Exascale Computing Project Update

Paul Messina, ECP Director

Stephen Lee, ECP Deputy Director

ASCAC Meeting, Arlington, VA

DoubleTree Hilton Washington-Crystal City
September 27, 2017

EXASCALE COMPUTING PROJECT

Transition of leadership

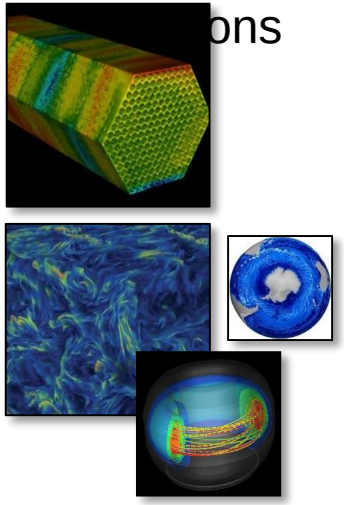
- Doug Kothe has been selected by the ECP Board of Directors as the new ECP Director effective October 1, 2017
- Doug has been involved in the ECP since the beginning and has done a stellar job of leading the Application Development focus area. I am confident that the project will thrive in Doug's capable hands

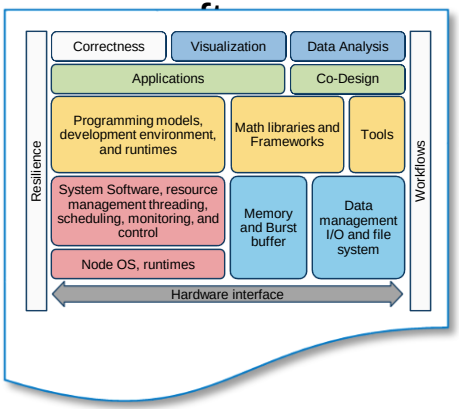
Staffing changes

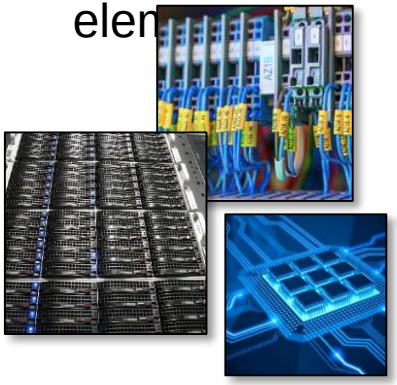
- Stephen Lee will continue serving as ECP Deputy Director. He has been invaluable as my deputy and I am delighted that he will be Doug's deputy as well.
- Other leadership positions need to be filled
  - some members of the ECP leadership team announced months ago that they are stepping down from their positions
  - We are fortunate to have many great leadership candidates in the 6 ECP core labs and are working with lab leadership to fill those roles rapidly within the next 2-3 weeks.

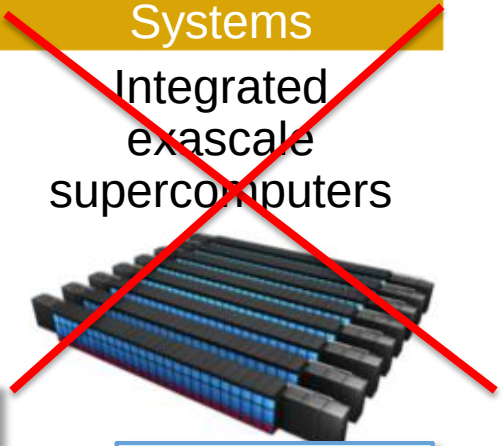
Changes in the ECP technical focus areas

Application Development	Software Technology	Hardware Technology	Exascale Systems
Science and mission ons	Scalable and productive	Hardware technology elem	Integrated exascale supercomputers









- Exascale Systems
  1.5 Terri Quinn, Susan Coghlan
- NRE
  1.5.1 Terri Quinn
- Testbeds
  1.5.2 Terri Quinn
- Co-design and Integration
  1.5.3 Susan Coghlan

Closer engagement and integration with facilities

- Due to the phasing out of the Exascale Systems focus area we are considering expanding the scope of the Hardware Technology focus area to include WBS elements for integration and collaboration with the supercomputing facilities at the six ECP core labs
  - This is a work in progress
- In the meantime we are having in-person meetings and jointly drafting high-level principles of engagement and detailed documents describing facility-ECP integration plans with each facility

ECP Apps also in ALCF Early Science Program

HACC	ECP
USQCD	ECP
QMCPACK	ECP
XGC	ECP
NWCHEMX	ECP
PHASTA	
NAQMD/RMD	
UINTAH	
SU2/PadeOps	
NAMD	

ECP Apps also in OLCF CAAR

ACME	ECP
DIRAC	
FLASH	ECP uses components from FLASH in CLASH
GTC	
HACC	ECP
LSDALTON	
NAMD	
NUCCOR	
NWCHEM	ECP
QMCPACK	ECP
RAPTOR	
SPECFEM	
XGC	ECP

ECP Apps also in NERSC NESAP

1. Computing the Sky at Extreme Scales, Salman Habib (ANL)
2. Exascale Lattice Gauge Theory Opportunities and Requirements for Nuclear and High Energy Physics, Paul Mackenzie (FNAL)
3. Molecular Dynamics at the Exascale: Spanning the Accuracy, Length and Time Scales for Critical Problems in Materials Science, Arthur Voter (LANL)
4. Exascale Modeling of Advanced Particle Accelerators, Jean-Luc Vay (LBNL)
5. An Exascale Subsurface Simulator of Coupled Flow, Transport, Reactions and Mechanics, Carl Steefel (LBNL)
6. NWChemEx: Tackling Chemical, Materials and Biomolecular Challenges in the Exascale Era, T. H. Dunning, Jr. (PNNL)
7. High-Fidelity Whole Device Modeling of Magnetically Confined Fusion Plasma, Amitava Bhattacharjee (PPPL)
8. Transforming Combustion Science and Technology with Exascale Simulations, Jackie Chen (SNL)
9. Cloud-Resolving Climate Modeling of the Earth's Water Cycle, Mark Taylor (SNL)
10. Exascale Solutions for Microbiome Analysis, Kathy Yelick (LBNL)
11. High Performance, Multidisciplinary Simulations for Regional Scale Seismic Hazard and Risk Assessments, David McCallen (LBNL)

Update on Application Development

EXASCALE COMPUTING PROJECT

The seven seed apps have been renewed

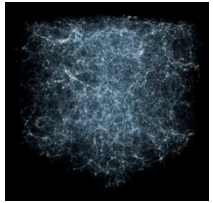
- After a thorough review, all seven seed apps will be renewed for FY 2018 and their funded increased to $1M/yr. (was $300K), as planned
  - Enabling GAMESS for Exascale Computing in Chemistry & Materials, Mark Gordon (Ames)
  - Multiscale Coupled Urban Systems, Charlie Catlett (ANL)
  - Exascale Models of Stellar Explosions: Quintessential Multi-Physics Simulation, Daniel Kasen (LBNL)
  - Exascale Solutions for Microbiome Analysis, Kathy Yelick (LBNL)
  - High Performance, Multidisciplinary Simulations for Regional Scale Seismic Hazard and Risk Assessments, David McCallen (LBNL)
  - Performance Prediction of Multiphase Energy Conversion Devices with Discrete Element, Particle-in-Cell, and Two-Fluid Models (MFIX-Exa), Madhava Syamlal (NETL)
  - Optimizing Stochastic Grid Dynamics at Exascale, Henry Huang (PNNL)

ECP Applications Adopt New Infrastructure for Block-Structured Adaptive Mesh Refinement Developed by AMReX Co-Design Center

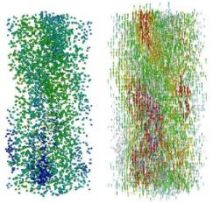
ECP WBS 1.2.5.03: AMReX
PI: John Bell, LBNL
Members: LBNL, ANL, NREL

Scope & Objectives

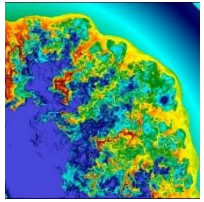
- Develop infrastructure to enable block-structured adaptive mesh refinement on exascale architectures
  - Core mesh, particle & particle-mesh operations on adaptive mesh hierarchy
  - Support for multiple time-stepping approaches
  - Embedded boundary representation of complex geometry
  - Performance portability for different architectures
- Current activities focused on:
  - Establishing support for core AMR functionality
  - Engagement of applications



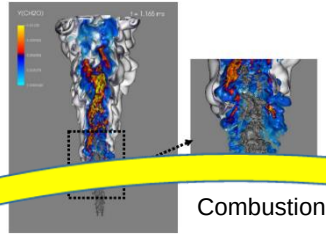
Cosmology



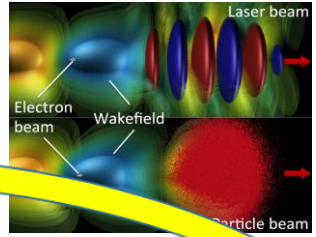
Multiphase flow in chemical reactors



Astrophysics



Combustion



Accelerator design

Impact

- Established a next-generation framework for developing block-structured adaptive mesh refinement algorithms for current and emerging architectures
- Provides a common framework for multiple ECP applications that use AMR
- Provides a common focal point for software technology, hardware technology and vendors to leverage activities over multiple applications
- Broad constituency within Office of Science and NNSA

Project Accomplishment

- New AMReX code framework adopted by multiple ECP applications
  - Accelerator modeling – WarpX
  - Astrophysics -- ExaStar (CASTRO)
  - Combustion – PeleC and PeleLM
  - Cosmology – ExaSky (Nyx)
  - Multiphase flow – MFIX-Exa
- AMReX code framework publicly released

Deliverables: AMReX software available at https://github.com/AMReX-Codes/amrex

Engaging a first wave of ECP applications in the Center for Efficient Exascale Discretizations (CEED)

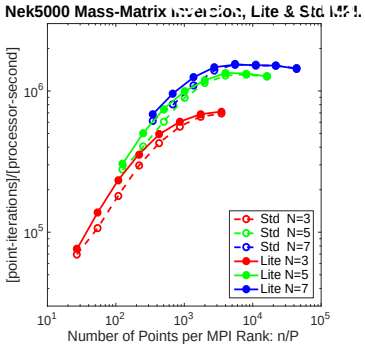
ECP WBS 1.2.5.3.4: CEED
PI: Tzanio Kolev, LLNL
Members: LLNL. ANL, UIUC, UTK, RPI, CU Boulder, Virginia Tech

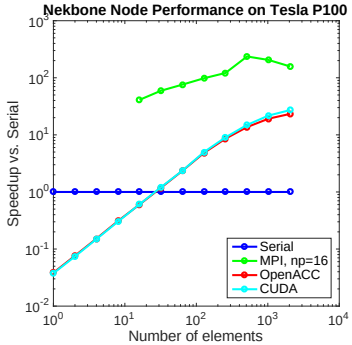
Scope & Objectives

- Deliver state-of-the-art high-order (FE/SE) discretization libraries to enable PDE-based applications to deliver accurate and efficient solutions at exascale
- Deliver fast algorithms for high-order operator evaluations with an order of magnitude performance gain over conventional methods
- Deliver efficient high-order representation to handle complex geometries

First Year Development Plan:

- Q1: Select the first wave of ECP/CEED applications and identify their needs
- Q2: Engage vendors and STs with initial miniapps and benchmarks
- Q3: Perform technology comparison with bake-off problems (benchmarks)
- Q4: Identify discretization-related performance bottlenecks in first wave apps

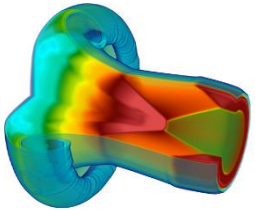


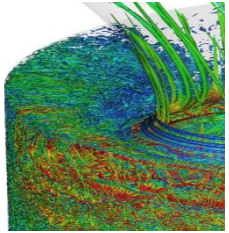


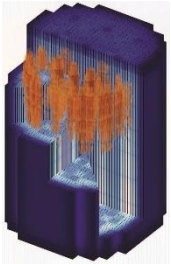
Nek5000 Lightweight MPI vs. standard MPI for mass-matrix inversion performance on 16384 MPI ranks; CPU vs. GPU performance on single node with Tesla P100s

CEED Impact

- Ultimate goal: discretization libraries supporting exascale-level performance with fast time-to-solution for target ECP applications: ExaSMR, MARBL, Combustion, Urban, ACME, …







Project Accomplishments

- Identified the needs for the first wave applications: ExaSMR, MARBL (Nek/MFEM based)
  - GPU-enabled Nek5000 pressure solver released: http://github.com/Nek5000/Nek5000
- Interactions with vendors, provided with miniapp
  - GPU-enabled miniapp Nekbone released: http://github.com/Nek5000/Nekbone
- Interactions with MPICH team: Nek5000 performed light-weight vs. standard MPI for bake-off problems (matrix inversion) – Results submitted as a paper to SC17
- Next: More miniapp releases, Perform bake-off problems and analysis
- Next: Fully extended GPU simulations for ExaSMR

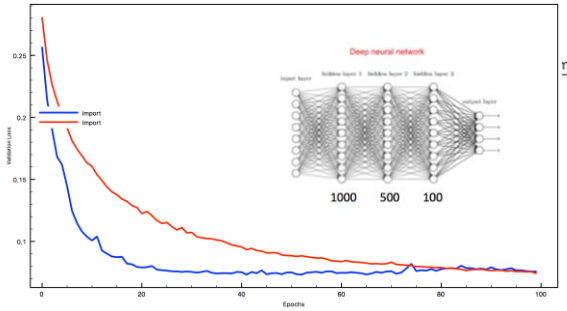
Deliverables: CEED-MS0 milestone report at https://confluence.exascaleproject.org/display/CEED/Milestones

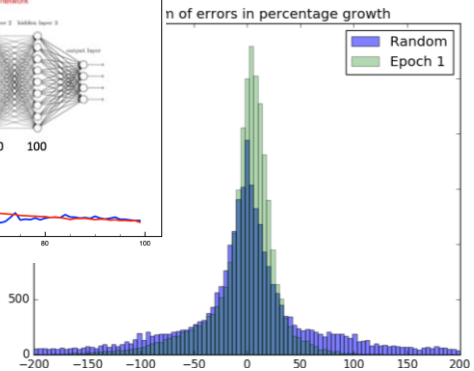
Formulation & Release of Seven Deep Learning Benchmarks from the Cancer Distributed Learning Environment (CANDLE)

ECP WBS 1.2.3.01: CANDLE
PI: Rick Stevens, ANL
Members: ANL, BNL, LANL, LLNL, ORNL, UIUC, NIH / NCI

Scope & Objectives

- DOE Labs are coupling their strengths in HPC, machine learning and data analytics with the domain strengths of the NCI (cancer biology and cancer healthcare delivery) to bring the promise of exascale computing to cancer and precision medicine
- Address 3 top NCI challenges: developing predictive models for drug response, understanding molecular basis of key protein interactions, and automating analysis and extraction of information from cancer patient records to direct optimal cancer treatment strategies
- Define representative problems at the core of the predictive oncology challenge, coded as deep learning (DL) problems selected as the common environment for scripting CANDLE problems





Impact

- Gather in one repository multiple cancer related deep learning problems
- Evolve benchmarks with an increasing problem and data sets to share and test CANDLE's ability to conduct large-scale deep learning
- Benchmarks are the baseline for developing deep hyperparameter searching and large-scale runs on LCF resources

Project Accomplishment and Next Steps

- Released 3 benchmark problems related to drug response prediction (an autoencoder for gene expression, cancer type prediction from SNPs, and a baseline drug response prediction from gene expression and drug descriptors), 2 problems related to the RAS challenge (an autoencoder for molecular states and a state predictor), and 2 problems from the patient trajectory (multi-task DNN for tumor identification and biomedical text)
- Evolve CANDLE with additional DL packages (MXNET, Neon, Pytorch) and dramatically scale-up DL problems; these benchmarks provide specific problem instances to guide benchmark scaling approaches

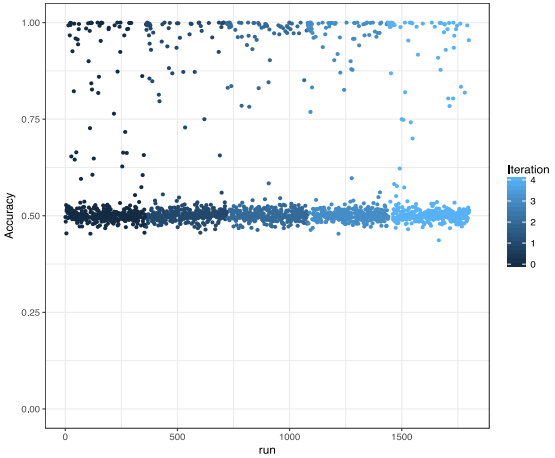
Deliverables: ECP milestone report ADOA01-2 and CANDLE benchmark code at https://github.com/ECP-CANDLE/Benchmarks (2017/03/31).

Formulation & Release of CANDLE Version 0.1 Architecture from the Cancer Distributed Learning Environment (CANDLE)

ECP WBS 1.2.3.01: CANDLE
PI: Rick Stevens, ANL
Members: ANL, BNL, LANL, LLNL, ORNL, UIUC, NIH / NCI

Scope & Objectives

- DOE Labs are coupling their strengths in HPC, machine learning and data analytics to develop scalable deep learning on DOE leadership machines to provide solutions for cancer and precision medicine
- Supports different approaches to the embedded learning problem (i.e., unsupervised, supervised and semi-supervised) with the same scalable deep learning code
- Using representative problems at the core of the predictive oncology challenge—coded as deep learning problems—provides a framework in this first release for large-scale model hyperparameter exploration



Model-based sampling accuracy values for NT3 benchmark on Cori. Samples are 360 design phase (Iteration 0) and samples of 360 for four subsequent iterations (1800 total)

Impact

- Components of the CANDLE architecture enable pluggable hyperparameter optimization and deep neural networks, workflows, data management and visualization
- CANDLE Version 0.0 implements the first version of the Supervisor: a system-level module responsible for orchestrating interactions between hyperparameter optimization algorithms, different implementations of deep neural networks, and the leadership computing resource

Project Accomplishment and Next Steps

- Designed an initial architecture and have begun writing an implementation that will enable model exploration at scale
- The deep neural networks used are examples from the CANDLE benchmarks and address representative problems relating to the three top challenges of the National Cancer Institute
- These also have the additional property of embodying different network architectures
- Future work on CANDLE will move toward data parallelism, thus allowing the training of a single model across several nodes

Deliverable: ECP milestone report ADOA01-6 and CANDLE Version 0.0 code at https://github.com/ECP-Candle/Candle (2017/07/14).

ExaBiome provides first scalable algorithms for high quality metagenome assembly and analysis

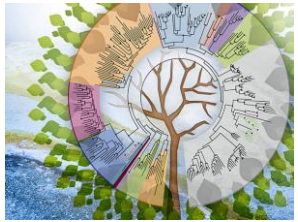
ECP WBS 1.2.1.20: ExaBiome
PI: Kathy Yelick, LBNL
Members: LBNL, LANL, JGI

Scope & Objectives

- Develop **genome assembly, protein clustering and comparative analysis** codes for exascale, using the aggregate memory and high speed networks
- **Assemble millions of metagenomes** without filtering data
- **Cluster billions of proteins** for discovery and to unlock functional behavior
- **Compare thousands of metagenomes** for environmental monitoring and analysis
- Discover new species and functions in large, complex metagenome data sets

Understanding ecosystems

Bio-synthesis



New life forms

Impact

- New scalable metagenome assembler (MetaHipmer) reduces runtime by orders of magnitude at petascale
- Demonstrated MetaHipMer assemblies have quality comparable to best assemblers used in production
- New scalable protein clustering algorithm sped runtime from 15 weeks on single node to 1 hour
- Largest protein clustering from JGI metagenome data

Project Accomplishment and Next Steps

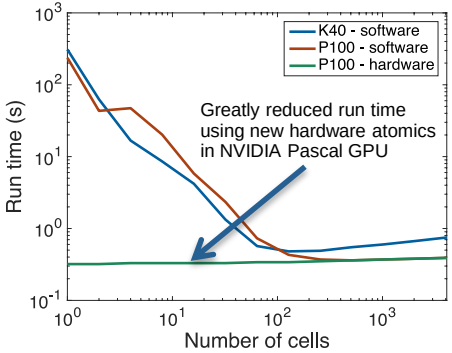
- MetaHipMer code consistently provides the lowest number of mismatches with the highest error-free contiguity compared to state-of-the-art assemblers, including metaSPAdes and MEGAHIT (Milestone 1)
- MetaHipMer scales linearly with cores on KNL (Cori) and to thousands of nodes
- Identified architectural features need for scalable assembly
- Protein clustering (HipMCL) uses communication-avoiding optimizations and novel parallelism for unprecedented performance at petascale (Milestone 2)
- First-of-kind clustering analysis enabled by HipMCL

First Ever Full Core Reactor Simulation with Continuous-Energy Monte Carlo on GPUs

ECP 1,2.1.08: ExaSMR
PI: Tom Evans, ORNL
Members: ORNL, ANL, MIT, INL

Scope & Objectives

- Small Modular Reactor (SMR) Challenge Problems require extreme-resolution models with hundreds of billions of particle trajectories simulated:
  - Requires Monte Carlo (MC) neutron transport solver that can utilize exascale computing resources
- Random access patterns of continuous-energy MC transport is problematic for vectorized computing platforms.
- Goal is to analyze and optimize performance of MC solvers for full core reactor benchmarks on two next-generation computing architectures: the Intel Xeon Phi and the GPU



***GPU Results*:** New Pascal-generation GPU offers built-in double-precision atomic updates. This greatly improves performance when tallying billions of particle trajectories over millions of compute threads.

Fuel Type	Intel Xeon (10^3 n/sec)		Intel Xeon Phi (10^3 n/sec)	
	Inactive	Active	Inactive	Active
Fresh	287	174	67.50	10.30
Depleted	79.2	9.7	13.70	2.63

Phi Results: Improving particle tracking rates is the ultimate measure of performance. Tracking rates are observed to decrease dramatically when tallies are included (so-called "active cycles"), indicating the importance of careful handling of atomic operations. Performance on Phi lags the standard Xeon despite higher theoretical floating-point performance. Calculation simulated 5 million particle histories on a single compute node.

Impact

- First ever full core reactor simulation with continuous-energy MC on GPUs
- Establishes baseline performance for quantifying improvements on existing and future architectures
- Identifies bottlenecks limiting ability to efficiently exploit pre-exascale and exascale systems
- Offers insight into expected performance for other applications with challenging memory access patterns

Project Accomplishment

- Developed GPU version of MC transport code using CUDA programming language
- Performed initial profiling and optimization on both the Phi and GPU architectures
- Improved cache usage/vectorization (Phi) and reduced thread divergence (GPU) determined to be most important areas for future improvement
- Encountered several challenges:
  - A bug in the Intel compiler was discovered and reported to Intel
  - Limited useful information obtained from NVIDIA profiler due to variability between runs

Deliverables: S. Hamilton, T. Evans, P. Romano, and W. Boyd. *Monte Carlo Device Implementation*. WBS 1.2.1.08 ECP-SE-08-2000, 1/18/2017.

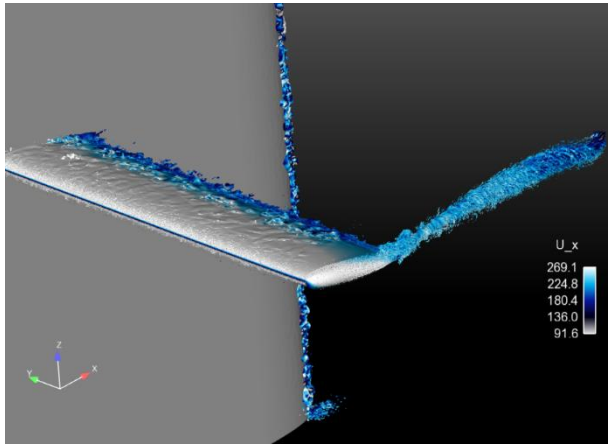
P. Romano, S. Hamilton, T. Evans, J. Liang, and J.S. Perez. *Monte Carlo Full Core Performance Baseline*. WBS 1.2.1.08 ECP-SE-08-36, 7/12/2017.

Blade-Resolved Simulation in Non-Rotating Turbulent Inflow

ECP WBS: 1.2.1.07 ExaWind
PI: Michael Sprague, NREL
Members: NREL, SNL, ORNL, UT

Scope & Objectives

- **ExaWind Objective:** Create a computational fluid and structural dynamics platform for exascale predictive simulations of wind farms
- **Challenge Problem:** Predictive simulation of a wind plant composed of O(100) wind turbines sited over O(100) km^2 with complex terrain
- FY17 Activities focused on
  - Establishing best practices for Nalu code development in multi-institution environment
  - Establishing baseline timing tests for atmospheric-boundary-layer and blade-resolved simulations
  - Deploying Nalu/Kokkos algorithmic infrastructure with performance benchmarking
- This milestone established baseline performance measurements for simulation time, strong scaling, and turbulence models for a fully resolved blade in a non-rotating configuration



Nalu simulation of of the McAlister-Takahashi blade, which has extensive wind-tunnel experimental data. Meshes with 68M and 300M elements were examined, with several turbulence-modeling approaches.

Impact

- Established baseline capabilities for blade-resolved simulations in terms of turbulence modeling and linear-solver performance for realistic meshes
- This performance benchmarking is critical to demonstrating improvements as we move to exascale platforms
- Accurate prediction of blade tip-vortices and wakes is a critical requirement for predictive wind-plant simulations

Project Accomplishment

- In this milestone, we
  - Established and verified turbulence models
  - Established best-performing linear-solver configurations/settings for complex meshes and turbulent flow under MPI-only communication
  - Established baseline timing and strong-scaling performance on Cori under ALCC allocation
- Future work: Improve turbulence models; increase resolution; test performance gains under MPI+OpenMP parallelism via the Kokkos abstraction layer

Deliverables: FY17 Q3 milestone report, "Demonstrate single-blade-resolved simulation in non-rotating turbulent inflow."

Update on Software Technology

EXASCALE COMPUTING PROJECT

Interagency Review of ECP Software Stack

- An interagency review of the ECP software stack was held on June 7, 2017 by the Joint Program Office
- ECP ST leads (Rajeev Thakur and Pat McCormick) presented the ECP Software Technology status and plans to representatives from various other agencies (DoD, NASA, NIST, NOAA, NITRD)
- The review went well
- The review team wanted to be kept informed of ECP progress periodically

Software Technology Gap Analysis

- A draft of the ECP Software Technology Gap Analysis has been completed and sent for review and feedback to various people at the labs (LOTF, CRLC, Facilities) and ECP program managers
- Feedback received will be incorporated and a first version released in a couple of weeks
- The gap analysis is a living document that will be frequently updated as new information is known, such as
  - Details of the software requirements of the first two exascale systems at ALCF and OLCF
  - Details of the vendor software plans for those systems
  - Requirements of new applications selected in areas of data analytics and machine learning
  - New requirements of the existing set of applications
  - Updated information about hardware technologies

Plan for Regular ECP Software Stack Releases

- The ECP Software Technology portfolio aims to be a comprehensive and coherent set of software products designed to work together, robust and production quality, and running on DOE pre-exascale and exascale systems
- A plan for a coordinated release of the ECP software stack has been developed, which includes a process for
  - Testing and continuous integration of the individual components on platforms at DOE facilities
  - Semi-annual releases of the software stack, starting with an initial set of products in FY18
  - Monthly bug-fix releases
- Feedback on the plan is being gathered from ST project participants, after which it will be put into practice

Examples of Milestones Completed

- Improved OpenMP 4.5 implementation in LLVM compiler
- Scalable memory usage in MPI implementation
- New release of HPCToolkit for ECP testbeds (Titan, Mira, Theta, Cori)
- C++ API specification for BLAS and LAPACK math libraries
- API specification for scalable checkpoint-restart
- New prototype in situ algorithms to derive and abstract select data characteristics from data types critical to ECP applications
- Node power abstraction layer for low-level hardware drivers

Update on Hardware Technology

EXASCALE COMPUTING PROJECT

Hardware Technology Activities

- PathForward: support DOE-vendor collaborative R&D activities required to develop exascale systems with at least two diverse architectural features; quote from RFP:
  - PathForward seeks solutions that will improve application performance and developer productivity while maximizing energy efficiency and reliability of exascale systems.
- Design Space Evaluation
  - Apply laboratory architectural analysis capabilities and Abstract Machine Models to PathForward designs to support ECP co-design interactions

PathForward contracts awarded

- Reminder: the ECP PathForward program supports DOE-vendor collaborative R&D activities required to develop exascale systems with at least two diverse architectural features; quote from RFP:
  - PathForward seeks solutions that will improve application performance and developer productivity while maximizing energy efficiency and reliability of exascale systems.
- PathForward contracts were awarded to these companies:
  - Advanced Micro Devices (AMD)
  - Cray Inc. (CRAY)
  - Hewlett Packard Enterprise (HPE)
  - International Business Machines (IBM)
  - Intel Corp. (Intel)
  - NVIDIA Corp. (NVIDIA)

ALCC allocation for ECP activities

2017 ALCC Allocation Amount: **969,000,000 processor hours**

Oak Ridge Leadership Computing Facility	300,000,000	Titan-core hours
Argonne Leadership Computing Facility	400,000,000	Mira-core hours
	130,000,000	Theta-core hours
National Energy Research Scientific Computing Center	14,000,000	Cori Phase 1/ Edison NERSC-core hours
	125,000,000	Cori Phase 2 NERSC-core hours

Upcoming events

- The ALCF has selected its first exascale system and in the next N months OLCF, and LLNL will have selected their first exascale systems
- Those selections will provide detailed information about the characteristics of those systems that will be used to further focus the ECP activities
- Independent Project Review January 9-11, 2018
- Second annual meeting in Knoxville February 6-8, 2018

Summary

- The ECP is making excellent progress
- There will continue to be substantial management and technical challenges – that is why the ECP is a project – but I am confident they will be tackled effectively

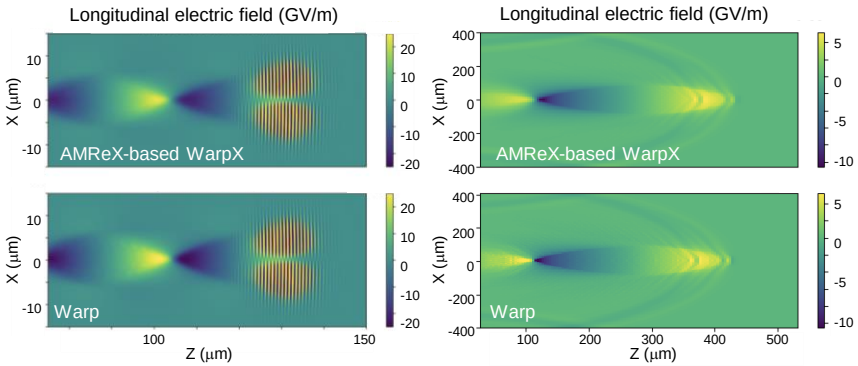
A few more application highlights

Port of WarpX to AMReX

ECP WBS: 1.2.1.06 WarpX
PI: J.-L. Vay, LBNL
Members: SLAC, LLNL

Scope & Objectives

- **Goal:** speedup development of particle accelerators that are significantly shorter and cheaper.
- **Challenge:** modeling in 3-D of up-to 100 plasma accelerator stages.
- **Year 1:** 1) couple Warp+Boxlib+PICSAR, 2) demonstrate coupling, 3) add mesh refinement, 4) demonstrate new code with mesh refinement.
- **Main drivers:** implement and test the fundamental building blocks for mesh refinement in WarpX, starting with a single level of refinement, in preparation for application of mesh refinement to the modeling of plasma accelerators.



Validation of AMReX-based version of WarpX (top) against Warp (bottom) on the modeling of plasma wakefields generated by lasers (left) or particle beams (right).

Impact

- New code has broad application across laboratory and space computational plasma and beam physics.
- Within ECP, this accomplishment establishes the basis of a relativistic electromagnetic Particle-In-Cell code with high performance, scalability and AMR.
- DOE HEP plasma accelerator projects need fast, scalable tools. It will also benefit BES, FES and NNSA activities.

Project Accomplishment

- Converted WarpX data structure to conform to AMReX needs.
- Adapted subroutine calls to AMReX.
- Tested implementation on unit tests for key subroutines and on standard test problems.
- Validated transition of WarpX from BoxLib to AMReX on modeling of plasma wakefields.
- Released tagged version of AMReX-based WarpX.

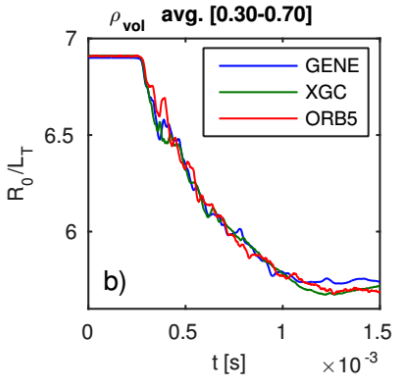
Deliverables: tagged version of WarpX on Bitbucket repository.

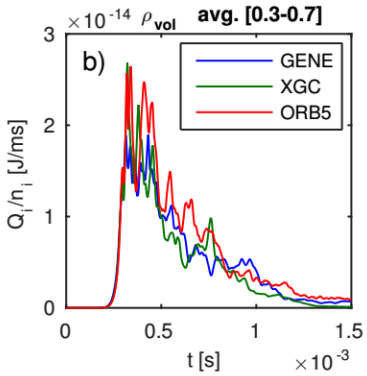
Successful GENE and XGC Benchmark Comparison of the ITG Turbulence Regime

ECP WBS: 1.2.1.12 Whole Device Modeling
PI: A. Bhattacharjee, PPPL
Institutions : PPPL, UCLA, UC-Boulder

Scope & Objectives

- Create a Whole Device Model of a tokamak with the ultimate goal of simulating future fusion reactors, including ITER
- Couple an Eulerian core code, GENE, with a Lagrangian edge code XGC, to simulate plasma turbulence across the entire device
- Benchmark the two codes in a realistic challenge problem
- Establish the scaling properties of the two codes
- Provide a robust basis and identify potential risks





Excellent agreement is shown between GENE and XGC results a nonlinear simulation of turbulence relaxation.

Impact

- This is the first benchmark exercise between GENE and XGC (also joined by the ORB5 code)
- This provides a verified starting point for coupling the codes together
- Produced a well documented benchmark test case available for the fusion community

Project Accomplishment

- First benchmark between the GENE and XGC codes
- Successfully verified the codes in ITG turbulent regime relevant for the coupling
- Identified subtle differences in the numerical schemes which are to be accounted in order to accurately couple the codes
- This accomplishment is the necessary basis for the up-coming next steps of the project

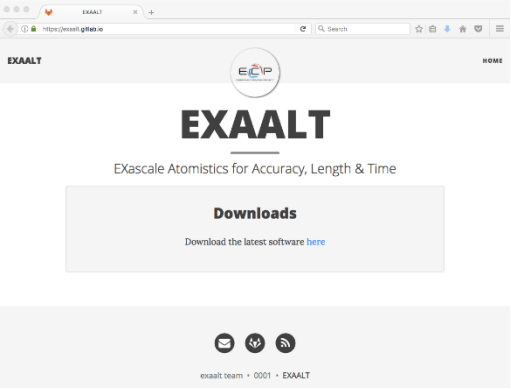
Deliverables: ECP-WDM AD Y1-Q1, ECP-WDM AD Y1-Q2, S. Ethier Talk at PASC 2017 Lugano Switzerland, and paper to be submitted to Phys. Of Plasmas.

First Release of the EXAALT Package

ECP WBS: 1.2.1.04 EXAALT
PI: Arthur F. Voter, LANL
Members: LANL, SNL, UTK

Scope & Objectives

- The goal of EXAALT is to provide a scalable atomistic simulation capability that can access the whole accuracy/size/time simulation space theoretically accessible at the exascale.
- We target material's problems in nuclear energy:
  - Evolution of first-wall materials in fusion reactors
  - Evolution of defects in nuclear fuels
- First FY concentrates on integrating the codes and building capability-base (interatomic potentials) that can be used to support further development and start addressing challenge problems



Snapshot of EXAALT's software distribution site. EXAALT's code-base for the first time integrates the three major simulation capabilities (long-time, large-size, high accuracy).

Impact

- Can analyze performance of the integrated code; this will drive further developments
- Can start sharing the code with external and external collaborators. Will facilitate evaluation of different software solutions.
- Code can be used "in the wild" by alpha users. This will help improve robustness and robustness
- Can be used by to start addressing challenge problems

Project Accomplishment

- Integration of the long-time Accelerated MD capabilities (LANL), Massively-parallel MD (SNL), and Quantum-based DFTB MD (LANL)
- Can now access very large areas of the accuracy/size/time simulation space
- Inputs from the software productivity team to manage integrated repository
- This is the codebase that will be the cornerstone of all further development

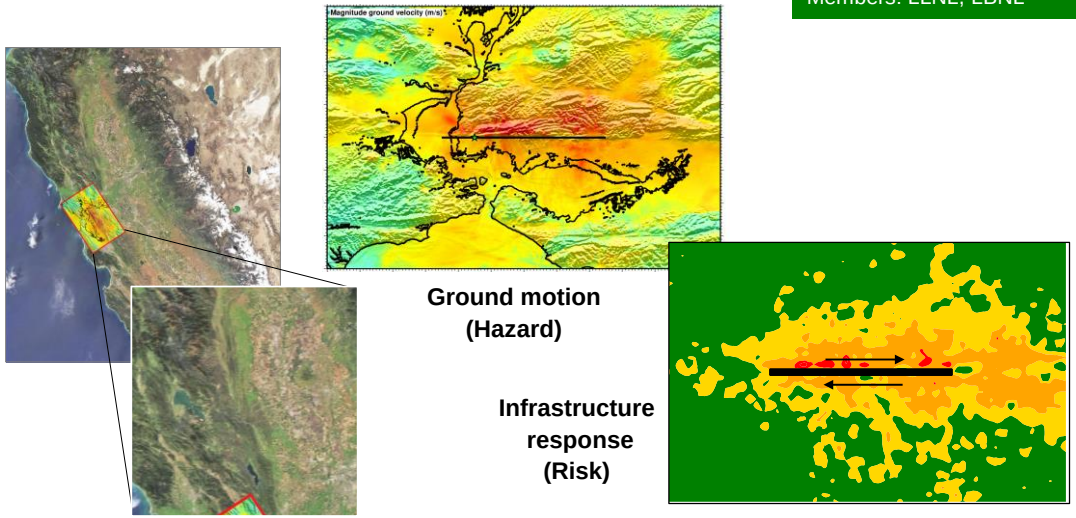
- **Deliverables:** Initial distribution available at https://exaalt.gitlab.io/

First Coupled, Regional-Scale Simulation of EQ *Hazard* and *Risk*

ECP WBS:1.2.1.19 EQSIM
PI: David McCallen LBNL/ UC
Members: LLNL, LBNL

Scope & Objectives

- Overarching goal is to develop a simulation capability for physics-based computations of seismic hazard and risk
- Resolving frequencies of engineering interest (0 – 10 Hz) and providing demonstrably realistic and accurate estimations of hazard and risk is essential
- It will be essential to include computational features for characterizing complex geologic heterogeneities
- This will support the thrust for transformational hazard and risk assessments that better account for site-specific features of ground motions



Using the seismically active San Francisco Bay Area as a numerical laboratory, simulations of earthquake hazard and risk have been successfully completed for a M=7 Hayward fault event

Impact

- These results are establishing a global computational approach to combined hazard and risk
- Successful resolution to 5-10 Hz will enable significant new science and engineering understanding
- Major implications for DOE's critical infrastructure ($5B), energy systems infrastructure, and many other infrastructure types (transportation, commercial etc.)

Project Accomplishment

- Completed a 60 billion grid point ground motion simulation resolving 0 - 2.5 Hz
- Translated the hazard simulation into a risk simulation for infrastructure risk
- Utilized SW4 code ported and improved for CORI-II platform
- Looking ahead - Implementing mesh refinement and robust code restart will provide a path for utilizing 8192 nodes in a 240 billion billion simulation in a push for 0 - 5 Hz

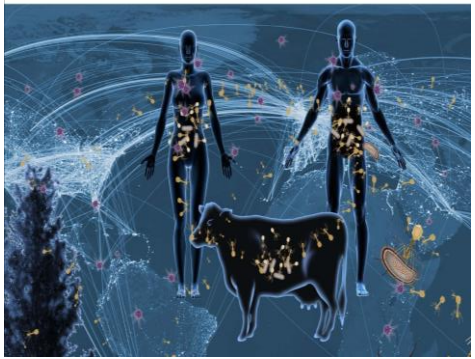
Deliverables: Third quarter ECP project milestone report summarizes these developments

MetaHipMer: High Metagenome Assembly Quality & Scalability

ECP WBS: 1.2.1.20 Exabiome
PI: Kathy Yelick, LBNL
Members: LBNL, JGI, LANL

Exabiome Scope & Objectives

- Metagenomic datasets are large and rapidly growing, and future assembly and analysis of these will require exascale resources.
- The Exabiome project will provide scalable tools for core metagenomics computations: assembly (MetaHipMer), protein clustering (HipMCL), and comparative analysis (GOTTCHA)
- Recently we developed MetaHipMer assembler and demonstrated quality comparable to other state-of-the-art assemblers.
- Current FY plan: demonstrate HipMCL clustering algorithm, demonstrate MetaHipMer performance, evaluate merAligner for GOTTCHA, release MetaHipMer software.



Metagenomics: genome sequencing communities of 1000s of microbial species. Key in environment, climate, agriculture, health and biomanufacturing.

Impact of MetaHipMer Quality Milestone

- MetaHipMer is of sufficient quality to be a viable alternative to current state-of-the art assemblers, and could be widely used by biologists for research into metagenomes.
- This is an important milestone because it demonstrates that the algorithms used are sufficiently accurate to form the basis for an assembler that can run on future exascale systems.

Project Accomplishments demonstrating Quality

- **MetaHipMer quality is as good as and often better than the best state-of-the-art assemblers**, including metaSPAdes and MEGAHIT
- For several commonly used evaluation datasets, MetaHipMer consistently provides the lowest number of mismatches with the highest error-free contiguity
- MetaHipMer can run efficiently on large-scale HPC systems, **reducing assembly times by orders of magnitude compared to other assemblers**. This enables more aggressive assembly approaches and accelerates scientific discovery.

Thank you!

www.ExascaleProject.org

EXASCALE COMPUTING PROJECT