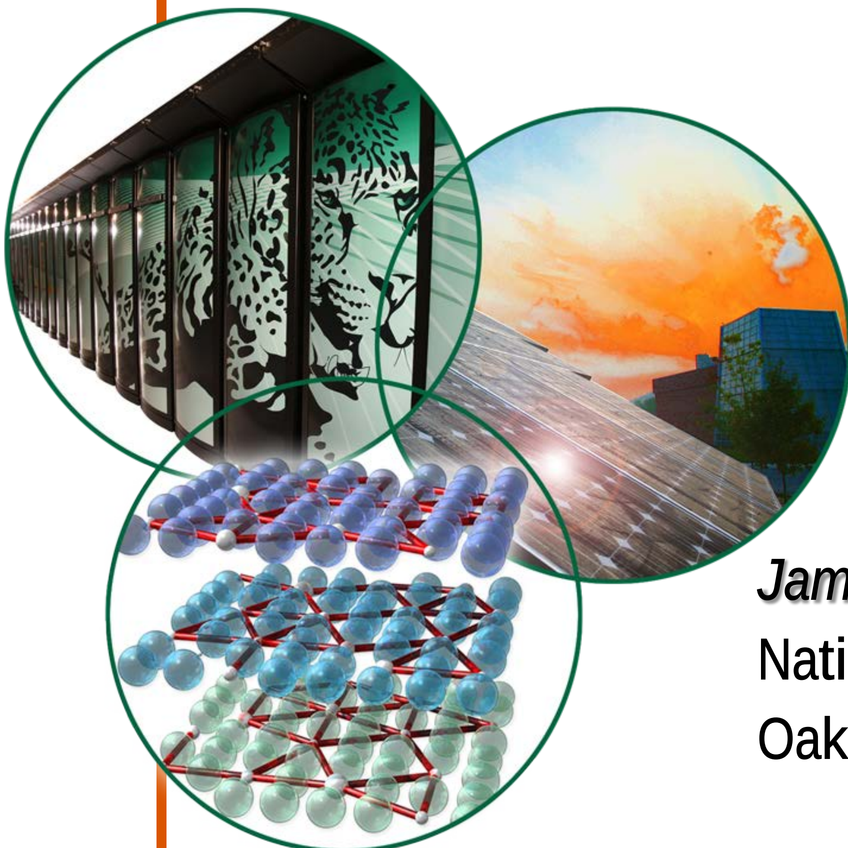
Some Climate and Computing Challenges and Opportunities on the path to Exascale



Biological and Environmental Research Advisory Committee (BERAC) Meeting
March 3-4, 2014

James J. Hack, Director

National Center for Computational Sciences

Oak Ridge National Laboratory

Arthur S. "Buddy" Bland, OLCF Project Director

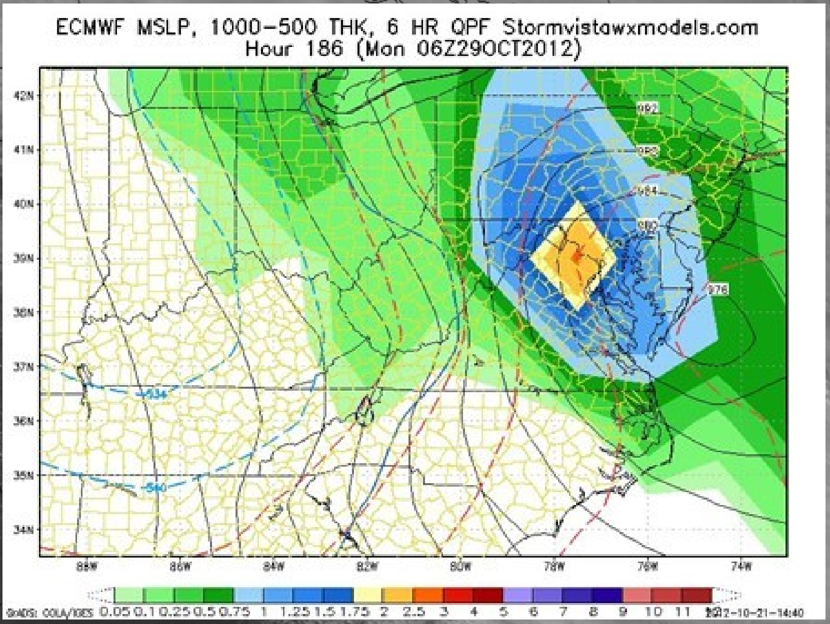
Jack Wells, NCCS Director of Science

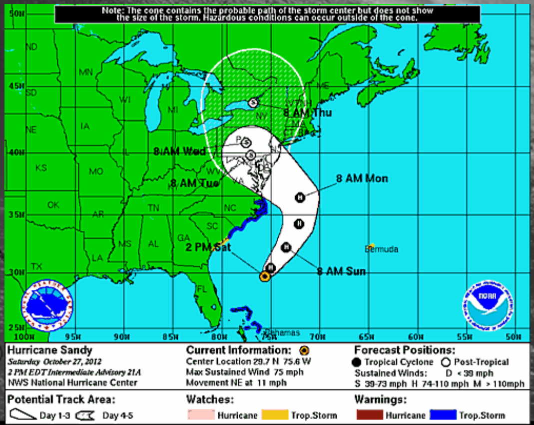
Galen Shipman, Chief Data Architect


New York Magazine

Example of the virtualization challenge

And a great example of success





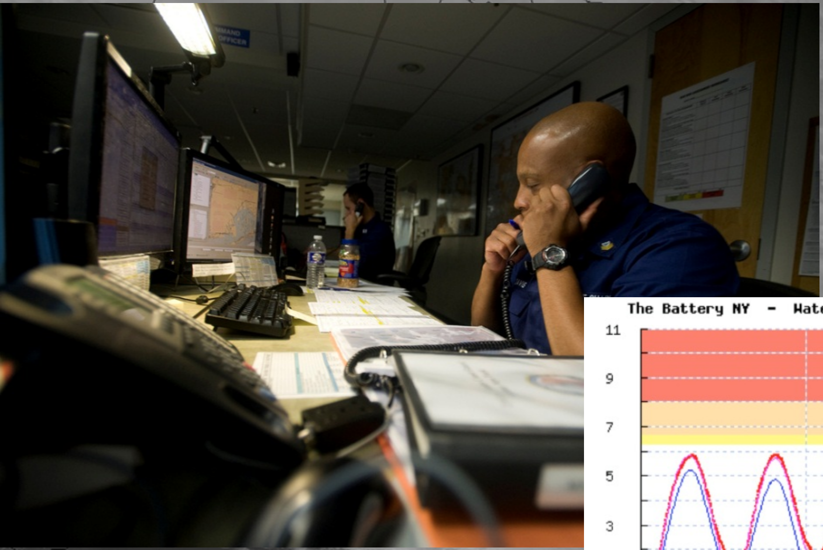
Big storm early next week (Oct 29) with wind and rain??? A number of computer models today were hinting at that, suggesting a tropical system now in the Caribbean may merge with a cold front in that time table, close to the East Coast. At the same time, 8 days is an eternity in storm forecasting, so plenty of time to watch this. Consider this an early "heads-up" that we'll have an interesting an weather feature to monitor this week... details on what to expect and whether we'll even deal with a storm still way up in the air.

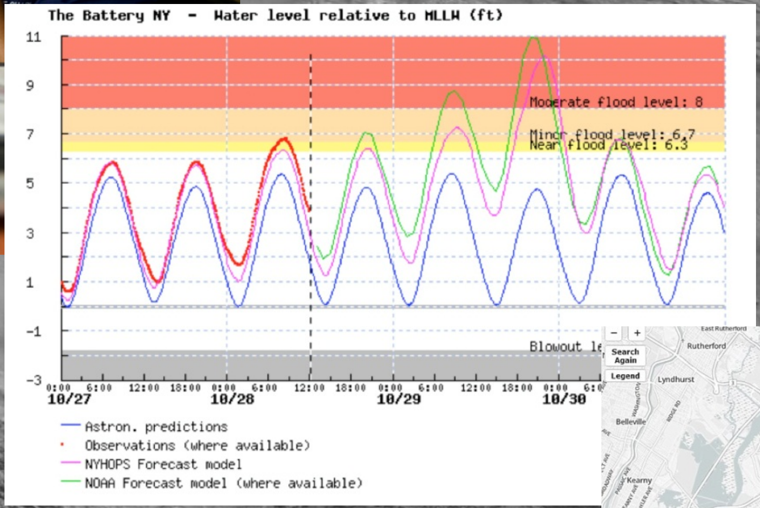
Jason Samenow, Capital Weather Gang, Washington Post 21 October 2012

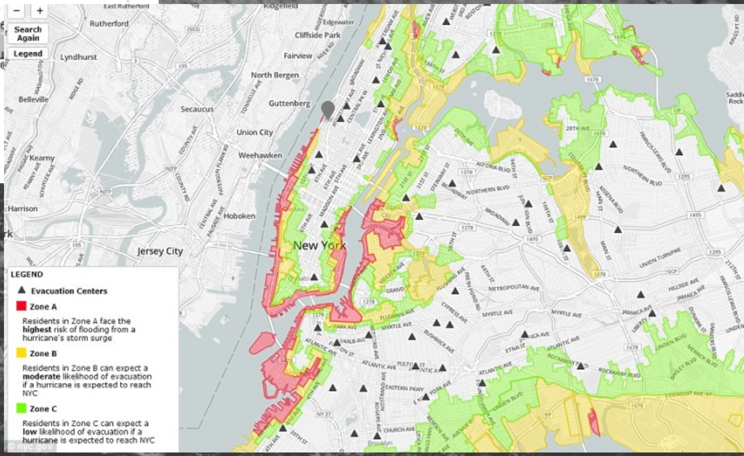
GOES-14 1-KM VISIBLE ON 28 OCT 12 AT 19:41 UTC (SSEC:UW-MADISON)

Example of the virtualization challenge

Simulation science has begun to master deterministic time scales and regional space scales





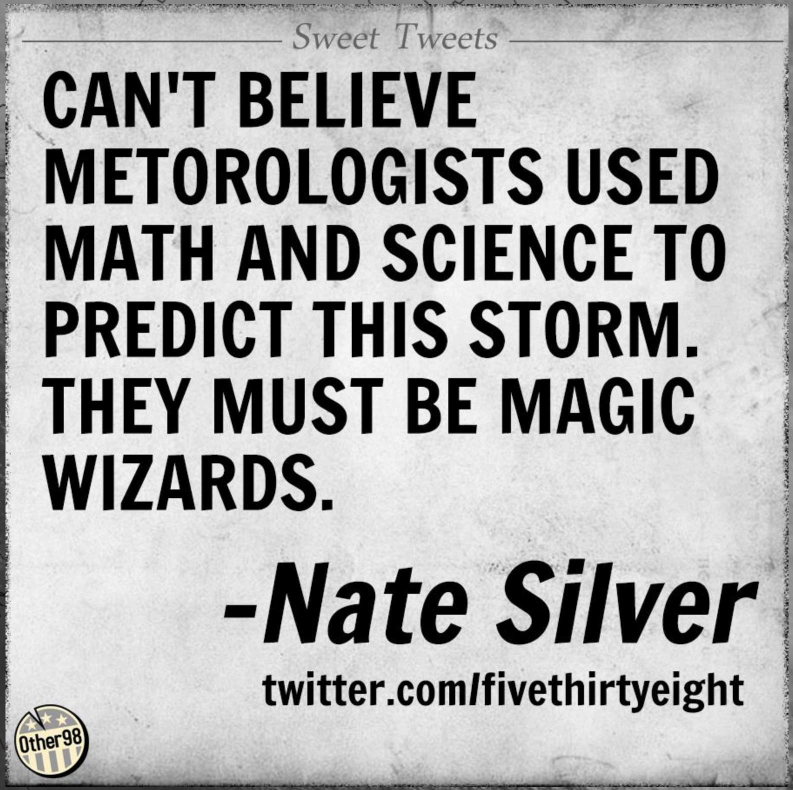


Also a demonstration of the scale challenge

GOES-14 1-KM VISIBLE ON 28 OCT 12 AT 19:41 UTC (SSEC:UW-MADISON)

Example of the virtualization challenge

Surprise!??


Sweet Tweets

CAN'T BELIEVE METOROLOGISTS USED MATH AND SCIENCE TO PREDICT THIS STORM. THEY MUST BE MAGIC WIZARDS.

-Nate Silver

twitter.com/fivethirtyeight

More of a Surprise!??

There has been a series of extreme weather incidents. That is not a political statement, that is a factual statement," Cuomo said. "Anyone who says there is not a dramatic change in weather patterns is denying reality.“

"I said to the president kiddingly the other day, 'We have a one-hundred year flood every two years.'“

Governor Andrew Cuomo, 30 October 2012

GOES-14 1-KM VISIBLE ON 28 OCT 12 AT 19:41 UTC (SSEC:UW-MADISON)

Examples of climate consequences questions

- **Water Resources**
  - management and maintenance of existing water supply systems, development of flood control systems and drought plans
- **Agriculture and food security**
  - Erosion control, dam construction (irrigation), optimizing planting/harvesting times, introduction of tolerant/resistant crops (to drought, insect/pests, etc.)
- **Human health**
  - Public health management reform, improved urban and housing design, improved disease/vector surveillance and monitoring
- **Terrestrial ecosystems**
  - Improvement of management systems (deforestation, reforestation,…), development/improvement of forest fire management plans
- **Coastal zones and marine ecosystems**
  - Better integrated coastal zone planning and management
- **Human-engineered systems**
  - Better planning for long-lived infrastructure investments

Multiscale Turbulent Phase Change of Water Drives *Global* Climate System

Turbulence

Diffusion

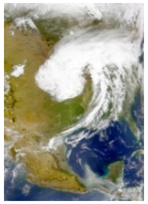
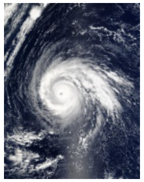
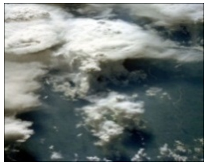
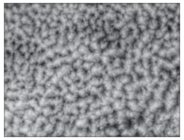
Kinetics

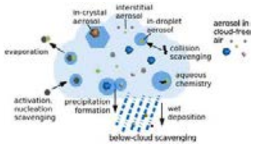
Planetary waves
Hurricanes
boundary layer clouds
chemistry

Cyclones
cloud systems
condensation/collision/coalesence

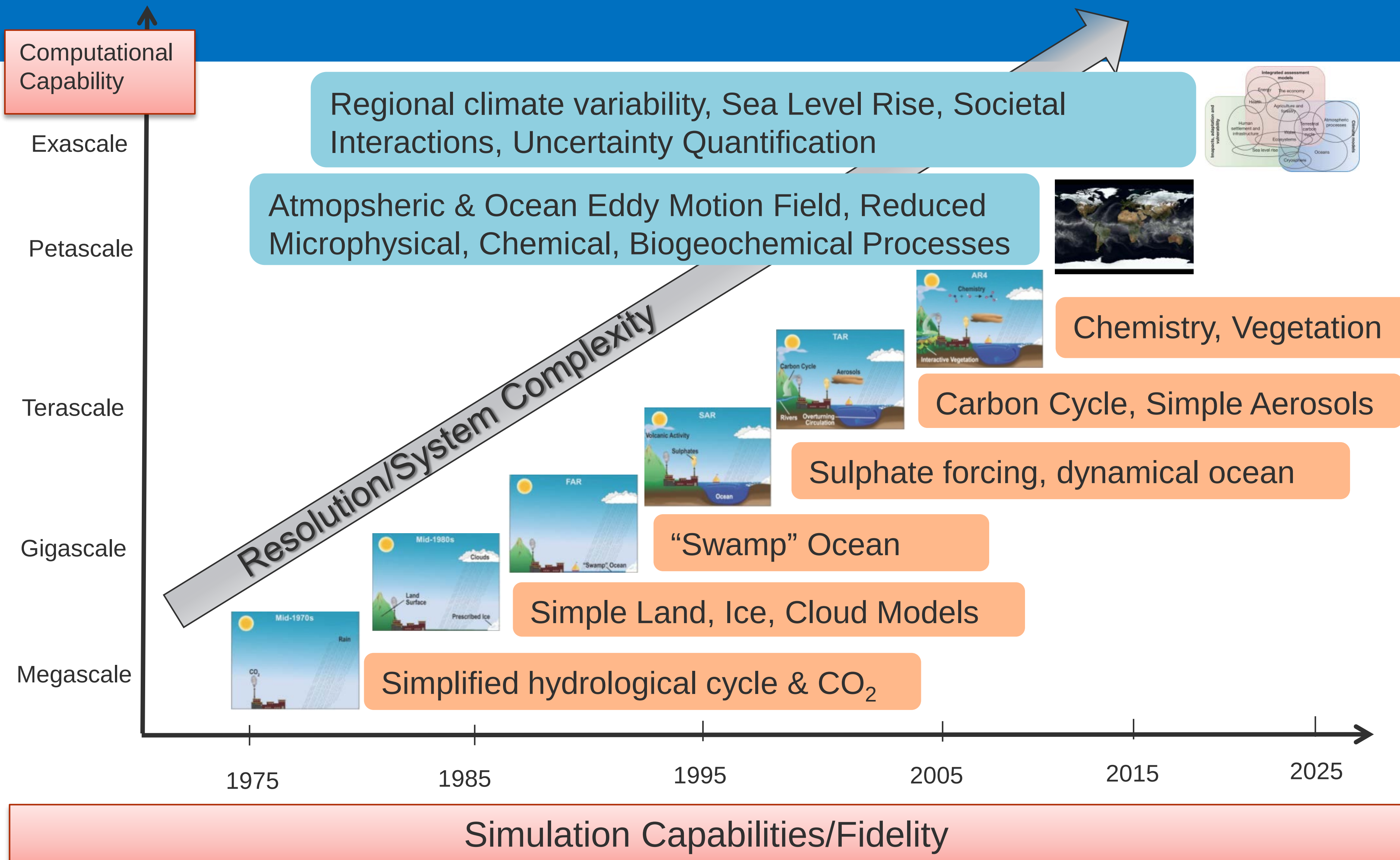
Mm
km
m
mm
µm

Gigascale
Terascale
Petascale
Exascale
Beyond Exascale ⟹

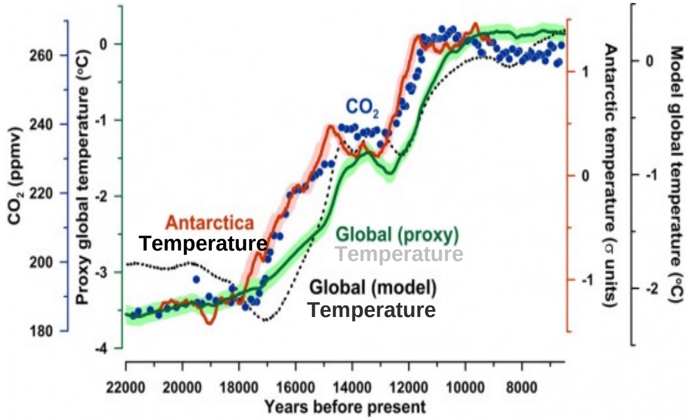


Global Modeling Complexity has Evolved with Improvements in Computational Capabilities



Steady capability-constrained scientific progress

Simulation central to developing an understanding of planetary circulation response to radiative forcing changes



22,000 year simulation explains lag of CO_2 behind Antarctic temperature records over glacial cycles (Shakun et al., Nature)

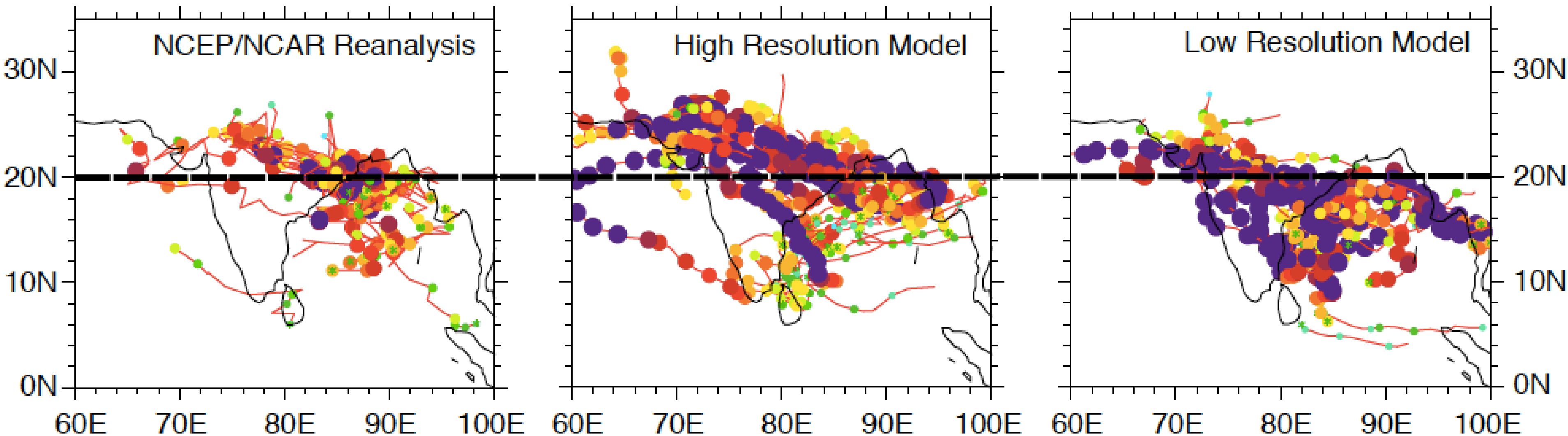
Petascale capabilities critical to enabling treatment of eddy processes required to characterize regional climate

Petascale computing provides the capability to include motion scales in atmosphere and ocean required to address regional climate science questions in more complex frameworks. Snapshots from climate simulations illustrating simulation fidelity now within reach for climate applications

South Asian Monsoon Depressions

Ashfaq and collaborators

South Asian Summer Monsoon Depressions in 1999-2009



Movement of monsoon depressions over land improves in the high-resolution model

High Resolution Spectral Configuration of CAM4

Mahajan and collaborators

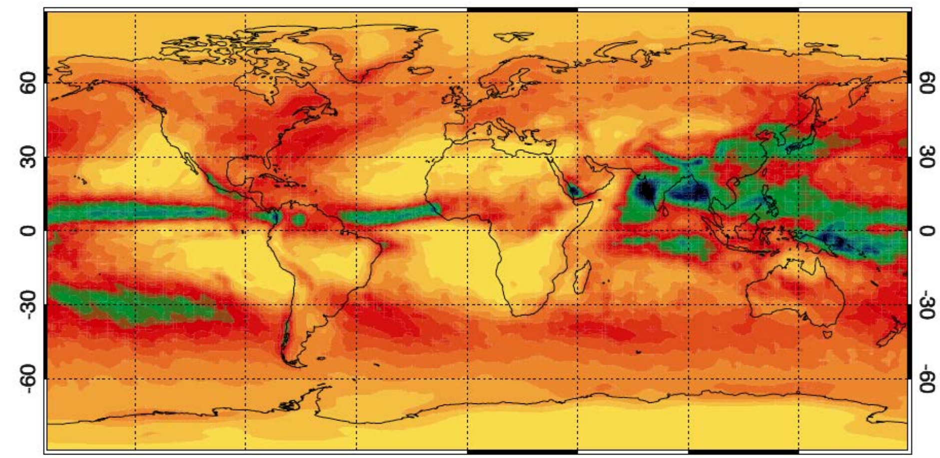
Experiments with T341 Spectral Configuration of CAM4 (ORNL):

- AMIP (1979-2005)
- CAM4 stand-alone pre-industrial control
- Fully Coupled pre-industrial control
- Fully Coupled present day

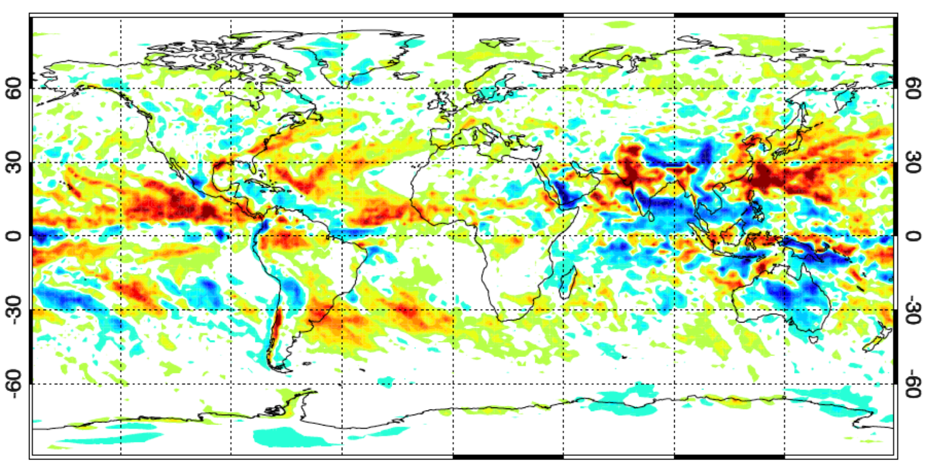
Preliminary Results:

* General increase in variance of precipitation

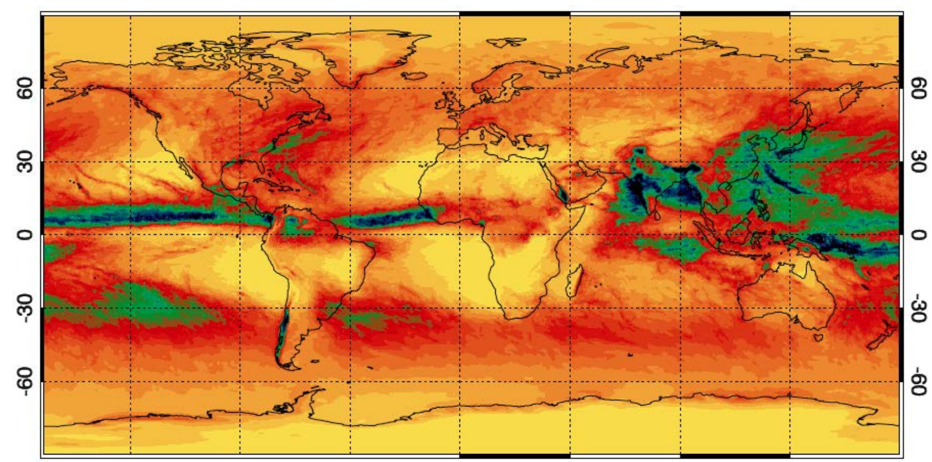
T85 Model (AMIP) Standard Deviation: Precipitation



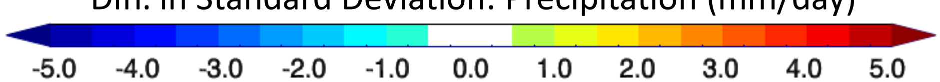
T341 Model – T85 Model



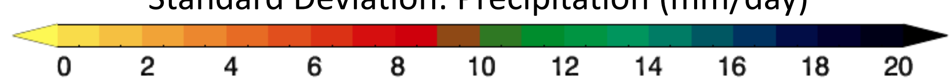
T341 Model (AMIP) Standard Deviation: Precipitation



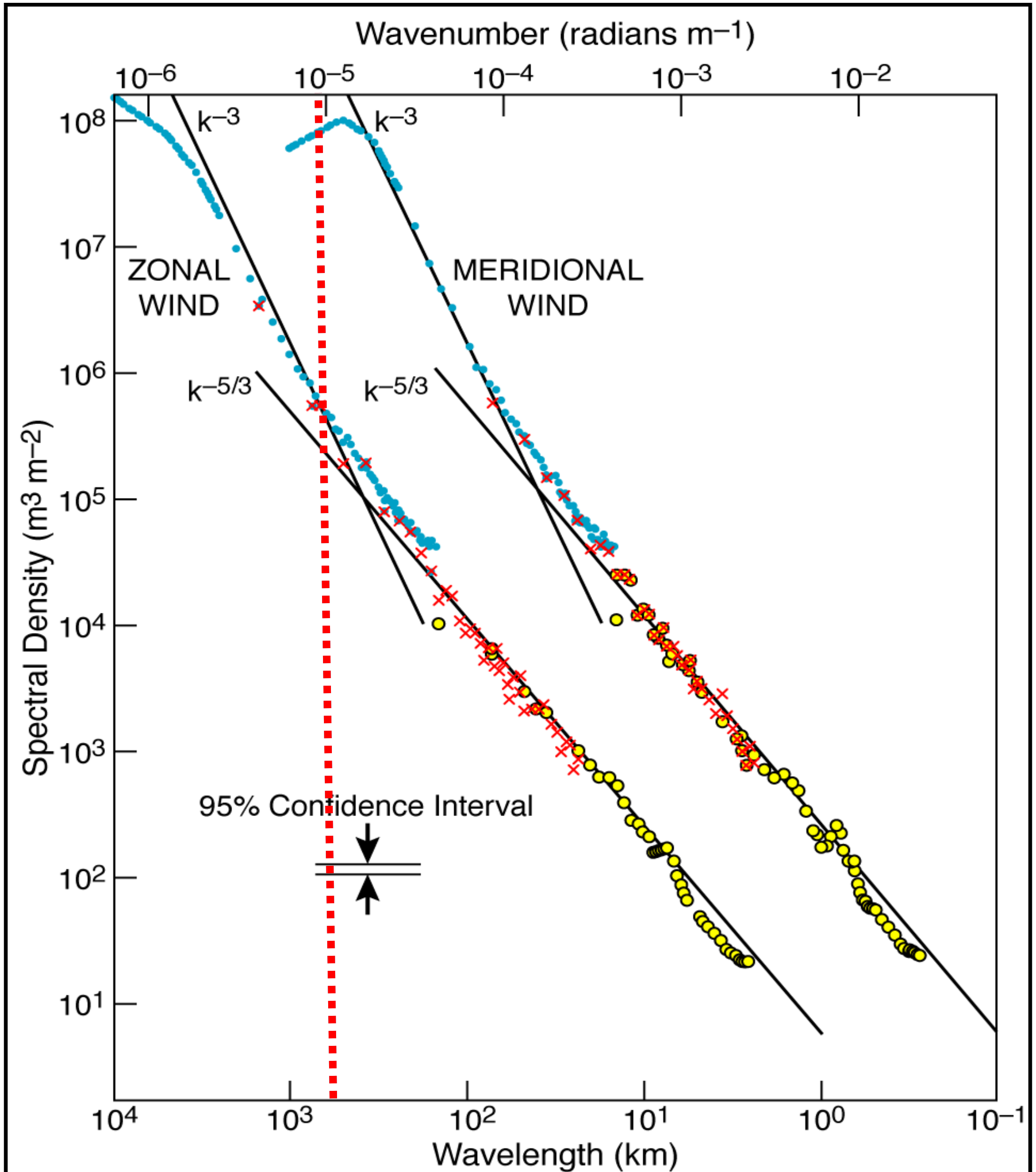
Diff. in Standard Deviation: Precipitation (mm/day)



Standard Deviation: Precipitation (mm/day)



Kinetic energy spectra from aircraft



Unresolved processes and upscale energy transports in models

- Traditionally assume that effects of unresolved scales can be represented through parametrizations based on bulk formulae.
- Inherently assumes that there is no coupling between dynamics and physics on these unresolved scales.
- Essentially ignores upscale energy cascades

 Nastrom and Gage, 1985

Original image from Julia Slingo

Simulating multi-scale multi-physics systems (climate models)

- Parameterized physics behavior varies with resolution
  - changes with horizontal resolution (difficult challenge)
    - time and space truncation properties
  - changes with vertical resolution (extremely difficult challenge)

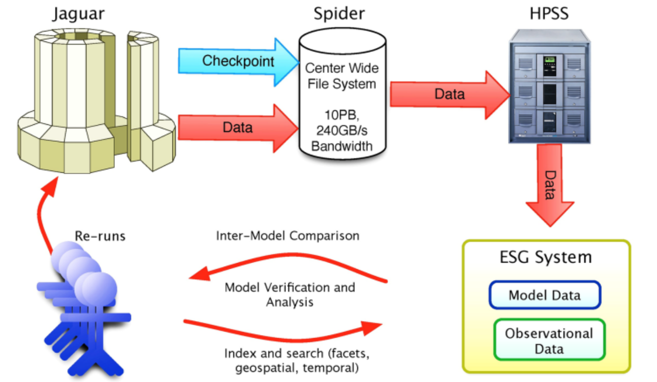
- Tuning of parameterized processes reaching formulation limits
  - main constraint is mean statistical behavior
  - other constraints relate to behavior of individual processes
    - unfortunately observational constraints are limited

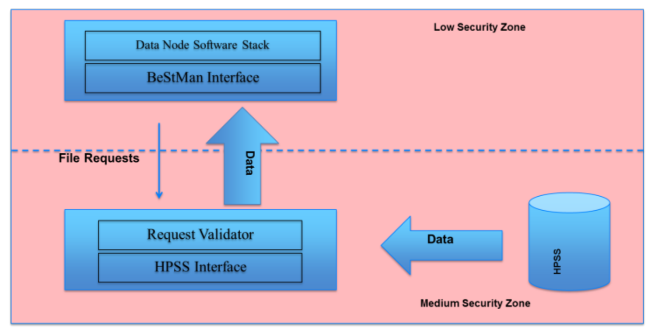
Need a new approach to move these difficult problems forward

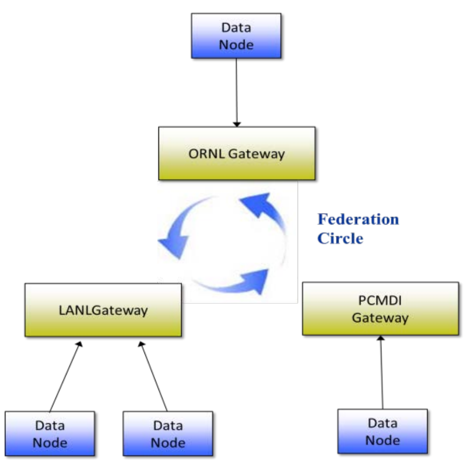
Oak Ridge National Laboratory

High Resolution Simulation Workflow 2009

- Simulation and sampling plan
  - Need to archive and export >300 TB to partners
- Development of automated workflow for high resolution production simulations
  - Manual system not scalable
  - Initial configuration for Jaguar, but can be hardened and exported



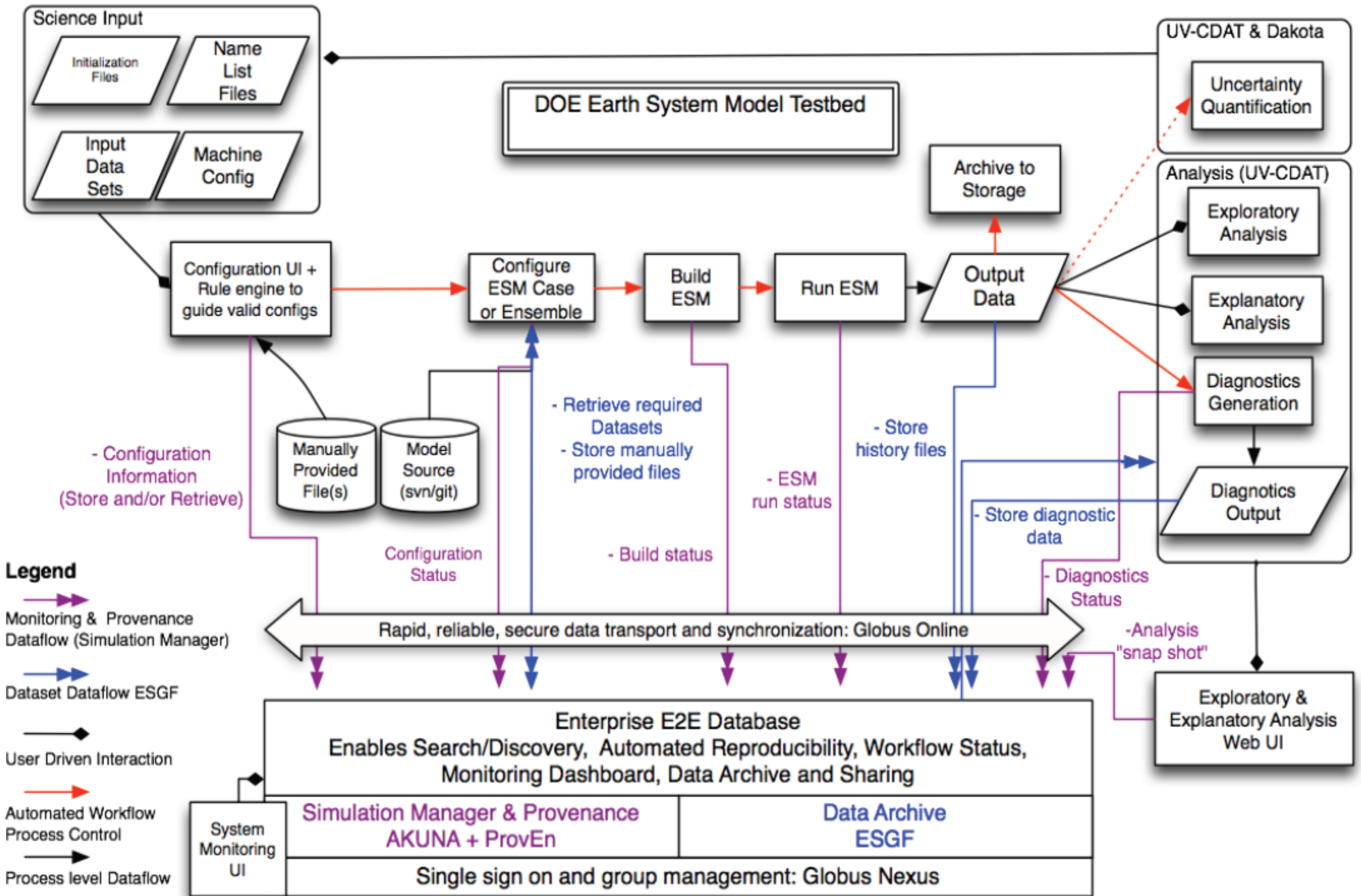




- High Volume Climate Data Server
  - http://cds.ccs.ornl.gov
  - Integrated with Earth System Grid (ESG)

Shipman and collaborators leveraging ESG effort

High Resolution Simulation Workflow 2014



- Workflows that enable uncertainty quantification as part of the development process
- Comprehensive multi-variate optimization, with formal parametric uncertainty estimates and characterization of error propagation

Model-Observation Integration

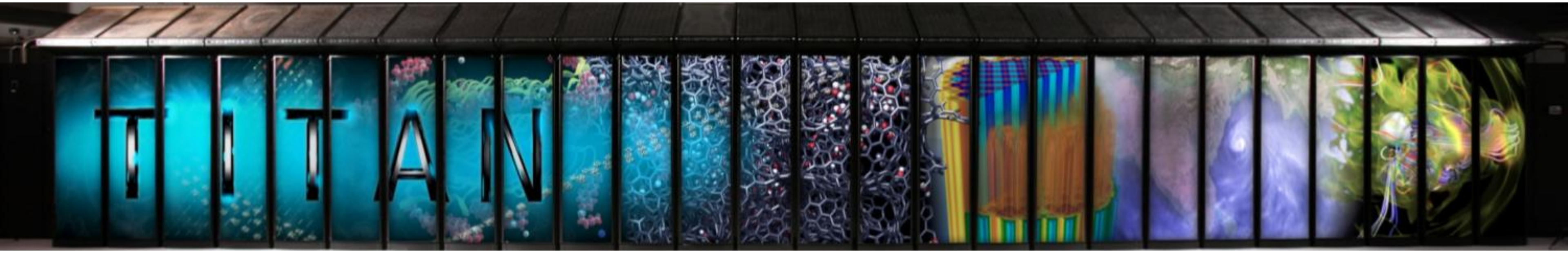
- Identify the scientific problems that would benefit from daily (or regular) Large Eddy Simulation (LES), single column modeling (SCM) and perhaps cloud resolving modeling (CRM)
- Explore ways to maximize the benefits of regular LES/SCM/CRM, confronted with observations
- Drive measurement and modeling strategies and needs to advance specific science problems

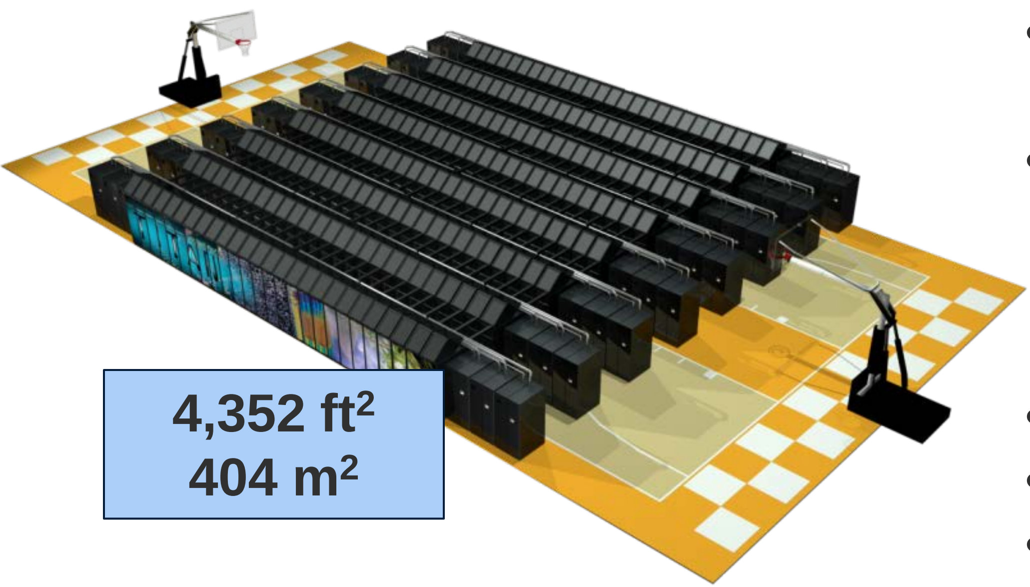
Why?

- Subgrid variability for the thermodynamic variables needs to be taken into account in any GCM for parameterizations of convection, clouds and radiation in a consistent way.
- Large Eddy Simulations (LES) in combination with observations is a useful tool to obtain this subgrid variability and to help develop GCM parameterizations for these cloud related processes.

The Next Order of Magnitude in Performance

ORNL's "Titan" Hybrid System:
Cray XK7 (AMD x86 Opteron & NVIDIA GPUs)





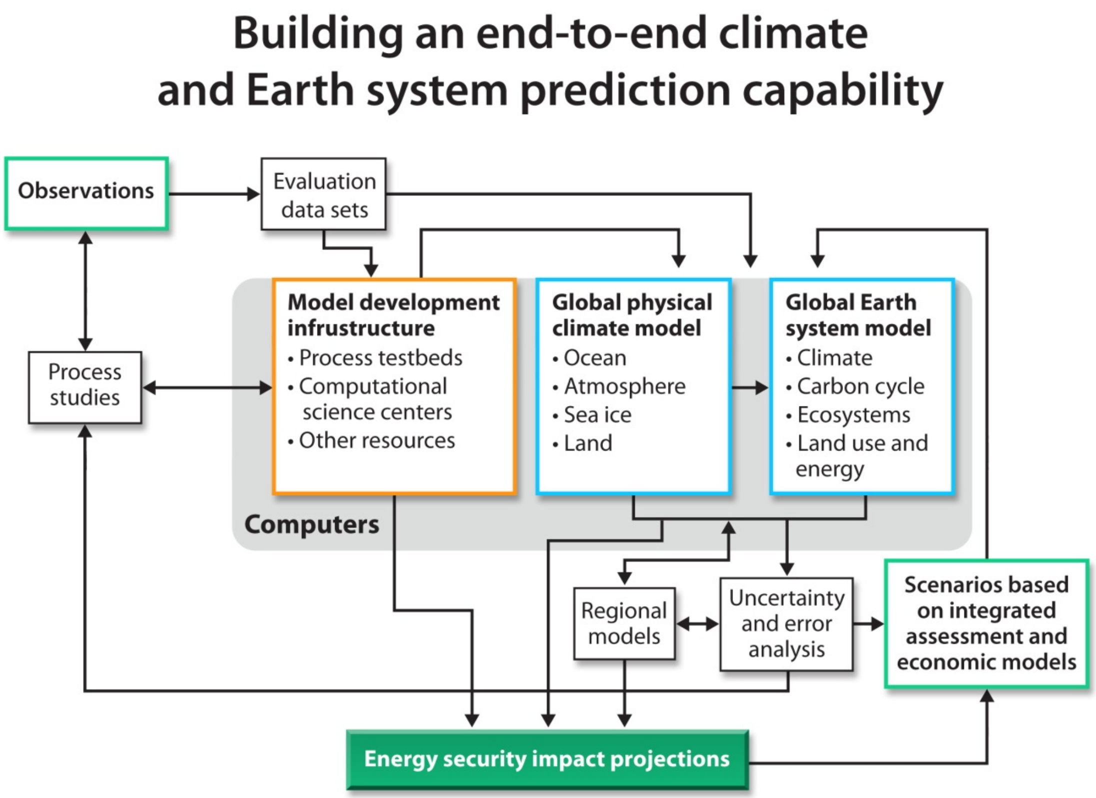
4,352 ft^2
404 m^2

SYSTEM SPECIFICATIONS:

- Peak performance of 27.1 PF
  - 24.5 GPU + 2.6 CPU
- 18,688 Compute Nodes each with:
  - 16-Core AMD Opteron CPU
  - NVIDIA Tesla "K20x" GPU
  - 32 + 6 GB memory
- 512 Service and I/O nodes
- 200 Cabinets
- 710 TB total system memory
- Cray Gemini 3D Torus Interconnect
- 8.9 MW peak power

The Accelerated Climate Modeling for Energy (ACME) Project

- Earth system model that will run efficiently on DOE computing systems by 2017 (5 simulated years per day)
- Designed to meets Science, Energy Mission and Computing Technology
- Model released with ensemble of simulations for the period from 1970-2050 running at ~20 km horizontal grid resolution in atmosphere and ocean
- 7 Labs and 6 Academic Partners
- $21M per year - $19 M in funding at labs that will be redirected from current foundational projects



Data Science and Scientific Discovery

- The rate of scientific progress is increasingly dependent on the ability to efficiently capture, integrate, analyze, and steward large volumes of diverse data
- Increasing data volume, variety, and velocity are creating a new environment for scientific discovery
- But many facilities and research programs across the Office of Science are not prepared for this challenge

National Dialog on Scientific Data

- America COMPETES Reauthorization Act of 2010 - P.L. 111-358
  - Interagency public access committee would coordinate Federal science agency research and policies related to the dissemination and long-term stewardship of the results of federally supported unclassified research
- Office of Science and Technology Policy – Big Data Across the Federal Government March 29, 2012
  - "Today, the Obama Administration is announcing the Big Data Research and Development Initiative. By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning."
- Office of Science and Technology Policy – Memo on data access February 22, 2013
  - Increasing Access to the Results of Federally Funded Scientific Research
  - The Office of Science and Technology Policy (OSTP) hereby directs each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government. This includes any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds, as defined in relevant OMB circulars (e.g., A-21 and A-11).
  - Each agency shall submit a draft plan by August 2013.
- DOE Office of Science
  - Requests all facilities to document their data policy
  - Requires all new proposals to develop a data management plan

Data Science Center – Motivation

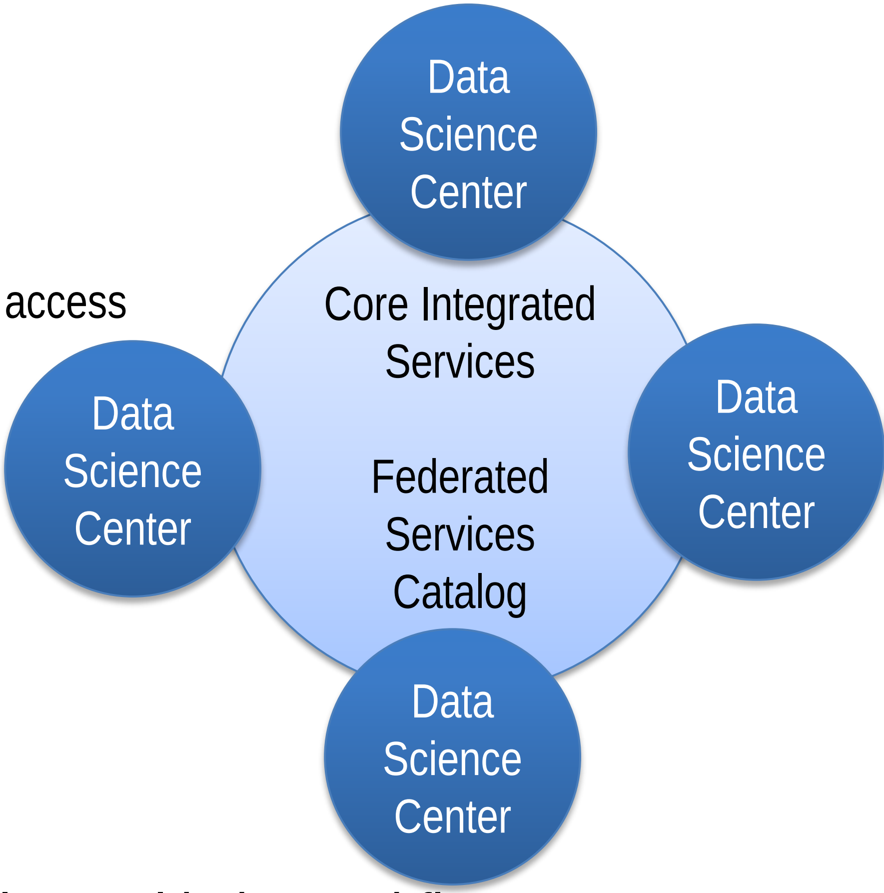
- DOE SC facilities and research teams have increasing storage and analysis needs that outstrip their abilities to handle independently.
- Research teams are increasingly geographically distributed and require collaborative tools and shared access to data and analysis infrastructure.
- Consolidated resources provides economies of scale and opportunities for integrating data sets across projects and experiments and access to capabilities that few if any projects could build in isolation.
- Research teams and facility users require access to and assistance from data science expertise to effectively conduct their research.
- Increased value and productivity from existing facilities because of easy and long-term access to data
- Increasing recognition of the value and importance for improved data management and data curation that extend beyond the life-time of the experiments and projects.
- Leveraging significant investments in infrastructure and expertise to deliver these capabilities

A Data Science Center for SC

- The ORNL Data Science Center will provide these capabilities
  - Rich data analysis environment - tightly coupling compute and data storage
  - Coupling simulation, experiment, and observation data
    - Model improvement, validation, steering, site selection, etc.
  - A flexible compute and data environment that can be curtailed to specific needs
  - Cataloging and long-term stewardship of scientific datasets
  - Long-term (many year) allocations for major facilities and projects
- The ORNL Data Science Center will leverage significant investments
  - 65,000 square feet of datacenter space
  - 40 Megawatts of power, 6K tons of cooling (after latest upgrade)
  - Connectivity to every major research network
  - Significant expertise across ORNL in delivering solutions to data challenges

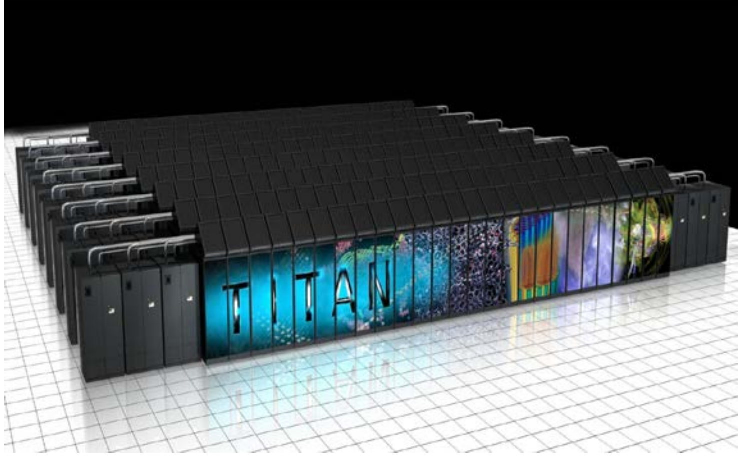
Data Science Centers - Services

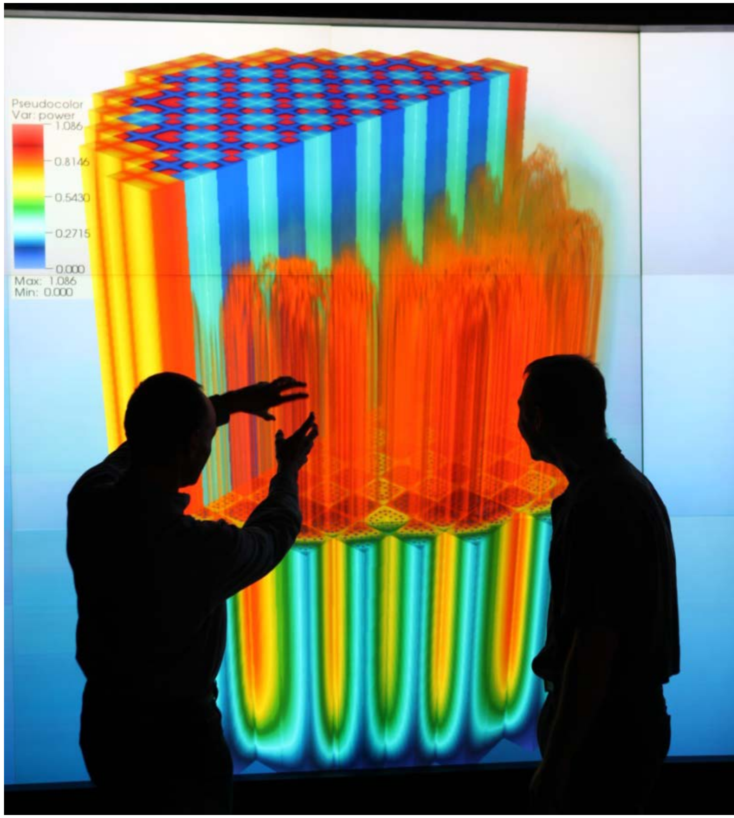
- Complimentary data science services from across lab complex
- Staff with expertise in all areas of data science
  - Partnering with Domain Scientists
- Core integrated services
  - Universal account
  - Data replication and inter-site data & metadata access
  - Data publication and digital curation services
  - Inter-site workflows
  - Other critical replicated services
- Federated services catalog
  - Core common services
  - Site-specific services
  - Common service provisioning API
- Providing the ability to construct complex multi-site workflows



OAK RIDGE National Laboratory

Oak Ridge Leadership Computing Facility Mission





The OLCF is a DOE Office of Science National User Facility whose mission is to enable breakthrough science by:

- Fielding the most powerful capability computers for scientific research,
- Building the required infrastructure to facilitate user access to these computers,
- Selecting a few time-sensitive problems of national importance that can take advantage of these systems,
- And partnering with these teams to deliver breakthrough science.

Extreme-scale » Exascale Systems

- 1-10 billion way parallelism
  - Requires hierarchical parallelism to manage
    - MPI between nodes
    - OpenMP or other threads model on nodes
    - SIMD / Vector parallelism within thread
- Power will dominate architectures
    - Takes more power to go to memory for data than to recompute it
- Traditional “balance ratios” are erroding
    - Memory size is not growing as fast as processor performance
    - Memory bandwidth is growing even more slowly
    - Floating point operations cheap; memory access and data movement rate limitors

Some Lessons Learned

- Exposure of unrealized parallelism
  - Identifying the opportunities is often straightforward
  - Making changes to exploit it is hard work (made easier by better tools)
  - Developers can quickly learn, e.g., CUDA and put it to effective use
  - A directives-based approach offers a straightforward path to portable performance
- For those codes that already make effective use of scientific libraries, the possibility of continued use is important.
  - HW-aware choices
  - Help (or, at least, no hindrance) to overlapping computation with device communication
- Ensuring that changes are communicated back and remain in the production “trunk” is every bit as important as initially thought
  - Other development work taking place on all CAAR codes could quickly make acceleration changes obsolete/broken otherwise

All Codes Will Need Rework To Scale!

- Up to 2-4 person-years required to port each code from Jaguar to Titan
  - Refactoring effort will be required for exascale performance regardless of the specific architecture
  - Also pays off for other systems—the ported codes often run significantly faster CPU-only (Denovo 2X, CAM-SE >1.7X)
- Experience demonstrates 70-80% of developer time is spent in code restructuring, regardless of whether using OpenMP / CUDA / OpenCL / OpenACC / …
- Each code team must make its own choice of using OpenMP vs. CUDA vs. OpenCL vs. OpenACC, based on the specific case—may be different conclusion for each code
- The user community and sponsors must plan for this expense

All Codes Need Error Recovery at Scale

- Simple checkpoint / restart is a minimum
  - At the scale of Titan, we are seeing several nodes fail per day
  - Jobs running on the full system for several hours should expect to have a node fail during job execution and be prepared to recover
- More advanced error detection and recovery techniques will be required as parallelism increases
  - FT-MPI, algorithms that can ignore faults, and other research techniques for error containment and recovery mandatory for larger systems

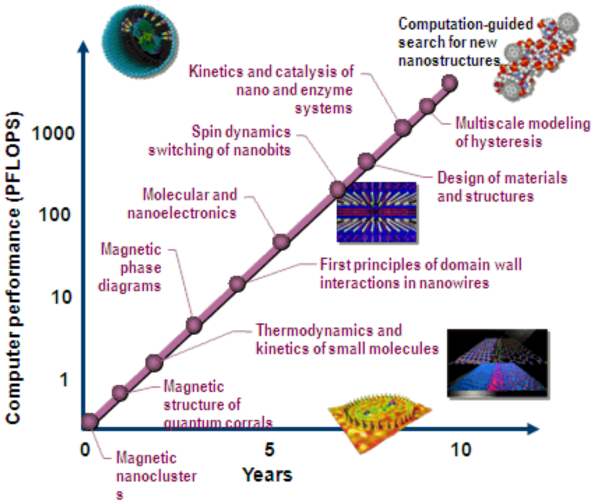
Need for a richer programming environment

- Tools are critical to success
  - **complex hierarchical parallelism and heterogeneous processors ≠ the days** of debugging with print statements
- Ongoing investments in good tools essential
  - debuggers, performance analysis, memory analysis, and the training

Rethink fundamental algorithmic approach

- Heterogeneous architectures can make previously intractable or inefficient models and implementations viable
  - Alternative methods for electrostatics that perform slower on traditional x86 can be significantly faster on GPUs *(Nguyen, et al. J. Chem. Theor. Comput. 2013. 73-83)*
  - 3-body coarse-grain simulations of water with greater concurrency can allow > 100X simulation rates when compared to fastest atomistic models even though both are run on the GPUs *(Brown, et al. Submitted)*

DOE-SC Science Drivers



Expected outcomes

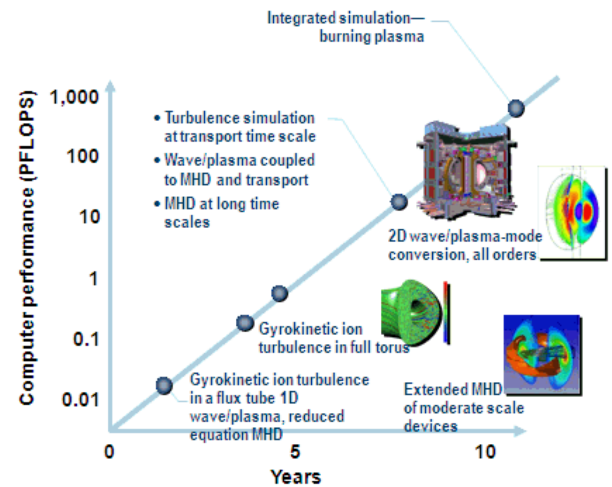
5 years

- Realistic simulation of self-assembly and single-molecule electron transport
- Finite-temperature properties of nanoparticles/quantum corrals

10 years

- Multiscale modeling of molecular electronic devices
- Computation-guided search for new materials and nanostructures

Materials Science



Expected outcomes

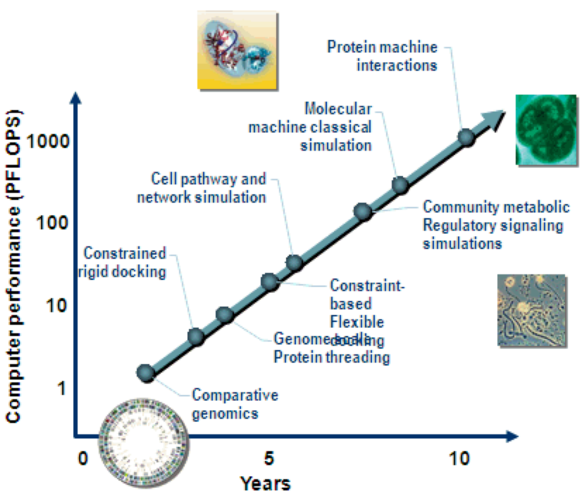
5 Years

- Full-torus, electromagnetic simulation of turbulent transport with kinetic electrons for simulation times approaching transport time-scale
- Develop understanding of internal reconnection events in extended MHD, with assessment of RF heating and current drive techniques for mitigation

10 years

- Develop quantitative, predictive understanding of disruption events in large tokamaks
- Begin integrated simulation of burning plasma devices –

Fusion



Expected outcomes

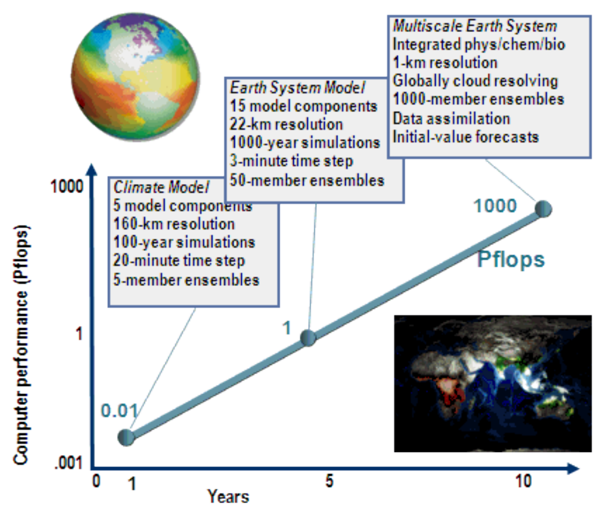
5 years

- 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

Biology



Expected outcomes

5 years

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

10 years

- Cloud-resolving 1-km spatial resolution atmosphere
- Fully coupled, physics, chemistry, biology earth system model

Climate

What is CORAL (Partnership for 2017 System)

- CORAL is a **C**ollaboration of **O**ak **R**idge, **A**rgonne, and **L**awrence Livermore Labs to acquire three systems for delivery in 2017.
- DOE's Office of Science (DOE/SC) and National Nuclear Security Administration (NNSA) signed an MOU agreeing to collaborate on HPC research and acquisitions
- Collaboration grouping of DOE labs was done based on common acquisition timings. Collaboration is a win-win for all parties.
  - It reduces the number of RFPs vendors have to respond to
  - It improves the number and quality of proposals
  - It allows pooling of R&D funds
  - It strengthens the alliance between SC/NNSA on road to exascale
  - It encourages sharing technical expertise between Labs

CORAL COLLABORATION
OAK RIDGE ARGONNE LIVERMORE

CORAL Joint NNSA & SC Leadership Computing Acquisition Project

Current DOE Leadership Computers

Titan (ORNL)
2012 - 2017

Sequoia (LLNL)
2012 - 2017

Mira (ANL)
2012 - 2017

Objective - Procure 3 leadership computers to be sited at ANL, ORNL and LLNL in CY17

Leadership Computers run the most demanding DOE mission applications and advance HPC technologies to assure continued US/DOE leadership

Approach

Competitive process - one RFP (issued by LLNL) leading to 2 R&D contracts and 3 computer procurement contracts

For risk reduction and to meet a broad set of requirements, 2 architectural paths will be selected

Once Selected, Multi-year Lab-Awardee relationship to co-design computers

Both R&D contracts jointly managed by the 3 Labs

Each lab manages and negotiates its own computer procurement contract, and may exercise options to meet their specific needs

Understanding that long procurement lead-time may impact architectural characteristics and designs of procured computers

CORAL Management Structure

Top three levels of management and their responsibilities

CORAL Executives	
Rick Stevens, Associate Laboratory Director, Argonne **Jeff Nichols**, Associate Laboratory Director, ORNL **Mike McCoy**, ASC Program Director, LLNL	• Laboratory leadership oversight • Final decision authority

CORAL Directors	
Michael E. Papka, LCF Director, Argonne **Jim Hack**, LCF Director, ORNL **Terri Quinn**, LC Program Lead, LLNL	• Concur on system selection and ensures selected system meets DOE and facility (LC, ALCF, OLCF) mission needs • Coordination at facility level

CORAL Acquisition Leads	
Susan Coghlan, ALCF-3 Project Director, Argonne **Buddy Bland**, OLCF-4 Project Director, ORNL **Bronis de Supinski**, LC CTO, LLNL	• Responsible for the acquisition • Work with technical and procurement teams to gather requirements, develop RFP, prepare for and respond to IPRs, issue RFP, and recommend system selection

CORAL Procurement Model

Two Diverse Architecture Paths

RFP

R&D contract

SC Lab #1 computer contract (2017 delivery)

R&D contract

SC Lab #2 computer contract (2017 delivery)

LLNL computer contract (2017 delivery)

Summary

- Simulation science is at multiple inflection points
  - new, growing, and demanding expectations from simulation tools
  - multiscale/multiphysics simulation challenges
  - new directions in HPC computer architectures/environments
- Strategy to survive and thrive?
  - continued investment in basic scientific understanding
    - investment in methodologies for abstracting multiscale behavior
    - development of methods for quantifying sources of uncertainty
  - creation of multi-disciplinary project-like teams with clear science goals
- Partnerships will be crucial for sustained progress
  - responsive data infrastructure and services
  - methods and disruptive architectures
  - responsive development environments

Questions?

http://www.nccs.gov
http://www.nics.tennessee.edu/