Implementing Effects due to Stellar Evolution in a Cosmological Simulation Code

A Proposal to the NCSA Strategic Applications Program

Paul Ricker
University of Illinois at Urbana-Champaign
National Center for Supercomputing Applications

SCIENTIFIC GOALS

As the most massive collapsed objects in the Universe, clusters of galaxies offer particularly important insights into the cosmos as a whole. Because they require billions of years to virialize, clusters' structural evolution is set by the rate of expansion of the Universe, which in turn is determined by the amount of matter and energy in the Universe. Statistical measures of clusters such as their mass distribution and power spectrum also reflect the influence of cosmology, since clusters are the most recent structures to have separated from the cosmic expansion. Their lateness on the cosmic stage makes them especially sensitive to the properties of dark energy.

Clusters consist primarily of three components: dark matter, which interacts gravitationally with itself and ordinary matter and makes up 80-85% of the total mass; the diffuse, hot ($> 10^7$ K) gas of the intracluster medium (ICM), which contributes 15-20% of the total mass; and galaxies, which contribute only about 1% to the total.

The ICM can be treated as a collisional fluid because tangled magnetic fields on spatial scales $\ll 1$ kpc suppress the free streaming of electrons, which would otherwise have a Coulomb mean free path of several kpc (Sarazin 1988). Mergers between clusters and the motions of galaxies within the ICM produce transonic to mildly supersonic gas motions, resulting in shocks, which raise the entropy of the gas. Typical Reynolds numbers in the ICM should be $\sim 100-1000$, implying the presence of turbulence (Norman & Bryan 1999; Ricker & Sarazin 2001).

However, this simple picture of shocked ideal hydrodynamics fails to reproduce several important statistical scaling properties of clusters at low redshift, most notably the luminosity-temperature relation (Kaiser 1986; Arnaud & Evrard 1998; Markevitch 1998; Allen & Fabian 1998). This discrepancy has been attributed to the need to consider additional sources and sinks of entropy in the ICM. Several models have been proposed, including ICM preheating and energy injection from supernovae or galactic winds (e.g., Metzler & Evrard 1997; Bialek et al. 2001; Borgani et al. 2001), removal of low-entropy gas via radiative cooling (Bryan 2000), and some combination of these models (Voit et al. 2002). Similar discrepancies are found when considering the size-temperature (Mohr & Evrard 1997) and mass-temperature (Mohr et al. 1999) relations. Moreover, we now know from high-resolution X-ray observations of nearby cooling-flow clusters that active galactic nuclei (AGN) can play a significant role in disturbing and heating the ICM (e.g., McNamara et al. 2000, 2001; Forman et al. 2002; Blanton et al. 2003).

This additional gas physics involves processes on unresolvably small scales. For example, directly including stars in a simulation would require resolving regions smaller than 10^{-5} pc, whereas large cosmological simulations are just now resolving kpc scales. Fortunately, the details of very small-scale processes should be relatively unimportant. On kpc scales and larger, they represent sources and sinks of energy, entropy, and metals. For our purposes we can therefore hope to treat them using hydrodynamical source terms based on appropriate subgrid modeling. This is similar to the way in which the universality of isotropic turbulence on very small scales is used to include turbulent dissipation in large eddy simulations of turbulence. As with turbulence models, it is important that subgrid models for stellar evolution reproduce the observed statistical behavior of stellar populations when averaged over kpc-size volumes.

Subgrid modeling helps us to cope with the enormous dynamic range separating cluster length scales from scales characteristic of stellar evolution. However, clusters form by gravitationally accumulating matter from regions several tens of Mpc on a side, so a gas-dynamics simulation with a spatial resolution of 10 kpc already requires a spatial dynamic range > 1000 just to handle a single cluster properly. If ensembles of clusters are to be simulated — as we must do if we are to make accurately statistical comparisons with observational data — we need resolution equivalent to a uniform mesh with 10^4 or 10^5 zones on a side. Fortunately, this level of resolution is required only in a small fraction of the total volume, so techniques such as adaptive mesh refinement (AMR) can be employed profitably. Even with AMR, however, ensemble simulations of galaxy cluster evolution are necessarily large. Given the processor and memory capacity of modern computers, parallel computers with hundreds to thousands of processors are essential for these simulations.

COMPUTATIONAL GOALS AND METHODS

To study cluster evolution, we use a simulation framework called FLASH that was originally developed by the ASCI Center for Astrophysical Thermonuclear Flashes at the University of Chicago. FLASH (Fryxell et al. 2000) was designed to study X-ray bursts, novae, and Type Ia supernovae using AMR on large parallel computers. FLASH has been used to perform some of the largest AMR calculations ever attempted, earning its authors the 2000 Gordon Bell Prize (Calder et al. 2000). It is nearly unique among astrophysical codes in having been extensively validated against laboratory experiments (Calder et al. 2002). Since its beginning, FLASH has evolved into a general-purpose astrophysical simulation tool, including modules for gravity, magnetohydrodynamics, and particles.

For cosmological simulations with FLASH, all calculations are assumed to take place in comoving coordinates $\mathbf{x} = \mathbf{r}/a$, where $\mathbf{r}$ is a proper position vector and $a(t)$ is the time-dependent cosmological scale factor. The present epoch is defined to correspond to $a = 1$; in the following discussion we use $t = t_0$ to refer to the age of the Universe today.

The evolution of gas (baryons) is described by the Euler equations of hydrodynamics. A special coordinate transformation casts these equations into a relatively simple form when cosmological expansion is included. The gas velocity $\mathbf{v}$ is taken to be the comoving peculiar velocity $\dot{\mathbf{x}}$. The comoving gas density, pressure, temperature, and internal energy density are defined to be

$$\begin{aligned}
\rho &\equiv a^3\tilde{\rho} \\
p &\equiv a\tilde{p} \\
T &\equiv \frac{\tilde{T}}{a^2} \\
\rho\epsilon &\equiv a\tilde{\rho}\tilde{\epsilon}\ .
\end{aligned} \tag{1}$$

The quantities marked with a tilde ($\tilde{\rho}$ etc.) are the corresponding "proper" or physical quantities. Note that, in terms of comoving quantities, the equation of state has the same form as for the proper quantities in noncomoving coordinates. For example, the equation of state for an ideal gas with adiabatic index γ is

$$\rho\epsilon = \frac{p}{\gamma - 1} = \frac{\rho k T}{(\gamma - 1)\mu}\ , \tag{2}$$

where μ is the average mass per particle and k is Boltzmann's constant. With these definitions, the Euler equations for a perfect gas can be written in the form

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = \Gamma_M \tag{3}$$

$$\frac{\partial\rho\mathbf{v}}{\partial t} + \nabla\cdot(\rho\mathbf{v}\mathbf{v}) + \nabla p + 2\frac{\dot{a}}{a}\rho\mathbf{v} + \rho\nabla\phi = \mathbf{\Gamma_P} \tag{4}$$

$$\frac{\partial\rho E}{\partial t} + \nabla\cdot[(\rho E + p)\mathbf{v}] + \frac{\dot{a}}{a}[(3\gamma - 1)\rho\epsilon + 2\rho v^2] + \rho\mathbf{v}\cdot\nabla\phi = \Gamma_E\ . \tag{5}$$

Here E is the specific total energy, $\epsilon + \frac{1}{2}v^2$. Γ_M, $\mathbf{\Gamma_P}$, and Γ_E are local net mass, momentum, and energy source terms, respectively. The appropriate specification of these source terms is half of our proposed subgrid modeling effort.

Since astrophysical flows are usually highly compressible, FLASH uses the piecewise-parabolic method (PPM) to solve the Euler equations (Colella & Woodward 1984). PPM is optimized for flows that include sharp flow discontinuities such as shocks, resolving them with 1-2 mesh zones and avoiding the spurious oscillations and excessive dissipation that plague less sophisticated methods.

For collisionless species, including dark matter and stars, we use an N-body or particle representation. Each computational particle corresponds to a large number of physical particles. The ith computational particle is characterized by a comoving position $\mathbf{x}_i$, a comoving velocity $\mathbf{v}_i$, a mass m_i, and perhaps additional quantities such as a unique identifier (tag). The particles move in the same potential as the gas, obeying the equations

$$\begin{aligned}
\frac{d\mathbf{x}_i}{dt} &= \mathbf{v}_i \\
\frac{d\mathbf{v}_i}{dt} &= -\nabla\phi - 2\frac{\dot{a}}{a}\mathbf{v}_i
\end{aligned} \tag{6}$$

The first term on the right-hand side of the second equation represents the gravitational force on the particles, while the second term represents the redshift effect due to cosmic expansion. Because star formation converts gas into collisionless stars, the second half of our subgrid modeling effort involves the specification of rules for creating and destroying particles intended to follow stellar populations.

The most expensive part of a collisionless particle solver is generally the computation of the force on the particles. Since gravity is a long-range force, a naive calculation of the force on N particles requires of order N^2 operations. For this reason several methods to speed up the force calculation have been developed. FLASH uses an N-body solver based on the particle-mesh method (Hockney & Eastwood 1988). This method speeds force calculations by converting particle positions to densities on a grid, solving the Poisson equation on the grid, and finally interpolating the potential from the grid to obtain forces at particle locations. Since we already need to solve the Poisson equation on a grid for our hydrodynamical code, the particle-mesh method is ideally suited for integration with it. While the FLASH framework enables other methods such as tree-based methods to be implemented straightforwardly, we have chosen the particle-mesh method because of its simplicity and because the AMR capabilities of the code can be used to yield the correct smoothing length in low- and high-density regions without further changes (Knebe, Green, & Binney 2001).

The comoving potential ϕ in the above equations is the solution to the Poisson equation in the form

$$\nabla^2\phi = \frac{4\pi G}{a^3}(\rho_{\rm tot} - \bar{\rho}) \ , \tag{7}$$

where $\rho_{\rm tot}$ is the comoving matter density (gas plus dark matter), $\bar{\rho}$ is the comoving mean matter density, and G is Newton's gravitational constant. The comoving mean matter density is defined in terms of the present-day critical density $\rho_{\rm crit}$ by

$$\bar{\rho} \equiv \Omega_{\rm m}\rho_{\rm crit} \ , \tag{8}$$

where $\Omega_{\rm m}$ is the matter density parameter (whose value is specified as part of the cosmological model), and

$$\rho_{\rm crit} \equiv \frac{3H_0^2}{8\pi G} \ . \tag{9}$$

Here H_0 is the Hubble constant. The Hubble parameter $H(t)$ ($H_0 \equiv H(t_0)$) characterizes the rate of expansion of the Universe and is given by the Friedmann equation:

$$H^2(t) \equiv \left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left(\frac{\Omega_{\rm m}}{a^3} + \frac{\Omega_{\rm r}}{a^4} + \Omega_{\rm de}(a) - \frac{\Omega_{\rm c}}{a^2}\right) \ . \tag{10}$$

Here $\Omega_{\rm r}$ is the present-day density parameter of radiation, and $\Omega_{\rm de}(a)$ is the model-dependent density parameter of dark energy. $\Omega_{\rm de}$ is thought to be roughly constant at the present day, and it is this lack of dilution as the Universe expands that is driving the expansion to accelerate. The present-day contribution of the overall spatial curvature of the Universe (thought to be zero) is given by

$$\Omega_{\rm c} \equiv \Omega_{\rm m} + \Omega_{\rm r} + \Omega_{\rm de} - 1 \ . \tag{11}$$

The Friedmann equation is solved using a standard ordinary differential equation solver.

At each timestep we solve the Poisson equation using a multigrid elliptic solver (Brandt 1977). Multigrid algorithms solve elliptic equations by accelerating the convergence of relaxation methods. Relaxation methods are easy to implement but converge very slowly. They accomplish the global coupling implied by an elliptic equation using a series of iterations that communicate information across the grid one zone at a time. Hence their convergence rate (fractional reduction in error per iteration) decreases with increasing grid size. The longest-wavelength components of the error require the most iterations to decrease to a given level. By iterating on a sequence of coarser grids, multigrid algorithms bring all wavelengths into convergence at the same rate. This works because long wavelengths on a fine mesh appear to be short wavelengths on a coarse mesh.

Our multigrid solver is based on a multilevel adaptive refinement scheme described by Martin and Cartwright (1996) and Martin (1998). Adaptive mesh refinement provides many benefits in conjunction with a multigrid solver. Where errors are unlikely to have short-wavelength components it makes sense to avoid using fine grids, thus reducing storage requirements and the cost of relaxations on fine levels. The AMR package manages the multilevel mesh data structures and can handle all parallel communication, allowing the multigrid solver to be implemented independently of such details. The AMR package supplies many of the basic functions required by multigrid algorithms in addition to the mesh data structure, including prolongation, restriction, and boundary condition updates. Therefore we use a mesh hierarchy defined by the AMR package.

To implement adaptive mesh refinement, FLASH uses a customized version of the block-structured AMR package PARAMESH (MacNeice et al. 2000). Berger and co-workers (Berger & Oliger 1984, Berger & Colella 1989) pioneered block-structured algorithms, which use a hierarchy of logically Cartesian grids to cover the computational domain. This approach is flexible and memory-efficient, but the resulting code is complex and can be difficult to parallelize efficiently. Quirk (1991) and De Zeeuw & Powell (1993) implemented simplified versions which develop the hierarchy of nested grids by bisecting blocks in each coordinate direction and placing each block at a node of a tree. Blocks are distributed among processors using a Morton space-filling curve (Warren & Salmon 1993). This is the approach on which PARAMESH is based. PARAMESH differs from other AMR libraries in that it was designed with much less abstraction and a greater emphasis on achieving parallel performance. Abstraction and control are implemented at compilation time within the FLASH framework, enabling these performance advantages to be preserved within a flexible application environment.

PARAMESH uses two main data structures: one to hold the solution and another to store the tree information describing the mesh hierarchy. FLASH makes these structures available to the rest of the application through a variety of accessor methods, enabling most physics modules to be coded without explicit reference to AMR. Each block in the mesh hierarchy contains the same number of interior and guard cells. Blocks on different levels are nested. Between any two adjacent levels the spatial resolution changes by a factor of two, and any block's immediate neighbors can be at most one level finer or coarser than it. In exchange for some loss of flexibility, this method gives very good performance, because it gives modern compilers the best opportunity to efficiently manage cache use.

The criteria for refining blocks are user-defined. For hydrodynamical simulations, we use a normalized second-derivative criterion (Löhner 1987) to capture shocks. It is important to ensure that flow discontinuities are detected by a block before they move into the interior cells of that block. We therefore test for refinement every four timesteps.

Refinement of cosmological simulations must also respond to the need for high resolution in collapsed regions such as clusters without wasting resources in voids. One straightforward approach is to use fine meshes in regions where the overdensity is highest. However, because the density fluctuation field possesses power on all length scales, new refined meshes must contain the correct small-scale power. Since these fluctuations may be nonlinear already at the time of refinement, the only correct approach at present is to include them in the initial conditions. This requires either refining all of the initial volume — defeating the purpose of AMR — or somehow predicting in advance which regions will end up in collapsed structures and refining them from the beginning. To do this, we are investigating several methods of predicting where structure will form, including resimulation methods (e.g., Kravtsov 1999) and using the adhesion approximation (Shandarin 1987). We are also studying methods for initializing structure without large uniform-grid Fourier transforms (Pen 1997; Bertschinger 2001). The results of this study will form the basis of a forthcoming paper (Ricker et al., in preparation).

POTENTIAL BENEFITS

We will develop a subgrid module for FLASH that will enable the following physics to be accurately modeled within hydrodynamical simulations having spatial resolutions of order 100 pc or greater. As a starting point we will use the hybrid multiphase star formation model described by Springel & Hernquist (2003) for smoothed-particle hydrodynamics.

- **Cold clouds:** Below a temperature of about 10^7 K, diffuse gas in pressure equilibrium rapidly cools through radiation and condenses into dense molecular clouds, which are the sites for star formation. Although we already implement radiative cooling within FLASH, we cannot follow the short time and length scales on which these clouds form. Therefore we will treat cold clouds as a separate fluid with a source term based on the local cooling rate and a sink term based on the photoionization heating rate (see below).

- **Star formation:** Given a specified dependence of star formation rate (SFR) and initial mass function (IMF) on local conditions in the diffuse gas, we will convert diffuse gas continuously into stars, which will be represented by collisionless particles. These particles will represent definite ranges of stellar mass and have built-in 'clocks' that track the evolutionary stage of the corresponding stars.

- **Stellar winds:** Stars in certain stages of evolution (e.g., asymptotic giant branch stars) can produce powerful gaseous outflows ('winds') that return material to the interstellar medium. This material has been enriched in fusion products relative to the material from which the stars formed. These winds will be modeled using gas-phase source terms tied to the presence of star particles at the appropriate stages of evolution.

- **Photoionization heating:** Massive main-sequence stars are hot and produce significant amounts of ionizing radiation. In addition to changing the local ionization state of the diffuse gas, this radiation deposits energy into the gas as heat. Modeling this effect requires adding an energy source term tied to the presence of particles representing massive main-sequence stars.

- **Supernova explosions:** At the end of their main-sequence lifetimes, massive stars undergo catastrophic core collapse resulting in powerful explosions known as supernovae. Supernovae release substantial amounts of enriched material moving at high speeds into the interstellar medium. These explosions will be modeled using gas-phase source terms tied to the presence of massive star particles whose clocks have reached the end of their main-sequence lifetimes.

Parameters for each of these models will be calibrated by comparing simulation results to galaxies with observed star formation rates, including both ‘quiescent’ star formation, which tends to produce stars like our Sun, and ‘starbursts’, which tend to produce massive stars whose supernovae drive galactic winds.

This module will enable us to include our best current knowledge of stellar evolution into our galaxy cluster simulations, allowing us to determine the extent to which stellar processes can realistically explain the excess entropy needed in cluster cores. Assuming we can reproduce the observed scaling relations for clusters at low redshift, we should then with some confidence be able to produce ensemble simulations producing thousands of clusters for direct comparison with observational surveys. This module will also enable FLASH to handle problems in galaxy formation and evolution. For example, studies of the reionization and enrichment of the primordial intergalactic medium will come within reach, as will studies of galaxy evolution in group and cluster environments. All of these are current topics of great interest to workers in cosmological simulation.

The proposed project will also benefit NCSA and its user community. We will develop and share useful experience in several areas:

- Modeling stellar evolution processes on large scales.

- Developing subgrid models for adaptive mesh refinement codes.

- Parallel load balancing for particle applications.

- Mixing particle- and mesh-based representations of multiphase flows.

- Developing user physics modules for FLASH.

- Deploying FLASH on new platforms (tungsten and mercury).

FLASH has become a widely-used astrophysical community code, and demonstrating its use on NCSA platforms will encourage others in the community to consider using it for their simulation projects at NCSA.

COMPUTATIONAL RESOURCES

FLASH is an MPI-based parallel application and should port readily to NCSA clusters. We will target two platforms for porting, development, and optimization: tungsten, because of the large number of processors on this machine; and mercury, because of the large amount of memory available. This strategy will also allow us to compare the code's performance on 32-bit and 64-bit platforms.

We will demonstrate the scalability of our code through a suite of benchmark cluster and galaxy formation calculations with effective spatial resolutions up to 512^3 zones. Scientific utility will be demonstrated using a simulation of galaxy cluster formation in the ΛCDM cosmological model, computed both with and without the subgrid model, using an adaptive mesh equivalent (at least) to a 2048^3-zone uniform grid.

In addition to the PI, who has a half-time appointment with NCSA as a research scientist in the Applications Group, this project will employ the efforts of Dr. Ramesh Balakrishnan and Dr. Gregory Bauer, NCSA staff members in the Performance Engineering Group. The PI will lead the development and calibration of the subgrid physics models and conduct the target scientific calculations. Dr. Balakrishnan will lead the development of a consistent mathematical representation of the physics models and code the subgrid module. Dr. Balakrishnan and Dr. Bauer will port and optimize the code and document benchmark results.

REFERENCES

Allen, S. W. and Fabian, A. C. 1998, MNRAS, 297, L57

Arnaud, M. and Evrard, A. E. 1998, MNRAS, 305, 631

Berger, M. J. and Oliger, J. 1984, JCP, 53, 484

Berger, M. J. and Colella, P. 1989, JCP, 82, 64

Bertschinger, E. 2001, ApJS, 137, 1

Bialek, J. J., Evrard, A. E., and Mohr, J. J. 2001, ApJ, 555, 597

Blanton, E. L., Sarazin, C. L., and McNamara, B. R. 2003, ApJ, 585, 227

Borgani, S., et al. 2001, ApJ, 559, L71

Brandt, A. 1977, Math. Comp., 31, 333

Bryan, G. L. 2000, ApJ, 544, L1

Calder, A. C., et. al. 2000, in Proc. Supercomputing 2000

________. 2002, ApJS, 143, 201

Colella, P. and Woodward, P. 1984, JCP, 54, 174

De Zeeuw, D. and Powell, K. G. 1993, JCP, 104, 56

Forman, W., et al. 2002, in Lighthouses of the Universe (ESO), 51

Fryxell, B., et al. 2000, ApJS, 131, 273

Hockney, R. W. and Eastwood, J. W. 1988, Computer Simulation using Particles (IOP)

Kaiser, N. 1986, MNRAS, 222, 323

Knebe, A., Green, A., and Binney, J. 2001, MNRAS, 325, 845

Kravtsov, A. V. 1999, PhD thesis, New Mexico State University

Löhner, R. 1987, Comp. Meth. App. Mech. Eng., 61, 323

MacNeice, P., et al. 2000, Comp. Phys. Comm., 126, 330

Markevitch, M. 1998, ApJ, 504, 27

Martin, D. 1998, PhD thesis, Univ. of California at Berkeley

Martin, D. and Cartwright, K. 1996, Tech. Rep. UCB/ERL M96/66 (UC Berkeley)

McNamara, B. R., et al. 2000, ApJ, 534, L135

________. 2001, ApJ, 562, L149

Metzler, C. and Evrard, A. E. astro-ph/9710324

Mohr, J. J. and Evrard, A. E. 1997, ApJ, 491, 38

Mohr, J. J., Mathiesen, B., and Evrard, A. E. 1999, ApJ, 517, 627

Norman, M. L. and Bryan, G. L. 1999, in The Radio Galaxy Messier 87, H.-J. Röser and K. Meisenheimer, eds. (Springer), 106

Pen, U. 1997, ApJ, 490, L127

Quirk, J. J. 1991, PhD thesis, Cranfield Institute of Technology

Ricker, P. M. and Sarazin, C. L. 2001, ApJ, 561, 621

Sarazin, C. L. 1988, X-Ray Emission from Clusters of Galaxies (Cambridge)

Shandarin, S. F. 1987, in Large Scale Structures of the Universe: Proc. 130th IAU Symposium (Kluwer), 273

Springel, V. and Hernquist, L. 2003, MNRAS, 339, 289

Voit, G. M., et al. 2002, ApJ, 576, 601

Warren, M. S. and Salmon, J. K. 1993, in Proc. Supercomputing 1993 (IEEE Comp. Soc.), 12