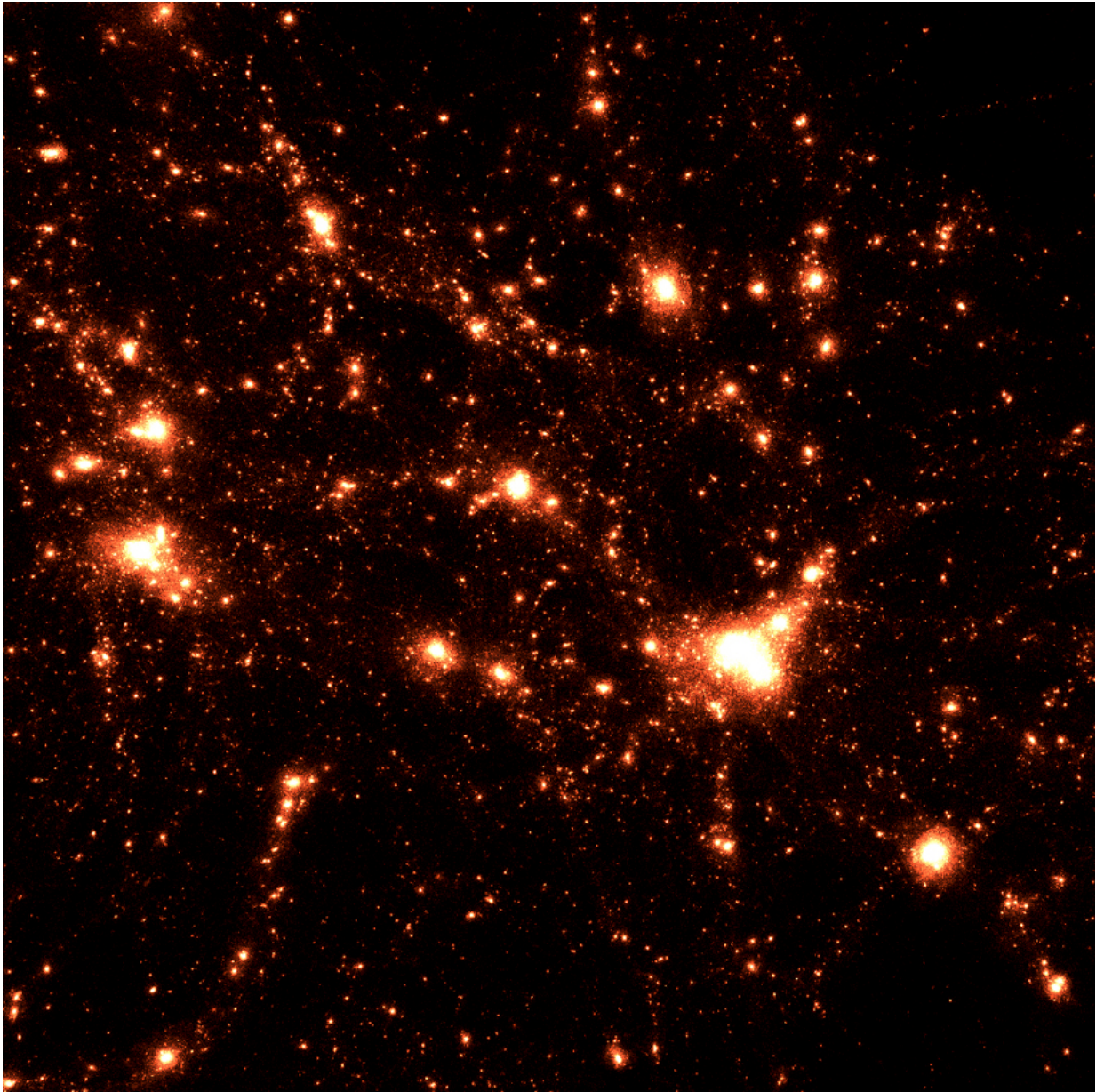# Cartography of the cosmos

September 27 2017

This shows the HACC cosmology simulation, which combines high spatial and

temporal resolution in a large cosmological volume. The high temporal resolution tracks the evolution of structures in great detail and correlates formation histories to the environments in which the structures form. Credit: Argonne National Laboratory

There are hundreds of billions of stars in our own Milky Way galaxy. Estimates indicate a similar number of galaxies in the observable universe, each with its own large assemblage of stars, many with their own planetary systems. Beyond and between these stars and galaxies are all manner of matter in various phases, such as gas and dust. Another form of matter, dark matter, exists in a very different and mysterious form, announcing its presence indirectly only through its gravitational effects.

This is the universe Salman Habib is trying to reconstruct, structure by structure, using precise observations from telescope surveys combined with next-generation data analysis and simulation techniques currently being primed for exascale computing.

"We're simulating all the processes in the structure and formation of the universe. It's like solving a very large physics puzzle," said Habib, a senior physicist and computational scientist with the High Energy Physics and Mathematics and Computer Science divisions of the U.S. Department of Energy's (DOE) Argonne National Laboratory.

Habib leads the "Computing the Sky at Extreme Scales" project or "ExaSky," one of the first projects funded by the recently established Exascale Computing Project (ECP), a collaborative effort between DOE's Office of Science and its National Nuclear Security Administration.

From determining the initial cause of primordial fluctuations to measuring the sum of all neutrino masses, this project's science objectives represent a laundry list of the biggest questions, mysteries and challenges currently confounding cosmologists.

There is the question of dark energy, the potential cause of the accelerated expansion of the universe, while yet another is the nature and distribution of dark matter in the universe.

These are immense questions that demand equally expansive computational power to answer. The ECP is readying science codes for exascale systems, the new workhorses of computational and big data science.

Initiated to drive the development of an "exascale ecosystem" of cutting-edge, high-performance architectures, codes and frameworks, the ECP will allow researchers to tackle data and computationally intensive challenges such as the ExaSky simulations of the known universe.

In addition to the magnitude of their computational demands, ECP projects are selected based on whether they meet specific strategic areas, ranging from energy and economic security to scientific discovery and healthcare.

"Salman's research certainly looks at important and fundamental scientific questions, but it has societal benefits, too," said Paul Messina, Argonne Distinguished Fellow. "Human beings tend to wonder where they came from, and that curiosity is very deep."

HACC'ing the night sky

For Habib, the ECP presents a two-fold challenge—how do you conduct cutting-edge science on cutting-edge machines?

The cross-divisional Argonne team has been working on the science through a multi-year effort at the Argonne Leadership Computing Facility (ALCF), a DOE Office of Science User Facility. The team is running cosmological simulations for large-scale sky surveys on the facility's 10-petaflop high-performance computer, Mira. The simulations are designed to work with observational data collected from specialized survey telescopes, like the forthcoming Dark Energy Spectroscopic Instrument (DESI) and the Large Synoptic Survey Telescope (LSST).

Survey telescopes look at much larger areas of the sky—up to half the sky, at any point—than does the Hubble Space Telescope, for instance, which focuses more on individual objects. One night concentrating on one patch, the next night another, survey instruments systematically examine the sky to develop a cartographic record of the cosmos, as Habib describes it.

Working in partnership with Los Alamos and Lawrence Berkeley National Laboratories, the Argonne team is readying itself to chart the rest of the course.

Their primary code, which Habib helped develop, is already among the fastest science production codes in use. Called HACC (Hardware/Hybrid Accelerated Cosmology Code), this particle-based cosmology framework supports a variety of programming models and algorithms.

Unique among codes used in other exascale computing projects, it can run on all current and prototype architectures, from the basic X86 chip used in most home PCs, to graphics processing units, to the newest Knights Landing chip found in Theta, the ALCF's latest supercomputing system.

As robust as the code is already, the HACC team continues to develop it further, adding significant new capabilities, such as hydrodynamics and

associated subgrid models.

"When you run very large simulations of the universe, you can't possibly do everything, because it's just too detailed," Habib explained. "For example, if we're running a simulation where we literally have tens to hundreds of billions of galaxies, we cannot follow each galaxy in full detail. So we come up with approximate approaches, referred to as subgrid models."

Even with these improvements and its successes, the HACC code still will need to increase its performance and memory to be able to work in an exascale framework. In addition to HACC, the ExaSky project employs the adaptive mesh refinement code Nyx, developed at Lawrence Berkeley. HACC and Nyx complement each other with different areas of specialization. The synergy between the two is an important element of the ExaSky team's approach.

A cosmological simulation approach that melds multiple approaches allows the verification of difficult-to-resolve cosmological processes involving gravitational evolution, gas dynamics and astrophysical effects at very high dynamic ranges. New computational methods like machine learning will help scientists to quickly and systematically recognize features in both the observational and simulation data that represent unique events.

A trillion particles of light

The work produced under the ECP will serve several purposes, benefitting both the future of cosmological modeling and the development of successful exascale platforms.

On the modeling end, the computer can generate many universes with different parameters, allowing researchers to compare their models with

observations to determine which models fit the data most accurately. Alternatively, the models can make predictions for observations yet to be made.

Models also can produce extremely realistic pictures of the sky, which is essential when planning large observational campaigns, such as those by DESI and LSST.

"Before you spend the money to build a telescope, it's important to also produce extremely good simulated data so that people can optimize observational campaigns to meet their data challenges," said Habib.

But the cost of realism is expensive. Simulations can range in the trillion-particle realm and produce several petabytes—quadrillions of bytes—of data in a single run. As exascale becomes prevalent, these simulations will produce 10 to 100 times as much data.

The work that the ExaSky team is doing, along with that of the other ECP research teams, will help address these challenges and those faced by computer manufacturers and software developers as they create coherent, functional exascale platforms to meet the needs of large-scale science. By working with their own codes on pre-exascale machines, the ECP research team can help guide vendors in chip design, I/O bandwidth and memory requirements and other features.

"All of these things can help the ECP community optimize their systems," noted Habib. "That's the fundamental reason why the ECP science teams were chosen. We will take the lessons we learn in dealing with this architecture back to the rest of the science community and say, 'We have found a solution.'"

Provided by Argonne National Laboratory