
The guide to the Santa Fe Light Cone Simulation research project files RCIDC.0001

Finding aid prepared by Finding aid prepared by Rick Wagner and Bradley Westbrook
UC San Diego Office of the Chancellor.
UC San Diego Library

Title: Santa Fe Light Cone Simulation research project files

Identifier/Call Number: RCIDC.0001

<https://doi.org/10.5060/D2DN4300>

Contributing Institution: UC San Diego Library

Language of Material: English

Physical Description: 683.0 Gigabyte(s)39 digital objects collectively containing 1,797 digital files of various types.

Date (bulk): Bulk, 2005-2007

Date (inclusive): 2005-2012, Bulk 2005-2007

creator: Burns, Jack O.

creator: Hallman, Eric J.

creator: Harkness, Robert

creator: Norman, Michael L.

creator: O'Shea, Brian W., (Brian William), 1978-

creator: So, Geoffrey

creator: Wagner, Rick, 1972-

Cite This Work

Wagner, Rick; Hallman, Eric J.; O'Shea, Brian W.; Burns, Jack O.; Norman, Michael L.; Harkness, Robert; So, Geoffrey (2013): The Santa Fe Light Cone Simulation research project files. Data version 1.0. UC San Diego Library Digital Collections. <https://doi.org/10.5060/D2DN4300>

Project Background

The Santa Fe Light Cone Simulation project was the result of an ongoing effort by the Laboratory for Computational Astrophysics, beginning with the LUSCID Project in 2005. This led to the development of the ENZO simulation software to the point where it was able to complete a seven-level adaptive mesh refinement (AMR) cosmology simulation.

LUSCID

During the 1990s, observational cosmology became “big science,” involving expensive instruments (e.g., the Hubble Space Telescope) and large teams (e.g., the Sloan Digital Sky Survey [SDSS]) attacking fundamental questions about the origin and evolution of the universe. Progress was astonishing and included the discovery of the accelerating universe (Riess et al. 1998, Perlmutter et al. 1999); precision measurements of the global geometry, age, and composition of the universe (de Bernardis et al. 2000); and deep images of galaxies at the dawn of time (Beckwith et al. 2004). These and other observations have narrowed the range of acceptable theoretical models for cosmological structure formation to a single model called the concordance model (Bahcall et al. 1999), for which free parameters are now known to high precision (Spergel et al. 2003). Cosmology thus finds itself in a place not unlike particle physics, where the goal going forward is to refine and test the standard model with yet higher precision measurements. Fundamental science questions driving the field include the nature of dark energy and dark matter, the formation and evolution of galaxies and quasars, and how and when the intergalactic medium was re-ionized.

Future progress requires ambitious observational surveys of the universe of unprecedented depth and breadth. The SDSS is collecting megabytes of data per galaxy on nearly 1 million galaxies distributed throughout a volume of space many billions of light years on a side. Currently over 2 TB of data has been collected and archived. This number is expected to grow to 5 TB by project's end. Several similarly sized surveys are underway, and much larger ones are planned. In particular, the Large aperture Synoptic Survey Telescope [LSST] will collect 15 TB of image data every night for a year, amassing a collection of tens of petabytes over several years. The LSST will produce an object catalog of a billion galaxies—a thousand-fold increase over the SDSS. Coping with this “data flood” requires advanced scientific data management technologies.

In order to maximize the science return, results from massive surveys need to be compared to the detailed predictions of the concordance model. These take the form of massive cosmological simulations of the formation of galaxies and large scale structure. Just as Moore's Law is the force behind the data explosion in astronomy, it has also enabled numerical simulations of unprecedented size and complexity on massively parallel supercomputers.

ENZO is a parallel cosmology application developed at the Laboratory for Computational Astrophysics (LCA) at UCSD, directed by Michael Norman. ENZO solves the equations of dark matter dynamics, multi-species hydrodynamics, non-equilibrium chemical and ionization kinetics, and self-gravity in an expanding universe dominated by dark energy. Parameterized models of star formation and feedback effects allow the simulation of the formation and evolution of galaxies on cosmic length scales and time scales. The state-of-the art is shown in Fig. 1. The simulation shown in the left

panel evolves a concordance model with 1 billion Lagrangian dark matter particles and the equations of Eulerian hydrodynamics and self-gravity on a uniform grid of 1 billion (1024^3) cells. The calculation was done on 512 processors of SDSC's IBM Blue Horizon computer, and produced 10TB of raw data and 6 TB of derived data. This calculation serves as a survey volume for follow-on adaptive mesh refinement (AMR) simulations which resolve the galaxies' internal structure. At right is shown an old AMR simulation of galaxy formation done at NCSA in 1998. Due to computer power and data handling limitations at the time, only 1/64 of the survey volume (2563 base grid) could be simulated at high resolution. Now, with more powerful parallel computers and data management technologies, we can in principle simulate the entire volume at high spatial resolution. Making that a practical reality is the overarching goal of the cosmology simulation data grid project, which we shall henceforth refer to as the Cosmic Simulator.

The specific goals of the Cosmic Simulator project are threefold:

1. use the LLNL-SDSC-UCSD data grid to be deployed to enable cosmological simulations of unprecedented size and physical realism;
2. improve the physical realism of cosmological modeling through the inclusion of radiation transfer on adaptive meshes;
3. generate simulated sky maps and galaxy catalogs using automated processing pipelines for LSST applications.

January 2005	LUSciD (LLNL UCSD Scientific Data Management) proposal submitted
January 2006	The LRAC (Large Resource Allocations Committee) proposal is submitted by Michael Norman, requesting time to run the low redshift tiles of the Santa Fe Light Cone.
April 2007	Submission of "The Santa Fe Light Cone Simulation Project. I. Confusion and the Warm-Hot Intergalactic Medium in Upcoming Sunyaev-Zel'dovich Effect Surveys"
January 2008	A second LRAC (Large Resource Allocations Committee) proposal is describing planned analysis of the simulation in the area of weak gravitational lensing.
June 2008	Submission of "Cosmological Shocks in Adaptive Mesh Refinement Simulations and the Acceleration of Cosmic Rays"
March 2009	Submission of "The Santa Fe Light Cone Simulation Project: II. The Prospects for Direct Detection of the WHIM with SZE Surveys"
August 2010	Submission of "Quantifying the collisionless nature of dark matter and galaxies in A1689"
October 2010	Submission of "The Properties of X-ray Cold Fronts in a Statistical Sample of Simulated Galaxy Clusters"
June 2011	Submission of "Profiles of Dark Matter Velocity Anisotropy in Simulated Clusters"

Key Personnel (including institutional affiliations and project positions)

1. Michael L. Norman, University of California, San Diego, Principal Investigator
2. Jack O. Burns, University of Colorado Boulder, Co-Principal Investigator
3. Eric J. Hallman, University of Colorado Boulder, Postdoctoral Fellow
4. James Bordner, University of California, San Diego, Scientist and Programmer
5. Robert Harkness, University of California, San Diego, Scientist and Programmer
6. Brian W. O'Shea, University of California, San Diego, Graduate Student
7. Geoffrey So, University of California, San Diego, Graduate Student
8. Rick Wagner, University of California, San Diego, Graduate Student

Scope and Contents note

The project files consists of data in three broad categories: the simulation data ("Data at Redshift" components); analysis tools and example scripts (Data Processing Tools) for processing the data; and project administration and background documents (Historical Documents) related to the project. All these materials were created between 2005 and 2012, beginning with a proposal for the LUSciD Project, continuing on to the simulation data, and ending with the recent analysis tools. The historical documents are proposals and progress reports that were part of grants or requests for computational resources supporting the research. The component for analysis tools and example scripts contains the source code to yt (<http://yt-project.org/>), which was used to produce the example data analysis results. The results are a combination of structured text, binary files, and images. The historical documents and analysis tools are described in greater detail in their component descriptions.

The scientific motivations for the light cone simulation are described in the Project Background. Here we describe how the simulation data was generated. The simulation was the final in a group of simulations, with each one designed to meet

certain requirements, such as resolution. Earlier simulations tied to the LUSciD Project were performed on Thunder, a Lawrence Livermore National Laboratory cluster. This calculation for the Santa Fe Light Cone Simulation was a demonstration of the software's ability to perform adaptive refinement throughout the volume, and as a result, was run on the San Diego Supercomputer Center's DataStar system and the National Center for Supercomputing Applications Altix, Cobalt.

The simulation was initialized at high redshift, assuming a standard cosmological model incorporating dark energy and cold dark matter. The physical volume represented was a periodic cube 512 comoving megaparsecs on a side. The simulation was evolved to the present day, using models for gravity and adiabatic gas dynamics. At specific points, snapshots of the simulation were saved, and a representative subset of those are contained in this collection.

These snapshots are organized by time (or, equivalently, redshift) at the top level, and named from RD0009 to RD0036; lower numbers (e.g., RD0009) represent earlier times in the universe's evolution, while higher numbers are later times and ones closer to the present day. Each snapshot has an archive (tar) file of the original data, a checksum of the archive, and text files of the parameters, grid hierarchy, and boundary conditions. The parameter, hierarchy, and boundary files are also in the archive file, but are available separately for convenience in a component named "Parameters."

The contents of each project component labeled RD00## are the same:

- * RD00## (parameters, ASCII): All of the simulation parameters are listed in these files as key-value pairs, using a "key = value" format. The input parameter are identical across all parameter files, while variables such as the current time, or redshift, change.

- * RD00##.hierarchy (grid metadata, ASCII): A list of the grid data structures, their spatial position, file names, and numerical size.

- * RD00##.cpu0XXX (physical data, HDF5): These files hold the physical fields (density, velocity, etc.) for each grid.

- * RD00##.boundary (boundary conditions, ASCII): Boundary metadata.

- * RD00##.boundary.hdf (boundary conditions, HDF5): Boundary data for necessary fields

Use References

Referenced below are articles and other publications identified at the end of 2011 as having used the data generated by Santa Fe Light Cone Simulation project.

Hallman, Eric J.; Skillman, Samuel W.; Jeltama, Tesla E.; Smith, Britton D.; O'Shea, Brian W.; Burns, Jack O.; and Norman, Michael L. "The Properties of X-ray Cold Fronts in a Statistical Sample of Simulated Galaxy Clusters." *The Astrophysical Journal*, Vol. 725, Issue 1: 1053-1068 (Dec. 2010); <http://dx.doi.org/10.1088/0004-637X/725/1/1053>; <http://iopscience.iop.org/0004-637X/725/1/1053>.

Hallman, Eric J.; O'Shea, Brian W.; Burns, Jack O.; Norman, Michael L.; and Harkness, Robert; Wagner, Rick. "The Santa Fe Light Cone Simulation Project. I. Confusion and the Warm-Hot Intergalactic Medium in Upcoming Sunyaev-Zel'dovich Effect Surveys." *The Astrophysical Journal*, V. 671, Issue 1: 27-39 (12/2007). <http://dx.doi.org/10.1086/522912>; <http://adsabs.harvard.edu/abs/2007ApJ...671...27H>

Hallman, Eric J.; O'Shea, Brian W.; Smith, Britton D.; Burns, Jack O.; and Norman, Michael L. "The Santa Fe Light Cone Simulation Project. II. The Prospects for Direct Detection of the Whim with SZE Surveys." *The Astrophysical Journal*, Vol. 698, Issue 2: 1795-1802 (2009); <http://dx.doi.org/10.1088/0004-637X/698/2/1795>; <http://iopscience.iop.org/0004-637X/698/2>.

Lemze, Doron; Rephaeli, Yoel; Barkana, Rennan; Broadhurst, Tom; Wagner, Rick; and Norman, Mike L. "Quantifying the Collisionless Nature of Dark Matter and Galaxies in A1689." *The Astrophysical Journal*, Vol. 728, Issue 1, article id 40 (2011); <http://dx.doi.org/10.1088/0004-637X/728/1/40>; <http://iopscience.iop.org/0004-637X/728/1/40>.

Lemze, Doron; Wagner, Rick; Rephaeli, Yoel; Sadeh, Sharon; Norman, Michael L.; Barkana, Rennan; Broadhurst, Tom; Ford, Holland; and Postman, Marc. "Profiles of Dark Matter Velocity Anisotropy in Simulated Clusters." eprint arXiv:1106.6048 (June 2011).

Skillman, Samuel W.; O'Shea, Brian W.; Hallman, Eric J.; Burns, Jack O.; and Norman, Michael L. "Cosmological Shocks in Adaptive Mesh Refinement Simulations and the Acceleration of Cosmic Rays." *The Astrophysical Journal*, Vol. 689, Issue 2: 1063-1077 (Dec. 2008); <http://dx.doi.org/10.1086/592496>; <http://iopscience.iop.org/0004-637X/689/2/1063>

Arrangement Note

The data set is arranged into 31 components: 1: Data processing tools; 2: Initial conditions for simulation; 3-30: Data at redshift=3.0 to Data at redshift=0.0, and 31: Historical documents.

Immediate Source of Acquisition note

Rick Wagner, 2012.

Processing Information note

The project lead collected, on the Triton Resource at the San Diego Supercomputer Center, all data generated by the Santa Fe Light Cone Simulation project deemed essential to representing the simulation project and facilitating re-use of the data. Data files were categorized and arranged to represent each snapshot (Data at redshift) comprising the simulation. The files for each snapshot include files specifying the parameters for each snapshot, binary data files constituting the results of applying the parameters, and derived data products generated from processing of the results. Files deemed irrelevant to representation of the project and / or use of the data were removed from the data set. In addition to data files, scripts necessary for processing the data were added to the collection, as were products generated using the scripts. The former are included in the component labeled "Data Processing Tools, " whereas the latter are typically included in a sub-component labeled "Derived Products" for each of the primary "Data at redshift" components. Finally, a variety of project files, primarily proposals and project status reports, have been incorporated and are listed in the component labeled "Historical documents." The Santa Fe Light Cone simulation files were then transferred from the SDSC server to the Research Data Curation Program data storage space. The transfer of all files were monitored for accuracy.

The entire collection was arranged into thirty-one components and described completely using the Archivists' Toolkit application. Component and sub-component descriptions were linked to digital object records composed in the AT and containing links to the files constituting the data set, or snapshot. The AT description was used to generate an Encoded Archival Description (EAD) document for the complete set of files for the Santa Fe Light Cone Simulation project data set and a METS document for each primary component. The EAD is to be uploaded to the Online Archive of California (OAC), whereas the METS records and the digital content files they reference are to be uploaded to the UC San Diego Digital Asset Management System (DAMS). A researcher will thus be enabled to access the data files either through the OAC or the UCSD DAMS. Finally, all files and descriptive records for the simulation project are to be deposited in the Chronopolis digital preservation network for long-term preservation management.

Access

This data set is available for use by the general research community, via the UC San Diego Library Digital Collections. Inquiries about using the dataset may be directed to the Research Data Curation Program at research-data-curation@ucsd.edu

Rights

The information contained in this set of research project files is the property of its creators and the Regents of the University of California. Some or all of the materials in the project files may be protected by copyright law. Use of this work beyond that allowed by "fair use" requires the written permission of the copyright holders(s). Responsibility for obtaining permissions and any use and distribution of this work rests exclusively with the user and not the UC San Diego Library. Inquiries can be made to the UC San Diego Library unit having custodial responsibility for the work (<https://lib.ucsd.edu/rdcp>).

License

This work is licensed under a Creative Commons Attribution 4.0 International Public License (<http://creativecommons.org/licenses/by/4.0/>).

Subjects and Indexing Terms

Los Alamos National Laboratory. Theoretical Astrophysics Group T-6.

San Diego Supercomputer Center.

University of California, San Diego. Center for Astrophysics and Space Sciences.

University of Colorado (System). Dept. of Astrophysics and Planetary Sciences. Center for Astrophysics and Space Astronomy.

Wagner, Rick, 1972-

Cosmic background radiation

Cosmology

Cosmology--Observations

Galaxies--Clusters

Hydrodynamics

Research data

Research-Data processing
