From *SIAM News*, Volume 42, Number 5, June 2009

CSE 2009

Petascale Challenges for Cosmological Simulation

By Paul Ricker

To simulate the entire Universe at one go would be an insane undertaking. Fortunately, cosmologists have been able to make progress without including the full range of astrophysical length scales in a single calculation. Simulations covering segments of this range have given us important insights into phenomena ranging from black hole event horizons (on the order of kilometers) to the cosmological horizon (on the order of 10^{23} km). Increasingly, however, questions of interest in cosmology link multiple scales and require coupling of physical processes that, until recently, had to be neglected.

Driving this development is the dramatic expansion under way in the size, precision, accessibility, and multiwavelength character of astronomical datasets. In the coming decade, several large new telescopes will begin operation, including the Large Synoptic Survey Telescope (visible light), the James Webb Space Telescope (near infrared), and the Atacama Large Millimeter Array (millimeter wavelengths) [17]. Because of the finite speed of light, these instruments will be able to probe the history of the Universe from the present back to the first few hundred million years after the Big Bang. The size of the resulting datasets (> 1 petabyte) poses its own challenges for observational data analysis and archiving [10]. The focus in this article is on the implications of these datasets for cosmological simulation.

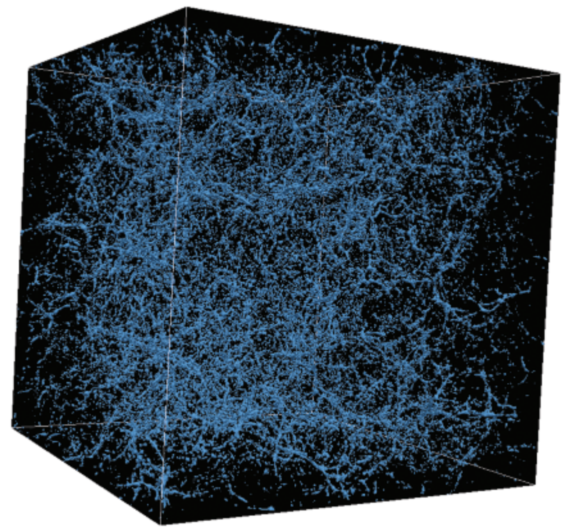

Figure 1. *Gas density isosurface in a 10^9-particle FLASH cosmological simulation spanning 365 megaparsecs on a side, showing the filamentary structure of the matter.*

The Current State of the Art

Current observations [6] suggest that only 27% of the Universe consists of matter. Most of this is dark matter, a hypothetical form of matter that interacts only very weakly with ordinary matter but feels gravity as an attractive force. Its existence is inferred from the dynamics of galaxies and clusters of galaxies. Baryons (protons and neutrons) make up only 4% of the total. The remaining 73% of the Universe consists of dark energy, about which we know little other than that it causes the expansion of the Universe to accelerate. Discovering the nature of dark matter and dark energy is one of the central problems in cosmology.

On the largest observable scales, the Universe is on average homogeneous and isotropic. Below ~100 megaparsecs* (Mpc), however, it is noticeably inhomogeneous, with a filamentary structure that on scales smaller than ~10 Mpc resolves into blobs called *halos*, which contain galaxies and clusters of galaxies (Figure 1). These structures form via the action of gravitational instability on small, random density perturbations that existed early in the Universe's history. Cosmological simulations solve the equations describing the growth of these structures in different cosmological models.

Usually, these simulations assume the validity of general relativity, which describes how the contents of the Universe influence its spatial geometry and time evolution. On scales far larger than the event horizons of black holes and smaller than the cosmological horizon, we can ignore most of the non-Newtonian aspects of general relativity. But the expansion of the Universe is a nonperturbative background effect that must be included. Dark energy enters simulations primarily via its effects on the expansion rate.

Collisionless dark matter, which dominates the density perturbations, is handled with *N*-body techniques. Hybrid tree/particle-mesh algorithms, as in Gadget-2 [11], and multilevel particle-mesh algorithms, as in Enzo [9] and FLASH [4], represent the state of the art in this area. The key metrics here are the gravitational force resolution, which determines the smallest halos that can be studied, and the number of particles, which determines the range of mass scales. To capture the dynamics of a galaxy like the Milky Way, a force resolution of ~1 kpc or smaller is needed, whereas the computational volume must span more than ~500 Mpc if "mock skies" are to be generated for comparison with galaxy surveys. To be useful, this spatial dynamic range must be matched by a correspondingly large number of particles. Simulations with 10^{10} particles are becoming routine [12], and those approaching 10^{11} particles are now possible [15]. To cover the complete survey volumes spanned by planned observation campaigns, such as the Dark Energy Survey [18], at least 10^{12} particles will be required.

Pure *N*-body simulations provide useful information about halo structure and clustering statistics. But because dark matter does not emit light, self-consistent predictions of observable quantities must include baryons. Most baryons exist in diffuse form, either as cold, low-density gas in the filaments, which is difficult to detect, or as hot, ionized plasma in galaxy clusters. The wide range of Mach numbers for the gas—from ~1–5 (in clusters) to more than 100 (for cold gas accreting onto clusters)—makes shock-capturing (magneto)hydrodynamics algorithms essential. Smoothed particle hydrodynamics (SPH), one of the two major categories, uses *N*-body particles as moving interpolation centers for fluid quantities like pressure. Eulerian Godunov-type methods with adaptive mesh refinement (AMR) are the main alternative. While SPH handles densi-

*1 parsec (pc) ≈ 3.09×10^{13} km.

ty peaks more efficiently, AMR methods offer superior treatment of shear flows and density gradients [1,14].

The greatest uncertainties in the behavior of the diffuse baryons arise because of gas physics on scales smaller than the force resolution. The gas, driven to higher densities by self-gravity, radiatively cools to form stars, some of which die in supernova explosions that inject entropy and nuclear fusion products into the intergalactic medium. The supermassive black holes found at the centers of most galaxies can also produce feedback when they accrete gas from their surroundings. To predict how galaxies light up and evolve and to correctly follow the thermodynamic state of the diffuse gas, we must treat these processes in some fashion. The state of the art is represented by direct simulations of the formation of the first stars [16], statistical subgrid models [13], and "semi-analytic models" based on phenomenological modeling of the gas [2]. None of the existing methods reproduces all of the observed properties of galaxies, however, and numerous fundamental questions remain, particularly regarding the origin of the supermassive black holes [7].

Getting to the Petascale

The forthcoming large observational datasets demand petascale cosmological simulations for two reasons. Because of the immense volumes probed by the new surveys, we will need to study the statistics of rare objects (massive clusters and quasars) and create accurate mock skies. This will require a large increase in dynamic range, with corresponding increases in simulation data volume. For accurate predictions of the internal structure and multiwavelength observable properties of clusters and galaxies, we will have to include more complete physics at the kpc scale and below. This will add to overall processor and memory requirements, and the numerical treatment will have to be validated by comparison with nearby galaxies and with direct simulations on smaller scales.

The high-node-count, multicore, heterogeneous petascale systems currently being designed and built present us with several challenges. Because the cosmic density fluctuations are random, algorithms that focus resources on high-density regions require dynamically varying load balancing. Our current algorithms are thus sensitive to communication latency and bandwidth, particularly when we are solving for the gravitational potential. One possible solution to this problem is to move load balancing away from the application and tie it to the communications library, as done with processor virtualization under CHARM++ [5]. This allows the load-balancing scheme to be tuned individually for each architecture without modifications to the application. At Illinois we are exploring this approach with FLASH.

The trend toward more cores per processor presents a challenge because memory bandwidth has not scaled with the number of cores. Godunov methods and subgrid models, in particular, can benefit from cache-optimizing code transformations on multicore systems. These transformations usually make codes difficult to maintain, however. Performance annotations [8], one approach we are exploring, allow programmers to suggest code transformations to an annotation preprocessor without changing the readability of their existing code. The preprocessor creates optimized source code that can significantly outperform compiler-based optimizations.

Hardware accelerators, such as the Cell processor and general-purpose graphics processing units, have shown promise in astrophysical *N*-body applications [3]. Taking advantage of these systems requires significant investment in code and algorithm development. Communication between accelerators and other nodes within a heterogeneous cluster is often slow compared to the accelerators' raw floating-point performance. Hence, the most successful codes will likely use accelerators for local, compute-intensive tasks, such as subgrid models or other physics that can be decoupled from distant regions of the computational volume.

Conclusions

The next decade should yield greatly improved constraints on the properties of dark matter and dark energy, a better understanding of how the first stars and galaxies formed, and substantially better understanding of the evolution of galaxies in different environments. If we are to extract this knowledge from observations, cosmological simulations will need to have the greatly enlarged dynamic range and improved baryonic physics fidelity that can be achieved by making full use of petascale resources.

References

[1] O. Agertz et al., *Fundamental differences between SPH and grid methods*, Monthly Notices Roy. Astron. Soc., 380 (2007), 963–978.

[2] C.M. Baugh, *A primer on hierarchical galaxy formation: The semi-analytical approach*, Rep. Prog. Phys., 69 (2006), 3101–3156.

[3] R.G. Belleman et al., *High performance direct gravitational N-body simulations on graphics processing units II: An Implementation in CUDA*, New Astron., 13 (2008), 103–112.

[4] B. Fryxell et al., *FLASH: An adaptive mesh hydrodynamics code for modeling thermonuclear flashes*, Astroph. J. Suppl., 131 (2000), 273–334.

[5] L.V. Kalé et al., in *Petascale Computing: Algorithms and Applications*, CRC Press, Boca Raton, Florida, 2007, 421–442.

[6] E. Komatsu et al., *Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological interpretation*, Astroph. J. Suppl., 180 (2009), 330–376.

[7] P. Madau et al., *Massive black holes across cosmic time*, white paper submitted to Astro2010 Decadal Rev., arXiv:0903.0097.

[8] B. Norris et al., in *Petascale Computing: Algorithms and Applications*, CRC Press, Boca Raton, Florida, 2007, 443–462.

[9] B.W. O'Shea et al., in *Adaptive Mesh Refinement: Theory and Applications*, T. Plewa, T. Linde, and V.G. Weirs, eds., Springer, New York, 2004, 341–350.

[10] R.C. Smith et al., T*he ground-based O/IR data preservation challenge at NOAO: Scaling up from terabytes to petabytes*, ASP Conf. Ser., 394 (2008), 446–449.

[11] V. Springel, *The cosmological simulation code GADGET-2*, Monthly Notices Roy. Astron. Soc., 364 (2005), 1105–1134.

[12] V. Springel et al., *Simulations of the formation, evolution and clustering of galaxies and quasars*, Nature, 435 (2005), 629–636.

[13] E.J. Tasker and G.L. Bryan, *The effect of the interstellar model on star formation properties in galactic disks*, Astroph. J., 673 (2008), 810–831.

[14] E.J. Tasker et al., *A test suite for quantitative comparison of hydrodynamic codes in astrophysics*, Monthly Notices Roy. Astron. Soc., 390 (2008), 1267–1281.

[15] R. Teyssier et al., *Full-sky weak lensing simulation with 70 billion particles*, Astron. Astroph., accepted in 2009, arXiv:0807.3651.

[16] J.H. Wise and T. Abel, *Resolving the formation of protogalaxies III: Feedback from the first stars*, Astroph. J., 685 (2008), 40–56.

[17]http://www.lsst.org; http://www.jwst.nasa.gov; http://almaobservatory.org.

[18] http://www.darkenergysurvey.org.

Paul Ricker is an assistant professor in the Department of Astronomy and the National Center for Supercomputing Applications, University of Illinois at Urbana–Champaign.