Berkeley Lab's Contributions to the DOE Exascale Computing Project

Joint Tsukuba – LBNL Meeting 5 March 2018

David L Brown, Director
Computational Research Division
Lawrence Berkeley National Laboratory
Berkeley, California, USA

Goals of US DOE 's Exascale Computing Project

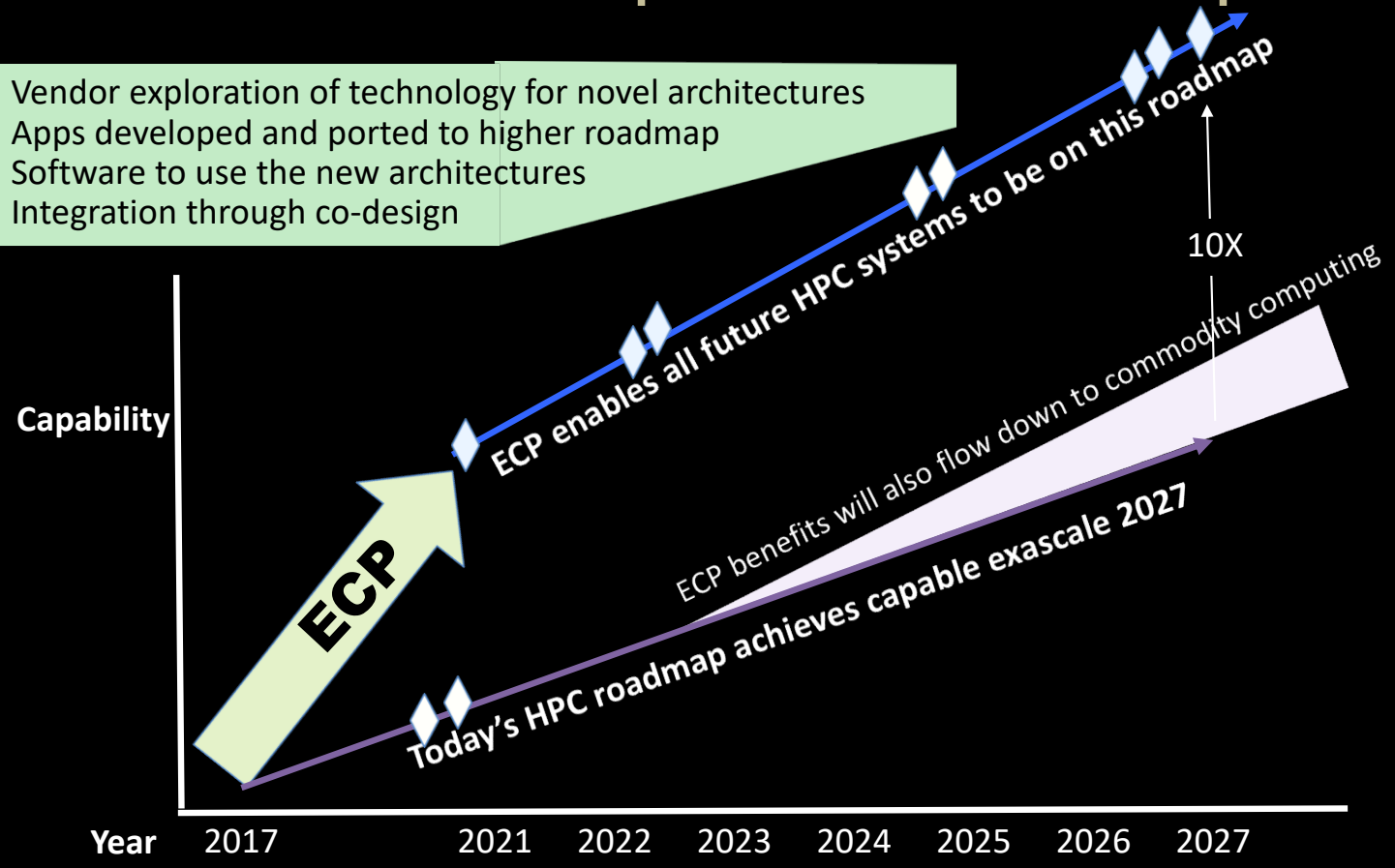
"The Exascale Computing Project (ECP) was established to accelerate delivery of a capable exascale computing system" to deliver approximately 50 times more performance than today's 20-petaflops machines on mission critical applications in under 30 Megawatts.

ECP is a 7-year project, which runs through 2023 to enable the deployment of at least two diverse, capable exascale systems.

Berkeley Lab's Exascale Goal: Maximize the Science Impact of Exascale

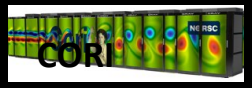
The US DOE ECP aims to impact Broad HPC landscape

- Vendor exploration of technology for novel architectures
- Apps developed and ported to higher roadmap
- Software to use the new architectures
- Integration through co-design

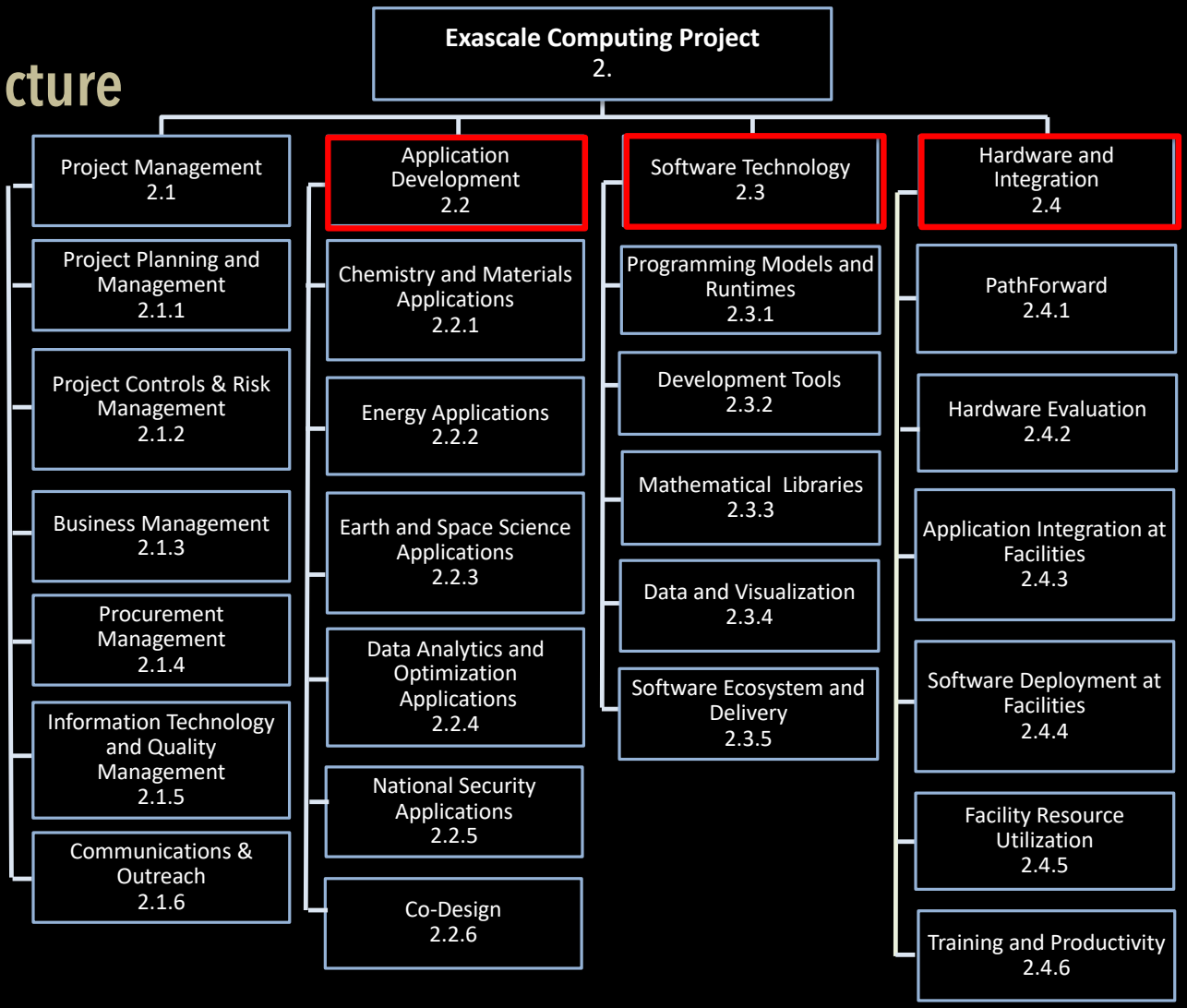


Pre-Exascale Systems

Exascale Systems

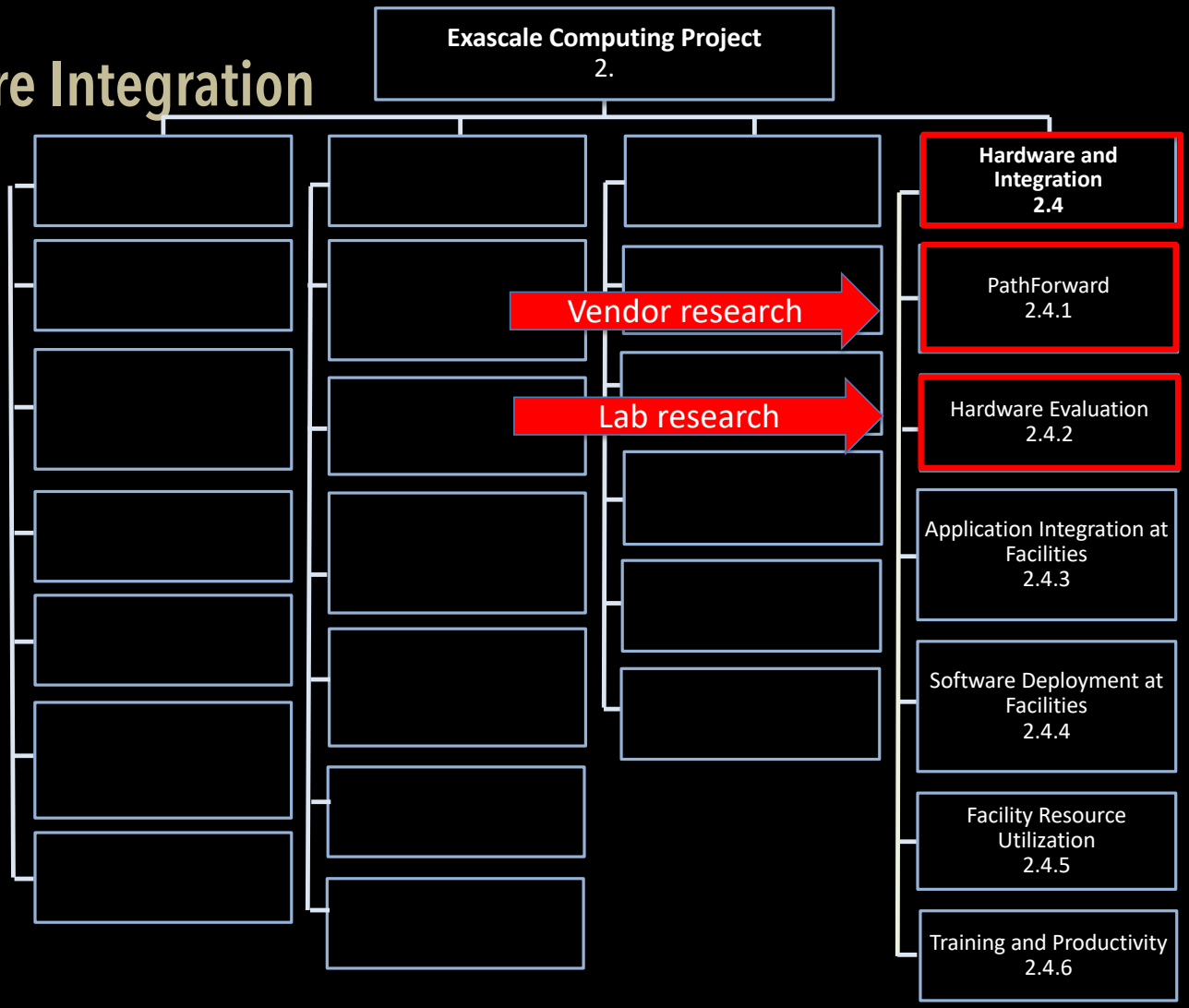
2013	2016	2018	2020	2021-2022
 Argonne IBM BG/Q Open	 Argonne Intel/Cray KNL Open	 ORNL IBM/NVidia P9/Volta Open	**NERSC-9** LBNL TBD Open	 Argonne Intel/Cray TBD Open
 ORNL Cray/NVidia K20 Open	 LBNL Cray/Intel Xeon/KNL Open			**Frontier** ORNL TBD Open
 LLNL IBM BG/Q Secure	 LANL/SNL Cray/Intel Xeon/KNL Secure	 LLNL IBM/NVidia P9/Volta Secure	**Crossroads** LANL/SNL TBD Secure	**El Capitan** LLNL TBD Secure

ECP Structure



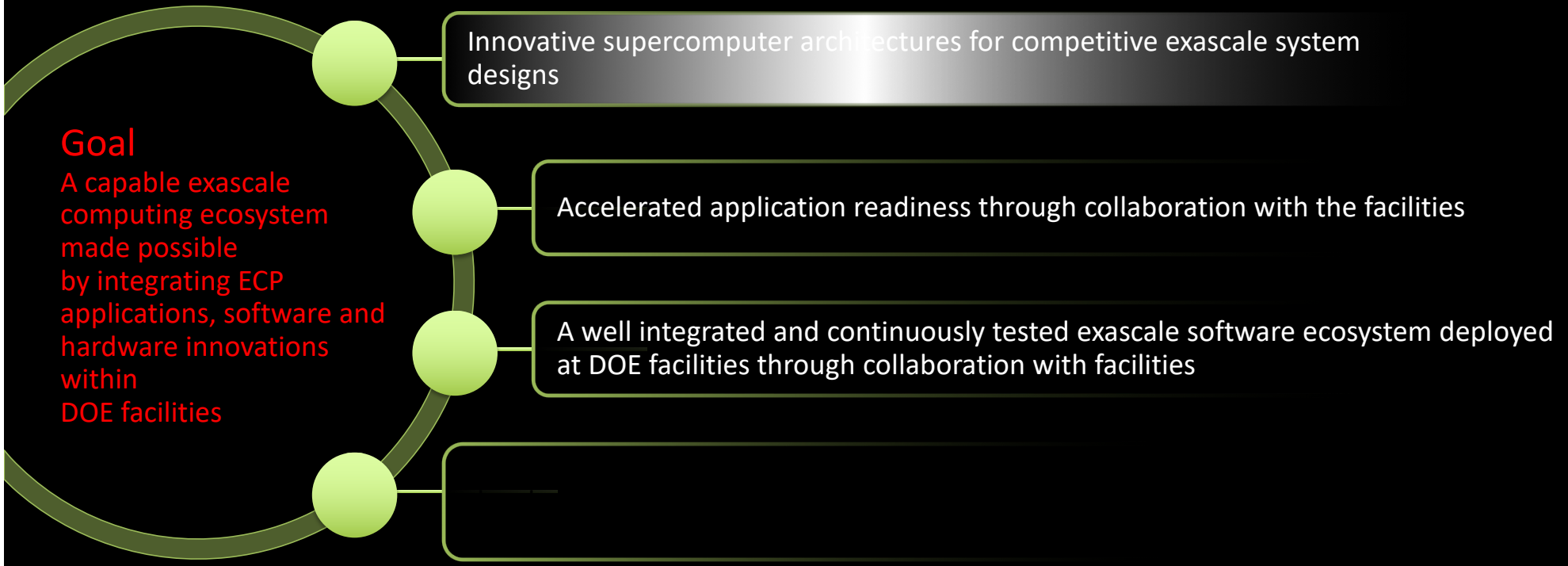
Source: *Doug Kothe, ECP Project Director*

Hardware Integration



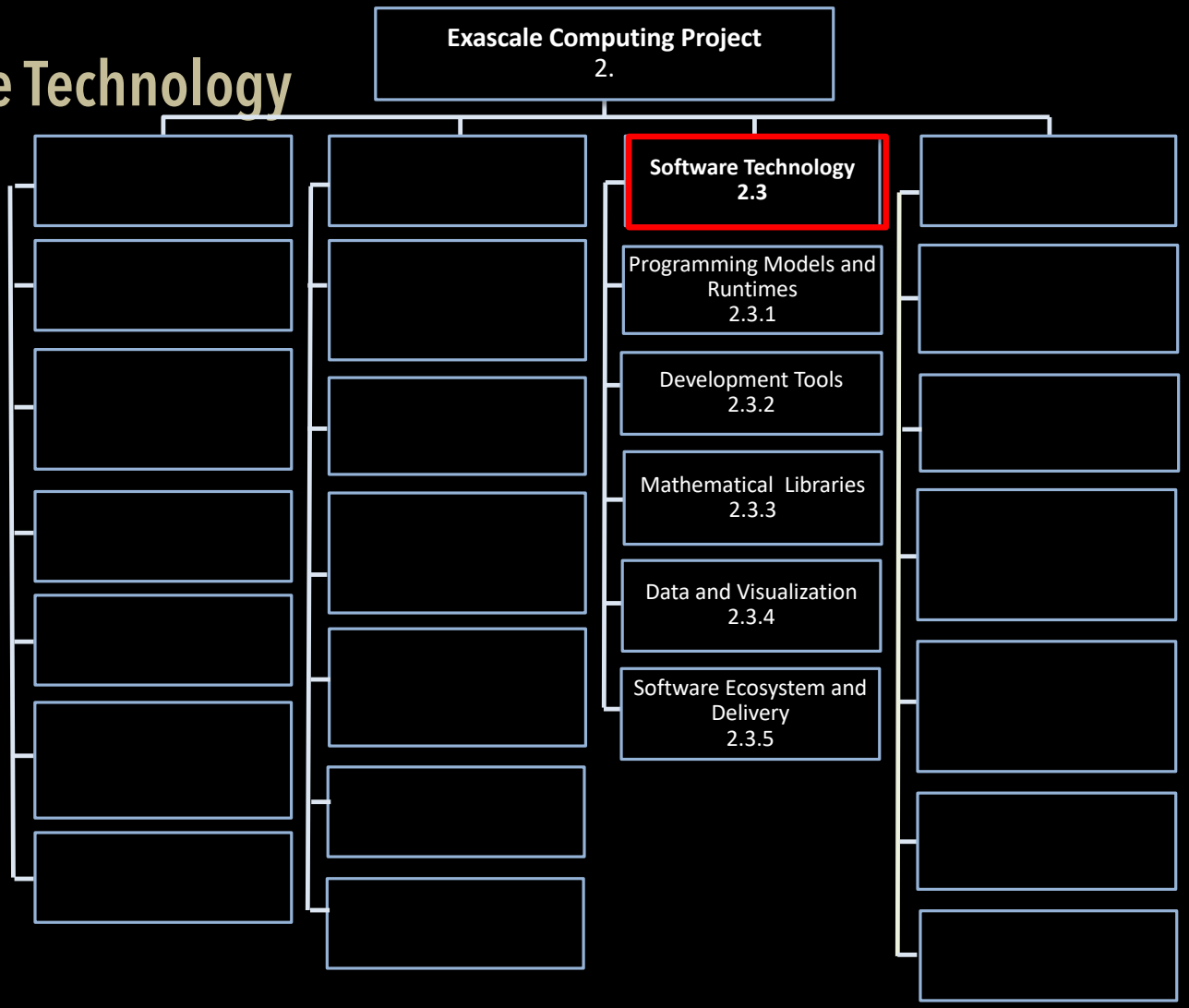
Source: *Doug Kothe, ECP Project Director*

ECP Hardware and Integration: Delivery of integrated ECP/DOE facility products



Source: *Doug Kothe, ECP Project Director*

Software Technology



Source: *Doug Kothe, ECP Project Director*

ECP Software: Productive, Sustainable Ecosystem

Goal

Build a comprehensive, coherent software stack that enables application developers to productively write highly parallel applications that effectively target diverse exascale architectures

- Extend current technologies to exascale where possible
- Perform R&D required for new approaches when necessary
- Guide, and complement, and integrate with vendor efforts
- Develop and deploy high-quality and robust software products

Source: *Doug Kothe, ECP Project Director*

Software Development Kits (SDKs): A Key ST Design Feature

An important delivery vehicle for software products with a direct line of sight to AD applications

ECP software projects

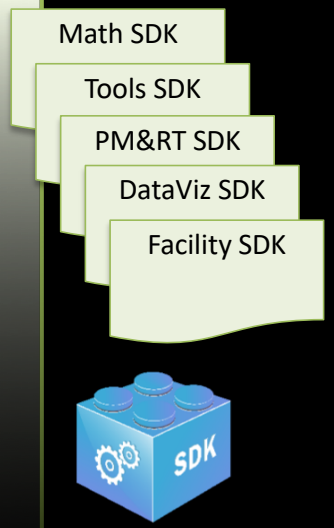
Each project to define (at least 2) release vectors

More projects ⟵ ⟶ Fewer projects

SDKs

Reusable software libraries embedded in applications; cohesive/interdependent libraries released as sets modeled on xSDK

- Regular coordinated releases
- Hierarchical collection built on Spack
- Products may belong to >1 SDK based on dependences
- Establish community policies for library development
- Apply Continuous Integration and other robust testing practices



OpenHPC

Potential exit strategy for binary distributions

- Target similar software to existing OpenHPC stack
- Develop super-scalable release targeting higher end systems

Direct2Facility

Platform-specific software in support of a specified 2021–2023 exascale system

- Software **exclusively** supporting a specific platform
- System software, some tools and runtimes

Assume all releases are delivered as “build from source” via Spack – at least initially

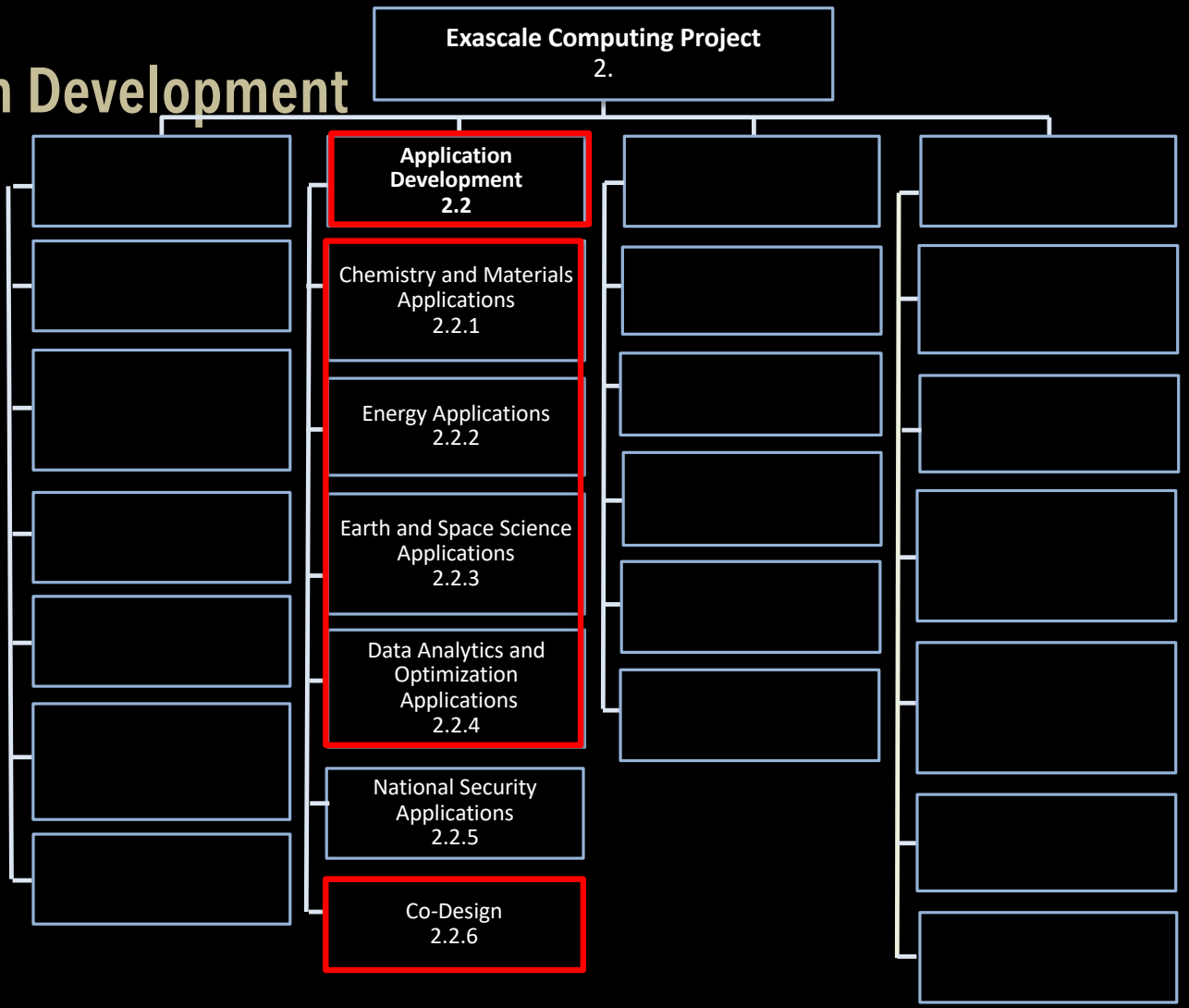
Focus on ensuring that software compiles robustly on all platforms of interest to ECP (including testbeds)

Source: *Doug Kothe, ECP Project Director*

LBNL Participation in ECP Software Technology Projects

- **[Programming Models and Runtimes] Pagoda project**
  - Scott Baden, Paul Hargrove leads (LBNL)
  - GASNet-EX, UPC++
- **[Development Tools] Y-TUNE: Autotuning compiler technology**
  - Mary Hall (Utah) PI, Sam Williams (co-PI) Protono Basu, Brian van Straalen, (LBNL participants), ANL,
  - Provide architecture-specific optimization and tuning underneath high-level language code implementations
- **[Math Libraries] STRUMPACK/SuperLU: Factorization based sparse solvers and preconditioners for Exascale**
  - Sherry Li, Peter Ghysels
  - Reduce arithmetic and communication complexity, develop parallelization strategies that reduce inter-process communication, integrate into higher-level algebraic solver packages
- **[Data/Viz] ExaHDF5: efficient parallel I/O for exascale**
  - Suren Byna (LBNL, PI), Quincey Kozio, HDF, ANL
- **ADIOS Framework for Scientific Data**
  - Scott Klasky (ORNL, PI), John Wu, Junmin Gu (LGNL, Kitware, Rutgers, GA Tech
- **ALPINE: in situ visualization**
  - Jim Ahrens (LANL, PI) , Gunther Weber, Oliver Ruebel (LBNL), Oregon
- **[Software Ecosystem and Delivery] xSDK for ECP**
  - Lois Curfman McInnes (ANL, PI), James Demmel, Sherry Li (UCB/LBNL), SNL, LANL, ORNL, PNNL, LLNL, Colo Sch of Mines, U Tenn,
  - Increase usability. Standardization, interoperability of DOE math and scientific libraries:
    - Hypre, PETSc, SUNDIALS, SuperLU, Trilinos, dense linear algebra packages

Application Development



Source: *Doug Kothe, ECP Project Director*

ECP Applications: Exascale-capable modeling, simulation, data

Goal

Develop and enhance the predictive capability of applications critical to DOE across science, energy, and national security mission space

- Targeted development of requirements-based methods

- Integration of software and hardware via co-design methodologies

- Systematic improvement of exascale readiness and utilization

- Demonstration and assessment of effective software integration

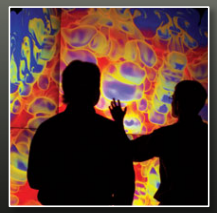
Source: *Doug Kothe, ECP Project Director*

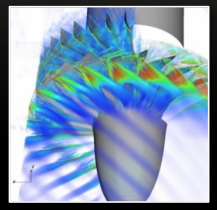
ECP Applications Target National Problems

National security

Stockpile stewardship

Next-generation electromagnetics simulation of hostile environment and virtual flight testing for hypersonic re-entry vehicles





Energy security

Turbine wind plant efficiency

Design and commercialization of SMRs

Nuclear fission and fusion reactor materials design

Subsurface use for carbon capture, petroleum extraction, waste disposal

High-efficiency, low-emission combustion engine and gas turbine design

Carbon capture and sequestration scaleup

Biofuel catalyst design

Economic security

Additive manufacturing of qualifiable metal parts

Urban planning

Reliable and efficient planning of the power grid

Seismic hazard risk assessment

Scientific discovery

Cosmological probe of the standard model of particle physics

Validate fundamental laws of nature

Plasma wakefield accelerator design

Light source-enabled analysis of protein and molecular structure and design

Find, predict, and control materials and properties

Predict and control stable ITER operational performance

Demystify origin of chemical elements

Earth system

Accurate regional impact assessments in Earth system models

Stress-resistant crop analysis and catalytic conversion of biomass-derived alcohols

Metagenomics for analysis of biogeochemical cycles, climate change, environmental remediation

Health care

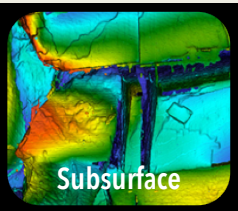
Accelerate and translate cancer research

Source: *Doug Kothe, ECP Project Director*

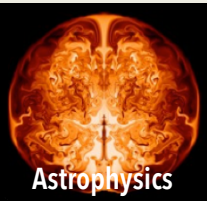
Berkeley Lab has demonstrated unsurpassed ability to harness power of advanced mathematics and computer science to deliver high-impact science

Adaptive Mesh Refinement

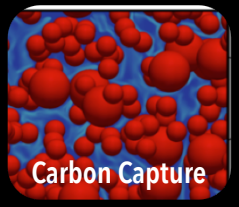
A 1 TeV electron-positron collider



Geo-mechanical chemical evolution of fracking

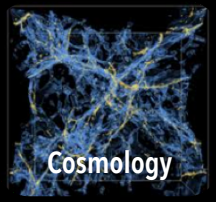


Source of heaviest elements



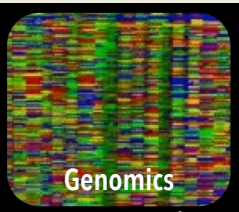
1MWe chemical looping reactor

Regional-scale model to simulate structures

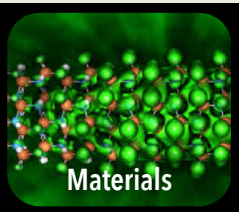


Dark energy equation of state

Scalable (Sparse) Solvers

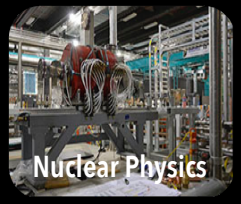


Gene clusters for biomanufacturing



Defects, interfaces and disorder in functional materials

Catalytic conversion of biomass-derived intermediates



Large neutron-rich nuclei and nuclear binding

AMReX Co-Design Center for Exascale Computing *PI: John Bell, LBNL*

The AMR Co-Design Center supports five current ECP Application Projects and provides support for:

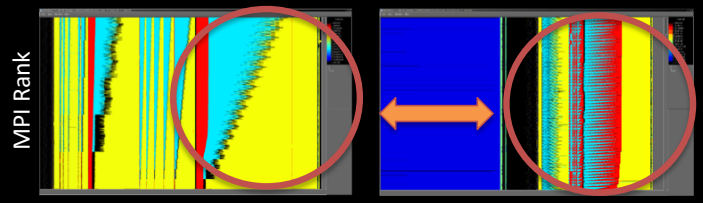
- Data layout and asynchronous data parallelism at the node level
- Internode-communication, dynamic load balancing of heterogeneous workloads, AMR-aware performance modeling of potential strategies
- Asynchronous coarse-grained (fork-join) task parallelism

Coming soon: support for hybrid CPU/GPU systems

Technical Areas:

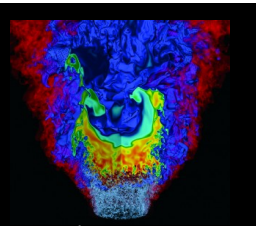
- AMR Mesh support
- Particle data structures and functionality
- Study and adoption of exascale programming models
- Linear solver support for AMR at exascale
- Embedded boundary support
- AMR-specific I/O and Data analysis

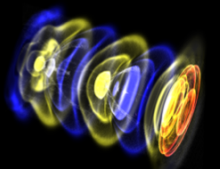
AMReX Profiling Tools:
Example performance visualization: removing collectives in I-O operations

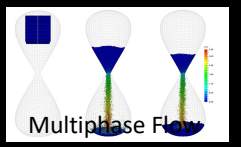


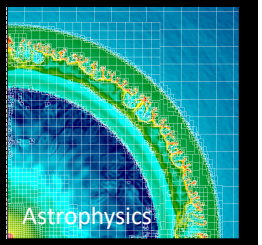
Interactions with ECP Software Technology Projects

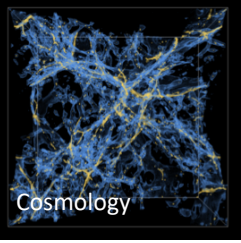
- **SUNDIALS**: Stiff ODE solvers for local physics
- **ALPINE** Visualization and in situ analysis of mesh and particle data
- **Pagoda**: UPC++ for alternative communication model
- **Exascale code generation toolkit**: code generation for stencil operations and automatic code generation for CPUs
- **LANL ATDM Dev.Tools Project**: OpenMP compiler issues


Combustion


Particle Accelerators

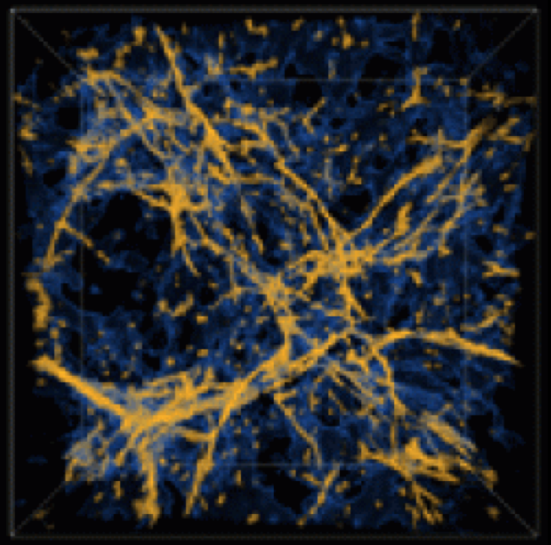

Multiphase Flow


Astrophysics


Cosmology

Cosmology at the Exascale

Synthetic galaxy catalog for LSST generated with HACC and Galacticus codes



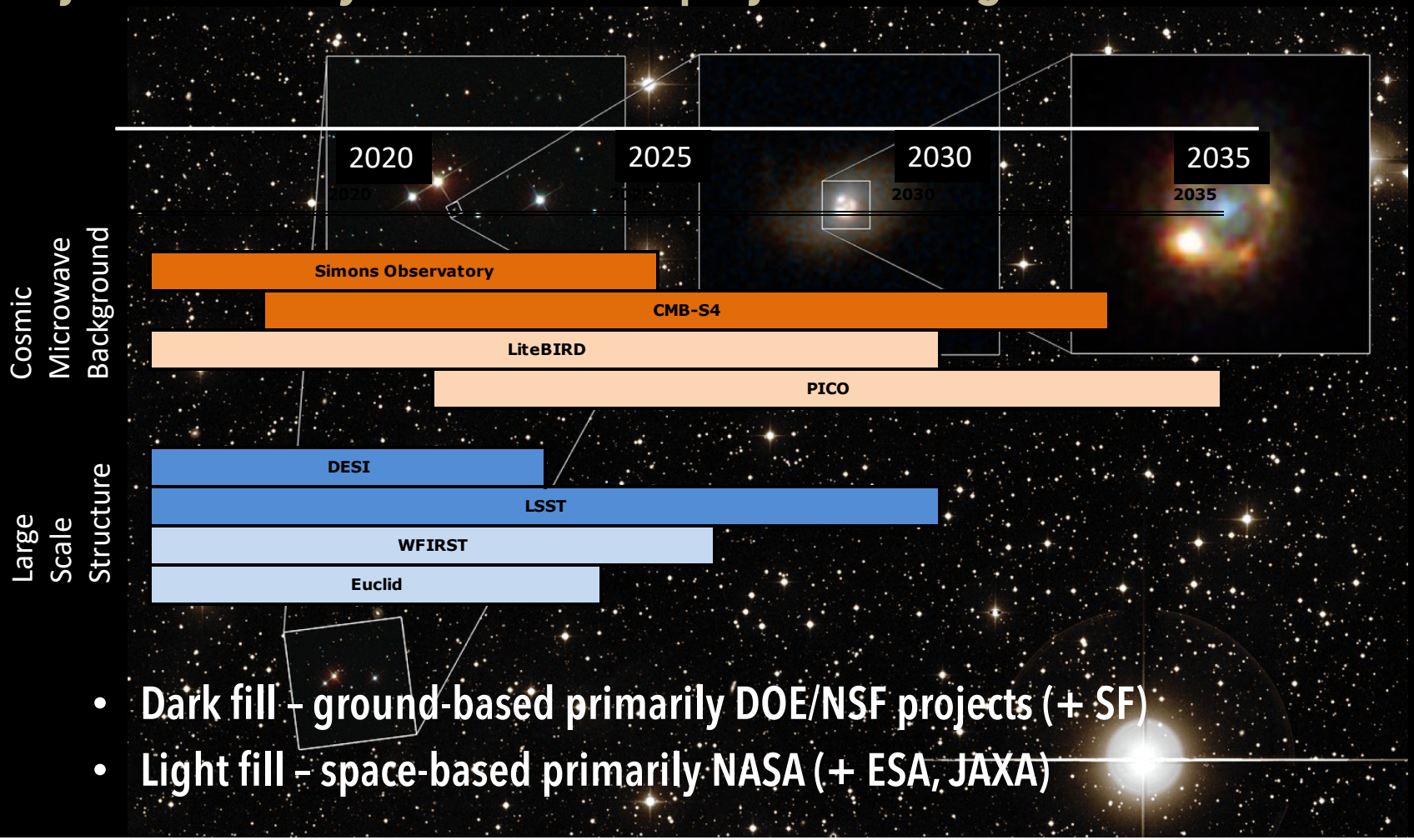
Simulation of Lyman-Alpha Forest with Nyx, used to estimate neutrino mass and as a standard ruler.

Exascale is needed to model and interpret the latest observations

Improve understanding of *Dark Energy, Dark Matter, Primordial Gravitational Waves, Neutrino Mass, and parametrics such as the Hubble Constant*

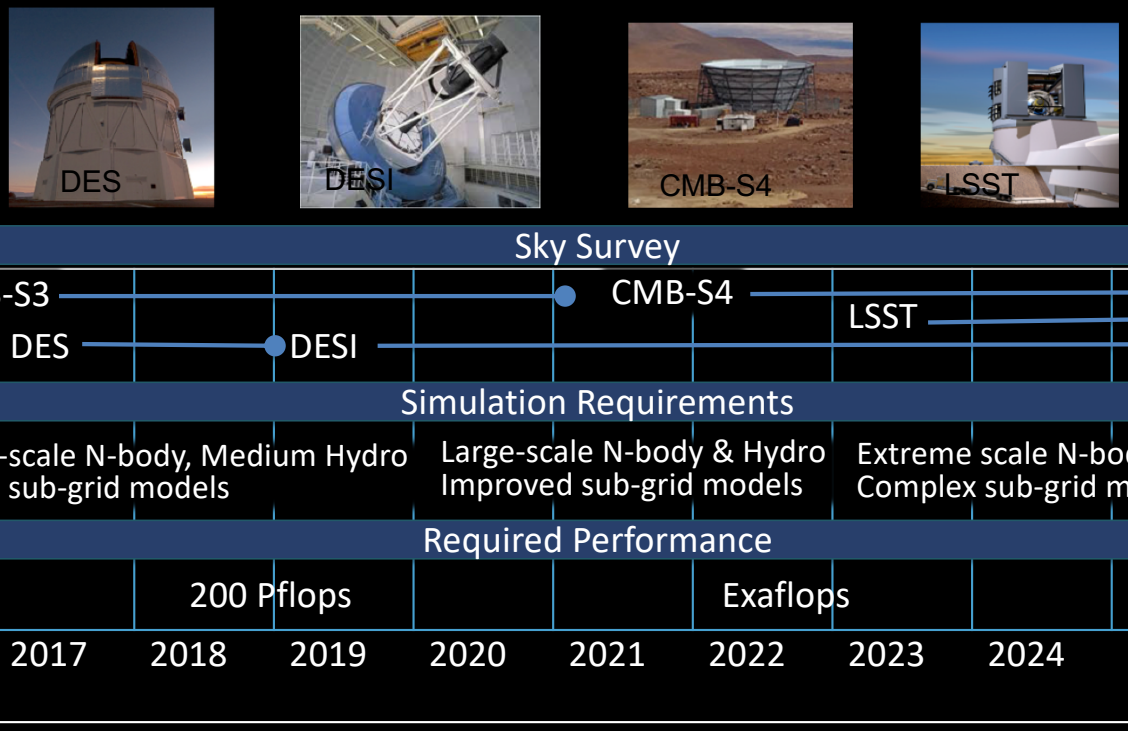
US DOE ECP PI: Salman Habib (ANL)

Major Astronomy observational projects through 2035



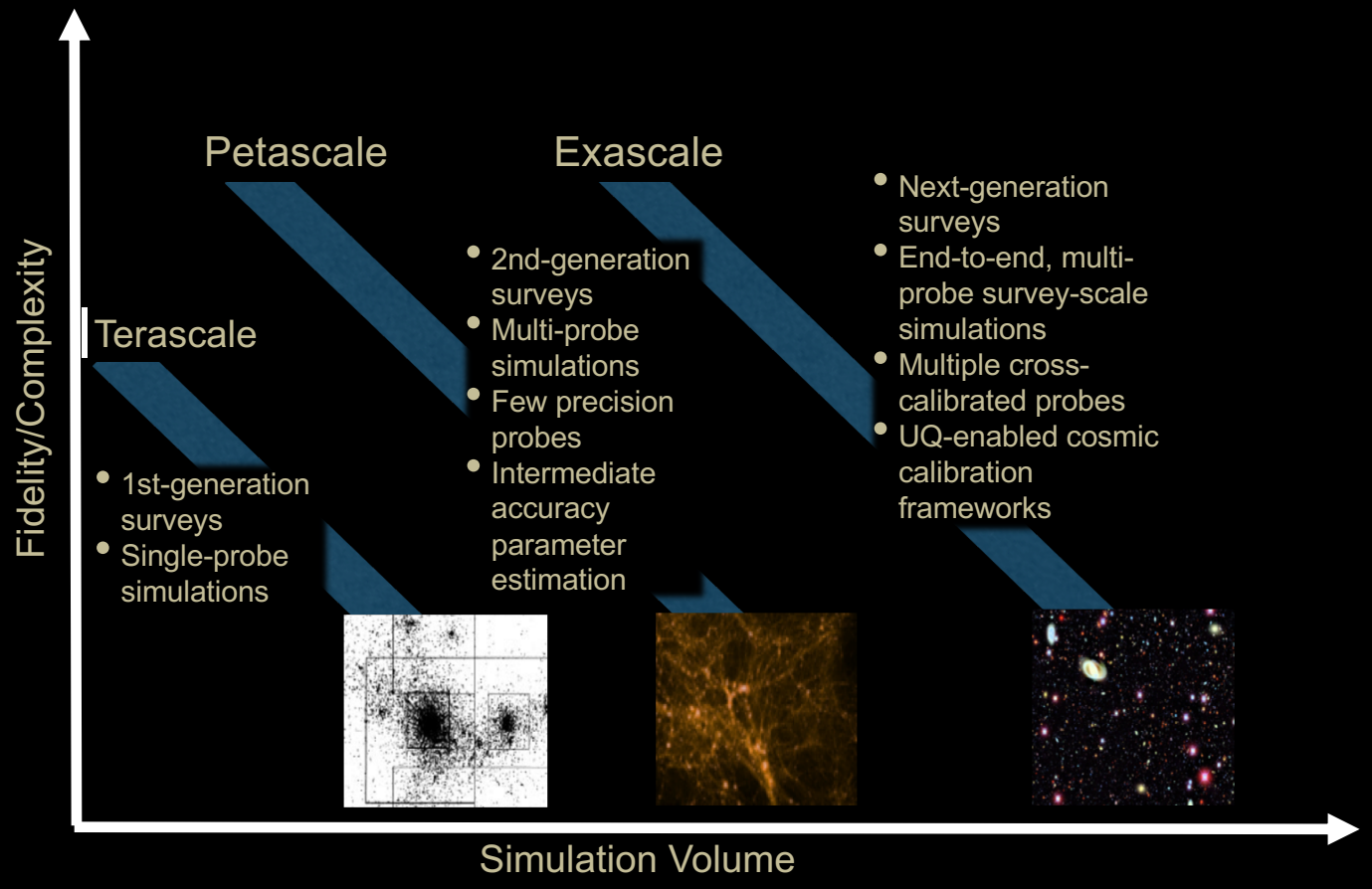
- Dark fill – ground-based primarily DOE/NSF projects (+ SF)
- Light fill – space-based primarily NASA (+ ESA, JAXA)

- **Science:** *Dark Energy, Dark Matter, Gravitational Waves, Neutrino Mass*
- **Computation:** factor of **X100** increase in science reach, order of magnitude improvement in modeling accuracy and predictability



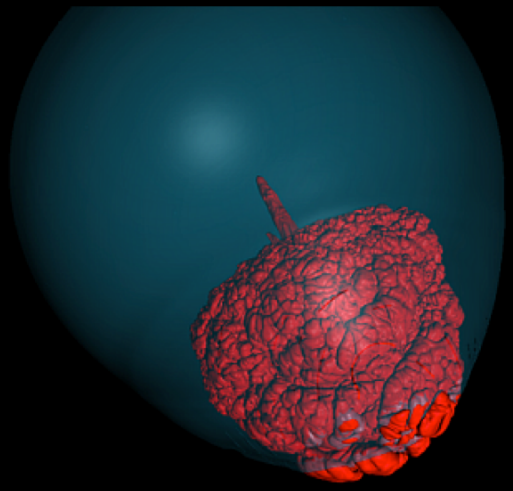
US DOE ECP PI: Salman Habib (ANL)

Precision Cosmology: Simulation Frontiers

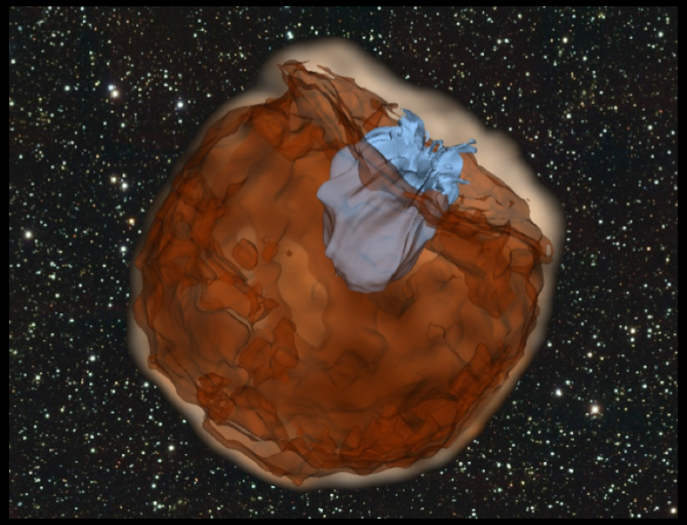


US DOE ECP PI: Salman Habib (ANL)

Astrophysics at the Exascale



Less than a second after ignition, the flame breaks through the surface of an expanded white dwarf (using AMR)



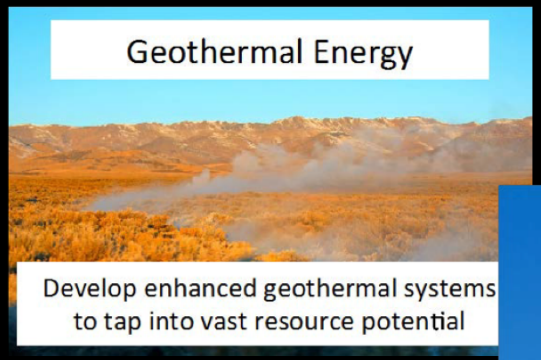
Expanding debris from a supernova explosion (red) running over and shredding a nearby star (blue)

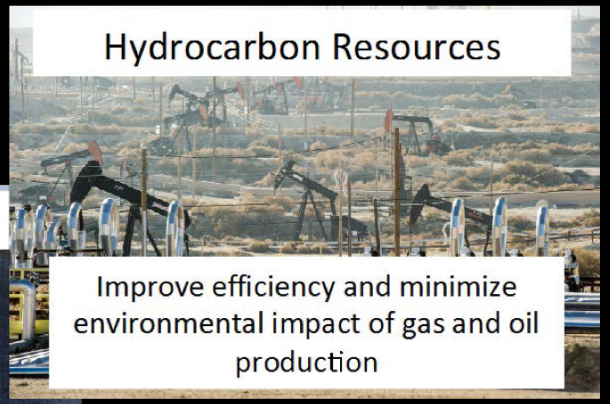
Exascale is needed to identify the source of the heaviest elements

Understand rapid neutron capture process (r-process) by simulating scenarios: core-collapse supernovae, neutron star mergers, and accreting black holes

US DOE ECP PI: Dan Kasen (LBNL)

Subsurface Science at the Exascale



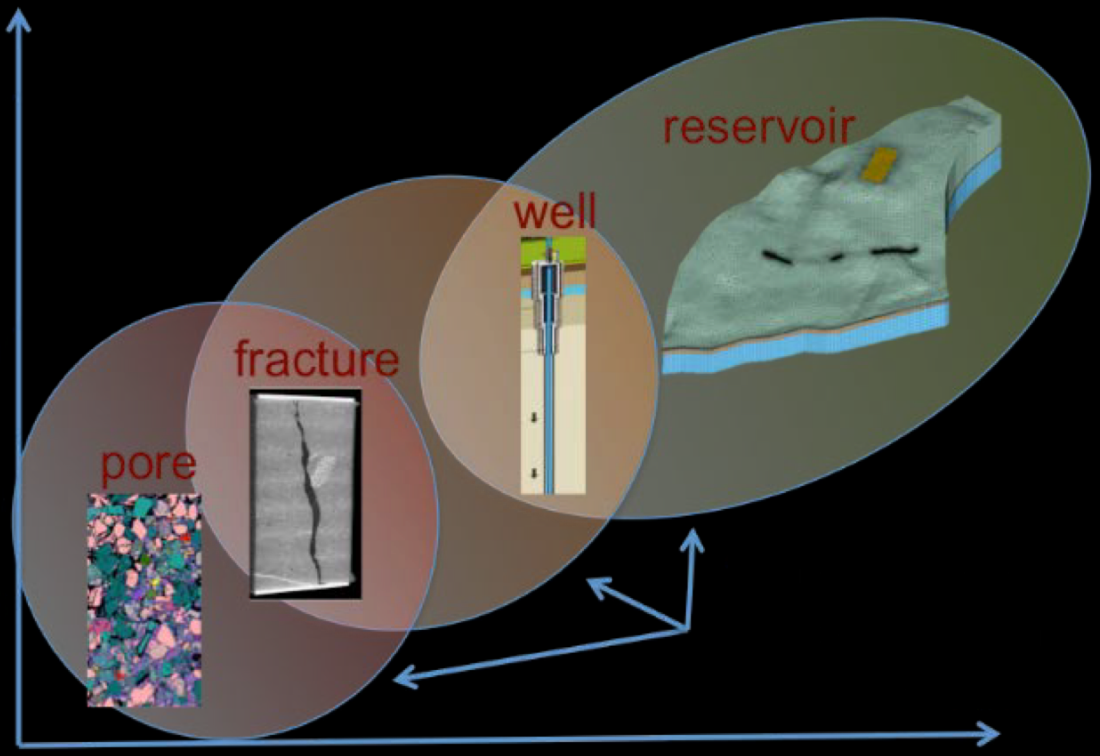


Exascale is needed for impacts of energy extraction and waste storage on subsurface integrity

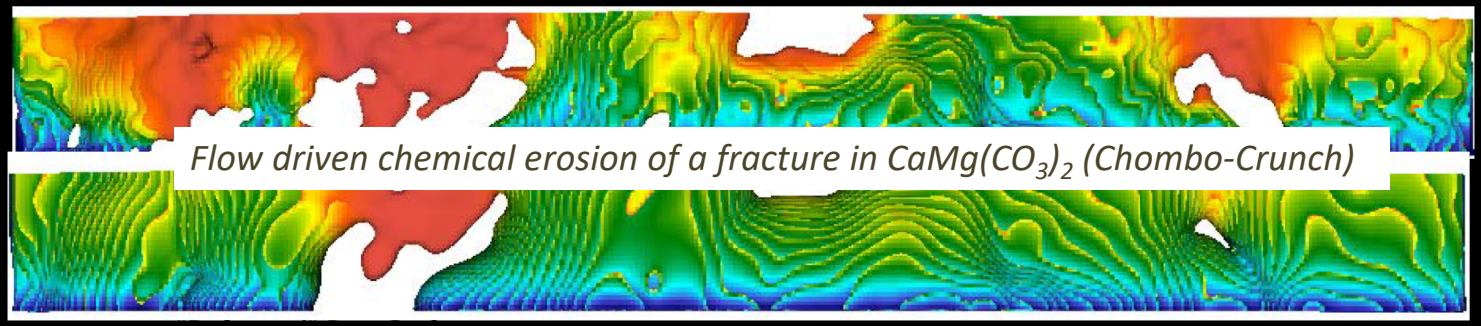
Simulate an entire field of well bores and their interaction through the reservoir over 100 year timescales. Simulate the evolution of a fracture system in caprock subject to geomechanical and geochemical stresses over scales from pore (micron) to 100 meters

US DOE ECP PI: Carl Steefel (LBNL)

Subsurface science requires modeling across scales

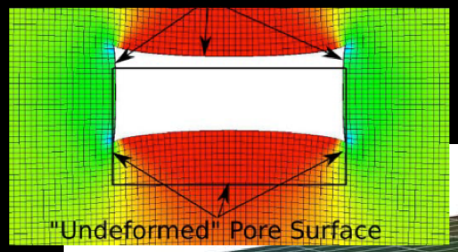


Combining subsurface codes to deliver new science capability



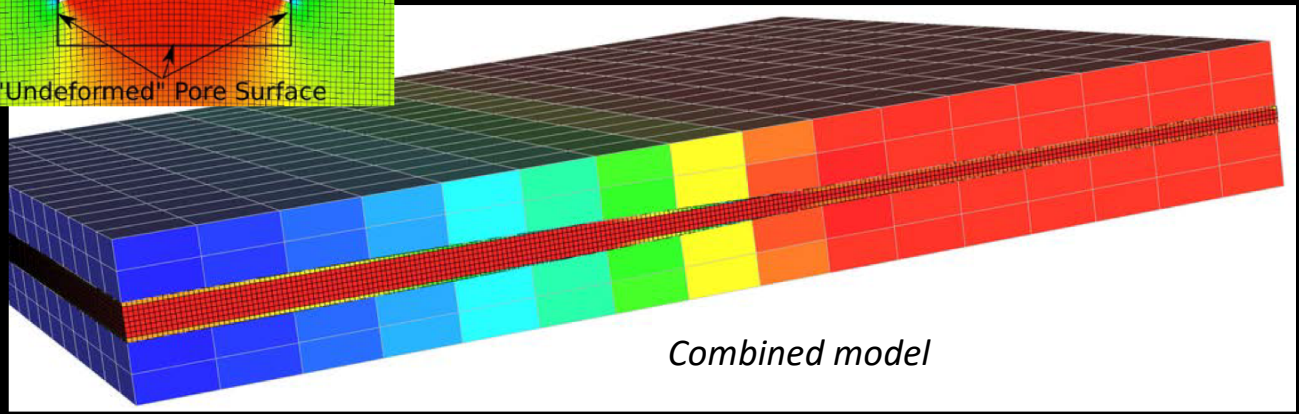
Flow driven chemical erosion of a fracture in $CaMg(CO_3)_2$ (Chombo-Crunch)

+



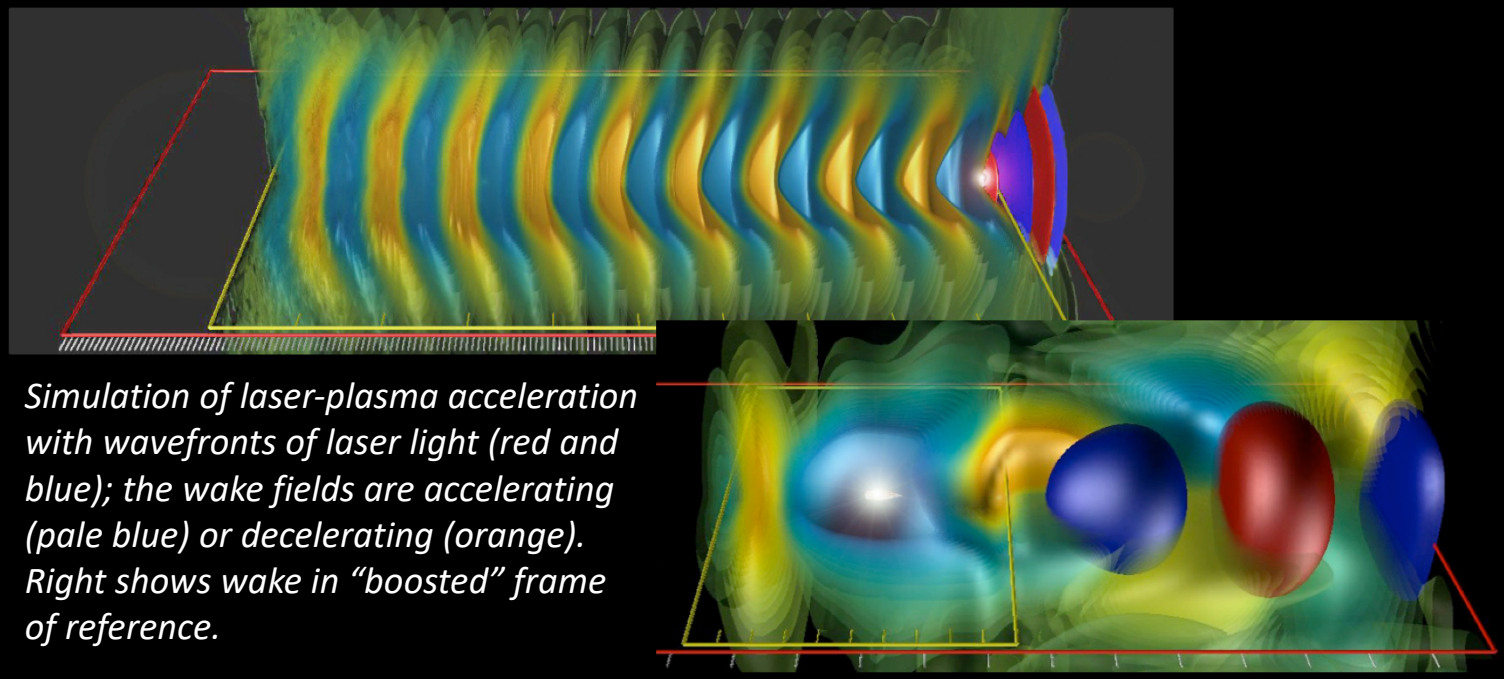
Pore deformation resulting from change in stress loading in a Lagrangian mechanics treatment (GEOS)

=



Combined model

Accelerator Science at the Exascale



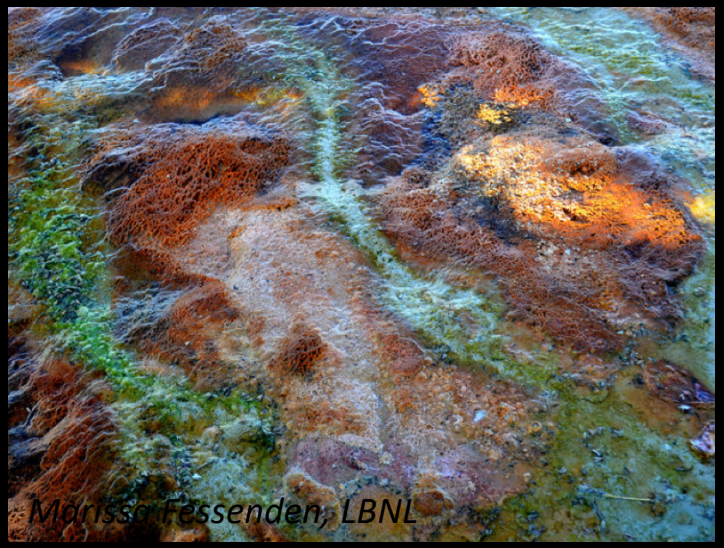
Simulation of laser-plasma acceleration with wavefronts of laser light (red and blue); the wake fields are accelerating (pale blue) or decelerating (orange). Right shows wake in "boosted" frame of reference.

Exascale is needed to simulate future accelerators

Goal: Model a chain of up to a hundred plasma acceleration stages in a few days, for the design of a 1 TeV electron-positron high-energy collider

US DOE ECP PI: Jean Luc Vay (LBNL)

Genome Science at the Exascale



Thermophilic microbial mat in West Thumb Geyser Basin, Yellowstone National Park

Compact CRISPR systems found in deep underground Crystal Geyser bacteria (Banfield)

Exascale is needed to characterize microbial communities

Metagenome analysis with high performance assembly and machine learning; identify gene clusters for energy, environment, biomanufacturing and health

US DOE ECP PI: Kathy Yelick (LBNL)

Powerful Antibiotics Found in Dirt

Posted on February 20, 2018 by Dr. Francis Collins

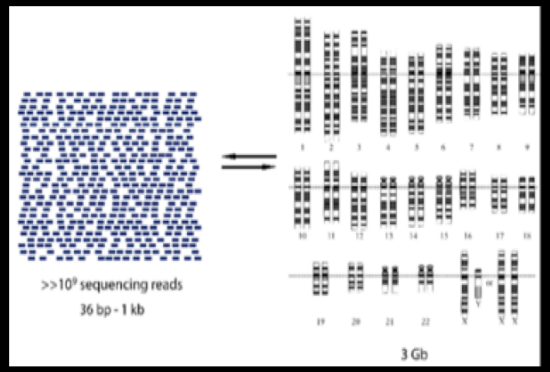
Caption: *Researchers found a new class of antibiotics in a collection of about 2,000 soil samples.*

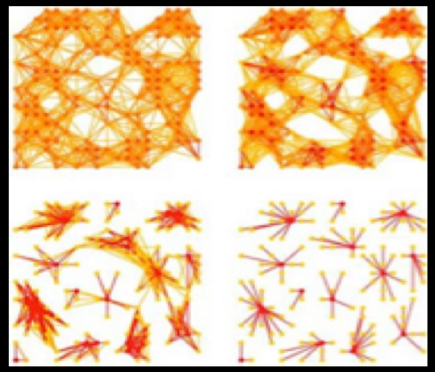
Credit: *Sean Brady, The Rockefeller University, New York*

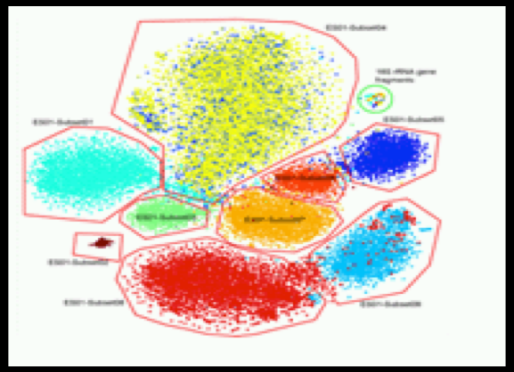
ExaBiome: Microbiome Analysis

Kathy Yelick, PI

Use HPC algorithms and systems for orders of magnitude speedup and to solve previously intractable problems







Metagenome Assembly	Protein Clustering	Comparative Analysis
Assemble millions of metagenomes based on complete data	Cluster billions of proteins	Use fast alignment and annotation for time-sensitive analyses
Graph algorithms, Hash Tables, alignment (Smith-Waterman)	Machine learning (clustering), sparse linear algebra / graphs	Alignment, Machine learning (dimensionality reduction), linear algebra

Summary

- **Exascale computing will deliver science breakthroughs**
  - In simulation and data analytics
  - **But requires advances in models, algorithms and software**
- **Exascale will impact a broad set of applications**
  - Science, health, manufacturing, environment, infrastructure
- **There are still many computer science challenges**
  - Architectures, code generation, algorithms, integration