Report for the BER Advisory Committee

Two BER computational workshops:

1. ASCR-BER Exascale Requirements Workshop

March 29-30, 2016; Rockville Hilton

2. Advancing Cross-Cutting Ideas for Computational Climate Science (AXICCS)

September 12-13, 2016; Rockville Hilton

Ramana Madupu
Biological Systems Sciences Division

Dorothy Koch
Climate and Environmental Sciences Division

April 20, 2017

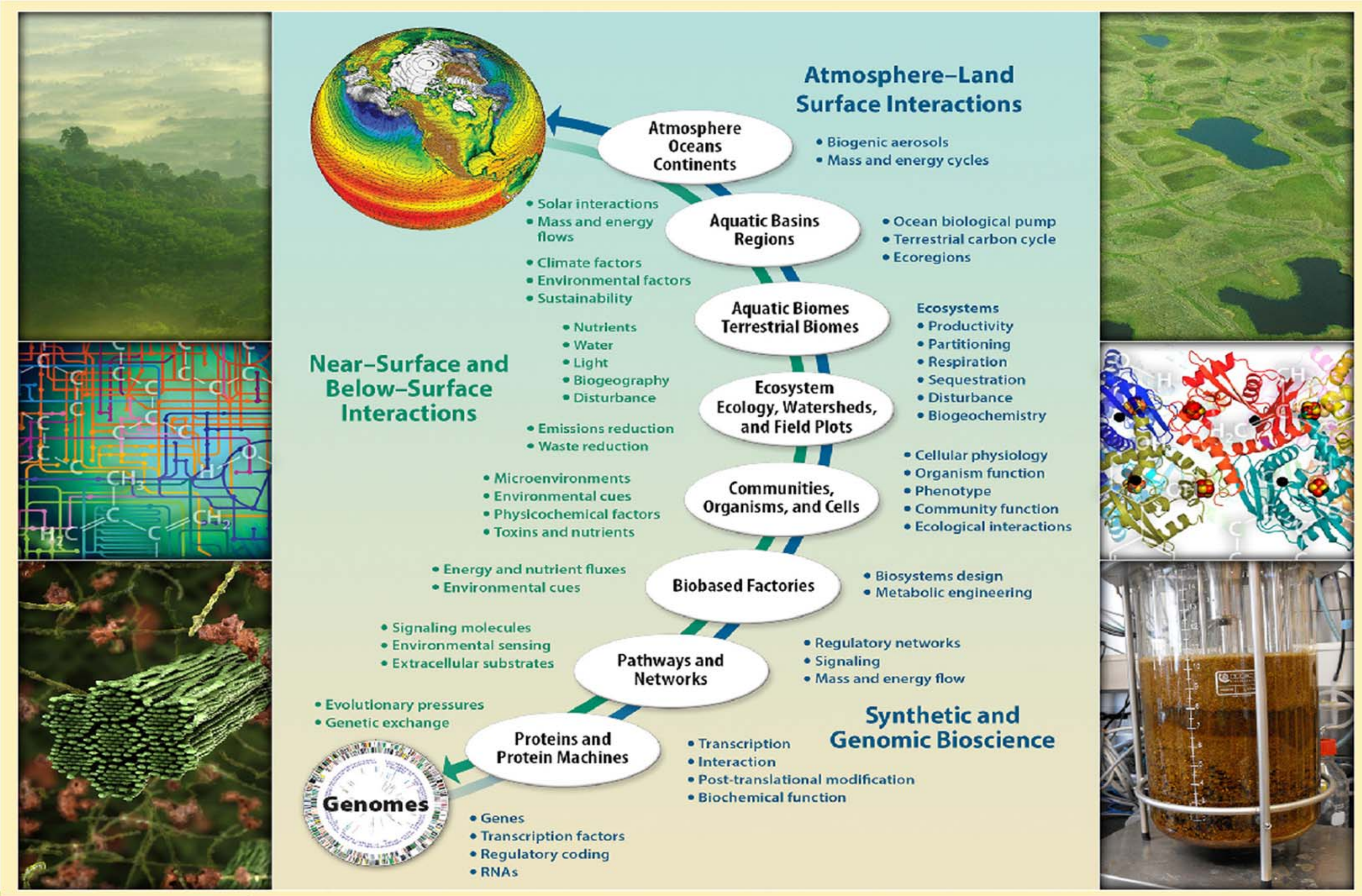
Overview

BER-Exascale Requirements

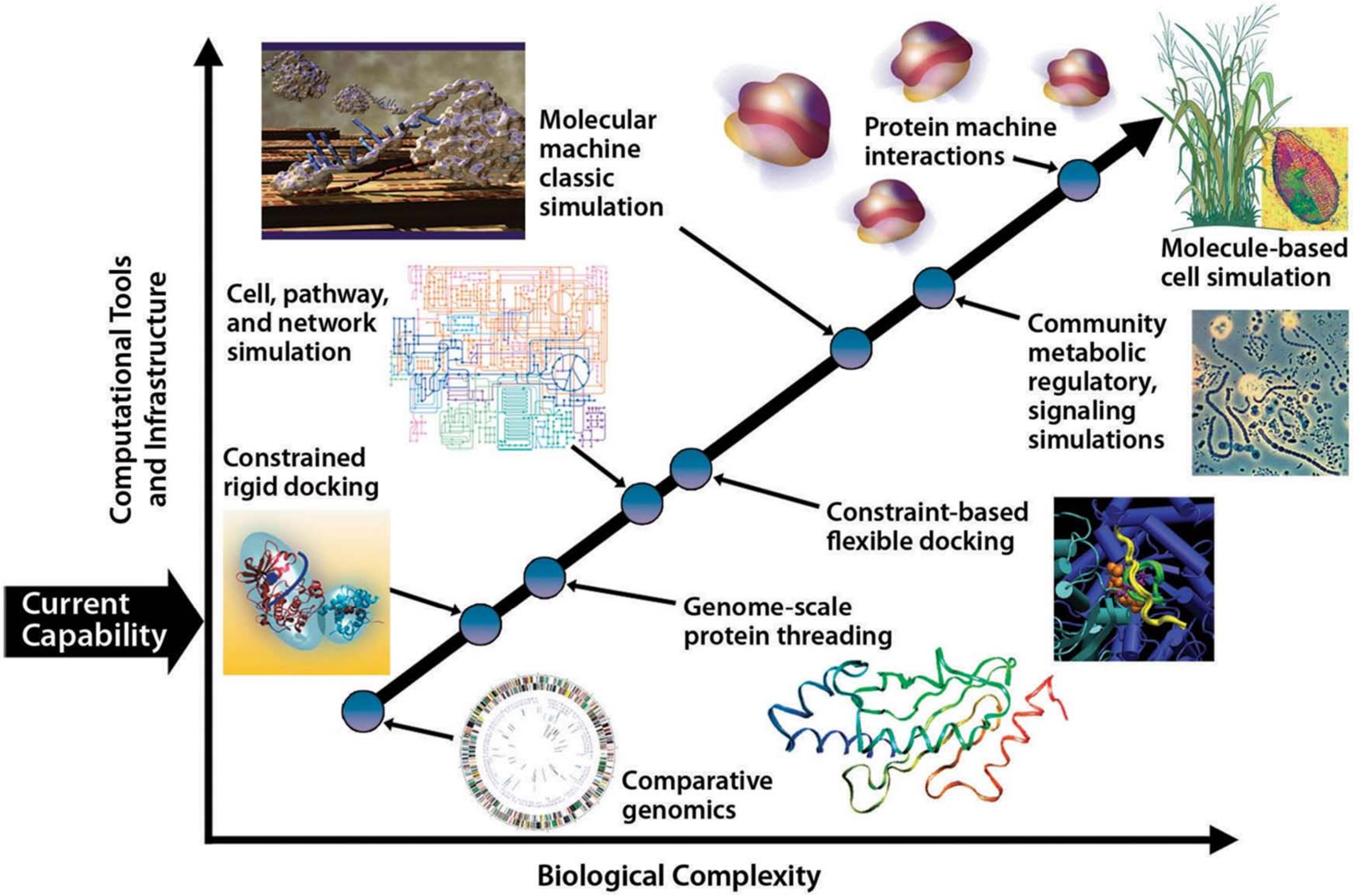
- **Purpose:** One of 7 Office of Science workshops led by ASCR-Facilities to determine decadal science requirements for hardware and software upgrades
- **Organization:** ASCR Facilities, BER & ASCR PMs (Koch, Madupu, Lauzon), BER-chairs (Bader and Arkin)
- **Sponsorship:** ASCR
- **Attendees:** Invited by BER, primarily BER-supported scientists; 50 attendees
- **Outcome:** BER Workshop report & Cross-cut report (pending)

BSSD's Big Data challenge

".. Building a foundation for **<u>a continuum of understanding</u>** across a range of spatiotemporal scales—from the molecular to the global scale and from picoseconds to millennia"



Biological Complexity: Genomes to Cell Models (nm-um)



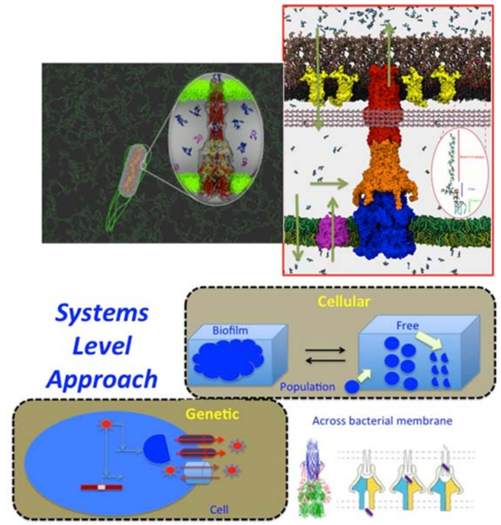
ASCR-BER Requirements Workshop (BSSD)

Biological Systems Science organized into Four Topics

1. Multiscale Biophysical Simulation from Molecules to Cells
2. Mapping Sequence to Models
3. Microbes to the Environment
4. Biological Big Data Challenges

Multiscale Biophysical Simulation from Molecules to Cells

- Multiscale Models crossing time and length scales with real time data analysis - Link genes to processes that take place in time and space.
- Molecular Dynamic models from single macromolecule to cellular complexes to whole cells



- Integration of large-scale biophysical codes- quantum mechanical (QM) methods through molecular dynamics (MD) to stochastic reactive transport codes to facilitate the multiscale predictions required at all levels of biological systems sciences

Mapping Sequence to Models

Priority Research Directions

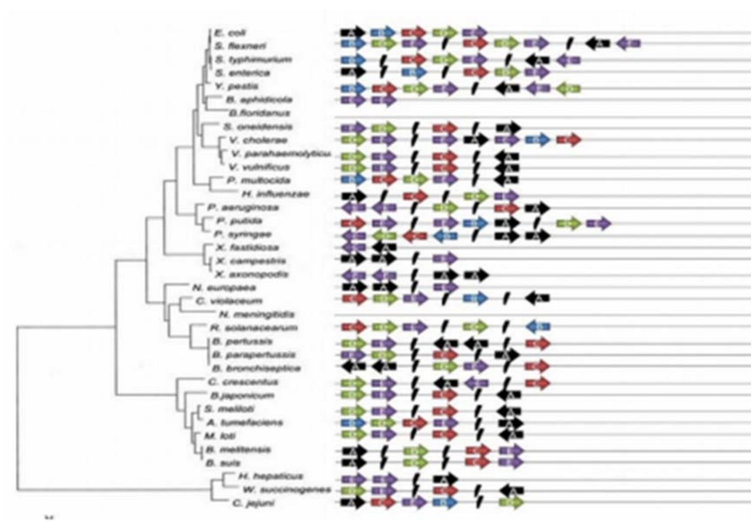
- Genomics - genome and metagenome assembly, and structural annotation of genes and operons, homology searches and evolutionary analysis

- Protein structure and function prediction, protein design tasks and modeling of larger complexes

- Network reconstruction and modeling, cell-level simulation, modeling, and prediction, interpretation of genetic variation in wild populations

Mapping Sequence to Models

Gene block evolution



- New innovative Technologies in **Sequencing, Molecular Imaging, Molecular Functional Assays** driving Computational Innovation
- Exponentially Scaling Data rates and Data sizes
- Computationally Intensive Data Processing Algorithms require large memory and disk space
- Multimodal data sets with spatial and temporal dimensions
- Data Mining (from new and extant datasets) on a massive scale with many different methods
- Critical need for Algorithms to move to Exascale machines
- Need for developing and maintaining key functional genomics resources that provide critical linkages to map genotype to phenotype processes

Microbes to Environment

Priority Research Directions

- Multiscale Simulation of Complex Biogeochemical Earth Systems
- Elucidating the supramolecular structure of organic matter with inorganic ionic species and interactions with mineral surfaces
- Understanding how the aggregation of organic matter influences its stability, i.e., protection from decomposition by biotic and abiotic processes

Scientific challenges frame the need for a multiscale predictive simulation capability that integrates BSSD and CESD science

Biological Big Data

Scientific Challenges

- Automated cognitive integration of biological ‘Big Data’

  Use of functional genomics is presently limited due to large number of genes have unknown function; integrative OMICS analysis
- Linking genes to complex phenotype

  Data from HTP experiments and associations between multiple SNPs and metabolites and phenomics data

Computational Challenges

- Data Storage
- Machine Learning Algorithms
- Deep Learning Algorithms

Programmatic Elements and Cross-Cutting Research Directions

1) Methods Development

- Conversion of primary measurement to physical quantities for comparison to models of function/dynamics, with integrated uncertainty quantification and analysis
- Integration of these diverse, heterogeneous measurements on natural and simulated systems to infer models
- **2) Computational Environment and Resources**
- A software stack for analysis, machine learning, and visualization of data
- Interactive access tested engineering frameworks for code development and execution
- Support for specific community codes for implementation and optimizations on new architectures

Programmatic Elements and Cross-Cutting Research Directions *contd*

3) Data

- Mirrored data facilities
- Data sharing and archiving capabilities
- New exploration, analysis and visualization tools
- Complex analysis workflows for in situ analysis

4) Communication and Community Involvement

- Computationally focused workforce development of the user community

Overview: 2 CESD computational climate workshops

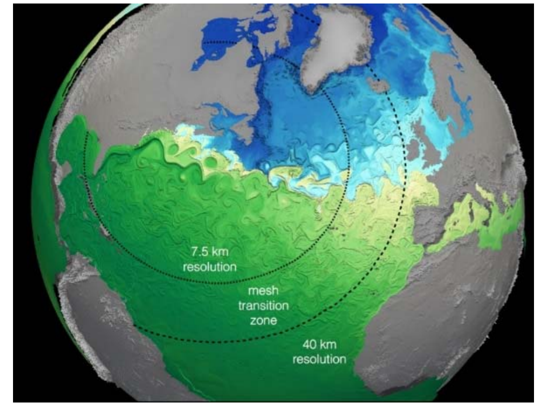
BER-Exascale Requirements (BSSD and CESD)

- **Purpose:** One of 7 Office of Science workshops led by ASCR-Facilities to determine decadal science requirements for hardware and software upgrades
- **Organization:** ASCR Facilities, BER & ASCR PMs (Koch, Madupu, Lauzon), BER-chairs (Bader and Arkin)
- **Sponsorship:** ASCR
- **Attendees:** Invited by BER, primarily BER-supported scientists; 50 attendees
- **Outcome:** BER Workshop report & Cross-cut report (pending)

AXICCS (Advancing Cross-Cutting Ideas for Computational Climate Science)

- **Purpose:** Elicit creative and exploratory ideas on climate modeling algorithms, software and methods for next-generation computers
- **Organization:** Organized by Evans (ORNL-BER) and Ng (LBNL-ASCR)
- **Sponsorship:** DOE-BER and ASCR
- **Attendees:** Open to broad, multidisciplinary, international community; 65 scientists
- **Outcome:** Workshop report on BER website

Report outlines (CESD)



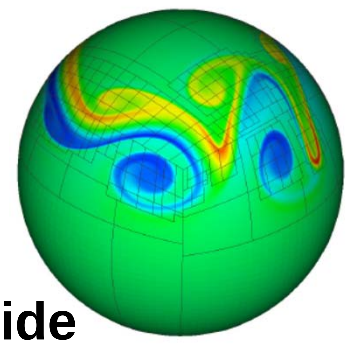
BER-Exascale Requirements

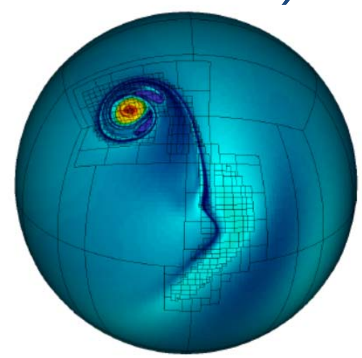
- Atmosphere
- Terrestrial/subsurface
- Ocean/cryosphere
- Earth system
- IA and IAV
- Model-data fusion
- Exascale capabilities: algorithms and computation

AXICCS (Advancing Cross-Cutting Ideas for Computational Climate Science)

“Big Ideas” – Earth system side

- Process-resolving capabilities
- Integrated model credibility
- Representing climate system complexity
- Continuum model framework





“Cross-cutting issues” – Computation side

- Model/software complexity
- Hardware
- Performance
- Data management

Exascale-Requirements - focus on DOE science priorities & challenges
AXICCS - community brainstorming challenges & solutions

BER Exascale Requirements: Terrestrial and Atmospheric

Atmosphere

- Multi-scale modeling for atmospheric and cloud processes, from DNS to LES to Global, to improve process understanding and parameterization
- Improved representation of cloud and aerosol physics and land-atmosphere interactions in models
- Improved and multi-scale data-model fusion/assimilation for validation and for initialization

Terrestrial and subsurface

- Multi-scale modeling of processes, from cellular to watershed scales, to improve process understanding and parameterization appropriate to various scales, including ever higher resolutions for global models (1 km grids)
- Improved representation of hydrologic and biological processes for climate and environmental models
- Model-observation integration across scales for testing and research

BER Exascale Req.: Ocean, cryosphere, coupled models

Ocean & Cryosphere

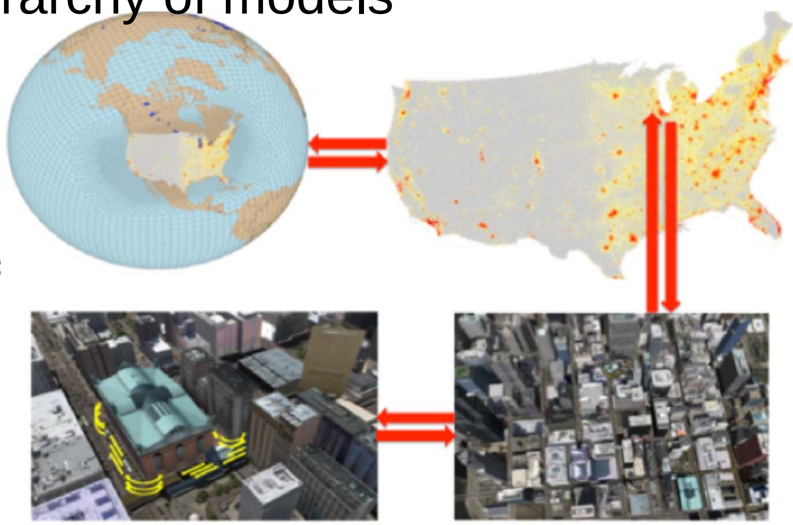
- Ice-ocean margin interactions, ice-sheet changes, to model sea-level
- Sea-level and storm-surge impacts and interactions with coastal regions
- Sea-ice modeling evolution ice-free Arctic

Earth system

- Simulation of extreme events – storms, droughts
- Narrow estimation of climate sensitivity (cloud feedbacks, land and ocean uptake of heat/carbon)
- Methods to connect projection uncertainty to Earth system elements
- Methods for model ensembles and of capturing statistics
- Theory and approach to interconnect hierarchy of models

Integrated Assessment, Impacts-Adaptation-Vulnerability

- Need for flexible interoperable model-data frameworks adaptable to a variety of climate-human connected impacts and adaptation questions, depending on sector and scale



BER Exascale Requirements: Model-data, computer science

Model-data

- Improved automated model validation and analysis workflows
- Development of in-situ diagnostics and “simulators” for high-frequency phenomena and more appropriate comparison with observations
- Improved assimilation for better initialization and identification of biases
- Development of regional refinement to very fine non-hydrostatic scales and comparison with e.g. ARM measurements

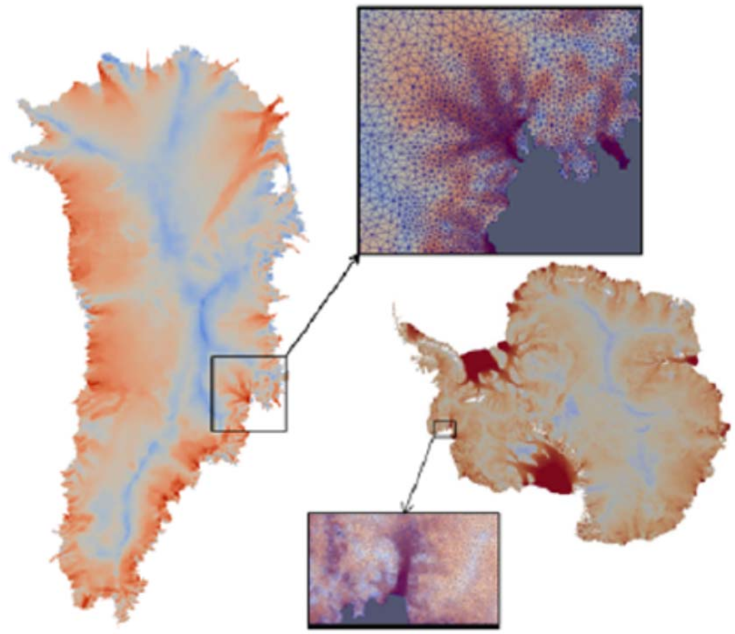
Algorithm and Computer Science

- Development and application of programming models that enable portability, ease of use and simulation speed
- Flexible algorithms for multiple machine-architectures, purposes, scales, and methods to address the “high-bandwidth” communication and complex systems couplings challenges inherent in Earth system research

AXICCS: “Big ideas” for Earth system science

Process resolving modeling capability: Creation of more specialized, multiscale model elements, with minimal assumptions and parameters

- Examples: cloud-resolving model, dynamic cryosphere, human-environment couplings
- Facilitated by: adaptive mesh, better scalability



Climate system complexity: subgrid process representations

- Improved algorithms and parameterizations, perform off-line or parallelized calculations, automated tuning

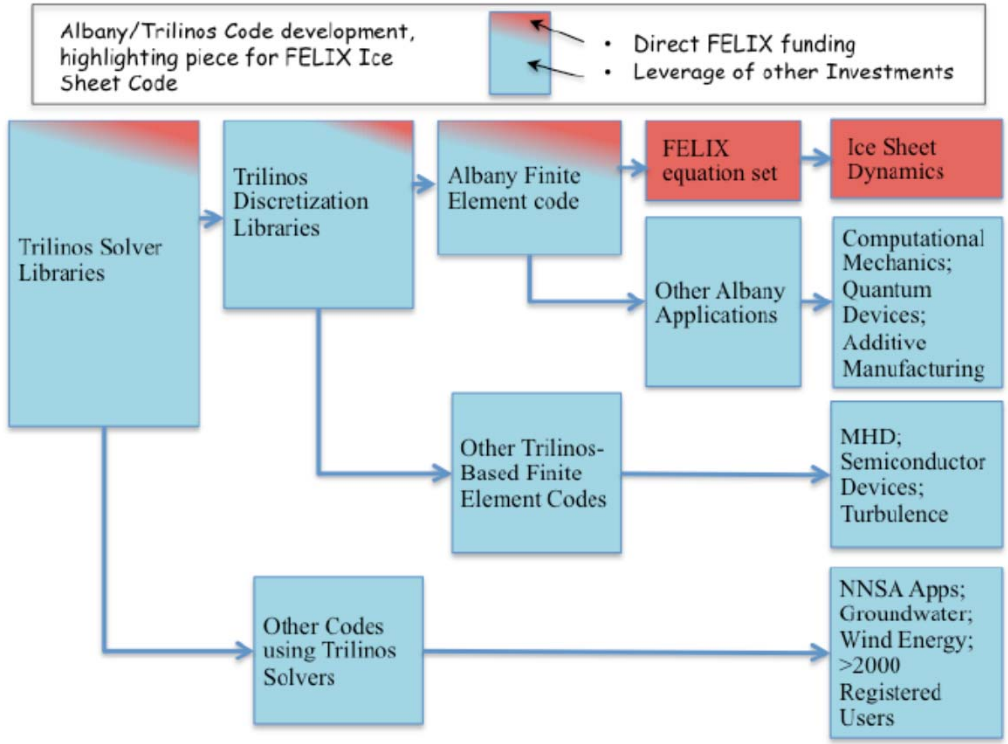
Integrated model credibility:

- Validation, verification and uncertainty quantification: identify model biases, trace to processes and effects on system projections
- Facilitated by: embedded error, multi-fidelity methods, emergent constraints

Continuum model frameworks: integration of model components into a single modeling framework

- Need to identify best coupling strategy (one-way vs full; tight vs loose)
- Develop functional or reduced-order “components” and processes as appropriate

AXICCS: Cross-cutting computational challenges



Software complexity: Need for more verification and unit testing, use of optimized general software (e.g. libraries), computer-science solutions for complex models on complex architectures

New Hardware: With current trajectory of hardware, climate model resolution probably cannot increase more than 2-4x without increase in power/cost.

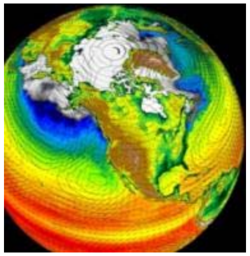
Solutions: Need for better methods and algorithms; need for new semiconductor technology or other hardware designs that are suited for climate-like codes.

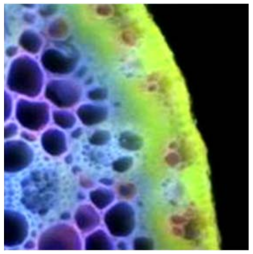
Model performance and time to solution:
Challenges: Reduce inter-node communication, preconditioner methods
Solutions: Better solver methods, (spatial and temporal) adaptive meshes and methods to focus computation where needed, improved parallelization, better ensemble methods

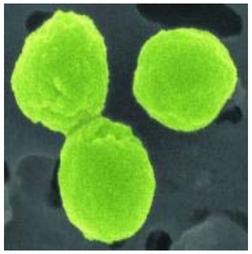
Data management and analysis: Excessive model output - need to explore parameter space more efficiently. Solutions: Reduce data with in-situ analysis, or functional data analysis - continuum process representation

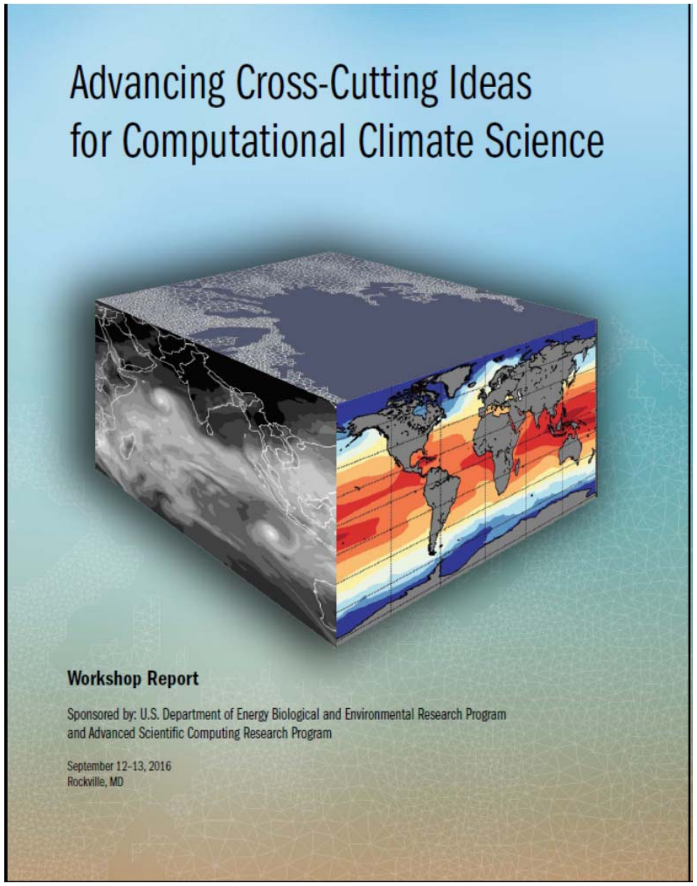
Programmatic elements (both workshops)

- **Hardware:** BER will need ecosystem of computing systems, including high performance computing as well as machines designed for large-data and I/O efficiency
- **Skill-development:** Need better training and reward structure for scientists with interdisciplinary computational and climate modeling skills
  - DOE Labs have unique environment for this, e.g. SC-Graduate Student Research Opportunity
  - Retention of software and cross-disciplinary researchers: recognition for technical contributions such as software releases or new capability developments, include co-authorship on science papers, special project-recognition
- **Programs** (e.g. SciDAC) that integrate a committed multidisciplinary team is required to address exascale challenges: domain (climate), performance, algorithm, software, facilities-hardware insights









CESD: Dorothy.Koch@science.doe.gov
BSSD: Ramana.Madupu@science.doe.gov

BER requirements website: http://exascaleage.org/ber/
AXICCS report at https://science.energy.gov/ber/community-resources/
AXICCS website: http://www.csm.ornl.gov/workshops/AXICCS/index.html