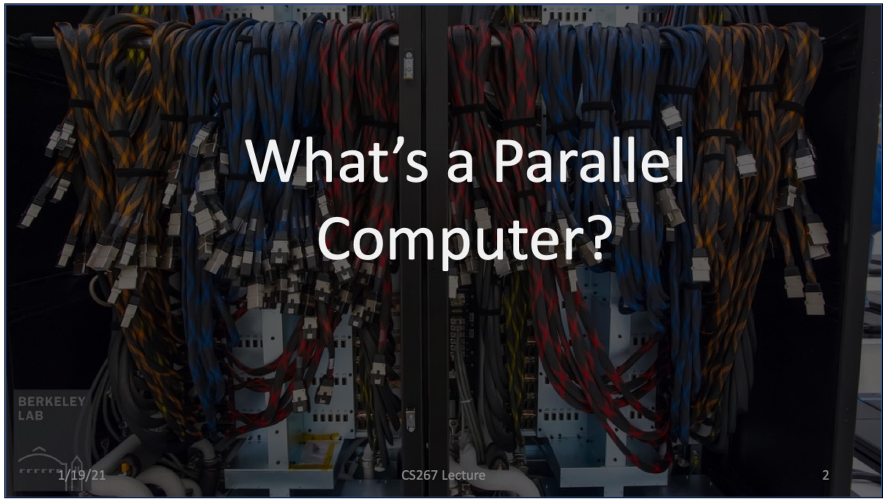
Applications of Parallel Computers

CS267/E233

https://sites.google.com/lbl.gov/cs267-spr2021

Survey: https://forms.gle/XNKNXhhH4mtibMtR8

What’s a Parallel Computer?



It’s all about the need for speed

Parallel Computing: Faster Solutions

Using multiple processors in parallel to solve problems more quickly than with a single processor

Compute the prime factors of 1 billion numbers:

45	12	66	. . .	13

If we had 1 million processors…

Sum (reduction) in parallel

- Add n values

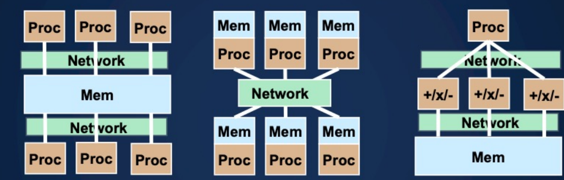
Serial: O(n)

Parallel: O(log n)

Uses n processors!

- Takes advantage of associativity in +

What is a Parallel Computer?

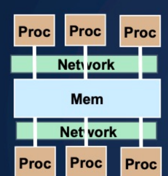


Shared Memory (SMP) or Multicore

High Performance Computing (HPC) or Distributed Memory

Single Instruction Multiple Data (SIMD)

What is a Parallel Computer?



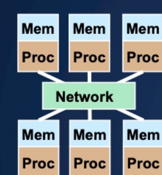
Shared Memory (SMP) or Multicore

A **shared memory multiprocessor** (SMP*) by connecting multiple processors to a single memory system

A **multicore processor** contains multiple processors (cores) on a single chip

* Technically, SMP stands for "Symmetric Multi-Processor"

What is a Parallel Computer?



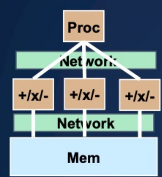
High Performance Computing (HPC) or Distributed Memory

A **distributed memory multiprocessor** has processors with their own memories connected by a high speed network

Also called a **cluster**

A **high performance computing (HPC)** system contains 100s or 1000s of such processors (nodes)

What is a Parallel Computer?

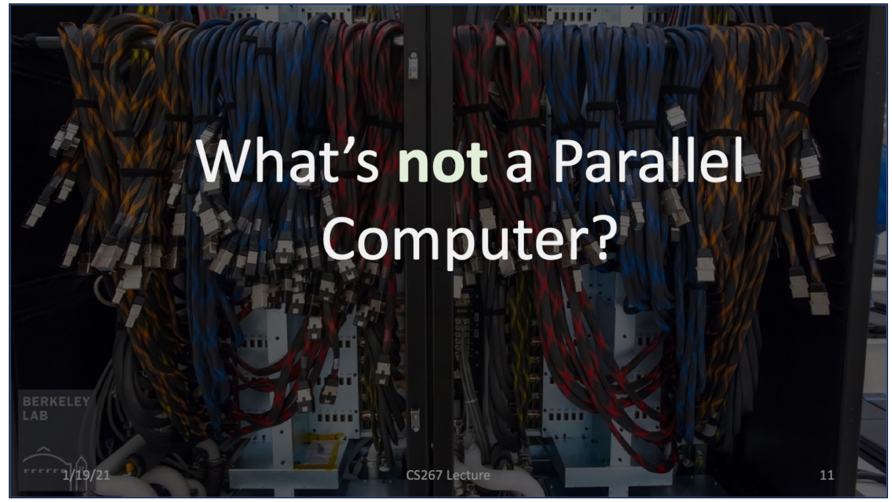


Single Instruction Multiple Data (SIMD)

A **Single Processor Multiple Data (SIMD)** computer has multiple processors (or functional units) that perform the same operation on multiple data elements at once

Most single processors have **SIMD units** with ~2-8 way parallelism

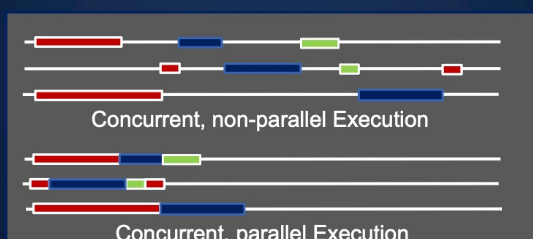
Graphics processing units (**GPUs**) use this as well



What's **not** a Parallel Computer?

Concurrency vs. Parallelism

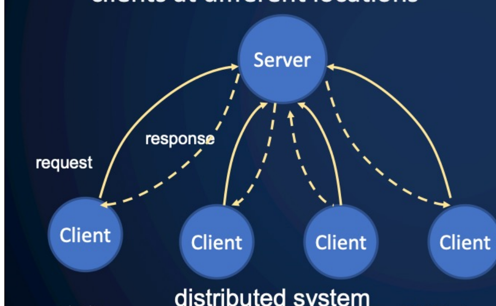
- Concurrency: multiple tasks are ***logically*** active at one time.
- Parallelism: multiple tasks are ***actually*** active at one time.



Slide Source: Tim Mattson, Intel

Parallel Computer vs. Distributed System

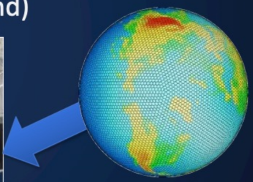
- A distributed system is ***inherently*** distributed, i.e., serving clients at different locations



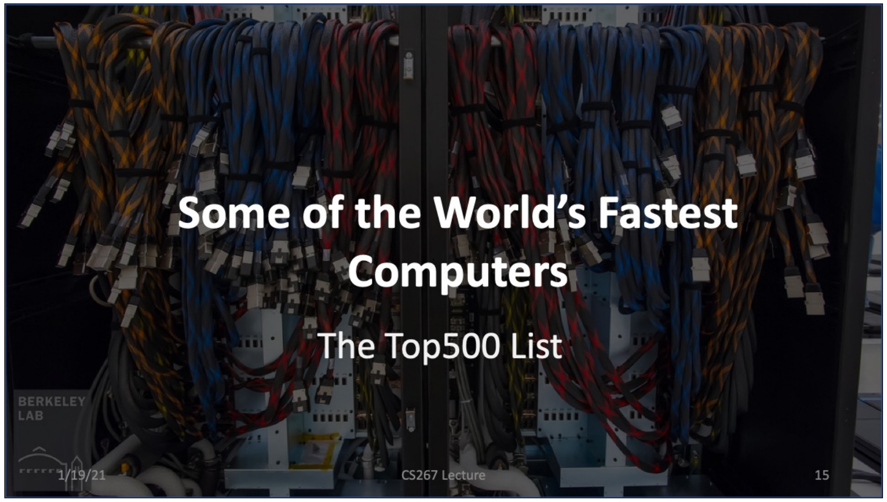
A parallel computer may use ***distributed memory*** (multiple processors with their own memory) for more performance

The Fastest Computers (for Science) Have Been Parallel for a Long Time

- Fastest Computers in the world: top500.org
- LBNL's Cori Computer has over 680,000 cores and ~30 Petaflops (10^{15} math operations / second)



Supercomputing is done by parallel programming



Some of the World's Fastest Computers

The Top500 List

Units of Measure for HPC

- **High Performance Computing (HPC) units are:**
  - **Flop: floating point operation, usually double precision unless noted**
  - **Flop/s: floating point operations per second**
  - **Bytes: size of data (a double precision floating point number is 8 bytes)**
- **Typical sizes are millions, billions, trillions…**

	Kilo	Kflop/s = 10^3 flop/sec	Kbyte = 10^3 ~ 2^{10} = 1024 bytes (KiB)
	Mega	Mflop/s = 10^6 flop/sec	Mbyte = 10^6 ~ 2^{20} bytes (MiB)
	Giga	Gflop/s = 10^9 flop/sec	Gbyte = 10^9 ~ 2^{30} bytes (GiB)
	Tera	Tflop/s = 10^{12} flop/sec	Tbyte = 10^{12} ~ 2^{40} bytes (TiB)
	Peta	Pflop/s = 10^{15} flop/sec	Pbyte = 10^{15} ~ 2^{50} bytes (PiB)
Goal	Exa	Eflop/s = 10^{18} flop/sec	Ebyte = 10^{18} ~ 2^{60} bytes (EiB)
	Zetta	Zflop/s = 10^{21} flop/sec	Zbyte = 10^{21} ~ 2^{70} bytes (ZiB)
	Yotta	Yflop/s = 10^{24} flop/sec	Ybyte = 10^{24} ~ 2^{80} bytes (YiB)

- **Current fastest (public) machines are petaflop systems**
  - **Up-to-date list at www.top500.org**

The TOP500 Project

- 500 most powerful computers in the world
- Updated twice a year:
  - ISC' xy in June in Germany
  - SCxy in November in the U.S.
- All information available from the TOP500 web site at: www.top500.org

Yardstick: Floating Point Operations per Second (FLOP/s) Rmax of Linpack
- Solve Ax=b, Matrix A is dense with random entries
- Dominated by dense matrix-matrix multiply

TOP500® SUPERCOMPUTER SITES

#	Site	Manufacturer	Computer	Country	Cores	Rmax [Pflops]	Power [MW]
1	RIKEN Center for Computational Science	Fujitsu	Fugaku Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Japan	7,630,848	442.0	29.9
2	Oak Ridge National Laboratory	IBM	Summit IBM Power System, P9 22C 3.07GHz, Mellanox EDR, NVIDIA GV100	USA	2,414,592	148.6	10.1
3	Lawrence Livermore National Laboratory	IBM	Sierra IBM Power System, P9 22C 3.1GHz, Mellanox EDR, NVIDIA GV100	USA	1,572,480	94.6	7.4
4	National Supercomputing Center in Wuxi	NRCPC	Sunway TaihuLight NRCPC Sunway SW26010, 260C 1.45GHz	China	10,649,600	93.0	15.4
5	NVIDIA Corporation	NVIDIA	Selene DGX A100 SuperPOD, AMD 64C 2.25GHz, NVIDIA A100, Mellanox HDR	USA	555,520	63.5	2.65
6	National University of Defense Technology	NUDT	Tianhe-2A ANUDT TH-IVB-FEP, Xeon 12C 2.2GHz, Matrix-2000	China	4,981,760	61.4	18.5
7	Forschungszentrum Jülich (FZJ)	Atos	JUWELS Booster Module BullSequana XH2000, AMD EPYC 24C 2.8GHz, NVIDIA A100, Mell. HDR	Germany	449,280	44.1	1.76
8	Eni S.p.A	Dell EMC	HPC5 PowerEdge C4140, Xeon 24C 2.1GHz, NVIDIA T. V100, Mellanox HDR	Italy	669,760	35.5	2.25
9	Texas Advanced Computing Center / Univ. of Texas	Dell	Frontera Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox HDR	USA	448,448	23.5	
10	Saudi Aramco	HPE	Dammam-7 Cray CS-Storm, Xeon G. 20C 2.5GHz, NVIDIA T. V100, IB HDR 100	Saudi Arabia	672,520	22.4	

Summit (#1 in US) System Overview

OAK RIDGE National Laboratory

System Performance
- Peak performance of 200 petaflops for modeling & simulation
- Peak of 3.3 ExaOps for data analytics and artificial intelligence

Each node has
- 2 IBM POWER9 processors
- 6 NVIDIA Tesla V100 GPUs
- 608 GB of fast memory
- 1.6 TB of NVMe memory

The system includes
- 4608 nodes
- Dual-rail Mellanox EDR InfiniBand network
- 250 PB IBM Spectrum Scale file system transferring data at 2.5 TB/s

Fugaku (#1 Worldwide) System Overview

RIKEN Center for Computational Science (R-CCS)

System Performance
- Peak performance of 442 petaflops (per TOP500 Rmax),
- 2.0 EFLOPS on a different mixed-precision benchmark

Each node has
- Fujitsu A64FX CPU (48+4 cores) per node
- HBM2 32 GiB

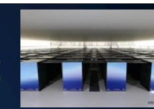
The system includes
- 158,976 nodes
- Custom Tofu Interconnect D
- 1.6 TB NVMe SSD/16 nodes (L1)
- 150 PB Lustre Filesystem (L2)
- Cloud storage (L3)

Cori at NERSC (#1 in Berkeley, CA)

Phase 1	**Peak:** 2.8 PFlop/s	**Proc:** 3800 Haswell (ph 1) **Cores:** 2.3 GHz, 16/proc **Proc Peak:** 0.6 TFlop/s	**Memory:** 0.3 PB **Storage:** shared
Phase 2	**Peak:** 28PFlop/s **LinPack: 14 Pflops/s** **Power:** 4 MW	**Processor:** 9300 KNL **Cores:** 1.4 GHz, 68/proc **Node Peak:** 3 TFlop/s	**Memory:** 1.31 PB **MemBW:** 1 PB/s HBM,DDR **Storage:** 28 PB

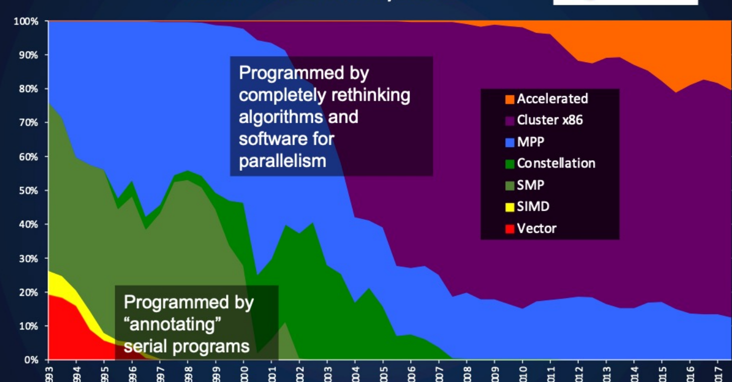
Performance History and Projection

Other Algorithms / Arithmetic

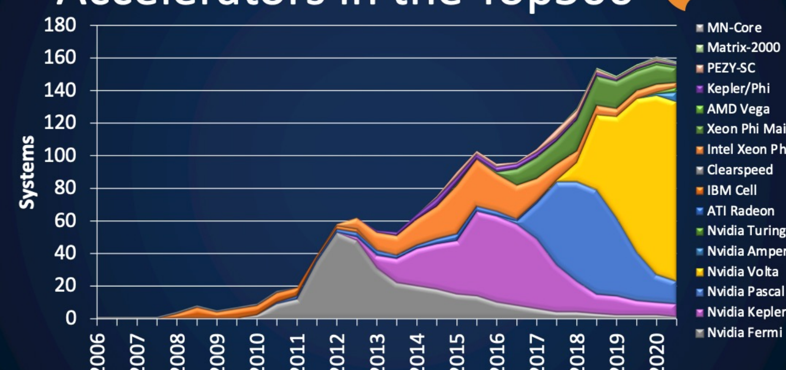


- Mixed precision iterative refinement approach solved a matrix of order 16,957,440 on Fugaku.
  - Composed of nodes made up of Fujitsu's ARM A64fx Processor
  - The run used 158,976 nodes of Fugaku, 7,630,848 cores
  - Used a random matrix with large diagonal elements to insure convergence.
- Mixed precision HPL achieved 2.004 Eflop/s
  - 4.5 X over DP precision HPL (442 PFLOPS).
  - 67 Gflops/Watt
- Same accuracy compared to full 64 bit precision

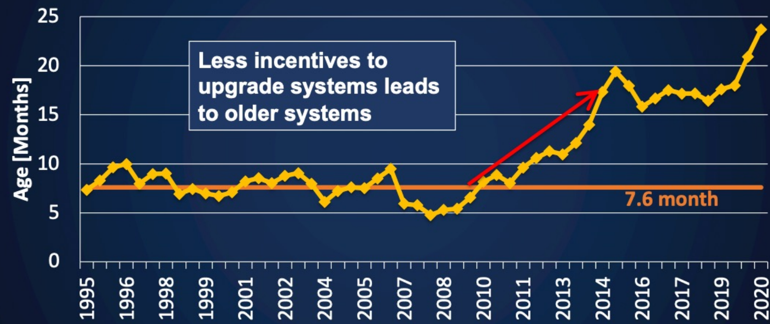
From Vector Supercomputers to Massively Parallel Accelerator Systems



Accelerators in the Top500



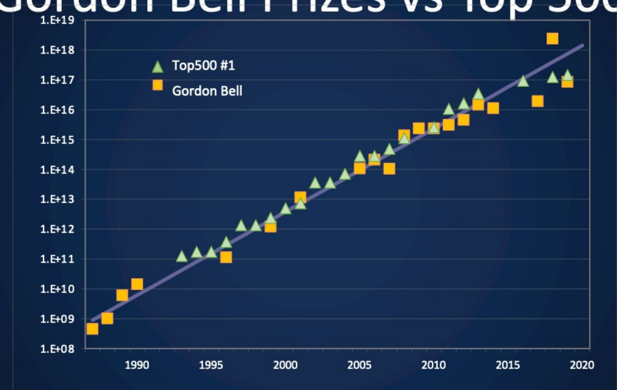
Average System Age

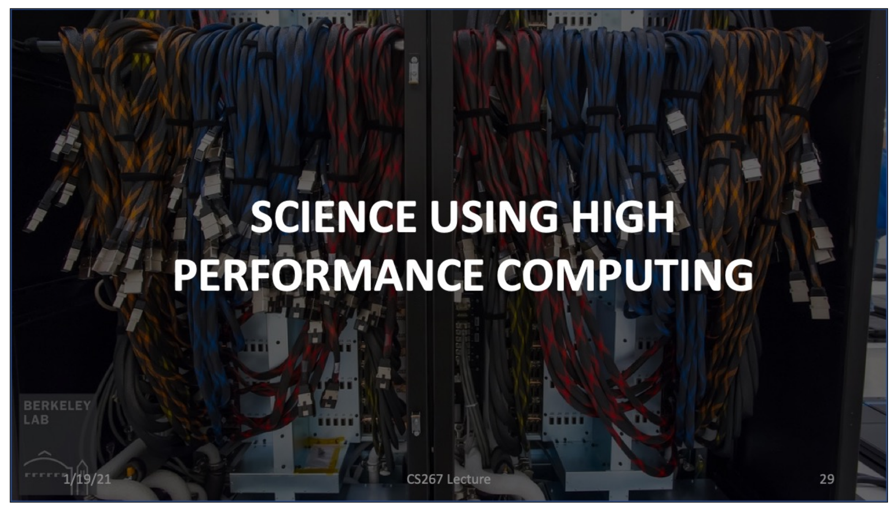


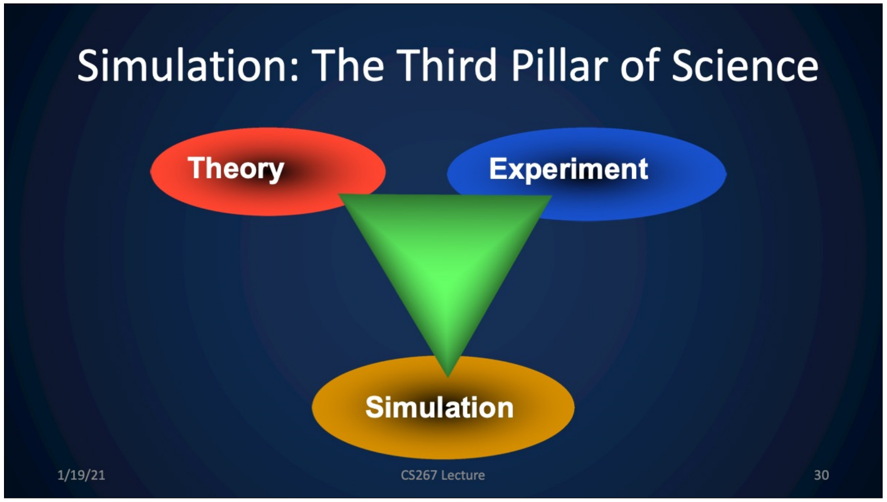
Gordon Bell Prizes: Science at Scale

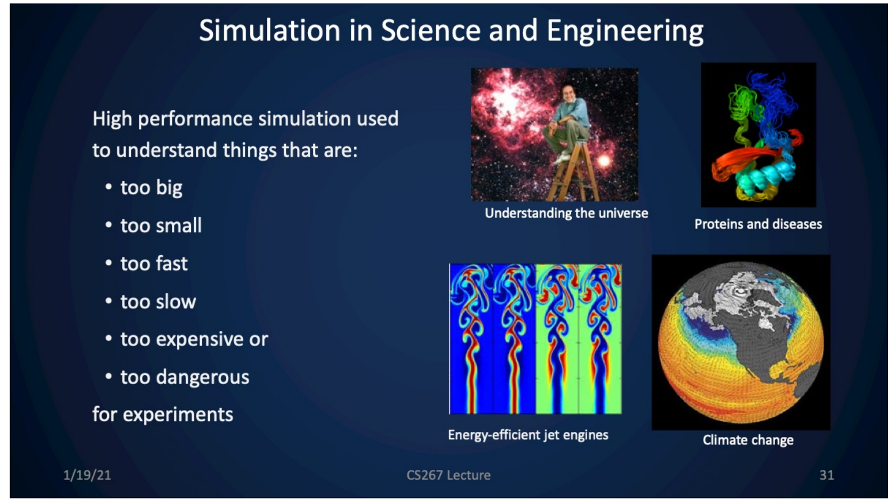
Established in 1987 with a cash award of $10,000 (since 2011), funded by Gordon Bell, a pioneer in HPC.
For innovation in applying *HPC to applications in science, engineering, and data analytics.*

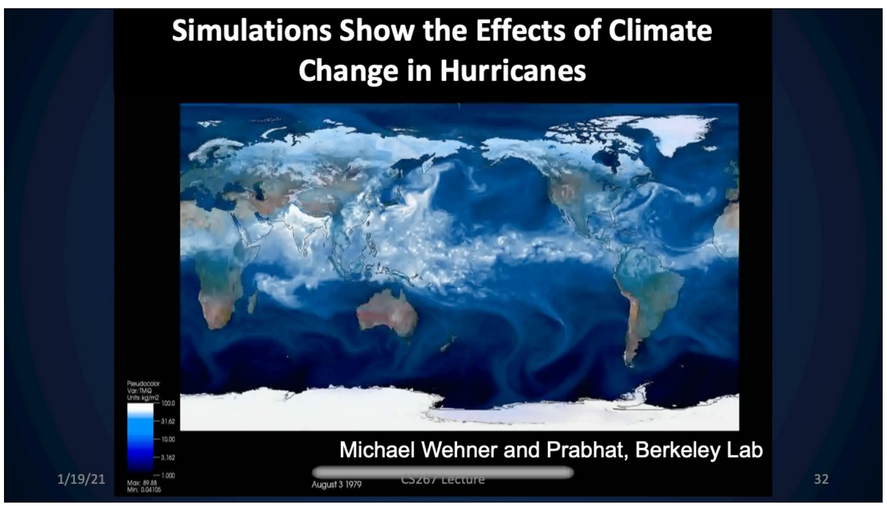
Gordon Bell Prizes vs Top 500

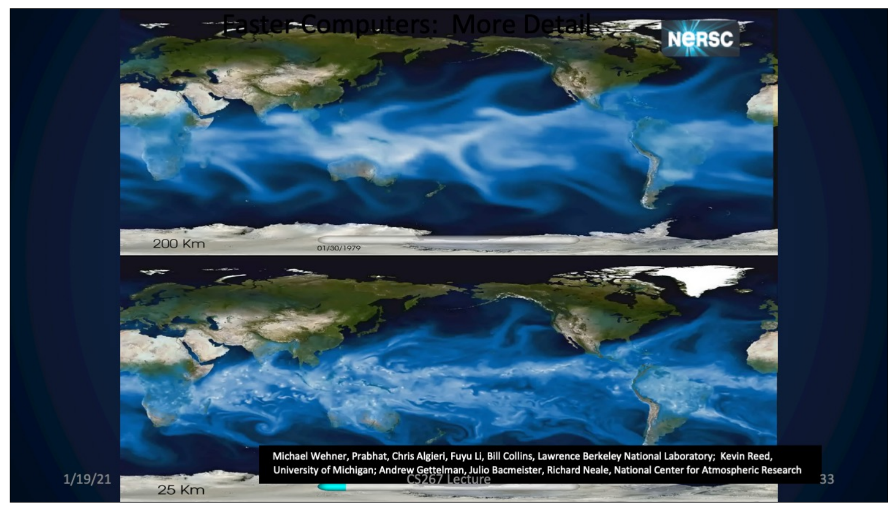




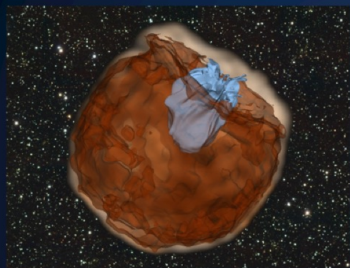

Simulation: The Third Pillar of Science
Theory
Experiment
Simulation
1/19/21
CS267 Lecture
30


Simulation in Science and Engineering
High performance simulation used to understand things that are:
• too big
• too small
• too fast
• too slow
• too expensive or
• too dangerous
for experiments
Understanding the universe
Proteins and diseases
Energy-efficient jet engines
Climate change
1/19/21
CS267 Lecture
31

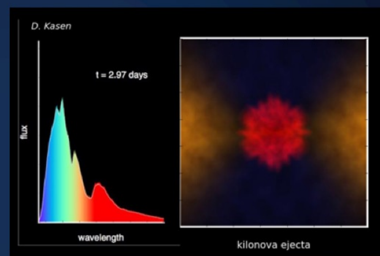

Simulations Show the Effects of Climate Change in Hurricanes
Michael Wehner and Prabhat, Berkeley Lab
August 3 1979
1/19/21
CS267 Lecture
32


Faster Computers: More Detail
NERSC
200 Km
25 Km
Michael Wehner, Prabhat, Chris Algieri, Fuyu Li, Bill Collins, Lawrence Berkeley National Laboratory; Kevin Reed, University of Michigan; Andrew Gettelman, Julio Bacmeister, Richard Neale, National Center for Atmospheric Research
1/19/21
CS267 Lecture
33

HPC for Astrophysics



Expanding debris from a supernova explosion (red) running over and shredding a nearby star (blue)

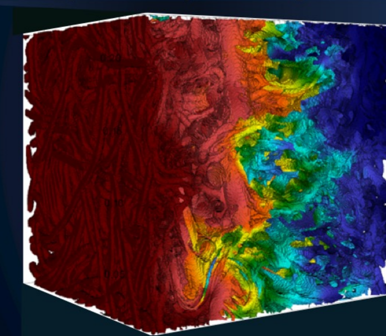


Ligo and Virgo observations match earlier simulations of gravitational waves from neutron star merger. Simulations predict ~200 earth masses of gold; ~500 of platinum

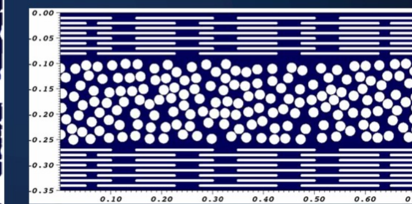
Dan Kasen, UCB Astronomy/Physics + LBNL

HPC for Energy Efficiency in Industry

Paper industry is 4th Largest Energy Consumer in US



Chombo-Pulp: Apply adaptive embedded boundary solver to resolve flow around pulp fibers and in felt pore space



Adaptive mesh refinement and interface tracking

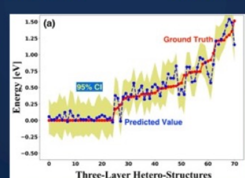
High Throughput HPC for Materials Design

Design of Materials for Batteries, Solar Panels and More

Software Supercomputers Screening > 40,000 Users

NANOPOROUS MATERIALS	530,243
INORGANIC COMPOUNDS	131,613
BAND STRUCTURES	76,194
MOLECULES	49,705

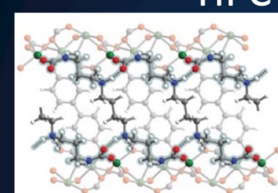
Data



- Use of Bayesian optimization for layered materials
- [Bassman et al, npj Computational Materials 2018]

Kristin Persson, Gerd Ceder, MSE UCB and LBNL, Materials Project

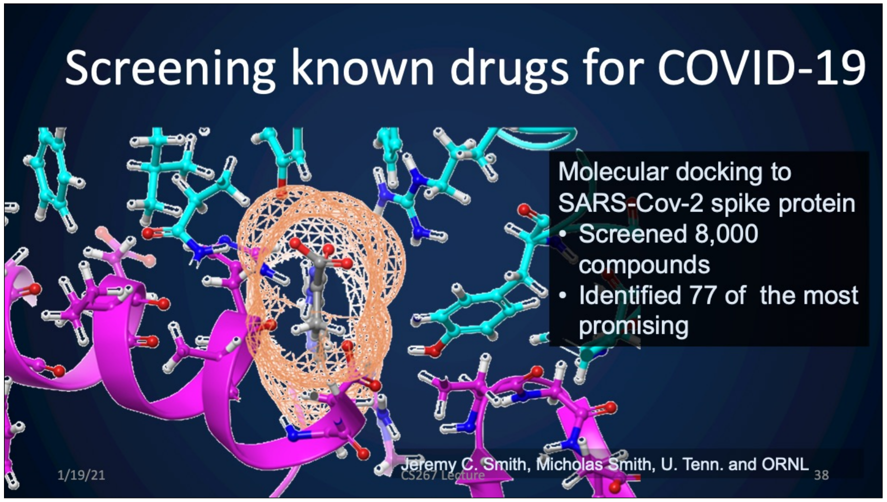
HPC for Carbon Capture

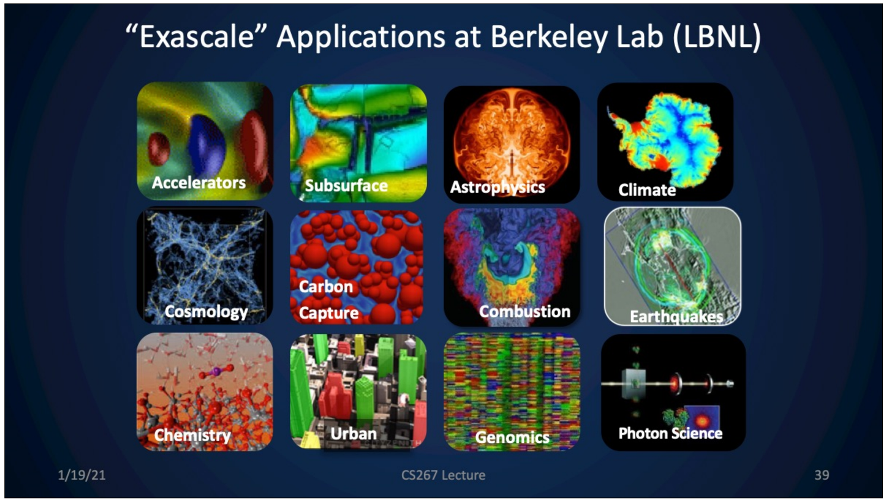


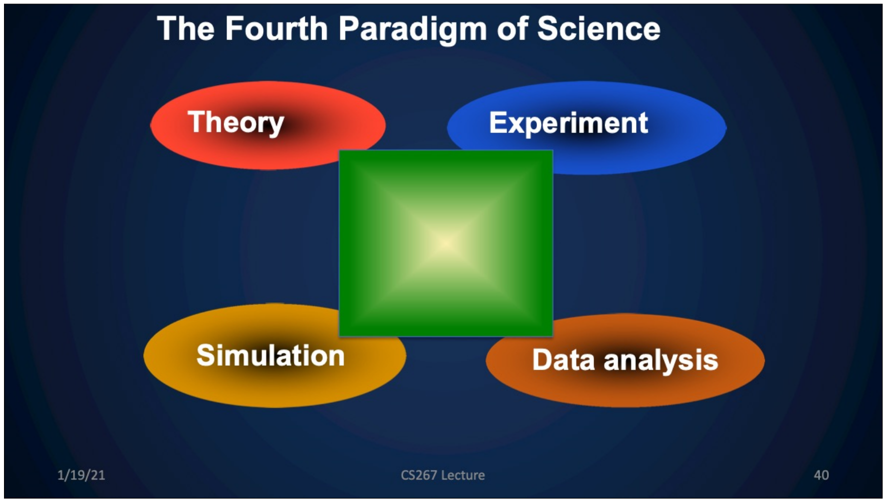
Metal Organic Frameworks (MOFs) to capture carbon in natural gas plants.

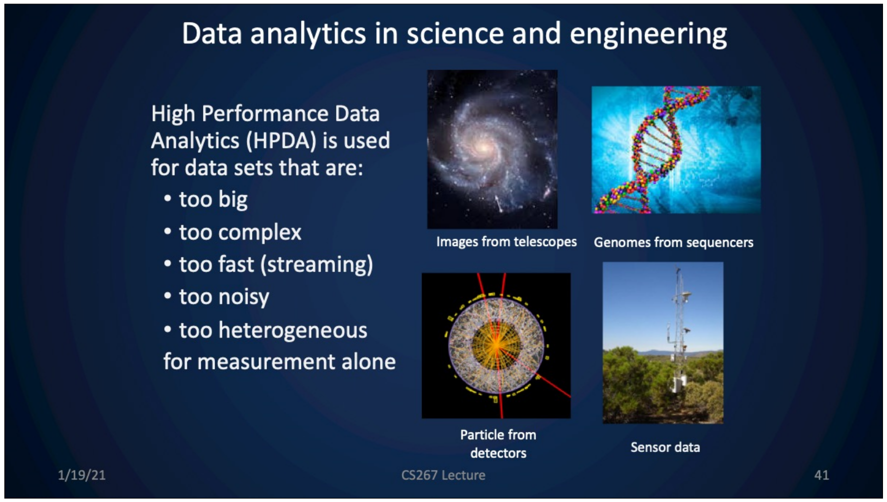
- Removes >90% of CO_2 from flue, 6X more than current (amine) technology.
- Steam to regenerate the MOF to reuse
- Exploring MOF design space with Density Functional Theory (DFT)

Jeff Long College of Chemistry / UC Berkeley and LBNL

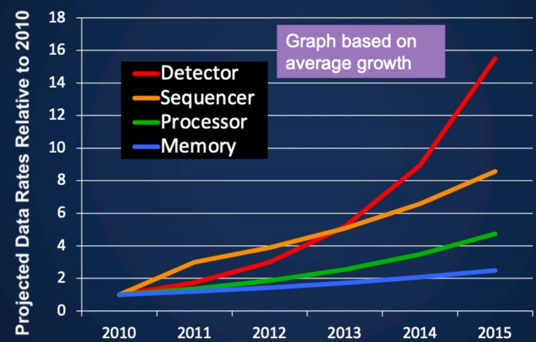

Screening known drugs for COVID-19
Molecular docking to SARS-Cov-2 spike protein
• Screened 8,000 compounds
• Identified 77 of the most promising
Jeremy C. Smith, Micholas Smith, U. Tenn. and ORNL
1/19/21
CS267 Lecture
38


"Exascale" Applications at Berkeley Lab (LBNL)
Accelerators
Subsurface
Astrophysics
Climate
Cosmology
Carbon Capture
Combustion
Earthquakes
Chemistry
Urban
Genomics
Photon Science
1/19/21
CS267 Lecture
39

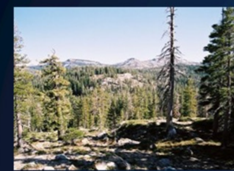

The Fourth Paradigm of Science
Theory
Experiment
Simulation
Data analysis
1/19/21
CS267 Lecture
40


Data analytics in science and engineering
High Performance Data Analytics (HPDA) is used for data sets that are:
• too big
• too complex
• too fast (streaming)
• too noisy
• too heterogeneous
for measurement alone
Images from telescopes
Genomes from sequencers
Particle from detectors
Sensor data
1/19/21
CS267 Lecture
41

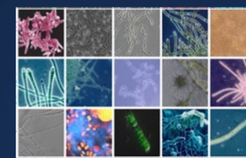
Data Growth is Outpacing Computing Growth



High Performance Data Analytics (HPDA) for Genomics



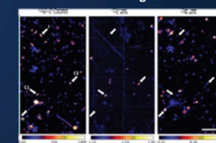
What happens to microbes after a wildfire? (1.5TB)



What are the microbial dynamics of soil carbon cycling? (3.3 TB)

How do microbes affect disease and growth of switchgrass for biofuels (4TB)

What at the seasonal fluctuations in a wetland mangrove? (1.6 TB)



Combine genomics with isotope tracing methods for improved functional understanding (8TB)

JGI-NERSC-KBase FICUS projects, MetaHipMer assembler ExaBiome project

Analysis of Genomic Data

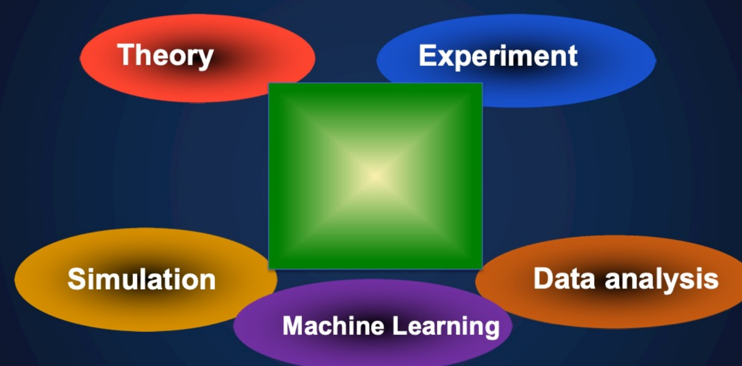
Dark green nodes: Kalanchoë genes
Yellow nodes: pineapple genes
Light green: model plant that uses a different photosynthesis strategy.

Blue edges: positive correlations
Red edges show negative correlations a

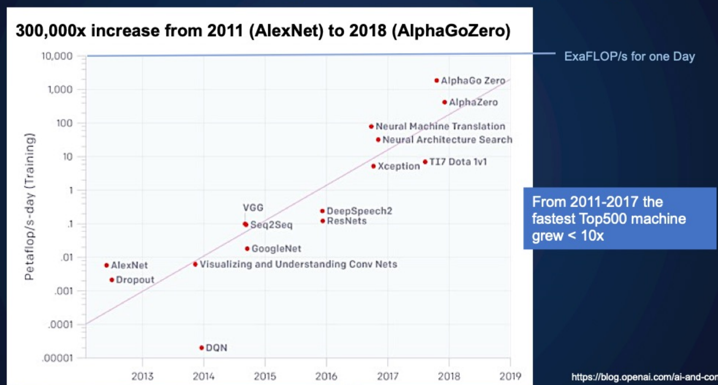
- Correlations of gene expression in plants that use different photosynthesis strategies.
- Kalanchoë and pineapple both use water-sparing photosynthesis
- 2.36 exaops / second on Summit computer
  - exaop = 10^{18} 16-bit ops
  - Gordon Bell Prize at SC18

Sharlee Climer, Kjiersten Fagnan, Daniel Jacobson, Wayne Joubert, Amy Justice, David Kainer, Deborah Weighill

The Fifth Paradigm of Science ?

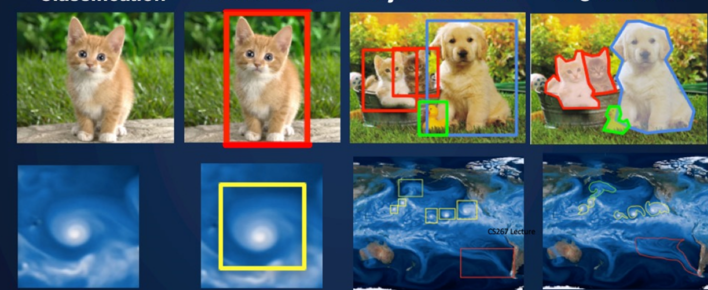


Machine learning demands more computing



https://blog.openai.com/ai-and-compute/

Data Analytics via Supervised Learning

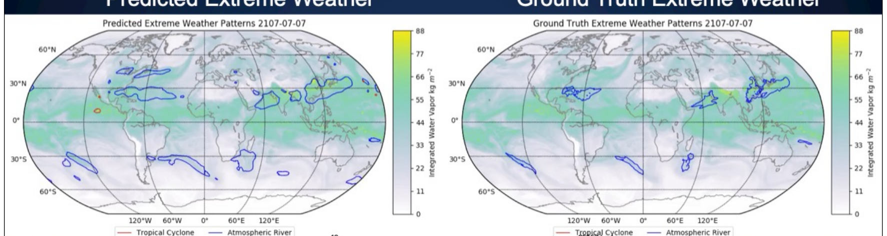


Extending image-based methods to complex, 3D, scientific data sets is non-trivial!

Slide source Prabhat

Big Data, Big Model, and Big Iron

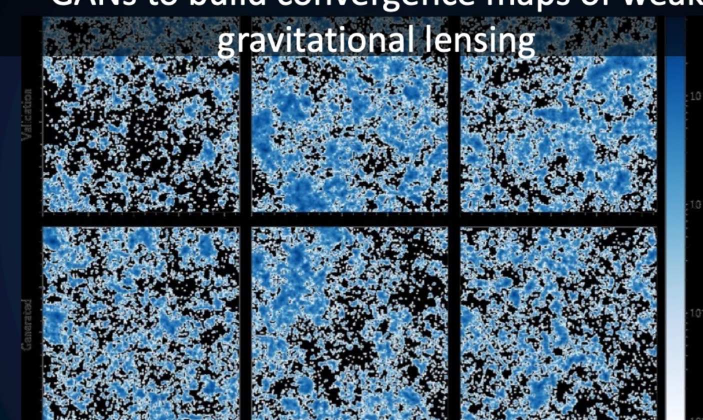
Predicted Extreme Weather Ground Truth Extreme Weather



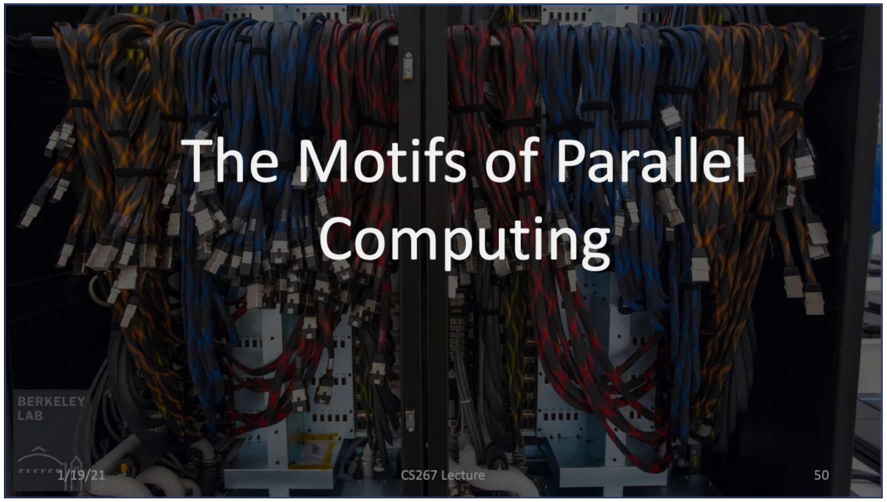
- **Deep learning results are smoother than heuristic labels**
- **Achieved over 1 EF peak on OLCF Summit: Gordon Bell Prize in 2018**

Thorsten Kurth, Sean Treichler, Joshua Romero, Mayur Mudigonda, Nathan Luehr, Everett Phillips, Ankur Mahesh, Michael Matheson, Jack Deslippe, Massimiliano Fatica, Prabhat, Michael Houston

GANs to build convergence maps of weak gravitational lensing



CosmoGAN: Mustafa Mustafa, Deborah Bard, Wahid Bhimji, Zarija Lukić, Rami Al-Rfou, Jan M. Kratochvil

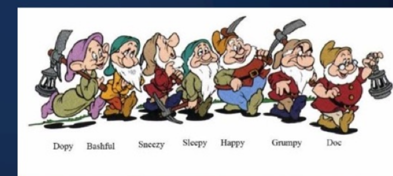


The Motifs of Parallel Computing

How to cover all applications?

- Phil Colella's famous 7 "dwarfs" of scientific computing (simulation)

Dense Linear Algebra
Sparse Linear Algebra
Particle Methods
Structured Grids
Unstructured Grids
Spectral Methods (e.g. FFT)
Monte Carlo

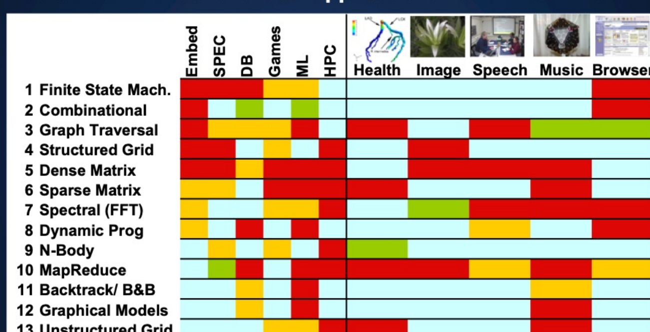


Colella's 2004 DARPA presentation "Defining Software Requirements for **Scientific** Computing"

Motifs: Common Computational Methods

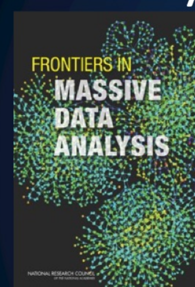
(Red Hot Important → Blue Cool Not Important)

What do commercial and CSE applications have in common?



2006 Berkeley View paper -- A View of the Parallel Computing Landscape

Analytics vs. Simulation Motifs

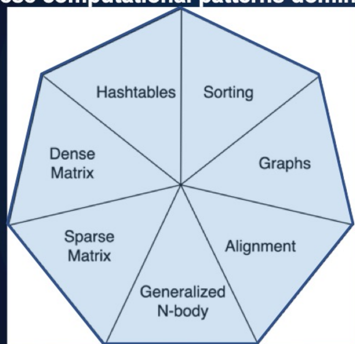


National Academies 2013

7 Giants of Data	7 Dwarfs of Simulation
Basic statistics	Monte Carlo methods
Generalized N-Body	Particle methods
Graph-theory	Unstructured meshes
Linear algebra	Dense Linear Algebra
Optimizations	Sparse Linear Algebra
Integrations	Spectral methods
Alignment	Structured Meshes

Motifs of Genomic Data Analysis

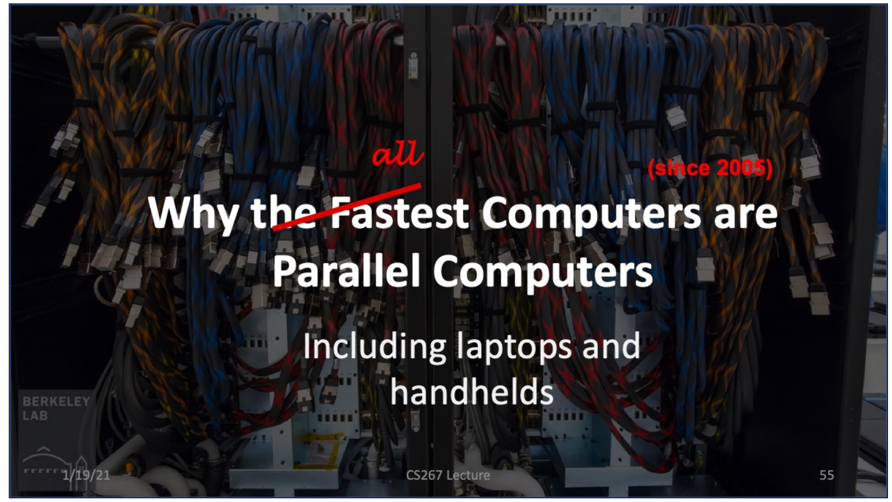
These computational patterns dominate ExaBiome



Application problems
- Overlap: Find all overlaps in a set
- Assembly: find / correct overlaps
- Distance: how good are overlaps
- Index: lookup in database

Large shared memory platforms were most common – limits science questions and approaches

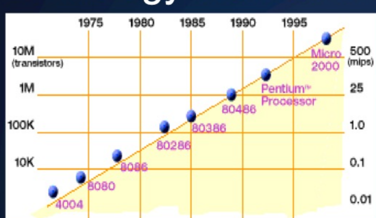
Yelick, et al. "The Parallelism Motifs of Genomic Data Analysis", Philosophical Transactions A, 2020



Why the ~~Fastest~~ all Computers are Parallel Computers (since 2005)

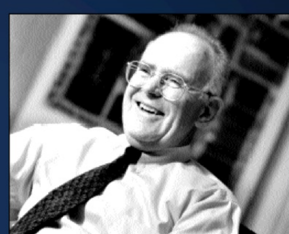
Including laptops and handhelds

Technology Trends: Microprocessor Capacity



2X transistors/Chip Every 1.5 years
Called "Moore's Law"

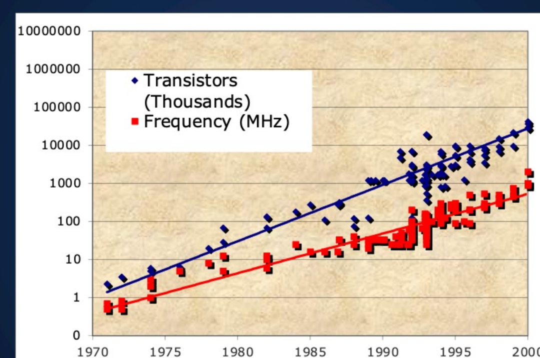
Microprocessors have become smaller, denser, and more powerful.



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

Microprocessor Transistors / Clock (1970-2000)



Historical Impact of Device Shrinkage

- What happens when feature size (transistor size) shrinks by a factor x ?
- Clock rate goes up by x because wires are shorter
  - actually less than x, because of power consumption
- Transistors per unit area goes up by x^2
- Die size has also increased
  - typically another factor of ~x
- Raw computing power of the chip goes up by ~ x^4 !
  - typically x^3 is devoted to either on-chip
    - parallelism: hidden parallelism such as ILP
    - locality: caches
- So some programs got x^3 times faster, without changing them

Limits: How fast can a serial computer be?

1 Tflop/s, 1 Tbyte sequential machine

r = 0.3 mm

- Consider the 1 Tflop/s (10^{12}) sequential machine:
  - Data must travel distance, r, from memory to processor.
  - To get 1 data element per cycle, this means 10^{12} times per second at the speed of light, c = 3×10^8 m/s. Thus r < c/10^{12} = 0.3 mm.
- Now put 1 Tbyte of storage in a 0.3 mm x 0.3 mm area:
  - Each bit occupies about 1 square Angstrom, or the size of a small atom.
- No choice but parallelism

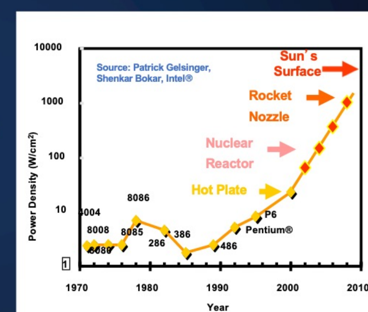
But What about Heat Density

Image by Sam Spratt

Power Density Limits Serial Performance

Scaling clock speed (business as usual) will not work

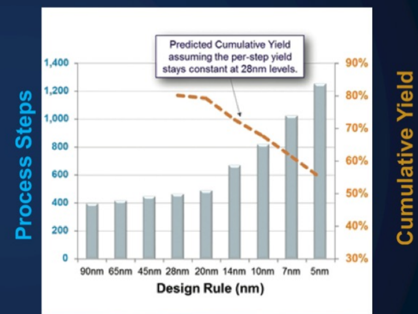
- *Faster processors → increase power*
  - Dynamic power is proportional to V^2fC
  - Increasing frequency (f) also increases supply voltage (V) → cubic effect
  - Increasing cores increases capacitance (C) but only linearly
  - Save power by lowering clock speed and adding parallelism
- High performance serial processors waste power
  - Speculation, dynamic dependence checking, etc. burn power
  - Implicit parallelism discovery
- More transistors, but not faster serial processors



More drivers for parallelism: Chip yield

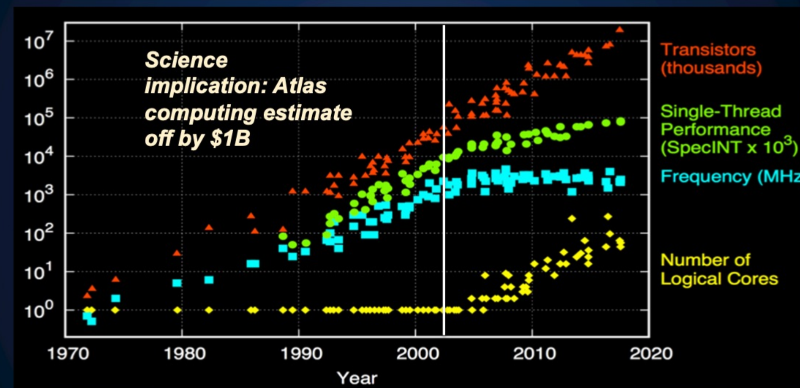
Yield

- What % of chips are usable?
- Complexity of fabrication (decreased size and number of steps) increases errors
- Parallelism helps, e.g., KNL (in Cori) sold with only 68 out of 76 "on" to improve yield



http://electroiq.com/blog/2016/02/yield-and-cost-challenges-at-16nm-and-beyond/

Dennard Scaling is Dead; Moore's Law Will Follow



M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, C. Batten, and K. Rupp

Moore's Law reinterpreted

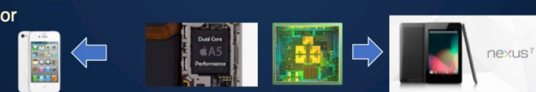
- **Number of cores per chip can double every two years**
- **Clock speed will not increase (possibly decrease)**
- **Need to deal with systems with millions of concurrent threads**
- **Need to deal with inter-chip parallelism as well as intra-chip parallelism**
- **But Moore's Law is not forever… industry consortium predicts end in 2021**

Parallel hardware is everywhere

It's just about all you can find today in laptops, servers etc.

Intel® Knights Landing (68 cores) | Intel® Haswell (16 cores) | AMD Ryzden (8 cores) | NVIDIA V100 | Fujitsu ARM (48 cores)

- Even my cell phone has a parallel processor



There's no escape!

Caution: Amdahl's Law

- Suppose only part of an application is parallel
- Amdahl' s law
  - **s** = fraction of work done sequentially (Amdahl fraction)
  - 1-s is fraction parallelizable
  - **P** = number of processors

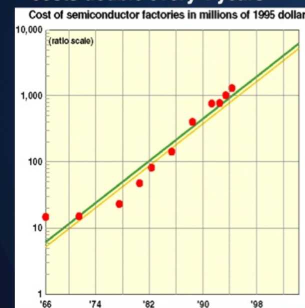
Speedup(P) = Time(1)/Time(P)
$\quad\quad$ <= 1/(s + (1-s)/P)
$\quad\quad$ <= 1/s

- Even if the parallel part speeds up perfectly, performance is limited by the sequential part
- E.g., 1/10th of your code's runtime is serial → max speedup is 10x (Cori has 65K cores)

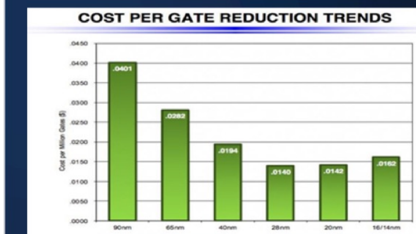
Moore's Law is Techno-Economic

approaching atomic scale, but economics is first

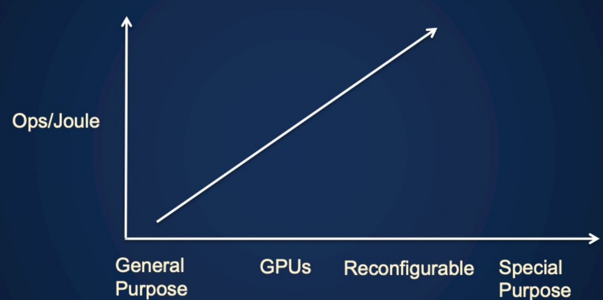
Cost of fabrication facilities:
Moore' s 2nd law (Rock' s law):
costs double every 4 years



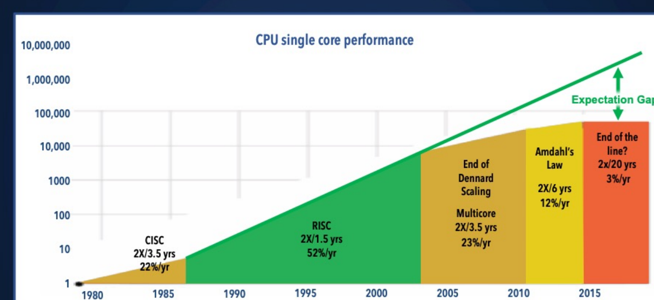
Cost per gate not improving:
lithography and fab costs



Specialization: End Game for Moore's Law



Traditional Scaling is Coming to an End!



Hennessy/Patterson Turing Award presentation: Golden Age of Computer Architecture Research

Moore's Law

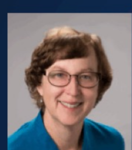
It's hard to think exponentially

But it's also hard to stop

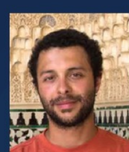
Who's Who and What's What in CS267

Course Staff

Instructors



Kathy Yelick

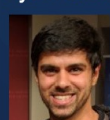


Aydın Buluç

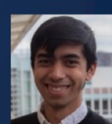
Jim Demmel

GSIs

Giulia Guidi



Melih Elibol



Alok Tripathy

See website for contact information and office hours. Email us at cs267@lists.eecs.berkeley.edu

Universities that have offered this class

NSF XSEDE Program

- University of Kentucky
- Indiana University Southeast
- Penn State University
- New Mexico State University
- George Mason University
- Oregon State University
- Bluffton University
- Gonzaga University
- Norfolk State University
- University of Toronto
- Texas A&M University San Antonio
- Washington State University
- East Carolina University
- Morehouse College Saint Martin's University in Washington
- University of Arizona
- Adrian College in Michigan
- Clarkson University in New York
- University of Rhode Island
- University of Arkansas
- University of Cincinnati in Ohio
- University of Nevada, Las Vegas
- Fordham University in New York
- Wofford College in South Carolina
- Morgan State University in Maryland
- New Jersey Institute of Technology
- Brown University in Rhode Island
- University of Pennsylvania
- University of Toronto in Canada
- Francis Marion University in South Carolina
- University of Kentucky
- Louisiana State University
- Indiana University Southeast
- Portland State University in Oregon
- New Mexico State University
- Jarvis Christian College in Texas
- Hampden–Sydney College in Virginia
- University of Nevada, Reno
- Universidad de Puerto Rico
- Texas A&M U.-Corpus Christi
- Universidad EAFIT in Columbia

Overview of the course (not in order)

- Parallel Programming Models and Machines (plus some architecture, e.g., caches)

Algorithm/machine model	Language / Library skills
Shared memory	OpenMP (pThreads)
Distributed memory	MPI
	PGAS
Data parallel	SPARK
	CUDA

- Parallelization Strategies for the "Motifs" of Scientific Computing (and Data)

Dense Linear Algebra	Monte Carlo
Sparse Linear Algebra	Spectral Methods
Particle Methods	Graphs
Structured Grids	Sorting
Unstructured Grids	Hashing

- Performance models:
  - Roofline
  - α-β (latency/bandwidth)
  - (LogP)
- Cross-cutting topics:
  - Communication avoiding
  - Load balancing
  - Hierarchical algorithms
  - Autotuning
- The Laws
  - Moore's Law
  - Amdahl's Law
  - Little's Law
- Applications (in some detail)
  - Machine Learning
  - Biology
  - Cosmology