Nyx: A New Lyman-alpha Forest Cosmology Code

Peter Nugent (LBNL/UC Berkeley)

Computational Cosmology

Currently we are in the middle of a series of large multi-agency experimental programs designed to investigate the "Dark Universe."

- In these experiments, the Universe is part of the apparatus, blurring the line between theory and experiment.

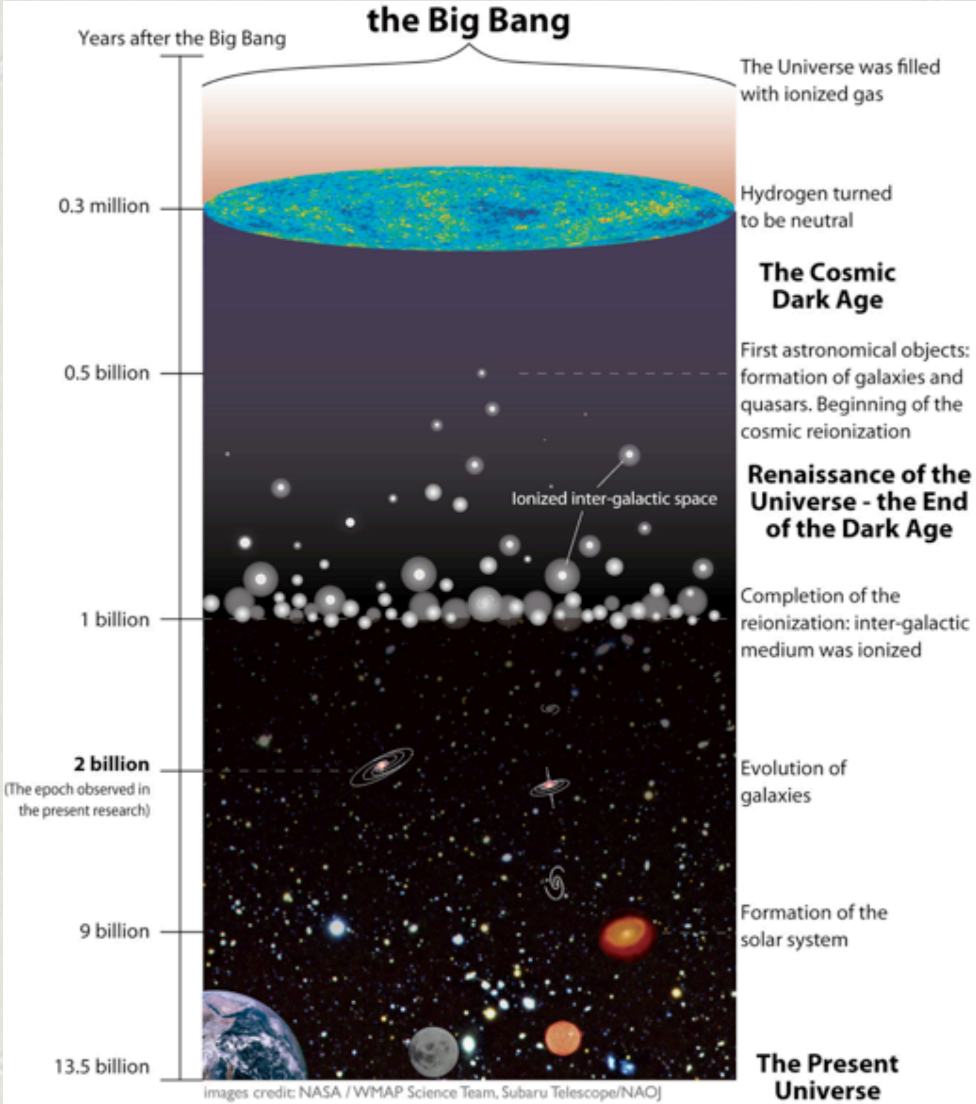
- To understand and calibrate the response of the apparatus to the fundamental physics requires a coordinated and disciplined program of cosmological simulations.

- A large-scale computational cosmology program is necessary to analyze and interpret the results of upcoming experiments such as BOSS, DES and LSST; it will also extend their science reach...

Structure Formation

Solid understanding of structure formation, success underpins most cosmic discovery:

- Initial conditions laid down by inflation.
- Initial perturbations amplified by gravitational instability in a dark-matter dominated Universe.
- Relevant theory is gravity and atomic physics - from 'first principles'
- Early Universe: Linear perturbation theory very successful-Cosmic Microwave Background
- Latter half of the history of the Universe: Nonlinear domain of structure formation, impossible to treat without large-scale computing.



Role of Simulations

Cosmic Frontier: In uncontrolled experiments, where the Universe is the apparatus, the role of computational theory and modeling is pervasive, complex, and crucial to the success of the entire enterprise.

The Key Tools

All structure formation is, in essence, a probe of density. Hence we desire robust ways to characterize clustering statistics of the underlying mass field and its tracers (e.g., galaxies, gas clouds, etc.)

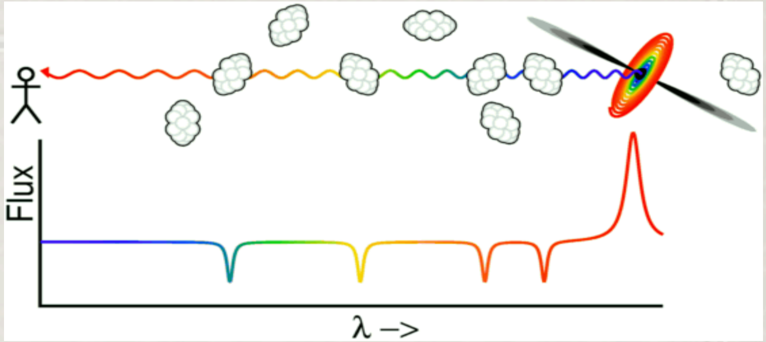
- The (2-point) correlation function is the excess probability of finding an object pair separated by a distance r_{12} compared to that for a random distribution: $dP=n^2(1+\xi(r_{12}))dV_1dV_2$
where n is the mean density; the power spectrum $P(k)$ is the Fourier transform of the correlation function

- The primordial fluctuations, as best known currently, are Gaussian, and completely specified by 2-point statistics

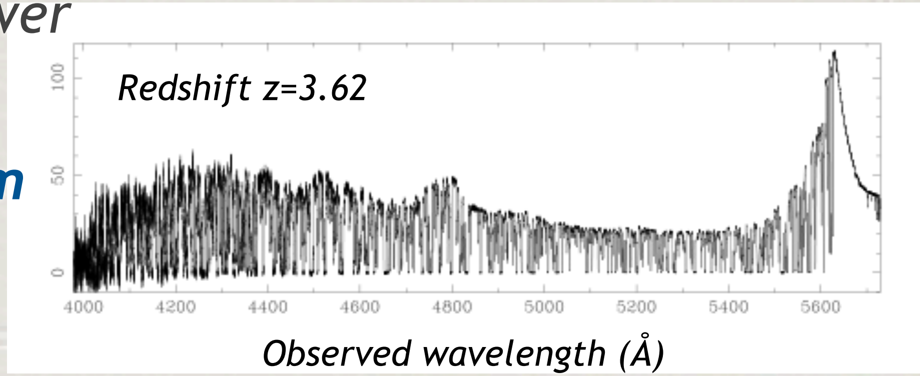
- Nonlinear structure formation induces nonzero higher point correlation functions

Lyman-alpha Forest

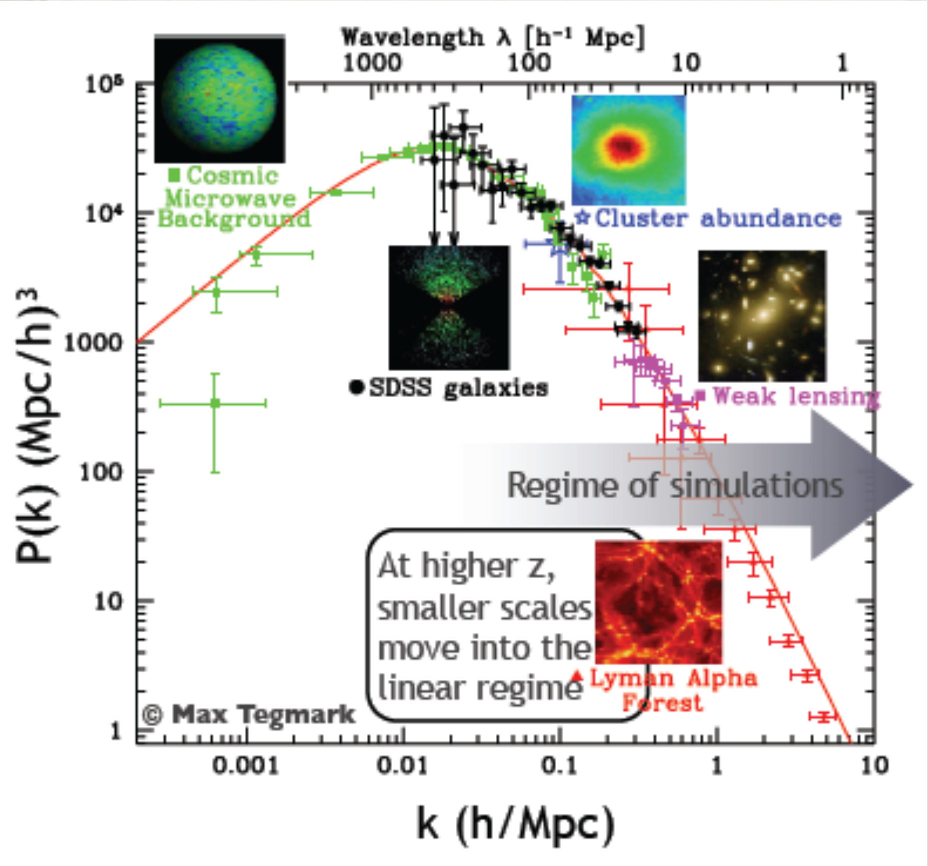
- ***Background:*** *Use bright distant sources as "skewers" through the cosmic density field*
  - ***Skewers:*** *Gas probed at near mean density along 1-D sightlines*
  - ***High-z:*** *Signal shifts into visible at z>2*
  - ***Scales:*** *Along a sightline, probe scales >~100 kpc (comoving), across sightlines, depends on inter-skewer separation*
  - ***Systematics very different from galaxy surveys***



Spectrum '=' Density



Power Spectrum



Baryon Acoustic Oscillations (BAO) from galaxy surveys (BOSS, DES, LSST)

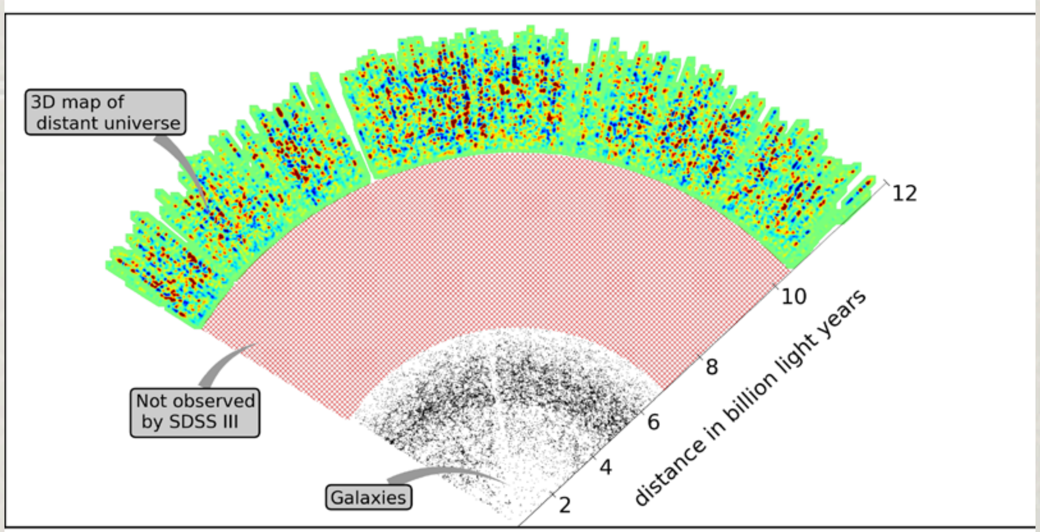
- **Measurement:** Geometry at $z<1$
- **Challenge:** Large volume N-body simulations to precisely determine BAO 'wiggles' in P(k) or peak in the correlation function.

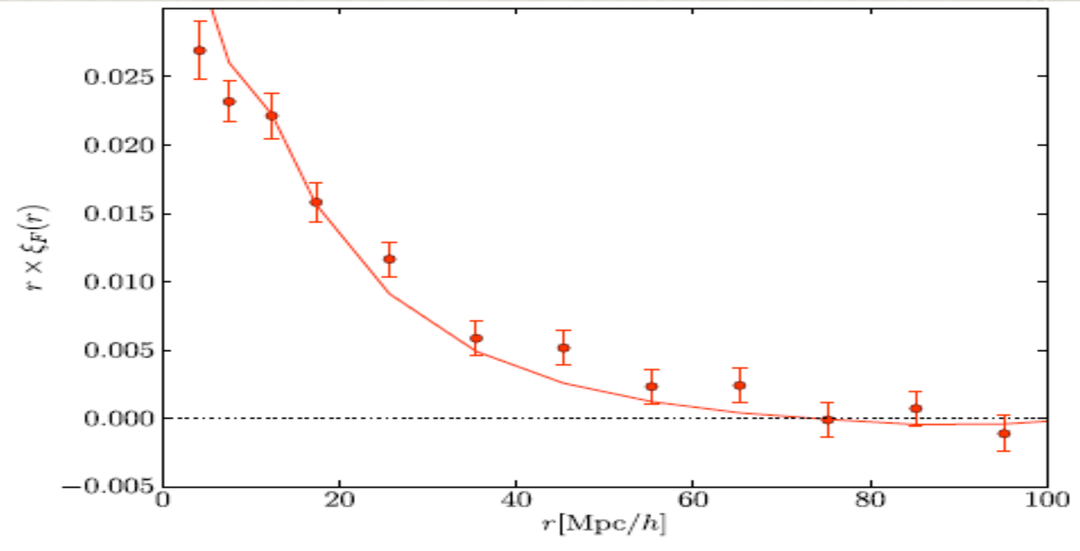
BAO from Lyman-alpha (BOSS)

- **Measurement:** Geometry at $z>2$
- **Challenge:** Large volume N-body +hydro to model the neutral hydrogen distribution.

The Experiments

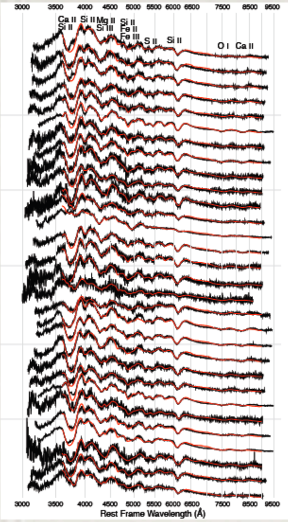
- First 3-D results from BOSS set the stage for cosmological parameter determination (BAO, neutrino masses --)
- To understand these and post-BOSS measurements, requires sophisticated modeling of the intergalactic plasma
  - **Dynamic range:** Controllable since high-density regions give complete absorption
  - **Multi-physics:** Gravity, hydrodynamics, atomic physics, radiative transfer, reionization sources, galactic winds --
  - **Only need to go to z~2**

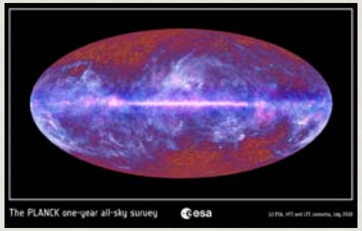


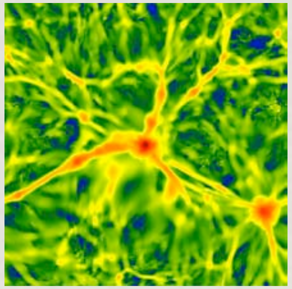


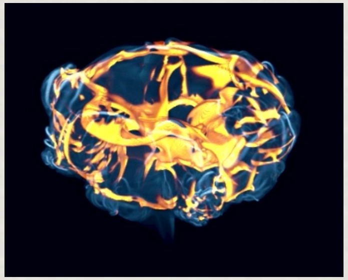
The Group

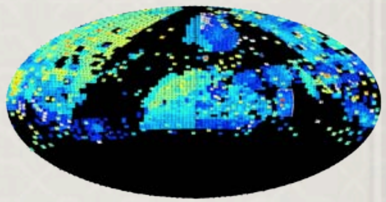
In recent years astrophysics has undergone a renaissance, transforming from a data-starved to a data-driven science requiring leading-edge high performance computing resources. The Computational Cosmology Center is a focused collaboration of astrophysicists and computational scientists whose goals are to develop the tools, techniques and technologies to meet the analysis challenges posed by present and future cosmological data sets.











From radiation-hydrodynamic simulations of supernovae and large-scale structure formation in the universe to the analysis/simulation of 100 of TBs of data from ground- and space-based observatories, C^3 tackles problems that span history from the Big Bang to the present and scales from the quantum level to the observable universe. Research in C^3 spans nuclear and atomic physics, detector simulations, general relativity and radiation transport.

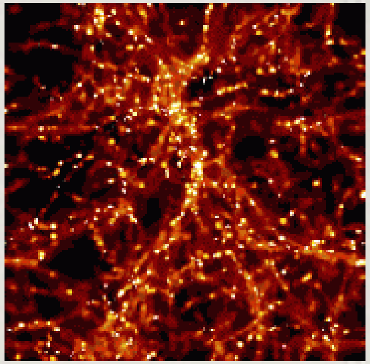
The "Perfect" Code

Our ultimate goal is to calculate absorption spectra caused by neutral hydrogen in the cosmological environment, with all the relevant physical processes.

Simulating the whole observable universe requires a box size of 6000 Mpc/h.

In order to produce a realistic spectrum we need to resolve the Jeans length, the critical radius of a cloud where expansion by thermal energy is balanced by contraction due to gravity, around 60kpc/h.

This requires more than 1000 trillion particles in a single run and/or box sizes 100,000 on a side. At 100 bytes per particle we need 10^5 TB RAM. Such memory-limited computations can be performed only on future Exascale systems (*maybe*), with fault-tolerant algorithms, advanced I/O and sophisticated communication strategies.



Nyx

Nyx is a massively parallel code that solves the

- multicomponent compressible hydrodynamic equations

with a

- general equation of state

and includes

- self-gravity,
- nuclear reactions, and
- radiation.

It is an offshoot of the CASTRO code found in Almgren *et al.*, (2010) ApJ

The Hydrodynamics

The continuity equation

$$\frac{\partial(a\rho_b U)}{\partial t} = -\nabla \cdot (\rho_b UU) - \nabla p + \rho_b \mathbf{g} \ ,$$

The energy equation

Dual-Energy Formalism due to KE >> internal energy

$$\frac{\partial(a^2\rho_b E)}{\partial t} = -a\nabla \cdot (\rho_b UE + pU) + a\rho_b U \cdot \mathbf{g} + a\dot{a}\left((2 - 3(\gamma - 1))\rho_b e\right) + a^2(\Lambda^H - \Lambda^C)$$

The evolution equation for internal energy

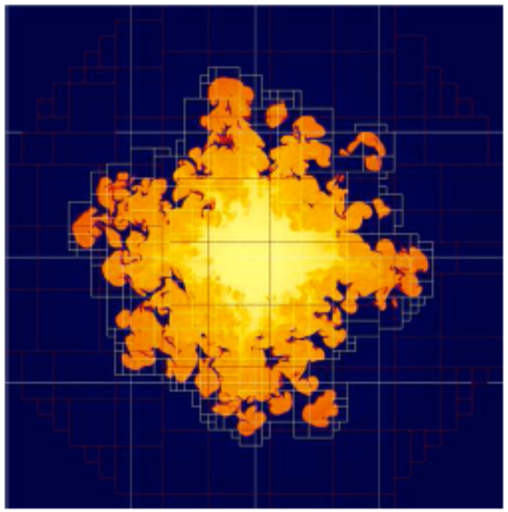
$$\frac{\partial(a^2\rho_b e)}{\partial t} = -a\nabla \cdot (\rho_b Ue) - ap\nabla \cdot U + a\dot{a}\left((2 - 3(\gamma - 1))\rho_b e\right) + a^2(\Lambda^H - \Lambda^C) \ .$$

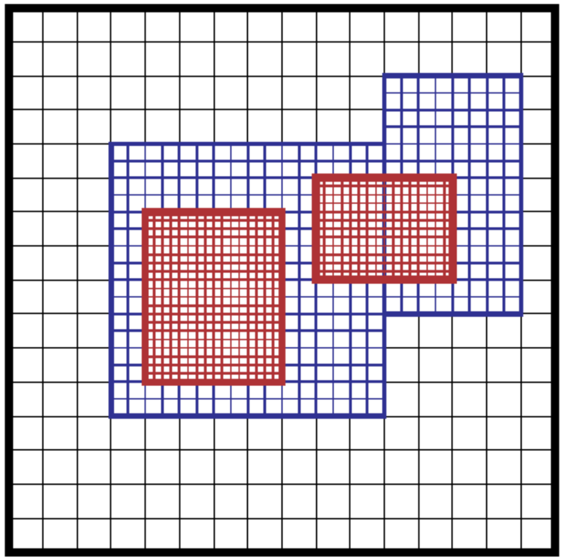
The Poisson equation for gravitational potential

$$\nabla^2 \phi = \frac{4\pi G}{a}(\rho_b + \rho_{dm} - \rho_0)$$

The Mesh

Nyx uses an Eulerian grid with adaptive mesh refinement (AMR). Our approach to AMR uses a nested hierarchy of logically-rectangular grids with simultaneous refinement of the grids in both space and time - the Boxlib formalism.





Comparison to other codes

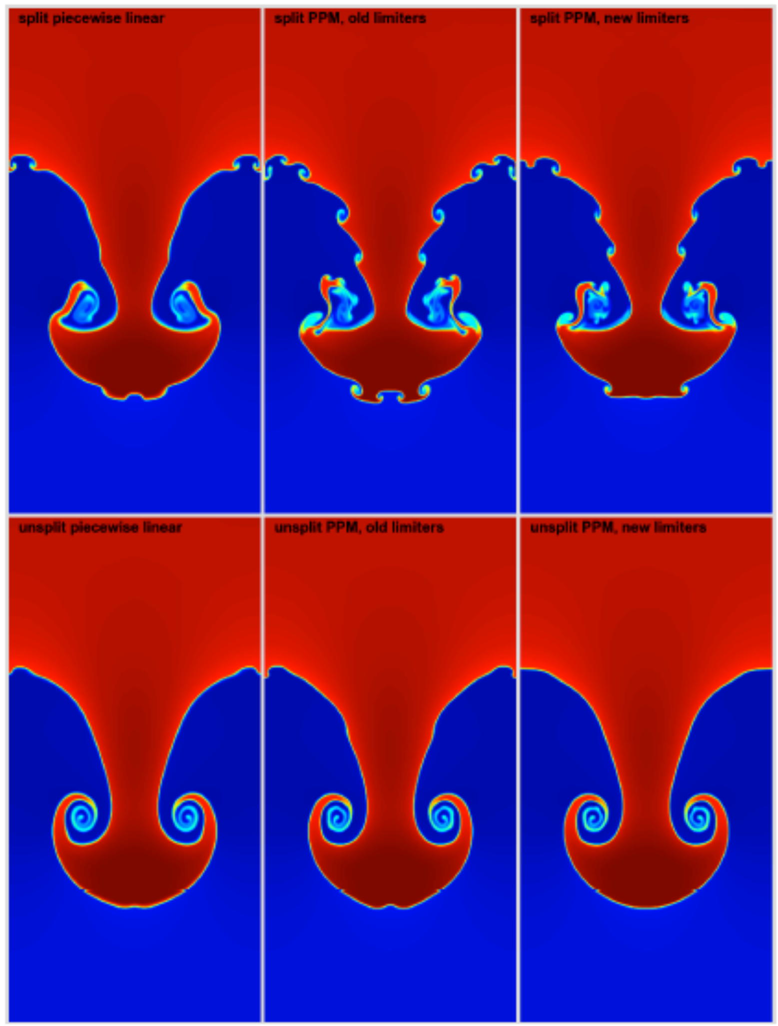
Code	Subcycling?	Split / Unsplit	Refinement
CASTRO	Y	Unsplit	patch-based
FLASH	N	Split	patch-based
ENZO	Y	Split	patch-based
RAGE	N	Split	cell-by-cell

All four codes have structured grid AMR, but differ in how they construct and handle their patches

Nyx vs. Other Codes

A hybrid MPI and OpenMP approach is used, so that we achieve the coarse-grained parallelization of distributing grids to MPI processes and fine-grained parallelization of threading individual loops over many cores.

Nyx uses an unsplit version of the piecewise parabolic method. (Right) Density in a single-mode RT simulation for a variety of advection schemes. Dimensionally split method results are shown on the top row; unsplit method results are shown on the bottom row. The unsplit methods do better at suppressing the growth of high-wavenumber instabilities resulting from grid effects.



Enzo enforces a strict parent-child relationship between patches; i.e., each refined patch is fully contained within a single parent patch; Nyx requires only that the union of fine patches be contained within the union of coarser patches with a suitable proper nesting.

A Good First Step...

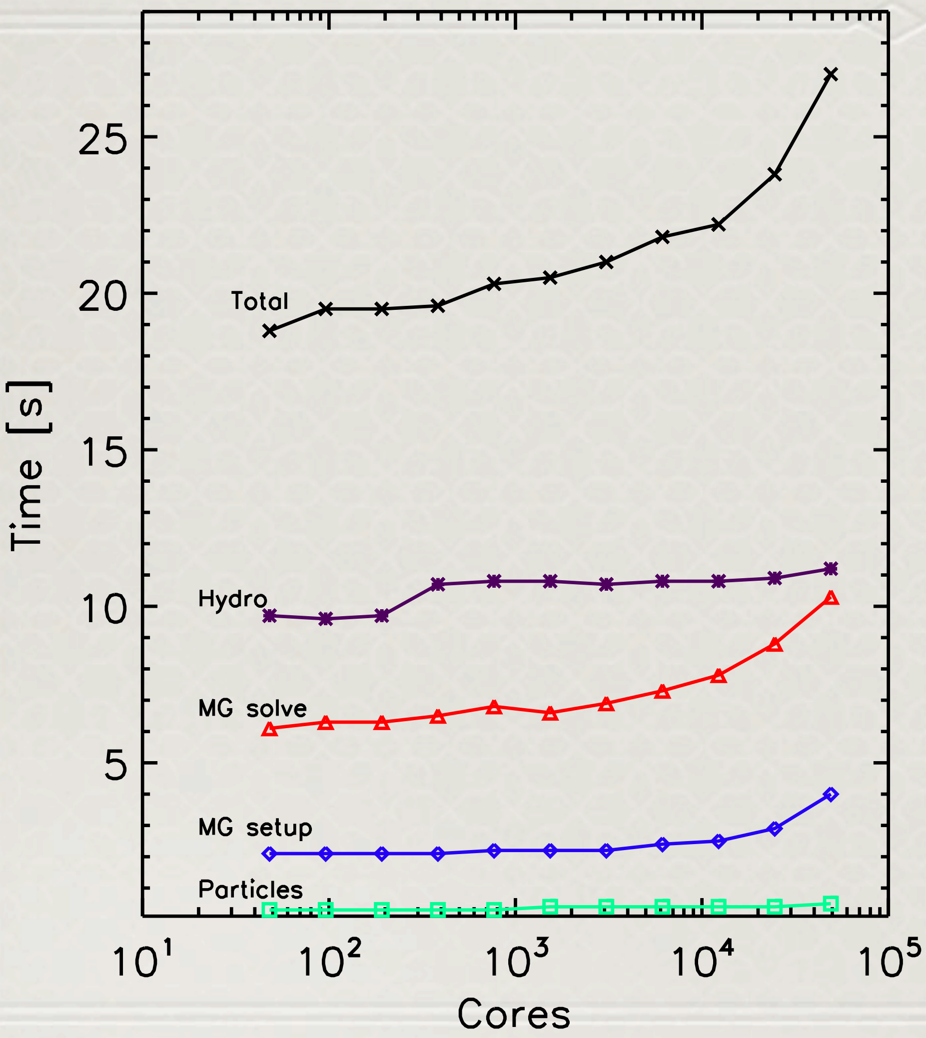
Advantages over existing codes such as Enzo and Gadget-2

Gadget-2 is a smooth particle hydrodynamics (SPH) code. SPH automatically adds refinement (in the form of mass particles) where the mass is. The Lyman-alpha forest problem requires computational resources to be focused on the vast regions of low density rather than the relatively few massive objects.

Nyx includes the capability to solve the Poisson equation for self-gravity on the full adaptive mesh hierarchy with the appropriate matching conditions enforced between grids and between levels. In Enzo, these critical matching conditions are not currently enforced.

Finally neither Enzo nor Gadget-2 were designed to solve the particular problem of the Lyman-alpha forest from day one, with an eye towards Exascale computing.

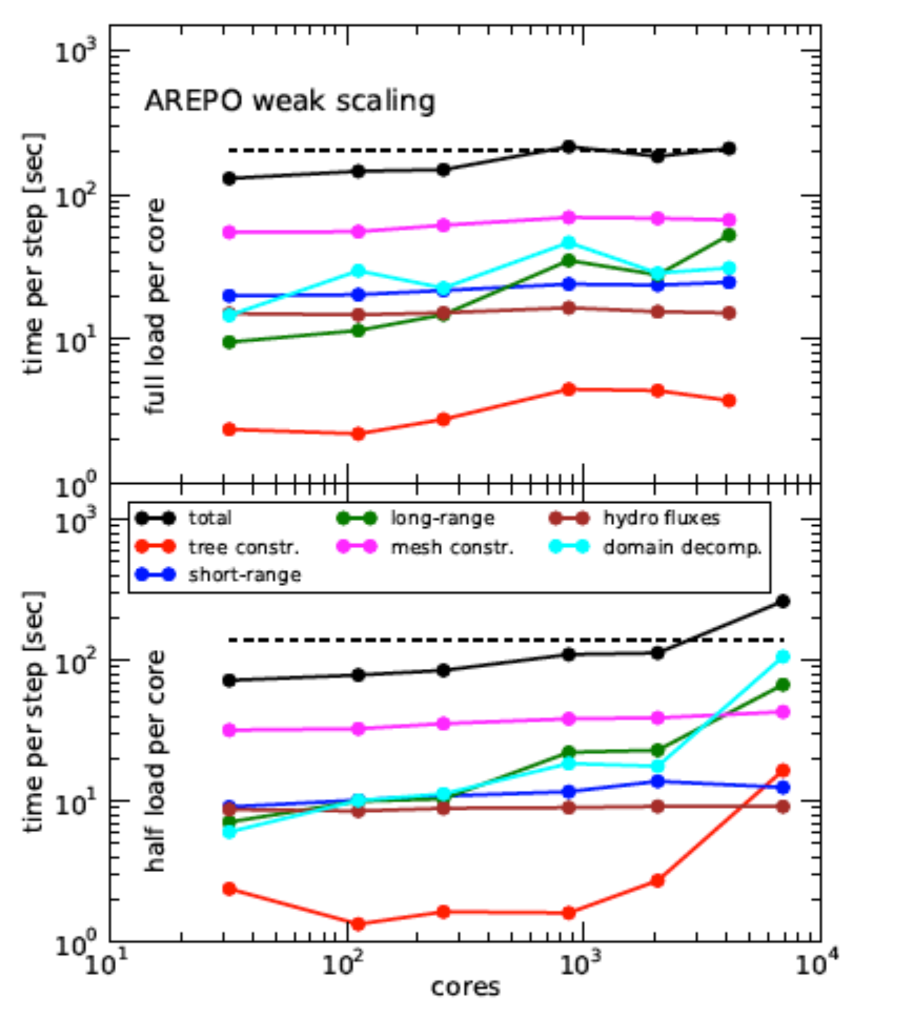
Scaling



Weak Scaling runs done on NERSC's Hopper, a Cray XE6, with 153,216 compute cores. With 2 twelve-core AMD 'MagnyCours' 2.1-GHz processors per node and only 32 GB of ram per node (1.3 GB/core) the machine is good, but not ideal for our setup. Ideal would be at least 2GB ram/core and the number of cores per node = 2^n.

The ***realistic*** results presented here (an actual cosmology calculation which was artificially replicated to keep the size/core the same throughout) show that Nyx can scale to 1/3 of Hopper (~50,000 cores) with just a 40% hit in time for the Multi-Grid Solve (gravity). I/O was able to get 97% of peak (35 GB/s).

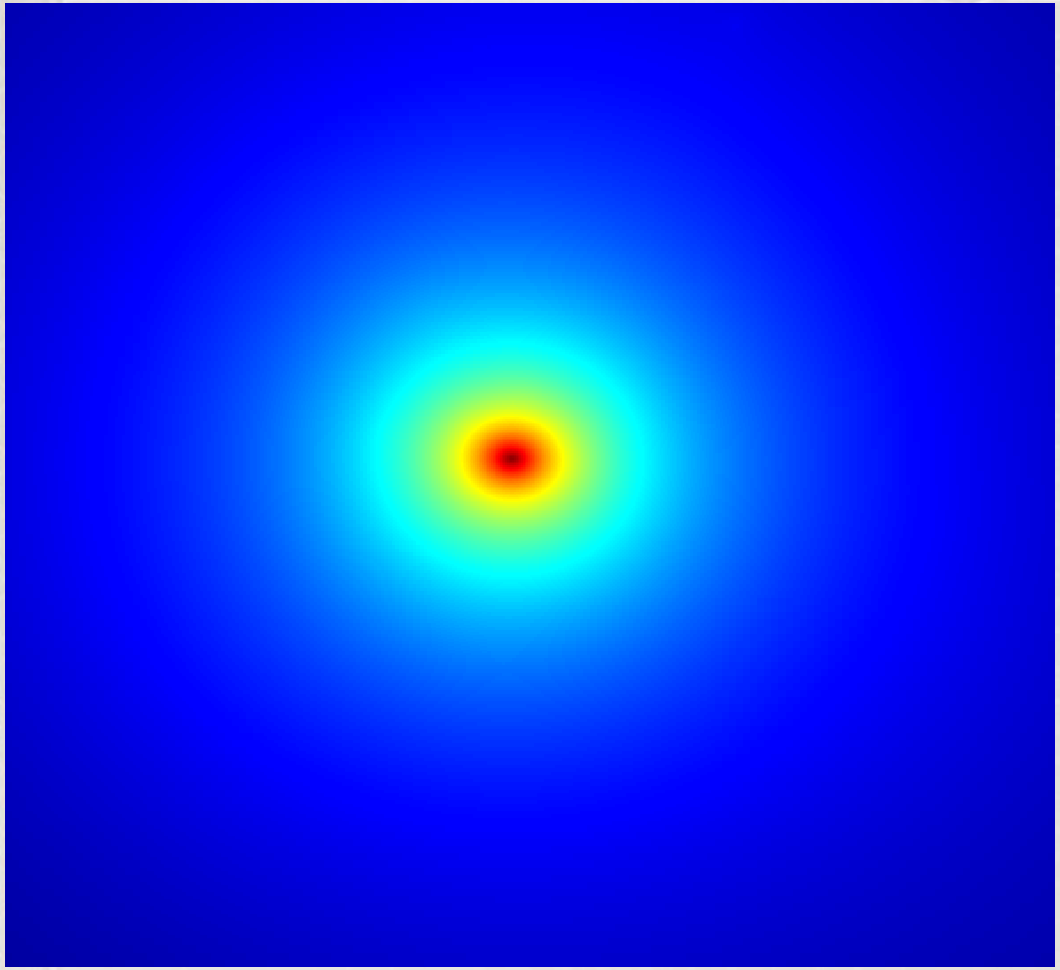
Other Codes



AREPO, a moving mesh N-body+hydro code, is the only other code to date, which has presented scaling tests which approach 10,000 cores. However, due to the mesh construction taking the majority of the time, it is a factor of 10 slower than Nyx per timestep (200s compared to 20s). AREPO is also limited by their FFT solve, running out of memory on Ranger beyond 4096 cores.

Other Hydro codes such as ART and Enzo have not presented scaling beyond ~512 cores, though efforts are underway to improve this.

Santa Barbara Run

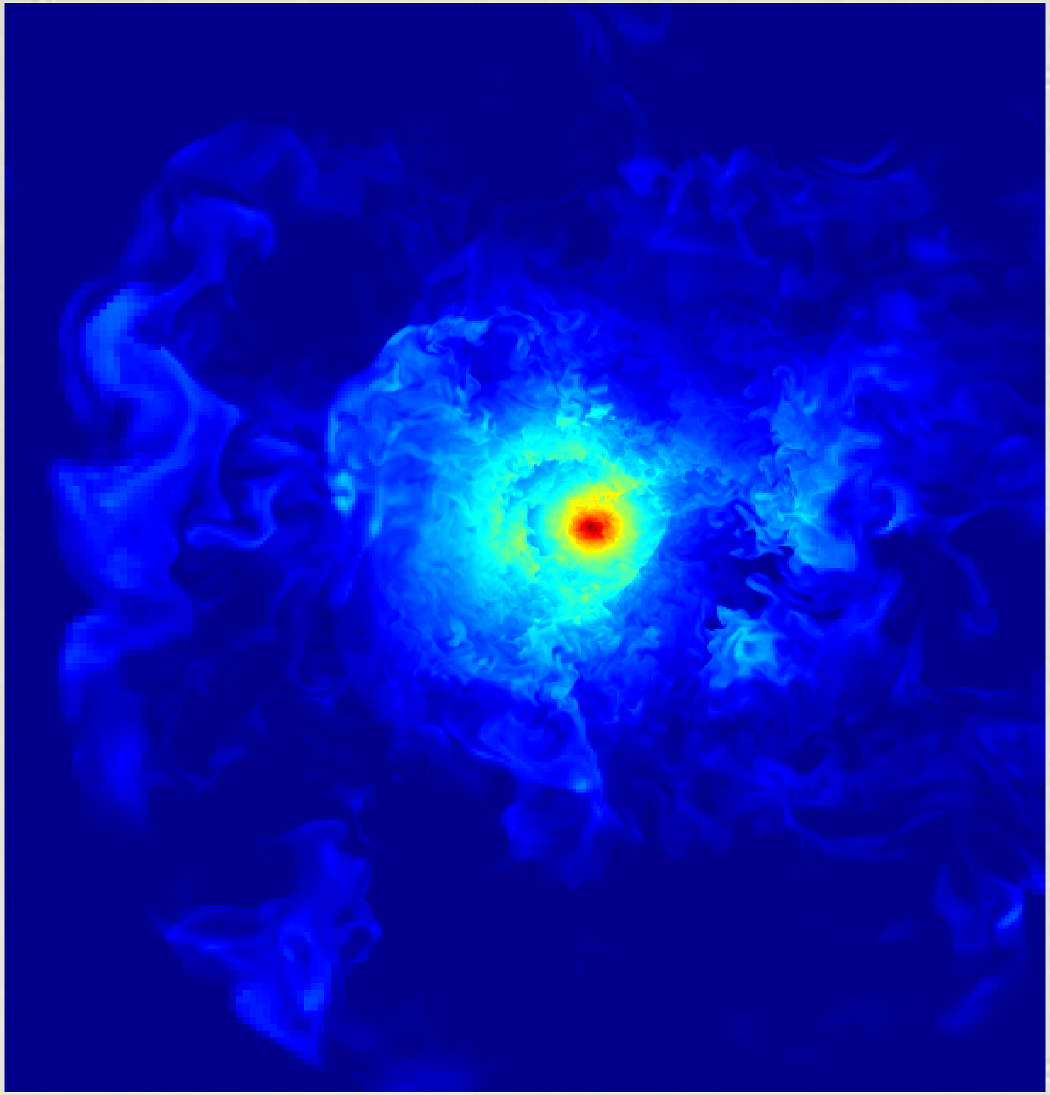


Gravitational Potential (zoom-in)

To obtain a handle on code's accuracy, at a workshop in Santa Barbara in 1995, a plan for detailed test of cluster formation in realistic cosmological setting was laid out.

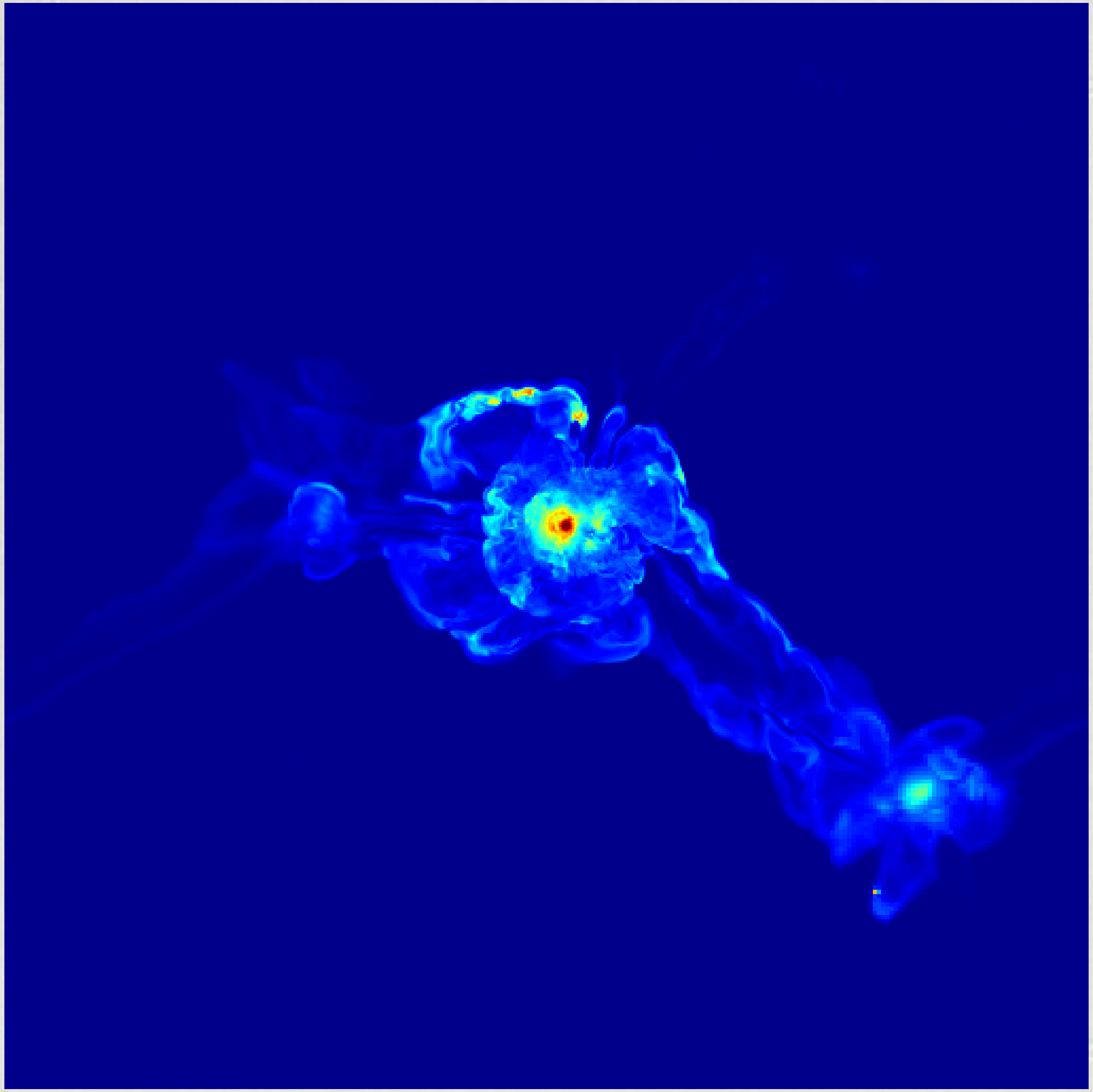
The idea was that all codes would start from the same set of initial conditions, which would be evolved via gravity and hydrodynamics assuming ideal gas equation of state to the redshift z = 0. Initial conditions are constrained such that a large cluster forms in the central part of the box (see Frenk *et al*. 1999).

Santa Barbara Run



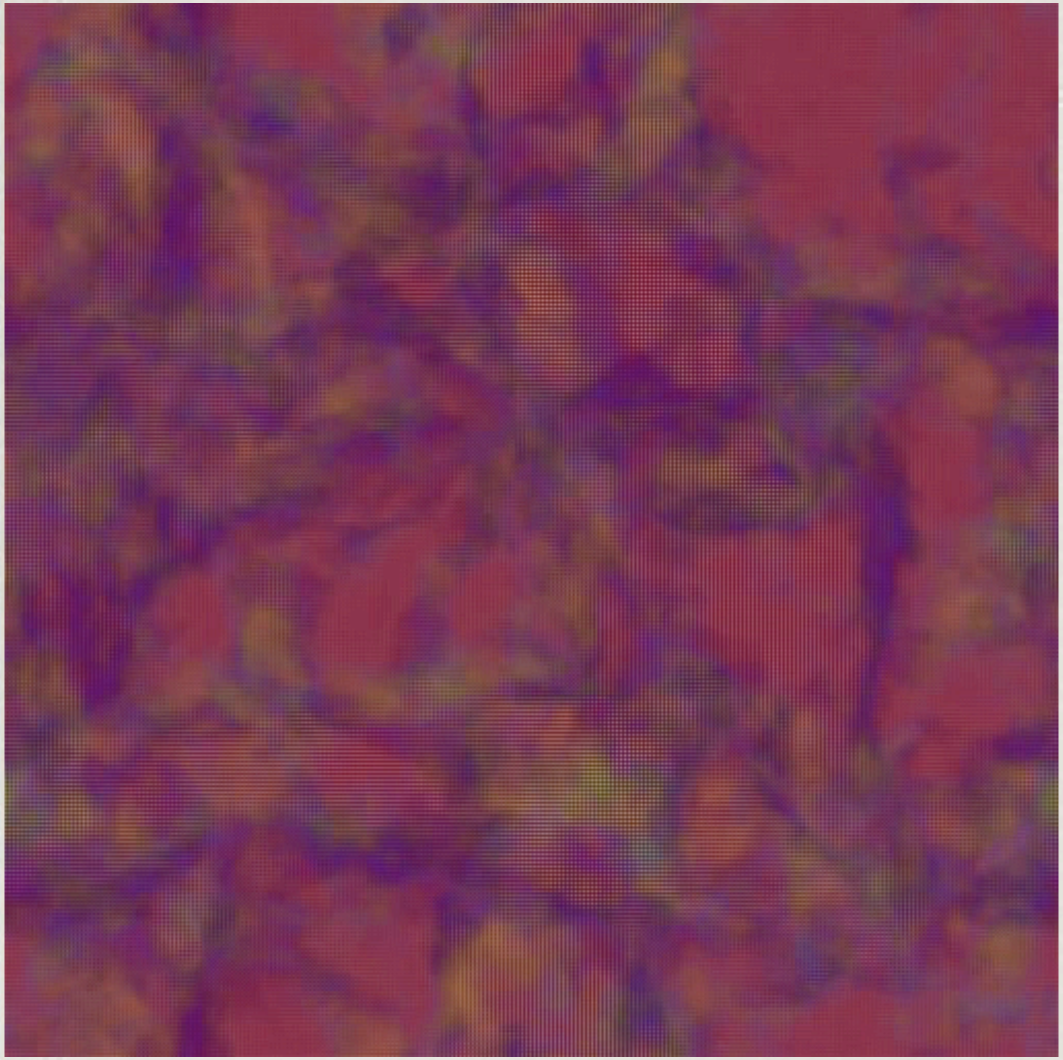
Temperature (zoom-in)

Santa Barbara Run



Temperature (full view)

Conclusions



Nyx is well-positioned to be a competitive N-body+hydro code for the purpose of performing cosmology on the Lyman-alpha forest and cluster formation for some time to come.

First paper submitted this week, first major science runs on Hopper over the next month, including heating+cooling & subgrid models.

It scales well to ~ 10^5 cores as long as the memory per core is reasonable (> 1GB / core).

It is very efficient (run-time wise) given that it is utilizing an unsplit integrator, which may be key in resolving turbulence accurately.