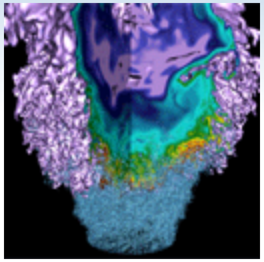
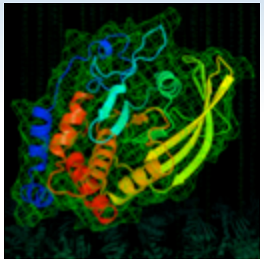
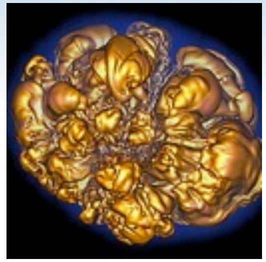
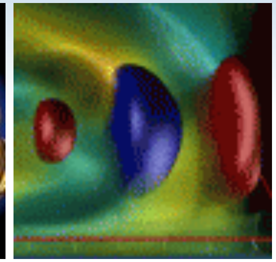
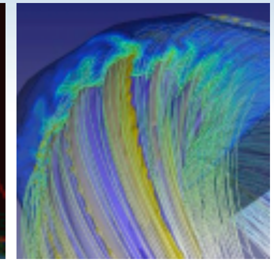
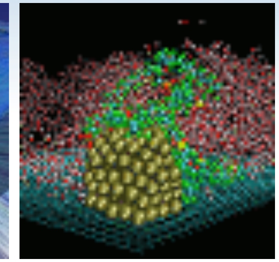
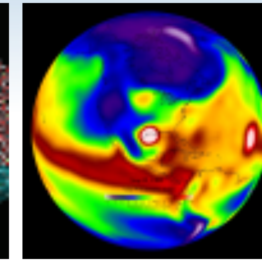
      

Saving the World with Computing

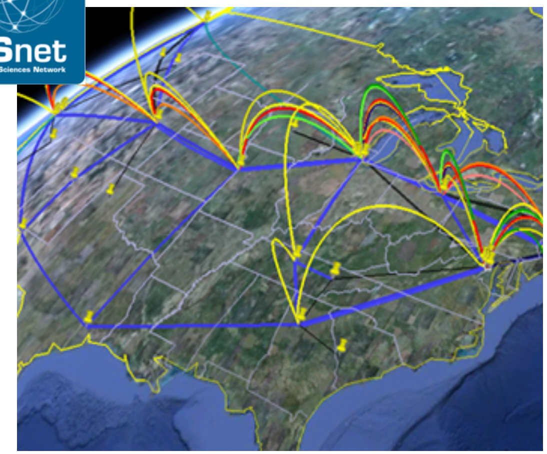
Kathy Yelick

EECS Professor, U.C. Berkeley

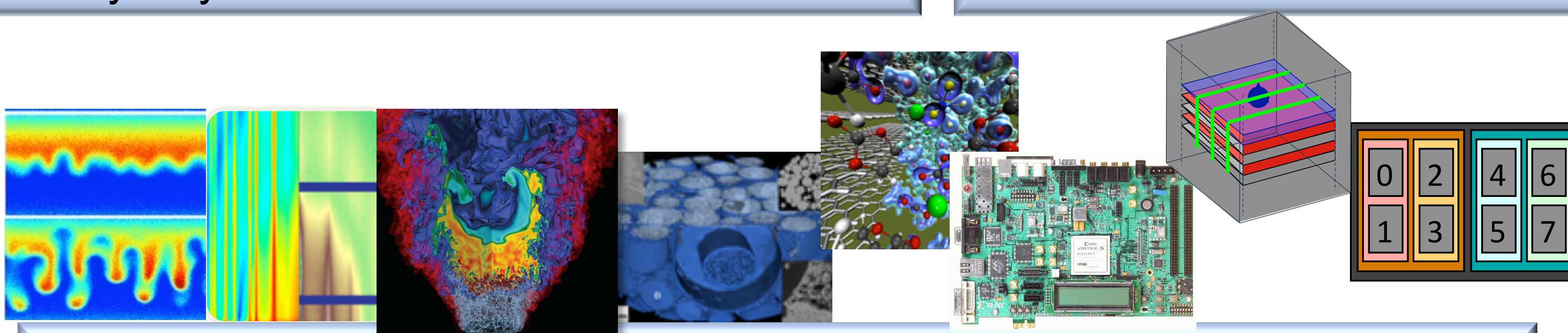
Associate Laboratory Director for Computing Sciences and Lawrence Berkeley National Laboratory

Computing at Berkeley Lab

NERSC Supercomputing Center: Petaflop and Petabyte systems for science



ESnet: An instrument for Data-Driven Science



Applied Mathematics and Computer Science

Why are you Interested in Computer Science?

I want to:

A. Build computer hardware and software

B. Create new companies and industries

C. Solve important problems facing the world

D. Work on teams with other creative people

E. All of the above

Using Computers for Science and Engineering

Computers are used to understand things that are:

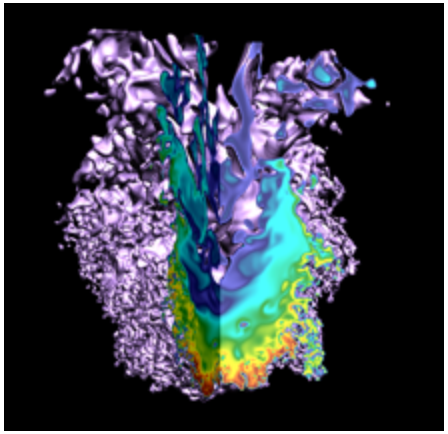
- too big
- too small
- too fast
- too slow
- too expensive or
- too dangerous

for experiments

Understanding the universe

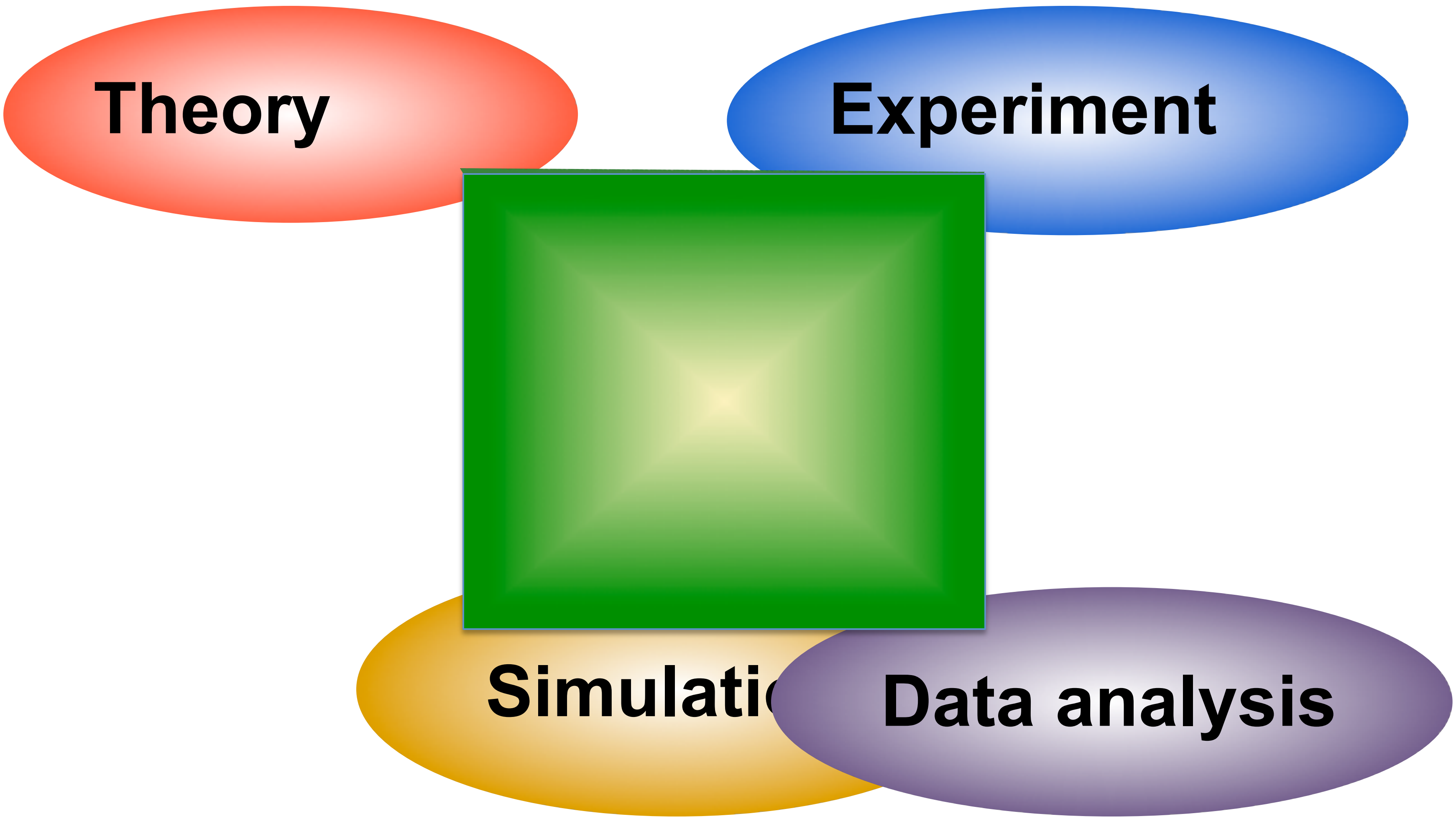
Proteins and diseases like Alzheimer's

Industrial products and processes



Energy-efficient combustion engines

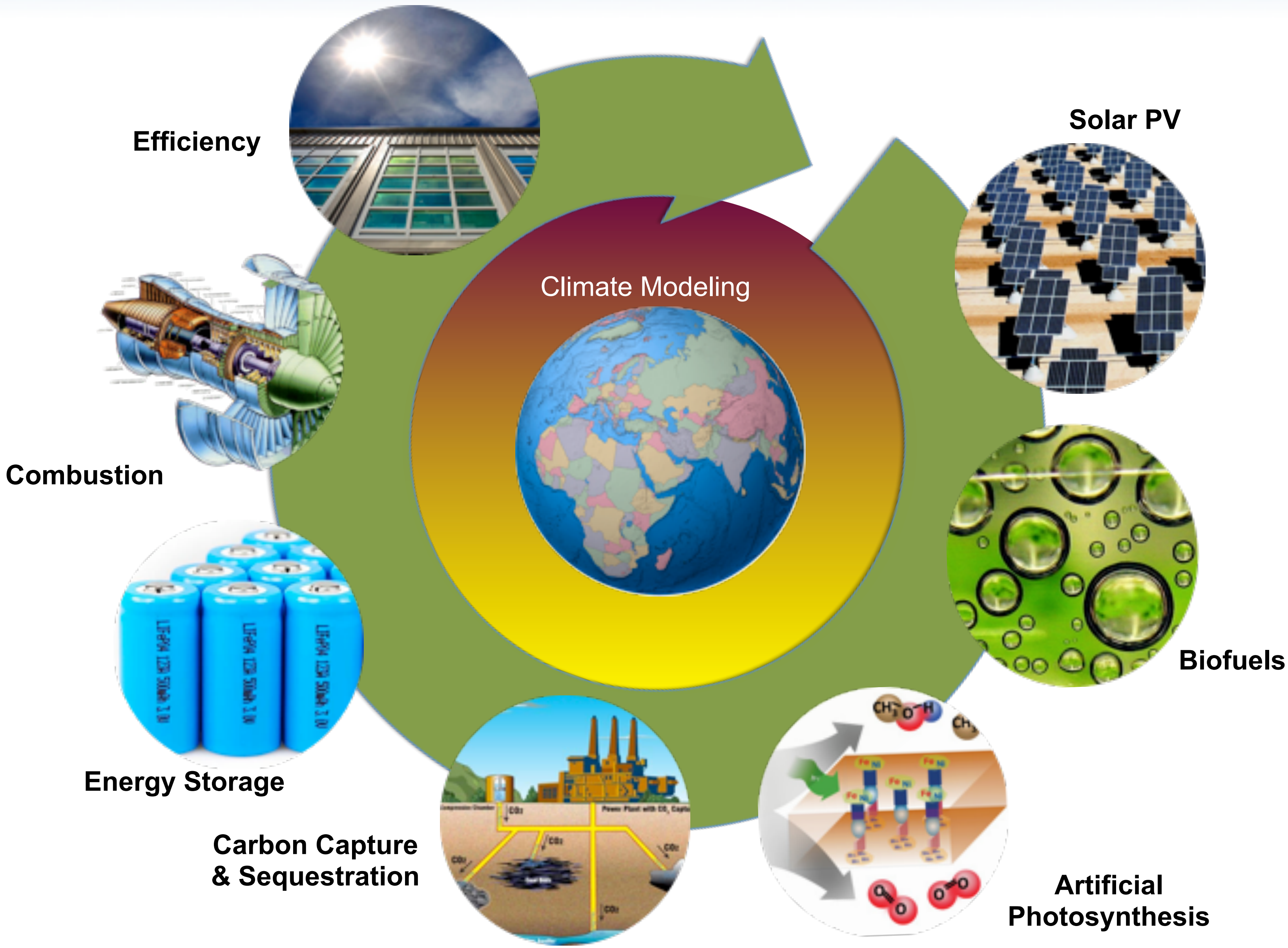
The ~~“Third Pillar”~~ of Science
“Four Paradigms”



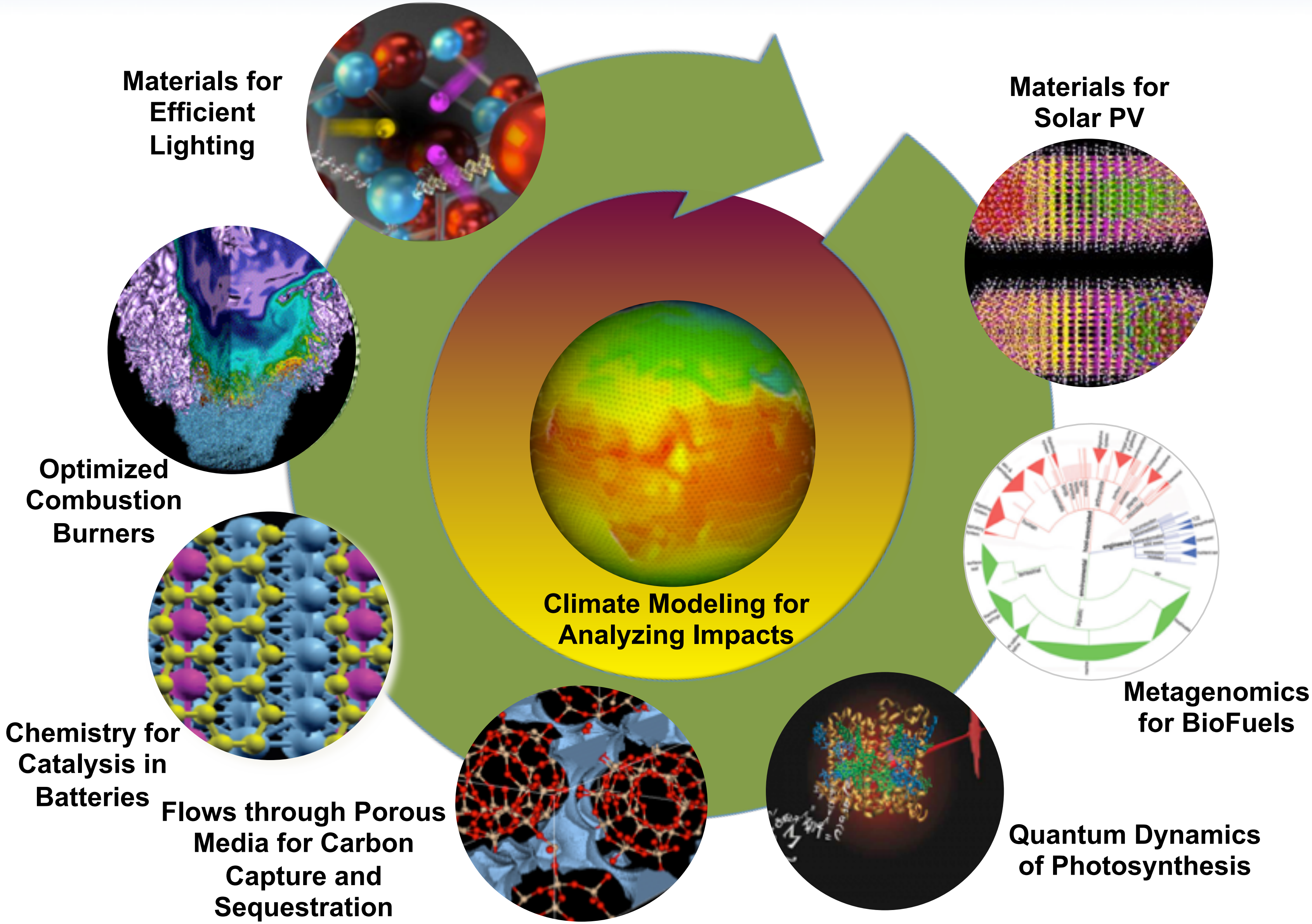
Addressing Challenges using Computing

- **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

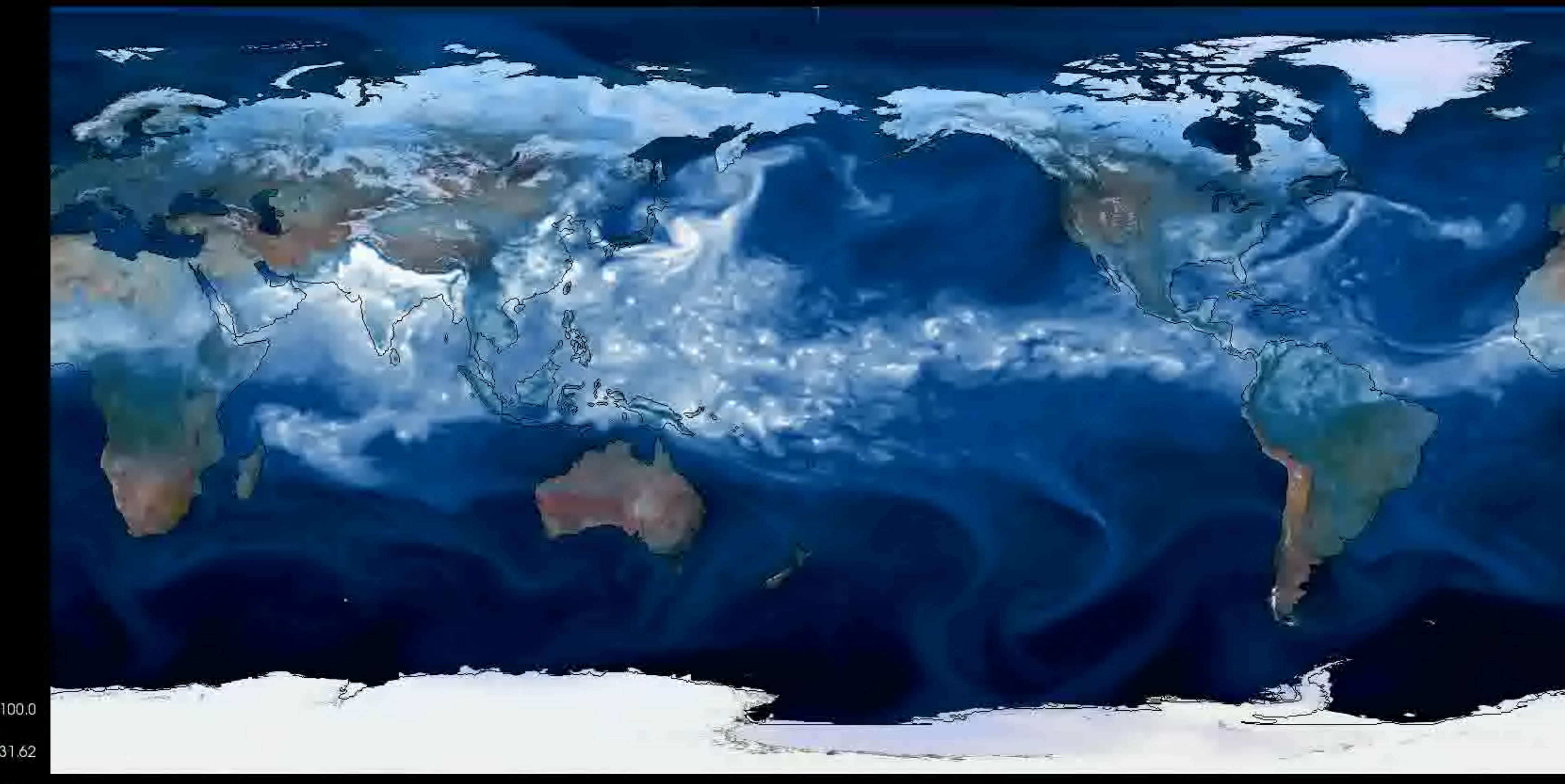
Carbon Cycle 2.0 at Berkeley Lab



Computing in Carbon Cycle 2.0



Simulations Show the Effects of Climate on Hurricanes

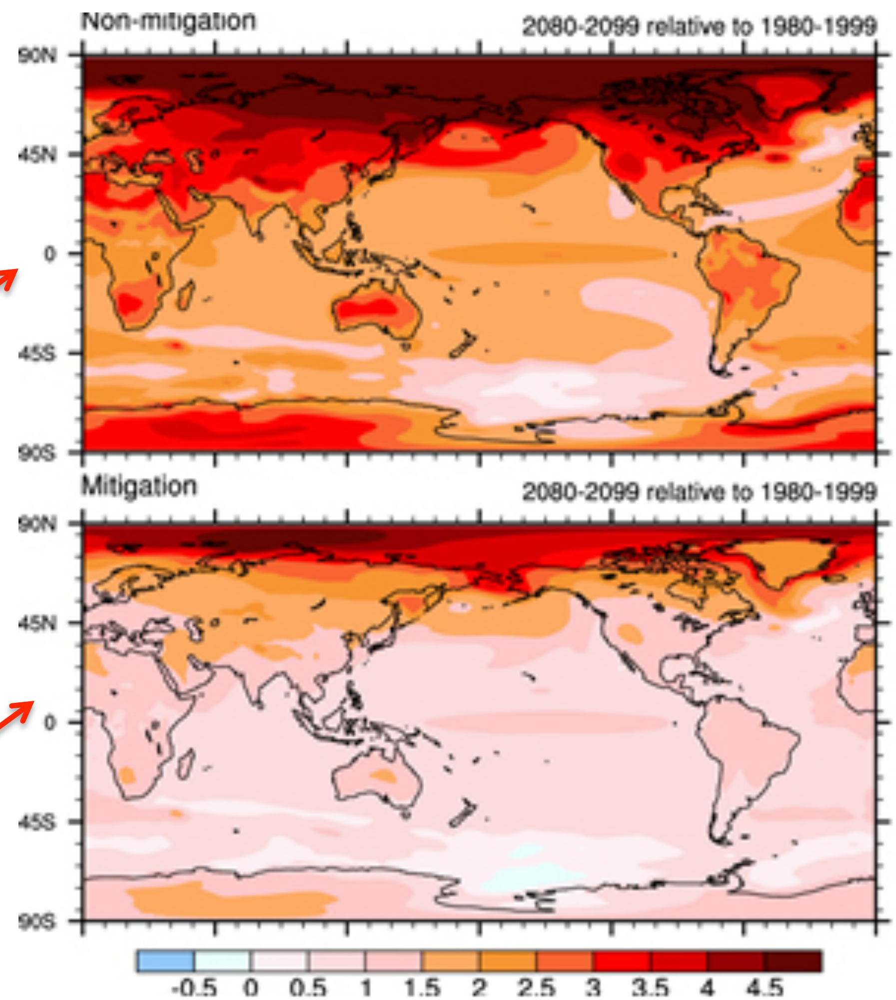


Mitigating Global Climate Change

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

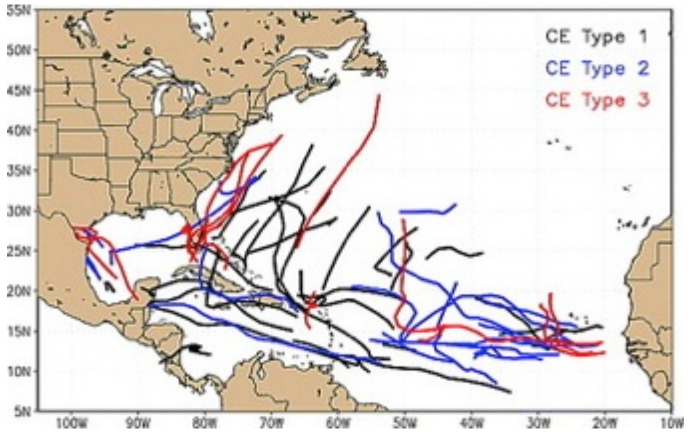
Temperatures rise by <2°C across nearly all populated areas if emissions are cut by 70%



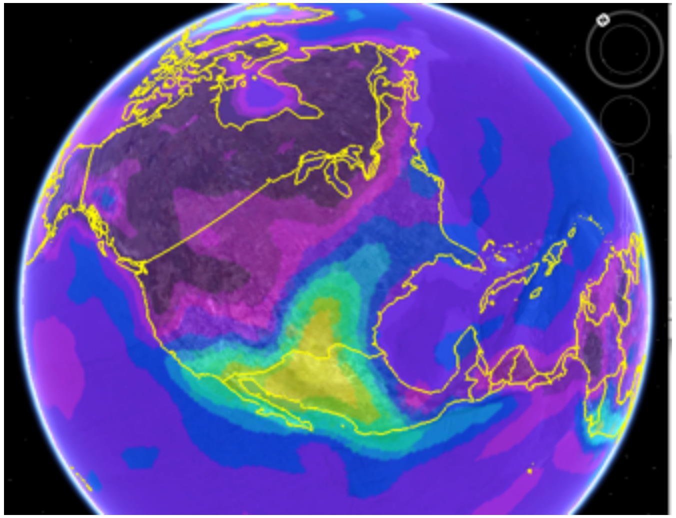
20^{th} Century Climate Data Reconstructed

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

- Data from meteorologists, military, volunteers and ships



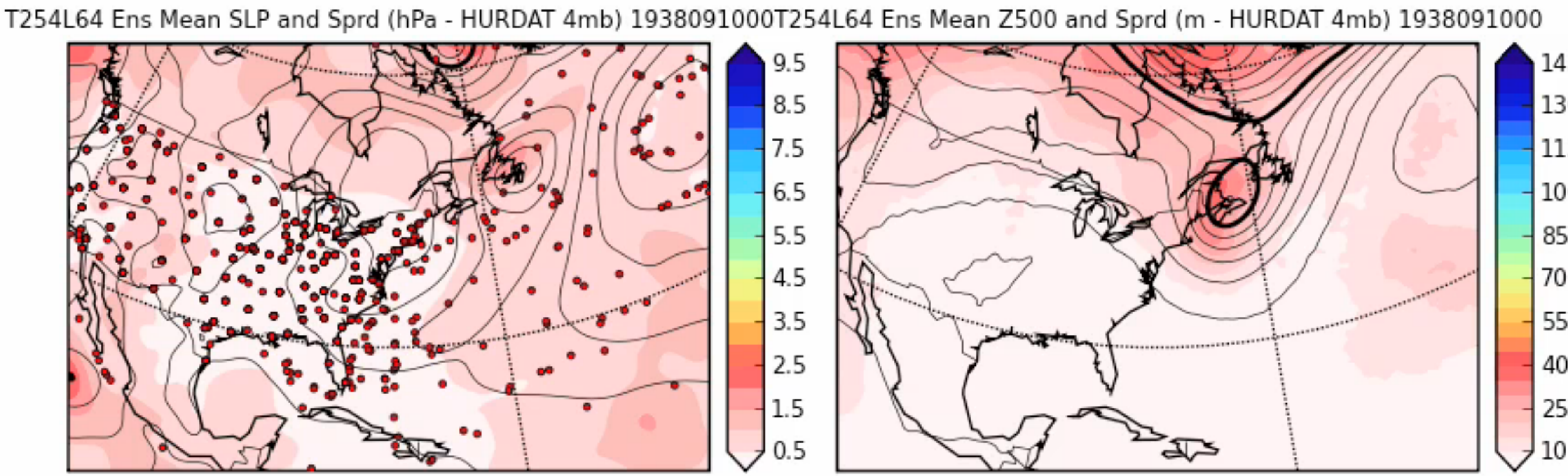
Previously undetected warm-core cyclones, *Geophys. Res. Letters*, 2011



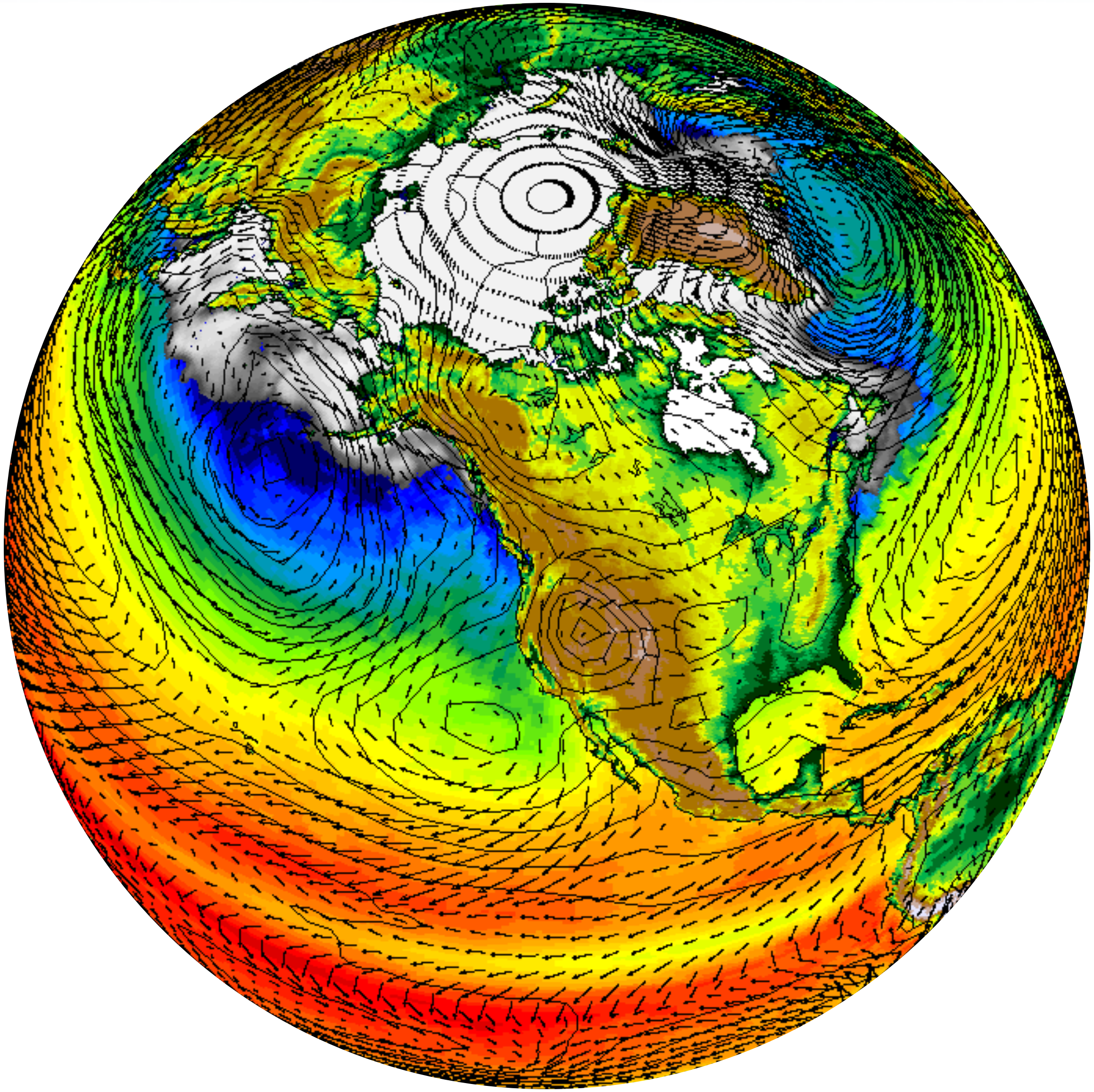
Relative Humidity for 1920-1929
Gil Compo, PI (U. Colorado)

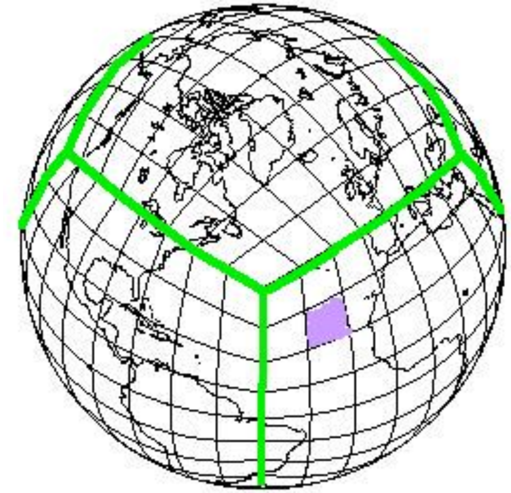
Climate Change Requires Lots of Data

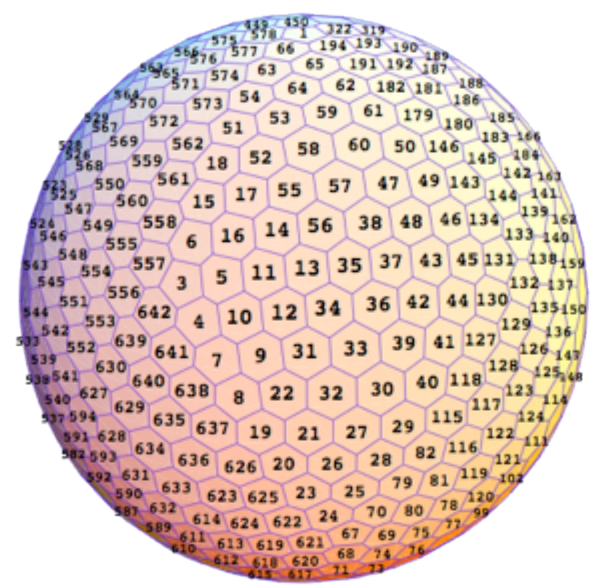
"validate" that the computer models are working as expected



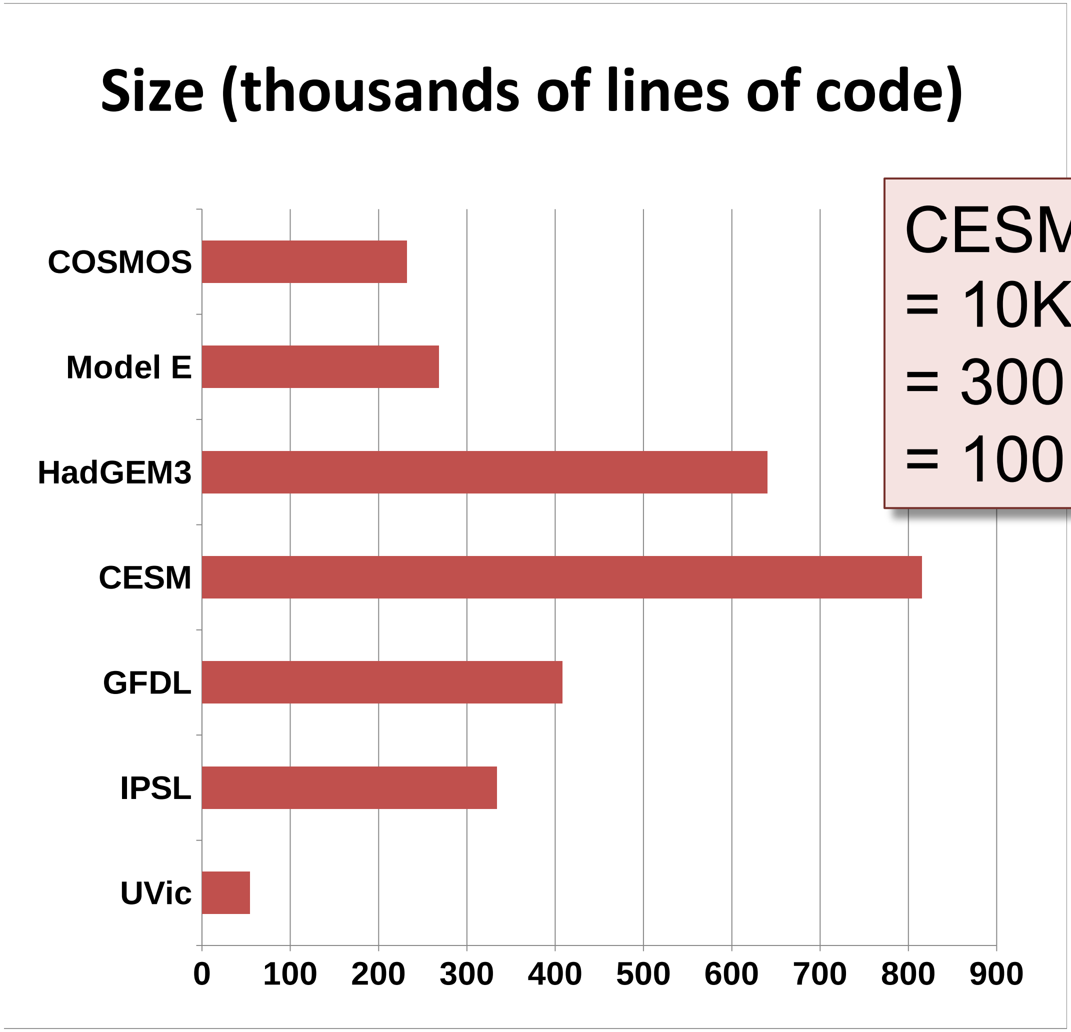
Data Structures for Simulations







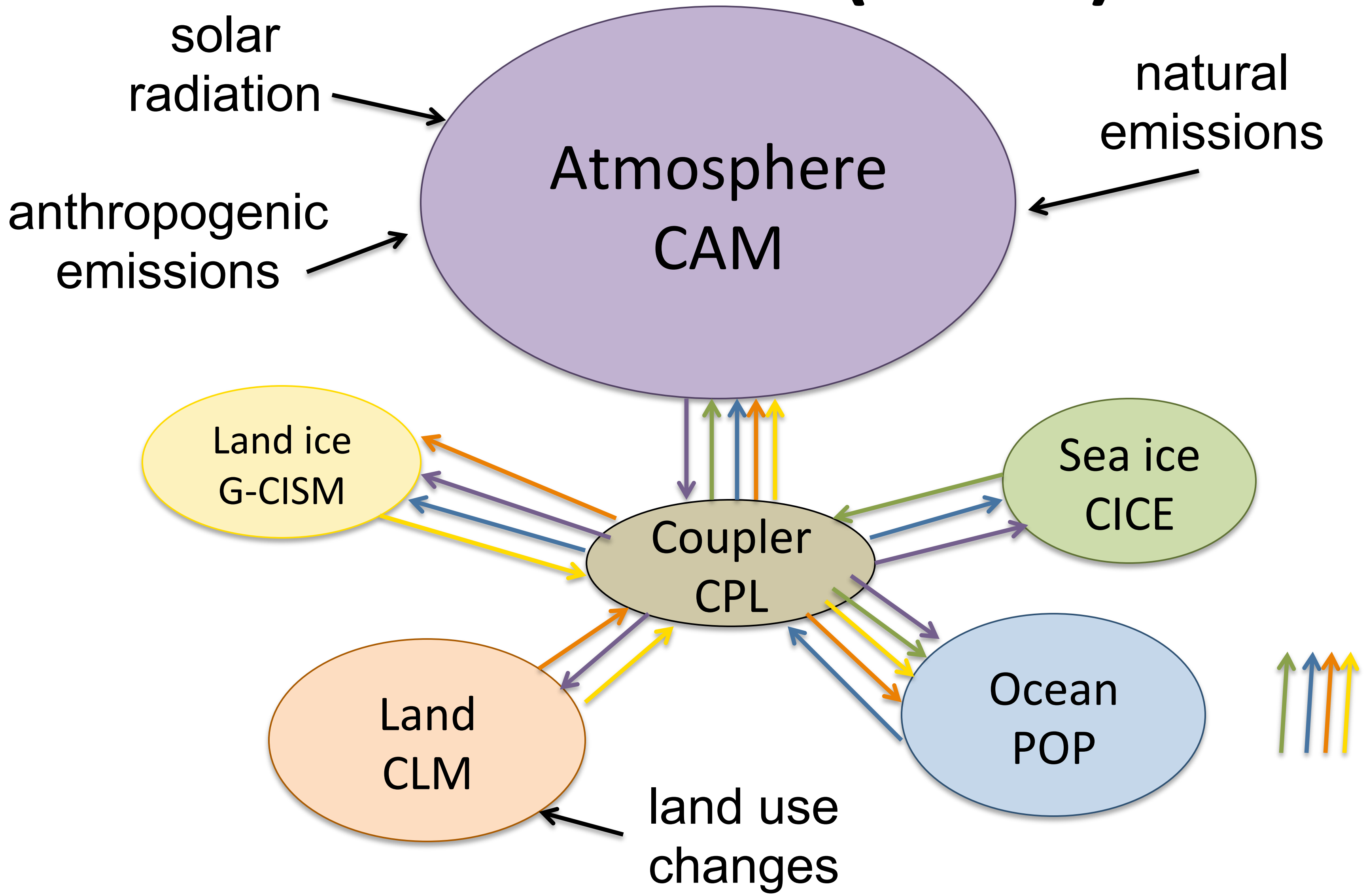
How big are these applications



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

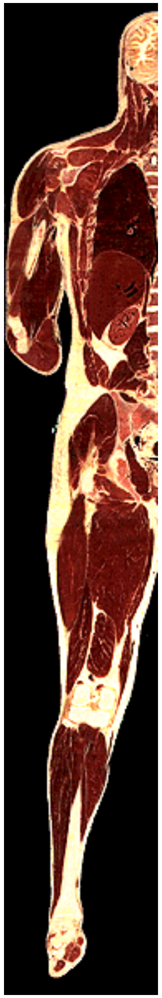
Video Edited by Sebastian Burke

From the proceedings of SCA 2009, New Orleans

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

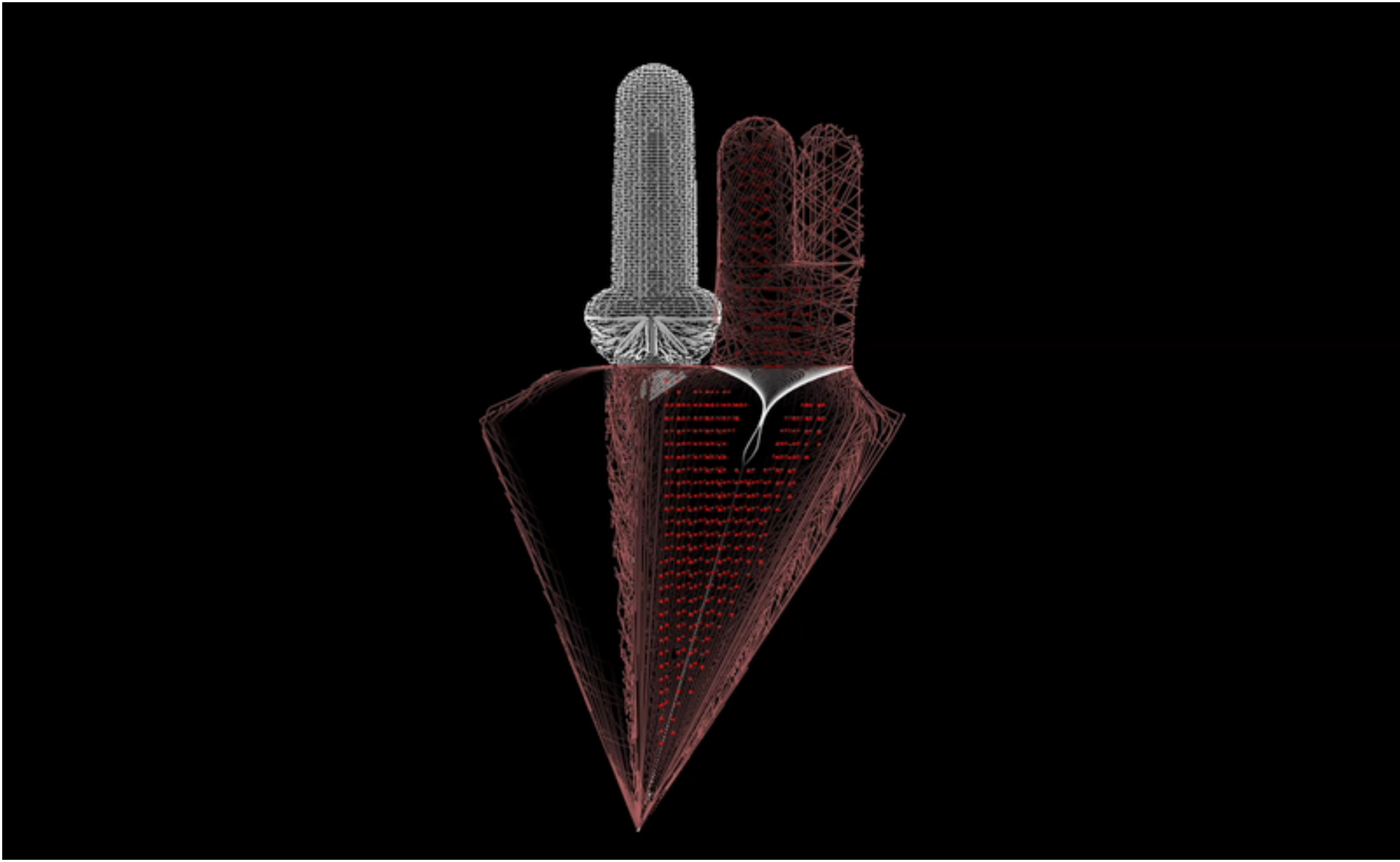
THE VISIBLE HUMAN PROJECT®

Image Source: www.madsci.org

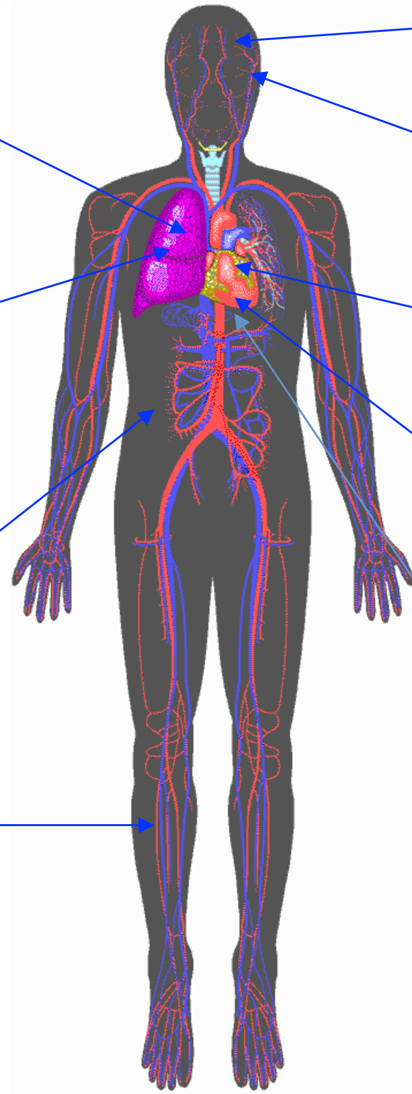
Experimental Data: Visible Human

Heart Simulation

Movie from Boyce Griffith's PhD thesis, NYU



Organ Simulation



Brain
UCSD (Ellisman), IBM/Lausanne/LBNL

Lung transport
Vanderbilt

Cochlea
Caltech, UMichigan

Lung flow
U. Iowa (Lin & Hoffman)

Cardiac flow
NYU, UCB, UCD…

Cardiac cells/muscles
SDSC, Auckland, UW, Utah

Kidney mesh generation
Dartmouth

Electrocardiography
Johns Hopkins,…

Skeletal mesh generation

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

The New York Times

Science

Search All NYTimes.com

WORLD | U.S. | N.Y. / REGIO | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | **JOBS** | **RE**

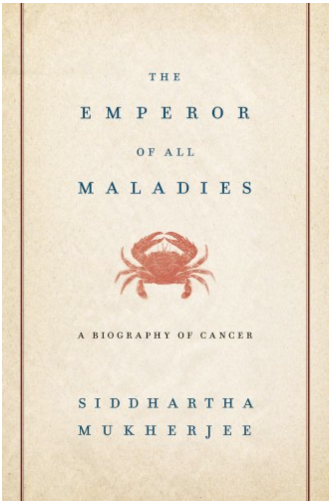
ESSAY

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.

David Patterson



See page 464

2010 article in Nature:

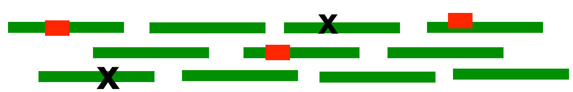
- Most scientists are self-taught in programming
- Only ⅓ think formal training in SW Eng is important
- < ½ have a good understanding of SW testing

Bug in SW supplied by another research lab forced UCSD Scripps Prof to retract 5 papers!

Two Problems in Genomics

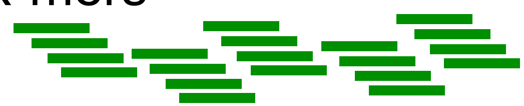
ASSEMBLY

reads



New fast I/O using SeqDB over HDF5

k-mers



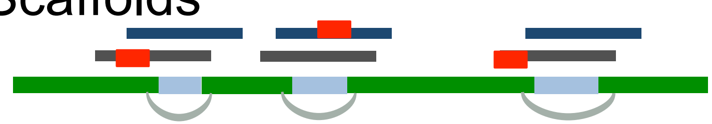
New analysis filters errors using probabilistic "Bloom Filter"

contigs

Graph algorithm scales to 15K cores

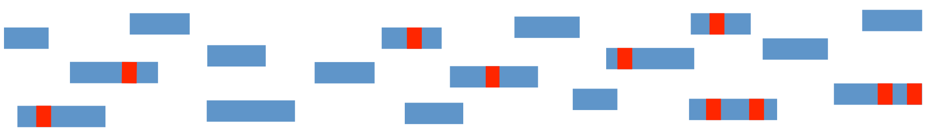
Human genome goes from: 44 hours to 20 secs

Scaffolds



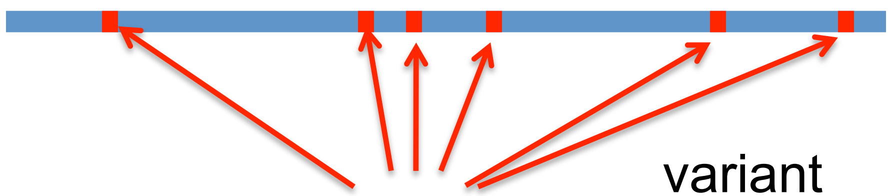
ALIGNMENT

reads



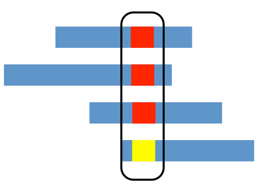
reconstructed genome

reference genome



variants

variant



consensus

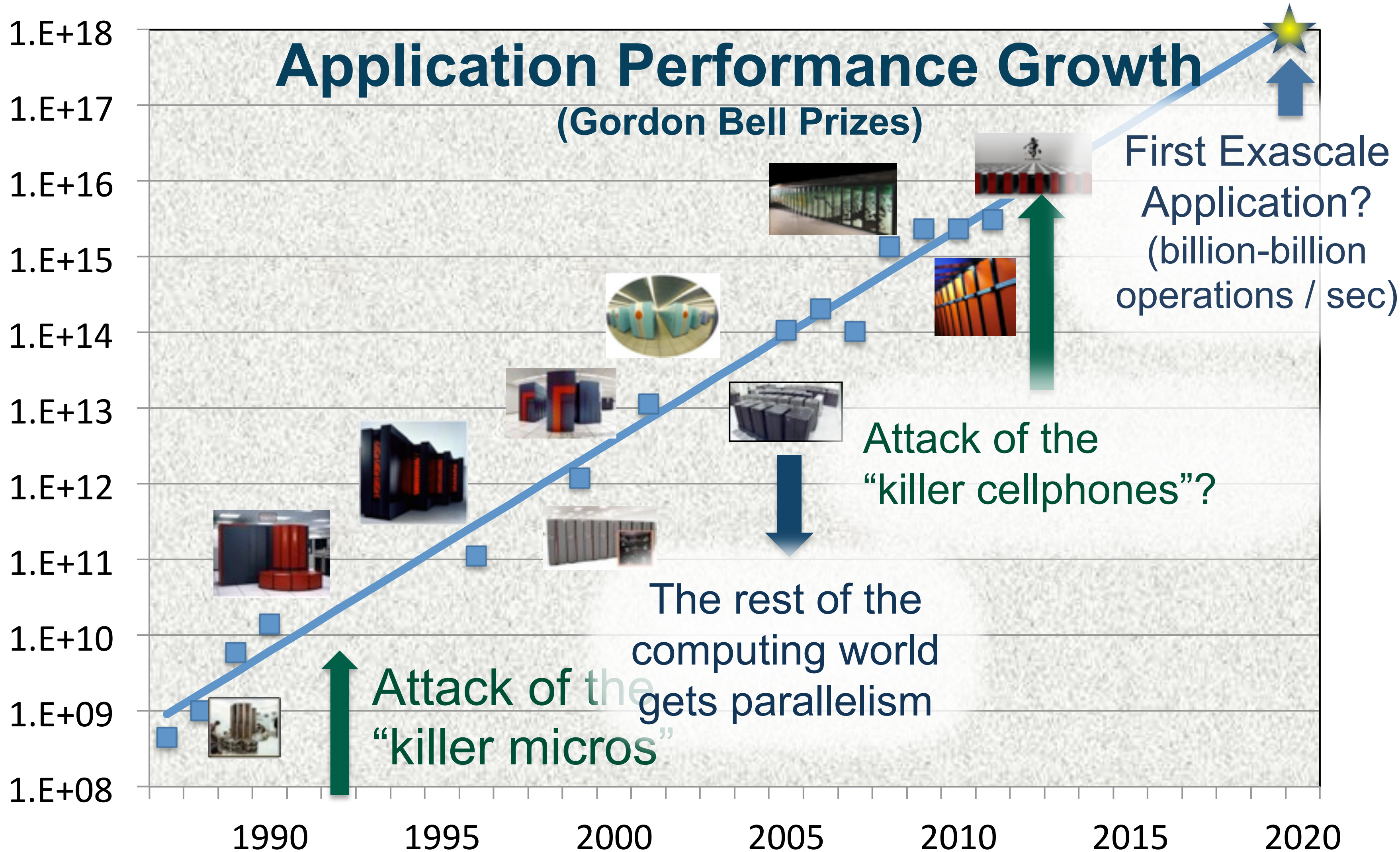
Runs on AMPLab software stack in the Cloud for $25

Trends in Computer Science

Which of the following are true?

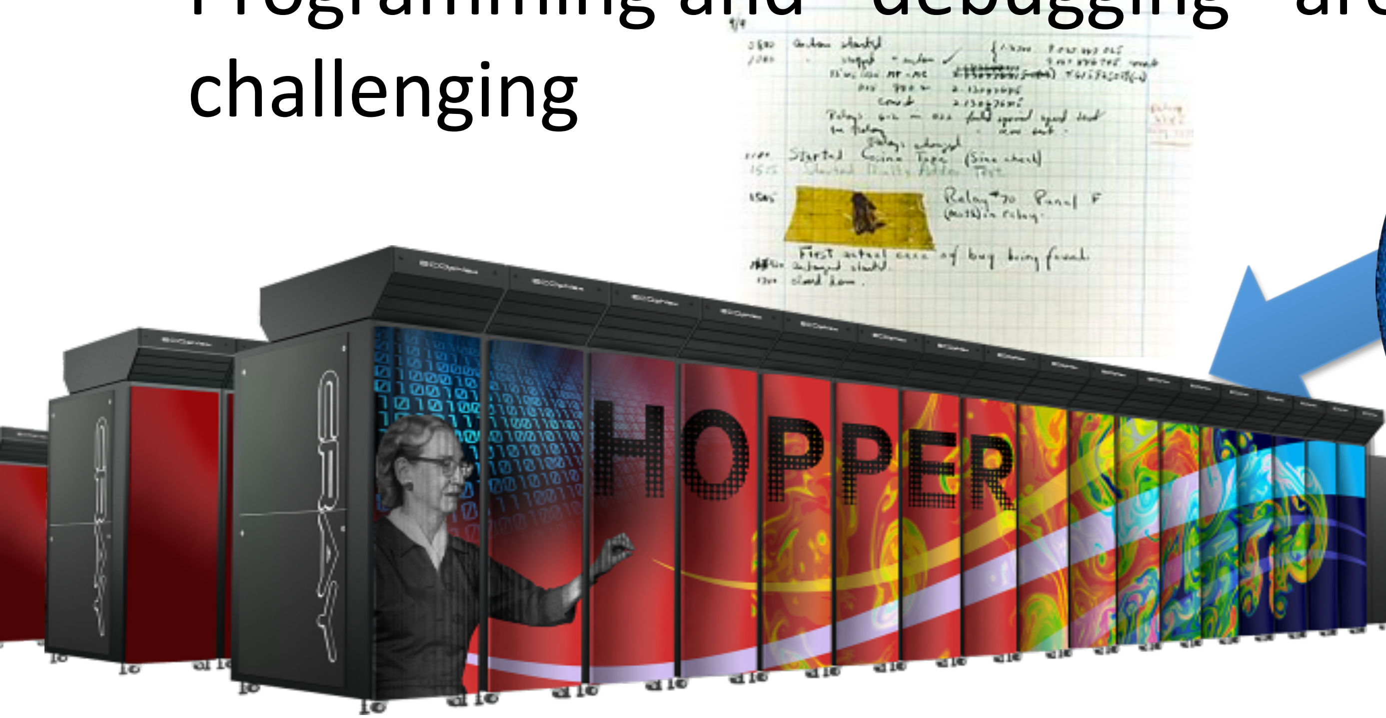
A. Moore's Law says that processor performance doubles every 18 months
B. Moore's Law has ended
C. Most of the time in scientific codes is spent doing arithmetic
D. None of the above
E. All of the above

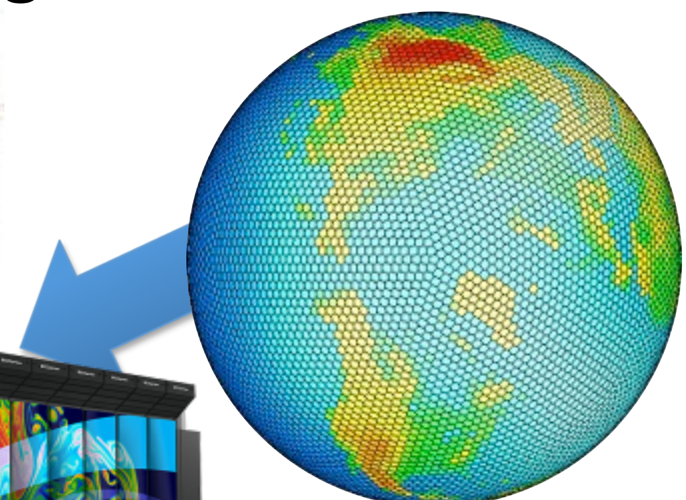
High End Computing Revolutions



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

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

Writing Software

Which of the following are true?

A. Most computer software is written by brilliant hackers, working alone

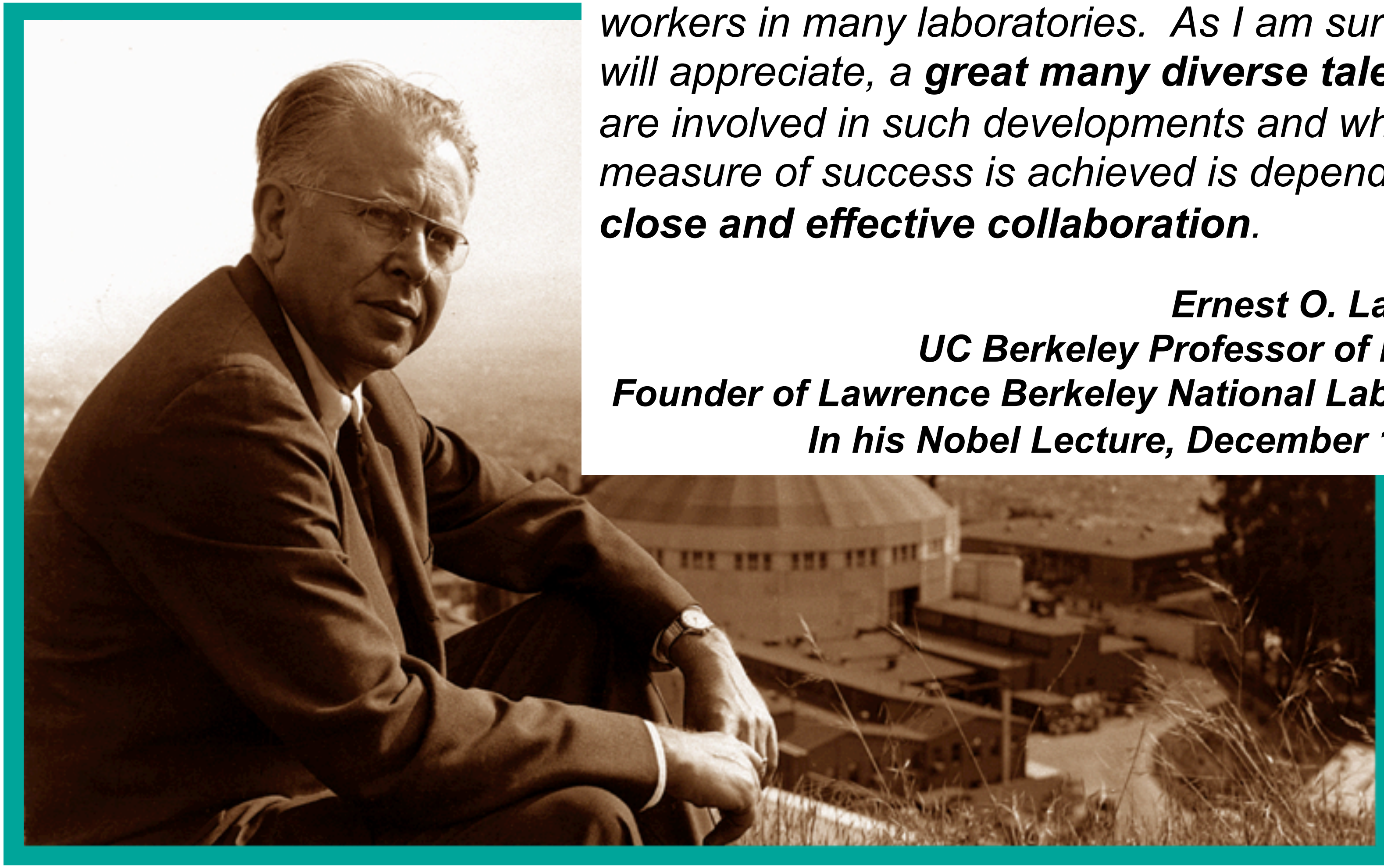
B. Parallel programming is a *solved problem*

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

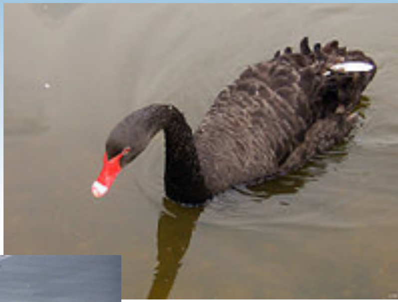
Computational Science is Necessarily Collaborative

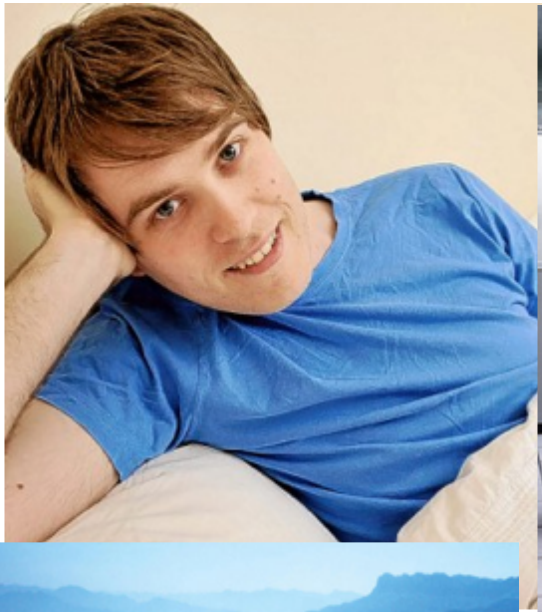


… as from the beginning the work has been a team effort involving many able and devoted co-workers in many laboratories. As I am sure you will appreciate, a ***great many diverse talents*** *are involved in such developments and whatever measure of success is achieved is dependent on* ***close and effective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

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

Black Swans of Computing
with 1992 Technology





Google™