EXASCALE COMPUTING PROJECT

ADDRESSING A NATIONAL IMPERATIVE

Application Development Update
September 2019

exascaleproject.org

U.S. DEPARTMENT OF ENERGY | Office of Science

NNSA National Nuclear Security Administration

Table Of Contents

Letter from the Application Development Leadership Team

This project overview and update report is focused on Exascale Computing Project (ECP) Application Development. It is the first in a series of three documents to provide an in-depth update on activities related to the ECP's three technical focus areas. The remaining documents—Software Technology and Hardware and Integration—will be released prior to the 2019 annual US high performance computing (HPC) conference, SC19.

The ECP was launched in 2016, bringing together a vastly experienced team of collaborating project management experts, scientists, researchers, and technical experts from participating DOE national laboratories and US HPC companies. Based on an advanced understanding of developing software and applications for high-performance computing, this unprecedented effort to deliver significant research, development, and deployment capabilities will ensure an enduring, capable exascale computing ecosystem for the nation.

The ECP's applications directive is to deliver specific exascale-ready applications critical to the US Department of Energy's mission, integrated with exascale-capable software products and hardware technologies necessary to ensure the success of forthcoming exascale systems from day one.

Currently, the ECP applications development team is working on a total of 24 application projects, each addressing an exascale challenge problem—a high-priority strategic problem of national interest that is intractable without at least 50 times the computational power of the leading computing systems available at the time of the project's inception in 2016. These applications span chemistry, materials, energy, earth and space science, data analytics and optimization, and national security.

The ECP's applications experts are creating and enhancing the predictive capability of these targeted applications through software advances and targeted development of requirements-based models, algorithms, and methods, often in close collaboration with the ECP's six co-design centers.

Bear in mind that these applications could not be effective without the exascale software stack and software development kits currently being developed by the ECP and the coordinated integration of specifically designed hardware components necessary to push the limits of future exascale systems.

We hope you find this overview and update report focused on the ECP's Application Development efforts informative, and we welcome your inquiries.

For the Exascale Computing Project,

Doug Kothe
ECP Director

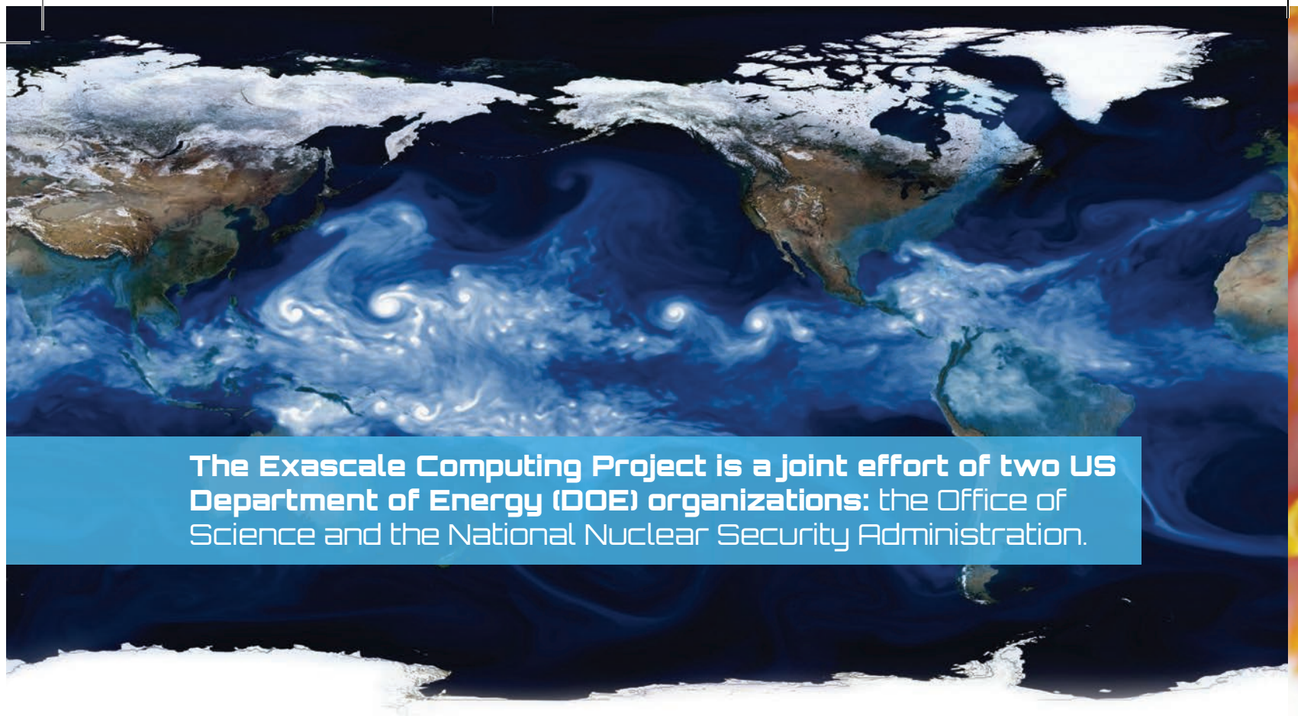
Lori Diachin
ECP Deputy Director

Andrew Siegel
Applications Development Director

Erik Draeger
Applications Development Deputy Director

ACRONYMS

ALE	Arbitrary Lagrangian-Eulerian
ALT	Accuracy, Length, and Time Simulation Space
AM	Additive Manufacturing
AMR	Adaptive Mesh Refinement
AMReX	Adaptive Mesh Refinement for Exascale
ANL	Argonne National Laboratory
API	Application Programming Interface
ARC	Advanced Reactor Concept
ATDM	Advanced Technology Development and Mitigation
BOD	Board of Directors
CANDLE	CANcer Distributed Learning Environment
CC	Coupled Cluster
CCS	Carbon Capture and Storage
CCSNe	Core-Collapse Supernovae
CEED	Center for Efficient Exascale Discretizations
CFD	Computational Fluid Dynamics
CI	Continuous Integration
CLR	Chemical Looping Reactor
CODAR	Center for Online Data Analysis and Reduction at the Exascale
CoPA	Co-design Center for Particle Applications
CPU	Central Processing Unit
DEM	Discrete Element Model
DESI	Dark Energy Spectroscopic Instrument
DFT	Density Functional Theory
DNN	Deep Neural Network
DNS	Direct Numerical Simulation
DOE	US Department of Energy
DTK	Data Transfer Kit
E4S	Extreme-scale Scientific Software Stack
ECI	Exascale Computing Initiative
ECP	Exascale Computing Project
EF	Exaflop
ELGs	Emission Line Galaxies
EOS	Equation of State
ExaAM	Exascale Additive Manufacturing Project
FleCSI	Flexible Computer Science Infrastructure
FNAL	Fermi National Accelerator Laboratory
FRIB	Facility for Rare Isotope Beams
GPU	Graphics Processing Unit
HEDP	High-Energy-Density Physics
HF	Hartree-Fock
HPC	High Performance Computing
HPO	Hyperparameter Optimization
I/O	Input/Output
ICF	Inertial Confinement Fusion
IEMP	Ionizing Electromagnetic Pulse
IPAMS	Integrated Platform for Additive Manufacturing Simulation
ITER	International Tokamak Experimental Reactor
LANL	Los Alamos National Laboratory
LBNL	Lawrence Berkeley National Laboratory
LCLS	Linac Coherent Light Source
LES	Large-Eddy Simulation
LLNL	Lawrence Livermore National Laboratory
LOTF	Laboratory Operations Task Force
LRGs	Luminous Red Galaxies
LSST	Large Synoptic Survey Telescope
MAPP	Multi-Physics on Advanced Platforms Project
MC	Monte Carlo
MD	Molecular Dynamics
MHD	Magneto-Hydrodynamics
ML	Machine Learning
MMF	Multiscale Modeling Framework
MPI	Message Passing Interface
MSN	Mesoporous Silica Nanoparticle
MTIP	Multi-Tiered Iterative Phasing algorithm
NCI	National Cancer Institute
NETL	National Energy Technology Laboratory
NIH	National Institutes of Health
NNSA	National Nuclear Security Administration
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
PF	Petaflop
PIC	Particle-in-Cell
PMR	Programming Models and Runtimes
PNNL	Pacific Northwest National Laboratory
PPPL	Princeton Plasma Physics Laboratory
QC	Quantum Chemistry
QCD	Quantum Chromodynamics
QMC	Quantum Monte Carlo
RCCI	Reactivity-Controlled Compression Ignition
RNAS/LES	Reynolds-Averaged-Navier-Stokes/Large-Eddy-Simulation
SC	Office of Science
SDK	Software Development Kit
SFBA	San Francisco Bay Area
SGD	Stochastic Grid Dynamics
SGEMP	System-Generated-Electromagnetic-Pulse
SLAC	Stanford Linear Accelerator Center
SMR	Small Modular Reactor
SNAP	Spectral Neighbor Analysis Potentials
SNL	Sandia National Laboratories
SPARC	Sandia Parallel Aerodynamics and Reentry Code
SPI	Single-Particle Imaging
SREMP	Source-Region-Electromagnetic-Pulse
SSP	Stockpile Stewardship Program
TN	Thermonuclear
WDM	Whole Device Model
WDMApp	Whole Device Model Application



The Exascale Computing Project is a joint effort of two US Department of Energy (DOE) organizations: the Office of Science and the National Nuclear Security Administration.

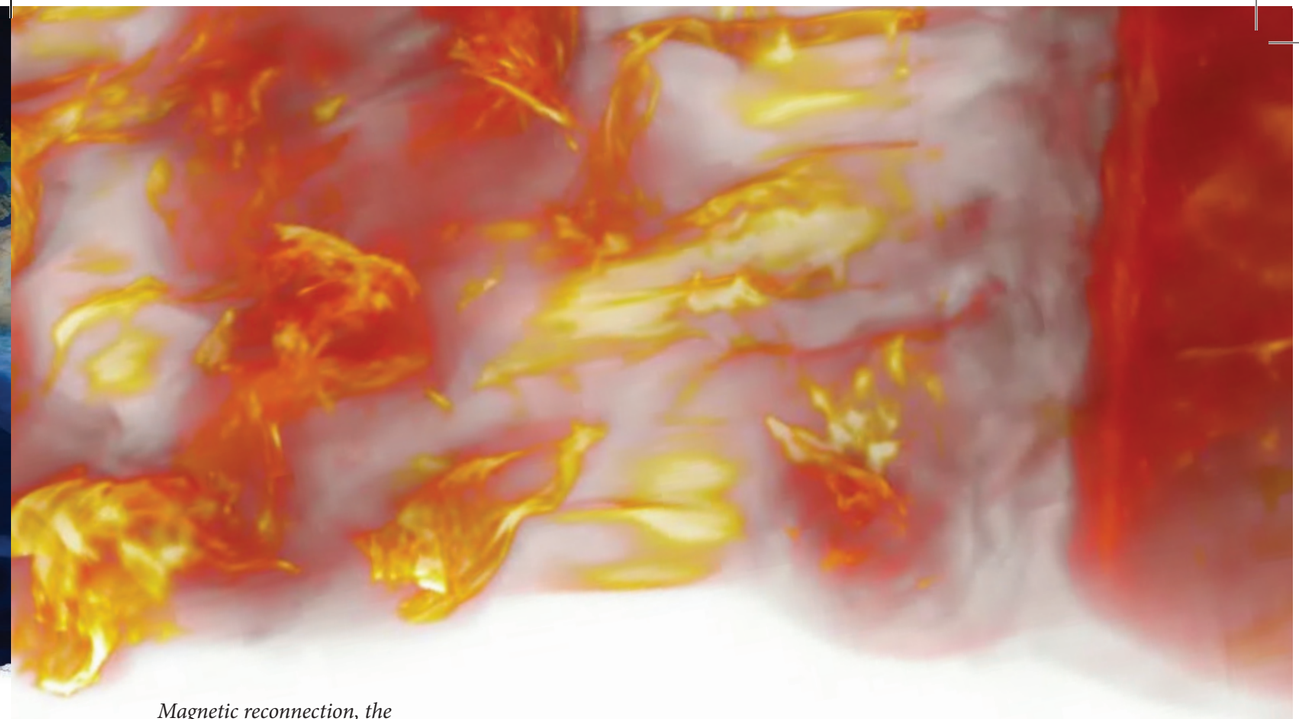
Community Atmospheric Model used in attributing changes in the risk of extreme weather and climate. Source: Department of Energy, Argonne Leadership Computing Facility.

US Department of Energy
Office of Science
Advanced Scientific Computing Research (ASCR) Program

The US Department of Energy's Advanced Scientific Computing Research (ASCR) program is one of six interdisciplinary scientific program offices within the Office of Science along with Basic Energy Sciences, Biological and Environmental Research, Fusion Energy Sciences, High Energy Physics and Nuclear Physics.

The ASCR program leads the nation and the world in supercomputing, high-end computational science, and advanced networking for science with its mission to discover, develop, and deploy computational and networking capabilities to analyze, model, simulate, and predict complex phenomena important to the US Department of Energy (DOE).

ASCR has already initiated investments to address the challenges of hybrid, multi-core computing up to the exascale (capable of an exaflop, or 10^{18} floating point operations per second). However, there are significant technological challenges that must be addressed at the exascale to reduce the energy demands and increase the memory available so the systems will be useful for science and engineering. Addressing these challenges will result in not only exascale systems but also in affordable, energy-efficient petascale systems and high-end desktops to drive scientific and engineering discovery across the country. With this integrated approach, ASCR will continue to deliver scientific insights to address national problems in energy and the environment while advancing US competitiveness in information technology and the nation's high-tech industry.



Magnetic reconnection, the continuous breaking and rearrangement of magnetic field lines in a plasma, is a fundamental process in physics. Understanding reconnection phenomena has broad implications and may eventually help us protect astronauts, communications satellites, and electrical power grids. Source: Department of Energy, Los Alamos National Laboratory.

US Department of Energy
National Nuclear Security Administration
Advanced Simulation and Computing (ASC) Program

Established in 1995, the Advanced Simulation and Computing (ASC) Program provides the National Nuclear Security Administration (NNSA) Office of Defense Programs the simulation-based predictive science capabilities for the stewardship of the US nuclear weapons stockpile. Under ASC, high-performance simulation capabilities are developed to analyze and predict the performance, safety, and reliability of nuclear weapons and to certify their functionality.

Beyond the stewardship of today's stockpile, the rapidly growing capability of potential adversaries to defeat US weapons using advanced defensive systems represents a sobering threat. ASC applications must also target performance assessment of current and life-extended weapon systems subject to a wide variety of hostile environments and potential threat scenarios. To execute its mission for the NNSA Stockpile Stewardship Program (SSP), ASC oversees the high-performance simulation and computing work of three NNSA laboratories—Los Alamos National Laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), and Sandia National Laboratories (SNL) as a nationally coordinated program.

The ECP Core Partner Laboratories

Six US Department of Energy (DOE) laboratories manage and oversee the Exascale Computing Project (ECP) through the project's Board of Directors, Laboratory Operations Task Force (LOTF), and the ECP senior leadership team

Argonne National Laboratory (ANL) is a multidisciplinary research center with a pioneering history in high performance computing. Home to the Argonne Leadership Computing Facility, the laboratory provides supercomputing resources to the research community to accelerate scientific discovery and innovation. Argonne is preparing to deploy Aurora, one of DOE's three planned exascale systems, in 2021.

The Computing Sciences Area at Lawrence Berkeley National Laboratory (LBNL) provides the computing and networking resources and expertise critical to advancing Department of Energy Office of Science (DOE-SC) research missions: developing new energy sources, improving energy efficiency, developing new materials, and increasing our understanding of ourselves, our world, and our universe.

Oak Ridge National Laboratory (ORNL) is home to Summit, the world's most powerful computer. Since 2005, the Oak Ridge Leadership Computing Facility has deployed Jaguar, Titan, and Summit, each the world's fastest computer in its time, and will launch its first exascale system, Frontier, in 2021.

Founded in 1952, Lawrence Livermore National Laboratory (LLNL) will welcome ElCapitan, the first exascale supercomputer to support the National Nuclear Security Administration, in 2022–23. LLNL has a long pedigree of world-class supercomputing in the service of national security and basic science, from the Univac in 1953 to today's Sierra and Sequoia pre-exascale systems.

As the senior laboratory in the DOE system, Los Alamos National Laboratory (LANL) executes work in all of DOE's missions: national security, science, energy, and environmental management. LANL's contributions are part of what makes DOE a science, technology, and engineering powerhouse for the nation. LANL is strongly represented across the breadth of ECP governance, applications, and software technologies which form the exascale ecosystem.

Sandia National Laboratories (SNL) is a multidisciplinary national laboratory that develops advanced technologies to ensure global peace. SNL leads and participates in many projects across the spectrum of the ECP. Sandia's leadership spans the ECP software technology director role, combustion science and climate modeling applications, as well as software technologies ranging from performance portability to large-scale visualization.

1,000,000,000,000,000

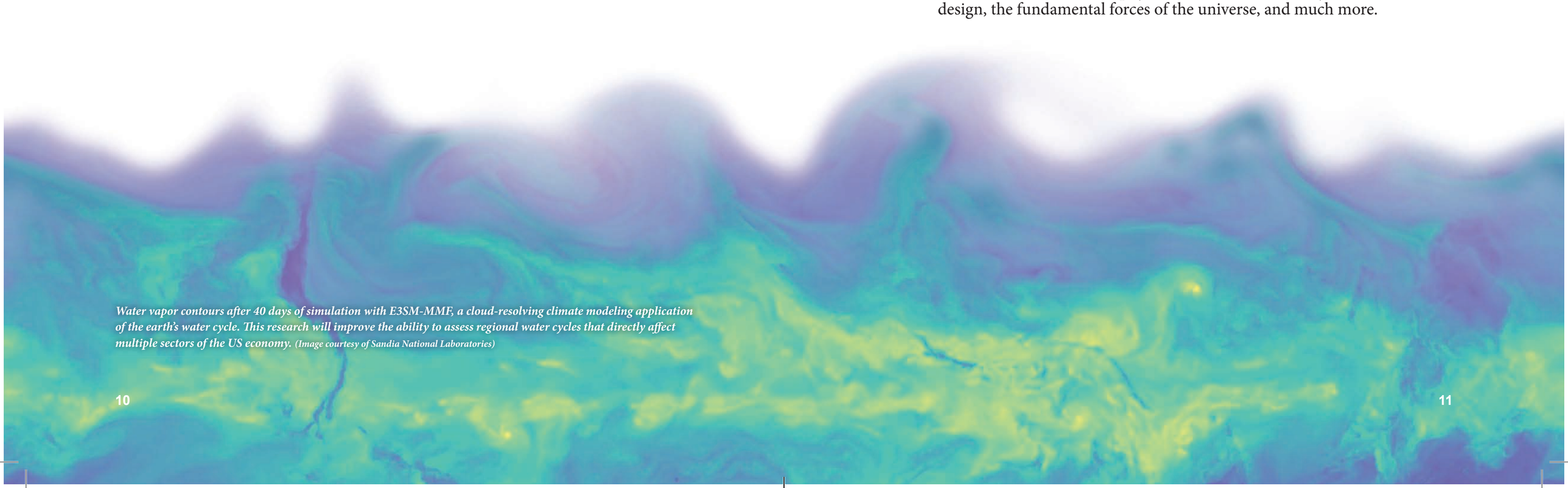
Exascale systems will perform more than a quintillion (a billion billion) operations per second.

Achieving exascale will have profound effects on the American people and the world—improving the nation's economic competitiveness, advancing scientific discovery, and strengthening our national security

The fastest supercomputers in the world today solve problems at the petascale—that is quadrillion (10^{15}) calculations each second.

While these petascale systems are quite powerful, the next milestone in computing achievement is the exascale—a higher level of performance in computing that will have profound impacts on everyday life.

At quintillion (10^{18}) calculations each second, exascale supercomputers will more realistically simulate the processes involved in scientific discovery and national security such as precision medicine, regional climate, additive manufacturing, the conversion of plants to biofuels, the relationship between energy and water use, the unseen physics in materials discovery and design, the fundamental forces of the universe, and much more.



Water vapor contours after 40 days of simulation with E3SM-MMF, a cloud-resolving climate modeling application of the earth's water cycle. This research will improve the ability to assess regional water cycles that directly affect multiple sectors of the US economy. (Image courtesy of Sandia National Laboratories)

Oak Ridge National Laboratory

Oak Ridge National Laboratory is the host site for the project office of the Exascale Computing Project.

ORNL's supercomputing program grew from humble beginnings to deliver the world's most powerful system several times over the past three decades. On the way, it has helped researchers deliver practical breakthroughs and new scientific knowledge in climate, materials, nuclear science, and a wide range of other disciplines.

INTRODUCTION

The quest to develop a capable exascale ecosystem is a monumental effort

The US Department of Energy (DOE) is a long-time global leader in the development and use of high performance computing (HPC). HPC-based modeling and simulation is vital to the execution of DOE missions in science and engineering and to DOE's responsibility for stewardship of the nation's nuclear weapons stockpile.

Over the past several decades, sustained technology investment has supported the development of increasingly powerful HPC systems and produced substantial benefits for the United States. However, as other nations have recognized the benefits of HPC and increased their investments, US leadership in this important area is no longer ensured—yet this leadership is essential to our economic, energy, and national security.

To maintain leadership and to address future challenges in economic impact areas such as national security, energy assurance, economic competitiveness, healthcare, and scientific discovery, as well as growing security threats, the United States is making a strategic move in HPC—a grand convergence of advances in co-design, modeling and simulation, data analytics, machine learning, and artificial intelligence.

Collaboration, Partnership, and a Strategic Focus

The DOE-led Exascale Computing Initiative (ECI), a partnership between two DOE organizations, the Office of Science (SC) and the National Nuclear Security Administration (NNSA), was formed in 2016 to accelerate research, development, acquisition, and deployment projects to deliver exacale computing capability to the DOE labs by the early to mid 2020s. The ECI includes three main components: (1) SC and NNSA computer facility site preparation investments, (2) computer vendor nonrecurring engineering activities needed for the delivery of exascale systems within this time fame, and (3) the Exascale Computing Project (ECP), which was launched in 2016 and brings together research, development, and deployment activities as part of a capable exascale computing ecosystem to ensure an enduring exascale computing capability for the nation.

The ECP is focused on delivering specific applications, software products, and outcomes on DOE computing facilities. Integration across these elements with specific hardware technologies for exascale system instantiations is fundamental to the success of the ECP. The outcome of the ECP is the accelerated delivery of a capable exascale computing ecosystem to provide breakthrough solutions addressing our most critical challenges in scientific discovery, energy assurance, economic competitiveness, and national security. This outcome is not simply a matter of ensuring more powerful computing systems. The ECP is designed to create more valuable and rapid insights from a wide variety of applications ("capable"), which requires a much higher level of inherent efficacy in all methods, software tools, and ECP-enabled computing technologies to be acquired by the DOE laboratories ("ecosystem").

ECP is built on a strategic collaboration of government, academia, and industry

Advanced leadership computing capabilities are required to

- discover new energy solutions needed for a sustainable future;
- extend our knowledge of the natural world through scientific inquiry;
- maintain a vibrant effort in science and engineering as a cornerstone of the nation's economic prosperity;
- deliver new technologies to advance DOE's mission; and
- sustain a world-leading workforce in advanced technology.

The ECP is led by a team of senior scientists, project management experts, and engineers from six of the largest DOE national laboratories. Working together, this leadership team has established an extensive network to deliver a capable exascale computing ecosystem for the nation.

The ECP enables US revolutions in technology development: scientific discovery; healthcare; energy, economic, and national security

ECP Mission

Develop exascale-ready applications and solutions that address currently intractable problems of strategic importance and national interest.

Create and deploy an expanded and vertically integrated software stack on DOE HPC pre-exascale and exascale systems.

Deliver **US HPC vendor technology advances and deploy ECP products** to DOE HPC pre-exascale and exascale systems.

ECP Vision

Deliver **exascale simulation and data science innovations** and solutions to national problems that enhance US economic competitiveness, improve our quality of life, and strengthen our national security.

Argonne National Laboratory, the nation's first national laboratory, will be home to Aurora, one of the US Department of Energy's three planned exascale supercomputers.

PREPARING FOR THE NATION'S EXASCALE SYSTEMS

Preparing for the Nation's Exascale Systems

Today, US scientists can take advantage of powerful pre-exascale systems at two of DOE's leading HPC facilities: the 200 petaflop IBM Summit system at Oak Ridge National Laboratory and the 125 petaflop IBM Sierra system at Lawrence Livermore National Laboratory. These systems serve numerous scientific and national security programs today and are critical stepping stones as we prepare for the nation's first exascale supercomputers to be procured between 2021 and 2023.

As of June 2019, the DOE had the top two systems on the world ranking of the TOP500 computing systems, a list that is updated twice yearly. Access to these systems is proving invaluable in helping scientists prepare their codes for the forthcoming exascale platforms. The planned US exascale systems are critically important for addressing the growing computational challenges and sustaining the nation's preeminence in technological advances and economic competitiveness. In addition to solving important computational grand challenge science problems and addressing serious economic, environmental, and national security challenges, these exascale systems and the technology that will ultimately be made available to the broad HPC community will ensure the United States is at the forefront of HPC on a global scale.

For the past few decades, DOE's continued commitment to advances in supercomputing has fueled a robust partnership with the computer industry in software and hardware, which has led to the development of new technologies and markets surrounding them in a variety of areas ranging from HPC architectures, memory technologies, high-speed interconnects, systems software, programming environments, and high-performance storage systems. This commitment to the delivery of new supercomputing technologies has served both to advance our energy and science missions as well as to advance the technology industry and sustain a world-leading workforce. Advances in hardware and software not only enable new and more advanced science but also lead to important advances in all areas touched by computing and computing technology, thereby driving innovations in many other industries such as automotive, aerospace, chemical production, enhanced oil recovery, precision medicine, and nuclear energy.

Supported by the efforts of the ECP, the United States is preparing for the arrival of three exascale systems starting in 2021—***Aurora*** at Argonne National Laboratory, ***Frontier*** at Oak Ridge National Laboratory, and ***El Capitan*** at Lawrence Livermore National Laboratory.

The success of these first US exascale systems will depend largely on the ECP to fulfill its mission of building exascale-ready software and applications and influencing the development of vendor hardware specifically needed to deploy capable exascale systems. In support of this mission, ECP is funding several efforts from national laboratories, industry, and academia to fill the gaps in hardware, software services, and applications.

Pre-Exascale Systems

2012

ORNL
Cray/AMD/
NVIDIA

ANL
IBM BG/Q

LLNL
IBM BG/Q

2016

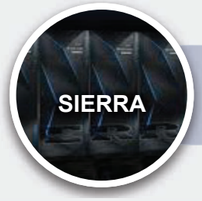
LBNL
Cray/Intel

ANL
Intel/Cray

LANL/SNL
Cray/Intel

2018

ORNL
IBM/NVIDIA



LLNL
IBM/NVIDIA

Future Exascale Systems

2020

LBNL
Cray/AMD/
NVIDIA

2021–2023

ORNL
Cray/AMD

ANL
Intel/Cray

LANL/SNL
TBD

LLNL
Cray

AURORA

Argonne National Laboratory

Argonne National Laboratory's next-generation supercomputer, Aurora, will be one of the nation's first exascale systems when it is delivered in 2021. Designed in collaboration with Intel and Cray, Aurora will help ensure continued US leadership in high-end computing for scientific research.

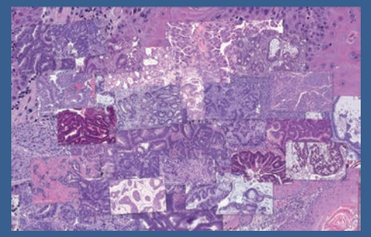
Scientists will use Aurora to pursue some of the farthest-reaching science and engineering breakthroughs ever achieved with supercomputing. From mapping the human brain and designing new functional materials to advancing the development of alternative energy sources, Argonne's forthcoming machine will enable researchers to accelerate discoveries and innovation across scientific disciplines.

Aurora will be based on Intel's Xeon Scalable processors, high-performance Intel X^e GPU compute accelerators, and Optane DC persistent memory. The system will rely on Cray's Shasta exascale-class architecture and Slingshot interconnect technology, which can provide concurrent support for advanced simulation and modeling, AI, and analytics workflows. Aurora will leverage historical advances in software investments along with increased application portability via Intel's OneAPI. The supercomputer will also introduce a new I/O system called Distributed Asynchronous Object Storage (DAOS) to meet the needs of exascale workloads.

SYSTEM SPECS	AURORA
Sustained Performance	≥ 1 EF
Cabinets	> 100
Node	Intel Xeon scalable processor, multiple X^e arch based GP-GPUs
Aggregate System Memory	> 10 PB
System Interconnect	Cray Slingshot providing 100 GB/s network bandwidth. Slingshot dragonfly network providing adaptive routing, congestion management, and quality of service.
High-Performance Storage	> 230 PB, > 25 TB/s (DAOS)
Programming Models	Intel OneAPI, OpenMP, DPC++/SYCL

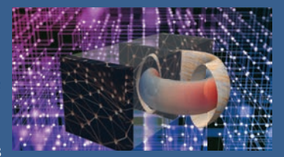
Aurora will usher in a new era of scientific discovery and innovation

"What excites me most about exascale systems like Aurora is the fact that we now have, in one platform and one environment, the ability to mix simulation and artificial intelligence. This idea of mixing simulation and data-intensive science will give us an unprecedented capability and open doors in research which were inaccessible before, like cancer research, materials science, climate science, and cosmology."



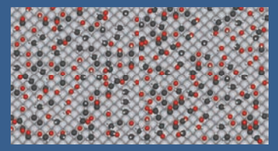
– Rick Stevens, Argonne Associate Laboratory Director for Computing, Environment, and Life Sciences; Aurora Early Science Program Principal Investigator

"Exascale computing's ability to handle much larger volumes of data unlocks our ability to prove what was once unprovable. Plus, the incredible speed of supercomputers shortens our time-to-discovery by a huge margin. Work that used to take months or years now takes hours or days. Therefore, we finally have the means to validate theories statistically and prove their reality."

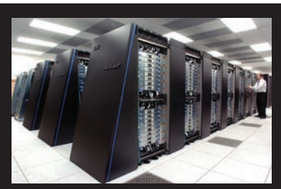


– William Tang, Principal Research Physicist at Princeton Plasma Physics Laboratory; Aurora Early Science Program Principal Investigator

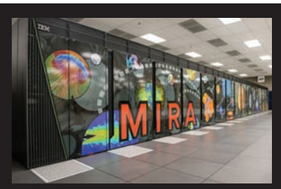
"Exascale will allow us to solve yesterday's insurmountable scientific problems. Until now, we have been forced to navigate trade-offs between the fidelity of a simulation, the time necessary to derive it, and the number of simulations we can perform with a finite number of cycles. With a greatly increased computational resource like Aurora, we can perform vastly more high-fidelity simulations. Therefore, we can get much better quantitative descriptions to fuel our work."



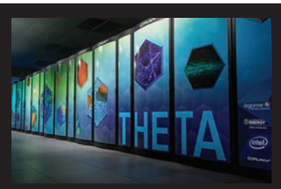
– David Bross, Argonne Computational Chemist; Aurora Early Science Program Principal Investigator



Intrepid
2007–2013



Mira
2012–2019



Theta
2016–TBD

Aurora
Coming in 2021

Leadership Computing

Aurora will be housed at the Argonne Leadership Computing Facility (ALCF), a DOE Office of Science User Facility that deploys and operates world-class supercomputers for open science research. The ALCF was established at Argonne National Laboratory in 2006 as part of a DOE initiative dedicated to providing leading-edge computing resources to the science and engineering community to advance fundamental discovery and understanding in a broad range of disciplines. From Intrepid to Mira to Theta, each new ALCF system brings advanced capabilities that enable researchers to expand their investigations in both scope and scale.

FRONTIER

Oak Ridge National Laboratory

Scheduled for delivery in 2021, Frontier is ORNL's exascale supercomputer. Frontier will accelerate innovation in science and technology and maintain US leadership in high-performance computing and artificial intelligence. Frontier will be able to simulate the detailed life cycle of a nuclear reactor, help to uncover the genetics of complex diseases, and allow scientists to build on recent developments in science and technology by further integrating artificial intelligence with more detailed data analytics coupled to new approaches to modeling and simulation.
The system will be based on Cray's new Shasta architecture and Slingshot interconnect with high-performance AMD EPYC CPU and Radeon Instinct GPU technology. The new accelerator-centric compute blades will support a 4:1 GPU-to-CPU ratio with high-speed links and coherent memory between them within the node. With Frontier, scientists will be able to pack in more calculations, identify new patterns in data, and develop innovative data analysis methods to accelerate the pace of scientific discovery.

SYSTEM SPECS	SUMMIT	FRONTIER
Peak Performance	200 PF	> 1.5 EF
Cabinets	256	> 100
Node	2 IBM POWER9 CPUs 6 NVIDIA Volta GPUs	1 HPC and AI Optimized AMD EPYC CPU 4 Purpose-built AMD Radeon Instinct GPU
CPU-GPU Interconnect	NVLINK Coherent memory across the node	AMD Infinity Fabric Coherent memory across the node
System Interconnect	2x Mellanox EDR 100G InfiniBand Non-Blocking Fat-Tree	Multiple Slingshot Network Interface Controller (NICs) providing 100 GB/s network bandwidth. Slingshot dragonfly network provides adaptive routing, congestion management, and quality of service.
Storage	250 PB, 2.5 TB/s, GPFS	2–4x performance and capacity of Summit's I/O subsystem. Frontier will have near node storage like Summit.

Frontier will help guide researchers to new discoveries at exascale.

"As a third-generation accelerated system—following the world-leading Summit system deployed at ORNL in 2018—Frontier will provide unmatched capability for modeling and simulation studies along with new capabilities for deep learning, machine learning and data analytics for applications ranging from manufacturing to human health."

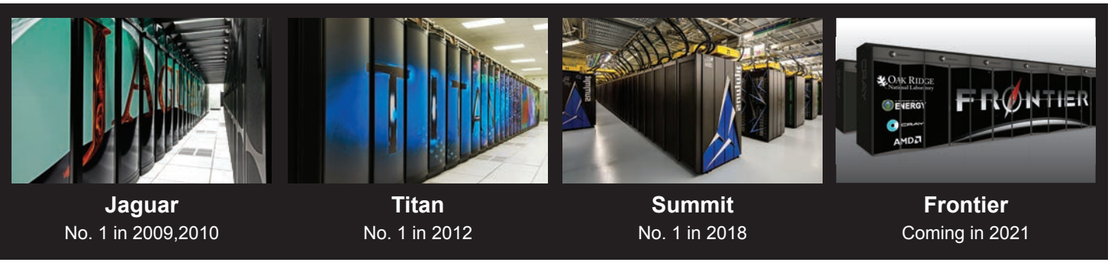
–James J. Hack, Director, National Center for Computational Sciences at ORNL

"The ability to have a large amount of very fast memory like we're going to have on Frontier will be a real boon to our simulations."

– Bronson Messer,
ORNL and the ECP ExaStar Team

"The thing that's really attractive about Frontier is the powerful nodes. Having fewer powerful nodes with a very tightly integrated set of CPUs and GPUs at the node level gives us the ability to distribute hundreds or thousands of microstructure and property calculations on one or a few nodes across the machine."

– John Turner, ORNL and the ECP ExaAM Team



World-Leading Systems

Oak Ridge National Laboratory has decades of experience in delivering, operating, and conducting research on world-leading supercomputers. Since 2005, ORNL has deployed Jaguar, Titan, and Summit, each the world's fastest computer in its time. Frontier will leverage ORNL's extensive experience and expertise in GPU-accelerated computing to become the US DOE's next record-breaking supercomputer when it debuts in 2021.

EL CAPITAN

Lawrence Livermore National Laboratory

El Capitan will be the National Nuclear Security Administration's (NNSA's) first exascale supercomputer. Scheduled for delivery to LLNL in late 2022, El Capitan will feature advanced capabilities for modeling, simulation and artificial intelligence.

Boasting a peak performance of more than 1.5 exaflops, El Capitan will perform essential functions for the NNSA's Stockpile Stewardship Program, which supports US national security missions through leading-edge scientific, engineering, and technical tools and expertise, ensuring the safety, security, and effectiveness of the nation's nuclear stockpile in the absence of underground testing. El Capitan will be used to make critical assessments necessary for addressing evolving threats to national security and other purposes such as nonproliferation and nuclear counterterrorism.

El Capitan will be built on Cray's Shasta supercomputing architecture and will be composed of Shasta compute nodes and a future generation of ClusterStor storage. This unique architecture will be connected with Cray's new Slingshot high-speed interconnect. The Shasta architecture can accommodate a variety of processors and accelerators, making it possible for Cray and LLNL to make a late-binding decision on CPU and GPU components in the coming months.

	SIERRA	EL CAPITAN	EL CAPITAN and SIERRA comparison
Peak (Exaflop/s)	125 PF	> 1.5	10.5× more performance
System Power (MegaWatts)	11.0	< 40	>3.3× more power efficient
Application Performance Improvement		6× to 12× over Sierra	

El Capitan will provide unprecedented capabilities in support of the nation's nuclear deterrent

"El Capitan will allow our scientists and engineers to get answers to critical questions about the nuclear stockpile faster and more accurately than ever before, improving our efficiency and productivity and enhancing our ability to reach our mission and national security goals."

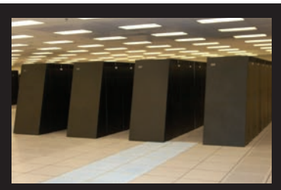
– Bill Goldstein, Director, Lawrence Livermore National Laboratory

"A machine of this magnitude will be key for the rapid, 3D iterative analyses required to modernize our deterrent, as adversaries are making rapid improvements in their defensive and offensive capabilities."

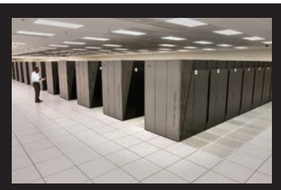
–Michel McCoy, LLNL Advanced Simulation & Computing Program Director

"El Capitan will continue the GPU-accelerated era begun at LLNL with Sierra. This system architecture offers outstanding price/performance that will ensure that ASC contributes critical computing cycles to NNSA's mission in FY24 and beyond. We are excited to resume the LLNL partnership with Cray after its long dormancy."

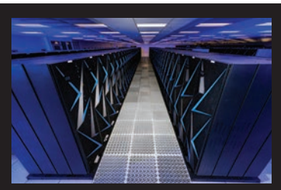
– Bronis R. de Supinski, Chief Technology Officer for Livermore Computing, Lawrence Livermore National Laboratory



BlueGene/L
No. 1 in 2004-2008



Sequoia
No. 1 in 2012



Sierra
No. 2 in 2018,2019

El Capitan
expected 2022

A History of Leading-Edge Supercomputing at LLNL

Since it was founded in 1952, Lawrence Livermore National Laboratory has prided itself on being the tip of the spear for cutting-edge computing. The first computer, the UNIVAC I, was ordered before the Lab even opened, marking the beginning of a decades-long mission to develop the world's fastest and most powerful computers and to use those machines to solve large, complex problems. Over the years, LLNL supercomputers have topped more Top500 lists of the world's fastest and most-powerful systems than any other computing facility on Earth, the most recent being the IBM/Blue Gene Sequoia in 2012. LLNL's current most powerful supercomputer Sierra is second only to Oak Ridge's Summit.

Lawrence Berkeley National Laboratory

Thirteen Nobel prizes are associated with Lawrence Berkeley National Laboratory, located on a 202-acre site in the hills above the UC Berkeley campus.

APPLICATION DEVELOPMENT

The Compelling Case for Exascale Applications

Exascale Is a Watershed for Extreme Scale Computing

To maintain leadership and to address future challenges in national security, energy assurance, economic competitiveness, healthcare, and scientific discovery, as well as growing security threats, the United States is making a strategic move in high performance computing (HPC)—a grand convergence of advances in co-design, modeling and simulation, data analytics, machine learning, and artificial intelligence.

The nation needs exascale computing to drive economic competitiveness and strengthen national security, and the key to ensuring that our forthcoming exascale systems are able to address science from day one lies in our ability to have a portfolio of exascale-ready applications.

Exascale-ready applications are a foundational element of the ECP and critical to achieving DOE's goal of science from day one on the forthcoming exascale systems. The ECP has launched its mission need application projects, each addressing an exascale challenge problem—a high-priority strategic problem of national interest that is intractable without at least 50 times the computational power of the majority of the DOE's current HPC systems. These applications span chemistry, materials, energy, earth and space science, data analytics and optimization, and national security. The ECP's Application Development team will create or enhance the predictive capability of these applications through algorithmic and software advances via co-design centers and targeted development of requirements-based models, algorithms, and methods. In addition, the ECP's application team will provide systematic improvement of exascale system readiness and utilization and demonstration and assessment of effective software integration.

With a suite of exascale-ready applications, scientists will be able solve problems 50 times faster or more complex than typically possible today. Exascale will enable new approaches to predictive analysis for scientific discovery and stockpile stewardship, new means of addressing the sobering threat posed by the rapidly growing capabilities of potential adversaries to use advanced defensive systems, and new solutions to complex data-driven engineering problems.

"Exascale will have a profound impact on the strength of the nation and the quality of life for all citizens."

Doug Kothe
ECP Director

ECP Application Development

The mission of the ECP Application Development group is to develop and enhance the predictive capability of targeted applications deemed critical to national interests in science, energy, and national security.

The ECP has formed a collaboration, representing 15 of the DOE's national laboratories, additional partner universities, and participating hardware technology manufacturers, to create or enhance these applications through algorithmic and software advances via the ECP co-design centers and targeted development of requirements-based models, algorithms, and methods.

Applications will also be integrated with the ECP software stack and vendor system architectures for maximum capability and efficiency. Currently there are 24 application teams engaged in this effort with access to pre-exascale systems for early development and performance assessment.

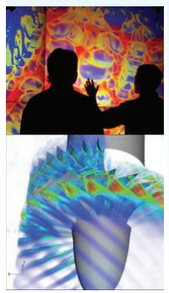
The critical importance of the ECP application portfolio can't be overstated.

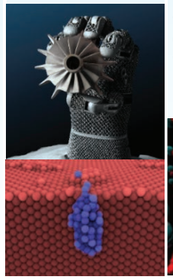
ECP application development targets science-based applications for which exascale will enable high-confidence insights and answers to critical problems that simply can't be addressed with today's computing platforms.

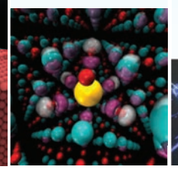
THE ECP Application Portfolio

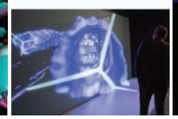
Applications are the bridge between the nation's scientists and the forthcoming exascale systems. The ECP portfolio of 24 applications being prepared for exascale is targeting national problems in six strategic areas: national security, energy security, economic security, scientific discovery, earth system, and health care.

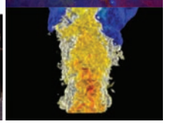
National security	Energy security	Economic security	Scientific discovery	Earth system	Health care
Stockpile stewardship Next generation simulation tools for assessing nuclear weapons performance Response to hostile threat environments and reentry conditions	Turbine wind plant efficiency High-efficiency, low-emission combustion engine and gas turbine design Materials design for extreme environments of nuclear fission and fusion reactors Design and commercialization of Small Modular Reactors Subsurface use for carbon capture, petroleum extraction, waste disposal Scale-up of clean fossil fuel combustion Biofuel catalyst design	Additive manufacturing of qualifiable metal parts Reliable and efficient planning of the power grid Seismic hazard risk assessment	Find, predict, and control materials and properties Cosmological probe of the standard model of particle physics Validate fundamental laws of nature Demystify origin of chemical elements Light source-enabled analysis of protein and molecular structure and design Whole-device model of magnetically confined fusion plasmas	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	Accelerate and translate cancer research











How ECP Applications Are Organized

To take advantage of common requirements, shared knowledge, and efficient project management, the ECP applications are organized into six major categories.

- Chemistry and MaterialsApplications
- Energy Applications
- Earth and Space Science Applications
- Data Analytics and Optimization Applications
- Co-design Centers
- National Security Applications

Chemistry and Material Applications focus on simulation capabilities that attempt to precisely describe the underlying properties of matter needed to optimize and control the design of new materials and energy technologies. These applications require the use of sophisticated models and algorithms to solve complex physics equations.

Energy Applications focus on the modeling and simulation of existing and future technologies for the efficient and responsible production of energy to meet the growing needs of the United States. These applications generally require detailed modeling of complex facilities and multiple coupled physical processes. Their goal is to help overcome obstacles to the efficient and safe delivery of energy. The Energy Applications activity encompasses the research areas of combustion, nuclear energy, fusion energy, wind energy, chemical energy, and particle accelerators.

Earth and Space Science Applications span fundamental scientific questions, from the origin of the universe and chemical elements to planetary processes and interactions affecting life and longevity. These application areas treat phenomena where controlled and fine resolution data collection is extremely difficult or infeasible, and, in many cases, fundamental simulations are our best source of data to confirm scientific theories and predict critical phenomena. This activity encompasses the research areas of earthquake simulations, subsurface geophysics, climate forecasting, cosmology, and stellar astrophysics.

Data Analytics and Optimization Applications is an emerging area whose predictive capability is partially based on modern data analysis and machine learning techniques rather than strictly on approximate solutions to equations that state fundamental physical principles or reduced semiempirical models. This activity encompasses a broad range of research areas and techniques, some of which are only recently coming into maturity in the context of high-end simulation. As such, they represent greater risk but also significant potential for new discovery.

Co-design Centers target crosscutting algorithmic methods that capture the most common patterns of computation and communication (known as motifs) in the ECP applications. The goal of the co-design activity is to integrate the rapidly developing software stack with emerging hardware technologies while developing software components that embody the most common application motifs.

The focus of the **National Security Applications** is to deliver comprehensive science-based computational weapons applications able to provide, through effective exploitation of exascale HPC technologies, breakthrough modeling and simulation solutions that yield high-confidence insights into at least three currently infeasible problems of interest to the NNSA Stockpile Stewardship Program (SSP).

Exascale will enable scientists to boost earth system models to enable the reliable assessment of regional agricultural and water cycles.

Application Portfolio

Chemistry and Materials Applications

- **LatticeQCD:** Exascale Lattice Gauge Theory Opportunities/Requirements for Nuclear and High-Energy Physics
- **NWChemEx:** Tackling Chemical, Materials, and Biomolecular Challenges in Exascale
- **GAMESS:** Enabling GAMESS for Exascale Computing in Chemistry and Materials
- **EXAALT:** Molecular Dynamics at the Exascale
- **ExaAM:** Transforming Additive Manufacturing through Exascale Simulation (ExaAM)
- **QMCPACK:** Predictive and Improvable Quantum-mechanics-Based Simulations

LatticeQCD:

Validate Fundamental Laws of Nature

Objective: Validate Fundamental Laws of Nature

Physical light quark masses; properties of light nuclei from first principles; <1% uncertainty in simple quantities

Lead: Fermi National Accelerator Laboratory (FNAL)

Principal Investigators: Andreas Kronfeld, lead, Fermilab; Carleton DeTar, University of Utah; Norman Christ, Columbia University; Richard Brower, Boston University; Robert Edwards, Jefferson Lab

NWChemEx:

Tackling Chemical, Materials, and Biomolecular Challenges in Exascale

Objective: Catalytic Conversion of Biomass-derived Alcohols

Simultaneously address time, length, and accuracy requirements for predictive microstructural evolution of chemicals and materials

Lead: Pacific Northwest National Laboratory (PNNL)

Principal Investigator: Thom Dunning Jr., Pacific Northwest National Laboratory

GAMESS:

General Atomic and Molecular Electronic Structure System

Objective: Biofuel Catalyst Design

Design more robust and selective catalysts orders of magnitude more efficient at temperatures hundreds of degrees lower

Lead: Ames Laboratory

Principal Investigator: Mark Gordon, Ames Laboratory

EXAALT:

Molecular Dynamics at Exascale

Objective: Simultaneously address time, length, and accuracy requirements for predictive microstructural evolution of materials

Enable atomistic simulations to assist in the development of novel materials for energy applications.

Lead: Los Alamos National Laboratory (LANL)

Principal Investigators: Danny Perez, Los Alamos National Laboratory; Anders Niklasson, Los Alamos National Laboratory; Steve Plimpton, Sandia National Laboratories

ExaAM:

Transforming Additive Manufacturing through Exascale Simulation

Objective: Additive Manufacturing of Qualifiable Metal Parts

Accelerate the widespread adoption of Additive Manufacturing by enabling routine fabrication of qualifiable metal parts

Lead: Oak Ridge National Laboratory (ORNL)

Principal Investigators: John Turner, lead, Oak Ridge National Laboratory; Jim Belak, Lawrence Livermore National Laboratory

QMCPACK:

Quantum Mechanics at Exascale

Objective: Find, predict, and control materials and properties at quantum level

Design and optimize next-generation materials from first principles with predictive accuracy

Lead: Oak Ridge National Laboratory (ORNL)

Principal Investigator: Paul Kent, Oak Ridge National Laboratory

Nuclear Physics

LatticeQCD: Exascale Lattice Gauge Theory Opportunities and Requirements for Nuclear and High-Energy Physics

Strong interactions between quarks and gluons represent 99% of the mass in the visible universe. Understanding these interactions and the phenomena that result from them is the central goal of nuclear physics. Over the past three decades, Quantum Chromodynamics (QCD) computations have been a driver of, and benefited from, the spectacular advances in HPC. Computing at the exascale is essential to reach two decadal challenges of central importance to nuclear and high-energy physics. The LatticeQCD project is implementing scalable QCD algorithms to realistically simulate the atomic nucleus to reveal a deeper understanding of the fundamental organization of matter at the subatomic level. These calculations will help us understand the fundamental interactions and nature of matter beyond "elementary" particles.

Atomic nuclei and most particles produced by high-energy accelerators are tightly bound composites of quarks and gluons. The fundamental interaction of these quarks and gluons is known as the strong (nuclear) force—one of the four fundamental forces of nature (i.e., strong, weak, electromagnetic, gravity). These nuclear interactions are explained with mathematical precision by QCD, and HPC is required to predict the consequences of this underlying theory. The properties of the resulting bound states and the nature of their strong, highly nonlinear interactions are the central focus of nuclear physics and an important context in which high-energy physics research must be conducted.

The couplings between the quarks and the W, Z, and Higgs bosons lie at the heart of the Standard Model of particle physics and can be studied, often with exquisite precision, by measuring the properties of the bound states formed from these quarks and gluons. Recent simulations of QCD on the previous generation of massively parallel computers have enabled a comparably precise theoretical understanding of these fundamental interactions of quarks and gluons.

Advances in exascale capability expected over the next decade offer to extend these exciting opportunities to even more groundbreaking discoveries in high-energy and nuclear physics. Exascale computing has the potential to realistically both simulate the atomic nucleus and discover the first harbingers of new laws of nature, revealing a deeper theory that underlies the present "elementary" particles. These possibilities can be achieved only if new and impending advances in computer science can be harnessed to provide a software framework that allows lattice QCD code to efficiently exploit exascale architectures, enabling application scientists to create and refine that code as new challenges and ideas emerge.

The challenge problem consists of six computations representative of three of the common fermion actions in current use by the worldwide lattice-QCD community. Each of these actions has specific advantages for different physical problems in nuclear and high-energy physics.

HISQ: The benchmark problem measures two rates—first, the rate of generating a new gauge configuration using a molecular dynamics (MD) algorithm for a Markov chain Monte Carlo, and second, the rate of making a representative set of "measurements" on the gauge–field configuration. Both steps are required for any campaign to calculate quantities of scientific importance.

DWF: As with the HISQ action, two figures of merit have been adopted for the DWF component of the application. The first measures the rate at which a current state-of-the-art gauge–field ensemble can be generated, and the second calculates a suite of observables using this ensemble.

Wilson-clover: The Clover benchmark has two components. The first is the rate at which dynamical Clover fermion lattices can be generated using a molecular dynamics (MD) algorithm. Several solutions of the Dirac equation are computed and contracted to construct observables as part of the second component of the benchmark.

PI: Andreas Kronfeld, Fermilab

Collaborators: Fermilab, Brookhaven National Laboratory, Jefferson Lab, Argonne National Laboratory, Boston University, Columbia University, State University of New York at Stony Brook, College of William and Mary, Indiana University, University of Illinois at Urbana-Champaign, University of Utah

Progress to date

- Incorporated new Wilson-clover adaptive multigrid solver and force gradient integrator and achieved 100× reduction in GPU hours to generate gauge configurations on Summit over Titan.
- Implemented graph algorithms to perform Wick contractions in light nuclear matrix elements that yield greater than a 4× reduction in time and 10× reduction in memory.
- Implemented a new, fast eigensolver that combines Chebyschev preconditional with Block Lanczos and split grids that achieves a 3× speedup over previously used, implicitly restarted Lanczos iteration.

LatticeQCD's development of scalable algorithms for exascale systems will enable detailed physics investigations at the subatomic level that will lead to fundamental advances in our knowledge of the interactions of matter.

Chemistry and Biochemistry

NWChemEx: Tackling Chemical, Materials, and Biomolecular Challenges in Exascale

A major goal of DOE's advanced biofuels program is to develop fuels that can be distributed using the existing infrastructure and replace existing fuels on a gallon-for-gallon basis. However, producing high-quality biofuels in a sustainable and economically competitive way is technically challenging, especially in a changing global climate. The design of feedstock for the efficient production of biomass and the design of new catalysts for the efficient conversion of biomass-derived intermediates into biofuels are two major science challenges involved in the development of advanced biofuels.

The NWChemEx project directly addresses a Priority Goal in DOE's 2014–2018 Strategic Plan, namely, developing high-performance computational "models demonstrating that biomass can be a viable, sustainable feedstock" for the production of biofuels and other bioproducts. In addition to providing the means to resolve these biofuel challenge problems, NWChemEx will enable exascale computers to be applied toward the development of new materials for solar energy conversion and next-generation batteries, simulation of chemical processes in combustion, prediction of transport and sequestration of energy by-products in the environment, and design of new functional materials.

The NWChemEx project is redesigning and reimplementing NWChem for pre-exascale and exascale computers. NWChemEx is based on NWChem, an open-source, high-performance parallel computational chemistry code funded by the DOE Biological and Environmental Research (BER) program that provides a broad range of capabilities for modeling molecular systems. NWChemEx will support a broad range of chemistry research important to DOE BER and Basic Energy Sciences (BES) on computing systems ranging from terascale workstations and petascale servers to exascale computers.

In Phase I (FY17–20), the NWChemEx project will develop high-performance, scalable implementations of three major physical models: (1) Hartree-Fock (HF) and Density Functional Theory (DFT) methods. These methods are the foundations for the physical models to be incorporated in the NWChemEx framework. Their implementation must be significantly revised to simulate the large molecular systems in the targeted science challenges on exascale computers.

(2) Coupled Cluster methods. A suite of canonical and reduced-scaling coupled cluster (CC) methods will be implemented in NWChemEx. These methods are the gold standard in electronic structure theory and provide the level of fidelity required to address the targeted science challenges.

(3) Density Functional Embedding Theory. Embedding techniques provide a natural and mathematically sound basis for seamlessly integrating subsystems with different electronic structure representations, enabling the active site of interest to be described with high-accuracy CC methods while using a lower fidelity method to describe the impact of the environment on the molecular processes in the active site.

In Phase II (FY21–23), the NWChemEx project will extend NWChemEx to enable it to model the elementary chemical processes involved in the catalytic conversion of propanol to propene in the H-ZSM-5 zeolite. These new capabilities will focus on developing the physical models and algorithms needed to characterize the reactants, products, intermediates, and transition states involved in the conversion process. Two benchmarks have been established for NWChemEx, one at the end of each of the two phases.

Phase I. To illustrate the performance of NWChemEx on biomolecular systems, the ubiquitin molecule was selected as the benchmark for assessing the performance of NWChemEx on ORNL's pre-exascale Summit computer. Ubiquitin is a protein molecule typical of many biomolecular molecules with an abundance of experimental data available from it as well as its fragments. Although it will not be feasible to run canonical coupled cluster calculations on ubiquitin, a 1,231-atom molecule, it will be possible to run reduced-scaling, CC calculations on it. The availability of both implementations along with the sequence of ubiquitin fragments will enable any inaccuracies in the reduced-scaling method to be identified and corrected.

Phase II. A number of elementary chemical transformations have been postulated for the conversion of propanol to propene in the H-ZSM-5 zeolite (basic unit cell: $Si_{96}O_{192}$) based on low-level calculations combined with the available experimental data. To illustrate the performance of the Phase I and II capabilities in NWChemEx on the two exascale computers to be deployed in FY22–23 at Argonne (Aurora) and ORNL (Frontier), reduced-scaling, CC calculations will be used to redefine the structures and energetics of the postulated elementary steps in the conversion of propanol to propene. Depending on the outcome of these calculations, additional work may be required to more fully characterize the mechanism of this conversion.

PI: Thom Dunning, Pacific Northwest National Laboratory

Collaborators: Pacific Northwest National Laboratory, Brookhaven National Laboratory, Ames Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Virginia Polytechnic Institute and State University

Progress to date

- Designed and implemented a simulation development environment as the framework for initiating and connecting modules. This component includes memoization and caching of results to decrease redundant computations, increase support for scripting, and enable application programming interfaces (APIs) for I/O.
- Implemented the LibChemist module that supports core data associated with quantum chemistry codes (e.g., molecular and basis set information).
- Completed the initial design and implementation on GPUs of the TAMM module for both dense and sparse tensor operations.
- The initial design and implementation of a number of the modules in NWChemEx, e.g., canonical HF, DFT, and CC, have been completed.

NWChemEx is implementing exascale algorithms for computational chemistry that will solve two DOE-recognized grand challenge problems to enable the development of advanced biofuels.

Chemistry

GAMESS: Enabling GAMESS for Exascale Computing in Chemistry and Materials

Heterogeneous catalysis and the design of new catalysts is a grand challenge problem in computational chemistry that will require the capabilities of exascale computing. The GAMESS project is extending methods and algorithms based on chemical fragmentation methods and coupling these with high-fidelity Quantum Chemistry (QC) and Quantum Monte Carlo (QMC) simulations to solve this problem. Through computation on a well-defined representative heterogeneous catalysis problem comprising mesoporous silica nanoparticles, GAMESS will demonstrate the capability to model physical systems requiring chemical interactions that involve many thousands of atoms, indicating a new ability to model complex chemical processes.

To take full advantage of exascale architectures, it is critical that application software be developed that can exploit multiple layers of parallelism and take advantage of emerging low-power architectures that dramatically lower energy and power costs without significant deterioration of time-to-solution. This work will develop ab initio methods for GAMESS based on fragmentation methods that have been shown to scale beyond the petascale combined with QMC. To attain exascale performance, GAMESS will be refactored to take advantage of modern computer hardware and software, and the capabilities of the C++ libcchem code that is codeveloped with GAMESS will be greatly expanded. Concurrently, performance analyses will be conducted for the broad array of electronic structure methods in GAMESS on current and emerging architectures to assess their ability to decrease time-to-solution. The improved codes that are developed will be brought to bear on the heterogeneous catalysis problem, specifically using mesoporous silica nanoparticles (MSNs), requiring thousands of atoms, as a template.

MSNs are highly effective, selective heterogeneous catalysts for a wide variety of important reactions including the production of carbinoalamine, which is a starter material for other structures. MSN selectivity is provided by "gatekeeper" groups that allow only desired reactants A to enter the pore, keeping undesirable species B from entering the pore. The presence of a solvent further complicates the computational problem of simulating the heterogeneous catalysis. Accurate electronic structure calculations are needed to deduce the reaction mechanism(s), including the effects of various solvents, and to subsequently design even more effective catalysts. The narrow pores (3–5 nm) can create a diffusion problem that can prevent product molecules from exiting the pore. Therefore, in addition to elucidating the reaction mechanism, the dynamics of the reaction process should be studied, and a sufficiently realistic cross section of the pore must be included. Small models are commonly used to approximate this type of system, with the expectation that a small model might provide insight into the system it represents. However, a recent computational study of the MSN catalysis of carbinolamine formation demonstrated that small proxy models are inadequate, both qualitatively and quantitatively.

This project involves computing both energetics and dynamics on a model reaction with a representative MSN. An adequate representation of the MSN pore requires thousands of atoms with an appropriate basis set. For example, 5,000 heavy atoms with the aug-cc-pVTZ basis set requires more than 500,000 basis functions, not including the hydrogen atoms, the reacting molecules, and especially the solvent molecules.

The energy surface will be mapped via GAMESS calculations using the EFMO + resolution of identity (RI)-MP2 methodology, with refined calculations using the EFMO+CR-CC coupled cluster approach or GAMESS EFMO + QMC approach for accurate reaction rates. The pore selectivity dynamics will be computed with an MD approach requiring approximately 10,000 energetics-type calculations using the GAMESS+FMO code.

PI: Mark Gordon, Iowa State University

Collaborators: Iowa State University, Ames Laboratory, Oak Ridge National Laboratory, Georgia Institute of Technology, Old Dominion University, EP Analytics, Australian National University

Progress to date

- Developed and optimized the underlying RI-MP2 energy + gradient code for both closed and open shells to GPU architectures.
- Developed a preliminary GAMESS-HF (Hartree-Fock) proxy application based on a slightly modified SIMGMS module that utilizes the SIMINT and OED integral libraries.
- Evaluated the performance of the LIBCCHEM GPU implementations of HF, MP2, and RI-MP2 methods on NVIDIA P100 (Pascal) and V100 (Volta) GPUs using a set of input files for calculations on linear alkenes of increasing size.
- Validated GAMESS (E)FMO methodology for fractional MSN systems and performed weak-scaling studies.

GAMESS will use new methods and algorithms to enable computations of chemical processes involving thousands of atoms that will lead to the solution of a grand challenge problem in computational chemistry using exascale computing.

Molecular Dynamics

EXAALT: Molecular Dynamics at the Exascale

Molecular dynamics (MD) is a cornerstone of computational science. However, MD is prevented from achieving complete scientific success by an inability to simultaneously reach the necessary length and timescales while maintaining sufficient accuracy. While the raw computing power available at the exascale should allow for a dramatic extension of the range of applicability of MD, conventional massively parallel codes suffer from poor strong scalability. In other words, a simple scale-up of current practices would enable only the simulation of much larger systems (i.e., containing billions or trillions of atoms) and would do little to improve current timescales (ns) and accuracy (empirical potentials). As most challenging problems require accessing different regions in the accuracy (A), length (L), and time (T) simulation space (ALT), one of the scientific community's key tools, MD, is in danger of missing out on the exascale revolution. EXAALT strives to fill this gap.

The EXAALT project combines three state-of-the-art codes—LAMMPS, LATTE, and ParSplice—into a unified tool that will leverage exascale platforms efficiently across all three dimensions of the ALT space. The new integrated capability is composed of three software layers. First, a task management layer enables the creation of MD tasks, their management through task queues, and the storage of results in distributed databases. It is used to implement various replica-based accelerated MD techniques, as well as to enable other complex MD workflows. The second layer is a powerful MD engine based on the LAMMPS code. It offers a uniform interface through which various physical models can be accessed. The third layer provides a wide range of physical models from which to derive accurate inter-atomic/molecular forces. In addition to the large number of empirical potentials implemented in LAMMPS, it provides high-performance implementations of quantum MD at the Density Functional Tight Binding level, as well as to Spectral Neighbor Analysis Potentials (SNAP), a set of high-accuracy machine-learned potentials.

The first challenge problem is related to nuclear fission. Nuclear energy based on fission provides about 16% of the world's electricity. However, only 4–6% of the uranium atoms in the primary fuel, UO_2, are burned, leaving behind a vast energy resource and creating a greater-than-necessary nuclear waste problem. One of the primary reasons is material integrity: as the fuel burns, radiation damage and fission gases accumulate, causing swelling of the fuel, pellet–clad interactions, and increased pressure on the clad. Because current burnup levels are predicated on our understanding of how the fuel evolves, improved models of fission gas evolution offer the potential for extracting more energy from the fuels.

Solving this grand challenge will require a significant advance in the scientific community's ability to carry out high-accuracy, electronic structure–driven MD simulations on the timescales needed to observe diffusion of defects. Given the size of these defects, relatively small systems (~100 atoms) are sufficient. However, given the high barriers for U-defect evolution, very long timescales will be required for the defects to move at the temperatures of interest. On petascale platforms, simulation rates are estimated at only 10 ns per day; thus, solving this problem requires development of a new simulation capability for the exascale.

The second challenge problem relates to nuclear fusion. Realizing the promise of fusion as a commercially attractive energy source for the 21st century requires advanced structural materials capable of sustained operation in an extreme environment with high temperatures and high fluxes of helium, hydrogen isotopes, and neutrons. The performance demands on plasma-facing components of future fusion power plants are beyond the capability of current materials. Tungsten will be the divertor material in ITER and is the leading candidate material for future fusion reactors. However, experiments indicate the possibility of substantial surface modification in tungsten exposed to low-energy plasma containing helium. Experiments show that nanostructured fuzz, a nanoporous phase with tendrils on the order of tens of nm in diameter, forms on the surface when the surface temperature is between 1000 and 2000 K and the incident ion energies are between 20 and about 100 eV. Such surface features will impact heat transfer and fuel retention, increase the rates of erosion through both sputtering and dust formation, and embrittle the divertor. These modifications to the microstructure can lead to premature materials failure or may quench the fusion reaction by cooling and destabilizing the plasma. However, at this time, the fundamental mechanisms that lead to fuzz formation are unknown, which makes the development of mitigation strategies very difficult. Solving this problem requires a dramatic extension of the reach of large-size extended-time MD simulations. The project simulates the evolution of a tungsten first wall in conditions typical of fusion reactor operation. The primary target is to simulate a 10^5 atom system with a quantum-trained SNAP potential.

PI: Danny Perez, Los Alamos National Laboratory

Collaborators: Los Alamos National Laboratory, Sandia National Laboratories, University of Tennessee

Progress to date

- Released baseline physical models for both challenge problems (i.e., tungsten for fusion, UO_2 for fission).
- Fission science-at-scale demonstration on homogeneous nodes (~270,000 cores on ALCF/Theta).
- Fusion science-at-scale demonstration on homogeneous nodes (~270,000 cores on ALCF/Theta).

EXAALT is creating an exascale MD application that will answer fundamental material problems for energy-relevant materials, especially for fission and fusion energy production. Knowledge gained will be used to design accident-tolerant fuels and will aid in the design of the first wall of fusion reactors.

Additive Manufacturing

ExaAM: Transforming Additive Manufacturing through Exascale Simulation

Additive Manufacturing (AM) is revolutionizing manufacturing, allowing construction of complex parts not readily fabricated by traditional techniques. Although there has been significant interest and investment in AM, the fraction of this investment devoted to modeling and simulation is relatively small and not focused on the development of high-fidelity predictive models but instead on reduced-order models for industry use. The Exascale Additive Manufacturing project (ExaAM) represents a unique opportunity to use exascale simulation to enable the design of AM components with location-specific properties and acceleration of performance certification.

ExaAM aims to develop the Integrated Platform for Additive Manufacturing Simulation (or IPAMS), a collection of capabilities that directly incorporate microstructure evolution and the effects of microstructure within AM process simulation. In AM, a geometric description of the part is processed into 2D slices. A feedstock material is melted, and the part is built layer by layer. In metal AM, the feedstock is often in wire or powder form, and the energy source is a laser or electron beam. ExaAM focuses on powder bed processes, where each layer is approximately 50 microns. A part that is 1 cm tall, for example, would require 200 layers, each requiring the spreading of new feedstock powder and one or more passes of the laser or electron beam to sinter and/or melt the powder in appropriate locations.

The physical processes involved in AM are similar to those of welding—a field with decades' worth of experimental, modeling, simulation, and characterization research. Unfortunately, the simulation tools developed for welding and other similar processes, while calibrated and approaching predictive capability, are inadequate for AM processes, as demonstrated by the inability to predict the failure rate for new AM parts, which can be as high as 80%. This is believed to be largely due to the fact that the process–structure–property–performance relationship is traditionally modeled in an uncoupled manner, relying on tabular databases that are unable to adequately capture the implicit, dynamic, nonequilibrium nature of AM processes.

One of the goals of ExaAM is to remove those limitations by coupling high-fidelity mesoscale simulations within continuum process simulations to determine the microstructure and properties using local conditions. Typically, thermomechanical finite element models are employed at the macroscopic part scale; finite volume or finite element models are used at millimeter scales for fluid dynamics and heat transfer to capture the melt pool dynamics and solidification; mesoscale approaches [e.g., discrete elements, cellular automata, kinetic Monte Carlo (MC), or phase field models] are used at the micron scale to simulate melting, solidification, and microstructure formation; and polycrystal plasticity models are used to develop the microscale mechanical property relationships.

ExaAM will develop and deploy a collection of simulation capabilities for performing process-aware performance modeling of additively manufactured parts using locally accurate properties predicted from microstructures that develop based on local processing conditions. ExaAM will demonstrate this capability by simulating the complex bridge structure developed for the 2018 National Institute of Standards and Technology AM-Bench Conference, known as AMB2018-01. The simulation will be performed where measurements were taken (e.g., "cut locations" for transverse and longitudinal scanning electron microscope specimens for microstructure images), and each location will require a four-stage sequence of simulations.

PI: John Turner, Oak Ridge National Laboratory

Collaborators: Oak Ridge National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, National Institute of Standards and Technology, University of Tennessee

Progress to date

- Leveraged other ECP projects to create new exascale-ready AM-specific capabilities—MFEM for polycrystal plasticity (ExaConstit), AMReX for both a new version of Truchas (TruchasPBF) and a new cellular automata capability for microstructure evolution (ExaCA).
- Implemented and demonstrated initial coupled physics capabilities—in-memory DTK-based integration of melt pool thermal fluids (TruchasPBF) + thermomechanics (Diablo), file-based coupling of melt pool thermal fluids + microstructure evolution (ExaCA), and microstructure + micromechanical properties (ExaConstit).
- Ported multiple components to Summit (ExaConstit, MEUMAPPS-SS, TruchasPBF, ExaCA), with others in progress (Diablo).

ExaAM is coupling multiple physics applications to enable an exascale-capable, multiscale, multiphysics simulation toolkit that can be used to accelerate certification of AM processes.

Nuclear Physics

QMCPACK: Predictive and Improvable Quantum Mechanics–Based Simulations

Materials design has progressed from the study of simple bulk properties to targeting collective effects in strongly correlated materials such as magnetic ordering, phase transitions, and quantum coherence. This requires a fundamentally different set of computational tools than have been used in the past. The QMCPACK project is employing QMC methods to attack this problem since they robustly deliver highly accurate calculations of complex materials that do not artificially bias solutions of a given character. Using exascale computing, QMC has the potential to attain a 18× increase in the number of modeled atoms in, as an example, metal oxide systems to fill missing gaps in experimental data and lead to advances in materials and nanoscience.

The ability to computationally design, optimize, or understand the properties of energy-relevant materials is fundamentally contingent on the existence of methods that can accurately, efficiently, and reliably simulate them. Quantum mechanics–based approaches must necessarily serve as a foundational role because only these approaches can describe matter in a truly first-principles (i.e., parameter free) and therefore robust manner. Quantum Monte Carlo (QMC) methods are ideal for the required simulations because they robustly deliver highly accurate calculations of complex materials that do not artificially bias solutions of a given character. Significantly, with increased computer power, the few approximations in these methods can be tested and systematically reduced, which is not possible with other first-principles methods.

The trade-off is that the computational demands of the QMC method are quite large. As an example, the use of petascale computers has allowed calculations of the magnetic exchange in a copper oxide, which is important for understanding the mechanism of high-temperature superconductivity. However, these calculations involved a highly symmetric supercell containing only 56 atoms, whereas a realistic model considering the defects and dopants of actual superconductors would require at least several hundreds of atoms. The 10 year challenge problem is to simulate transitional metal oxide systems of approximately 1,000 atoms to 10 meV statistical accuracy, such as complex oxide heterostructures that host novel quantum phases, using the full concurrency of exascale systems. The additional power and parallelism of exascale QMC will provide the essential predictive and quantitative capability for these and related materials that lie well beyond the capabilities of existing methods. Exascale provides the opportunity for highly impactful and enabling benchmark accuracy calculations on these materials, providing the reference calibration data that are missing from essentially all quantum mechanics–based materials calculations today. This capability will be highly useful across the materials sciences, nanoscience, and physics communities, particularly where experimental data are costly or difficult to obtain.

The challenge problem involves calculating the cohesive energy of a large supercell of nickel oxide (NiO) using QMCPACK and diffusion QMC to an accuracy of 0.010 eV per NiO formula unit at capability scale in a reasonable and scientifically productive amount of wall clock time (e.g., <1 day). The project is expected to yield a solution for a 1,024-atom supercell.

NiO has been selected because it is emblematic of science challenges addressing the complex physics of transition metal oxides. This classic Mott insulator (more accurately, a charge transfer insulator) defies nonempirical predictions by other methods. NiO is also part of the class of materials being studied by a DOE Basic Energy Sciences program–funded Computational Materials Sciences Center. Success addressing the NiO problem will indicate that a high and productive rate of computational work could be achieved for other challenging materials, including those with strong electronic correlations, novel magnetic states, and novel quantum phases.

PI: Paul Kent, Oak Ridge National Laboratory

Collaborators: Oak Ridge National Laboratory, Sandia National Laboratories, Lawrence Livermore National Laboratory, Argonne National Laboratory

Progress to date

- Developed a novel low-scaling implementation of auxiliary field QMC with reduced memory requirements and cubic scaling.
- Implemented a scheme to assemble atomic orbitals based on Gaussian basis sets that led to a 1500× reduction in memory.
- Achieved a 37× improvement on Titan using a prototype port of mainline QMCPACK developed in March 2019 to use OpenMP target/offload functionality on Summit.

QMCPACK is enabling the simulation and characterization of transition metal oxide transition systems containing up to 1,000 atoms that will improve physicists' understanding of these materials and supplement experimental investigations.

Application Portfolio

Energy Applications

- **ExaWind:** Exascale Predictive Wind Plant Flow Physics Modeling
- **Combustion-Pele:** Transforming Combustion Science and Technology with Exascale Simulations
- **ExaSMR:** Coupled Monte Carlo Neutronics and Fluid Flow Simulation of Small Modular Reactors
- **MFIX-Exa:** Performance Prediction of Multiphase Energy Conversion Device
- **WDMApp:** High-Fidelity Whole Device Modeling of Magnetically Confined Fusion Plasmas
- **WarpX:** Exascale Modeling of Advanced Particle Accelerators

ExaWind:

Exascale Predictive Wind Plant Flow Physics Modeling

Objective: Turbine Wind Plant Efficiency

Harden wind plant design and layout against energy loss susceptibility; higher penetration of wind energy

Lead: National Renewable Energy Laboratory (NREL)

Principal Investigators: Michael Sprague, lead, National Renewable Energy Laboratory; Paul Crozier, Sandia National Laboratories; John Turner, Oak Ridge National Laboratory; Shreyas Anathan, National Renewable Energy Laboratory; Robert Moser, University of Texas at Austin

Combustion-PELE:

High-efficiency, Low-emission Combustion Engine Design

Objective: Advance Understanding of Fundamental Turbulence-Chemistry Interactions in Device-relevant Conditions

Lead: Sandia National Laboratories (SNL)

Principal Investigators: Jacqueline Chen, lead, Sandia National Laboratories; Ray Grout, National Renewable Energy Laboratory; John Bell, Lawrence Berkeley National Laboratory; Ramanan Sankaran, Oak Ridge National Laboratory; Stephen Klippenstein, Argonne National Laboratory

ExaSMR:

Coupled Monte Carlo Neutronics and Fluid Flow Simulation of Small Modular Reactors

Objective: Design and Commercialization of Small Modular Reactors

Visual test reactor for advanced designs via experimental-quality simulations of reactor behavior

Lead: Oak Ridge National Laboratory (ORNL)

Principal Investigators: Steven Hamilton, lead, Oak Ridge National Laboratory; Paul Romano, Argonne National Laboratory; Kord Smith, Massachusetts Institute of Technology

MFiX-Exa:

Performance Prediction of Multiphase Energy Conversion Device

Objective: Scale-up of Clean Fossil Fuel Combustion

Commercial-scale demonstration of transformational energy technologies, curbing CO_2 emission at fossil fuel power plants by 2030

Lead: National Energy Technology Laboratory (NETL)

Principal Investigators: Madhava Syamlal, lead, National Energy Technology Laboratory; Jordan Musser, National Energy Technology Laboratory; Ann Almgren, Lawrence Berkeley National Laboratory; John Bell, Lawrence Berkeley National Laboratory; Christine Hrenya, University of Colorado Boulder; Thomas Hauser, University of Colorado Boulder

WDMApp:

High-fidelity Whole Device Modeling of Magnetically Confined Fusion Plasmas

Objective: High-fidelity Whole Device Modeling of Magnetically Confined Fusion Plasmas

Prepare for the International Thermonuclear Experimental Reactor (ITER) experiments and increase return of investment (ROI) of validation data and understanding; prepare for beyond-ITER devices

Lead: Princeton Plasma Physics Laboratory (PPPL)

Principal Investigators: Amitava Bhattacharjee, lead, Princeton Plasma Physics Laboratory; Chong-Seock Chang, Princeton Plasma Physics Laboratory; Andrew Siegel, Argonne National Laboratory

WarpX:

Exascale Modeling of Advanced Particle Accelerators

Objective: Plasma Wakefield Accelerator Design

Virtual design of 100-stage 1 TeV control; dramatically cut accelerator size and design cost

Lead: Lawrence Berkeley National Laboratory (LBNL)

Principal Investigators: Jean-Luc Vay, lead, Lawrence Berkeley National Laboratory; Ann Almgren, Lawrence Berkeley National Laboratory

Wind Power

ExaWind: Exascale Predictive Wind Plant Flow Physics Modeling

A key challenge to wide-scale deployment of wind energy in the utility grid without subsidies is predicting and minimizing plant-level energy losses, which are currently estimated to be 20% in relatively flat areas and much higher in regions of complex terrain. Current methods for modeling wind plant performance fall far short due to insufficient model fidelity and inadequate treatment of key phenomena, combined with a lack of computational power necessary to address the wide range of relevant length scales associated with wind plants. Thus, the ExaWind challenge problem is a predictive simulation of a wind farm with tens of megawatt-scale wind turbines dispersed over an area of 50 km^2, with complex terrain, involving simulations with O(100) billion grid points. These predictive, physics-based, high-fidelity computational models, validated with targeted experiments, will drive innovation in the blade, turbine, and wind plant design processes by providing a validated "ground truth" foundation for new turbine design models, wind plant siting, operational controls, and reliably integrating wind energy into the grid.

The scientific goal of the ExaWind project is to advance our fundamental understanding of the flow physics governing whole wind plant performance, including wake formation, complex terrain impacts, and turbine–turbine interaction effects. Greater use of the nation's abundant wind resources for electric power generation—reaching 30% of US electrical supply—will have profound societal and economic impacts: strengthening US energy security through greater diversity in its energy supply, providing cost-competitive electricity to key regions across the country, reducing greenhouse gas emissions, and reducing water used in thermoelectric power generation.

This multidisciplinary project embodies a systematic development of the modeling capability and computational performance and scalability required for effective exascale simulations. The project plan builds progressively from predictive petascale simulations of a single turbine, for which the detailed blade geometry is resolved, meshes rotate and deform with blade and tower motions, and atmospheric turbulence is realistically modeled, to a multi-turbine array in complex terrain.
This new modeling and simulation capability will establish a virtual wind plant test bed that will revolutionize the design and control of wind farms and result in a significant advance in the scientific community's ability to predict the response of wind farms to a wide range of atmospheric conditions.

The ExaWind challenge problem is a predictive simulation of a wind farm with tens of megawatt-scale wind turbines dispersed over an area of 50 km^2. The project goal is to capture crucial phenomena that are under-resolved in today's models, including wake formation, complex-terrain impacts, wake–atmosphere interaction, turbine–turbine interaction, and blade boundary layer dynamics. This target requires a modeling and simulation capability that resolves turbine geometry and utilizes adequate grid resolution (down to micron scales within the blade boundary layers). The resolution must capture the upstream chord-scale atmospheric turbulent eddies, generation of near-blade vorticity, and propagation and breakdown of this vorticity within the turbine wake to a distance of many rotor diameters downstream. This application uses the Nalu-Wind computional fluid dynamics (CFD) code and the OpenFAST turbine simulation code, both of which have been specifically designed for wind turbine and wind farm simulations. The simulation will require a hybrid Reynolds-averaged-Navier-Stokes/large-eddy-simulation (RANS/LES) turbulence model, fluid–structure interaction, and atmospheric turbulent flow.

The simulation will contain at least nine megawatt-scale turbines (e.g., NREL 5 MW reference turbines) organized in a 3×3 array and residing in a 4 km×4 km domain with a height of at least 1 km. A hybrid-RANS/LES model will be employed for which an unsteady RANS model will be used near turbine surfaces and an LES model will be used in the wake region. The simulation will have a mean wind speed at the turbines' rated speed (e.g., 11.4 m/s for the NREL 5 MW reference turbine). The model will require at least 30B grid points (and 150B degrees of freedom) to resolve the system, and near-blade grid spacing will be such that the viscous sublayer (within the RANS region) is resolved. A successful simulation will require an optimized solver stack that minimizes time per time step. A scientifically meaningful simulation duration will be for at least one domain transit time (~500 s for the 4 km×4 km domain at 11.4 m/s). The project will demonstrate that such a simulation is feasible within 4 weeks of system time. The simulation described here will require at least 150,000 time steps, which requires that the average time per time step must be no greater than 16 seconds per time step.

PI: Mike Sprague, National Renewable Energy Laboratory

Collaborators: National Renewable Energy Laboratory, Sandia National Laboratories, Oak Ridge National Laboratory, University of Texas, Parallel Geometric Algorithms LLC

Progress to date

- Completed transition of key low-Mach-number-CFD computational physics kernels to a highly portable, high-performance Kokkos-based design paradigm.
- Reduced cost of the pressure system solve to 17% of the total simulation time with algorithmic optimizations. In early ExaWind wind turbine simulations with the Nalu-Wind CFD solver stack, about 80% of the wall-clock simulation time was dedicated to the setup and solve of the momentum and pressure linear systems.
- In collaboration with the DOE Wind Energy Technologies Office High-Fidelity Modeling project, implemented and tested a full set of baseline physics models in the Nalu-Wind/OpenFAST modeling and simulation environment, including hybrid-RANS/LES models and fluid–structure interaction coupling. A demonstration calculation of a single NREL 5 MW turbine was performed for which inflow conditions were derived from an atmospheric boundary layer precursor simulation and overset meshes were used for individual blades, nacelle, and tower. The full set of turbine motions from OpenFAST were coupled to the Nalu-Wind fluid mesh, including blade pitch, variable rotation rate, blade deformation, and tower bending.

Exascale computing will help drive innovation in the design of wind farms resulting in increased efficiency and reduced cost per MW-hour of energy production.

Combustion

Combustion-Pele: Transforming Combustion Science and Technology with Exascale Simulations

Aggressive national goals for significantly reducing petroleum use and greenhouse gas emissions require major improvements in all aspects of our nation's energy use. Combustion processes have historically dominated electrical power production and transportation systems. Despite major advances in improving the efficiency and reducing the costs of alternative energy sources, combustion-based systems are projected to dominate the marketplace for decades. Consequently, these systems need to be optimized for energy efficiency and reduced emissions.

The motivating problem underlying this project is a sufficiently realistic simulation of the in-cylinder processes in an internal combustion engine using low-temperature combustion, for which reactivity-controlled compression ignition (RCCI) is the exemplar. The enabled exascale-era simulations will address key scientific questions regarding mixture formation effects, multistage ignition of a diesel surrogate fuel, lifted flame stabilization, jet reentrainment affected by cylinder–wall geometry, and emissions. The simulation will account for isentropic compression, subsequent injection of the high-reactivity fuel, and combustion processes in a compression ignition engine. Necessary physics include gas compression and models of fuel injection process; spray vaporization (i.e., injection of liquid fuel sprays into high-pressure conditions); and mixing. Combustion processes include autoignition, flame propagation, and soot/thermal radiation, all in a nontrivial engine geometry. The scenario involves kinetically controlled processes in turbulent combustion including ignition, extinction, and emissions. The application used for this project, Pele, implements a hybrid large-eddy simulation (LES)/direct numerical simulation (DNS) approach in both the compressible and low-Mach limits where, using the machinery of adaptive mesh refinement (AMR), the project team will refine the mesh to the DNS limit where necessary to capture turbulence/chemistry interactions, while restricting grid resolution to that required for a high-fidelity LES model far from the flame.

This project is structured around providing a combination of first-principles DNS and near-first-principles DNS/LES hybrid simulations to advance the scientific community's understanding of fundamental turbulence–chemistry interactions in device-relevant conditions. The exascale motivating problem is to perform high-fidelity simulations of the relevant processes in an RCCI internal combustion engine. The relevant processes include turbulence, mixing, spray vaporization, low-temperature ignition, flame propagation, and soot/radiation. RCCI is thermodynamically favorable relative to existing engines and hence holds the promise of groundbreaking efficiencies while operating in a regime that limits pollutant formation. The roadmap toward this exascale-era motivating problem includes simulations of a multi-injection low-temperature diesel jet into an open domain with a large alkane fuel undergoing two-stage ignition processes, dilute spray evaporation and mixing, and multi-injection with fuels of varying reactivity in a geometry that influences the mixing field. The multi-injection simulation forms the challenge problem to demonstrate new exascale capability.

The specific science-based challenge problem is derived from the roadmap toward the motivating exascale-era problem. Specifically, the challenge problem demonstrates the ability to simulate the interaction of two fuels with varying reactivity under a multipulse injection strategy into an engine-relevant geometry. It will serve as a baseline for a series of simulations that will enable isolating the impacts of effects such as spray evaporation on mixture fraction and temperature, alternative fuels, and design of strategies to control combustion phasing and subsequent combustion rate. The problem will be tractable under a realistic allocation using the full capabilities of an exascale machine.

PI: Jacqueline Chen, Sandia National Laboratories

Collaborators: Sandia National Laboratories, National Renewable Energy Laboratory, Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Argonne National Laboratory, Massachusetts Institute of Technology, University of Connecticut

Progress to date

- Extended Pele with tiling approach for data-parallel shared memory parallelism.
- Extended Pele with embedded boundary geometry capability.
- Extended Pele with non-ideal equation of state capability.
- Ported Pele compressible hydrodynamics to GPUs using OpenACC and CUDA.
- Completed demonstration calculations of multi-injection diesel ignition using tiled Pele.

Exascale computing is being used to investigate and validate new low-emission, high-efficient combustion engine designs.

Nuclear Energy

ExaSMR: Coupled Monte Carlo Neutronics and Fluid Flow Simulation of Small Modular Reactors

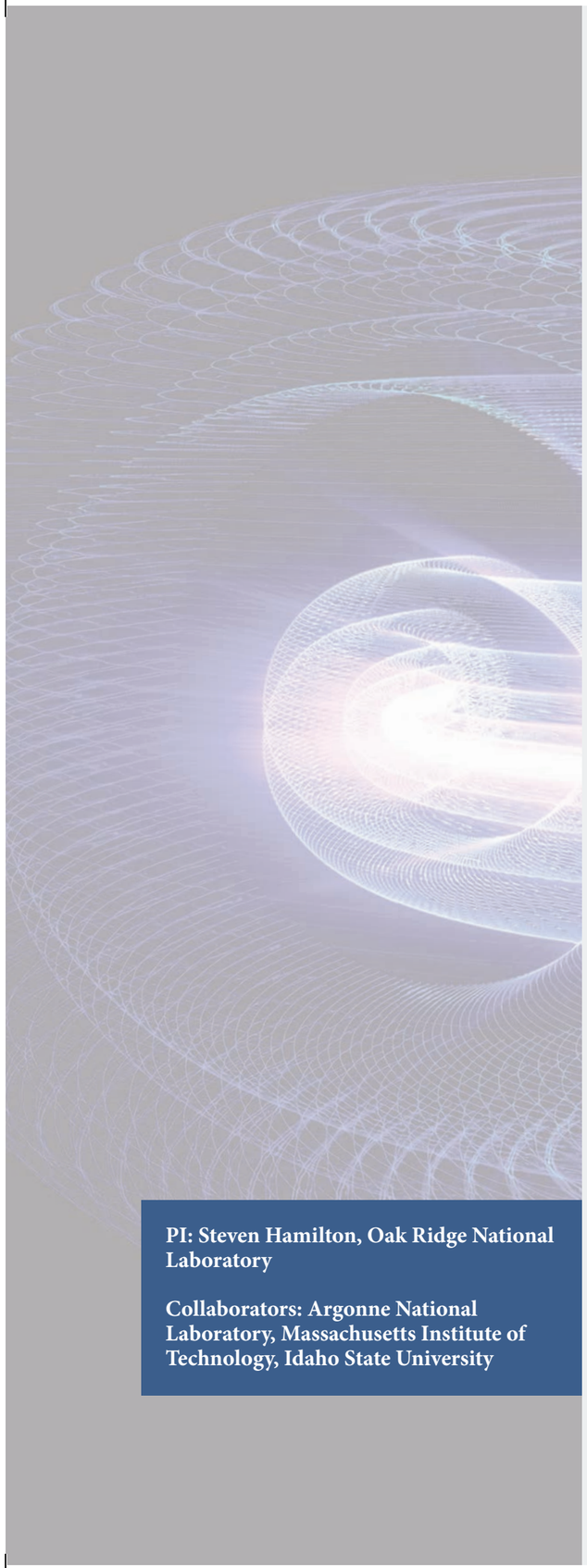
Small modular reactors (SMRs) and advanced reactor concepts (ARCs) will deliver clean, flexible, reliable, and affordable electricity while avoiding the traditional limitations of large nuclear reactor designs, including high capital costs and long construction timelines. Current advanced reactor design approaches leverage decades of experimental and operational experience with the US nuclear fleet and are informed by calibrated numerical models of reactor phenomena. The exascale SMR (ExaSMR) project generates virtual reactor design simulation datasets with high-fidelity, coupled physics models for reactor phenomena that are truly predictive, reflecting as much "ground truth" as experimental and operational reactor data. The ExaSMR virtual designs can accelerate the currently cumbersome advanced reactor concept-to-design-to-build cycle that has constrained the nuclear energy industry for decades. ExaSMR can also provide an avenue for validating existing industry design and regulatory tools.

ExaSMR integrates the most reliable and high-confidence numerical methods for modeling operational reactors, namely, the reactor's neutron state with Monte Carlo (MC) neutronics and the reactor's thermal fluid heat transfer efficiency with high-resolution computational fluid dynamics (CFD)—and all for efficient execution on exascale systems. ExaSMR builds on a base of simulation applications that have demonstrated high efficiency on current petascale-class leadership computing systems. The ExaSMR effort also provides value to US nuclear fuel providers and the broader nuclear community through the generation of highly detailed virtual datasets of operational and advanced concept nuclear reactors.

ExaSMR's exascale challenge problem will open the door to high-confidence prediction of advanced reactor conditions, such as during low-power conditions at startup via the initiation of natural circulation of the coolant flow through a small reactor core and its primary heat exchanger. The exascale software orchestrating this simulation, known as ENRICO, ensures intimate coupling of CFD (Nek5000) and MC neutron transport modules through a common interface that supports multiple exascale simulation technologies: one targeting the exascale Frontier architecture at ORNL (Shift) and another targeting the exascale Aurora system at Argonne National Laboratory (OpenMC).

Exascale neutron transport simulations for ExaSMR will accommodate a full-core SMR model, which typically has ~40 fuel assemblies (each with ~300 fuel rods). The MC portion of the simulation will orchestrate 10B particles per eigenvalue iteration with pin-resolved reaction rates having 3 radial tally regions and 20 axial levels and approximately 150 nuclides and 8 reactions per nuclide in each tally region. These calculations are expected to demonstrate unprecedented accuracy.

Exascale CFD requirements for ExaSMR will include assembly bundle mesh models with momentum sources from a representative resolved spacer grid and full-core mesh having at least 40M elements and 22B degrees of freedom.



PI: Steven Hamilton, Oak Ridge National Laboratory

Collaborators: Argonne National Laboratory, Massachusetts Institute of Technology, Idaho State University

Progress to date

- Completed novel implementation of the Shift MC neutron transport algorithm on GPUs with nearly linear weak-scaling behavior on Summit and a 60× per-node performance speedup over Titan.
- Developed a GPU-enabled version of the Nek5000 CFD code using OpenACC and optimized computational kernels from the Center for Efficient Exascale Discretizations' libParanumal library for a 20× performance improvement over Titan on Summit.
- A new windowed multipole method that provides on-the-fly temperature dependence in continuous-energy nuclear data implemented in both the OpenMC and Shift applications for use on accelerated hardware and a multiphysics driver capable of performing coupled neutronics and fluid flow simulations.

Exascale reactor modeling capabilities delivered by ExaSMR can help inform the design and licensing of advanced and SMRs with unprecedented resolution by improving the fidelity of the modeling of complex physical phenomena occurring within operating nuclear reactors.

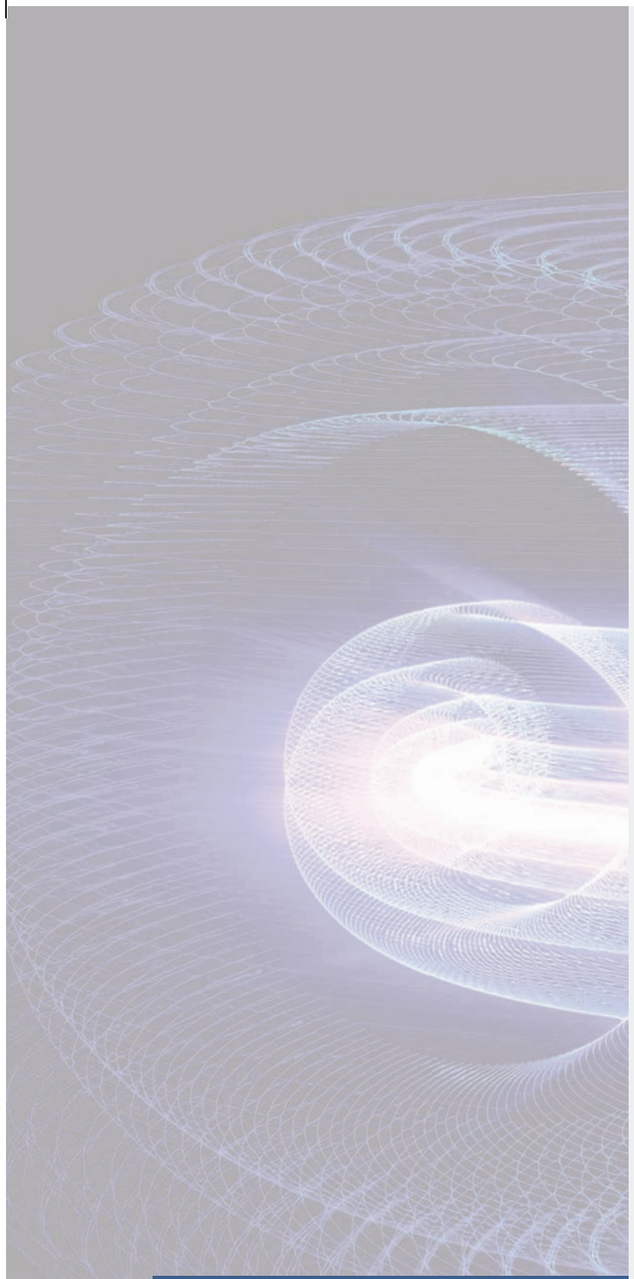
Fossil Energy with Carbon Capture

MFIX-Exa: Performance Prediction of Multiphase Energy Conversion Device

Carbon capture and storage (CCS) technologies such as oxy-fuel combustion, chemical looping combustion, and post-combustion capture systems offer the most promising approaches for reducing CO_2 emissions from fossil fuel power plants. Large-scale commercial deployment of CO_2 capture technologies requires an understanding of how to scale laboratory designs of multiphase flow reactors to industrial sizes. However, the direct scale-up of such reactors is known to be unreliable, and the current approach requires building and testing physical systems at increasingly larger intermediate scales. The cost in both dollars and development time of having to build and extensively test systems at multiple intermediate scales is prohibitive. High-fidelity computational tools that use exascale computing power can be used to model emerging CCS technologies to enable the design and optimization of these systems, which are critical to controlling costs and reducing the risk of designs failing to meet performance standards.

This work specifically targets scale-up of chemical looping reactors (CLRs) through the creation of MFIX-Exa, a scalable computational fluid dynamics–discrete element model (CFD-DEM) code, which is the next generation of the highly successful NETL-based MFIX code. CFD-DEM is an approach that allows for tracking of individual particles (DEM portion) within a continuum fluid phase (i.e., CFD portion). To date, the focus of existing MFIX CFD-DEM efforts has been on validation and development of physical models in the context of a relatively basic computational framework. MFIX-Exa will integrate expertise in HPC directly with expertise in multiphase modeling and will outperform the existing MFIX by orders of magnitude.

The challenge problem requires representing the full-loop CLR geometry, covering various gas-solids flow regimes occurring in the CLR (bubbling bed, riser, cyclone, standpipe, and L-valve), and including chemical reactions and interphase mass, momentum, and energy transfer. Without the capabilities of MFiX-Exa at exascale, it is not possible to resolve the distribution in particle-scale properties (size, density, chemical conversion) in simulations of gas-solids reactors as large as NETL's CLR.



PI: Madhava Syamlal, National Energy Technology Laboratory

Collaborators: National Energy Technology Laboratory, Lawrence Berkeley National Laboratory, University of Colorado – Boulder

Progress to date

- Enabled dynamic load-balancing strategies that account for particle density and a dual-grid approach that allows different domain decompositions for the mesh data and particle data.
- Replaced the SIMPLE algorithm for fluid flow with a modern projection method and explored two spatial discretization alternatives. The new projection formulation demonstrates a factor of four speed-up relative to the SIMPLE method in a weak scaling study from 1 to 4,096 cores.
- Implemented AMReX embedded boundary capability and local mesh refinement at and near solid boundaries and successfully demonstrated coupled fluid-particle flows in basic nonrectangular geometries

Using NETL's 50 kW CLR as an exemplar, MFIX-Exa is developing an exascale application that can be used in the design process of emerging carbon capture reactors to reduce technology development costs and ensure that scaled-up reactors meet performance targets.

Fusion Energy

WDMApp: High-Fidelity Whole Device Modeling of Magnetically Confined Fusion Plasmas

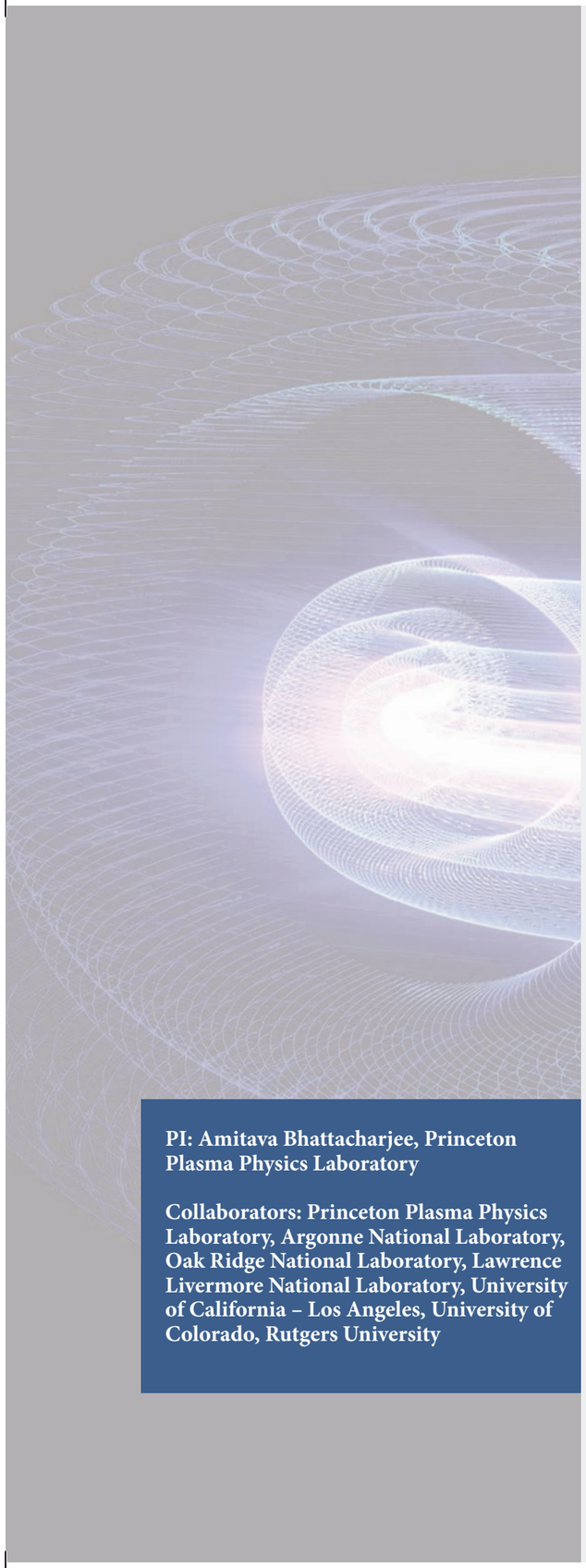
Magnetically confined fusion plasmas are being designed within the International Tokamak Experimental Reactor (ITER) and other projects that will operate in physics regimes never achieved through experiment. Accordingly, modeling and simulation activities that require exascale computational resources are required to design and optimize these new facilities. The WDMApp project is developing a whole device modeling approach that will provide predictive numerical simulations of the physics required for magnetically confined fusion plasmas to enable design optimization and fill in the experimental gaps for ITER and future fusion devices.

The Whole Device Model Application (WDMApp) project aims to develop a high-fidelity model of magnetically confined fusion plasmas, which is urgently needed to plan experiments on ITER and optimize the design of future next-step fusion facilities. These devices will operate in high-fusion-gain physics regimes not achieved by any current or past experiments, making advanced and predictive numerical simulation the best tool for the task. WDMApp is focused on building the main driver and coupling framework for the more complete Whole Device Model (WDM), with the ultimate goal of completing a comprehensive computational suite that includes all the physics components required to simulate a magnetically confined fusion reactor. The main driver for the WDM will be the coupling of two advanced and highly scalable gyrokinetic codes, XGC and GENE. The former is a particle-in-cell (PIC) code optimized for treating the edge plasma; the latter is a continuum code optimized for the core plasma. WDMApp takes advantage of the complementary nature of these two applications to build the most advanced and efficient whole device kinetic transport kernel for the WDM.

A major project thrust is the coupling framework EFFIS 2.0 (End-to-end Framework for Fusion Integrated Simulation 2.0), which will be further developed for operations exascale and optimized for coupling most of the physics modules that will be incorporated in the WDM. The current MPI+X implemented in the main GENE and XGC applications is to be enhanced with communication-avoiding methods, task-based parallelism, in situ analysis with resources for load optimization workflows, and deep memory hierarchy–aware algorithms.

The resulting exascale application will be unique in its computational capabilities and will have potentially transformational impact in fusion science, for example, by studying a much larger and more realistic range of dimensionless plasma parameters than ever before and by assessing the rich spectrum of kinetic microinstabilities that control the quality of energy confinement in a toroidal plasma (e.g., tokamaks, stellarators), with the core and the edge plasma strongly coupled at a fundamental kinetic level based on the gyrokinetic equations.

The exascale science challenge problem is the high-fidelity simulation of whole device burning plasmas applicable to a high-confinement (i.e., H-mode) advanced tokamak regime, specifically, an ITER steady-state plasma that aims to attain a tenfold energy gain. The physics objective is to predict one of the most important indicators for energy confinement in the H-mode—the plasma pressure "pedestal" height and shape. Realization of the H-mode with high-edge plasma pressure and mild pedestal gradient is critical to ITER's performance and success. Efficiency of the fusion burn is virtually determined by the height of the pressure pedestal at the edge. The strategy will involve using WDMApp, which is focused on coupling the continuum code GENE in the core region and the PIC code XGC at the edge.



PI: Amitava Bhattacharjee, Princeton Plasma Physics Laboratory

Collaborators: Princeton Plasma Physics Laboratory, Argonne National Laboratory, Oak Ridge National Laboratory, Lawrence Livermore National Laboratory, University of California – Los Angeles, University of Colorado, Rutgers University

Progress to date

- Benchmarked linear and nonlinear instability calculations of ion temperature gradient–driven instability in cyclone base case between GENE and XGC to remarkable accuracy (~5%).
- Created core-edge coupling algorithms with minimum data movement between the codes that led to 10× performance improvement.
- Collaborated with the Co-Design Center for Particle Applications Codesign Project to develop and use the Cabana particle library for more stable execution of XGC on Summit without performance degradation.

WDMApp will use exascale computing to provide a whole device modeling capability for magnetically confined fusion plasmas that, due to sparse experimental data at proposed operating conditions, is required to design ITER and future fusion power reactors.

Particle Accelerators

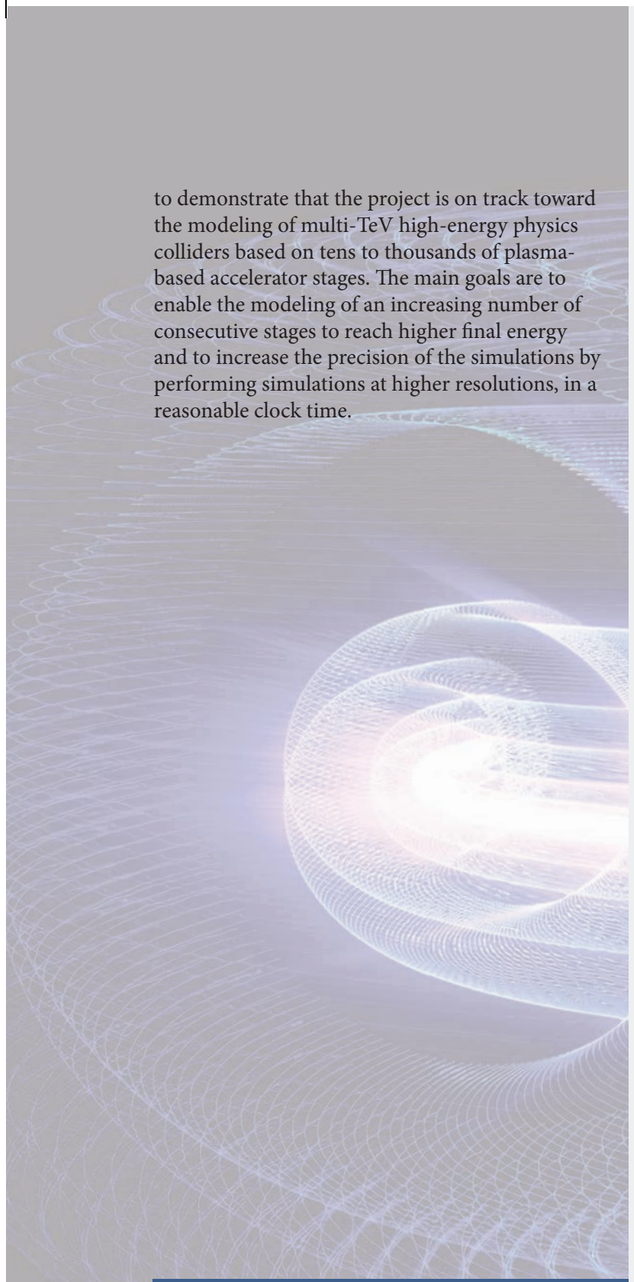
WarpX: Exascale Modeling of Advanced Particle Accelerators

Particle accelerators are used in many areas of fundamental research. A total of 30% of all Nobel prizes in physics since 1939, and four of the last 14 Nobel prizes in chemistry, have been enabled by this technology. Among the candidate new technologies for compact accelerators, the advent of plasma-based particle accelerators stands apart as a prime game-changing technology. The development of these devices depends critically on high-performance, high-fidelity modeling to capture the full complexity of acceleration processes that develop over a large range of space and timescales. WarpX is developing an exascale application for plasma accelerators that enables the exploration of outstanding questions in the physics of the transport and acceleration of particle beams in long chains of plasma channels. These new breeds of virtual experiments, which are not possible with present technologies, will bring huge savings in research costs, leading to the design of a plasma-based collider, and even bigger savings by enabling the characterization of the accelerator before it is built.

For most applications, the size and cost of particle accelerators are limiting factors that can significantly impact the funding of projects or adoption of solutions. Development of plasma-based particle accelerators depends critically on high-performance, high-fidelity modeling to capture the full complexity of acceleration processes that develop over a large range of space and timescales. However, these simulations are extremely computationally intensive, due to the need to resolve the evolution of a driver (laser or particle beam) and an accelerated beam into a structure that is orders of magnitude longer and wider than the accelerated beam. Studies of various effects, including injection, emittance transport, beam loading, tailoring of the plasma channel, and tolerance to nonideal effects (e.g., jitter, asymmetries), that are needed for the design of high-energy colliders, will necessitate a series of tens or hundreds of runs. This will require orders-of-magnitude speed-up over the present state of the art, which will be obtained by combining the power of exascale computing with the most advanced computational techniques.

This project is combining the AMR framework AMReX with novel computational techniques that were pioneered in the PIC code Warp to create a new code (WarpX) and porting the software to exascale platforms. WarpX's team is incorporating the most advanced algorithms in the code, including the optimal Lorentz boosted frame approach, scalable spectral electromagnetic solvers, and mitigation methods for the numerical Cherenkov Instability. To ensure speed and scalability, WarpX is taking advantage of the latest features in portable vectorization algorithms and hierarchical parallelism (on CPUs and GPUs), as well as AMReX's dynamic gridding capabilities, to load-balance the combined computational work associated with both the particles and the mesh. The new software will enable the exploration of outstanding questions in the physics of the transport and acceleration of particle beams in long chains of plasma channels, such as beam-quality preservation, hosing, and beam breakup instabilities.

The exascale challenge problem involves the modeling of a chain of tens of plasma acceleration stages. Realizing such an ambitious target is essential for the longer-range goal of designing a single- or multi-TeV electron–positron high-energy collider based on plasma acceleration technology. The WarpX application uses AMReX for AMR and employs PIC methodology to solve the dynamic Maxwell equations to model the accelerator system. The minimum completion criteria are designed to demonstrate that the project is on track toward the modeling of multi-TeV high-energy physics colliders based on tens to thousands of plasma-based accelerator stages. The main goals are to enable the modeling of an increasing number of consecutive stages to reach higher final energy and to increase the precision of the simulations by performing simulations at higher resolutions, in a reasonable clock time.



PI: Jean-Luc Vay, Lawrence Berkeley National Laboratory

Collaborators: Lawrence Berkeley National Laboratory, SLAC National Accelerator Laboratory, Lawrence Livermore National Laboratory

Progress to date

- Implemented parallel finite-difference 3D electromagnetic PIC solver on GPUs.
- Implemented parallel pseudo-spectral analytical 3D electromagnetic PIC solver on CPUs.
- Implemented and verified AMR (initially with single refined level) using AMReX.

The validation of plasma-based accelerators using exascale modeling through WarpX may lead to the development of tens of thousands of particle accelerators for various applications impacting our lives, from sterilizing food or toxic waste to implanting ions in semiconductors, treating cancer, or developing new drugs.

Application Portfolio

Earth and Space Science Applications

- **ExaStar:** Exascale Models of Stellar Explosions: Quintessential Multi-physics Simulation
- **ExaSky:** Computing the Sky at Extreme Scales
- **EQSIM:** High-Performance, Multidisciplinary Simulations for Regional-Scale Earthquake Hazard/Risk Assessments
- **Subsurface:** Exascale Subsurface Simulator of Coupled Flow, Transport, Reactions, and Mechanics
- **E3SM-MMF:** Cloud-resolving Climate Modeling of the Earth's Water Cycle

ExaStar:

Exascale Models of Stellar Explosions

Objective: Demystify Origin of Chemical Elements

What is the origin of the elements? Behavior of matter at extreme densities? Source of gravity waves?

Lead: Lawrence Berkeley National Laboratory

Principal Investigator: Daniel Kasen, Lawrence Berkeley National Laboratory

ExaSky:

Computing at the Extreme Scales

Objective: Cosmological Probe of the Standard Model of Particle Physics

Unravel key unknowns in the dynamics of the universe: dark energy, dark matter, and inflation

Lead: Argonne National Laboratory

Principal Investigator: Salman Habib, Argonne National Laboratory

EQSIM:

High-Performance, Multidisciplinary Simulations for Regional-Scale Earthquake Hazard/ Risk Assessments

Objective: Earthquake Hazard Risk Assessment

Replace conservative and costly earthquake retrofits with safe purpose-fit retrofits and designs

Lead: Lawrence Berkeley National Laboratory

Principal Investigator: David McCallen, Lawrence Berkeley National Laboratory

Subsurface:

Exascale Subsurface Simulator of Coupled Flow, Transport, Reactions, and Mechanics

Objective: Carbon Capture, Fossil Fuel Extraction, Waste Disposal

Reliably guide safe long-term consequential decisions about storage, sequestration, and exploration

Lead: Lawrence Berkeley National Laboratory

Principal Investigator: Carl Steefel, Lawrence Berkeley National Laboratory

E3SM-MMF:

Cloud-Resolving Climate Modeling of the Earth's Water Cycle

Objective: Accurate Regional Impact Assessment in Earth Systems

Forecast water resources and severe weather with increased confidence; address food supply changes

Lead: Sandia National Laboratories

Principal Investigator: Mark Taylor, Sandia National Laboratories

Astrophysics

Exastar: Exascale Models of Stellar Explosions: Quintessential Multiphysics Simulations

Astronomical observations have confirmed that the production of heavy elements occurred early in galactic history. Yet many details remain outside the purview of direct observation. Exascale computing, through the Exastar Clash code, can be used to address fundamental questions in astrophysics including understanding the origin of elements. The Exastar project is using exascale computing to gain a fuller understanding of where heavy elements are born.

Exastar focuses on developing a new code suite, Clash, which will be a component-based multiphysics adaptive mesh refinement (AMR)–based toolkit that can accurately simulate coupled hydrodynamics, radiation transport, thermonuclear kinetics, and nuclear microphysics for stellar explosion simulations. Clash will reach exascale efficiency by building upon current multicore and many-core efficient local physics packages integrated into a task-based asynchronous execution framework based on current AMR technology. The fundamental goal in the development of Clash is to understand the production of the chemical elements found in these explosions, particularly those heavier than iron. While astronomical observations reveal that the production of the heaviest nuclei began early in galactic history, it is not known how and where these elements were formed. To address this topic via laboratory measurements, a series of nuclear science long-range plans have supported construction of radioactive ion beam facilities, culminating in the Facility for Rare Isotope Beams (FRIB). While FRIB is designed to acquire extensive data on the nuclei relevant for astrophysical nucleosynthesis, its end science goal cannot be met unless those experimental data are integrated into high-fidelity simulations of stellar explosions, such as supernovae and neutron star mergers, that define the conditions under which such heavy element production most likely takes place. Through a better understanding of the sites where the heaviest elements are made, Clash can help focus experimental efforts at FRIB on those reactions of greatest influence.

The Exastar challenge problem is a 3D simulation of the first 2 seconds of evolution after iron-core bounce of core-collapse supernovae (CCSNe). The progenitor star model will be chosen at run time from the best available models. The most likely progenitor models include (1) the solar metallicity 12 solar mass progenitor of Sukhbold et al. (2016), chosen because it represents, in some sense, the "center" of the distribution of massive stars that produce CCSNe, or (2) the binary merger model of Menon and Heger (2017), chosen because it is believed to closely mimic the progenitor system of Podsiadlowski (1992), the only CCSNe from which we have multimessenger signals to date.

The physical domain will extend from the center of the star out to fully enclose the helium shell of the evolved star. The precise location of this radius is progenitor dependent, but it is always more than 10,000 km. The maximum spatial resolution (enabled with AMR) will be at least 1 km at the surface of the proto-neutron star (i.e., in the inner 100 km or so of the event). At least 20 energy groups will be used to resolve the spectra of neutrinos of all flavors (i.e., electron, mu, tau, and their antiparticles) from 0 to 300 MeV. An approximation to general relativistic gravity using at least 12 moments in a multipole approach will be used, with the option to have a more realistic treatment if possible (e.g., conformally flat approximation). A set of tabulated neutrino–matter interaction rates that include emission, absorption, scattering, and pair production from various nuclear and nucleonic processes will be used. This table will be coupled to a set of tabulated quantities derived from a high-density equation of state (EOS) that will provide pressures, entropies, and all other required thermodynamic values (e.g., hydrodynamics). The available set of coupled rates and EOS tables will include, at minimum, the SHF0 EOS of Steiner et al. (2012).

References

A. Menon and A. Heger (2017). " The quest for blue supergiants: binary merger models for the evolution of the progenitor of SN1987A," Monthly Notices of the Royal Astronomical Society, 469(4), 4649–4664, April.

P. Podsiadlowski (1992). "The Progenitor of SN 1987A," Publications of the Astronomical Society of the Pacific, 104(679), 717–729, September.

A. W. Steiner et al. (2012)."Core-collapse supernova equations of state based on neutron star observations," The Astrophysical Journal, 774(1), July. DOI: 10.1088/0004-637X/774/1/17.

T. Sukhbold et al. (2016). "Core-Collapse Supernovae from 9 to 120 Solar Masses Based on Neutrino-Powered Explosions," The Astrophysical Journal, 821, April. DOI: 10.3847/0004-637X/821/1/38.

PI: Daniel Kasen, Lawrence Berkeley National Laboratory

Collaborators: Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Argonne National Laboratory, State University of New York at Stony Brook

Progress to date

- Verified a new neutrino transport module in Clash.
- Completed a GPU implementation of nuclear kinetics and nonpolytropic equation of state.
- Integrated adaptive meshing from AMReX into Clash and performed baseline calculations on Titan.

Using exascale computing to model the physics of stellar explosions, Clash will inform planned and future astronomical experimental observations to answer questions about the origin of heavy nuclides in the universe.

Cosmology

ExaSky: Computing the Sky at Extreme Scales

Modern cosmological observations carried out with large-scale sky surveys are unique probes of fundamental physics. They have led to a remarkably successful model for the dynamics of the universe as well as a number of breakthrough discoveries. Three key ingredients—dark energy, dark matter, and inflation—are signposts to further breakthroughs, as all reach beyond the known boundaries of the Standard Model of particle physics. Sophisticated, large-scale simulations of cosmic structure formation are essential to this scientific enterprise. They not only shed light on some of the deepest puzzles in all of physical science but also rank among the very largest and most scientifically rich simulations run on supercomputers today. The ExaSky project is extending existing cosmological simulation codes to work on exascale platforms in order to address this challenge.

A new generation of sky surveys will provide key insights into questions raised by the current paradigm as well as provide new classes of measurements, such as neutrino masses. They may lead to exciting new discoveries, including that of primordial gravitational waves and modifications of general relativity. Existing machines do not have the performance and the memory needed to run the next-generation simulations that are required to meet the challenge posed by future surveys, whose timelines parallel that of the ECP. The ExaSky project extends the HACC and Nyx cosmological simulation codes so as to efficiently utilize exascale resources as they become available. The Eulerian AMR code Nyx complements the Lagrangian nature of HACC; the two codes are being used to develop a joint program for verification of gravitational evolution, gas dynamics, and subgrid models in cosmological simulations at very high dynamic range.

In order to establish accuracy baselines, there are statistical and systematic error requirements on a large number of cosmological summary statistics. The accuracy requirements are typically scale-dependent, large spatial scales being subject to finite-size effects and small scales being subject to a number of more significant problems such as particle shot noise and code evolution errors (including subgrid modeling biases). Strict accuracy requirements have already been set by the observational requirements for DOE-supported surveys such as the CMB-Stage 4 (CMB-S4), Dark Energy Spectroscopic Instrument (DESI), and the Large Synoptic Survey Telescope (LSST), which typically are sub-percent (statistical) over the range of well-observed scales. Systematic errors need to be characterized, and controlled where possible, to the percent level or better. The final challenge problem runs will be carried out with a new set of subgrid models for gas cooling, UV heating, star formation and supernova and AGN feedback, now under active development.

The simulation sizes are set by the scales of the cosmological surveys. The challenge problem simulations must cover boxes of linear sizes up to the few gigaparsecs (Gpc) scale, with galaxy formation-related physics modeled down to roughly 0.1 kiloparsecs (kpc) (a dynamic range of one part in 10 million, improving the current state of the art by an order of magnitude). Multiple-size boxes will be run to cover the range of scales that need to be robustly predicted. The mass resolution of the simulations (in the smaller boxes) will go down to roughly a million solar masses for the baryon tracer particles and about five times this value for the dark matter particles. The final dynamic range achieved depends on the total memory available on the first-generation exascale systems.

The ExaSky science challenge problem will eventually consist of a small number of very large cosmological simulations run with HACC that simultaneously address many science problems of interest. Setting up the science challenge problem in turn requires multiple simulations—building subgrid models by matching against results from very high-resolution galaxy formation astrophysics codes via a nested-box simulation approach, a medium-scale set for parameter exploration, and, based on these results, designing and implementing the final large-scale challenge problem runs on exascale platforms.

Project simulations are classified in three categories: (1) large-volume, high-mass and force resolution gravity-only simulations, (2) large-volume, high-mass and force resolution hydrodynamic simulations including detailed subgrid modeling, (3) small-volume, very high-mass and medium/high-force resolution hydrodynamic simulations including subgrid modeling. The first set of simulations is targeted at DESI observations of luminous red galaxies (LRGs), emission line galaxies (ELGs), quasars (the simulations are also relevant to the recently approved NASA SPHEREx mission and to modeling the cosmic infrared background for CMB-S4), and for end-to-end simulations for LSST. The second (main) set of simulations will include hydrodynamics and detailed subgrid modeling with the resolution and physics reach improving over time as more powerful machines arrive. The main probes targeted with these simulations are strong and weak lensing shear measurements, galaxy clustering, clusters of galaxies and cross-correlations (internal to this set as well as with CMB probes, such as CMB lensing and thermal and kinematic Sunyaev-Zel'dovich effect observations). A set of smaller volume, hydrodynamic simulations will be carried out in support of the program for convergence testing and verification and for developing and testing a new generation of subgrid models based on results from high-resolution, small effective volume, galaxy formation studies carried out by other groups (high-resolution boxes).

PI: Salman Habib, Argonne National Laboratory

Collaborators: Argonne National Laboratory, Los Alamos National Laboratory, Lawrence Berkeley National Laboratory

Progress to date

- High-performance hybrid N-body gravity solver for cosmological simulations proven at scale on manycore (Cori II, Theta) and large-scale CPU/GPU systems (Cooley, Summit, Titan) in full production mode; algorithms proven for challenge problem.
- New, improved Lagrangian hydrodynamics method (CRK-SPH) integrated into HACC for both manycore and GPU systems and Eulerian cosmological hydrodynamics capability with Nyx run at scale on manycore systems (Cori II, Theta); includes significant work on performance optimization and improved deep AMR capabilities.
- Development of a data reduction capability (lossy compression) for cosmological simulations that reduces storage and IO requirements by factors ranging from ~5 to on the order of ~100.

ExaSKY is enabling large-scale cosmological simulations that, when combined with exascale computing and next-generation sky surveys, will improve our understanding of the large-scale physical processes that drive the universe.

Natural Hazards

EQSIM: High-Performance, Multidisciplinary Simulation for Regional-Scale Earthquake Hazard and Risk Assessments

Large earthquakes present a significant risk around the world and are a major issue across the DOE mission space ranging from the safety of DOE's own inventory of one-of-a-kind mission critical facilities to all major US energy systems. Beyond the DOE enterprise, addressing earthquake risk, both from the standpoint of life safety and damage/economic impact, is a major societal challenge for virtually every element of the built environment including transportation, health, data/commerce, and all urban infrastructure. EQSIM is tapping the tremendous developments occurring in high performance computing, data collection, and data exploitation to help advance earthquake hazard and risk assessments. EQSIM application codes are removing the reliance on simplifying idealizations, approximations, and sparse empirical data; instead, they focus on resolving with the fundamental physics uncertainties in earthquake processes. Through EQSIM, regional-scale ground motion simulations are becoming computationally feasible, and simulation models that connect the domains of seismology and geotechnical and structural engineering are becoming within grasp.

The EQSIM application development project is focused on creating an unprecedented computational toolset and workflow for earthquake hazard and risk assessment. Starting with a set of the existing codes, SW4 (a fourth-order, 3-D seismic wave propagation model), NEVADA (a nonlinear, finite displacement program for building earthquake response), and ESSI (a nonlinear finite-element program for coupled soil-structure interaction), EQSIM is building an end-to-end capability to simulate from the fault rupture to surface ground motions (earthquake hazard) and ultimately to infrastructure response (earthquake risk). The ultimate goal of the EQSIM development is to remove computational limitations as a barrier to scientific exploration and understanding of earthquake phenomenology, as wells as to practical earthquake hazard and risk assessments.

Traditional earthquake hazard and risk assessments for critical facilities have relied on empirically based approaches that use historical earthquake ground motions from many different locations to estimate future earthquake ground motions at a specific site of interest. Given the fact that ground motions for a particular site are strongly influenced by the physics of the specific earthquake processes including the fault rupture mechanics, seismic wave propagation through a heterogeneous medium, and site response at the location of a particular facility, earthquake ground motions are very complex with significant spatial variation in both frequency content and amplitude. The homogenization of many disparate records in traditional empirically based ground motion estimates cannot fully capture the complex site-specificity of ground motion. Over the past decade, interest in utilizing advanced simulations to characterize earthquake ground motions (earthquake hazard) and infrastructure response (earthquake risk) has accelerated significantly. However, the extreme computational demands required to execute hazard and risk simulations at the regional scale have been prohibitive. A fundamental objective of the EQSIM application development project is to advance regional-scale ground motion simulation capabilities from the historical computationally limited frequency range of ~0–2 Hz to the frequency range of interest for a breadth of engineered infrastructure of ~0–10 Hz. A second fundamental objective of this project is to implement an HPC framework and workflow that directly couples earthquake hazard and risk assessments through an end-to-end simulation framework that extends from earthquake rupture to structural response, thereby capturing the complexities of interaction between incident seismic waves and infrastructure systems.

To achieve the overall goals, two fundamental challenges must be addressed. First, regional-scale forward ground motion simulations must be effectively executed at an unprecedented frequency resolution with much larger, much faster models. Achieving fast earthquake simulations is essential to allowing the parametric variations necessary to span critical problem parameters (e.g., multiple fault rupture scenarios). Second, as the ability to compute at higher frequencies progresses, there will be a need for better characterization of subsurface geologic structure at finer and finer scales; thus, a companion schema for representing fine-scale geologic heterogeneities in massive computational models must be developed. For the purpose of evaluating regional-scale simulations and assessing progress on the application developments of this project, a representative large regional-scale model of the San Francisco Bay Area (SFBA) has been created for the domain. This model includes all necessary geophysics modeling features (3D geology, earth surface topography, material attenuation, nonreflecting boundaries, fault rupture models). For a 10 Hz simulation, the computational domain includes approximately 203 billion grid points in the finite difference domain. The SFBA model provides the basis for testing and evaluating both advanced physics algorithms and computational implementations.

PI: David McCallen, Lawrence Berkeley National Laboratory

Collaborators: Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory

Progress to date

- Porting of SW4 to CORI and implementation of optimized loops and a hybrid MPI/OpenMP implementation.
- Successful porting of SW4 to GPU-based platforms including Sierra and Summit; this included utilization of the RAJA libraries to enable platform independence.
- Implementation of a capability to overlay statistically representative stochastic geology variations across a large SW4 geologic model to represent fine-scale geologic variability.

EQSIM will use exascale to help scientists understand the fundamental geologic physics of earthquakes that will inform civil engineering and emergency response on better ways to counter threats from these potential natural disasters.

Subsurface Wellbores

Subsurface: An Exascale Subsurface Simulator of Coupled Flow, Transport, Reactions, and Mechanics

An urgent challenge in the field of subsurface wellbores involves understanding and predicting the behavior of hundreds of thousands of deep wells drilled to locate and extract natural resources. The performance of a wellbore hinges on the behavior of very thin interface features controlling the leakage of fluids along the well casing–cement boundary. Similarly, leakage of buoyant fluids (e.g., CO_2) through caprocks may be controlled by micron-scale asperities in fracture networks that are themselves subject to geomechanical and geochemical modification. At the reservoir or field scale (~1–10 km domain size), multiphase flow and reactions in fractured porous media are typically modeled using continuum models that use averaged quantities and bulk parameters that do not fully take into account thermal, hydrological, chemical, and mechanical-related heterogeneity at different spatial and temporal scales. A more rigorous treatment is to resolve the pore-scale (0.1–10 micron) physical and geochemical heterogeneities in wellbores and fractures so as to improve their ability to predict the evolution of these features when subjected to geomechanical and geochemical stressors. The Subsurface project is using exascale to integrate the complex multiphysics processes occurring at multiple scales, from the micro to the kilometer scale, in a high-resolution reservoir simulator.

A wide range of processes take place in the subsurface that involve the evolution of fractures, including both opening and closing due to some combination of mechanical and chemical stresses. This project focuses on the science challenge of overcoming the failure of a wellbore for CO_2 sequestration in saline reservoirs, with consideration of a wellbore segment of up to 100 m and times up to 1 year. Wells are considered to be high-risk pathways for fluid leakage from geologic CO_2 storage reservoirs because breaches in this engineered system have the potential to connect the reservoir to groundwater resources and the atmosphere. The geologic carbon storage community has raised further concerns about wellbore stability because of acidic fluids in the CO_2 storage reservoir, alkaline cement meant to isolate the reservoir fluids from the overlying strata, and steel casings in wells that are inherently reactive systems. This is of particular concern for the storage of CO_2 in depleted oil and gas reservoirs with numerous legacy wells engineered to variable standards.

In contrast to the conventional treatment of wellbore failure currently modeled at large scales on the order of 100 m to 1 km and 10 years, accurate prediction of fracture evolution depends on microscale resolution of fracture asperities (i.e., pillars) controlling permeability and chemical reactivity. Microscale resolution is also needed to accurately predict fracture permeability because very rough fractures are typically held open by pillars of this scale. Chemical corrosion (i.e., dissolution) or mechanical corrosion (i.e., pressure solution) of these asperities occurs at the same micron scale. The localized subdomain needed to resolve reactive transport processes at microscale resolution during fracture propagation is a domain size up to 10 cm (in the length of the wellbore) × 1 cm (along an azimuth in the cement annulus) × 1mm (in the radial direction) with 1 micron grid resolution. This domain size is assumed to be the minimum domain needed to capture coupled reactive transport and mechanics effects in a fracture (e.g., pillar collapse).

The Subsurface project addresses this exascale computing challenge by coupling two mature code bases: (1) Chombo-Crunch, developed at LBNL, which currently handles Navier-Stokes and Darcy flow coupled to multicomponent geochemical reaction networks, and (2) the GEOSX code, developed at LLNL, which handles geomechanical deformation and fracture+Darcy flow at a variety of scales.

A science challenge problem has been developed that focuses on the evolution of a single fracture in wellbore cement, beginning at Stage 1 with diffusion-controlled reaction and a weakening of the cement that leads to fracturing. The propagation of the fracture as a result of further chemical reaction and fluid pressure–driven deformation is simulated with 1 micron resolution within the fracture and is coupled to a coarser resolution (10 micron) representation of the porous cement adjacent to the evolving fracture. The resulting challenge problem is estimated to require 1 trillion grid cells with 16 trillion degrees of freedom once the hydraulic, mechanical, and chemical variables are included. Based on prior experiments and modeling, the challenge problem is estimated to extend for 10 days of simulation to capture the evolving fracture and associated reaction fronts.

PI: Carl Steefel, Lawrence Berkeley National Laboratory

Collaborators: Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, National Energy Technology Laboratory

Progress to date

- Hybrid Raja (OpenMP) + MPI implementation of Geos completed.
- Coupling of Geos and Chombo-crunch via HDF5 API done.
- Design specification, baseline implementation, and Kernel implementation on GPU done for DSL Proto.

The Subsurface project will use exascale computing to develop coupled applications for resolving fracture problems in subsurface models ranging from 1 micron fracture scales up to 100 m reservoir scales.

Climate

E3SM-MMF: Cloud-Resolving Climate Modeling of Earth's Water Cycle

The impact of climate change on the global and regional water cycles is one of the highest priorities and most difficult challenges in climate change prediction. Current earth system models possess limited ability to model the complex interactions between the large-scale, mostly 2D baroclinic atmospheric motions and the smaller-scale 3D convective motions found in clouds and individual storms. Because full-resolution climate simulation requires zetascale computation, E3SM is employing a superparameterized model to accurately incorporate cloud physics at a resolution previously unobtainable at petascale computing levels. This next-generation model will improve the scientific community's ability to assess regional impacts of climate change on the water cycle that directly affect multiple sectors of the US and global economies.

The goal of the E3SM-MMF project is to develop a cloud-resolving earth system model with throughput necessary for multidecade, coupled high-resolution climate simulations. This next-generation model has the potential to substantially reduce major systematic errors in precipitation found in current models because of its more realistic and explicit treatment of convective storms. These motions and their interactions, to first order, determine the spatial distributions and characteristics of regional precipitation. Complexities include the microscale chemistry and physics of cloud formation and the impacts of anthropogenic climate change on cloud formation. Properly resolving the key processes involved in cloud formation requires resolution (i.e., grid spacing) on the order of 1 km in the atmosphere. Today's petascale computing systems are capable of such resolution but only at great expense and for very short times (i.e., several simulated days). Running conventional climate models at this resolution, for 100 year simulations, requires a 5000× increase in computing resources.

The project will examine a multiscale modeling framework (MMF) approach to cloud resolution modeling. Often referred to as superparameterization, it offers significant opportunities for unprecedented improvement for a model that has yet to be fully explored due to limited computing resources. This project will integrate a cloud-resolving convective parameterization (i.e., superparameterization) into the DOE E3SM Earth System model using MMF and explore its full potential to scientifically and computationally advance climate simulation and prediction. The superparameterization will be designed to make full use of GPU-accelerated systems and will involve refactoring and porting other key components of the E3SM model for GPU systems. The acronym E3SM-MMF is used to refer to the superparameterized version of the E3SM model being developed under this ECP effort.

The challenge problem has several aspects: (1) achieving cloud-resolving resolution in the atmosphere with superparameterization, defined as at least 1 km grid spacing in both horizontal and vertical directions; (2) achieving weather-resolving resolution in the global atmosphere model, defined as 50–25 km average grid spacing in the horizontal directions with ~1 km grid spacing in the vertical directions (i.e., the resolution of today's global operational forecast models); (3) achieving an eddy-resolving ocean/ice model, defined as a minimum 18 km resolution in equatorial regions, decreasing to 6 km in polar regions; and (4) achieving model throughput necessary to perform the simulation campaign for the challenge problem in the course of one calendar year on the Frontier supercomputer.

PI: Mark Taylor, Sandia National Laboratories

Collaborators: Sandia National Laboratories, Lawrence Livermore National Laboratory, Pacific Northwest National Laboratory, Los Alamos National Laboratory, Argonne National Laboratory, Oak Ridge National Laboratory, University of California –Irvine, Colorado State University

Progress to date

- Completed an initial port of superparameterization subcomponent to GPUs.
- Completed an initial port of E3SM-MMF atmosphere/land "F" compsets to Summit.
- Completed (in FY19) the first port of the atmosphere model to Summit, running at 0.6 simulation year per day and achieving a speedup of 31× relative to Titan.

The enhancements to E3SM enabled through the ECP will improve the scientific community's ability to predict, assess, and respond to the challenges imposed by local variations in the water cycle caused by global climate change.

Application Portfolio

Data Analytics and Optimization Applications

- **ExaSGD:** Optimizing Stochastic Grid Dynamics at Exascale
- **CANDLE:** Exascale Deep Learning–Enabled Precision Medicine for Cancer
- **ExaBiome:** Exascale Solutions for Microbiome Analysis
- **ExaFEL:** Data Analytics at the Exascale for Free Electron Lasers

ExaSGD:

Optimizing Stochastic Grid Dynamics at Exascale

Objective: Reliable and Efficient Planning of the Power Grid

Optimize power grid planning, operation, and control and improve reliability and efficiency

Lead: Pacific Northwest National Laboratory

Principal Investigator: Zhenyu (Henry) Huang, Pacific Northwest National Laboratory

CANDLE:

Exascale Deep Learning–Enabled Precision Medicine for Cancer

Objective: Accelerate and Translate Cancer Research

Develop pre-clinical drug response models, predict mechanisms of RAS/RAF driven cancers, and develop treatment strategies

Lead: Argonne National Laboratory

Principal Investigator: Rick Stevens, Argonne National Laboratory

ExaBiome:

Exascale Solutions for Microbiome Analysis

Objective: Metagenomics for Analysis of Biogeochemical Cycles

Discover knowledge useful for environmental remediation and the manufacture of novel chemicals and medicines

Lead: Lawrence Berkeley National Laboratory

Principal Investigator: Katherine Yelick, Lawrence Berkeley National Laboratory

ExaFEL:

Data Analytics at Exascale for Free Electron Lasers

Objective: Light Source–Enabled Analysis of Protein and Molecular Structures and Design

Process data without beam time loss; determine nanoparticle size and shape changes; engineer functional properties in biology and material science

Lead: SLAC National Accelerator Laboratory

Principal Investigator: Amedeo Perazzo, SLAC National Accelerator Laboratory

Power Grids

ExaSGD: Optimizing Stochastic Grid Dynamics at ExaScale

A critical national security challenge is the maintenance of the integrity of national power grids under adverse conditions imposed by natural or man-made causes. When the grid is subject to localized stresses, load imbalances can occur between generation sources and global demand. ExaSGD is developing algorithms that can optimize the grid's response to a large number of potential disruption events to compute a risk profile for grid operations. This application will harness the power of exascale computing to help civil planners assess alternative grid management and response strategies to best maintain the integrity of the national power grid under emergency conditions.

Energy delivery systems such as national power grids operate by maintaining a balance between energy supply and demand. Energy is produced at generators and via renewables and other sources, and it is then transmitted through a bulk power system. Attacks via physical or cyber means and hazards on the grid can create an imbalance between supply and demand, which can result in drops in voltage or frequency, both of which can permanently damage very large and expensive components. As a result, great care is taken to operate the grid with very high reliability within narrow operating voltage and frequency ranges.

Recovering from generation/load imbalance can be achieved by shedding load (i.e., deliberately allowing some load to go unserved, creating a partial blackout) to preserve the functionality of the remainder of the power grid. However, the behavior of the power grid can be influenced at many points within the system because of the increasing prevalence of cyber-enabled control and sensing, renewables (e.g., transient wind or solar power), plug-in storage devices (e.g., electric vehicles that can put power into the grid, or remove it), smart meters that can control load at a fine granularity (e.g., throttling home appliances or air-conditioning at times of peak demand), and other sensored elements that can be controlled remotely. The conventional load-shedding approach may miss more efficient strategies that make use of a more complete spectrum of grid elements. A capability for discovering more optimal configurations to recover from generation/load imbalance will improve our national readiness to recover from a variety of hazards to the power grid.

The ExaSGD challenge problem is to optimize the grid's response in a near-term time frame (e.g., 30 minutes per North American Electric Reliability Corporation operating standards) to a variety of underfrequency hazards via physical and control threat scenarios using comprehensive modeling that includes generation, transmission, load, and cyber/control elements. The ExaSGD team is comparing the frequency recovery performance of a complex grid plus control system in the presence and absence of smart devices, stored energy reserves, renewables, and demand response technologies. This will involve at least two calculations of the distribution of severity of frequency response to grid hazards/effects relevant to national grid response. Estimating these distributions involves solving a large number of optimal power flow calculations that consider different underfrequency scenarios. Each optimal power flow calculation requires the solution to a large-scale nonlinear optimization problem. Additionally, this challenge problem will consider the integrated execution of these optimization problems to warm-start subsequent power flow calculations across scenarios.

PI: Zhenyu (Henry) Huang, Pacific Northwest National Laboratory

Collaborators: Pacific Northwest National Laboratory, National Renewable Energy Laboratory, Argonne National Laboratory, Lawrence Livermore National Laboratory, Oak Ridge National Laboratory

Progress to date

- Integrated GridPACK and xGA.
- Generated a renewable energy scenario model.
- Performed optimization of solvers.

ExaSGD is developing an exascale application that can be used to provide near-real-time responses to both physical and man-made stresses on the national energy grid to maintain electricity generation integrity.

Cancer Research

CANDLE: Exascale Deep Learning–Enabled Precision Medicine for Cancer

The US Department of Energy (DOE) has entered into a partnership with the National Cancer Institute (NCI) of the National Institutes of Health (NIH). This partnership has identified three key science challenges that the combined resources of DOE and NCI can accelerate. The first challenge (called the "drug response problem") is to develop predictive models for drug response that can be used to optimize preclinical drug screening and drive precision medicine–based treatments for cancer patients. The second challenge (called the "RAS pathway problem") is to understand the molecular basis of key protein interactions in the RAS/RAF pathway that is present in 30% of cancers. The third challenge (called the "treatment strategy problem") is to automate the analysis and extraction of information from millions of cancer patient records to determine optimal cancer treatment strategies across a range of patient lifestyles, environmental exposures, cancer types, and health care systems. While these challenges are at different scales and have specific scientific teams collaborating on the data acquisition, data analysis, model formulation, and scientific runs of simulations, they also share several common threads. The CANDLE project focuses on the machine learning aspect of the challenges and in particular builds on a single scalable deep neural network (DNN) code called CANDLE (CANcer Distributed Learning Environment).

The CANDLE challenge problem is to solve large-scale machine learning problems for three cancer-related pilot applications: the drug response problem, the RAS pathway problem, and the treatment strategy problem. For the drug response problem, unsupervised machine learning methods are used to capture the complex, nonlinear relationships between the properties of drugs and the properties of tumors to predict response to treatment, with the goal of developing a model that can provide treatment recommendations for a given tumor. For the RAS pathway problem, multiscale MD runs are guided through a large-scale state-space search using unsupervised learning to determine the scope and scale of the next series of simulations based on the history of previous simulations. For the treatment strategy problem, semi-supervised machine learning is used to automatically read and encode millions of clinical reports into a form that can be computed upon. Each problem requires a different approach to the embedded learning problem, but all approaches are supported with the same scalable deep learning code in CANDLE.

The CANDLE software suite broadly consists of two distinct, interoperating levels: the DNN codes and the Supervisor portion, which handles work distribution across a distributed network. At the DNN level, the CANDLE utility library provides a series of utility functions that streamline the process of writing CANDLE-compliant code. This enables the essential functionality for network hyperparameters to be set either from a default-model file or from the command line. This in turn enables experiments to be designed that sweep across a range of network hyperparameters in an efficient manner. The Supervisor framework provides a set of modules to enable various hyperparameter optimization (HPO) schemes and to automatically distribute the workload across available computing resources. Together, these capabilities allow users to efficiently perform HPO on the large compute resources available across the DOE complex, as well as on any local compute resources.

The challenge for exascale manifests in the need to train large numbers of models. A need inherent to each pilot application requires production of high-resolution models that cover the space of specific predictions (i.e., individualized in the precision medicine sense). Take, for example, training a model that is specific to a certain drug and individual cancer. Starting with 1,000 different cancer cell lines and 1,000 different drugs, a leave-one-out strategy to create a high-resolution model for each drug by cancers requires approximately 1 million models. These models are similar enough that we can use a transfer learning strategy, where weights are shared during training in a way that avoids information leakage, which significantly reduces the time needed to train a large set of models.

PI: Rick Stevens, Argonne National Laboratory

Collaborators: Argonne National Laboratory, Oak Ridge National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Frederick National Laboratory for Cancer Research, National Cancer Institute

Progress to date

- Demonstrated an improved DNN that adds drug target descriptions to the input and improves cell line properties set. This work extends CANDLE's ability to integrate multimodal molecular and drug feature types across multiple data sources in a deep learning framework for drug response.
- Delivered an initial implementation of the CANDLE library to streamline the writing of CANDLE-compliant codes, as well as provide time-saving functionality to improve programmer productivity.
- Delivered a capability to simultaneously extract four critical pieces of information from unstructured text documents, namely, identifying (1) the cancer primary subsite (six classes, three breast cancer subsites, and three lung cancer subsites); (2) histologic grade (three classes); (3) behavioral type (two classes); and (4) laterality (two classes).
- Demonstrated a new multitask deep learning model for document classification with a hierarchical convolutional attention network (MT-HCAN).
- Demonstrated the design of optimal hyperparameters for the MT-HCAN using the CANDLE's mlr-MBO and HyperSpace workflows on Summit (Oak Ridge Leadership Computing Facility).

CANDLE, a partnership between DOE and NCI, is developing highly efficient DNNs optimized for the unique architectures provided by next-generation exascale platforms to address three significant science challenge problems in cancer research.

Bioinformatics

ExaBiome: Exascale Solutions for Microbiome Analysis

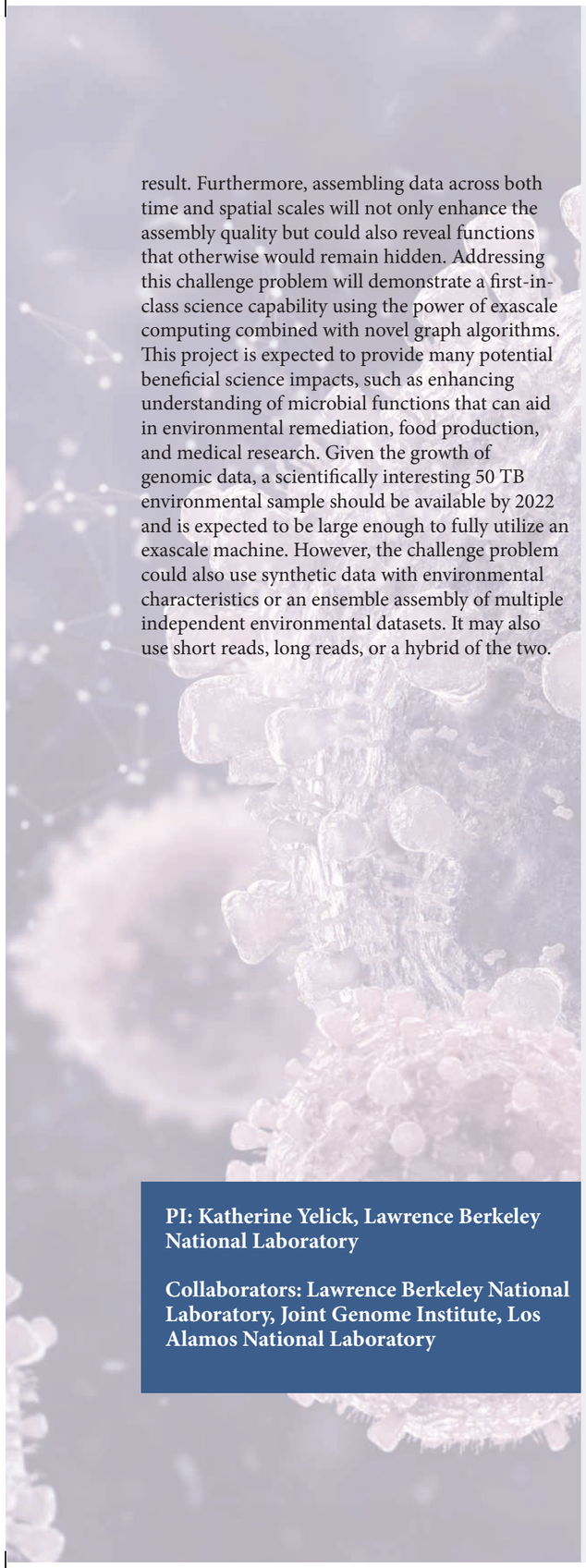
Genome sequencing on DNA extracted from microbiomes is used to study the diversity, integration, and dynamics of organisms in the microbiomes. Due to the size and complexity of the datasets involved, assembly and comparative analysis are the most computationally demanding aspect of this branch of bioinformatics. Furthermore, as more data become available, this cost will only grow. The ExaBiome project is developing scalable data assembly and analysis tools to address current needs and, through the use of exascale computing power, provide solutions for anticipated increases in biological data.

Metagenomics—the application of high-throughput genome sequencing technologies to DNA extracted from microbiomes—is a powerful and general method for studying microbial diversity, integration, and dynamics. Since the introduction of metagenomics over a decade ago, it has become an essential and routine tool. Assembly and comparative analyses of metagenomic datasets are among the most computationally demanding tasks in bioinformatics. The scale and rate of growth of these datasets will require exascale resources to process (i.e., assemble) and interpret through annotation and comparative analysis.

The ExaBiome project aims to provide scalable tools for three core computational problems in metagenomics: (1) metagenome assembly, which takes raw sequence data and produces long genome sequences for each species; (2) protein clustering, which finds families of closely related proteins; and (3) signature-based approaches to enable scalable and efficient comparative metagenome analysis, which may show, for example, variability of an environmental community over time.

The ExaBiome team has developed a scalable metagenome assembler, MetaHipMer, which scales well on thousands of compute nodes on today's petascale architectures and has already assembled large environmental datasets that had not been possible with previous tools. They continue to work on further scalability improvements across nodes and new node-level optimizations to take advantage of fine-grained on-node parallelism and memory structures including GPUs. MetaHipMer exhibits competitive quality with other assemblers, and the team continues to add innovations and parameters to control various aspects of how data are analyzed, driven by the experience of science teams. MetaHipMer is designed for short reads (Illumina) data, but a second assembler for long reads is also under development and shows even higher computational intensity, which may be a good fit for exascale systems. A second ExaBiome code, HipMCL, provides scalable protein clustering. HipMCL runs on thousands of nodes and has already been used to provide insight on the structure of protein families across hundreds of millions of proteins, a dataset that was previously intractable. These codes and comparative analysis tools use some common computational patterns, including dynamic programming for string alignment (either DNA or proteins) with minimal edits, counting and analysis of fixed-length strings (k-mers), and a variety of graph and sparse matrix methods.

ExaBiome's challenge problem is to demonstrate a high-quality assembly or set of assemblies on at least 50 TB of environmental data (reads) that runs across a full-exascale machine. The intent is to use a scientifically interesting environmental sample that may include multiple temporal or spatial samples, which will be processed as a single assembly using complete sequence data. In contrast, current state-of-the-art assembly pipelines are forced to use subsampling when datasets get large, which limits researchers' ability to assemble rare, low-coverage species, and confusing duplications of genomes can result. Furthermore, assembling data across both time and spatial scales will not only enhance the assembly quality but could also reveal functions that otherwise would remain hidden. Addressing this challenge problem will demonstrate a first-in-class science capability using the power of exascale computing combined with novel graph algorithms. This project is expected to provide many potential beneficial science impacts, such as enhancing understanding of microbial functions that can aid in environmental remediation, food production, and medical research. Given the growth of genomic data, a scientifically interesting 50 TB environmental sample should be available by 2022 and is expected to be large enough to fully utilize an exascale machine. However, the challenge problem could also use synthetic data with environmental characteristics or an ensemble assembly of multiple independent environmental datasets. It may also use short reads, long reads, or a hybrid of the two.



PI: Katherine Yelick, Lawrence Berkeley National Laboratory

Collaborators: Lawrence Berkeley National Laboratory, Joint Genome Institute, Los Alamos National Laboratory

Progress to date

- Demonstrated scalable HipMer and MetaHipMer peformance on over 1,000 nodes.
- Implemented overlap/aligner for long reads.
- Completed assembly of a 3 TB large soil data set, the largest metagenome ever assembled. Overall the computation required approximately 4.5 hours on 1,024 Cori KNL nodes, in addition to an hour on a single node (for the memory-intensive scaffolding bottleneck). Memory issues are being addressed by a new scaffolding version.

The ExaBiome project is providing exascale solutions for the assembly and analysis of metagenomic data that will address both current and future data processing needs in bioinformatics.

Data Analytics

ExaFEL: Data Analytics at the Exascale for Free Electron Lasers

The SLAC Linac Coherent Light Source (LCLS) facility uses x-ray diffraction to image individual atoms and molecules to observe fundamental material processes. Near-real-time interpretation of molecular structure revealed by x-ray diffraction will require computational intensities of unprecedented scales coupled with a data path of unprecedented bandwidth. Detector data rates at light sources are advancing exponentially: with the LCLS-II-HE upgrade, LCLS will increase its data throughput by three orders of magnitude by 2025. The objective of the ExaFEL project is to leverage exascale computing to reduce, from weeks to minutes, the time to analyze molecular structure x-ray diffraction data generated by LCLS.

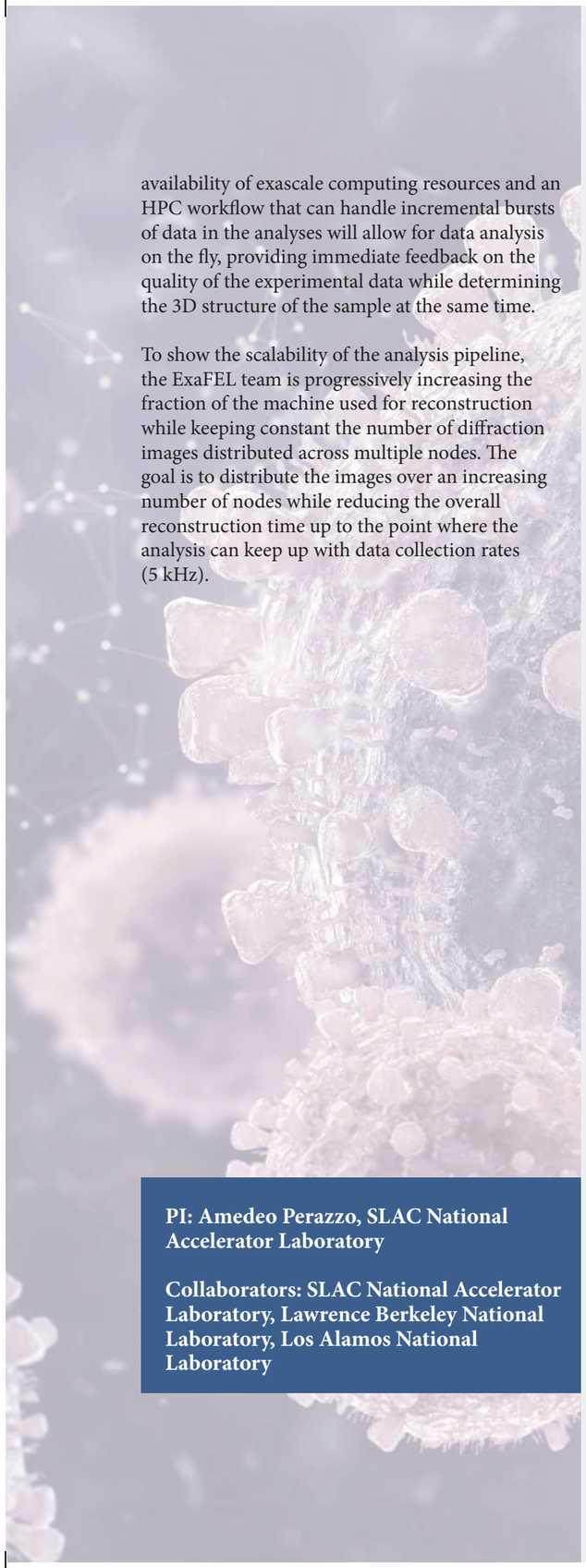
Users of the LCLS require an integrated combination of data processing and scientific interpretation, where both aspects demand intensive computational analysis. The ultrafast x-ray pulses are used like flashes from a high-speed strobe light to produce "stop-action movies" of atoms and molecules. The analysis must be carried out quickly to allow users to iterate their experiments and extract the most value from scarce beam time. Enabling new photon science from the LCLS will require near-real-time analysis (~10 min) of data bursts, requiring commensurate bursts of exascale-class computational intensities.

The high repetition rate and ultra-high brightness of the LCLS make it possible to determine the structure of individual molecules, mapping out their natural variation in conformation and flexibility. Structural dynamics and heterogeneities, such as changes in size and shape of nanoparticles, or conformational flexibility in macromolecules, are at the basis of understanding, predicting, and eventually engineering functional properties in the biology, material, and energy sciences. The ability to image these structural dynamics and heterogeneities using noncrystalline-based diffractive imaging, including single-particle imaging (SPI) and fluctuation x-ray scattering, has been one of the driving forces behind the development of x-ray free-electron lasers. However, efficient data processing, classification of diffraction patterns into conformational states, and subsequent reconstruction of a series of 3D electron densities, which allow for visualization of how the structure is changing, are vital computational challenges in diffractive imaging.

The ExaFEL challenge problem is the creation of an automated analysis pipeline for imaging of single particles via diffractive imaging. This entails the reconstruction of a 3D molecular structure from 2D diffraction images using the new Multi-Tiered Iterative Phasing (MTIP) algorithm. In SPI, diffraction images are collected from individual particles and are used to determine molecular (or atomic) structure, even from multiple conformational states (or nonidentical particles) under operating conditions. Determining structures from SPI experiments is challenging because orientations and states of imaged particles are unknown and images are highly contaminated with noise. Furthermore, the number of useful images is often limited by achievable single-particle hit rates, currently much less than 1. The MTIP algorithm introduces an iterative projection framework to simultaneously determine orientations, states, and molecular structure from limited single-particle data by leveraging structural constraints throughout the reconstruction, offering a potential pathway to increasing the amount of information that can be extracted from single-particle diffraction.

Rapid feedback is crucial for tuning sample concentrations to achieve a sufficient single-particle hit rate, ensuring that adequate data are collected and available to steer the experiment. The availability of exascale computing resources and an HPC workflow that can handle incremental bursts of data in the analyses will allow for data analysis on the fly, providing immediate feedback on the quality of the experimental data while determining the 3D structure of the sample at the same time.

To show the scalability of the analysis pipeline, the ExaFEL team is progressively increasing the fraction of the machine used for reconstruction while keeping constant the number of diffraction images distributed across multiple nodes. The goal is to distribute the images over an increasing number of nodes while reducing the overall reconstruction time up to the point where the analysis can keep up with data collection rates (5 kHz).



PI: Amedeo Perazzo, SLAC National Accelerator Laboratory

Collaborators: SLAC National Accelerator Laboratory, Lawrence Berkeley National Laboratory, Los Alamos National Laboratory

Progress to date

- Implemented Psana tasking, a port of Psana to the Legion programming system for exascale computing, that demonstrated performance comparable to MPI when running on up to 2,048 nodes of the Cori supercomputer.
- Demonstrated that using Legion in Psana tasking enables GPU use, a critical step toward readiness for upcoming exascale architectures.
- Optimized the merging step in the nanocrystallography pipeline to first read in the reduced data files in parallel and then to merge all duplicate observations together. Critically, the new algorithm allows the project team to merge a data set of $>10^9$ observations, which was previously intractable with a single process.
- Extended LUNUS diffuse scattering data processing to handle 10^3–10^4 diffraction images in parallel across multiple nodes. Developed MPI LUNUS for performance analysis and optimization and defined image processing requirements for SFX diffuse scattering data.

ExaFEL will enable real-time analysis of the thousandfold planned increase in x-ray diffraction data at LCLS, which will greatly increase the facility's ability to answer fundamental science questions about the nature of matter.

Application Portfolio

Co-design Centers

- Adaptive Mesh Refinement
- Efficient Exascale Discretizations
- Online Data Analysis and Reduction at the Exascale
- Particle-Based Applications
- Efficient Implementation of Key Graph Algorithms
- Exascale Machine Learning Technologies

Adaptive Mesh Refinement

Adaptive mesh refinement (AMR) is like a computational microscope; it allows scientists to "zoom in" on particular regions of space that are more interesting than others. Cosmologists might want to zoom in on detailed cosmic filaments; astrophysicists might focus on regions of nucleosynthesis; and combustion scientists may examine where the burning occurs.

Adaptive mesh refinement for exascale (AMReX) supports the development of block-structured AMR algorithms for solving systems of partial differential equations on exascale architectures and provides the basis for the temporal and spatial discretization strategy for a large number of applications relevant to DOE. AMR reduces the computational cost and memory footprint compared to a uniform mesh while preserving the local descriptions of different physical processes in complex multiphysics algorithms.

Principal Investigator: John Bell, Lawrence Berkeley National Laboratory

Collaborators: Lawrence Berkeley National Laboratory, Argonne National Laboratory, National Renewable Energy Laboratory

Efficient Exascale Discretizations

Efficient exploitation of exascale architectures requires a rethink of the numerical algorithms used in large-scale applications of strategic interest to the DOE. Many large-scale applications employ unstructured finite element discretization methods—the process of dividing a large simulation into smaller components in preparation for computer analysis—where practical efficiency is measured by the accuracy achieved per unit computational time.

The Center for Efficient Exascale Discretizations (CEED) is pursuing a crosscutting approach that includes working with hardware vendors, software developers, and computational scientists to meet the needs applications relevant to the DOE mission. The CEED co-design center is a focused team effort to develop the next-generation discretization software and algorithms that will enable a wide range of finite element applications to run efficiently on future hardware.

Principal Investigator: Tzanio Kolev, Lawrence Livermore National Laboratory

Collaborators: Lawrence Livermore National Laboratory, Argonne National Laboratory, Rensselaer Polytechnic Institute, University of Colorado Boulder, University of Illinois Urbana-Champaign, University of Tennessee, Virginia Tech

Online Data Analysis and Reduction at the Exascale

By 2024, computers are expected to compute at 10^{18} operations per second but write to disk only at 1012 bytes/sec: a compute-to-output ratio 200 times worse than on the first petascale systems. In this new world, applications must increasingly perform online data analysis and reduction—tasks that introduce algorithmic, implementation, and programming model challenges that are unfamiliar to many scientists and that have major implications for the design of various elements of exascale systems.

The goal of the Center for Online Data Analysis and Reduction at the Exascale **(CODAR)** is to produce infrastructure for online data analysis and reduction; provide valuable abstractions for implementing and customizing data analysis and reduction methods; import, integrate, and develop essential libraries implemented using these abstractions; incorporate the libraries into scientific applications and quantify accuracy and performance; release software artifacts; construct application-oriented case studies; document success stories and the process applied to obtain them; and report on codesign trade-off investigations.

PI: Ian Foster, Argonne National Laboratory

Collaborators: Argonne National Laboratory, Brookhaven National Laboratory, Oak Ridge National Laboratory, Brown University, Rutgers University, State University of New York – Stonybrook, University of Oregon

Particle-Based Applications

Particle-based simulation approaches are ubiquitous in computational science and engineering. The "particles" may represent, for example, the atomic nuclei of quantum and classical molecular dynamics methods or gravitationally interacting bodies or tracer particles in N-body simulations. In each case, every particle interacts with its environment by direct particle-particle interactions at shorter ranges and/or the particle-mesh interactions between a particle and a local field that is set up by longer range effects.

The Co-design Center for Particle Applications **(CoPA)** is addressing the challenges for particle-based applications to run on upcoming exascale computing architectures by partnering with ECP applications teams to create co-designed numerical recipes for particle-based methods: short-range particle–particle interactions (e.g., those which often dominate molecular dynamics methods), long-range particle-particle (e.g., electrostatic and gravitational) interactions, particle-in-cell methods, and linear-scaling electronic structure and quantum molecular dynamics algorithms.

PI: Susan Mniszewski, Los Alamos National Laboratory

Collaborators: Los Alamos National Laboratory, Argonne National Laboratory, Lawrence Livermore National Laboratory, Oak Ridge National Laboratory, Princeton Plasma Physics Laboratory, Sandia National Laboratories

Efficient Implementation of Key Graph Algorithms

Fundamental changes in the way electric power is generated, transmitted, and consumed today have resulted in an unprecedented need for computation to solve problems related to the design, planning, and operation of the power grid. This endeavor and the need to computationally design and model chemicals, materials, and biosystems at a molecular level are fundamental to accomplishing DOE's energy, environment, and national security missions. Recent advances in systems-based approaches coupled with ever-increasing data volumes require extreme-scale computing.

Combinatorial algorithms in general and graph algorithms in particular play a critical enabling role in numerous scientic applications. The irregular memory access nature of these algorithms makes them some of the most difficult algorithmic kernels to implement on parallel systems. Efficient implementation of key combinatorial (graph) algorithms chosen from four exascale application domains—smart grids, computational biology, computational chemistry, and climate science—will be captured in a unified soware framework, **ExaGraph**, that targets a diverse set of current and future extreme-scale architectures.

PI: Mahantesh Halappanavar, Pacific Northwest National Laboratory

Collaborators: Pacific Northwest National Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, Purdue University, Sandia National Laboratories

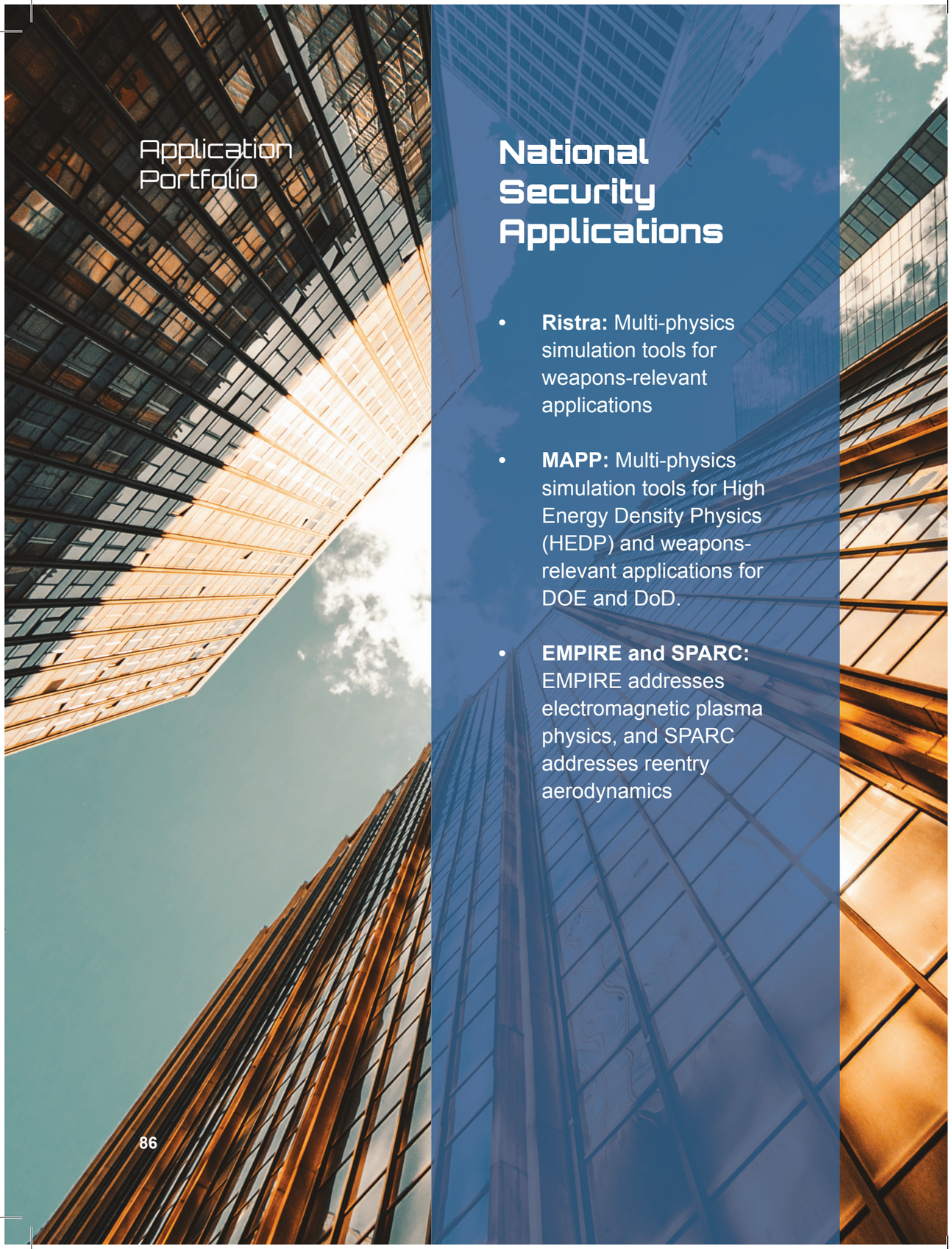
Exascale Machine Learning Technologies

The **ExaLearn** Co-design Center leverages the revolution in what is variously termed machine learning, statistical learning, computational learning, and articial intelligence. New machine learning technologies can have profound implications for computational and experimental science and engineering and thus for the exascale computing systems that DOE is developing to support those disciplines.

ExaLearn will identify the fundamental machine learning challenges associated with ECP applications and concentrate efforts on the development of scalable machine learning technologies for the analysis of data generated by exascale applications and DOE user facilities as well as guide the optimal selection and steering of (1) complex computer simulations (e.g., current exascale application projects) and (2) experiments (e.g., light sources, NIF, accelerators). The key to success in this endeavor is a deliberate focus on verification and validation and uncertainty quantification with a solid determination of generalization errors. A unifying principle is that of using exascale machine learning to improve the efficiency and effectiveness both of DOE computing resources and experimental facilities.

PI: Francis Alexander, Brookhaven National Laboratory

Collaborators: Brookhaven National Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Pacific Northwest National Laboratory, Sandia National Laboratories



Application Portfolio

National Security Applications

- **Ristra:** Multi-physics simulation tools for weapons-relevant applications
- **MAPP:** Multi-physics simulation tools for High Energy Density Physics (HEDP) and weapons-relevant applications for DOE and DoD.
- **EMPIRE and SPARC:** EMPIRE addresses electromagnetic plasma physics, and SPARC addresses reentry aerodynamics

Ristra:

The Ristra project is developing new multi-physics simulation tools that address emerging HPC challenges of massive, heterogeneous parallelism using novel programming models and data management.

Lead: Los Alamos National Laboratory

Principal Investigators: Aimee Hungerford and David Daniel, Los Alamos National Laboratory

MAPP:

The Multi-physics on Advanced Platforms Project (MAPP) is developing next generation multi-physics codes for simulating high-energy-density and focused physics experiments driven by high-explosive, magnetic or laser-based energy sources as well as weapons-relevant simulations. The MAPP approach includes a modular CS infrastructure that serves as the basis of multiple codes, as well as an emphasis on high-order algorithms that are expected to scale better on anticipated architectures, which are distinguishing characteristics of the project.

Lead: Lawrence Livermore National Laboratory

Principal Investigator: Rob Rieben, Lawrence Livermore National Laboratory

EMPIRE AND SPARC:

EMPIRE will deliver advanced electromagnetic and plasma physics code capabilities for next-generation hardware architectures, and SPARC will create a revolutionary reentry simulation capability well suited for effective use on next-generation HPC platforms

Lead: Sandia National Laboratories

Principal Investigators: EMPIRE, Matt Bettencourt, Sandia National Laboratories; SPARC, Micah Howard, Sandia National Laboratories

The goal of the ECP efforts in support of national security applications is the development of comprehensive science-based computational weapons applications able to provide, through effective exploitation of exascale HPC technologies, breakthrough modeling and simulation solutions that yield high-confidence insights into at least three problems of interest to the NNSA Stockpile Stewardship Program (SSP) that cannot be solved on the HPC systems available today. For two of the NNSA laboratories (Los Alamos and Lawrence Livermore), the demonstration applications include one or more weapons-relevant physics simulations. For the third NNSA laboratory (Sandia), the demonstration application embodies the weapons-relevant multi-physics phenomena in reentry aerodynamics and electromagnetics areas.

The outcomes and products of this activity will be integrated into the next generation of high-performance codes on advanced architectures. Together, the three NNSA laboratories aim to deliver applications that can address currently infeasible 3D problems of interest. The three approaches are different but complementary, providing both peer review and risk mitigation.

Los Alamos National Laboratory

NATIONAL SECURITY APPLICATIONS

Los Alamos National Laboratory's mission is to solve national security challenges through scientific excellence. The laboratory's strategic plan reflects US priorities spanning nuclear security, intelligence, defense, emergency response, nonproliferation, counterterrorism, energy security, emerging threats, and environmental management.

Stockpile Stewardship

Ristra: Multi-physics Simulation for National Security Problems

The property and behavior of various materials under a wide variety of extreme conditions are central to many applications within the realm of national security. Such modeling requires multiple length and timescales and drives requirements for exascale computing. Los Alamos National Laboratory is developing a next-generation multi-physics code for national security applications that focuses on 3D multi-physics, insight at the mesoscale for extreme condition materials, and high-energy-density physics simulations.

Computer science technologies that allow efficient use of emerging HPC architectures suggest a need for physics algorithms that permit increased concurrency at many scales. This motivates a fresh look at the numerical decisions made throughout the simulation process, from setup through analysis. With this in mind, Ristra is casting a wide net across available physics algorithms for multi-physics simulation and, at the same time, exploring novel programming models for emerging architectures.

Ristra's focus is on two application domains, both of which feature multiscale methods that will be an important component of extreme-scale multi-physics simulations of the future.

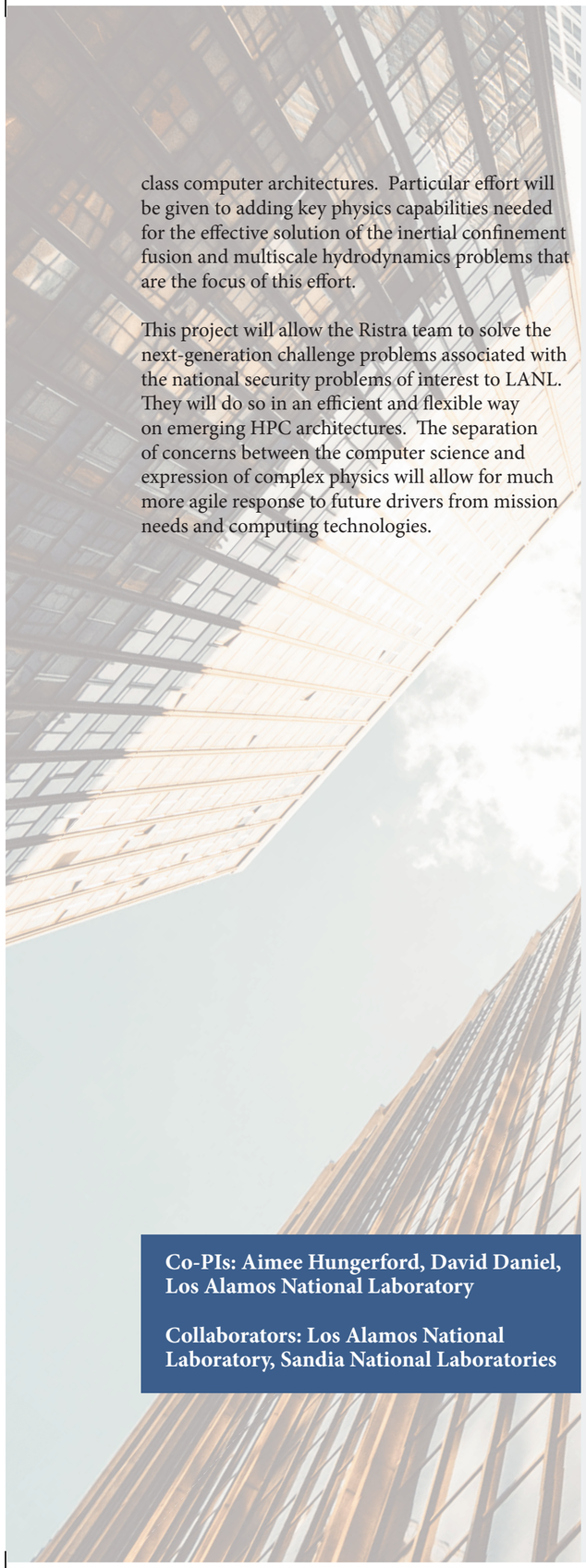
High-Energy-Density Physics for Inertial Confinement Fusion. Ristra's Symphony code is an unstructured multi-material radiation hydrodynamics application that features a multiscale algorithm for the radiation solve: a fully coupled low-order radiation hydrodynamics system is updated by a high-order radiation solver which has the potential to be executed asynchronously (work in progress).

Multiscale Hydrodynamics of Materials in Extreme Conditions. Ristra's FUEL code is an unstructured multi-material arbitrary Lagrangian-Eulerian (ALE) hydrodynamics code that can be coupled to complex material models to take account of mesoscale physics, such as grain structure, in the dynamic response of materials. Mesoscale modeling is computationally intensive, and the multiscale approach has potential for effective use of exascale-class systems, as well as providing a promising target for data-driven machine-learning (ML) techniques. Key to the architecture of Ristra's applications is FleCSI (Flexible Computer Science Infrastructure), an abstraction layer that provides the desired separation of concerns between computational physics and computer science.

FleCSI is a compile-time configurable framework designed to support multi-physics application development. As such, FleCSI provides a very general set of infrastructure design patterns that can be specialized and extended to suit the needs of a broad variety of solver and data requirements. FleCSI provides an abstract data model supporting compile- and run-time configurability for implementing a variety of discretizations (mesh and mesh-free) and physics fields and operators over them. FleCSI also provides an abstract execution model that can target a variety of underlying parallel programming runtimes from well-established options such as MPI to ambitious new programming systems such as Legion, a data-centric model with out-of-order task execution. The intent is to provide developers with a concrete set of user-friendly programming tools that can be used now, while allowing flexibility in choosing runtime implementations and optimizations that can be applied to future architectures and runtimes. This effort also provides a realistic infrastructure for the evaluation of programming models and data management technologies.

Over the course of the ECP project, the Ristra team will continue to push the boundaries on the development of multiscale, multi-physics applications as well on the programming models needed to demonstrate performance on exascale-class computer architectures. Particular effort will be given to adding key physics capabilities needed for the effective solution of the inertial confinement fusion and multiscale hydrodynamics problems that are the focus of this effort.

This project will allow the Ristra team to solve the next-generation challenge problems associated with the national security problems of interest to LANL. They will do so in an efficient and flexible way on emerging HPC architectures. The separation of concerns between the computer science and expression of complex physics will allow for much more agile response to future drivers from mission needs and computing technologies.



Co-PIs: Aimee Hungerford, David Daniel, Los Alamos National Laboratory

Collaborators: Los Alamos National Laboratory, Sandia National Laboratories

Progress to date

The initial features of the FleCSI infrastructure are complete, and the Symphony and FUEL codes are maturing to support the required physics modules for a target challenge problem. Recent accomplishments of the project include the following.

- Demonstrated first results from Ristra's Symphony code, an unstructured mesh, multi-material radiation hydrodynamics code using a novel high-order/low-order (HO LO) multiscale method for the radiation solver. Symphony was demonstrated using a variety of test problems including 2D and 3D Inertial Confinement Fusion (ICF) calculations.
- Enhanced fidelity in ICF simulations through adoption of new Discontinuous Galerkin low-order radiation solver.
- FleCSI's parallel capability was showcased using Ristra's simpler single material gas hydrodynamics code FleCSALE, demonstrating the ability of a code to connect through FleCSI to distinct parallel backends (MPI and Legion) with no required changes to the physics implementation.
- Enhancements to improve support for multi-material capability were added to FleCSI and demonstrated in Symphony and FleCSALE-mm.
- Demonstrated first results from a new FleCSI-based, serial, unstructured Eulerian and Lagrangian low-energy-density code FUEL.

Lawrence Livermore National Laboratory

For more than 60 years, the Lawrence Livermore National Laboratory (LLNL) has applied science and technology to make the world a safer place. LLNL will be home to El Capitan, one of the US Department of Energy's three planned exascale supercomputers.

NATIONAL SECURITY APPLICATIONS

Stockpile Stewardship

MAPP: Multi-physics on Advanced Platforms Project

LLNL is developing next-generation multi-physics simulation capabilities for national security applications and has adopted a modular approach to code development through the Multi-physics on Advanced Platforms Project (MAPP). The software that will be developed as part of MAPP will address the modeling needs of the high-energy-density physics (HEDP) community for simulating high-explosive, magnetic or laser-driven experiments such as inertial confinement fusion (ICF), pulsed-power magneto-hydrodynamics (MHD), equation of state (EOS) and material strength studies as part of the NNSA's stockpile stewardship program (SSP).

Fundamental to MAPP is the Axom computer science (CS) toolkit which provides a library of shared software components that provide various services for the development of modular, multiphysics application codes. MARBL, a next-generation code focused on ICF and pulsed power applications, is one code in the project that is built on top of the Axom toolkit. MARBL exemplifies the overall philosophy within the project of extreme modularity in physics and CS capabilities and includes multiple options for every major physics capability.

The Axom library consists of a collection of software components that provide core computer science infrastructure capabilities that can be shared by diverse high performance computing (HPC) applications. The current set of capabilities that Axom provides includes customizable support for error/warning and diagnostic message reporting; coordination among components of integrated applications (e.g., physics packages, libraries, etc.); and an in-memory datastore for hierarchical, mesh-aware simulation data. The Axom datastore supports data description, allocation, deallocation, and parallel l/0, along with mesh data model abstractions that enable the development of computational algorithms that work with many different mesh types.

The MARBL application code is designed from inception to support multiple diverse algorithms, including arbitrary Lagrangian-Eulerian (ALE) and direct Eulerian methods for solving the conservation laws associated with its various physics packages. A distinguishing feature of MARBL is the use of advanced, high-order numerical discretizations such as high-order finite element ALE and high-order finite difference Eulerian methods. This algorithmic diversity encompasses the ECP simulation motifs of unstructured and structured adaptive mesh refinement (AMR). High-order numerical methods were chosen because they have higher resolution/accuracy per unknown compared to standard low-order finite volume schemes and because they have computational characteristics which play to the strengths of current and emerging HPC architectures. Specifically, they have higher FLOP/byte ratios, meaning that more floating-point operations are performed for each piece of data retrieved from memory. This leads to improved strong parallel scalability on GPU platforms and increased computational efficiency. If successful, the advanced simulation capabilities provided by MARBL will improve user throughput along two axes: faster turnaround for multi-physics simulations on advanced architectures and less manual user intervention.

A key goal for MARBL is enhanced end-user productivity including improved workflow for problem setup and meshing, simulation robustness, support for UQ and optimization-driven ensembles, and in situ data visualization and analysis. High-order ALE and Eulerian schemes have proven to be more robust and should significantly improve the overall analysis workflow for users. As such, the MAPP project represents a massive software development effort, incorporating multiple physics, mathematics and computer science packages into the overall integrated code.

The team collaborates with multiple ST projects to integrate new production quality capabilities, including software developed both internally at LLNL and externally from the ECP and the broader open-source community.

The success of MAPP will ultimately be determined by the degree of adoption of its simulation tools by the LLNL user community. To this end, emphasis at this relatively early stage of development is being placed on adding physics and capabilities to meet the current state of the art that users demand from today's petascale production simulation codes. In the case of MARBL, this includes coupled multi-material radiation-magneto-hydrodynamics, thermonuclear burn for inertial confinement fusion calculations, general equations of state, material opacities and electrical conductivities, simulation diagnostics and queries, in situ analytics / rendering, and parallel computational and file IO performance at a massive scale. In addition, performance of the new codes on advanced architectures like the GPU-based Sierra system at LLNL is critical. Portability of the software stack and long-term maintainability are critical as well, placing stringent demands on the integration and interoperability of high-quality production-level software libraries and tools. Finally, MARBL will be the first demonstration of the viability of advanced high-order numerical approaches for production multi-physics simulation at scale in the NNSA and has already produced first-of-a-kind simulation results using such methods.

PI: Rob Rieben, Lawrence Livermore National Laboratory

Collaborators: Lawrence Livermore National Laboratory

Progress to date

- Development and release of modular library for calculating thermonuclear (TN) reaction rates, electron-ion coupling coefficients, and other commonly used plasma physics properties
- Fully coupled, high-order finite element Arbitrary Lagrangian-Eulerian (ALE) radiation-hydrodynamics
- Modular physics packages combined with computer science infrastructure library (Axom)
- Seamless connection to ST libraries for checkpoint, in situ rendering and data transfer
- First-of-a-kind high-order ALE simulation results using novel non-linear mesh optimization plus high-order discontinuous Galerkin (DG) for ALE remesh/remap in large-scale 3D radiation-hydro simulations

NATIONAL SECURITY APPLICATIONS

Keeping the US nuclear stockpile safe, secure, and effective is a major part of Sandia's work as a multidisciplinary national security engineering laboratory. The laboratory's highly specialized research staff is at the forefront of innovation, collaborating with universities and companies and performing multidisciplinary science and engineering research programs with significant impact on US security.

MULTI-PHYSICS FOR NATIONAL SECURITY

Sandia's Advanced Technology Development and Mitigation (ATDM) components vision has enabled apps like EMPIRE and SPARC to build on foundational capabilities developed and deployed by other teams, providing great leverage and potential for reuse and increased impact.

EMPIRE: Preparing Electromagnetic Plasma Physics Codes for Exascale

Ionizing electro-magnetic pulse (IEMP) environments comprise both system-generated and source-region-generated electromagnetic pulse conditions. To meet mission requirements, these IEMP environments must be extrapolated from what can be realized with test facilities, and hence, validated computational simulation tools become critical. To meet this need, EMPIRE will deliver advanced electromagnetic and plasma physics code capabilities, will be performant on next-generation hardware architectures, and will provide analysis tool suites.

The team will focus on computing electronic effects induced by ionizing radiation interacting with materials under various reentry flight conditions. The project will develop a self-consistent plasma simulation including the radiation output of a hostile builder device, radiation transport, plasma generation and propagation down through the effects on nuclear system electronics.

The team will demonstrate the capabilities in EMPIRE with a simulation of an experiment that is relevant to the team's qualification mission: modeling a high-energy beam with particles and the background gas with a fluid at a higher fidelity than has been possible before. For the performance and code comparison metrics, the team will expand upon the validation work already performed and simulate a cavity plasma experiment. The EMPIRE results will be compared to legacy-code simulations of the same problem to show the performance and portability enhancements that have been enabled by the ATDM and ECP program.

SPARC: Sandia Parallel Aerodynamics and Reentry Code Virtual Flight Testing

Engineering and physics applications for hypersonic reentry have multiple national security implications and represent complex modeling of physical phenomena and engineering responses that significantly drive exascale computing requirements. SPARC will provide nuclear weapon qualification evidence for the random vibration and thermal environments created by reentry of a warhead into the earth's atmosphere. This state-of-the-art hypersonic flight simulation capability on next-generation hardware will include thermo-chemical nonequilibrium gas ablation models and hybrid RANS-LES turbulence models.

The pacing science challenge problem for SPARC is to perform a virtual flight test of a reentry vehicle, in its entirety, and to predict the structural and thermal response of the vehicle's components under simulated reentry environments. Performing this analysis includes simulation of the flow field around the vehicle (including the aft end and its wake) using a turbulence model suited for hypersonic, unsteady turbulent fluid dynamics. The thermal loads generated from the computational fluid dynamics simulation will be used to predict the ablation and thermal response of the vehicle's thermal protection system and internal components. The structural loads generated from pressure and shear stress fluctuations predictions by the turbulence models will be used to analyze the vibrational response of the vehicle and its internal components. This predictive capability, which is being validated simultaneously with the code's development, will give weapons engineers an ability to assess reentry vehicle response to trajectories where little flight test data exists.

Sandia's ATDM components vision has enabled apps like EMPIRE and SPARC to build on foundational capabilities developed and deployed by other, well-coordinated teams, providing great leverage and potential for reuse and increased impact. For example, EMPIRE has used discretization and linear solver technology deployed in Trilinos to make the development process more efficient. Both EMPIRE and SPARC incorporate innovative approaches on several fronts including effective utilization of heterogeneous compute nodes using Kokkos, uncertainty quantification through Sacado integration, embedded mesh refinement and geometry, implementation of state-of-the-art reentry physics and multiscale models, use of advanced verification and validation methods, and enabling of improved workflows for users.

Progress to date

EMPIRE:

- Assessed status of next generation components and physics models in EMPIRE. The assessment focused on the electromagnetic and electrostatic particle-in-cell solutions for EMPIRE and its associated solver, time integration, and checkpoint-restart components. The assessment included code verification, performance, and portability across available HPC architectures.
- Next-generation readiness was based on the incorporation of portable performance abstractions such as Kokkos (MPI+X), high performance I/O (FAODEL), and time integration libraries that allow for embedded sensitivity analysis (Tempus). Performance was tested on roughly half of each Trinity partition (KNL and HSW). The core PIC algorithm was shown to have nearly perfect weak and strong scaling, on up to 256k cores on Trinity and problems with up to 1.3B elements and 66B particles.

SPARC:

- Assessed the code performance and portability across Trinity-Haswell, Trinity-KNL, and GPU test platforms. Algorithmic improvements have led to significant strong-scaling speedups since the start of FY19 (e.g., ~4× for GPU performance).
- A new structured mesh refinement capability that honors CAD geometry has been integrated. By honoring the CAD geometry, each refinement of the mesh more accurately captures the shape of curved surfaces, which helps achieve more accurate simulation results. Embedding the refinement process in the application will allow SPARC to read in a coarse mesh and refine in parallel to produce dramatically larger meshes for detailed analyses on large problems.
- Conducted a comprehensive verification and validation study of hypersonic flow in SPARC validated by several experiments and reviewed an external review committee. The V&V process was thorough, including applicable frameworks, professional standards, code and solution verification, calibration, sensitivity analysis, and parametric uncertainty, and has provided a basis for using SPARC as credible analysis tool for hypersonic reentry flows.

PIs:

EMPIRE: Matt Bettencourt, Sandia National Laboratories

SPARC: Micah Howard, Sandia National Laboratories
Collaborators: Sandia National Laboratories

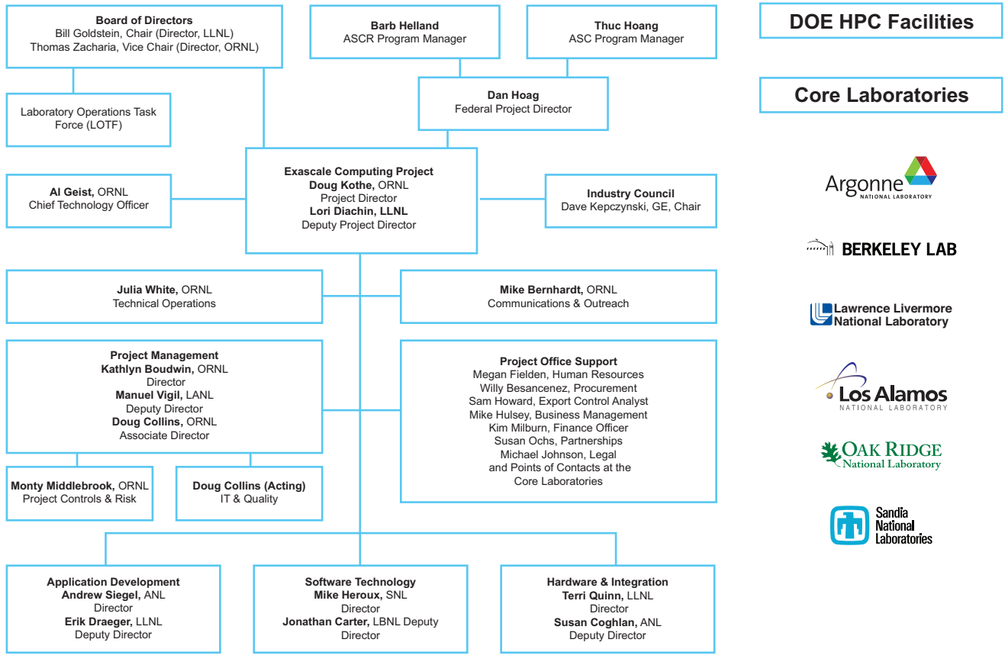
BRINGING IT ALL TOGETHER

BRINGING IT ALL TOGETHER

The ECP is led by a team of senior computer and computational scientists, engineers, and project specialists from six US Department of Energy (DOE) national laboratories that have historically maintained core competency and leadership in high performance computing (HPC), mathematics, and computer and computational science. Working together, the ECP leadership team has established an extensive network to deliver a capable exascale computing ecosystem for the nation, partnering with experts at other DOE national laboratories and HPC facilities, US HPC companies, and leading academic institutions.

The ECP Board of Directors (BOD) consists of the laboratory directors from the six core partner DOE laboratories, who have signed a memorandum of agreement. The Board has an active advisory, oversight, and line-management role. Within the Board is an executive committee that selects a chair and vice-chair from among its membership. One of the two will be from a DOE Office of Science laboratory and the other from a National Nuclear Security Administration laboratory. The Board's primary purpose is to provide strategic direction to the ECP project director and leadership team. The ECP Board also appoints a Laboratory Operations Task Force (LOTF), which is composed of associate laboratory directors with line management responsibility for HPC at their respective core partner DOE laboratories. The LOTF assists the Board in overseeing the operations of the ECP and supports and advises the ECP project director.

ECP Organization



Ensuring Strong, Competitive HPC Capabilities for US Industry in the Age of Exascale

The ECP Industry Council is an external advisory group that provides essential two-way communication and information exchange between ECP and the HPC industrial user community as well as the commercial HPC software community.

ECP's Industry Council is composed of senior executives from some of the nation's most prominent companies who share a collaborative interest in bringing the potential of exascale computing to a wide range of industry segments.

ECP Industry Council Member Organizations

Altair Engineering, Inc.
ANSYS, Inc.
Cascade Technologies, Inc.
Chevron Corporation
Cummins, Inc.
Animation Eli Lilly and Company
ExxonMobil Corporation
FedEx Corporation
General Electric
General Motors Company
KatRisk, LLC
Procter & Gamble
The Boeing Company
The Goodyear Tire & Rubber Company
Tri Alpha Energy, Inc.
United Technologies Corporation
Westinghouse Electric Company
Whirlpool Corporation

Maintaining the highest level of computational capability is critical to the nation's industrial competitiveness, and the ECP Industry Council is critically important to keeping the project in sync with the real world needs of the US industrial sector.

ECP by the Numbers

7 YEARS $1.8B

A 7 year, $1.8B R&D effort that launched in 2016

Six CORE DOE LABS

Six core DOE national laboratories: Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge and Sandia

Staff from most of the 17 DOE national laboratories take part in the project

Three FOCUS AREAS

Three technical focus area: Hardware and Integration, Software Technology, Application Development supported by a Project Management Office

80 R&D TEAMS

More than 80 top-notch R&D teams

1000 RESEARCHERS

Hundreds of consequential milestones delivered on schedule and within budget since project inception

The ECP's Enduring Legacy

The products and solutions generated by the Exascale Computing Project (ECP) will have a tremendous impact across the entire high performance computing (HPC) ecosystem, benefitting the US economy, scientific discovery, and national security for many years to come. Well into the next decade, scientists and researchers will take advantage of the ECP-driven advancements in numerous applications, software tools, and hardware innovations, accelerating a wide range of research efforts that address the toughest problems facing the DOE and the nation.

All Boats Will Rise

The emerging exascale ecosystem and exciting new capabilities made possible by the ECP's efforts will not be limited to applications running only on exascale platforms. The enduring legacy of the ECP will trickle down and benefit computing systems of all sizes (clusters to desktops), impacting R&D in scientific applications, as well as industrial and commercial high-end computing. The ECP's legacy for software will impact a new generation of HPC systems, well beyond Aurora, Frontier, and El Capitan.

While the legacy of the ECP will be most visible in applications, significant advances in software development and delivery, along with innovative hardware architecture enhancements, will be foundational to enabling those applications to meet their anticipated science capability and execute performance goals as the next generation of computational and data science tools. The software technology and hardware and integration efforts represent a behind-the-scenes herculean accomplishment that rounds out the capable exascale ecosystem. Ultimately, the ECP will transfer to the computational science and engineering community a portfolio of exascale-ready applications, a sophisticated, modern software stack, and a blueprint for best practices in co-design, continuous integration, and collaboration among government, academia, and industry enabling unprecedented HPC portability.

It is also worth noting that, with the added impetus from the ECP, the scientific and technical computing communities will experience widespread adoption of accelerators for enhancing the performance of exascale environments supported by documented best practices and extensive workforce training programs.

Our forthcoming project overview and update documents covering the ECP's software technology and hardware and integration efforts will give an in-depth look at these other technical focus areas and, together, will provide a great perspective on the nation's capable exascale ecosystem—the enduring legacy of the ECP.

Credits:

Project Management

Mike Bernhardt, Oak Ridge National Laboratory
Julia White, Oak Ridge National Laboratory

Graphics and Design

Adam Malin, Oak Ridge National Laboratory
Julie Russell, Lawrence Livermore National Laboratory
Meg Epperly, Lawrence Livermore National Laboratory

Production

Kase Clapp, Oak Ridge National Laboratory
Adam Malin, Oak Ridge National Laboratory

Content Contributors

Katie Bethea, Oak Ridge National Laboratory
Susan Coghlan, Argonne National Laboratory
James Collins, Argonne National Laboratory
Tom Evans, Oak Ridge National Laboratory
Michael Papka, Argonne National Laboratory
Terri Quinn, Lawrence Livermore National Laboratory
Jeremy Thomas, Lawrence Livermore National Laboratory

Application Development Principal Investigators:

(In order of applications within this document)

Andreas Kronfeld, Fermi National Accelerator Laboratory
Thom Dunning, Pacific Northwest National Laboratory
Mark Gordon, Iowa State University
Danny Perez, Los Alamos National Laboratory
John Turner, Oak Ridge National Laboratory
Paul Kent, Oak Ridge National Laboratory
Mike Sprague, National Renewable Energy Laboratory
Jackie Chen, Sandia National Laboratories
Steven Hamilton, Oak Ridge National Laboratory
Madhava Syamlal, National Energy Technology Laboratory
Amitava Bhattacharjee, Princeton Plasma Physics Laboratory
Jean-Luc Vay, Lawrence Berkeley National Laboratory
Daniel Kasen, Lawrence Berkeley National Laboratory
Salman Habib, Argonne National Laboratory
David McCallen, Lawrence Livermore National Laboratory
Carl Steefel, Lawrence Berkeley National Laboratory
Mark Taylor, Sandia National Laboratories
Zhenyu (Henry) Huang, Pacific Northwest National Laboratory
Rick Stevens, Argonne National Laboratory
Katherine Yelick, Lawrence Berkeley National Laboratory
Amedeo Perazzo, SLAC National Accelerator Laboratory
John Bell, Lawrence Berkeley National Laboratory
Tzanio Kolev, Lawrence Livermore National Laboratory
Ian Foster, Argonne National Laboratory
Susan Mniszewski, Los Alamos National Laboratory
Mahantesh Halappanavar, Pacific Northwest National Laboratory
Francis Alexander, Brookhaven National Laboratory
Aimee Hungerford, Los Alamos National Laboratory
David Daniel, Los Alamos National Laboratory
Rob Rieben, Lawrence Livermore National Laboratory
Matt Bettencourt, Sandia National Laboratories
Micah Howard, Sandia National Laboratories

The Exascale Computing Project (ECP) is a joint effort of two US Department of Energy (DOE) organizations: the Office of Science and the National Nuclear Security Administration.

The ECP is led by a team of senior scientists, project management experts and engineers from six of the largest DOE national laboratories.

Argonne NATIONAL LABORATORY

BERKELEY LAB

OAK RIDGE National Laboratory

Los Alamos NATIONAL LABORATORY

Lawrence Livermore National Laboratory

Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY | Office of Science

NNSA National Nuclear Security Administration

exascaleproject.org

Published: September 2019