15} 64-bit floating point operations / sec.

which takes more energy?

64-bit floating-point fused multiply add **or** moving three 64-bit operands 20 mm across the die

```
      934,569.299814557
x          52.827419489135904
-----------------------------
= 49,370,884.442971624253823
+           4.20349729193958
-----------------------------
= 49,370,888.64646892
```

this takes over 3x the energy!

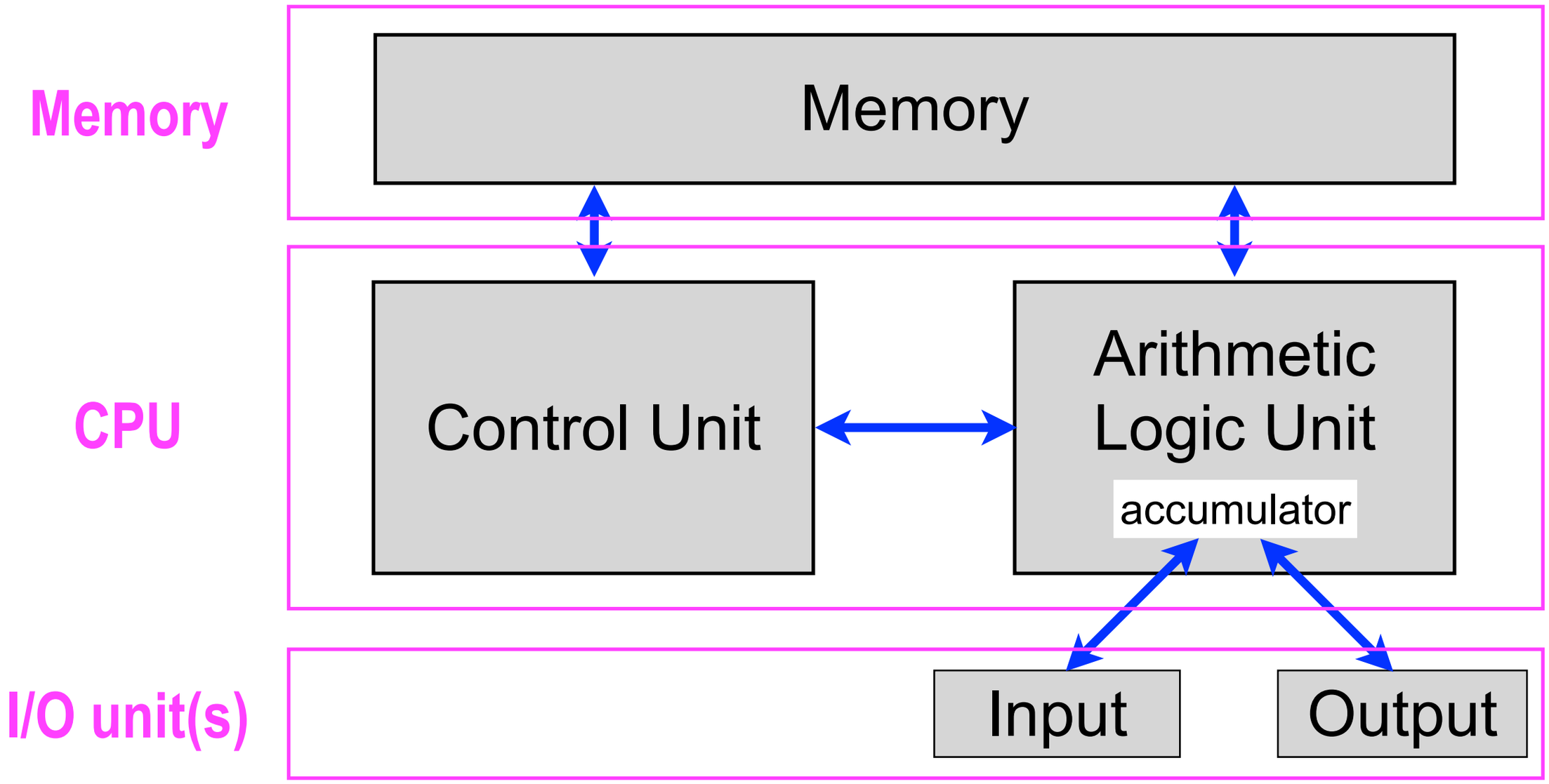
loading the data from off chip takes > 10x more yet

source: Steve Scott, Cray Inc.

moving data is expensive – exploiting data locality is critical to energy efficiency

If we care about energy consumption, we have to worry about these and other physical considerations of the computation – but where is the separation of concerns?

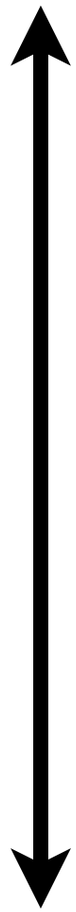
Von Neumann Architecture:



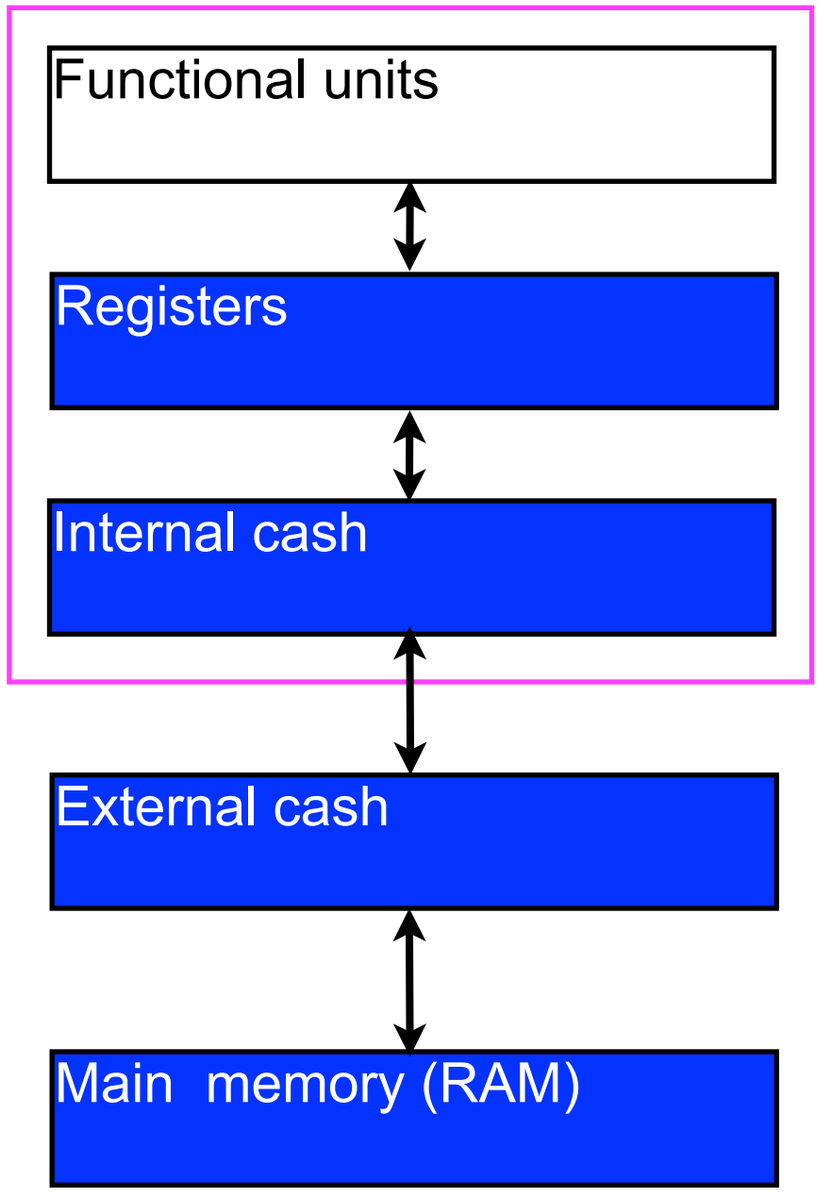
stored-program concept = general purpose computing machine

Memory hierarchy to work around latency and bandwidth problems

Expensive, fast, small



Cheap, slow, large



CPU

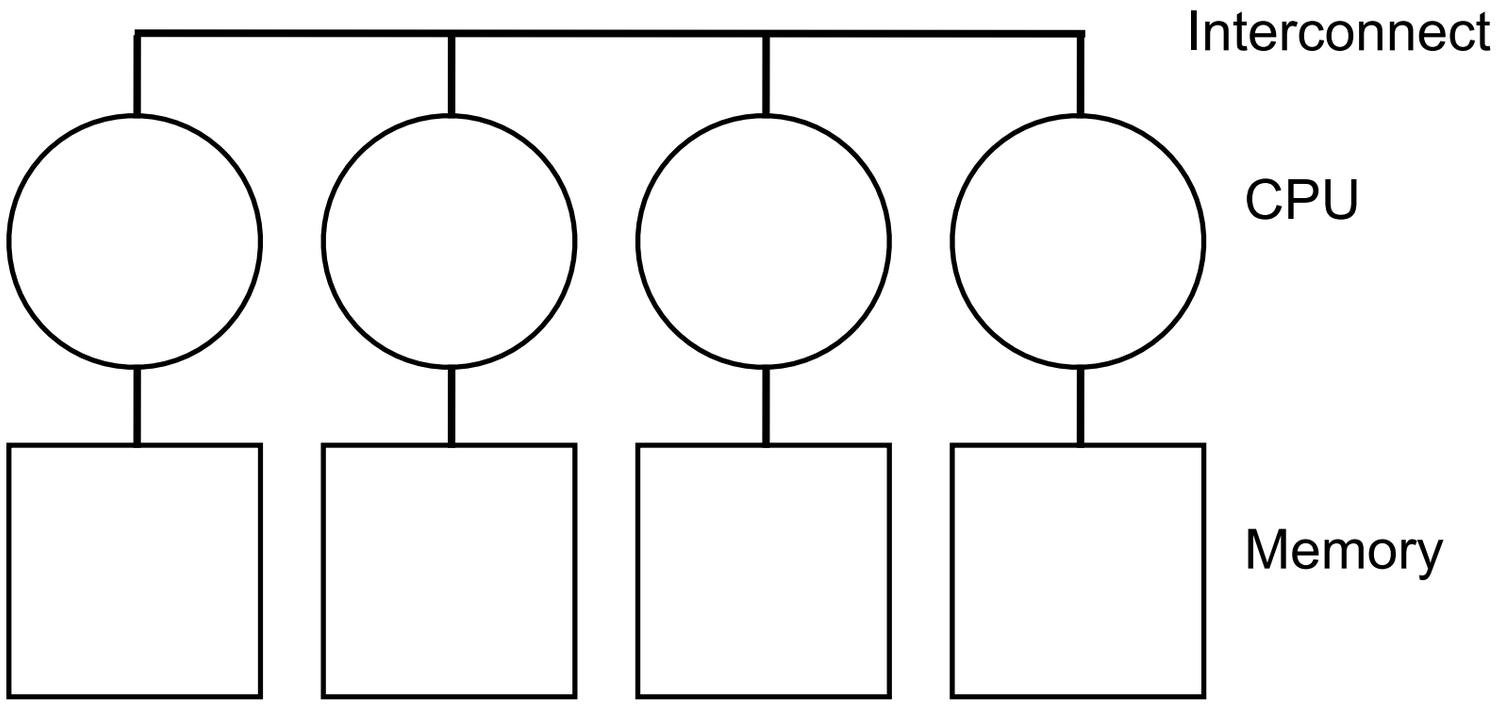
~100 GB/s
~ 6-10 ns

~50 GB/s

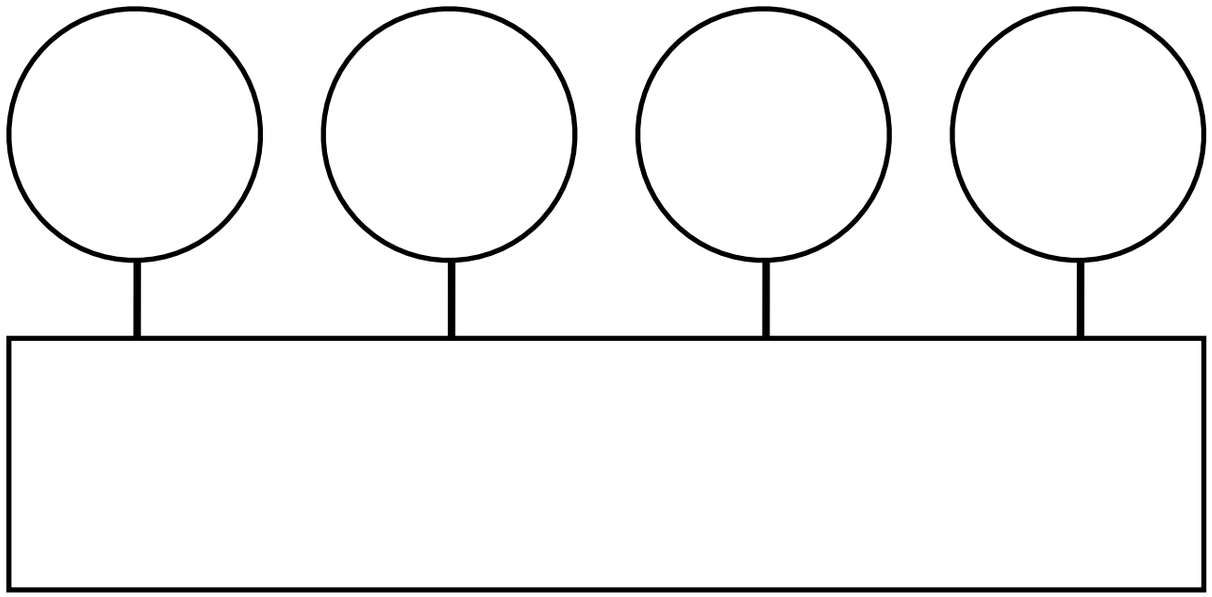
~10 GB/s
~ 75 ns

Distributed vs. shared memory architecture

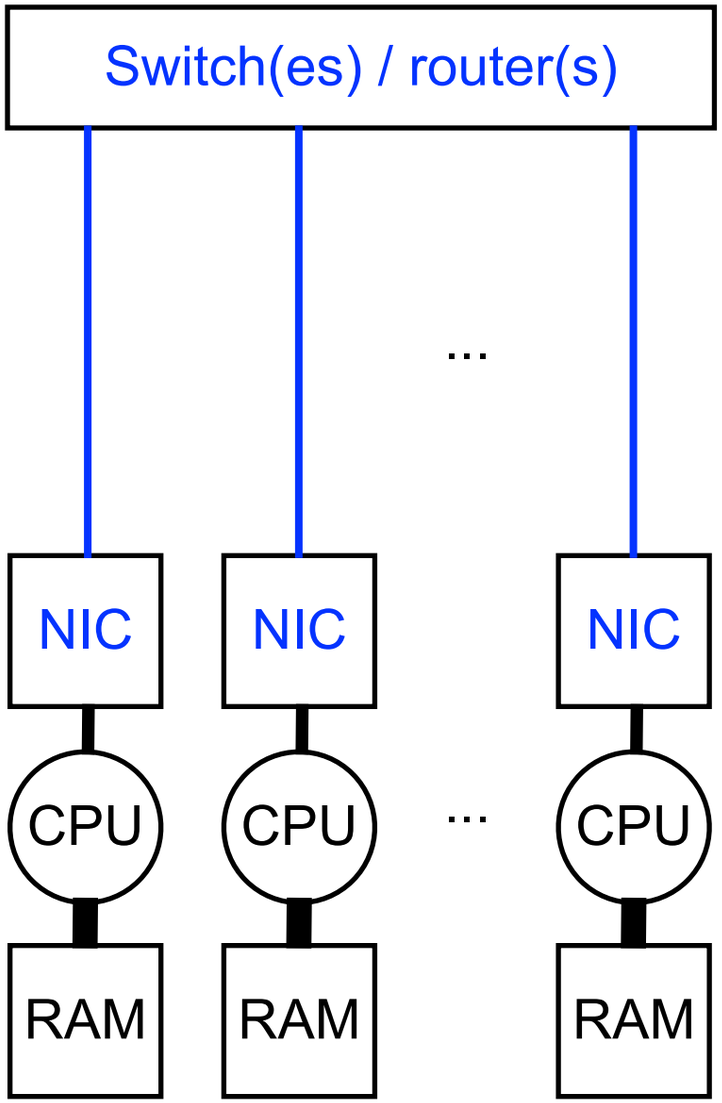
Distributed memory

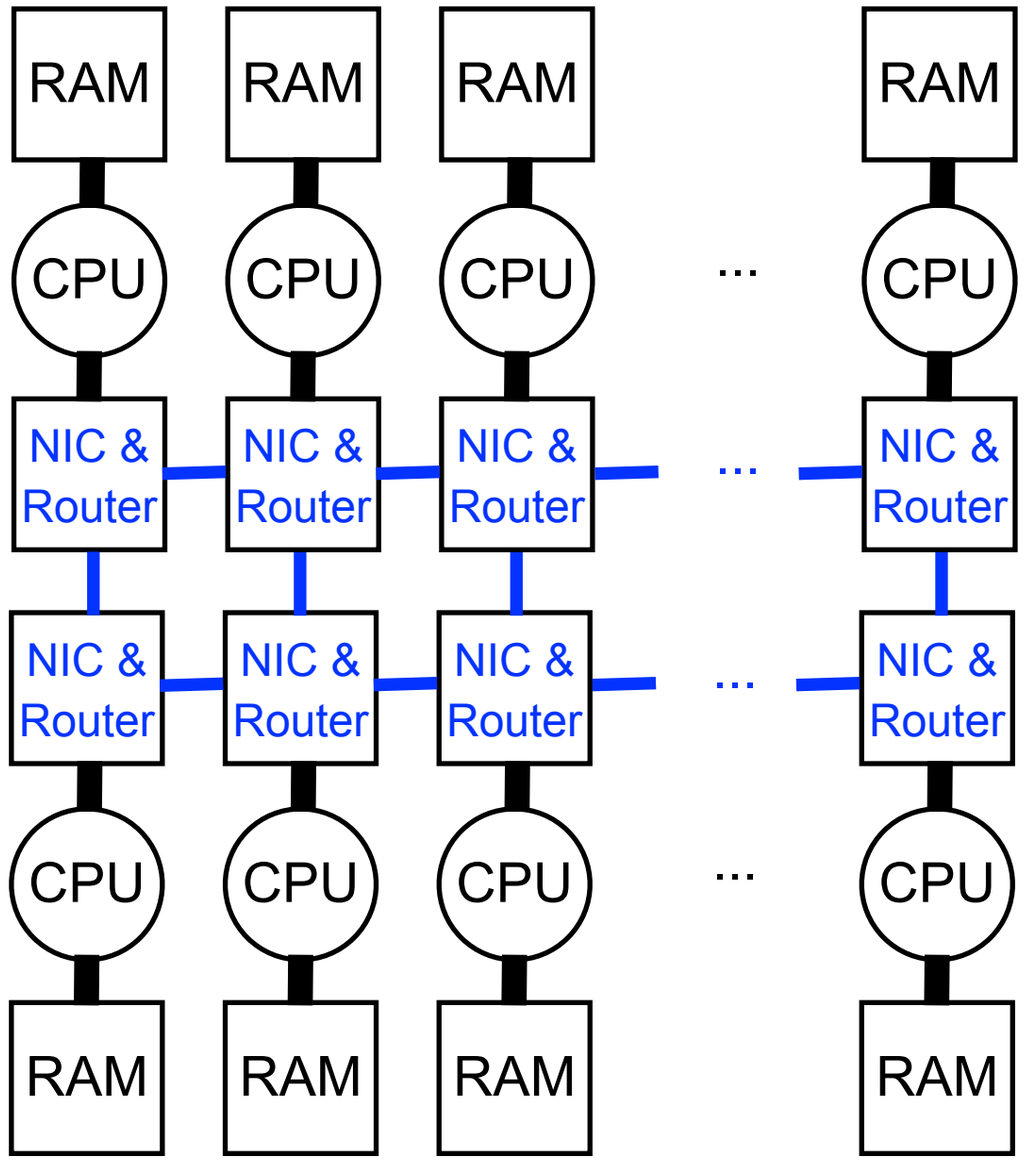


Shared memory

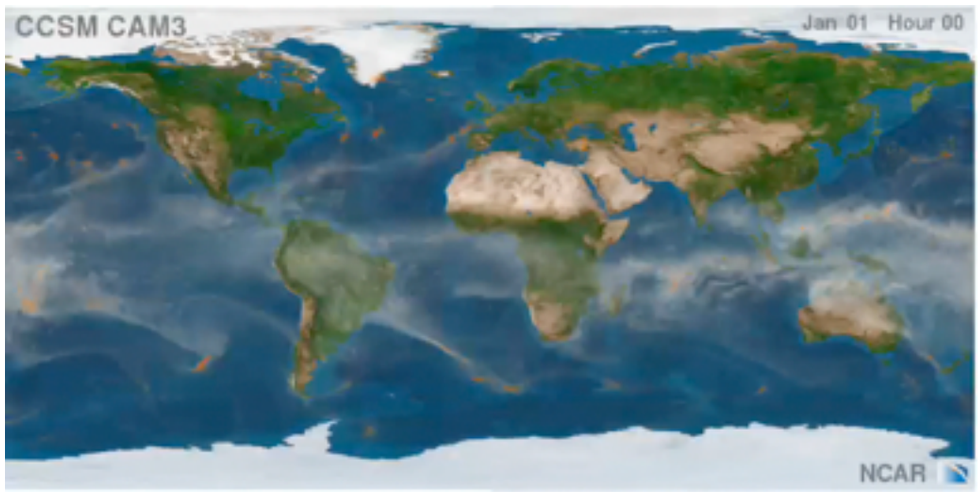


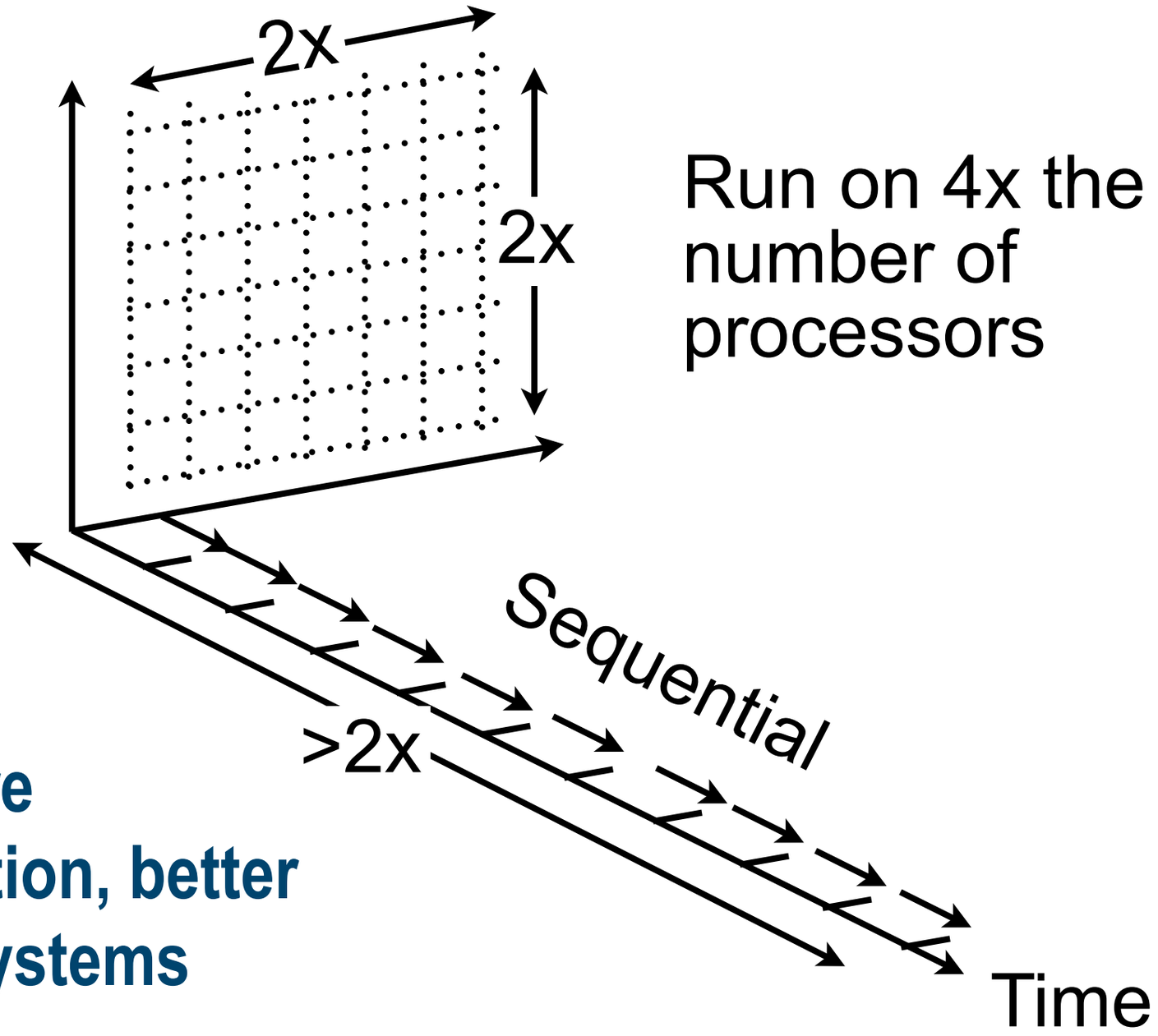
Interconnect types on massively parallel processing (MPP) systems – distributed memory





Larger parallel computers only solve part of the problem





Run on 4x the number of processors

Calculations have to be more efficient: better implementation, better algorithms, more suitable systems

Applications running at scale on Jaguar @ ORNL

Fall 2009

Domain area	Code name	Institution	# of cores	Performance	Notes
Materials	DCA++	ORNL	213,120	**1.9 PF**	2008 Gordon Bell Prize Winner
Materials	WL-LSMS	ORNL/ETH	223,232	**1.8 PF**	2009 Gordon Bell Prize Winner
Chemistry	NWChem	PNNL/ORNL	224,196	1.4 PF	2008 Gordon Bell Prize Finalist
Materials	OMEN	Duke	222,720	860 TF	
Chemistry	MADNESS	UT/ORNL	140,000	550 TF	
Materials	LS3DF	LBL	147,456	442 TF	2008 Gordon Bell Prize Winner
Seismology	SPECFEM3D	USA (multiple)	149,784	165 TF	2008 Gordon Bell Prize Finalist
Combustion	S3D	SNL	147,456	83 TF	
Weather	WRF	USA (multiple)	150,000	50 TF	

Algorithmic motifs and their arithmetic intensity

Arithmetic intensity: number of operations per word of memory transferred

Finite difference / stencil
in S3D and WRF (& COSMO)

Rank-1 update in HF-QMC → **Rank-N update in DCA++**

QMR in WL-LSMS

Sparse linear algebra
Matrix-Vector
Vector-Vector
BLAS1&2

Fast Fourier Transforms
FFTW & SPIRAL

Linpack (Top500)
Dense Matrix-Matrix
BLAS3

O(1) **O(log N)** **O(N)**

Supercomputers are designed for certain algorithmic motifs – which ones?

Relationship between simulations and supercomputer system

Simulations + Theory + Experiment

Science

?

Supercomputer

Model & method of solution

Mapping problem to supercomputer system
Port codes developed on workstations
> **Algorithm re-engineering**
> vectorize codes
> **Software refactoring**
> parallelize codes
> **Domain specific libraries/languages, etc.**
> petascaling and soon exascaling

> Focus on scientific / engineering problem
> Requires interdisciplinary effort / team

Basic numerical libraries
Programming environment
Runtime system
Operating systems
Computer Hardware

Co-Design

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CSCS
Swiss National Supercomputing Centre

Swiss Platform for High-Performance and High-Productivity Computing (HP2C, see www.hp2c.ch)

Scientific problem

Simulations **+ Theory + Experiment**

Swiss Universities / Federal Institutes of Technology
(presently **12 domain science projects** in HP2C Platform)

Interdisciplinary teams consisting of:
> model & method development
> application software design / engineering
> system software (everything between apps & hardware)
> numerical libraries / programming environments
> mapping methods onto computer hardware/ systems
> hardware design / engineering

Supercomputer

Swiss National Supercomputing Center (CSCS) & U. of Lugano (USI)
(collaboration with computer industry: Cray, IBM, Mellanox, SCS)

IT manufacturers – system integrators

Cray Exascale Center of Excellence in Lugano

IBM-ZRL in Rüschlikon

SuperComputing Systems im Technopark

Projects of the HP2C platform (see www.hp2c.ch)

- Gyrokinetic Simulations of Turbulence in **Fusion Plasmas** (ORB5) – Laurent Villard, EPF Lausanne
- **Ab initio Molecular Dynamics** (CP2K) – Jürg Hutter, U. of Zurich
- Computational **Cosmology** on the Petascale – Geoge Lake, U. of Zurich
- Selectome, looking for **Darwinian evolution** in the tree of life – Marc Robinson-Rechavi, Univ. of Lausanne
- **Cardiovascular Systems Simulations** (LifeV) – Alfio Quarteroni, EPF Lausanne
- Modern Algorithms for **Quantum Interacting Systems** (MAQUIS) – Thierry Giamarchi, Univ. of Geneva
- Large-Scale Parallel Nonlinear Optimization for High Resolution **3D-Seismic Imaging** (Petaquacke) – Olaf Schenk, Univ. of Basel
- 3D Models of **Stellar Explosions** – Matthias Liebendörfer, Univ. of Basel
- Large Scale **Electronic Structure Calculations** (BigDFT) – Stefan Gödecker, Univ. of Basel
- Regional **Climate & Weather** Model (COSMO) – Isabelle Bey, ETH Zurich/C2SM
- Lattice-Boltzmann **Modeling of the Ear** – Bastien Chopard, U. of Geneva
- Modeling **humans under climate stress** – Christoph Zollikhofer, U. of Zurich

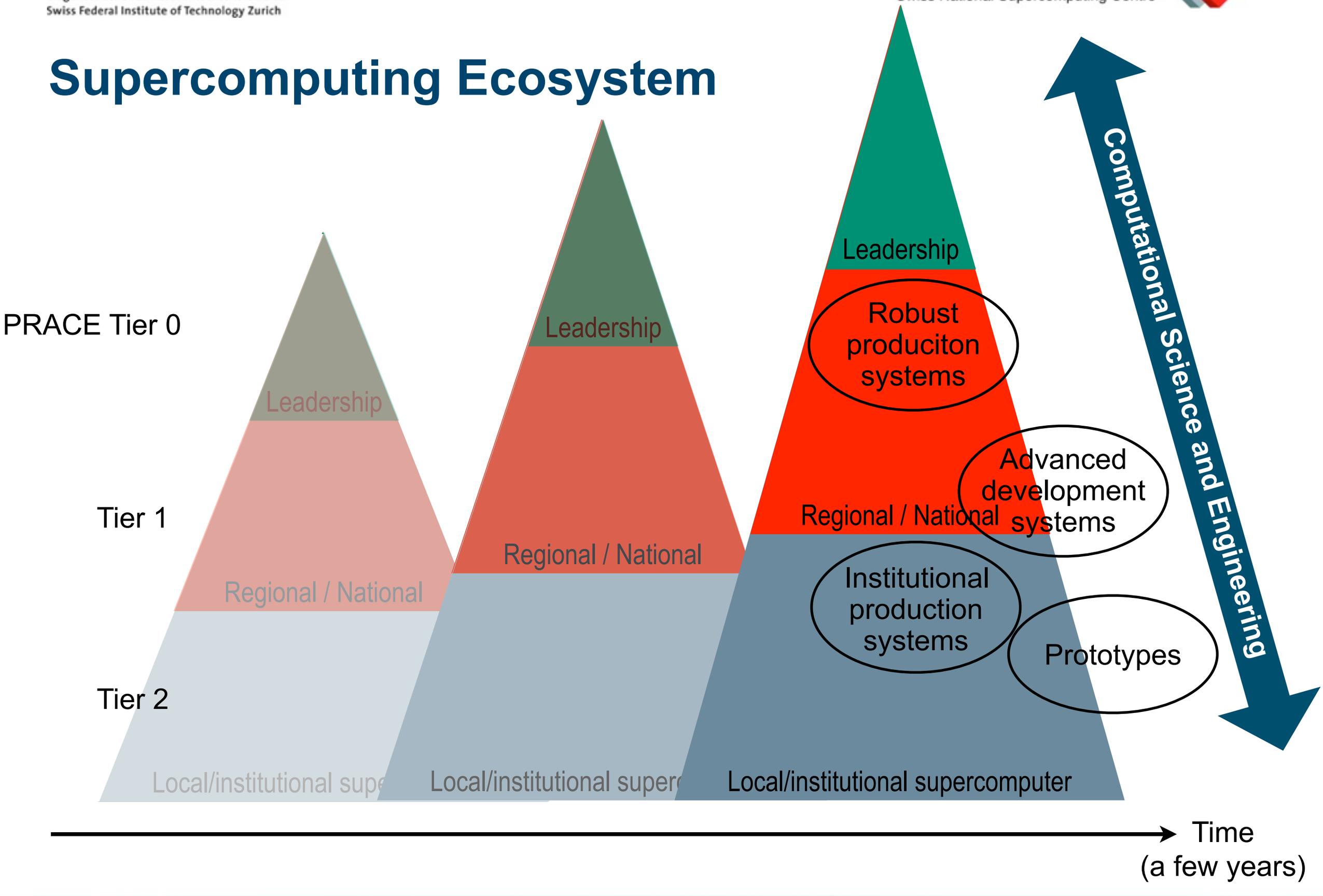
New building under construction in Lugano

- Computer room area (1500 m^2)
- Power & cooling ~ 12 MW (upgradable) (**PUE ~ 1.2**)
- Proximity to academic institution (USI)
- Extensible
- Facilitate seamless computer hardware upgrades/changes

http://webcam.cscs.ch/webcam/

Current CSCS building in Manno: **PUE ~1.7** – i.e. 1 MW delivered to computer requires 1.7 MW electrical power

Supercomputing Ecosystem



Swiss HP2C Platform
High-risk & high-impact projects of the
(www.hp2c.ch)

New procurement
Cray XT3
1'100 processors

Upgrade
Cray XT3
1'664 proc.

Dual core upgrade
Cray XT3
3'328 cores

Upgrade
Cray XT5
14'752 cores

Hex-core upgrade
22'128 cores

"Final" upgrade
Cray XT5

Procurement
next generation
supercomputer
HPCN initiative

Begin construction
of new building

New
building
complete

2005
2006
2007
2008
2009
2010
2011
2012
2013

Swiss HP2C Platform

Elements of the Swiss High-Performance and Networking (HPCN) Initiative & beyond

- Swiss Platform for HP2C (2009-12):
  - Simulation systems that make effective use of next gen. supercomputers
  - Establish HPC in CSE programs at Swiss universities
- Develop new building infrastructure by 2012:
  - Very advanced infrastructure that is energy efficient and supports a machine footprint that is about a factor 10 larger than today
- Hardware Investments (2009-11 and 2012-14):
  - Goal for CSCS is to host systems with performance of 20-25% compared to largest leadership system in the world
- Successor to HP2C (2013-16):
  - Focus on co-design targeted at scientific problems
- Next generation hardware investments (2015-17)
  - System generation leading towards exa-scale

Zusammenfassung und Schlussfolgerungen

- Wissenschaftliches Rechnen wird weiterhin die Zukunft der Informationstechnologie mitbestimmen
- Das Mooresche Gesetz ist nicht alleiniger Grund für die Leistungsverbesserung der Rechner und wird an Bedeutung verlieren – neue Gelegenheiten für Quereinsteiger!
- Physikalische Aspekte der Rechnungen gewinnen wieder an Bedeutung
- (Energie)Effizienz verlangt dass die Simulationssysteme den Problemen angepasste werden – Lösungsmethoden, Algorithmen, Software, und Hardware müssen aufeinander abgestimmt sein
- Nationale HPCN Initiative investiert in Leute (in der ganzen Schweiz), sowie in eine energieeffiziente Gebäudeinfrastruktur (in Lugano) und in ein ausgewogenes Ökosystem von Supercomputern – d.h. in eine Forschungsinfrastruktur für die Wissenschaft, von der aber auch der Technologiestandort Schweiz profitiert!

FRAGEN / KOMMENTARE?