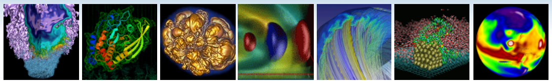


# *Saving the World with Computing*

Kathy Yelick

EECS Professor, U.C. Berkeley

Associate Laboratory Director for Computing Sciences and
Lawrence Berkeley National Laboratory

## Using Computers for Science and Engineering

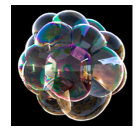
Computers are used to understand things that are:

- too big
- too small
- too fast
- too slow
- too expensive or
- too dangerous

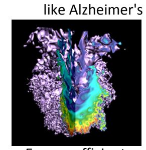
for experiments

Understanding the universe

Proteins and diseases like Alzheimer's

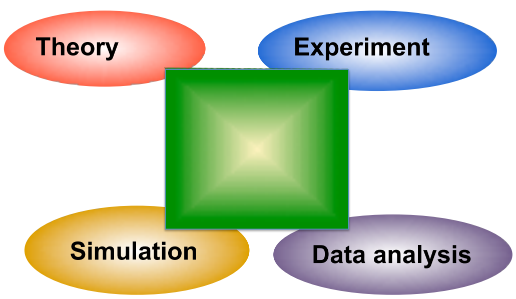


Industrial products and processes



Energy-efficient combustion engines

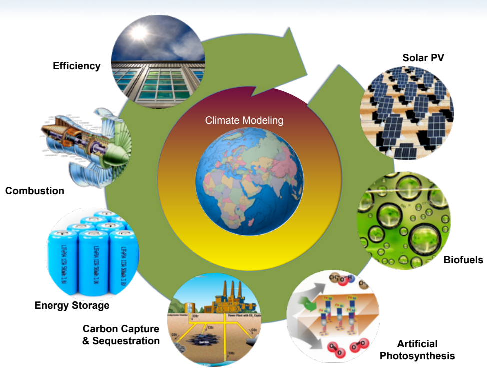
## The ~~"Third Pillar"~~ of Science "Four Paradigms"



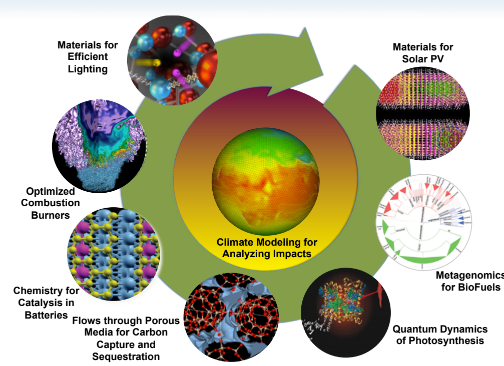

## Addressing Challenges using Computing

- **Two of the most significant challenges**
  - **Our changing world:** understanding climate change, alternative energy sources, mitigation techniques, etc.
  - **Health and medicine:** understanding the human body, development of treatments, and disease prevention

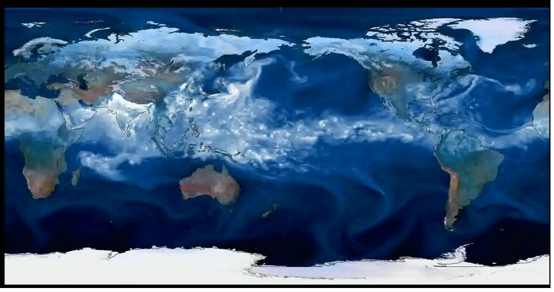
## Carbon Cycle 2.0 Initiative at Berkeley Lab




## Carbon Cycle 2.0 Initiative at Berkeley Lab



## Simulations Show the Effects of Climate on Hurricanes

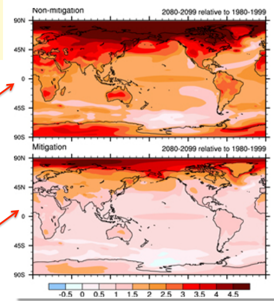


## Mitigating Global Climate Change

**Can global warming impacts be diminished if greenhouse gases are cut?**

*Average surface air temperatures rise by >3°C if emissions increase at current rate*

*Temperatures rise by <2°C across nearly all populated areas if emissions are cut by 70%*

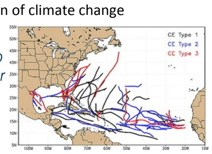



## 20th Century Climate Data Reconstructed

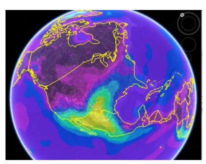
Reconstructed global weather conditions in 6-hour intervals from 1871-2010

- Based on data from meteorologists, military, volunteers and ships' crews
- Over 10M hours at NERSC using reverse Kalman filter algorithms
- Data used in 16 papers to date: reproduced 1922 Knickerbocker storm, understand causes of the 1930 Dust Bowls, and determine whether recent extremes are sign of climate change

*NERSC has 2PB of online storage and up to 44 PB of archive for scientific data sets. New "Science Gateways" make it easy to make data accessible on the web*



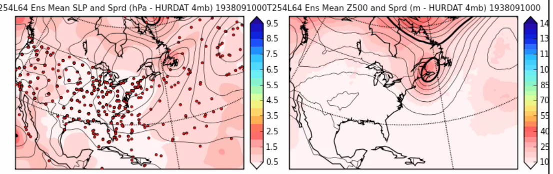

Previously undetected warm-core cyclones, *Geophys. Res. Letters*, 2011



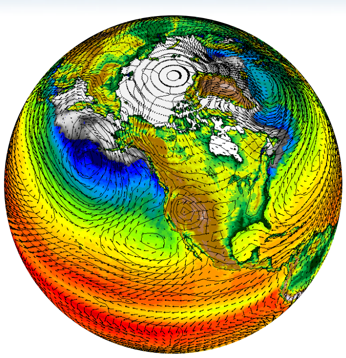
Relative Humidity for 1920-1929
Gil Compo, PI (U. Colorado)

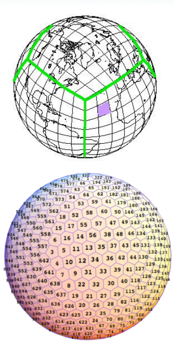
## Climate Change Requires Lots of Data

"validate" that the computer models are working as expected

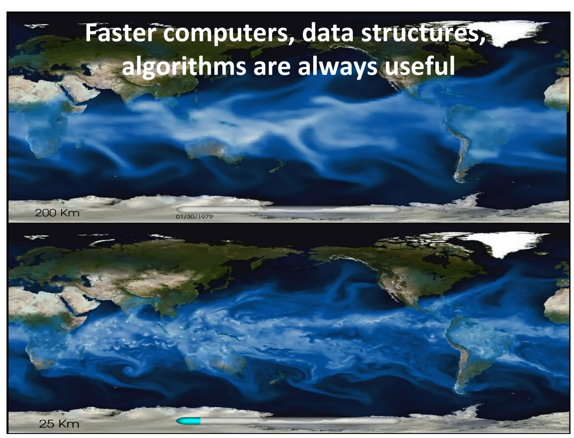


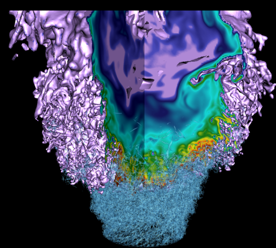

## Data Structures for Simulations





## Faster computers, data structures, algorithms are always useful





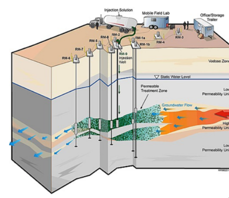
Simulations reveal features not visible in lab

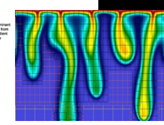
Experiments demonstrate feasibility

*Result: Low NOx burner technology licensed by industry*

## Simulations to Get Rid of $CO_2$

- Carbon sequestration: "The process of removing carbon from the atmosphere or from flue gasses and depositing it in a reservoir."
- $CaO + CO_2 \rightarrow CaCO_3$

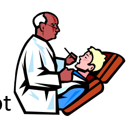




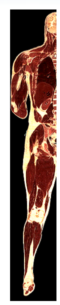
George Pau, LBNL

## Towards a Digital Human: The 20+ Year Vision

- Imagine a "digital body double"
  - 3D image-based medical record
  - Includes diagnostic, pathologic, and other information
- Used for:
  - Diagnosis
  - Less invasive surgery-by-robot
  - Experimental treatments



## Digital Human Today: Imaging



- The Visible Human Project
  - 18,000 digitized sections of the body
    - Male: 1mm sections, released in 1994
    - Female: .33mm sections, released in 1995
  - Goals
    - study of human anatomy
    - testing medical imaging algorithms
  - Current applications:
    - educational, diagnostic, treatment planning, virtual reality, artistic, mathematical and industrial
    - Used by > 1,400 licensees in 42 countries

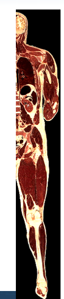


Image Source: www.madsci.org

THE VISIBLE HUMAN PROJECT®

## Experimental Data: Visible Human

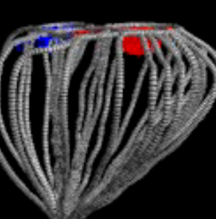
The National Library of Medicine's

Visible Human Project (TM)

Human-Computer Interaction Lab
Univ. of Maryland at College Park

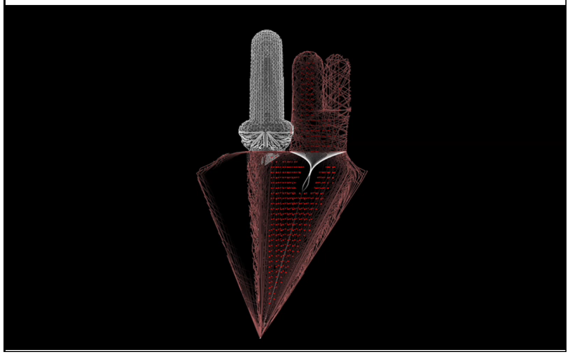
## Heart Simulation

Movie from Charles Peskin and Dave McQueen at NYU

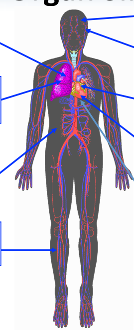


## Heart Simulation

Movie from Boyce Griffith's PhD thesis, NYU

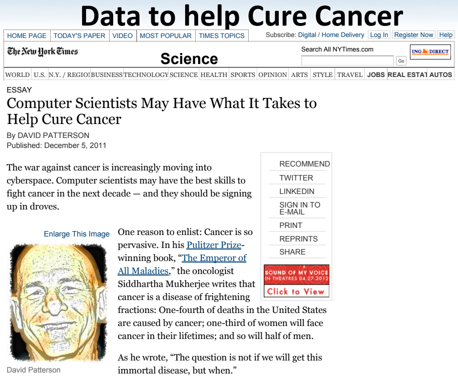


## Organ Simulation




Just a few of the efforts at understanding and simulating parts of the human body

## Big D and Big C: Computing on Big Data to help Cure Cancer



HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times

Science

ESSAY

Computer Scientists May Have What It Takes to Help Cure Cancer

By DAVID PATTERSON
Published: December 5, 2011

The war against cancer is increasingly moving into cyberspace. Computer scientists may have the best skills to fight cancer in the next decade — and they should be signing up in droves.

One reason to enlist: Cancer is so pervasive. In his Pulitzer Prize-winning book, "The Emperor of All Maladies," the oncologist Siddhartha Mukherjee writes that cancer is a disease of frightening fractions: One-fourth of deaths in the United States are caused by cancer; one-third of women will face cancer in their lifetimes; and so will half of men.

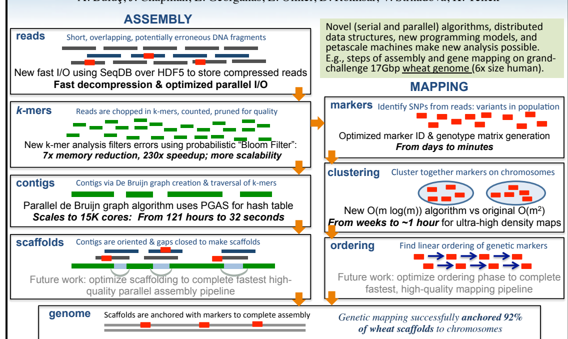
As he wrote, "The question is not if we will get this immortal disease, but when."

Enlarge This Image

David Patterson

## Optimization of Large-Scale Genome Assembly & Mapping

A. Buluç, J. Chapman, E. Georganas, L. Oliker, D. Rokhsar, V. Strnadova, K. Yelick



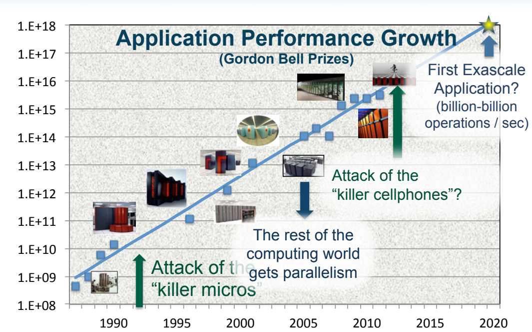

Funded under DOE LDRD, X-Stack DEGAS, JGI, Early Career, Math for DOE Data

## Trends in Computer Science

Which of the following are true?

A. Moore's Law says that processor performance doubles every 18 months
B. Moore's Law has ended
C. Most of the time in scientific codes is spent doing arithmetic
D. None of the above
E. All of the above

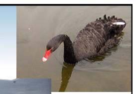
## High End Computing Revolutions




26

| # | Site | Manufacturer | Computer | Country | Cores | Rmax [Pflops] | Power [MW] |
|---|---|---|---|---|---|---|---|
| 1 | National University of Defense Technology | NUDT | Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi | China | 3,120,000 | 33.9 | 17.8 |
| 2 | Oak Ridge National Laboratory | Cray | Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x | USA | 560,640 | 17.6 | 8.21 |
| 3 | Lawrence Livermore National Laboratory | IBM | Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom | USA | 1,572,864 | 17.2 | 7.89 |
| 4 | RIKEN Advanced Institute for Computational Science | Fujitsu | K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect | Japan | 795,024 | 10.5 | 12.7 |
| 5 | Argonne National Laboratory | IBM | Mira BlueGene/Q, Power BQC 16C 1.6GHz, Custom | USA | 786,432 | 8.59 | 3.95 |
| 6 | Swiss National Supercomputing Centre (CSCS) | Cray | Piz Daint Cray XC30, Xeon E5 8C 2.6GHz, Aries, NVIDIA K20x | Switzer-land | 115,984 | 6.27 | 2.33 |
| 7 | Texas Advanced Computing Center/UT | Dell | Stampede PowerEdge C8220, Xeon E5 8C 2.7GHz, Intel Xeon Phi | USA | 462,462 | 5.17 | 4.51 |
| 8 | Forschungszentrum Juelich (FZJ) | IBM | JuQUEEN BlueGene/Q, Power BQC 16C 1.6GHz, Custom | Germany | 458,752 | 5.01 | 2.30 |
| 9 | Lawrence Livermore National Laboratory | IBM | Vulcan BlueGene/Q, Power BQC 16C 1.6GHz, Custom | USA | 393,216 | 4.29 | 1.97 |
| 10 | Leibniz Rechenzentrum | IBM | SuperMUC iDataPlex DX360M4, Xeon E5 8C 2.7GHz, Infiniband FDR | Germany | 147,456 | 2.90 | 3.52 |

## Black Swans of Computing with 1992 Technology




## Technology for Innovation

Which of the following are true?

A. Google developed its own programming language to hide machine failures

B. iPhones are programmed using Java

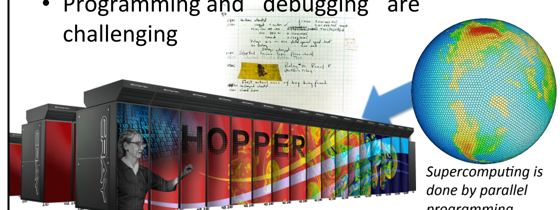
C. Web search algorithms use only integer arithmetic, not floating point (real) numbers

D. Scientific computing is done mostly using "Vector Supercomputers"

E. All of the above

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

- Fastest Computers in the world: top500.org
- Our Hopper Computer has 150,000 cores and > 1 Petaflop ($10^{15}$ math operations / second)
- Programming and "debugging" are challenging



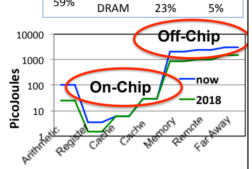

*Supercomputing is done by parallel programming*

30

## Challenge: Data Movement Dominates Cost

**Communication is expensive…**
**… time and energy**
**… processor to memory and processor to processor**

| Annual improvements | | | |
|---|---|---|---|
| Flops | | BW | Latency |
| 59% | Network | 26% | 15% |
| | DRAM | 23% | 5% |




**Cost components:**

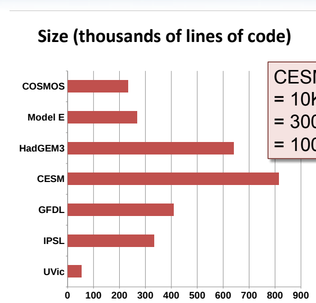
- **Bandwidth:** # of words
- **Latency:** # messages

**Strategies: hide latency, use new algorithms**

**Hard to change: Latency is physics; bandwidth is money!**

31

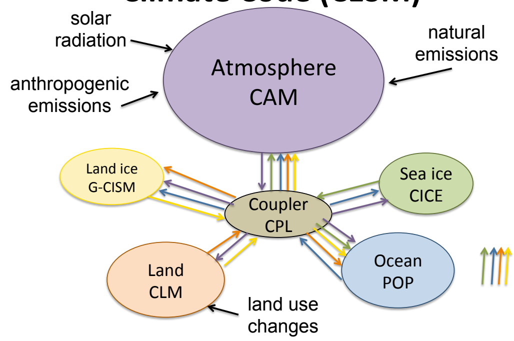
## How big are these applications




CESM: ~1M lines of code
= 10K programmer days?
= 300 programmer years
= 100 programmers, 3 years

Generated using David A. Wheeler's "SLOCCount".

## Abstractions in the Community Climate Code (CESM)




Real-Time Deformation and Fracture
in a Game Environment

Eric Parker
Pixelux Entertainment

James O'Brien
U.C. Berkeley

Video Edited by Sebastian Burke

From the proceedings of SCA 2009, New Orleans

## Writing Software

Which of the following are true?

A. Most computer software is written by brilliant hackers, working alone
B. Parallel programming is a *solved problem*
C. Speed of programming and speed of programs are the top goals in software
D. Most software is rewritten from scratch every few years
E. None of the above

## Computational Science is Necessarily Collaborative

*… as from the beginning the work has been a team effort involving many able and devoted co-workers in many laboratories. As I am sure you will appreciate, a* ***great many diverse talents*** *are involved in such developments and whatever measure of success is achieved is dependent on* ***close and effective collaboration****.*

***Ernest O. Lawrence***
***UC Berkeley Professor of Physics***
***Founder of Lawrence Berkeley National Laboratory***
***In his Nobel Lecture, December 11, 1951***

*Internships Available: http://csee.lbl.gov/*

## Why Study Computer Science?

1) Because computers can help solve important problems
2) Because computers are fun to program
3) Because computers make a good career
4) **Because you will get to work with lots of great people**