U.S. Department of Energy

Office of Science

Role of High Performance Computing in BER

Jerry Elwood
Acting Associate Director of Science
for Biological and Environmental Research

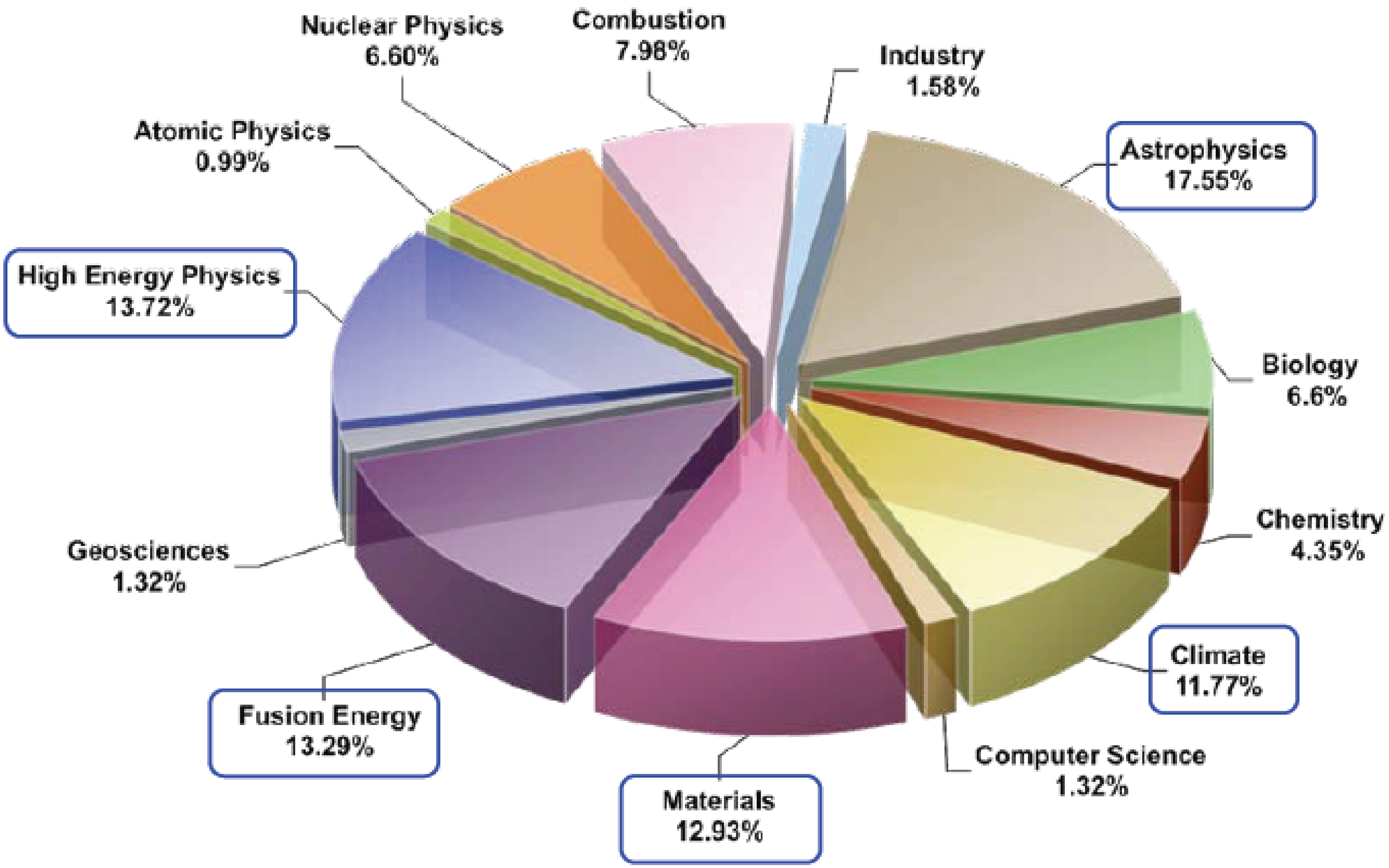
ASCAC Meeting
August 14, 2007

High Performance Computing (HPC) in BER

Three science application areas:

1. Computational Biology – still nascent but growing
2. Contaminant Transport and Fate in Subsurface Environments – expanding needs, e.g., geologic sequestration of carbon
3. Climate and Earth System Modeling – mature with expanding needs & applications

2006 INCITE Allocations at ORNL NCCS Breakdown by Discipline



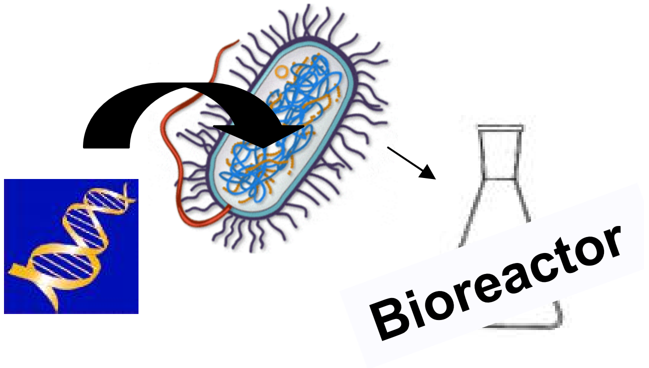
Computational Biology Areas in BER, e.g.,

- Using simulations to investigate complex actions of enzymes critical to producing biofuels, e.g., cellulosic ethanol
- Developing approaches for computational assignment of gene function(s)
- Simulating functioning of complex biological systems

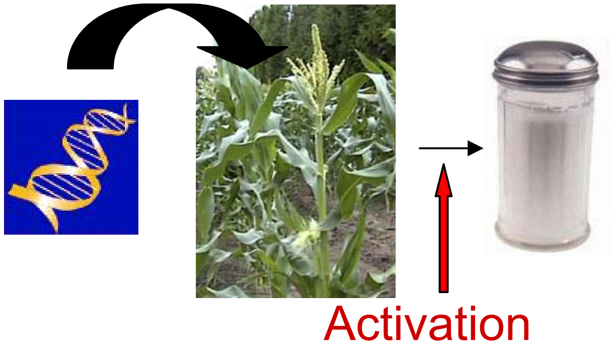
Computational Biology Role in Research on Bioenergy

- Develop more efficient enzymes for cellulose and hemicellulose hydrolysis

- Engineer fermenting organisms to secrete cellulases and hemicellulases,

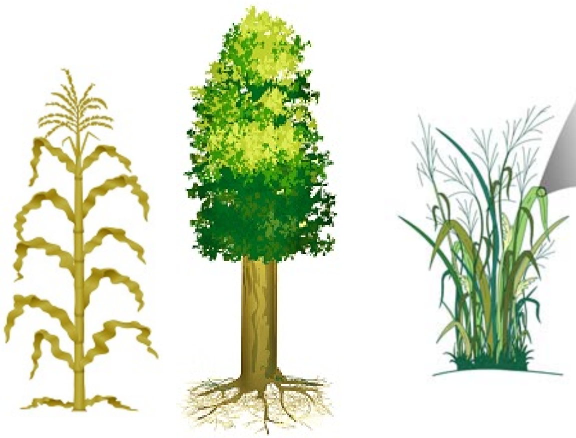


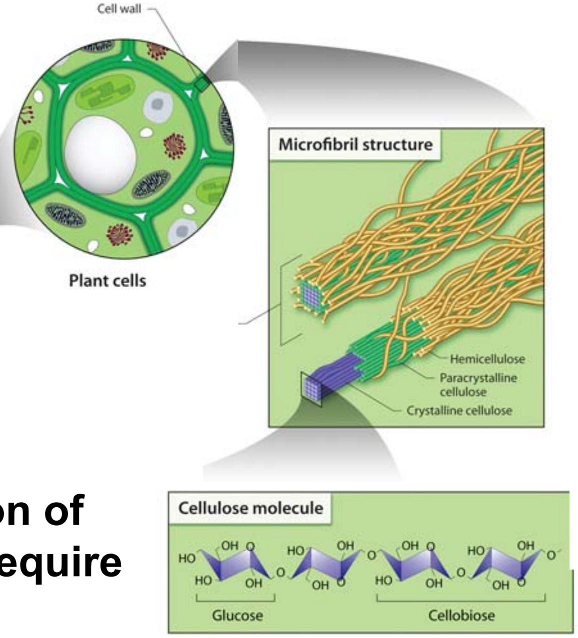
- Engineer feedstock crops with desired traits for producing biofuels

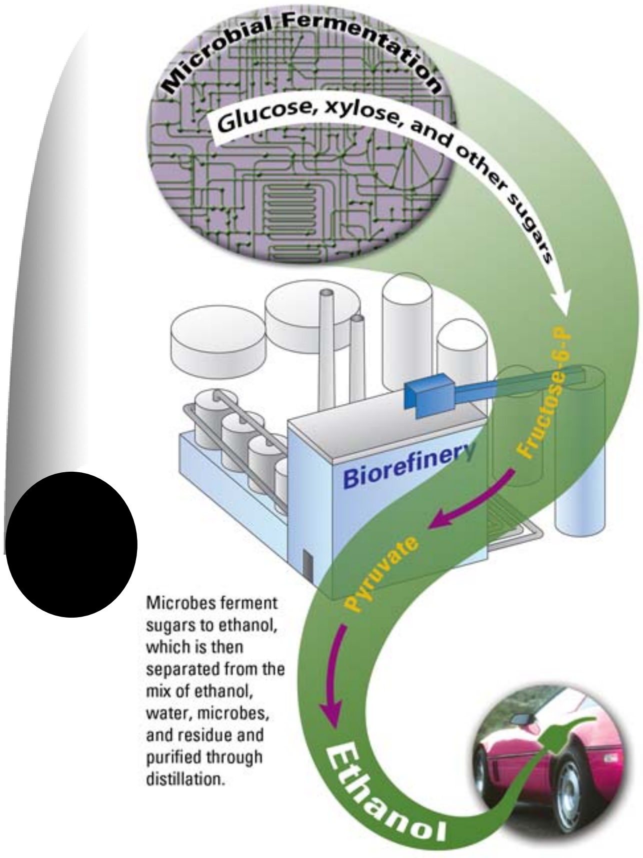


Basic Transformation

Lignocellulose → Sugars → Ethanol





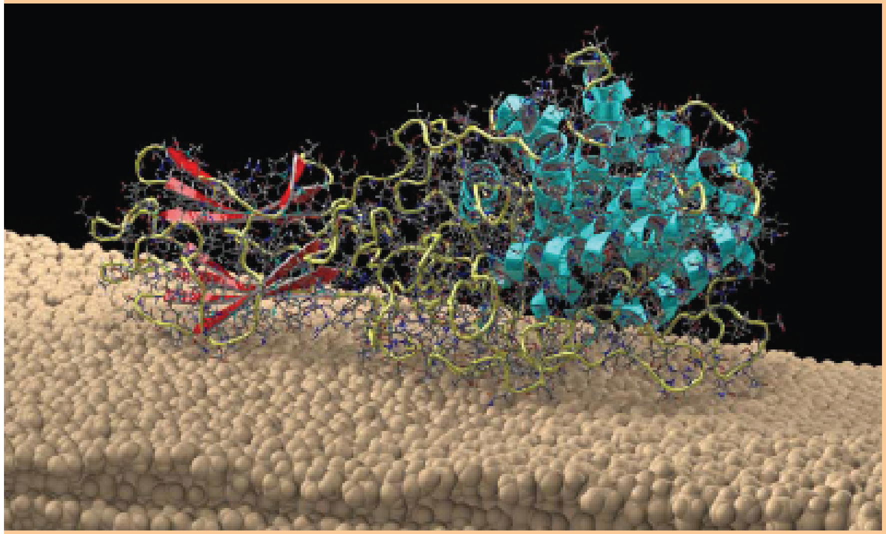


Microbes ferment sugars to ethanol, which is then separated from the mix of ethanol, water, microbes, and residue and purified through distillation.

- Cost-effective production of cellulosic biofuels will require transformational breakthroughs in basic science
- The scientific problem is highly challenging, but very well defined

Computational Biology with Implications for Biomass Conversion

- Exploring 300,000-atom cellulase enzyme converting cellulose to sugars
- Jaguar is enabling simulation of the activity of the cellulose enzyme on a time scale of 10 to 15 nanoseconds using LAMMPS and AMBER codes
  - 2000 processor simulations in ’06
  - 8000+ processor simulations in ’07



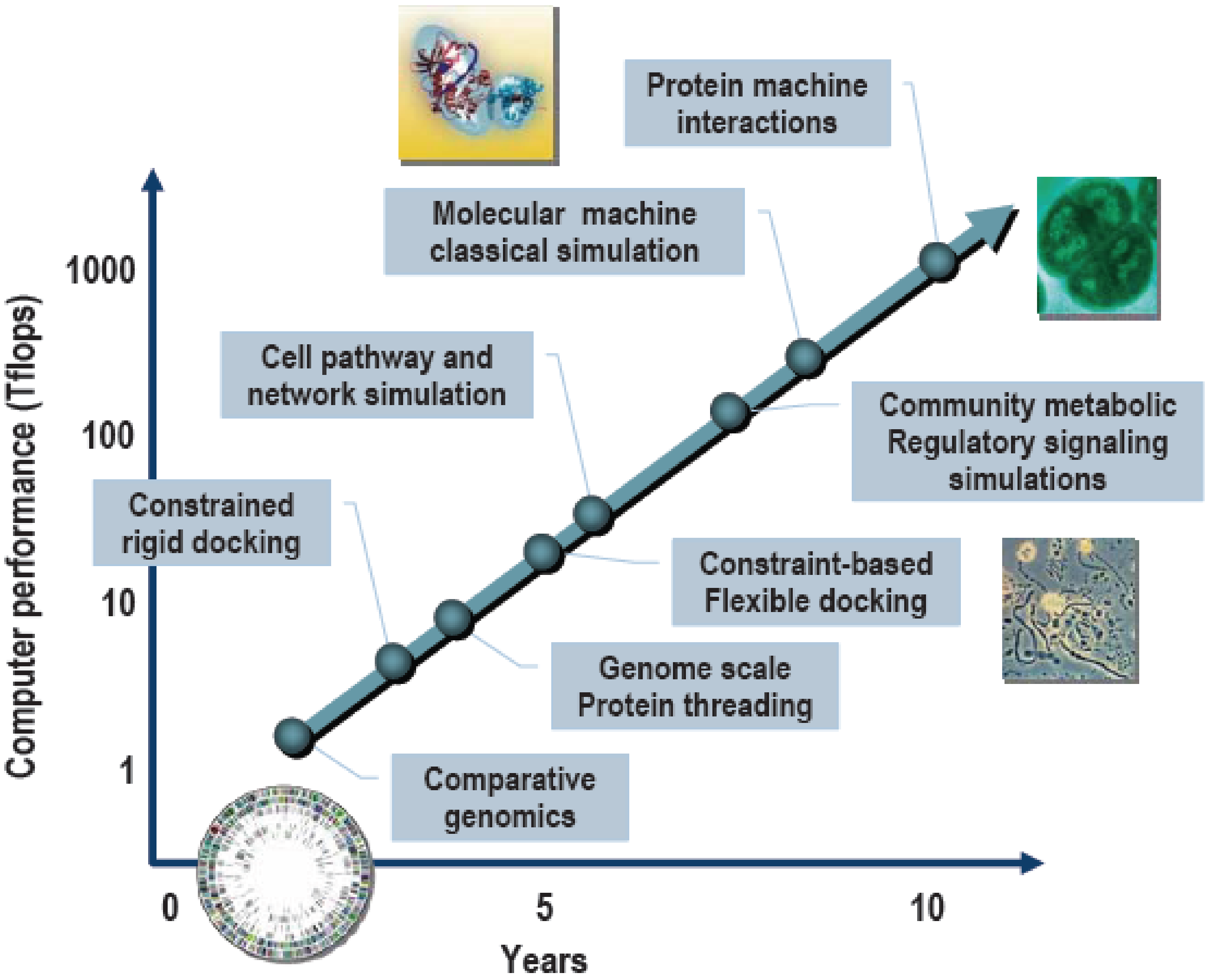
Cellulase enzyme in action on cellulose surface based on model by Brady (Cornell) and provided by Himmel (NREL).

- Simulations showed that interior “vibrations” in enzymes influence the rate at which they carry out chemical reactions
- Understanding these vibrations sheds insight on how to manipulate the makeup of enzymes to speed up or slow down their reactions
- Efficient, low-cost bio-ethanol conversion from improved cellulases is one driver for this research

2007 INCITE Projects on Computational Biology

- Next-generation simulations in biology: investigating biomolecular structure, dynamics, and function through multi-scale modeling
- Gating mechanisms of membrane proteins

Biology Roadmap



Comparative genomics

Constrained rigid docking

Genome scale Protein threading

Constraint-based Flexible docking

Cell pathway and network simulation

Community metabolic Regulatory signaling simulations

Molecular machine classical simulation

Protein machine interactions

Expected outcomes

5 years

- Metabolic flux modeling for hydrogen and carbon fixation pathways
- Constrained flexible docking simulations of interacting proteins

10 years

- Multiscale stochastic simulations of microbial metabolic, regulatory, and protein interaction networks
- Dynamic simulations of complex molecular machines

Environmental Remediation Science Division (ERSD)

Subsurface Reactive Transport Modeling and Simulation

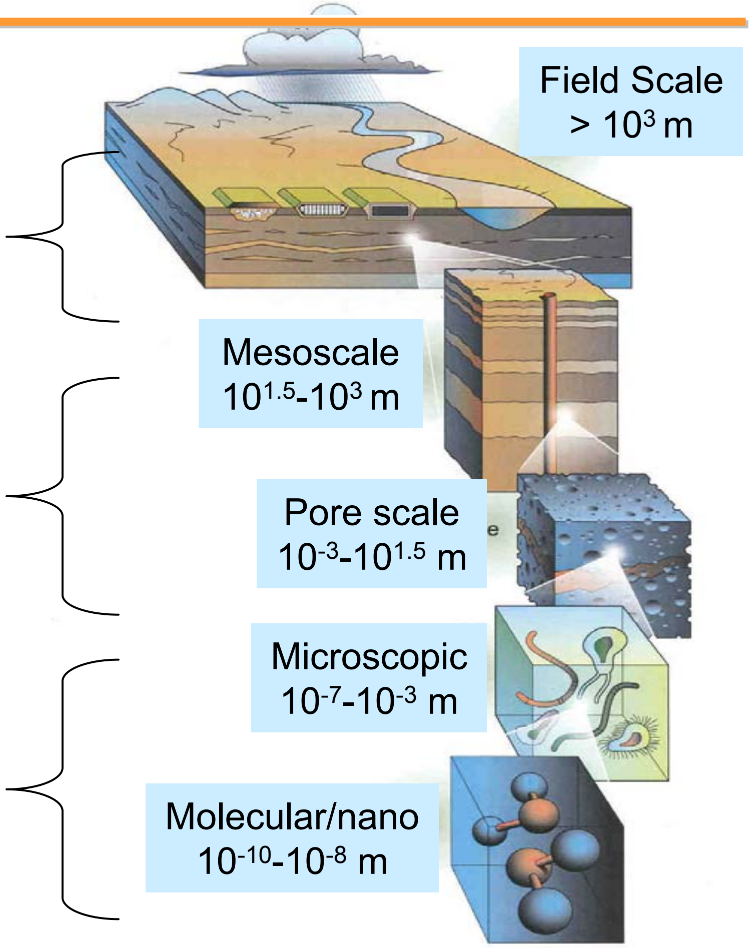
Predictions of contaminant transport at the field scale depend on processes occurring at smaller scales.

Characteristics of groundwater flow affecting contaminant transport at intermediate scales:

- advection/dispersion
- diffusion
- stratigraphy (heterogeneity)
- porosity

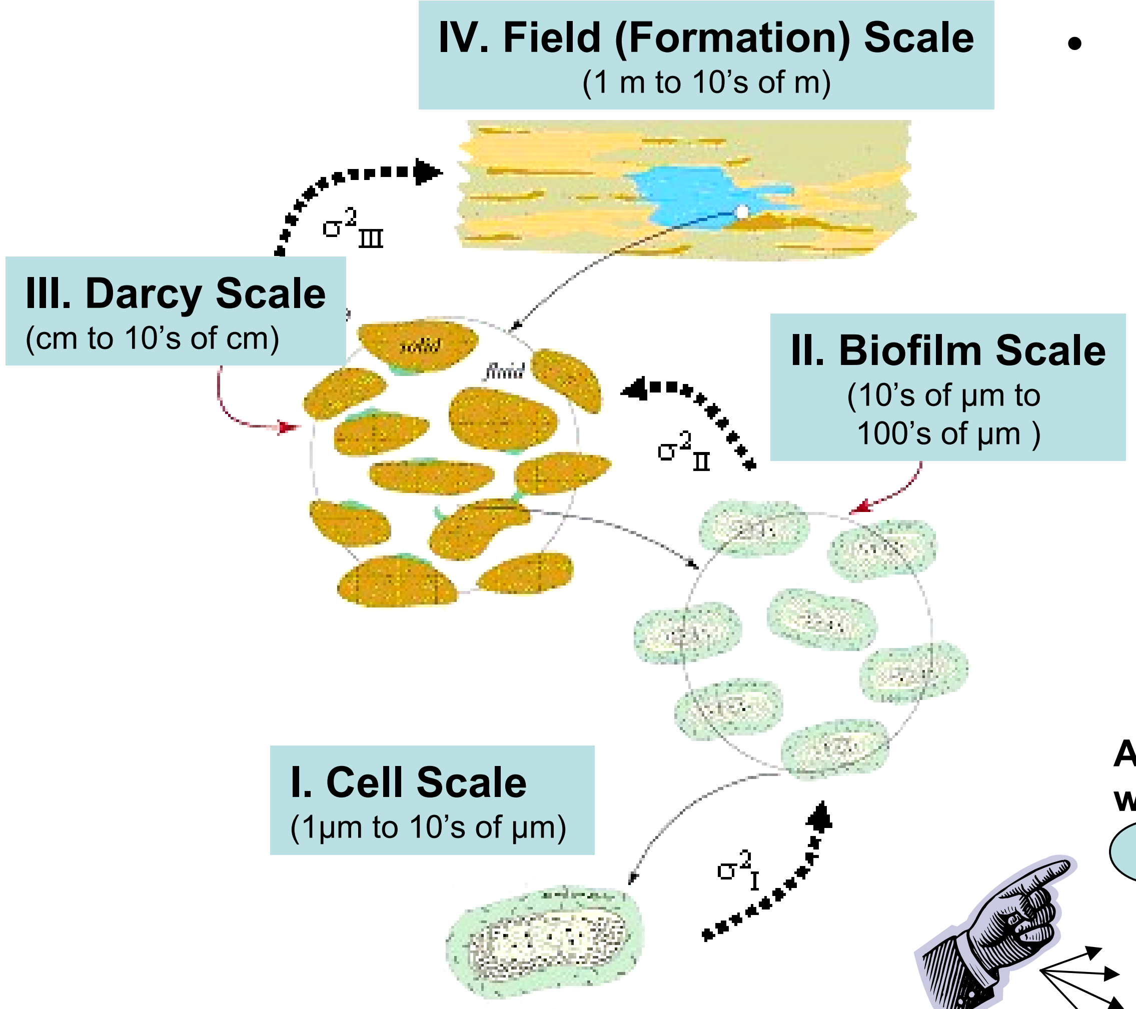
Reactive processes control contaminant transport occur at small scales:

- chemical oxidation/reduction
- biological oxidation/reduction
- adsorption
- precipitation/dissolution



High Performance Computing (HPC) needed to quantitatively describe and link processes affecting contaminant transport across many orders of magnitude (>10^{13}) change in scale in the environment

Hybrid Models to Describe Contaminant Transport Across Scales (SciDAC – T. Scheibe, PNNL)



IV. Field (Formation) Scale
(1 m to 10's of m)

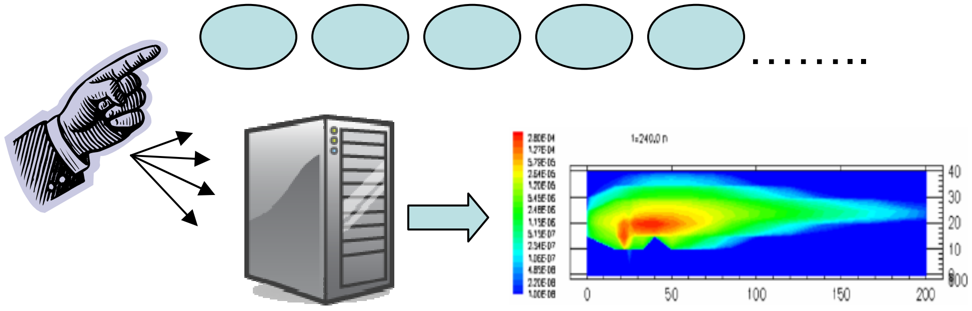
III. Darcy Scale
(cm to 10's of cm)

II. Biofilm Scale
(10's of µm to 100's of µm)

I. Cell Scale
(1µm to 10's of µm)

- **Hybrid Multiscale Modeling**
    - Directly couple models with different process representations at fundamental scales
    - Adaptively control which processes / models are applied in which parts of the model domain
    - Integrate process models into a component-based workflow environment
    - Wide potentially applicability to subsurface investigations

Available process models in workflow environment



Tartakovsky et al. 2007 (submitted)

High Performance Computing Needs for Contaminant Transport Research

Need HPC to:

1) Incorporate uncertainty quantification into contaminant transport predictions
   - incorporate parameter estimation and uncertainly analyses in subsurface simulations

2) Develop high resolution numerical systems of the subsurface for numerical experiments on classes of subsurface problems/processes
   - multiphase-multicomponent contaminant transport, CO_2 injection, heterogeneity effects, vadose zone transport.

3) Investigate a wider ranger of scales in the subsurface
   - for the Hanford 300 Area, a 50m x 50m x 15m test plot at 20cm resolution is approximately ~ 10^{10} degrees of freedom

4) Multi-media modeling

Joint ERSD/ASCR SciDAC Projects

Groundwater Reactive Transport Modeling & Simulation

Two new FY07 SciDAC “Groundwater” projects

- New Science Application for SciDAC
- Current models (Kd approach) are overly simplistic 1D approximations
- Few examples of HPC use in subsurface contaminant transport research

Modeling Multiscale-Multiphase-Multicomponent Subsurface Reactive Flows using Advanced Computing

Lead PI: Peter Lichtner, LANL (https://software.lanl.gov/pflotran)

Adapting the parallel code PFLOTRAN to a 3D, high resolution simulation of uranium transport in the subsurface at the Hanford 300 Area. This effort is fully integrative (Lichtner is a Co-PI) with a larger ERSD-funded, field scale, subsurface research project at the Hanford 300 Area.

Hybrid Numerical Methods for Multiscale Simulations of Subsurface Biogeochemical Processes

Lead PI: Tim Scheibe, PNNL (http://subsurface.pnl.gov/)

Develop an integrative modeling framework linking process models at different scales into a component-based, workflow environment to facilitate descriptions of biogeochemical reactions governing contaminant transport at DOE sites.

BER Climate Change Research

DOE Climate Change portfolio involves process research and modeling efforts …

- to improve understanding of factors affecting the Earth's radiant-energy balance;
- to predict accurately any global and regional climate change induced by increasing atmospheric concentrations of aerosols and greenhouse gases;
- to quantify sources and sinks of energy-related greenhouse gases, especially carbon dioxide; and
- to improve the scientific basis for assessing both the potential consequences of climate change, including the potential ecological, social, and economic implications of human-induced climatic changes caused by increases in greenhouse gases in the atmosphere and the benefits and costs of alternative response options.

Climate Change Research Questions:

- How do natural and anthropogenic factors influence past, present, and future climate?
- How will the hydrologic cycle and ecological systems respond to these influences?
- How will natural systems amplify or reduce human influences on climate?
- What are optimal (and sub-optimal) methods for adapting to and mitigating climate change?

Earth System Modeling (ESM)

Goal: Evaluate climate understanding with Earth system models

- Subsidiary goals are to understand
  - Sensitivity to climate forcings
  - Sensitivity to cloud processes
  - Interactions of biogeochemistry (e.g., carbon cycle) and physical climate
- Proposed simulation: global 0.1-degree ESM

  (Requirement: ~0.5 PF/s sustained)
- Extensions: economic/energy technology feedbacks

Drivers in Climate Modeling Research

- Basic research enterprise is producing a steady stream of new knowledge about climate processes, climate dynamics, numerical methods, etc.

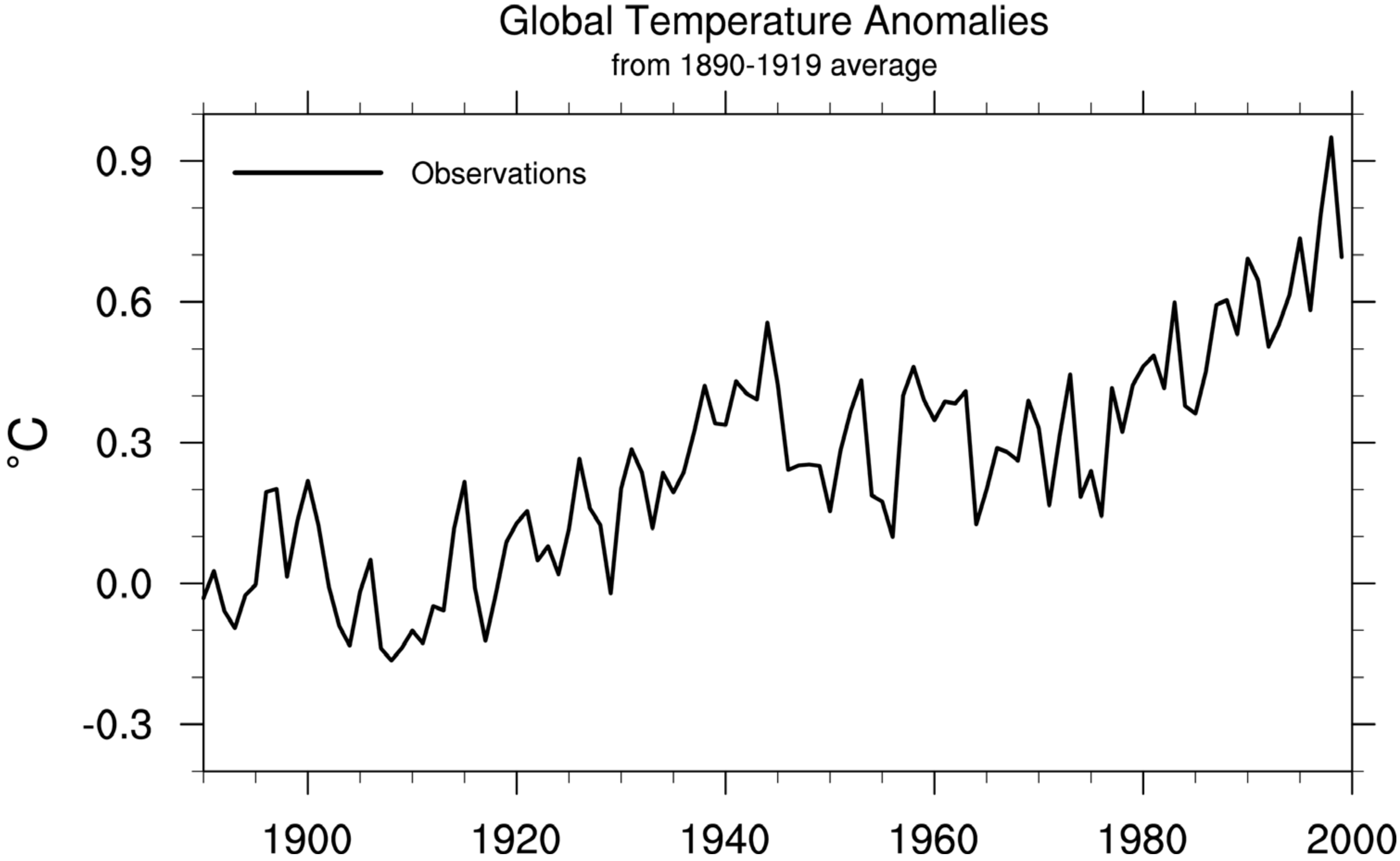
- Quantity, quality and diversity of observational data to support climate research is increasing

- Demand for model applications increasing in both magnitude and sophistication

- We are dependent on climate and Earth system models to answer questions about long-term climate response to human-induced forcing

- Computing capability is increasing dramatically,
  but …

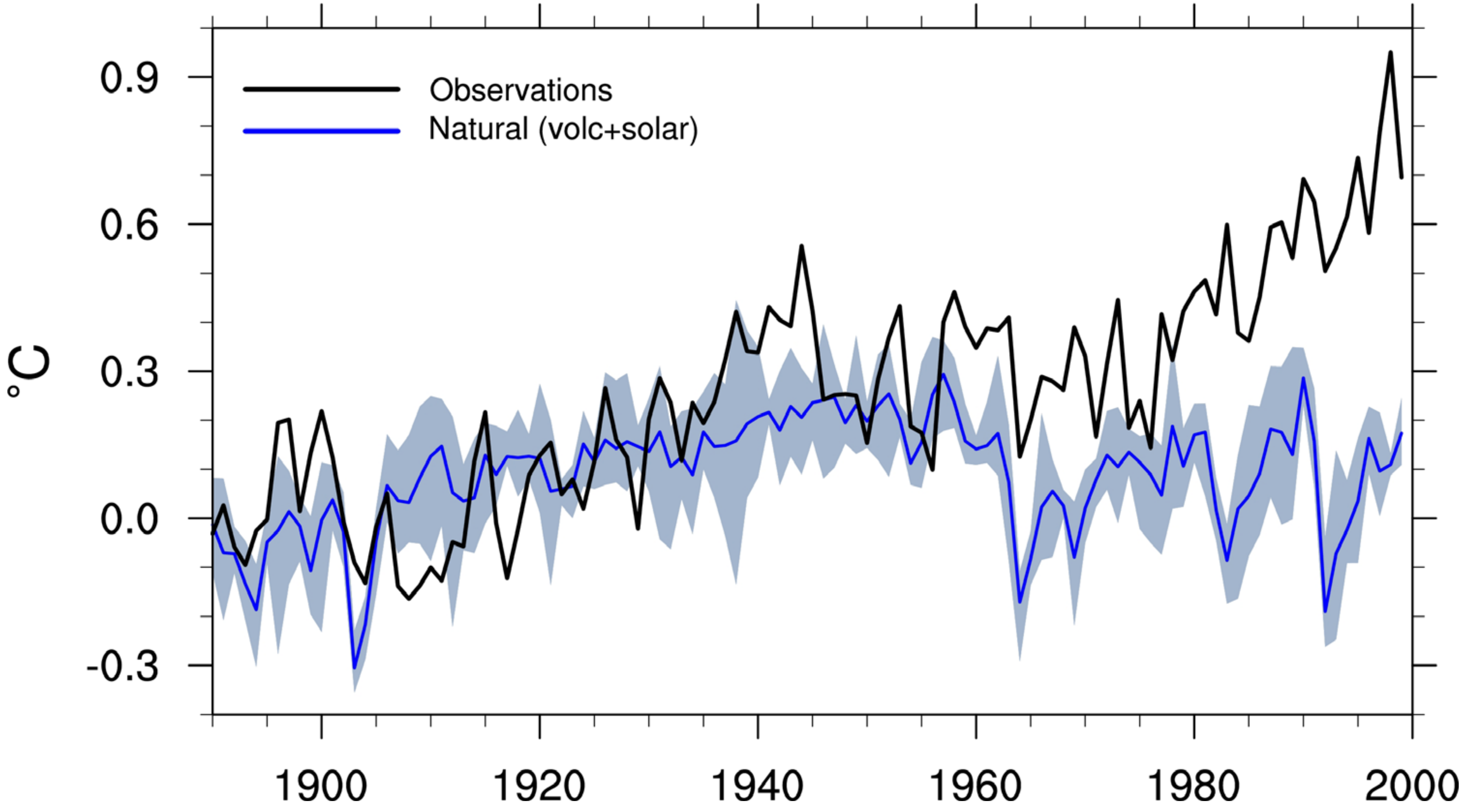
Use of models in climate change attribution – what caused the observed changes in global average temperatures over the past century?



Parallel Climate Model Ensembles

Global Temperature Anomalies

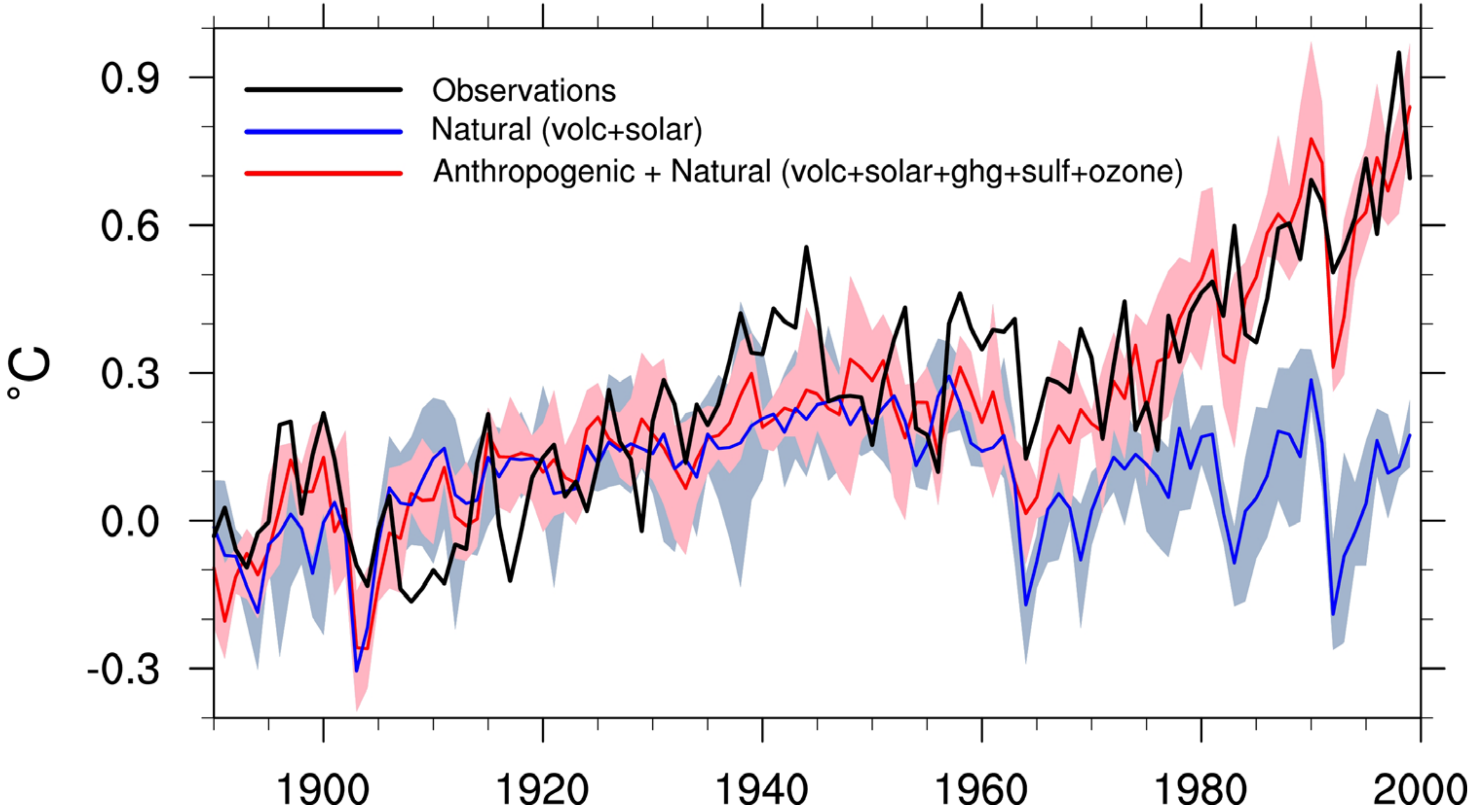
from 1890-1919 average



Parallel Climate Model Ensembles

Global Temperature Anomalies

from 1890-1919 average



(Meehl et al., 2004, *Journal of Climate,* **17**, 3721–3727.)

“Most of the observed increase in globally averaged temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations”

---IPCC Fourth Assessment Report, 2007

A key outstanding question that emerged from the IPCC AR4 Report:

Assuming the countries of the world undertake mitigation policies, and assuming we can target a certain amount of climate change with those mitigation policies (i.e. stabilized concentrations or radiative forcing or warming), then what is the exact time-evolving nature of the regional climate changes that human societies and natural systems must adapt to?

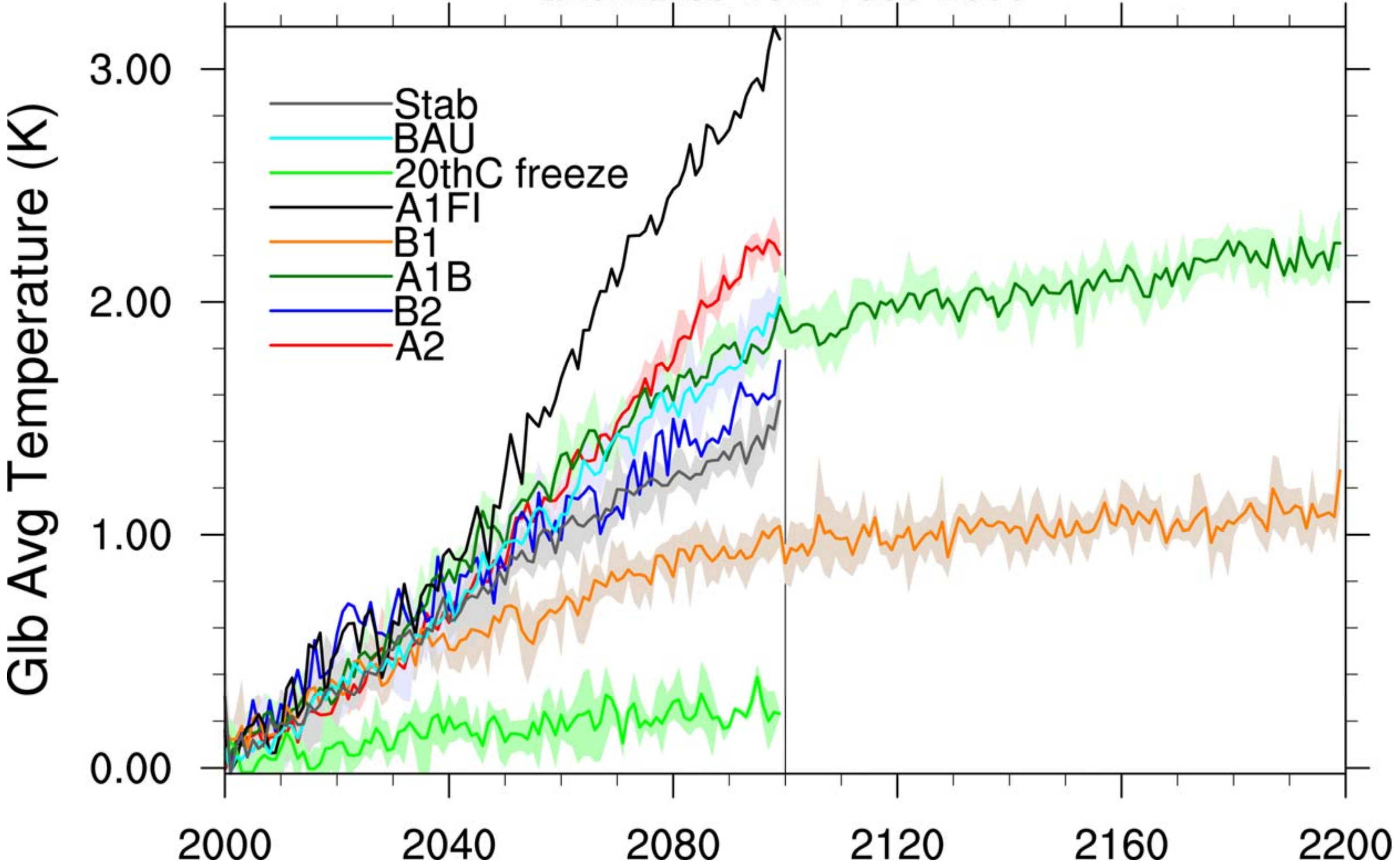
Puts greater pressure than ever before on climate models to produce “forecasts” for decades into the future

Requires close coordination across the climate science, impacts and scenario/integrated assessment communities

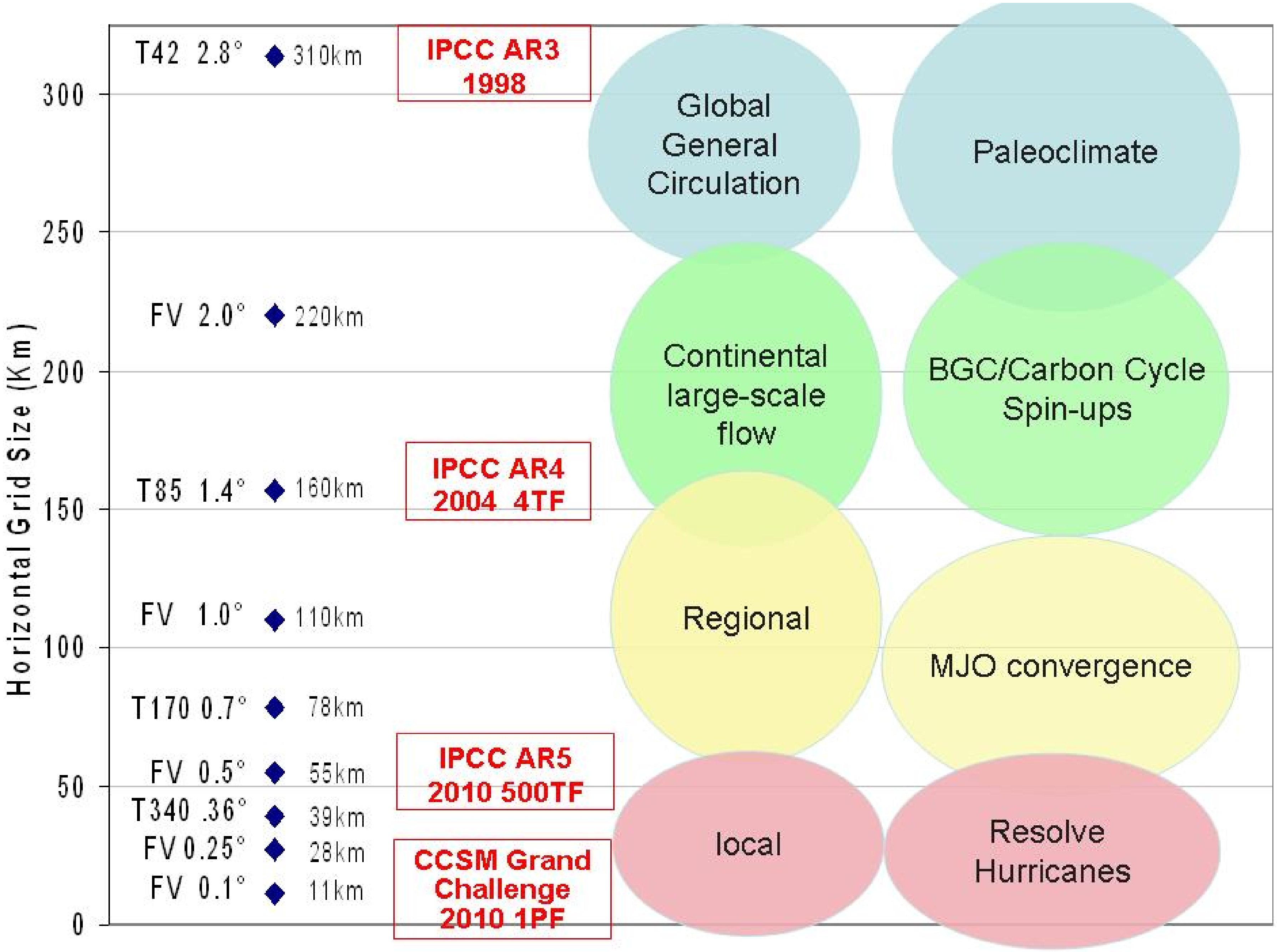
Requires enhanced research programs in climate and earth system modeling and associated computing and human analysis resources

PCM IPCC future scenarios

anomalies from 1990-2009



At 2000, we were already committed to 33% more warming (0.2°C) compared to that seen in the 20th century; future commitments past 2100 are comparable in magnitude

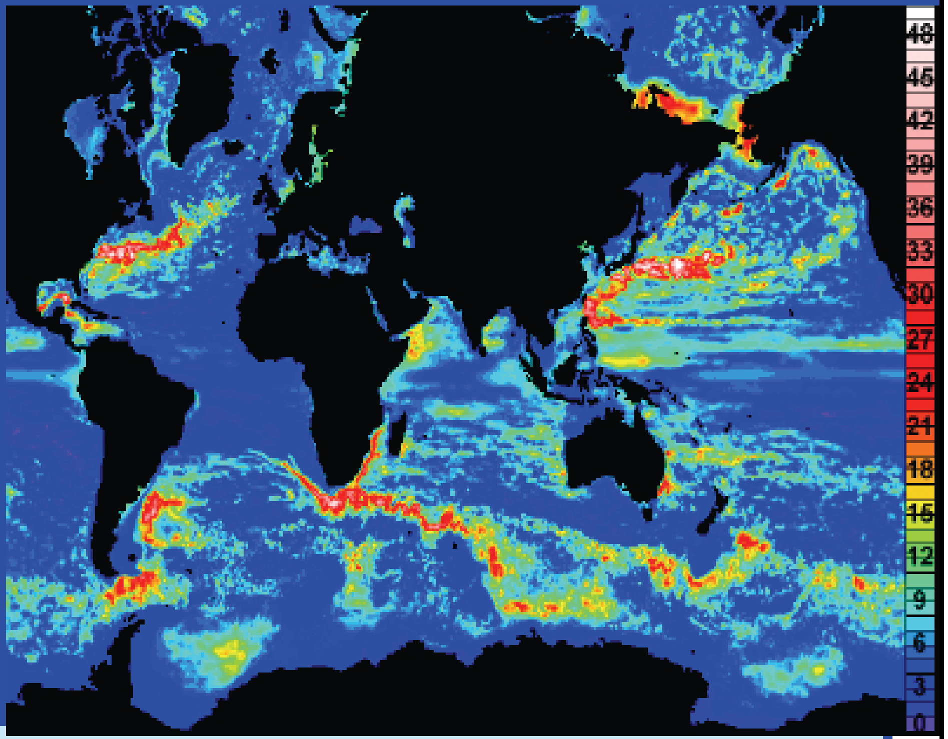


Horizontal grid size, time, and complexity of climate and earth system models

Understanding the Ocean's Role in Trapping Carbon Dioxide

> *"…half of the carbon dioxide that has been emitted over the last 100 years or so currently resides in the atmosphere. The rest is in the ocean…"*
>
> Synte Peacock, U. of Chicago, PI

- Simulation promises to increase understanding of the ocean's role in regulating climate, as a repository for greenhouse gases
- The most fine-grained, global-scale simulations ever of how the oceans work
  - **New knowledge of the currents and processes at work in the oceans**
  - **details about possible transport of gases and chemicals released into the ocean**



First-ever 100-year simulation of the ocean at a *fine enough scale to include the relatively small, circular currents known as eddies.* Until recently researchers lacked the computing power to simulate eddies directly on a global scale.

Project looks into the fate of trapped heat and greenhouse gases

Climate Change Simulations: IPCC AR4 Runs on DOE Machines

DOE ORNL Supercomputer:

- IBM p690 (Cheetah)
- 684 Processors
- Peak speed: 4.5 Teraflops

Characteristics of the Community Climate System Model (CCSM3):

- ~1 quadrillion operations/simulated year
- IPCC ~10,800 years
- Rate of simulation: 3.5 sim. years/day
- Output: 10 GB/simulated year
- Data volume for IPCC: ~110 TB (~200,000 Data CDs)

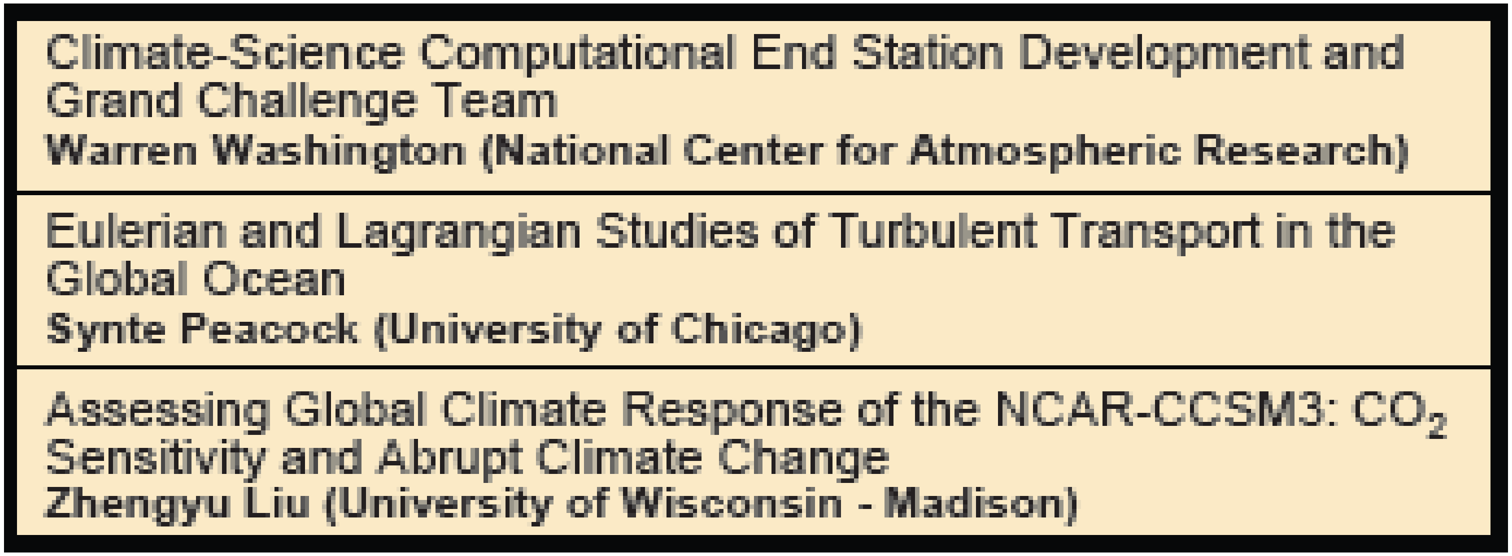
DOE Resources for IPCC runs (CCSM, PCM)

- ORNL 2.34M CPU hr on cheetah
- NERSC 1.4M CPU hr on seaborg

Substantial DOE prior developmental work:

- Ocean & Sea Ice model development at LANL
- Atmosphere & Land model testing/tuning at ORNL

2007 INCITE Projects on Climate Modeling

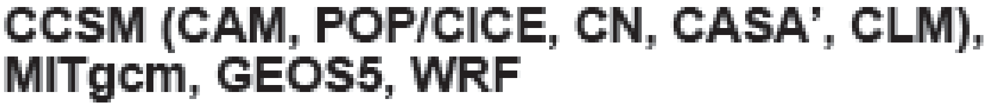


Climate-Science Computational End Station Development and Grand Challenge Team
Warren Washington (National Center for Atmospheric Research)

Eulerian and Lagrangian Studies of Turbulent Transport in the Global Ocean
Synte Peacock (University of Chicago)

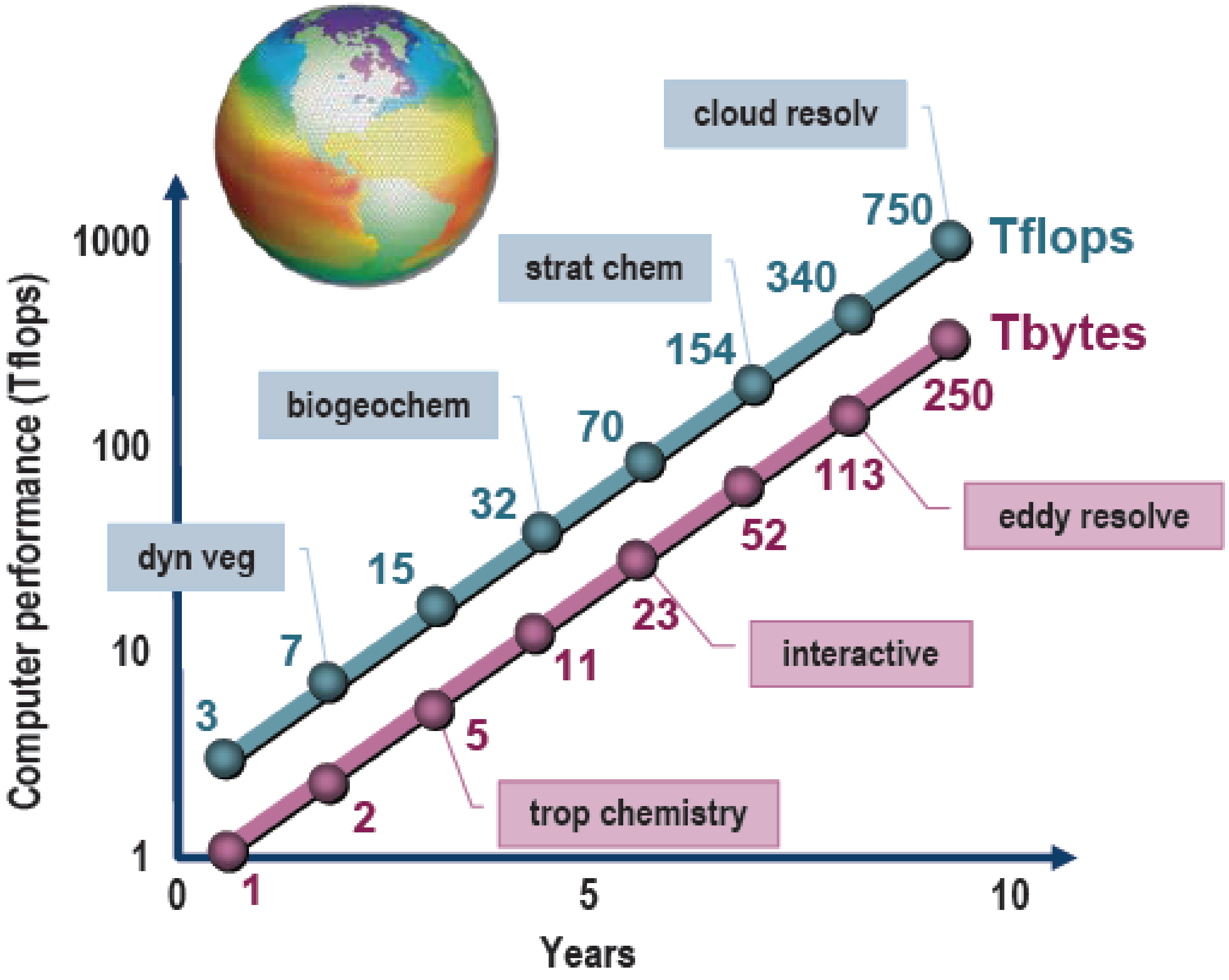
Assessing Global Climate Response of the NCAR-CCSM3: CO_2 Sensitivity and Abrupt Climate Change
Zhengyu Liu (University of Wisconsin - Madison)

CODES THAT HAVE DEMONSTRATED SCALABILITY:



CCSM (CAM, POP/CICE, CN, CASA', CLM), MITgcm, GEOS5, WRF

Climate Roadmap



Expected outcomes

5 years

- Fully coupled carbon-climate simulation
- Fully coupled sulfur-atmospheric chemistry simulation

10 years

- Cloud-resolving 30-km spatial resolution atmosphere climate simulation
- Fully coupled, physics, chemistry, biology earth system model