SDSC: Envisioning the Future
Research Advances 2007

*by Paul Tooby*

# Looking Back in Time

## SDSC computing and data resources power most detailed simulations of the early universe, helping astrophysicists understand key observations

Unlike most scientists, astronomers have a way to do "time travel," literally seeing billions of years back into the earliest history of the universe. A number of astronomical surveys such as the Sloan Digital Sky Survey are taking this trip, recording observations of objects in one section of the sky that are ever farther away—and therefore older—as their light travels longer distances and billions of years to reach the earth.

Hidden in the details of these surveys—counts of galaxies and galaxy clusters and the statistics of their scale and distribution—are clues to the underlying mechanisms that gave rise to this large-scale structure.

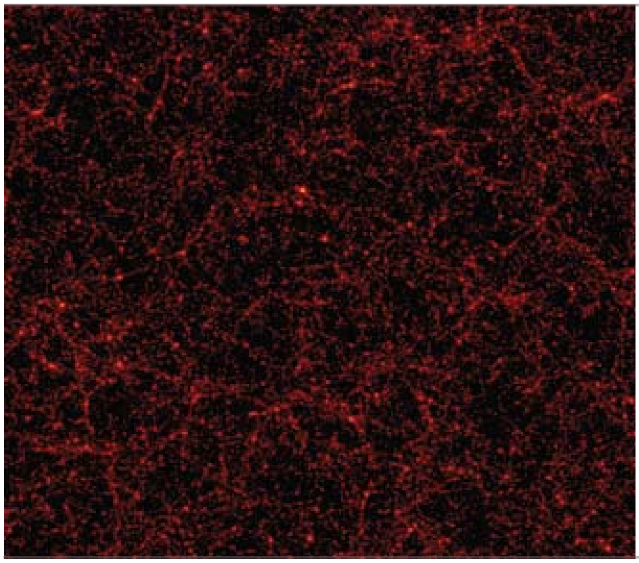
To help understand these observations, UC San Diego astrophysicist Michael Norman and collaborators are using the DataStar system at the San Diego Supercomputer Center (SDSC) at UC San Diego to run the ENZO cosmology code in simulations of the universe from first principles, starting near the Big Bang. In work accepted for publication in the *Astrophysical Journal*, the researchers have conducted the most detailed simulations ever of a significant region of the universe 500 megaparsecs across (more than 2 billion light years).

### Heating Things Up

The large-scale simulations make it possible for scientists to better understand factors related to both theories and observations of the formation of the universe, helping shed light on where we came from and how we've reached the galaxies, stars, and planets of today.

For example, the great detail of these ENZO simulations over a large region and range of scales lets researchers estimate more accurately than previously possible a slight heating of the cold cosmic microwave background—the remnant heat left over from the Big Bang—as it travels through the hot intergalactic gas present in the galaxy clusters that form at the intersections of cosmic filaments (visible as small bright knots in the images).

Being able to estimate this heating is important because when astronomers point a radio telescope at galaxy clusters in an astronomical survey that looks back into the universe's early history, they can observe this small heating effect. Although tiny, just a fraction of a percent, its telltale signature gives astrophysicists a wealth of useful information about galaxy cluster formation, content, growth, and evolution.



**Cosmic Microwave Background**
A map of the slight distortion in the cosmic microwave background as it travels through hot intergalactic gas in galaxy clusters, which gives astrophysicists a window into how clusters form and evolve. The ENZO simulations followed dark and normal matter condensing over five orders of magnitude in space and for 13.7 billion years from near the Big Bang to the present. Galaxy clusters are visible as bright white spots, with cosmic filaments traced by galaxies as lighter streamers. Image: Brian O'Shea, LANL.

"But it's not easy to unravel this small effect in the observations and understand the important information it can give us about things like galaxy cluster size," said astrophysicist Brian O'Shea of Los Alamos National Laboratory, who ran the simulations with SDSC computational scientist Robert Harkness. "Our simulations give astrophysicists an important tool to more accurately interpret what they're seeing, and make it possible to learn more about the structure of the universe through exciting new observational campaigns that use cutting-edge X-ray, radio, and optical telescopes."

The group's ENZO simulations play an important role in complementing observations. But the value of the computations relies on keeping the model faithful to reality, and to do this the researchers need to represent the extreme variability of matter as it coalesces under gravity, becoming many orders of magnitude more dense in local areas as it forms cosmic filaments and galaxy clusters.

## Zooming In

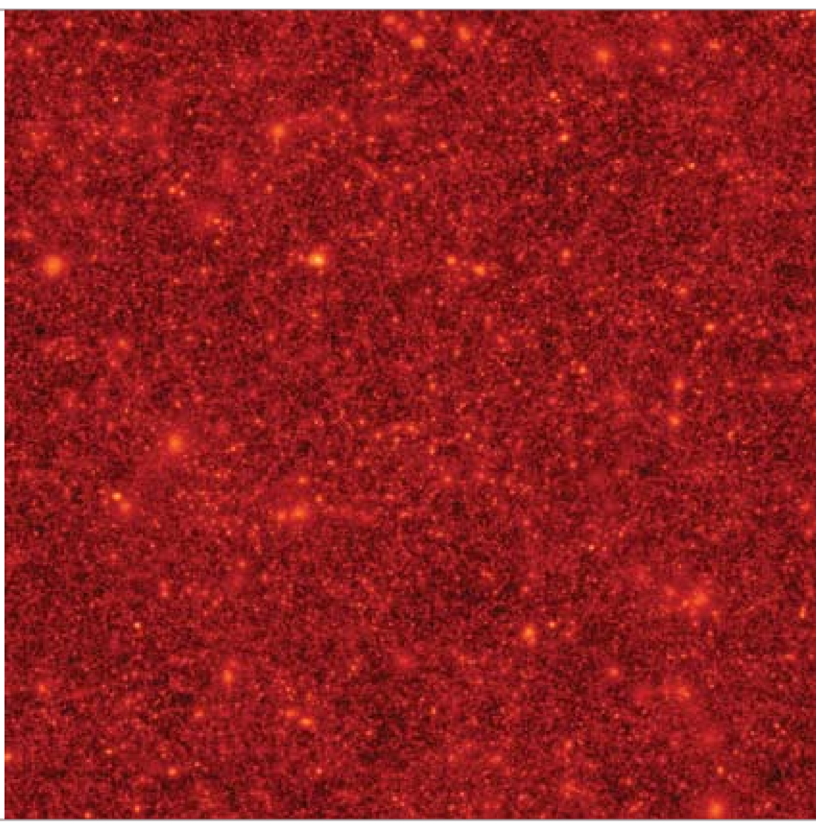
"We need to zoom in on these dense regions to capture the key physical processes—including gravitation, flows of normal and 'dark' matter, and shock heating and radiative cooling of the gas," said Norman, who is director of UCSD's Laboratory for Computational Astrophysics and Acting Deputy Director of Computational Sciences at SDSC. "This requires ENZO's adaptive mesh refinement capability."

Adaptive mesh refinement (AMR) cosmology simulations begin with a coarsely-spaced grid, and then spawn more detailed (and more computationally demanding) subgrids to track the process as gas collapses into high-density structures such as galaxies.

"We achieved unprecedented detail by reaching seven levels of adaptive mesh subgrids throughout the survey volume, something never done before," said SDSC's Harkness, who has long worked on development of the ENZO code. "This simulation utilized more than 400,000 subgrids at a time to model the evolving structures, which we could only do thanks to the two large-memory TeraGrid systems."

Running the code for a total of about 500,000 processor hours, the researchers used 1.5 terabytes of shared memory on the SGI Altix Cobalt system at the National Center for Supercomputing Applications (NCSA) for the initial part of the problem, with the bulk of the simulations then running on 2 terabytes of memory on SDSC's IBM DataStar, which has 32 gigabytes of memory per node.

To make it possible to attempt these computations, Harkness has participated in an ongoing SDSC Strategic Applications Collaboration with Norman's group to achieve several major improvements in scaling and efficiency of the ENZO code. The ability of the code to scale up has been increased by a factor of 512 in spatial resolution to a $2{,}048^3$ domain with more than 8 billion particles for some simulations, while also gaining the ability to scale up to run on 2,048 processors of DataStar. The effort also eliminated data input/output as a bottleneck, resulting in speed gains of up to 120 times on some systems.

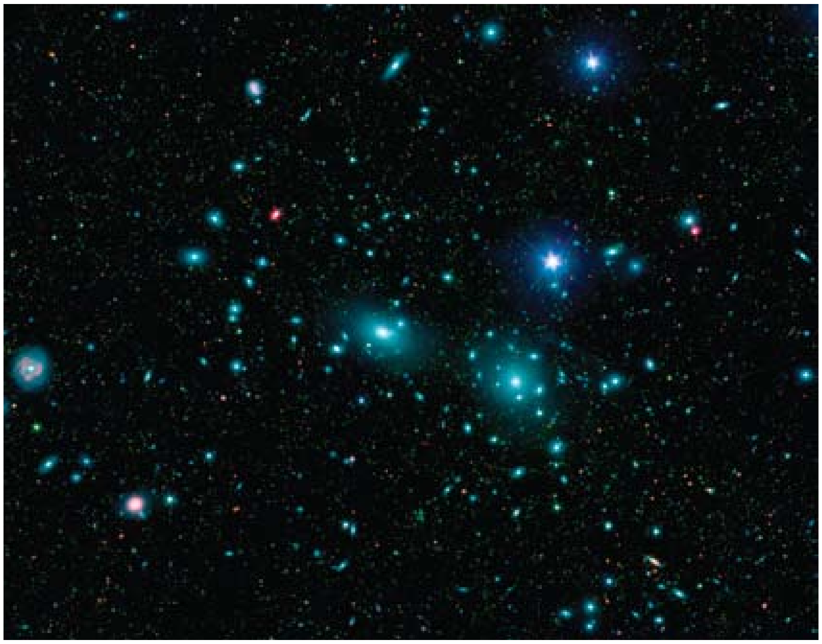


**Cosmological Light Cone**
A simulation of what a radio telescope will see when surveying a wide swath of sky 10 degrees square, which can help telescope builders calibrate their instruments and understand their observations. In the most detailed simulations ever using the ENZO code on SDSC's DataStar system, the effect of all cosmic structure from two billion years after the Big Bang to the present is added up, giving the "light cone" a length of over ten billion light years. Bright white spots are groups and clusters of galaxies; red background is from filamentary structure. Image: Brian O'Shea, LANL.

## Data for the Stars

As simulations like ENZO progress to larger scale and greater resolution while incorporating more sophisticated physics, they push the envelope not only of supercomputing power but also of data capabilities, generating some 10 terabytes of data for the current ENZO simulations alone. SDSC's robust data storage environment with 1.4 petabytes (a petabyte is one million gigabytes, equivalent to the storage on about 10,000 laptops) of online disk storage and 25 petabytes of archival tape storage—the largest of any academic institution in the world—allowed the researchers to efficiently store and manage the massive data collection. The Hierarchical Data Format (HDF) group at NCSA also provided support for handling the output.

Overall, the ENZO researchers have assembled one of the largest simulation data collections at SDSC and in computational astrophysics, more than 250 terabytes (more than a dozen times the digital text equivalent of the printed collection of the Library of Congress). Other computational astrophysicists also generate large data collections, and to help them manage, share, and analyze these data sets on a community basis, experts in SDSC's DataCentral have worked with Norman and colleagues to establish the Computational Astrophysics Data Analysis Center (see sidebar) at SDSC.



**Coma Galaxy Cluster**
As astronomers work to unravel the mysteries of how galaxy clusters formed, ENZO simulations help them understand their observations and basic cosmological parameters such as the amount of baryons (hydrogen, helium, metal), dark matter, and dark energy in the universe. This mosaic combines visible-light data from the Sloan Digital Sky Survey (blue) with long- and short-wavelength infrared views (red and green, respectively) from NASA's Spitzer Space Telescope. Credit: NASA/JPL-Caltech/GSFC/SDSS.

### DATA CYBERINFRASTRUCTURE FOR ASTROPHYSICS

Research in computational astrophysics involves massive, multi-terabyte data sets, but a lack of access to large-scale data and computational resources often hinders early publication and sharing of these large data collections, slowing scientific progress. To bring the power of data cyberinfrastructure to the astrophysics community, UCSD computational astrophysicists Michael Norman and Paolo Padoan and UCSD's Laboratory for Computational Astrophysics have collaborated with SDSC to develop the Computational Astrophysics Data Analysis Center (CADAC), a service that collects and stores results of large astrophysical simulations and provides data analysis resources to researchers worldwide. Through this SDSC-based data cyberinfrastructure, CADAC encourages early publication and sharing of complete numerical datasets, fostering scientific collaboration, increasing the impact of numerical experiments, and facilitating the review process for papers based on computational simulations. With the help of experts in SDSC's DataCentral, led by Natasha Balac, CADAC data is managed by the SDSC Storage Resource Broker (SRB) in a powerful system that combines file storage on disk with automatic replication of data to tape in SDSC's 25 petabyte-capacity High Performance Storage System (HPSS), as well as access to supercomputing resources for data analysis. The collection already has a 100 terabyte disk allocation, and may expand to 300 terabytes.

### Project Participants

Eric Hallman and Jack Burns, U. Colorado; Brian O'Shea, LANL; Michael Norman and Rick Wagner, UCSD; and Robert Harkness, SDSC/UCSD

### Related Links

ENZO cosmology simulation code http://lca.ucsd.edu/portal/software/enzo.

Computational Astrophysics Data Analysis Center (CADAC) http://cadac.sdsc.edu/index.html