Alleviating the scaling problems of cosmological hydrodynamic simulations with HECA

Renyue Cen (Princeton/Astrophysics)

PRAC Peta-Cosmology:
Galaxy Formation and Virtual Astronomy
NCSA, Blue Waters Symposium – May 21, 2013

on behalf of CAGE Consortium
Ken Nagamine (UNLV), Ludwig Oser (Columbia/MPA), Jeremiah P. Ostriker (Columbia), Thorsten Naab (MPA), Greg Bryan (Columbia), Eve Ostriker (Princeton), Renyue Cen (Princeton)
Goal: simulating the nonlinear universe
Galaxy formation and evolution: a multi-scale, multi-physics, multi-astrophysics problem

From stars to large-scale cosmic web: 17 decades in mass dynamic range, 16 decades in spatial dynamic range

Galaxies assemble and take shape

A bunch of equations to solve: Euler, Poisson, Einstein, Newton, Vlasov, atomic physics, molecular physics, ...

- Baryonic mass accretion?
- Angular momentum?
- Timescales/mergers?
- Star formation efficiency?
- Inside-out galaxy formation?
- Connection between bulge and disk formation?
- Feedback AGN, SNII, SNIa etc.? Dissipation?
- Environment? Evolution as function of mass?
- Relation between dark matter and baryons?
- Assembly of galaxies?

Images from NASA/WMAP Science Team and Hubble Heritage
$z=6$

12.7 Gyr

100 Mpc $= 3.3 \times 10^8$ ly
$z=4$

12.0 Gyr

100 Mpc $= 3.3 \times 10^8$ ly
$z = 2$

10.2 Gyr

$100 \text{ Mpc} = 3.3 \times 10^8 \text{ ly}$
$z=1$

7.7 Gyr

100 Mpc = $3.3 \times 10^8$ ly
$z=0$
today

$100\text{ Mpc} = 3.3 \times 10^8\text{ ly}$
Zoom-in Resimulations

$100^3 \text{ Mpc}$, $512^3$ particles
dark matter only, 100
snapshots (WMAP3: $\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, $h = 0.72$)

Trace back particles that
will form a gravitationally
bound structure at the
present day
Zoom-in Resimulations

Particles are replaced with gas and dark matter particles at a higher resolution level.

Simulation is redone including radiative cooling and star formation.

Cosmological context is preserved!
An Zoom-in Galaxy Formation Movie
Star formation history

(Cen 2011), Tonnesen & Cen (2013)
Galaxy Size evolution

(Oser et al. 2012)
Metals content evolution of damped Lyman alpha systems

Rafelski et al. 2012
Rafelski et al. 2012 $<Z>$
$<Z>$ best fit
This work

(Cen 2012)
Need Simulation Statistics

- Sloan digital sky survey (SDSS): spectra of nearly one million local galaxies and 10s million more with photos in multiple bands
- Need to take full advantage of the data to test statistically cosmological models and physics of galaxy formation
- Two approaches:
  Full-box simulation <> ensemble computing
- Including SPH, AMR
- Full-box simulations: $T_{cpu} \sim N^{1.73}$
Zoom-in

- Largest sample of cosmological zoom-in simulations so far (up to $3 \times 10^{14} \, M_\odot$)
- Successful in explaining a variety of properties of galaxies (sizes, star formation history, intergalactic and circumgalactic medium, age distribution of stars, kinematics)
- Still small number statistics (100-1000s), when compared much larger observational samples (10s millions), especially for massive galaxies that need large zoom-in boxes and expensive to compute

I.E.: sims are way behind observations in terms of statistics
Zoom-in Sims Pros and Cons

- Cons: “Sight lines” through simulation volume as well as 3-d regions not contiguous
- Pros: Long-range baryonic effects, e.g. reionization of the universe
- Much higher resolution possible in zoom-in simulations
- Different cosmological models (dark energy equation of state) and astrophysics (supernovae, supermassive black holes, etc) can be tested
HECA

- Hierarchical Ensemble Computing Algorithm

- Embarrassingly parallel problem:
  Instead of increasing the number of processors with the problem size, the number of simulations is increased, i.e. $T_{CPU} \sim N^1$

- Overhead for having to resimulate the background is negligible due to adaptive time stepping along with spatial AMR
Some Scalings

- We have already reduced $N^2$ scaling to Tree code scaling: $N_p \cdot \log(N_p)$
- Separation: $x \cdot N_z = N_p$

$$\frac{N_p \cdot \log(N_p)}{x \cdot N_z \log(N_z)} = \frac{N_p \cdot \log(N_p)}{x \cdot N_p/x \log(N_p/x)} = \frac{\log(N_p)}{\log(N_p) - \log(x)}$$

- The higher the number of zoom-simulations ($x$) the more time we save
Scheduler

Generate Initial Conditions

Assign CPUs
An Example
Scalings: separation saving

Wall-clock time

Combined: 165842.47
Separate: 41246.71, 37916.79, 26590.39, 19820.55

30% saving
Scalings: more saving /w more cores

Wall-clock time

Combined: 165842.47
Separate: 41246.71, 37916.79, 26590.39, 19820.55
Combined - 4x: 102140.19

60% saving
Scalings: more saving with high res

Wall-clock time

70% saving

good news
Scalings with 0.1 Gigahours

- A factor of 8 Increase in simulation volume

Graph showing CPU-h vs. Vol [Mpc^3] with lines for whole box and HECA regions.
Outlook

- Hybrid (OpenMP + MPI) approach, 20%-30% increase in performance
- Port to Blue Waters
- Run a very large parent simulation
- More, extended scaling tests
- Convert ICs for grid based codes (ENZO, AMR, TVD) as well as particle based codes
Conclusions

- Higher resolution, larger samples of galaxy formation simulations enabled in HECA than in full-box simulations
- Scalable up to arbitrarily large processor counts
  ➔ Statistical relevant sample of galaxies at high resolution
- Cosmological models (dark matter, dark energy, ...) and different physical/astrophysical models can be implemented and tested in a feasible time scale
- So if we have cent$ and computers, we will inform you what kind of universe you live in (and if you are unique)
Thank You

Acknowledgement:
This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (award number OCI 07-25070) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign, its National Center for Supercomputing Applications, Cray, and the Great Lakes Consortium for Petascale Computation.