Computation-Driven Discovery for the Dark Universe

Overview:

Project Director:
Salman Habib (ANL)

Institutional PIs:
James Ahrens (LANL)
Scott Dodelson (FNAL)
Katrin Heitmann (ANL)
Peter Nugent (LBNL)
Anze Slosar (BNL)
Risa Wechsler (SLAC)

Key Personnel:
Tom Abel (SLAC)
Ann Almgren (LBNL)
Joanne Cohn (UC Berkeley)
Hal Finkel (ANL)
Nick Gnedin (FNAL)
Mark Hereld (ANL)
David Higdon (LANL)
Nishikanta Khandai (BNL)
Eve Kovacs (ANL)
Juliana Kwan (ANL)
Robert Latham (ANL)
Earl Lawrence (LANL)
Zarija Lukic (LBNL)
Tom Peterka (ANL)
Adrian Pope (ANL)
Martin White (LBNL)
Jonathan Woodring (LANL)

The Cosmic Frontier presents a broad array of fundamental discovery opportunities related to dark energy, dark matter, the masses of neutrinos and their hierarchy, and unique probes of inflation and the early Universe. Large-scale surveys of the sky are collecting massive amounts of data that hold the key to a deeper understanding of the Universe. High performance computing is a powerful tool of discovery in extracting insights and making precision predictions for these observations. Next-generation surveys require a much tighter coupling to theory and significantly more end-to-end simulation capability than previous generations of experiment, bringing cosmology in line with the traditions of particle physics. This SciDAC-3 Partnership project brings together a National Laboratory-based collaboration – in partnership with the SciDAC Institutes – to build next-generation computational cosmology prediction and analysis frameworks that are directly targeted to current and future observations. We will make full use of DOE's Leadership class systems at the Argonne Leadership Computing Facility, the Oak Ridge Leadership Computing Facility, and the National Energy Research Scientific Computing Center.

The timeliness of this project is particularly notable as observations and high performance computing are both undergoing a period of rapid change, and will expand significantly in capability and range over the next several years; it is widely recognized that the paradigms of today will not be the paradigms of tomorrow. Driven by this imperative, our project will support the further development of three large-scale high-performance cosmology simulation codes, along with sophisticated analysis frameworks, including the solution of high-dimensional problems of statistical inference. These powerful computational 'tools of discovery' will be focused on developing and sharpening cosmological probes in the domains of optical galaxy surveys (both imaging and spectroscopic), weak gravitational lensing, and the Ly-alpha forest, all highly relevant to the DOE-supported surveys BOSS (Baryon Oscillations Spectroscopic Survey), DES (Dark Energy Survey), and LSST (Large Synoptic Survey Telescope).

Scope of work:

The project focuses on the further development of the parallel codes ART, HACC, and Nyx. ART and Nyx are AMR cosmological hydro simulation codes incorporating a variety of subgrid models, while HACC is a cosmological N-body code framework. ART (Adaptive Refinement Tree) is a proven and relatively mature AMR code – the main task in this project is to improve its scaling properties to beyond tens of thousands of cores. ART will be used to study the baryonic effects that influence weak lensing shear measurements. Nyx is a new code being developed using the block-structured BoxLib AMR framework supported by the FASTMath SciDAC Institute; it will initially be deployed for studies of the Lyman-alpha forest. The HACC (Hardware/Hybrid Accelerated Cosmology Codes) framework is a recently developed hybrid particle/grid code that has demonstrated scaling to ~750K cores at very high performance levels. Large-volume simulations with HACC will be used to obtain predictions for a variety of cosmological probes.

Because scientific inference in cosmology is in effect an inverse problem, a significant role for high-performance computing is to produce forward model predictions within a Markov chain Monte Carlo (MCMC) framework, where the predictions must be generated tens to hundreds of thousands of times. The complexity of a single prediction, both in terms of physics and numbers of parameters, obviously precludes brute force simulation runs as a viable approach. To overcome this problem, project members have recently developed the 'Cosmic Calibration Framework' (CCF) in close collaboration with participants in the QUEST SciDAC Institute. The main idea behind the framework is to cover the model space in an efficient manner by using sophisticated statistical sampling methods and techniques for interpolation over high-dimensional spaces. The power and importance of the CCF can be appreciated by realizing that running 10,000 - 100,000 cosmological models with full simulations for a MCMC analysis would take several decades with current computational power. With the new methodology, the time for the entire MCMC analysis is brought down to a matter of hours. Working with the QUEST statisticians, a key project task is to extend the capabilities of the CFCC in several new directions, and to carry out careful studies to understand the related UQ and V&V issues.

Finally, an important problem with the large-scale simulation capabilities that form the core of this project is in dealing with the large datasets that are routinely generated. In collaboration with members of the SDAV SciDAC Institute, on the fly (in situ) analysis techniques will be developed as part of this project, as also methods for better I/O performance. Remote visualization and visualization-aided analysis methods will be another focus of this work.

Challenges:

Broadly speaking, the mathematical challenges for the project relate to the development of methods for solving the associated dynamical equations at extreme scales, and in finding new analysis methods for large datasets. The UQ and V&V challenges are due to a combination of difficult problems in machine learning and statistics. Computer science challenges underlie many parts of the project, e.g., efficient implementations of parallel hybrid algorithms, new time integration methods, code scalability and performance optimization, new algorithms for data analysis, large I/O problems, and new visualization methods. Solving and combining all of these in one package to produce a capability that can exploit extreme scale architectures is the overall challenge.