Simulating the Universe

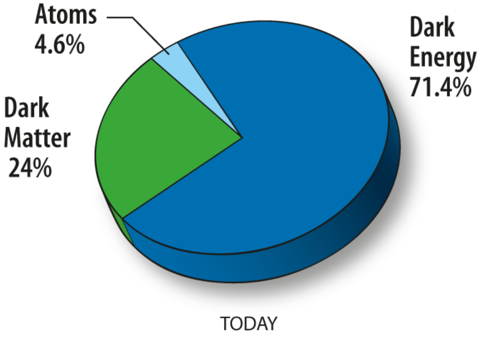
Outline

- What do we do?
- How do we make a Universe inside a computer?
- The current state-of-the-art.
- How can we do better in the future?

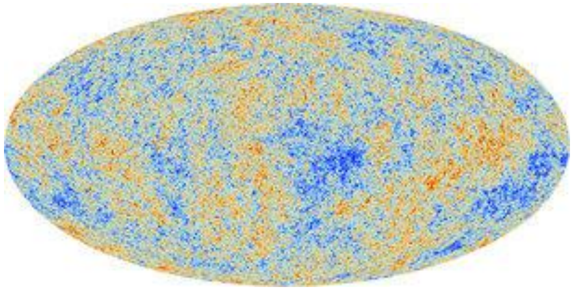
How to make a Universe

Ingredients

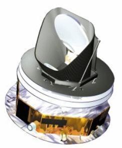
- Only ~20% of the matter in the Universe is ‘normal’: protons, neutrons, etc..
- The remaining ~80% is comprised of dark matter.x



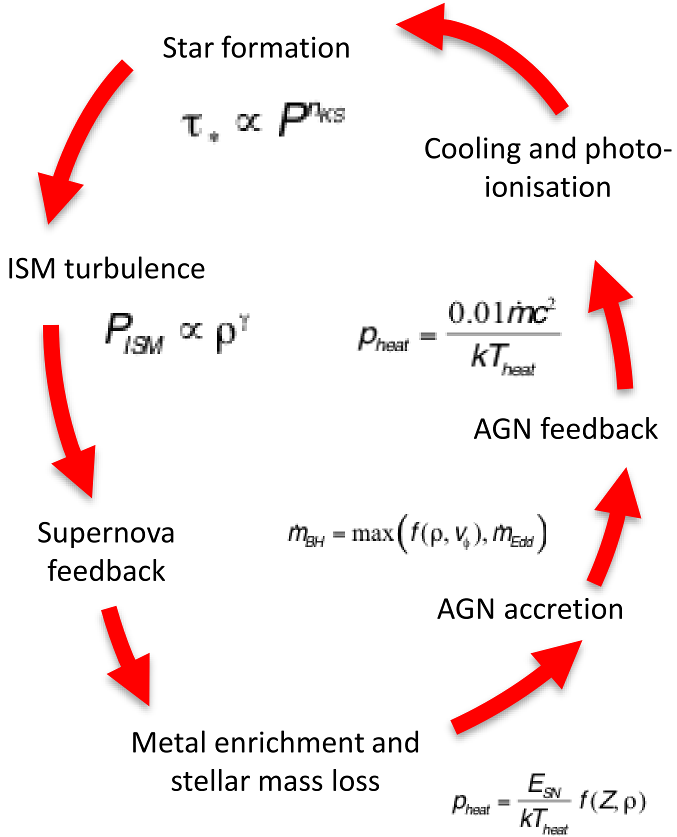
Initial conditions



- Light from the CMB tells us the conditions of the Universe when it was “only” 379,000 years old (About a day in the lifetime of a human).
- This tells us the initial conditions for our simulations.



The laws of physics



Star formation

$\tau_* \propto P^{n_{KS}}$

Cooling and photo-ionisation

ISM turbulence

$P_{ISM} \propto \rho^{\gamma}$

$p_{heat} = \frac{0.01\dot{m}c^2}{kT_{heat}}$

AGN feedback

Supernova feedback

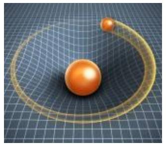
$\dot{m}_{BH} = \max\left(f(\rho, v_{\phi}), \dot{m}_{Edd}\right)$

AGN accretion

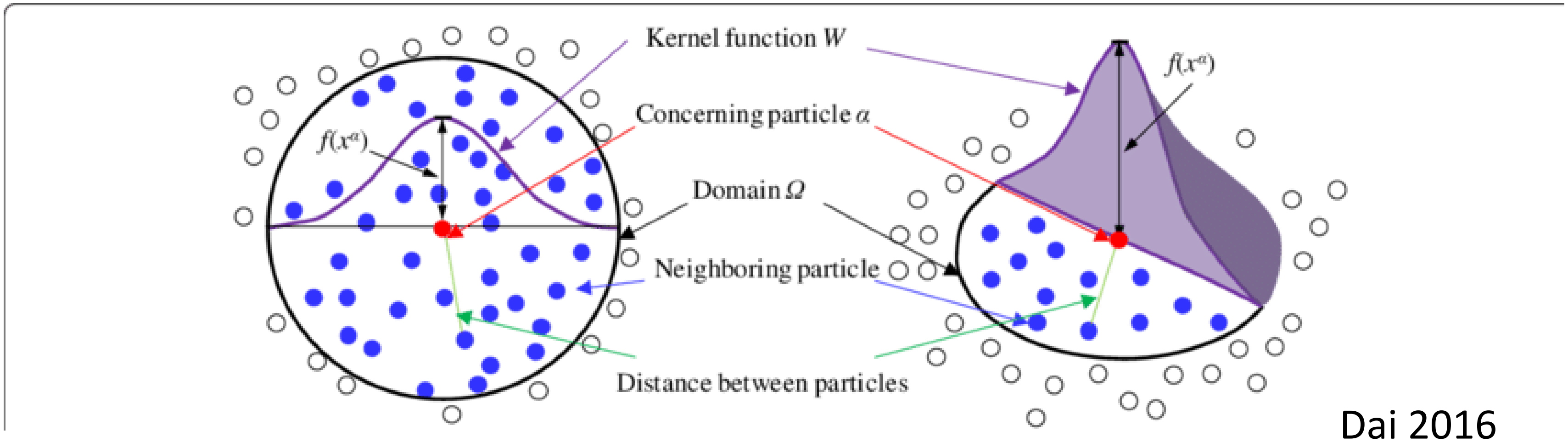
Metal enrichment and stellar mass loss

$p_{heat} = \frac{E_{SN}}{kT_{heat}} f(Z, \rho)$

Gravity

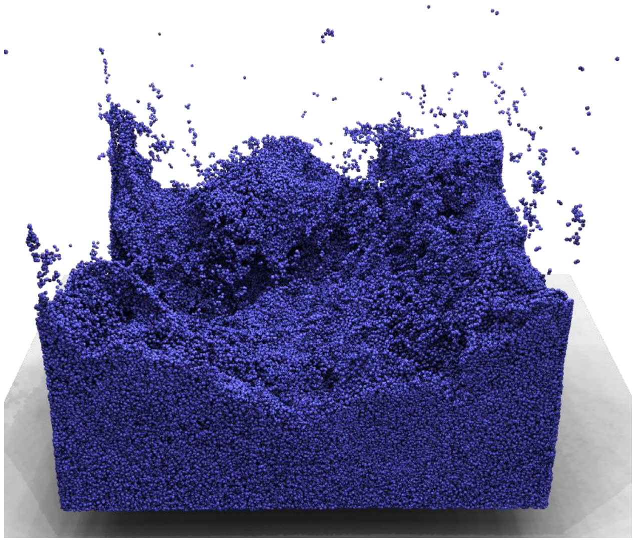


Smoothed particle hydrodynamics (SPH)



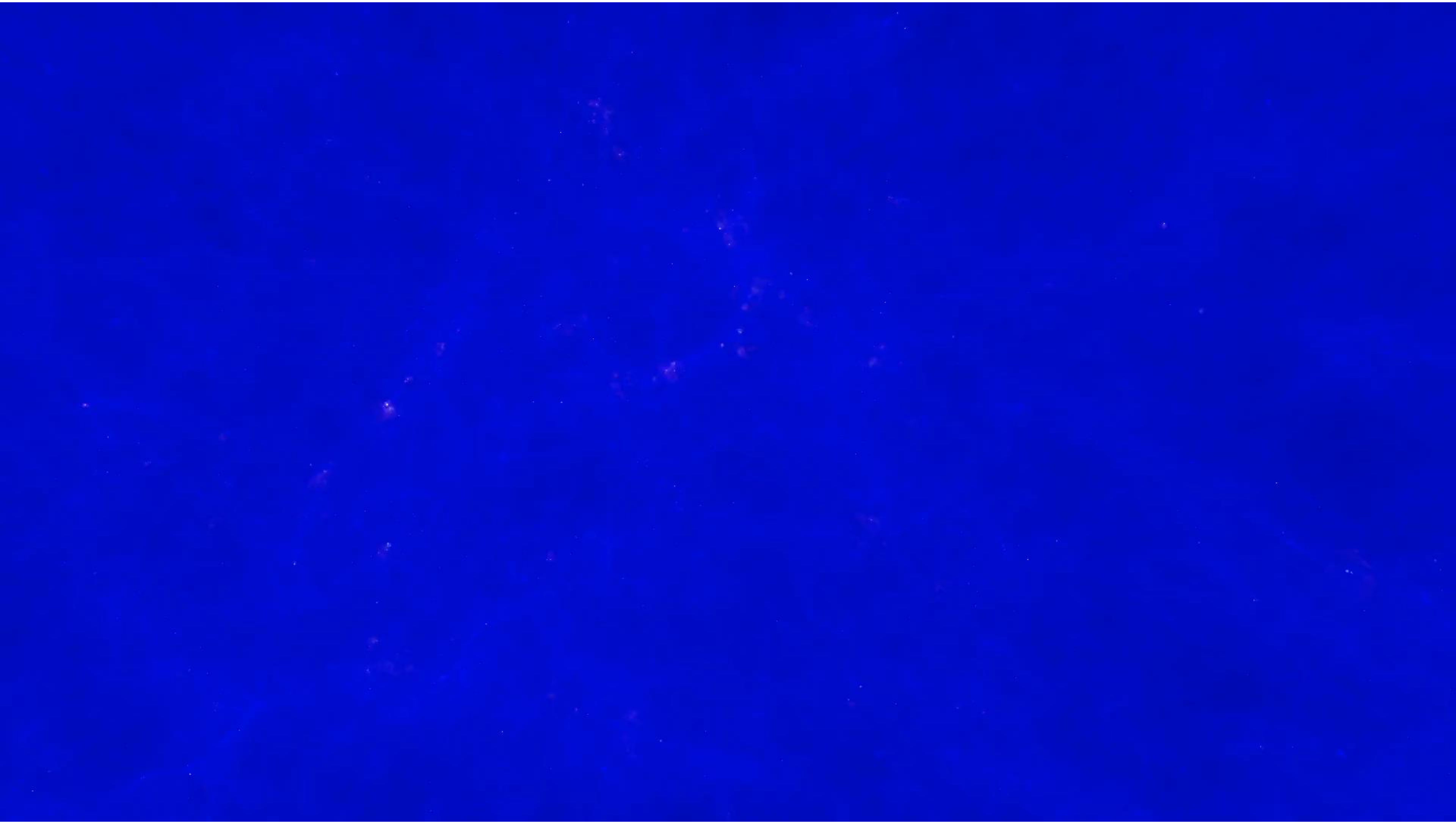
Dai 2016

- Aim to represent the contents of the Universe via a finite number of discrete "particles".
- These particles typically have a mass of around 1-10 million times the mass of our sun.



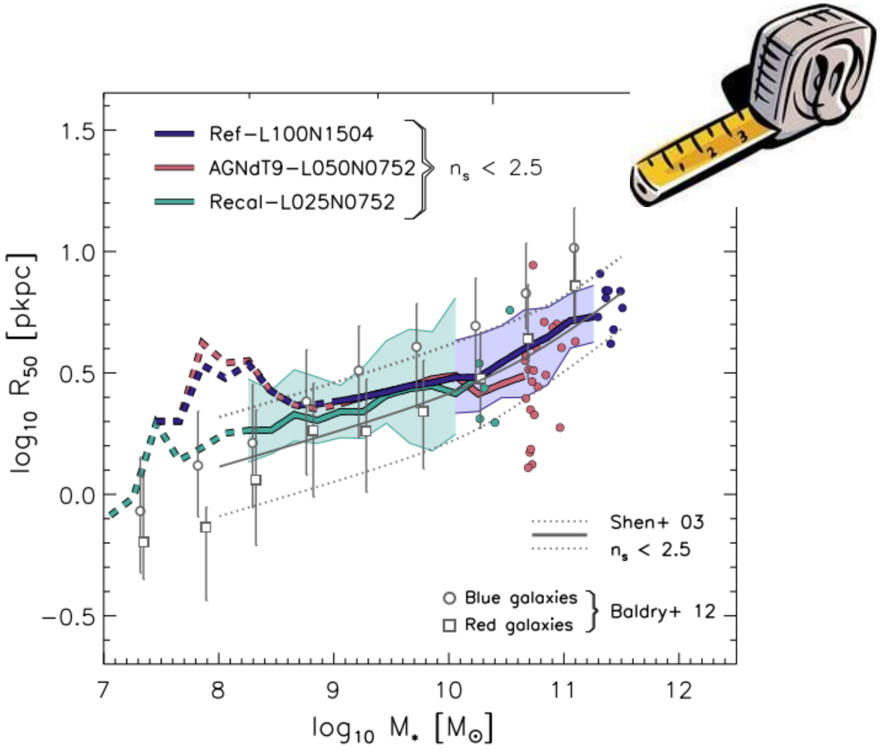
http://web.cse.ohio-state.edu/~wang.3602/courses/cse3541-2014-spring/proj_final/Ting-Chun_Sun/images/sph_particles2.png

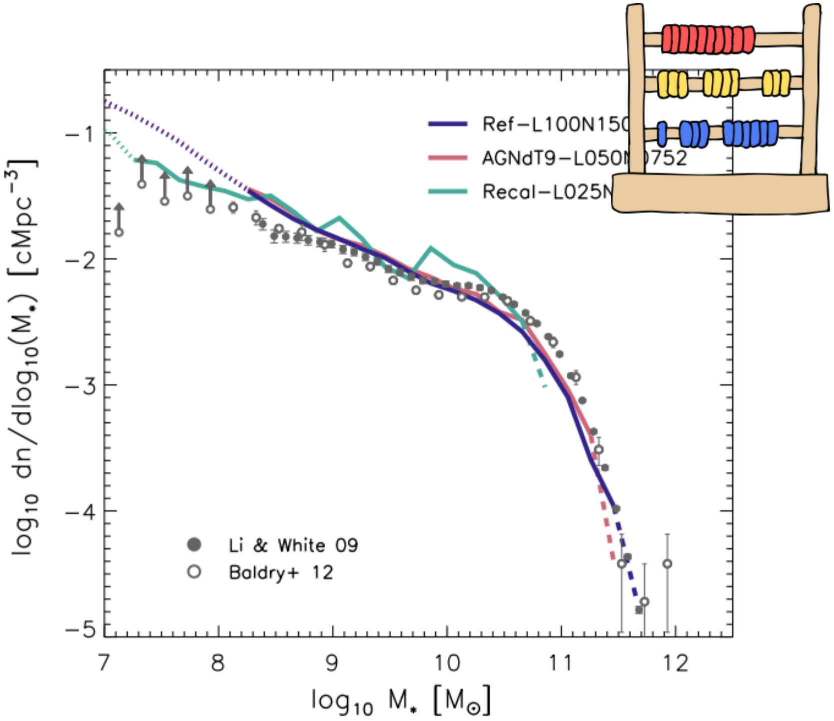
The EAGLE simulation



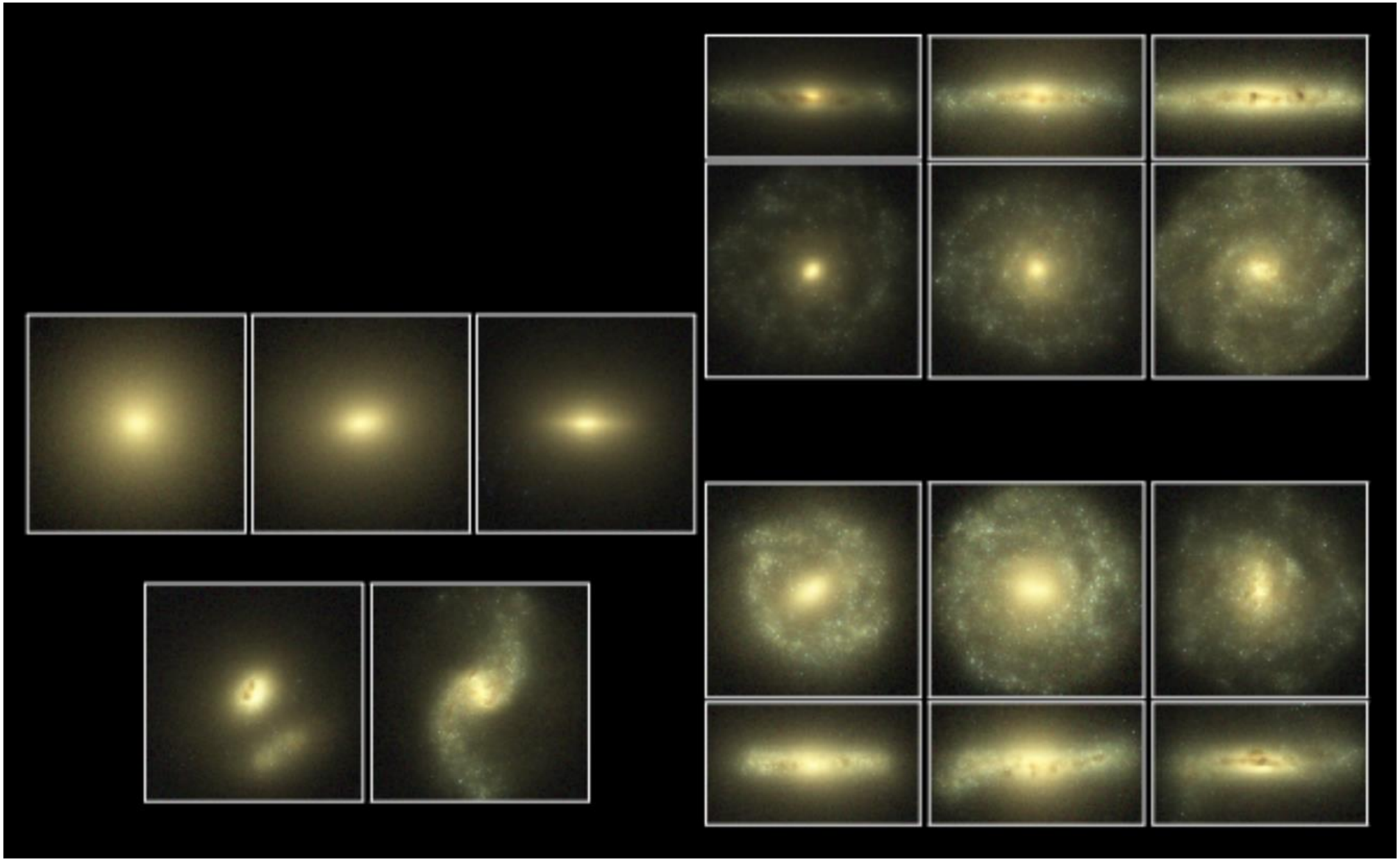
How do we know we’re right?

- We need to compare with empirical measurements of the Universe today.
- For example, do we have enough galaxies of the right mass, size, etc…



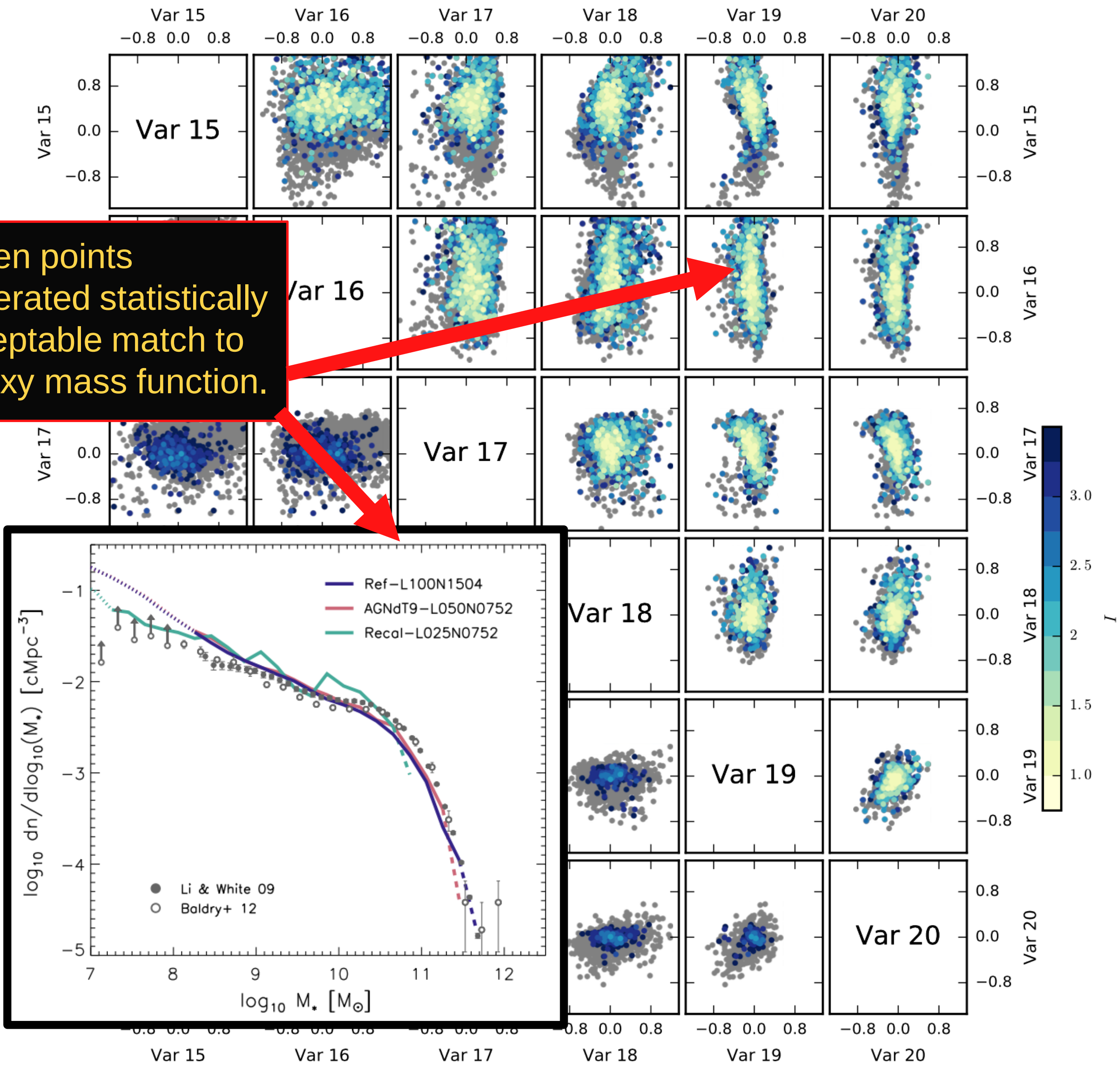


How do we know we're right?



Exploring the parameter space

- Acceptable luminosity function fits can be found over a wide range of parameter space.
- …but this is a projection effect, the acceptable space lies on thin planes (0.05% of volume).
- Each plot shows a pair of model parameters (uninteresting ones are suppressed).



Green points generated statistically acceptable match to galaxy mass function.

The computational cost

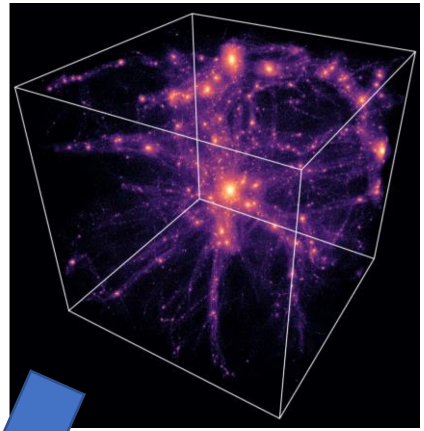
- The largest scientific run is 300,000,000 cubic light years in volume.
- ~3.4 billion resolution elements of gas and the same again for dark matter.
- Took 43 continuous days to run on 4096 cores, using 32 Tb of memory.

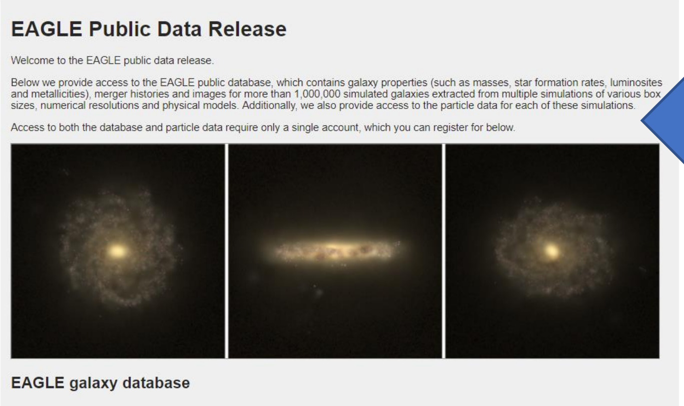
The cosmology machine

~0.5 Pb of disk space

A lot of post-processing

- After the simulation is run, we have to "find" the galaxies.
- This is computationally very expensive, and may not even be possible in the current state with the next generation of simulations.
- Many billion particles reduce to a few tens of thousands of galaxies.





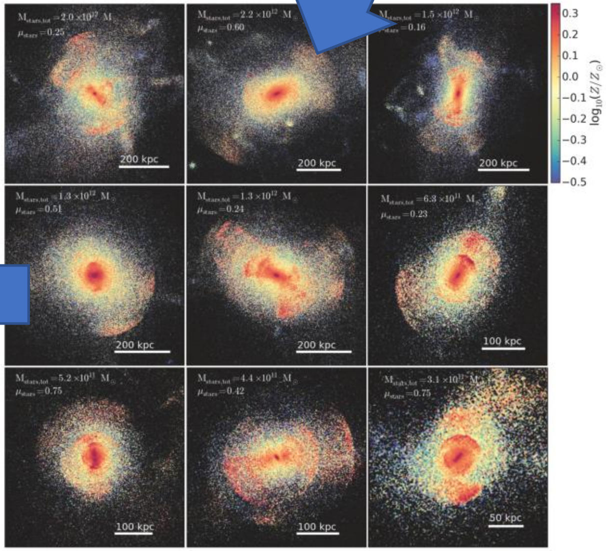
EAGLE Public Data Release

Welcome to the EAGLE public data release.

Below we provide access to the EAGLE public database, which contains galaxy properties (such as masses, star formation rates, luminosites and metallicities), merger histories and images for more than 1,000,000 simulated galaxies extracted from multiple simulations of various box sizes, numerical resolutions and physical models. Additionally, we also provide access to the particle data for each of these simulations.

Access to both the database and particle data require only a single account, which you can register for below.

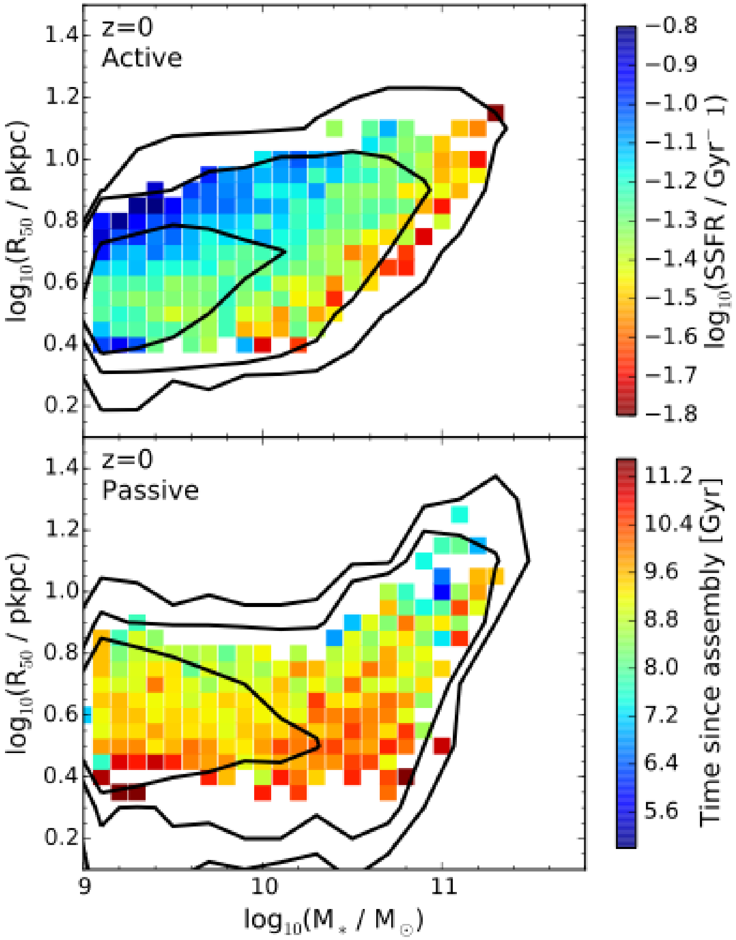
EAGLE galaxy database



What’s next?

EAGLE-XL

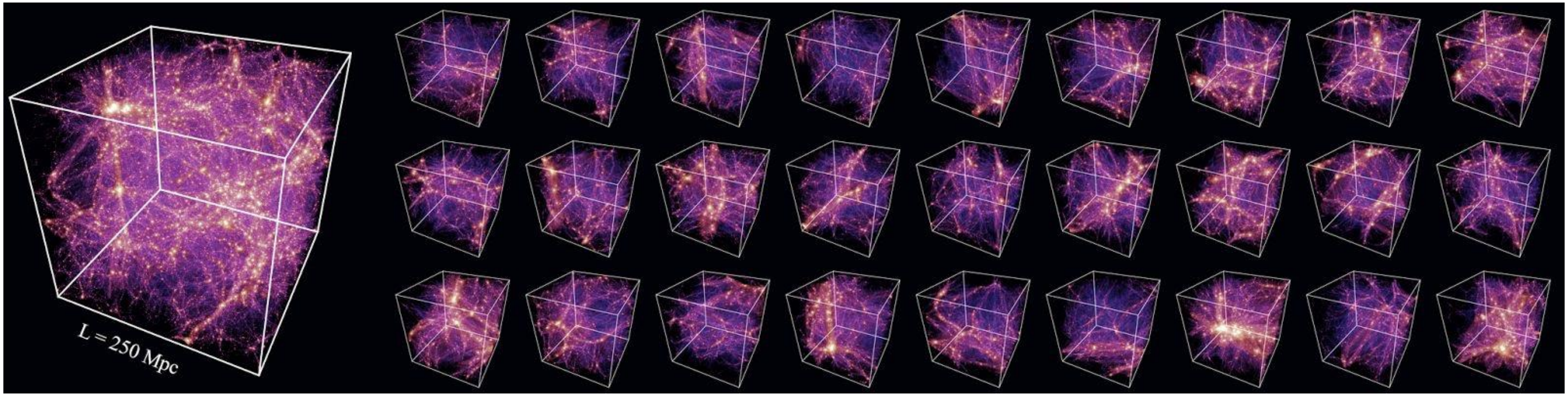
- Higher sampling of galaxies: 100,000 galaxies to understand what makes galaxies unique
- Enough volume to construct light-cones without excessive replication
- Enough volume to sample correlations and redshift space distortion on quasi-linear scales



In EAGLE, replicating even a modest 300 sq. deg. redshift survey of local galaxies, such as GAMA, requires each galaxy to be repeated 12 times. The SDSS survey requires each galaxy to be repeated 360 times!

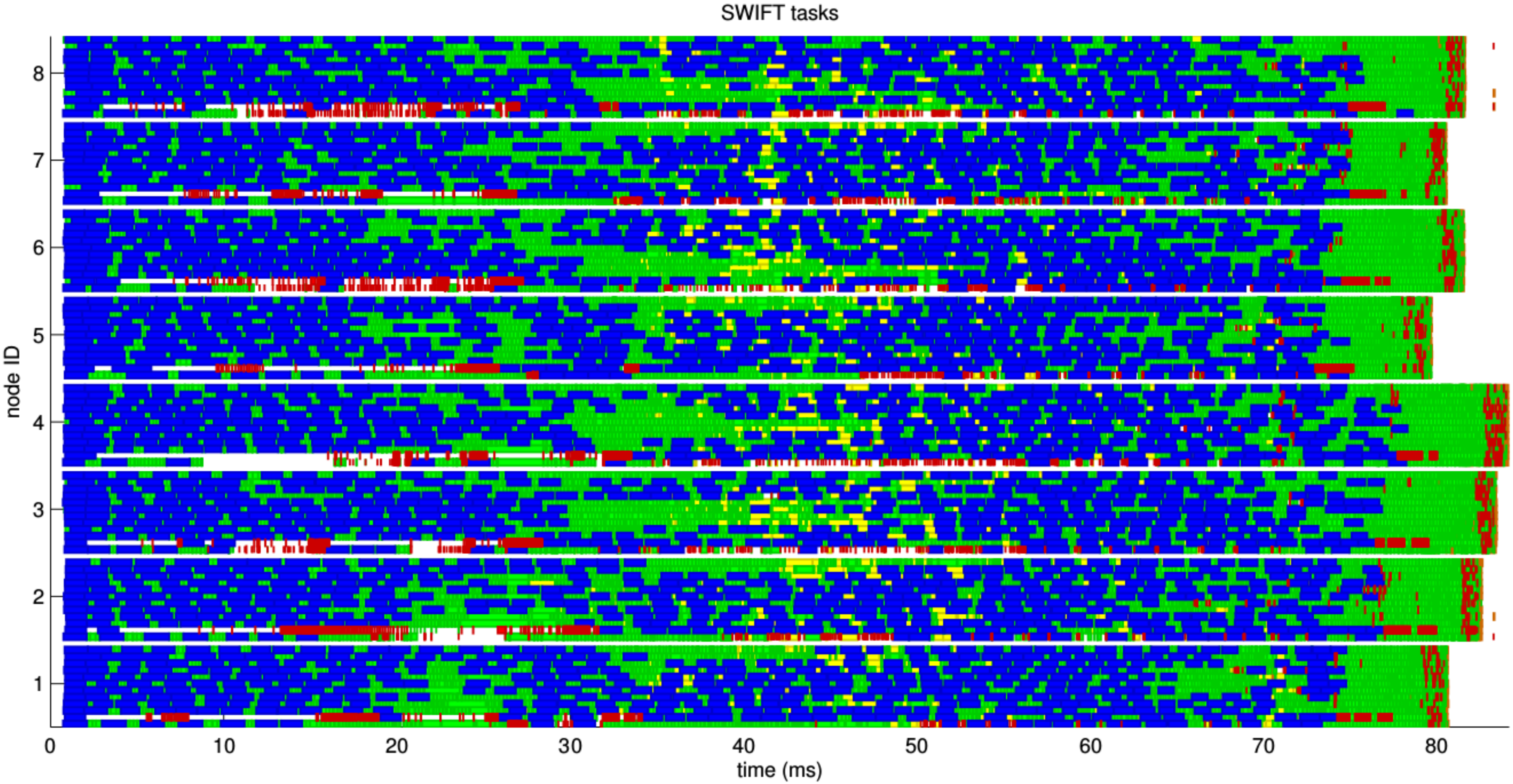
Challenges for EAGLE-XL

Simulating such a large volume (100 billion particles!)
- tiling
- in-flight analysis
- deleting the data!



To simulate a contiguous volume, we need to choose the boundaries between regions carefully

The next generation - SWIFT

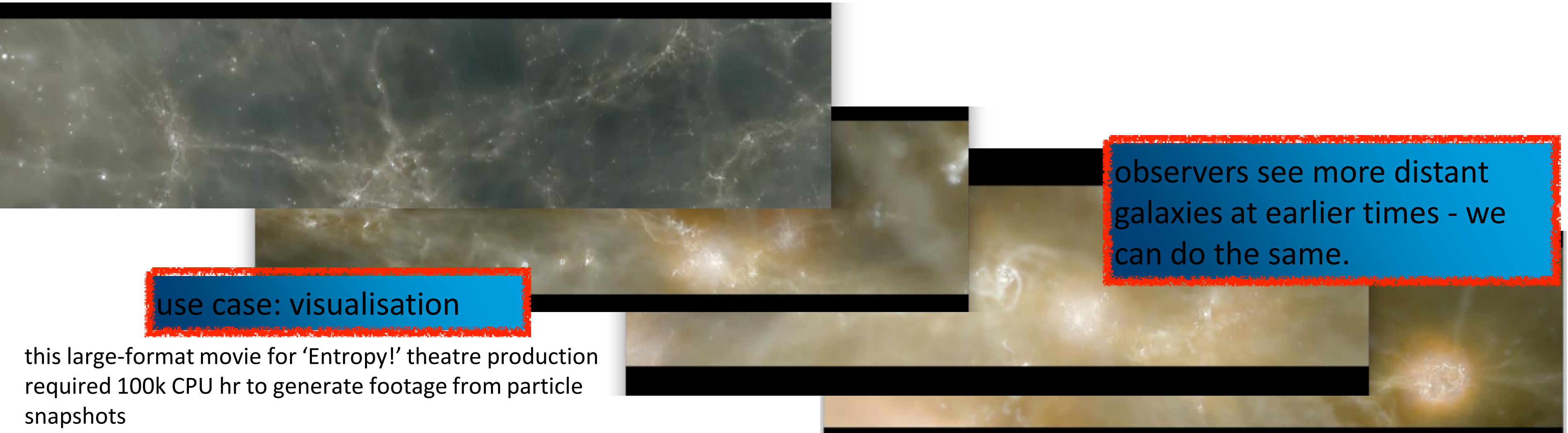


1M particle SPH simulation using SWIFT on 8 × 12-core nodes (Intel X5650 CPUs). Red indicates communication tasks.

Parallel Streaming I/O

- Simulation I/O is now a major bottle neck
  - speed gains from SWIFT means most of the time will be spent writing files
- Old method:
  - stop the simulation!!
  - output a “snapshot” of all particles
  - this is hopelessly inefficient!

- New method:
  - adaptively output particle updates only when needed
  - stream particle updates to local memory mapped file
  - integrate output into the task-graph
  - reconstruct “snapshots” for interesting regions in post-processing



use case: visualisation

this large-format movie for ‘Entropy!’ theatre production required 100k CPU hr to generate footage from particle snapshots

observers see more distant galaxies at earlier times - we can do the same.