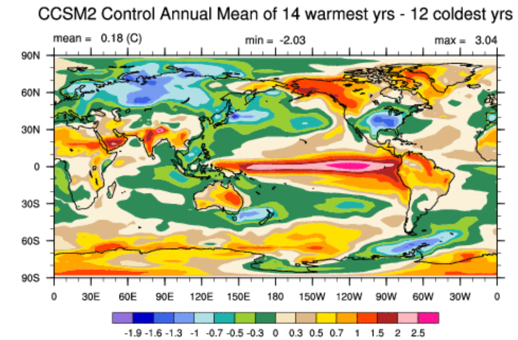
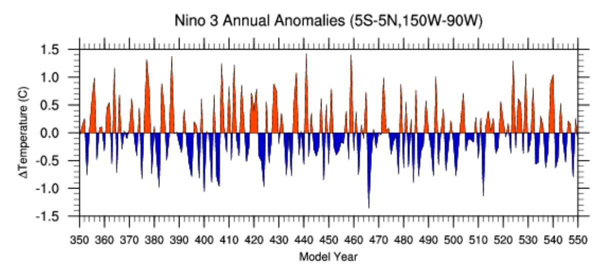
SC03 BOF on the State of the Art

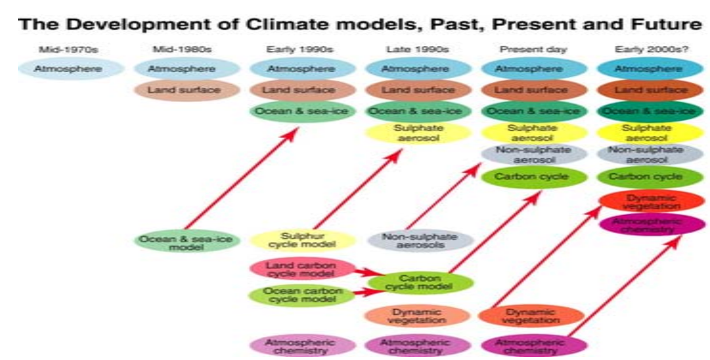
John Drake
Oak Ridge National Laboratory

Why Climate Prediction is Compute Limited

- Long time integrations:
  - Historical validation 1870-2000
  - Future scenarios 2000-2200
- Comprehensive, coupled processes
  - Models still under development
  - Nonlinear feedbacks and sensitivities
- Multi-scale interactions
- Need for ensemble forecasts
- Decision support scenarios



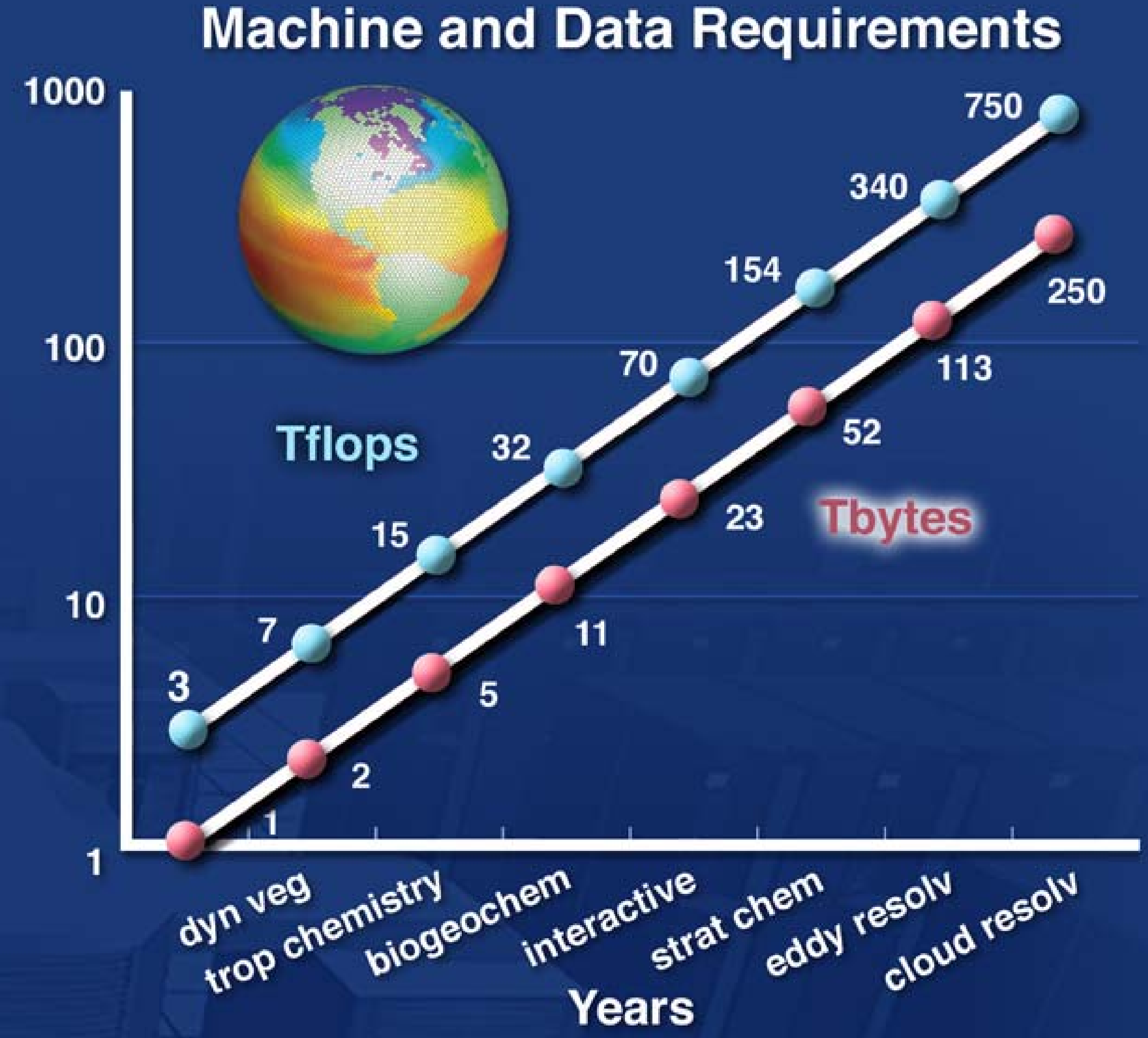




The climate simulation challenge:

- We must supply the type of information needed by decision-makers that can enable them to understand the impacts of climate change and the options for dealing with it.
- We must demonstrate the ability to simulate current and past climate on the regional scale (global averages do not tell the story).
- We must reduce discrepancies among climate models (or at least be able to explain them).
- We must be able to simulate the interactions among all aspects of the climate system in order to do this.

Climate



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

What new science could be achieved?

- Increased spatial resolution
  - Current model resolution (~ 250 km horizontal) is inadequate for simulating impacts, for example.
- Increased completeness
  - Important interactive physical, chemical, and biological processes are currently omitted because of computational complexity.
- Increased physical fidelity
  - More realistic and accurate representation of sub-grid processes (e.g., clouds) would be possible.
- Increased length of control and climate-change runs
  - Better understanding of natural climate variability, stability of climate system to greenhouse forcing, verification against paleoclimate record.
- Increased number of simulations
  - Climate change is not a deterministic process.
- Increased number of scenarios
  - Future greenhouse gas emissions are uncertain.

Computational power required?

Issue	*Motivation*	*Compute Factor*
Spatial resolution	Provide regional details	10^3-10^5
Model completeness	Add “new” science	10^2
New parameterizations	Upgrade to “better” science	10^2
Run length	Long-term implications	10^2
Ensemble size	Range of model variability	4
Multiple scenarios	Range of possible outcomes	2
Total Compute Factor		10^{10}-10^{12}

And storage?

Component	*Resolution*	*History output (Gbyte per simulated year)*	*Each additional tracer (Gbyte/year)*
Atmosphere	T42 (300 km)	7.5	0.02
Atmosphere	T85 (150 km)	29	0.08
Atmosphere	T170 (75 km)	110	0.3
Ocean	1 deg (100 km)	1.7	0.2
Ocean	1/10 deg (10 km)	120	17

What could we achieve with a factor of 100 increase?

- Modest increase in resolution of atmospheric component (300km to 100km).
- Significant improvements in representation of atmospheric chemistry.
- Addition of interactive carbon cycle and biogeochemistry.
- Inclusion of all types of aerosols (more than sulfate).
- Facilitate testing and evaluation of next-generation sub-grid parameterizations and other next-generation improvements.

How would you measure success? (Examples)

- Improved ability to simulate current climate, especially ENSO, PDO, and other climate cycles.
- Reduced biases between climate simulations and observations.
- Reduced discrepancies among models (e.g., climate sensitivity, regional impacts, etc.).
- More realistic physical processes; thus, less empirical tuning.
- Improved scalability (i.e., less or no need to retune model parameterizations as model resolution is changed).
- More useful information delivered to decision-makers.

Status Report

- MPP sustained simulations and allowed coupled models to become feasible
- Vector machines permit increased resolution
- Latency and bandwidth (memory and interconnect) control efficiencies < 10% efficiency
- Decomposed domain - hierarchical data structures addressed with hybrid MPI-OpenMP programming paradigm
- Codes expose parallelism, cache or vector processing opportunities with tunable chunking
- Better software practices are employing components and frameworks to advantage

Hardware and Software Issues

- Hardware top three
  - Performance
  - Stability
  - Usability
- Software top three
  - Compilers that see and manage the memory latencies
  - Standards based access to interconnect (MPI, MPI-2, CoArray?, MLP?, …)
  - Portable utility, mathematical and scientific components for high performance
- Modeling abstraction that better fits emerging hardware (e.g., IBM BlueGene)