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.