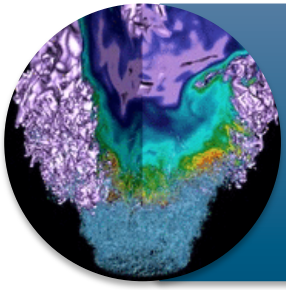
Challenges in HPC

U.S. DEPARTMENT OF ENERGY
Office of Science

National Energy Research
Scientific Computing Center

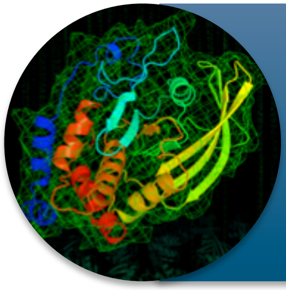
Lawrence Berkeley
National Laboratory

Thrusts in High Performance Computing



Science at Scale

Petaflops to Exaflops



Science through Volume

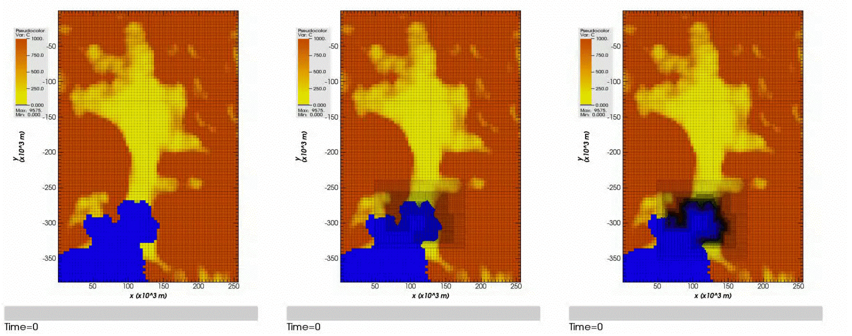
Thousands to Millions of Simulations

Science in Data

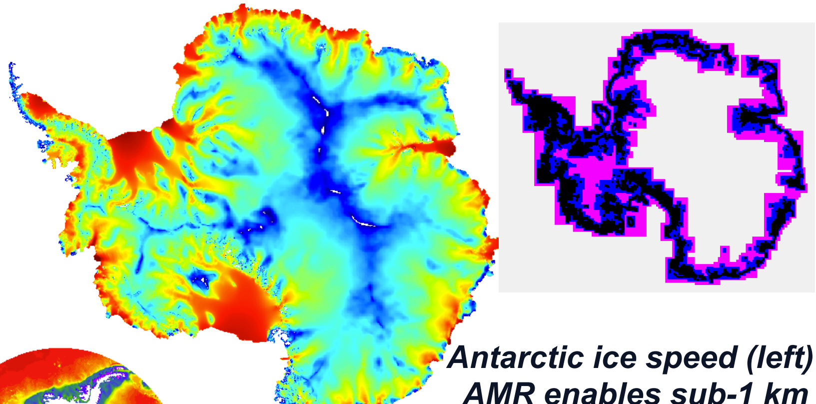
Petabytes to Exabytes of Data

Science at Scale: Simulations Aid in Understanding Climate Impacts

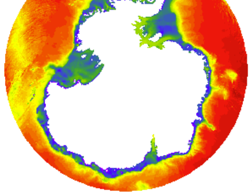
- **Warming ocean and Antarctic ice sheet key to sea level rise**
- **BISICLES ice sheet model uses AMR for ice-ocean interface.**
  - Dynamics very fine resolution (AMR)
  - Antarctica still very large (scalability)
  - Multi-institution (LANL, LBNL)
- **Ongoing collaboration to couple ice sheet and ocean models**
  - 10M ALCC H t NERSC



BISICLES Pine Island Glacier simulation – mesh resolution crucial for grounding line behavior.



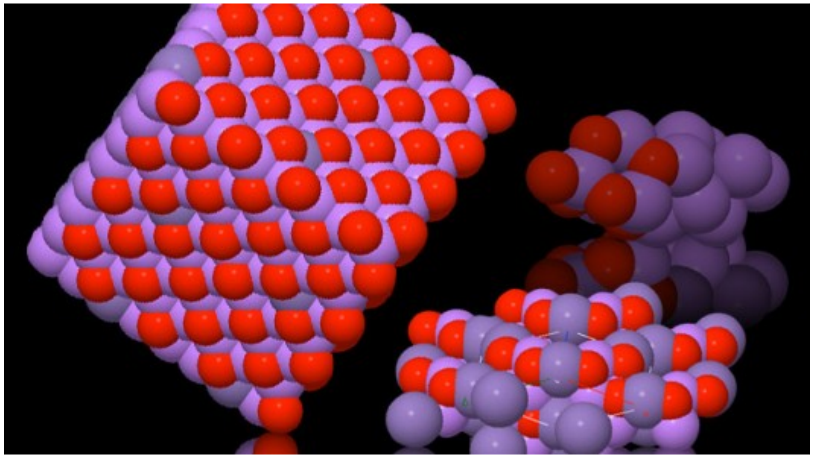
Antarctic ice speed (left): AMR enables sub-1 km resolution (black, above) *(Using NERSC's Hopper)*

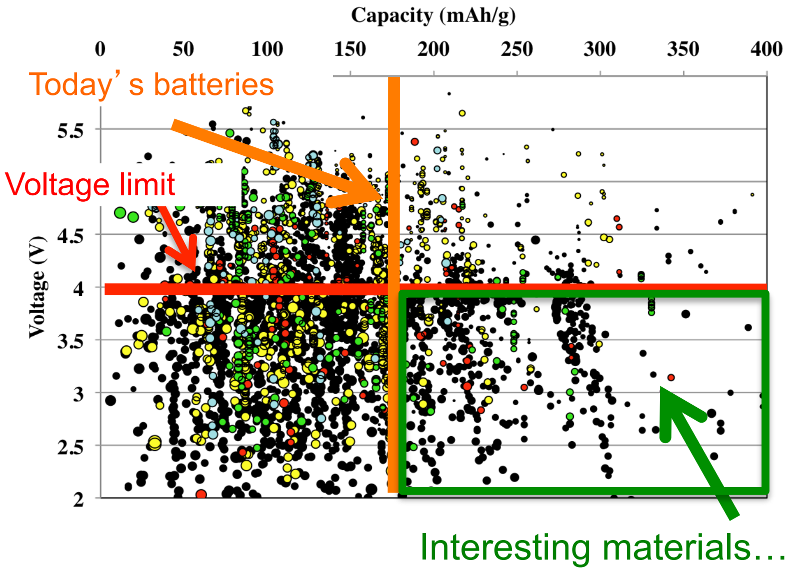


Enhanced POP ocean model solution for coupling to ice

Science through Volume: Large Numbers of Simulations for Materials

- Tens of thousands of simulations are used to screen related materials for use in battery design and other domains
- Goal: cut in half the 18 year from design to manufacturing





Today’s batteries

Voltage limit

Interesting materials…

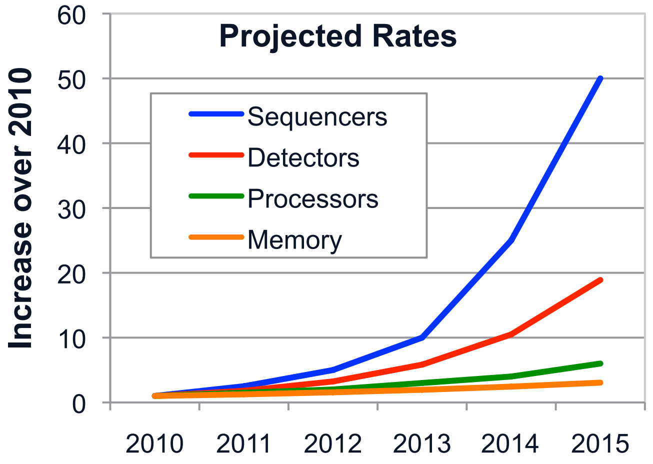
Materials Project, Gerd Ceder PI (MIT): *website has a database materials from simulations, e.g., over 20,000 potential battery materials.*

PIs: Gerd Ceder, MIT and Kristin Persson, LBNL

Science in Data: Automated Image Analysis in Astronomy

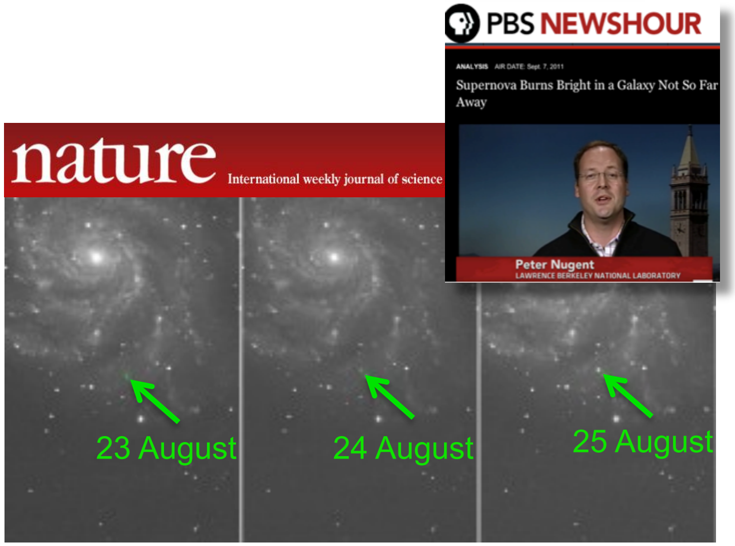
Data from scientific instruments is growing exponentially

- NERSC in 3 Nobel prizes, and 3 Science “best of decade” (CMB and Genomics)
- Far outpacing processor and memory performance growth



Astrophysics discover early nearby supernova

- Palamor Transient Factory runs machine learning algorithms on ~300GB/night delivered by ESnet “science network”
- Rare glimpse of a supernova within 11 hours of explosion, 20M light years away
- Telescopes world-wide redirected

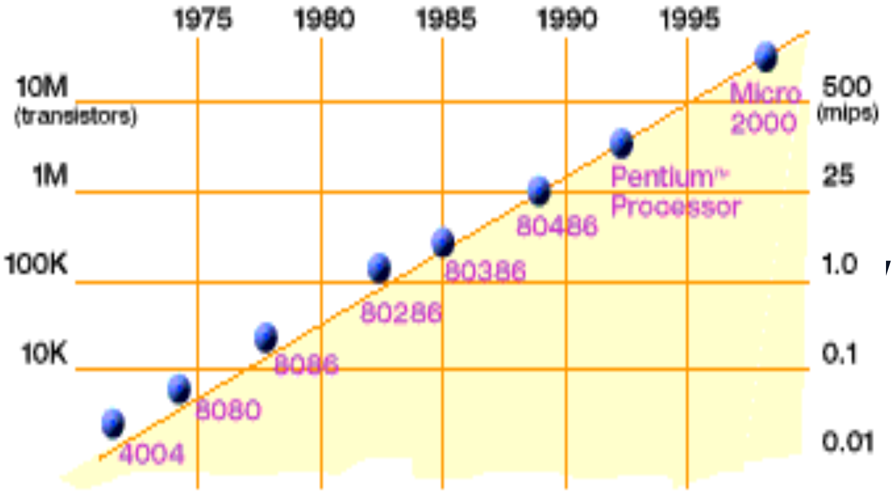


Biggest Challenge: Power

- **Engineering View**
  - Minimize power per computation
  - 1 Exaflop in 20 MW?
  - Goal: 1,000-fold performance increase with 5X power consumption by 2020

- **Programming View**
  - Past: minimize Flops
  - Future: minimize data movement

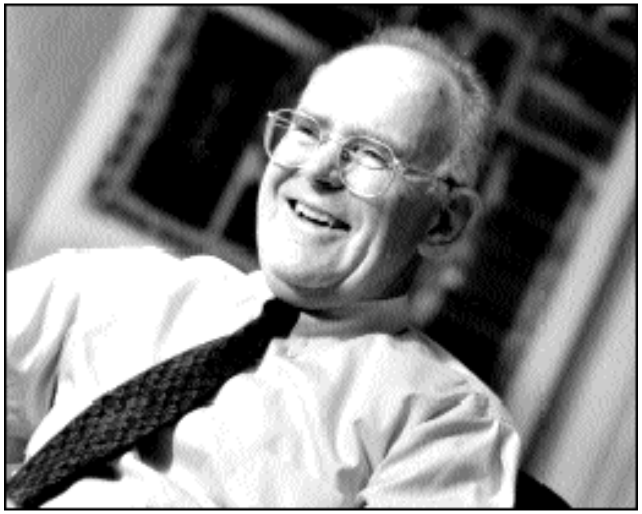
Moore’s Law



2X transistors/Chip Every 1.5 years

Called “Moore’s Law”

Microprocessors have become smaller, denser, and more powerful.

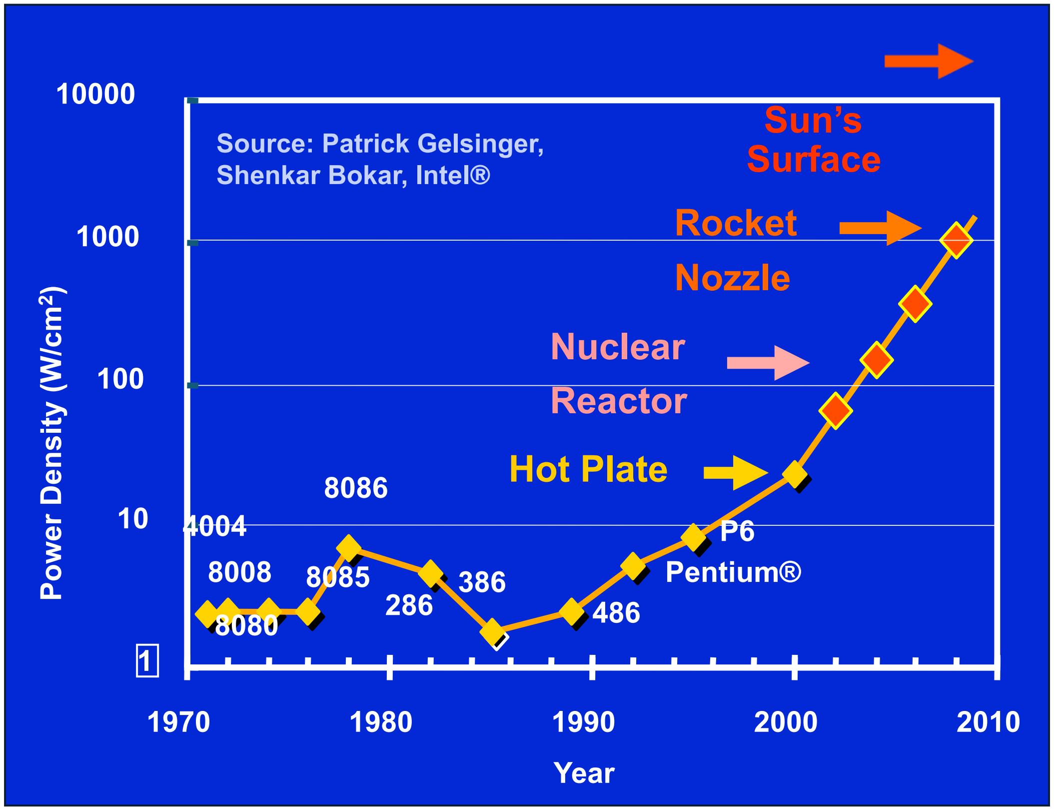


Gordon Moore (co-founder of Intel) predicted in 1965 that the transistor density of semiconductor chips would double roughly every 18 months.

Slide source: Jack Dongarra

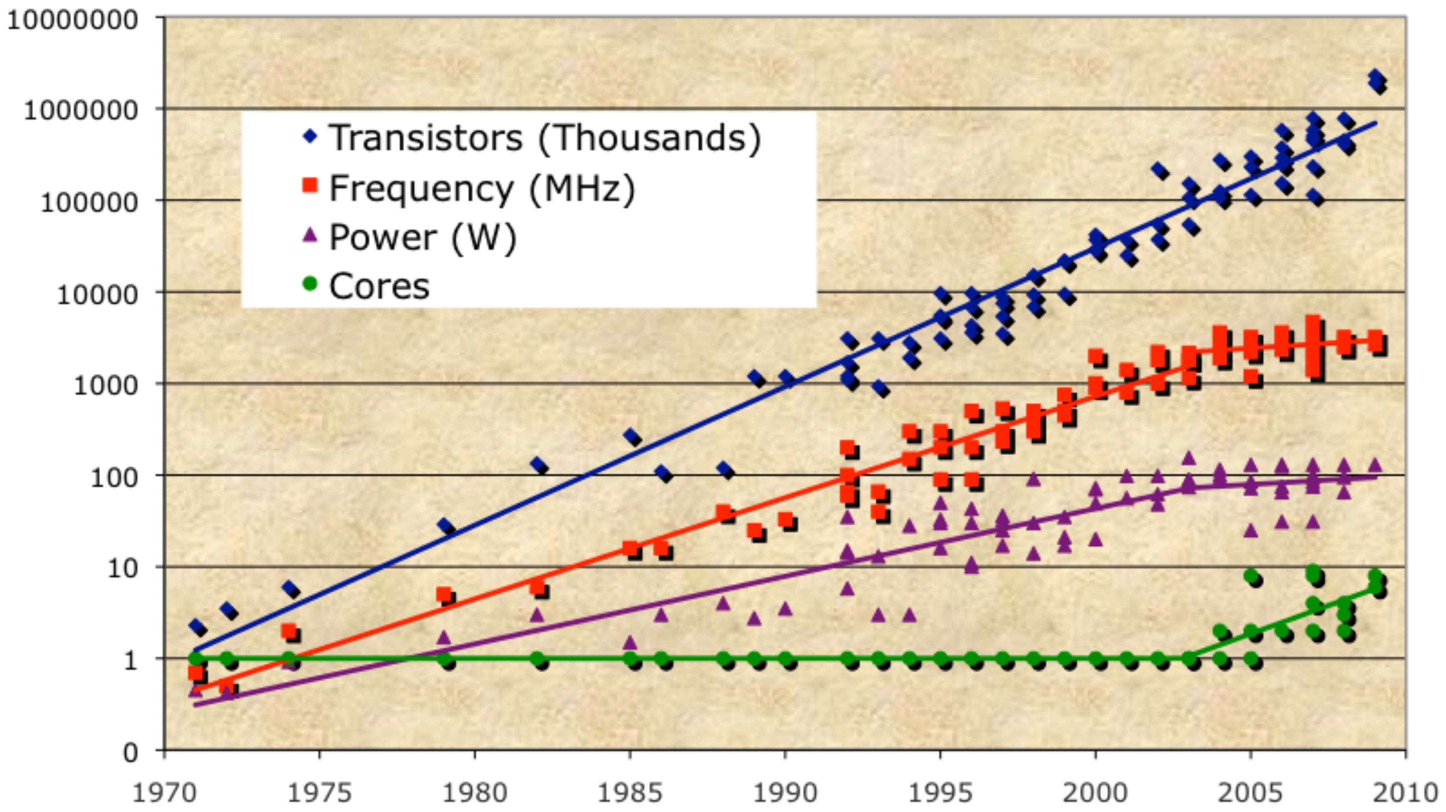
Power Density Limits Serial Performance

- Concurrent systems are more power efficient
  - Dynamic power is proportional to V^2fC
  - Increasing frequency (f) also increases supply voltage (V) → cubic effect
  - Increasing cores increases capacitance (C) but only linearly
  - Save power by lowering clock speed



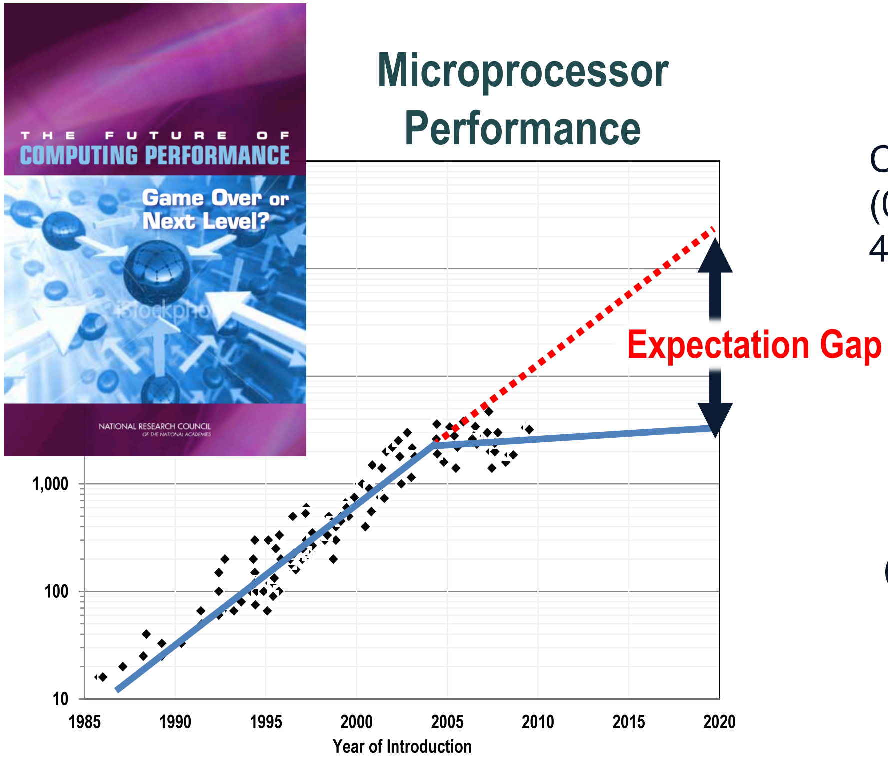
- High performance serial processors waste power
  - Speculation, dynamic dependence checking, etc. burn power
  - Implicit parallelism discovery
- More transistors, but not faster serial processors

Revolution in Processors

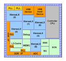


- Chip density is continuing increase ~2x every 2 years
- Clock speed is not
- Number of processor cores may double instead
- Power is under control, no longer growing

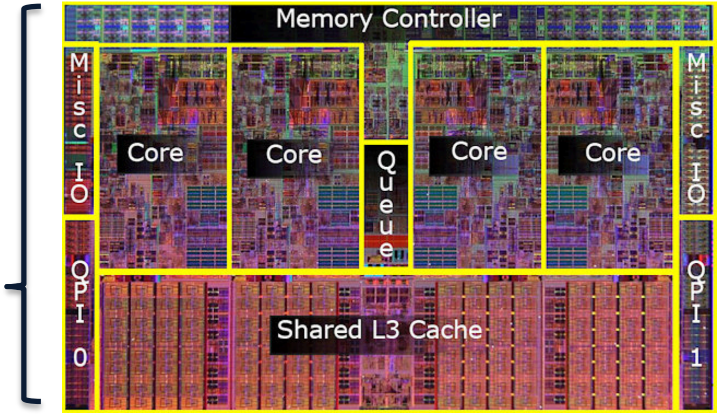
Major Innovations Needed to Sustain Performance Growth



Cell phone
(0.1 Watt,
4 Gflop/s)



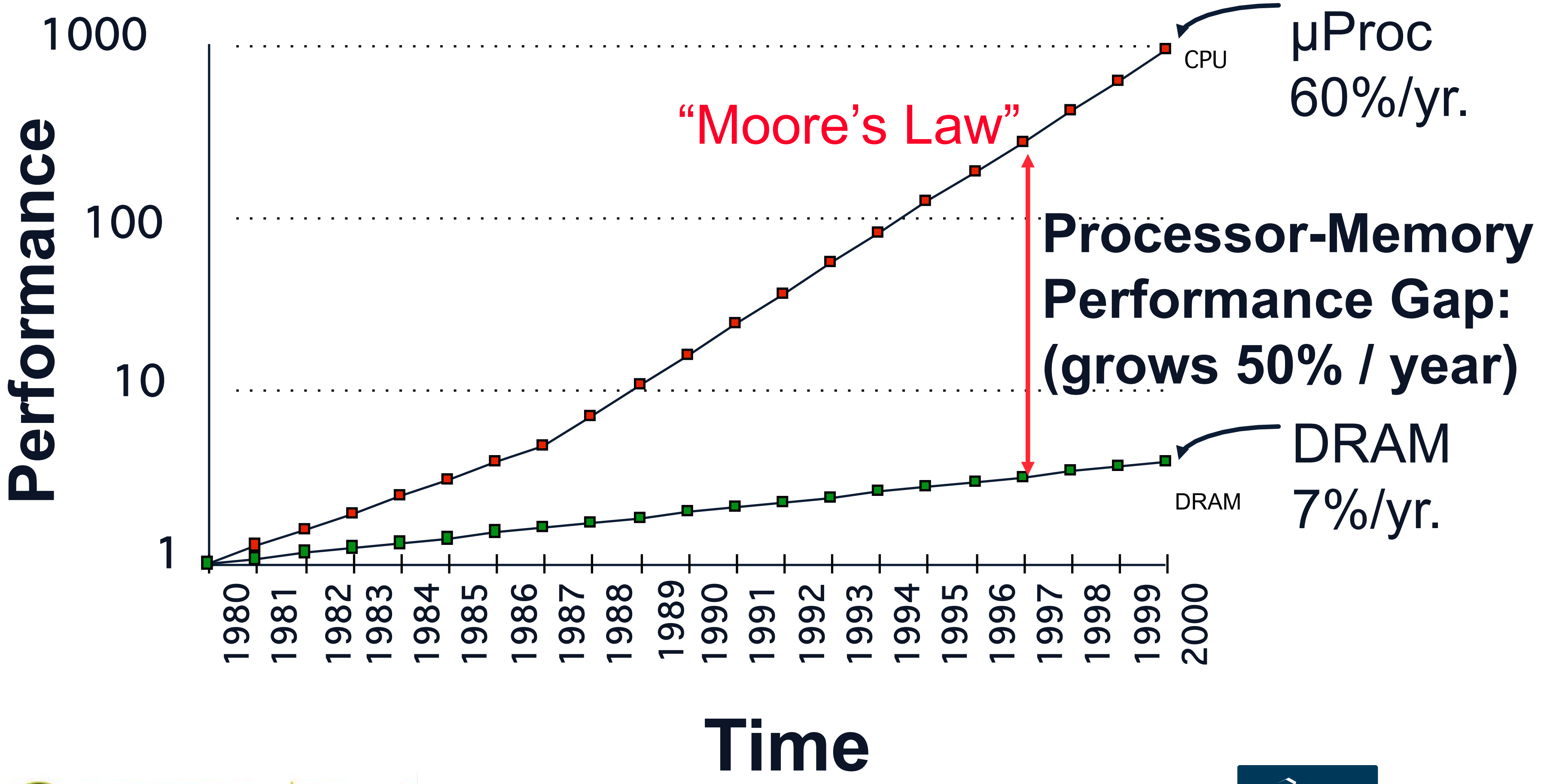
Server
(100 Watts,
50 Gflop/s)



- Processor performance growth is limited by power
- Exascale computers (1000x Hopper) in next decade:
  - Manycore processors using graphics, games, embedded cores, or other low power designs offer 100x in power efficiency
  - Facilities will need 10x more power (Hopper is 3MW)

Processor-DRAM Gap (latency)

Goal: find algorithms that minimize communication, not necessarily arithmetic

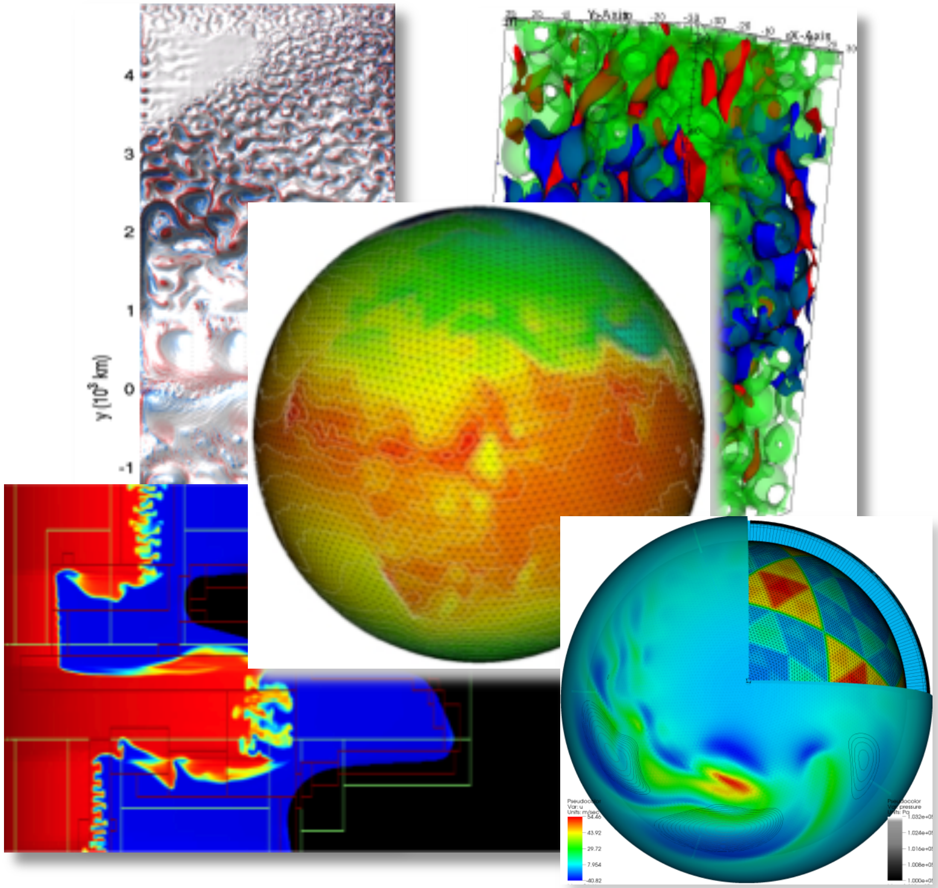


Can Accelerators Solve the Problem?

- **Accelerator configuration**
  - Many small, energy-efficient cores (GPUs)
  - GPU have private memory space
  - Attached to motherboard via PCI interface currently
- **Case for heterogeneity**
  - Accelerators are theoretically very fast
  - Much better theoretical Flop/Watt
- **Challenges**
  - Need one fat core (at least) for running the OS
  - Data movement from main memory to GPU memory kills performance
  - Programmability is very poor
  - Most codes will require extensive overhauls

Data: Getting bigger all the time

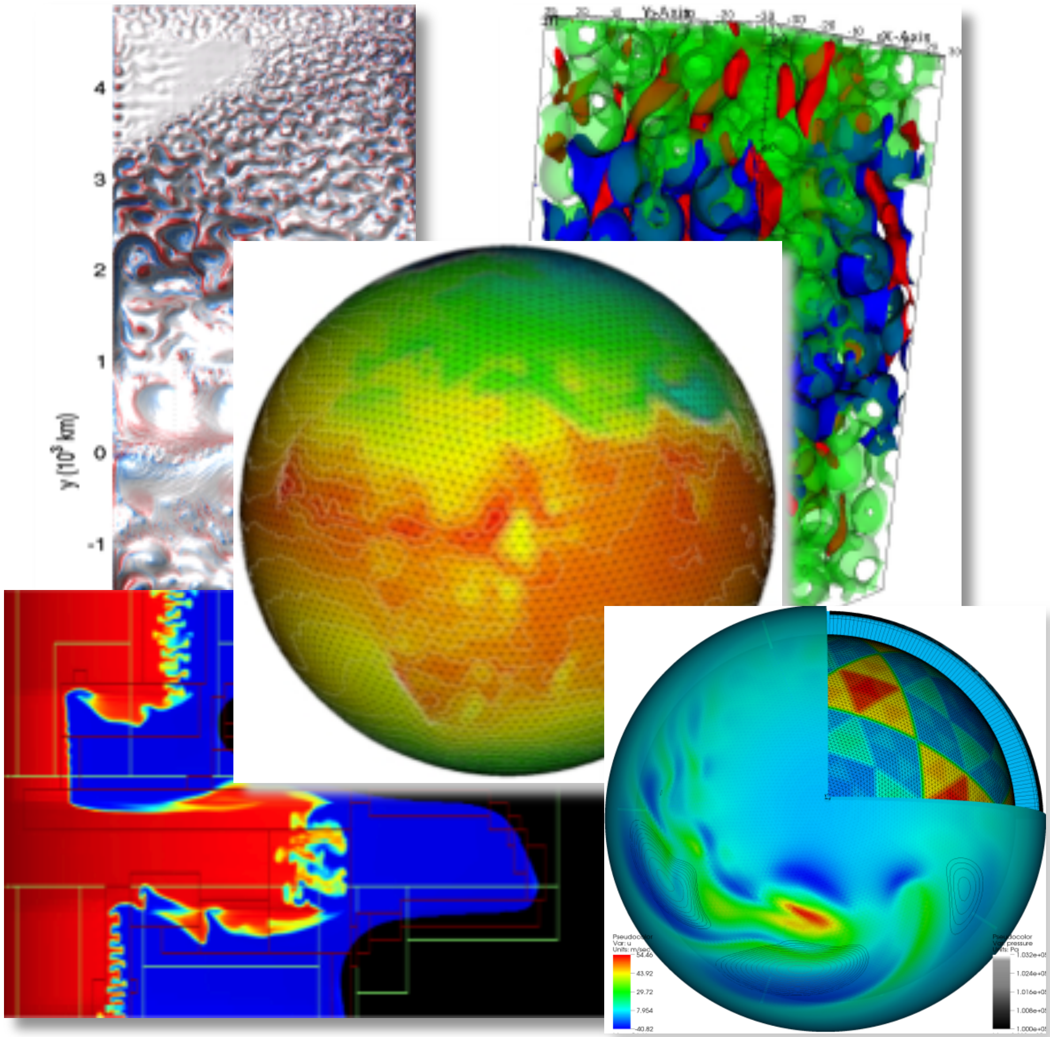
- I/O needs growing each year in scientific community
- For our largest users I/O parallelism is mandatory
- I/O remains a bottleneck for many users
- Early 2011 – Hopper: 2 PB /scratch (we thought that was huge!)
- New systems at TACC and NCAR have ~ 18 PB /scratch!!!!



Images from David Randall, Paola Cessi, John Bell, T Scheibe

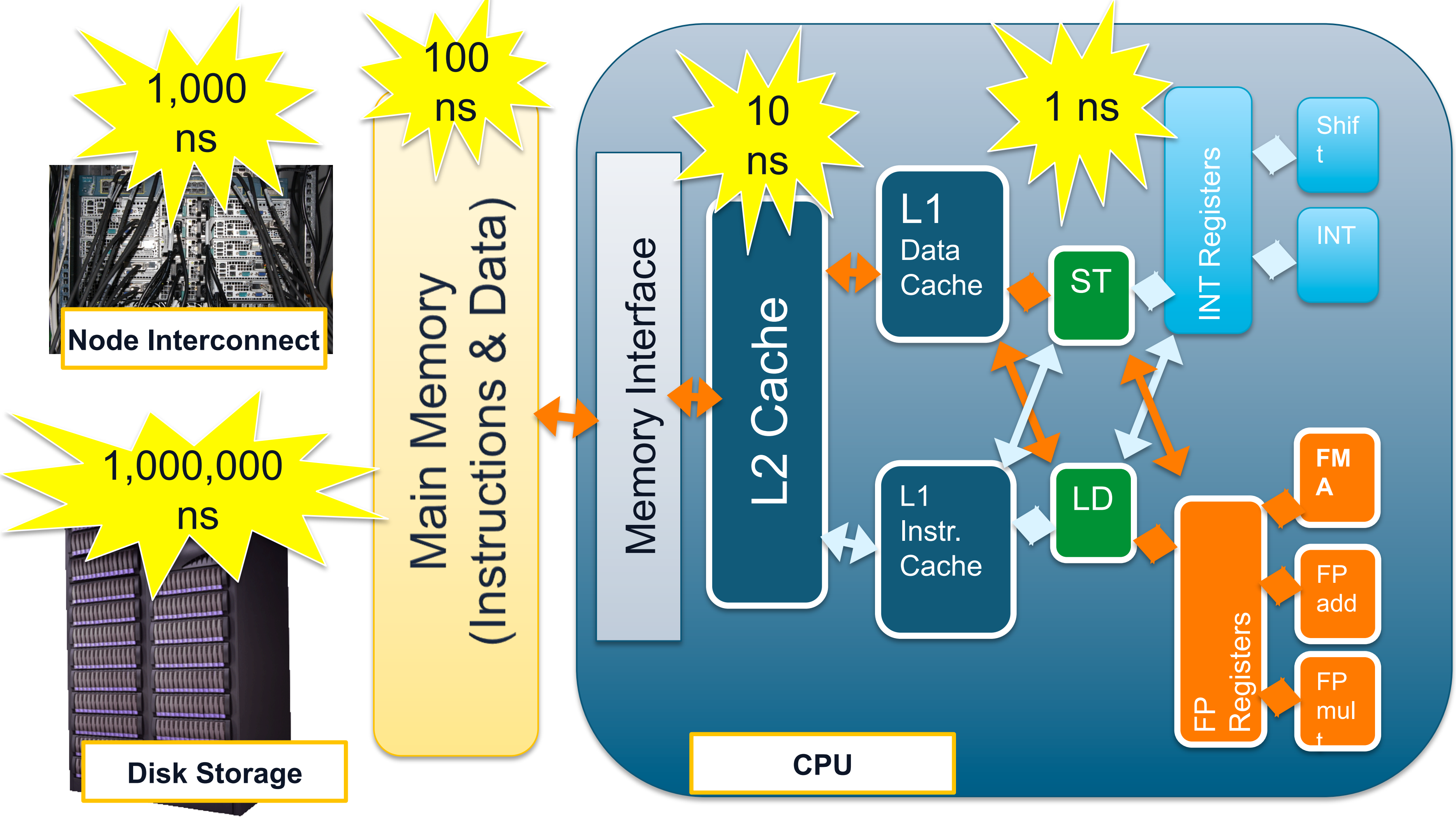
Why is Parallel I/O for science applications difficult?

- Scientists think about data in terms of how a system is represented in the code: as grid cells, particles, …
- Ultimately, data is stored on a physical device
- Layers in between the application and the device are complex and varied
- I/O interfaces and configurations are arcane and complicated



Images from David Randall, Paola Cessi, John Bell, T Scheibe

Latencies



1,000 ns

Node Interconnect

1,000,000 ns

Disk Storage

100 ns

Main Memory (Instructions & Data)

Memory Interface

10 ns

L2 Cache

L1 Data Cache

L1 Instr. Cache

1 ns

ST

LD

INT Registers

Shift

INT

FMA

FP add

FP mult

FP Registers

CPU