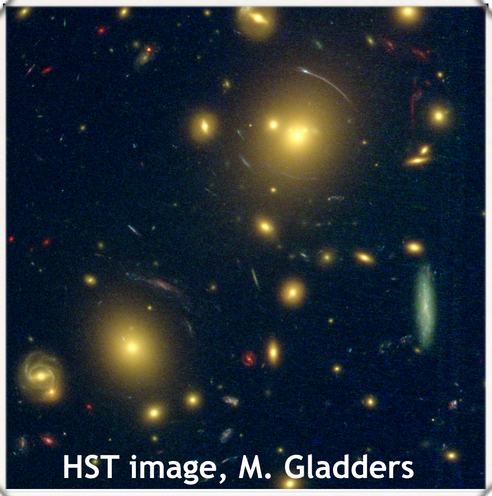
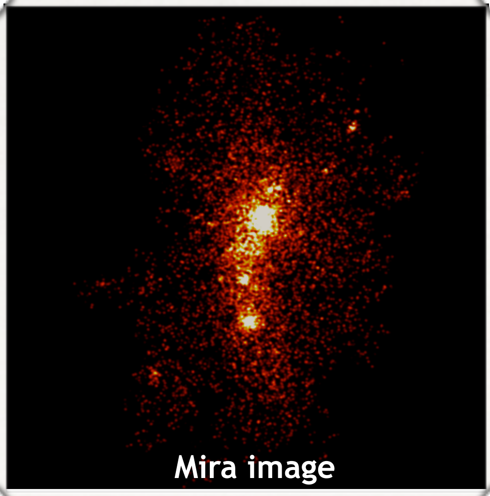
Cosmic Structure Probes of the Dark Universe

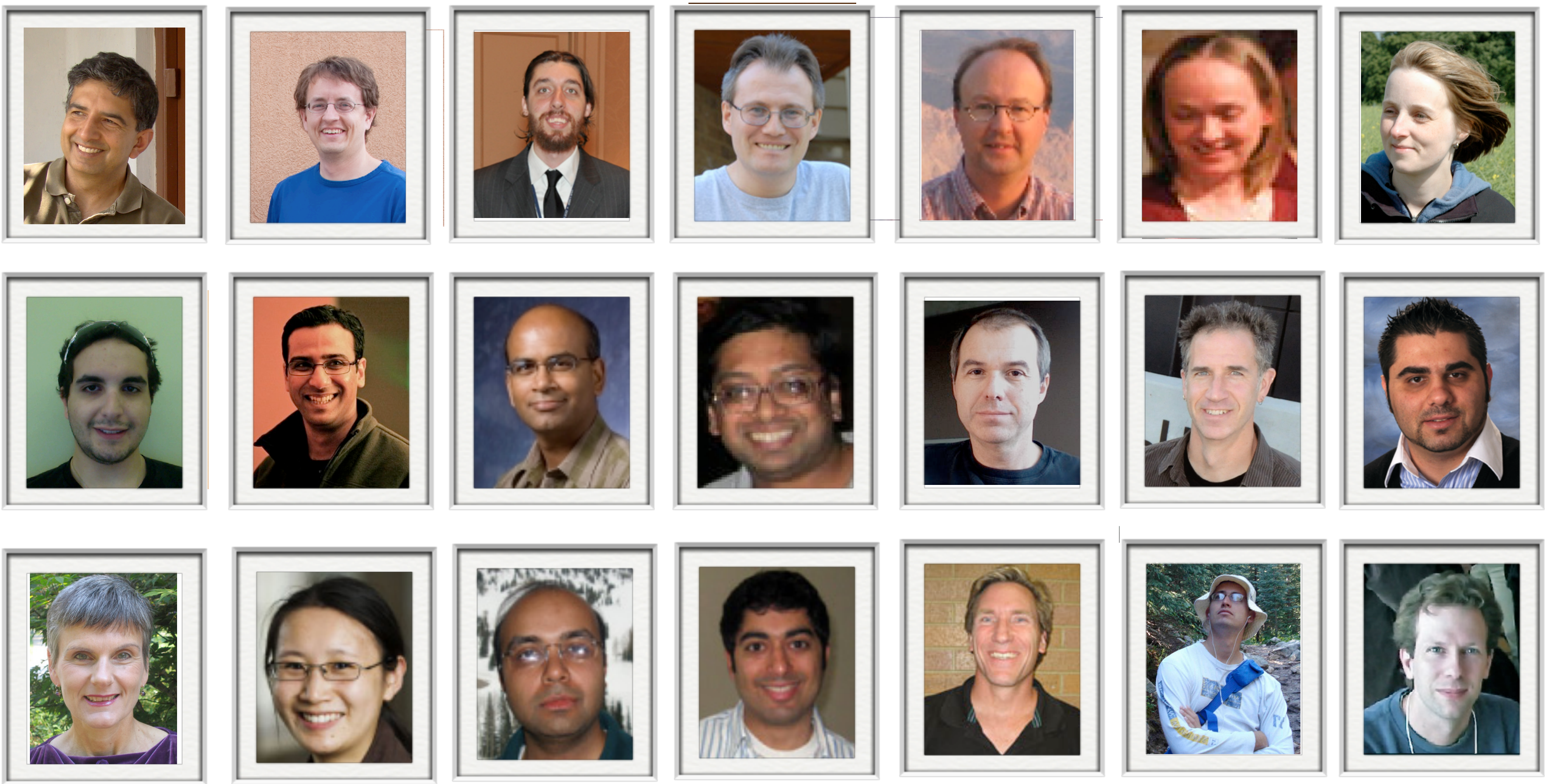
Katrin Heitmann
Argonne National Laboratory





An Exciting Journey from Porting HACC to Mira to First Science Results

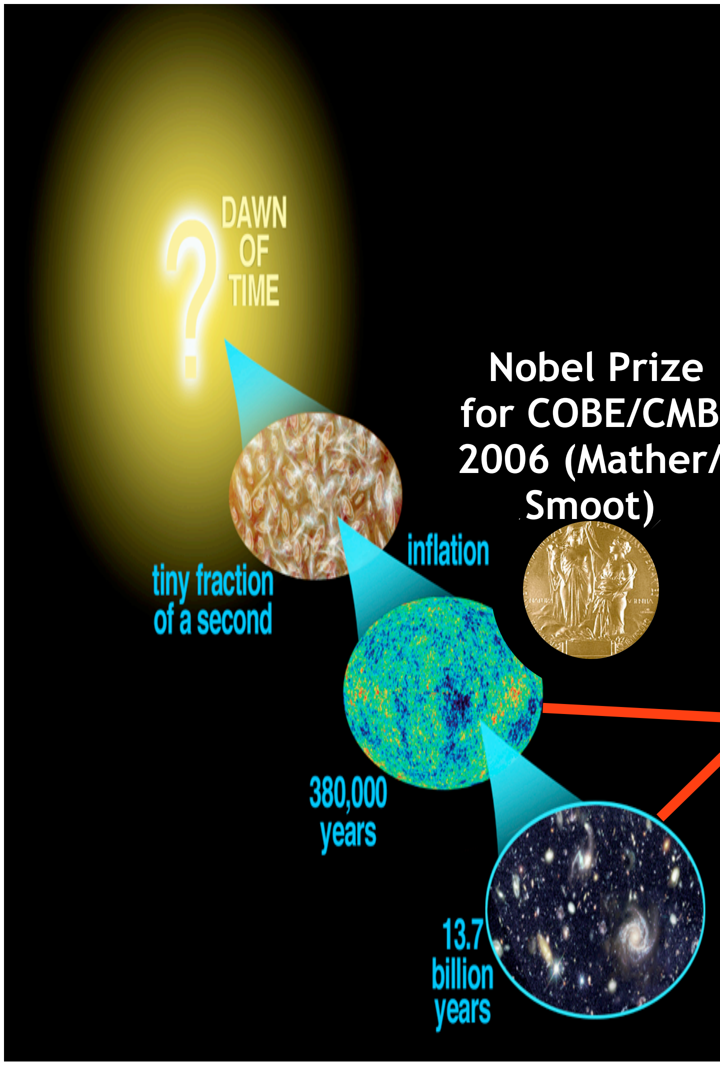
With contributions from many collaborators ...



.... and the ALCF Staff

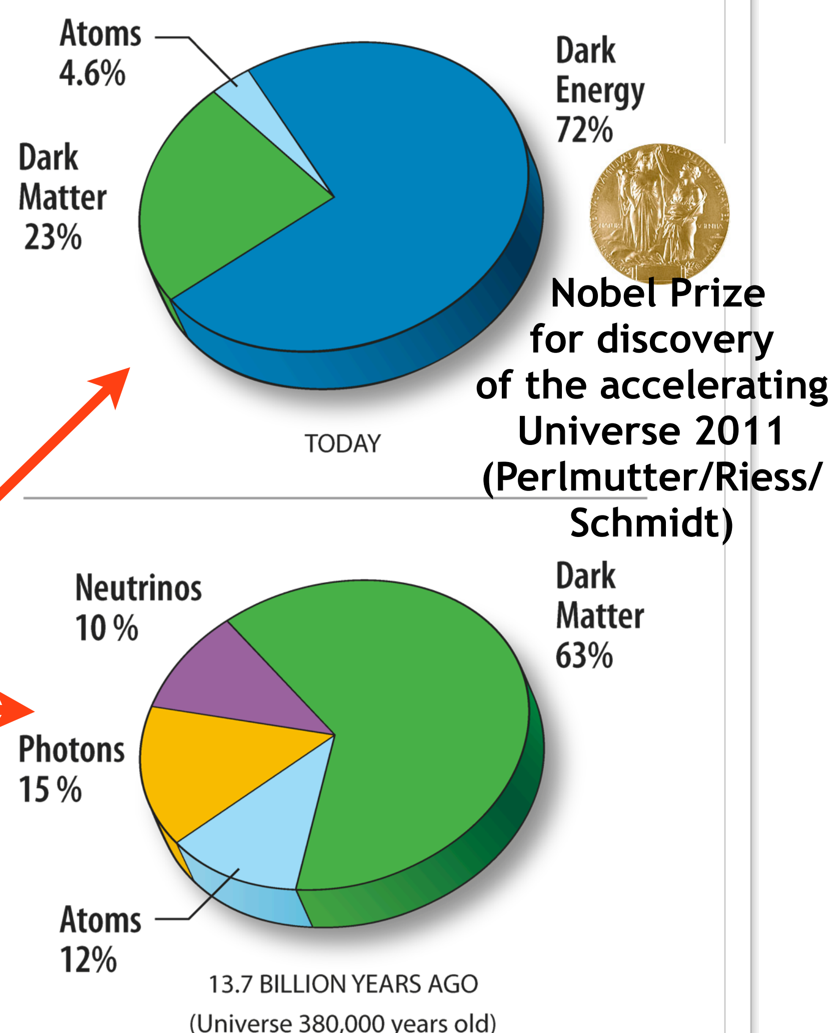
Cosmology

Evolution of the Universe



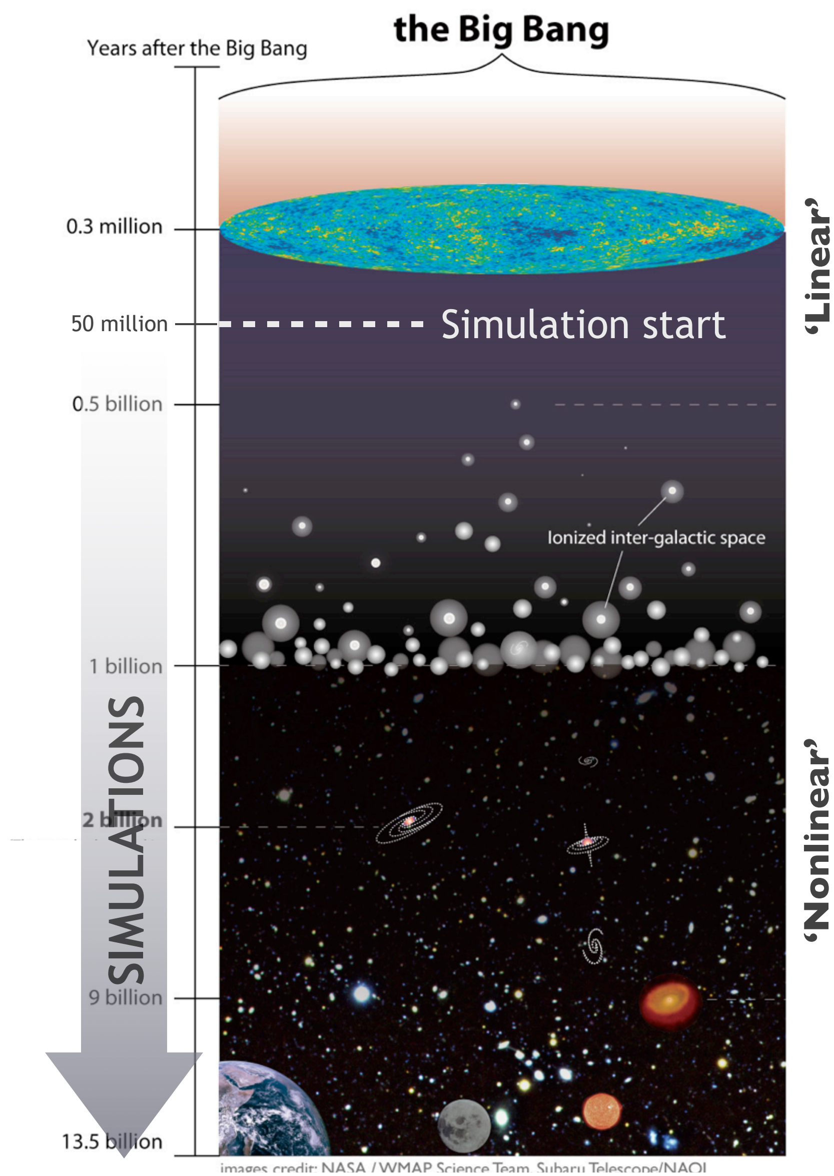
Credit: NASA / WMAP Science Tea

Content of the Universe

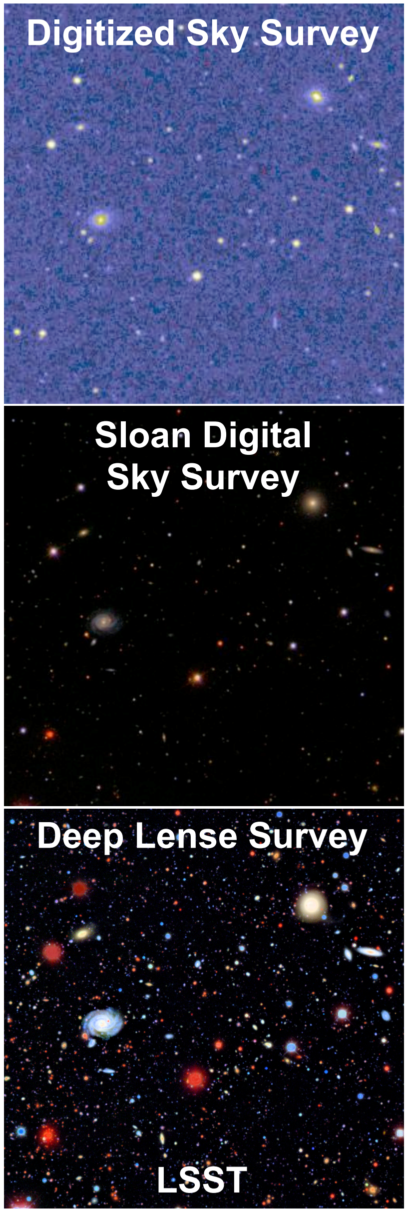


Structure Formation: The Basic Paradigm

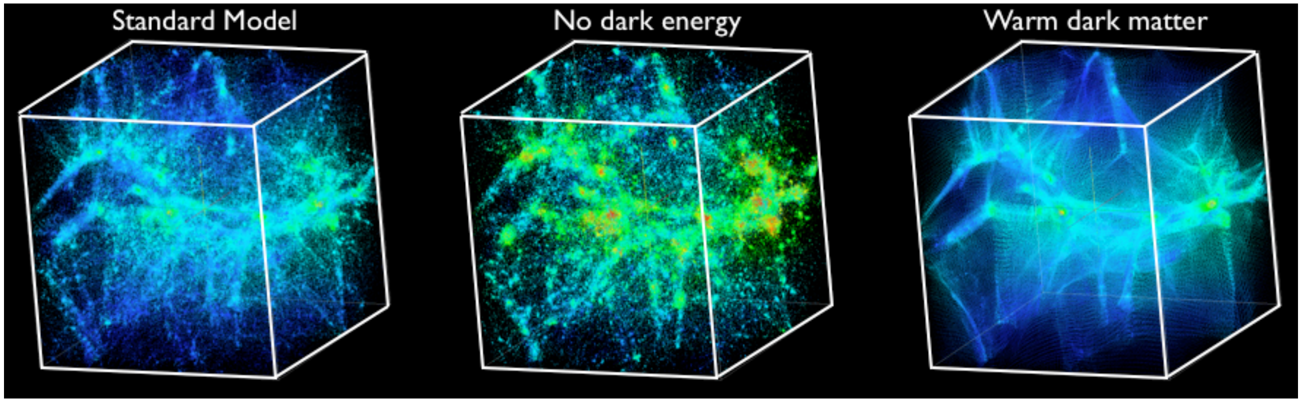
- **Solid understanding of structure formation; success underpins most cosmic discovery**
  - Initial conditions determined by primordial fluctuations
  - Initial perturbations amplified by gravitational instability in a dark matter-dominated Universe
  - Relevant theory is gravity, field theory, and atomic physics (‘first principles’)
- **Early Universe:** Linear perturbation theory very successful (CMB)
- **Latter half of the history of the Universe:** Nonlinear domain of structure formation, impossible to treat without large-scale computing



Precision Cosmology: "Inverting" the 3-D Sky



- **Standard Model of Cosmology:** Verified at the 5-10% level across multiple observations, describes make-up and evolution of the Universe
- **Next generation observatories:** aim to push the current boundaries by orders of magnitude
- Scales that are resolved by future surveys become smaller and smaller, demanding (i) ever larger simulations with increased mass and force resolution; (ii) more detailed physics
- **Next frontier:** Nonlinear regime of structure formation
- **Future Targets:** Aim to control survey measurements to the ~1% level, **can theory and simulation keep up?**

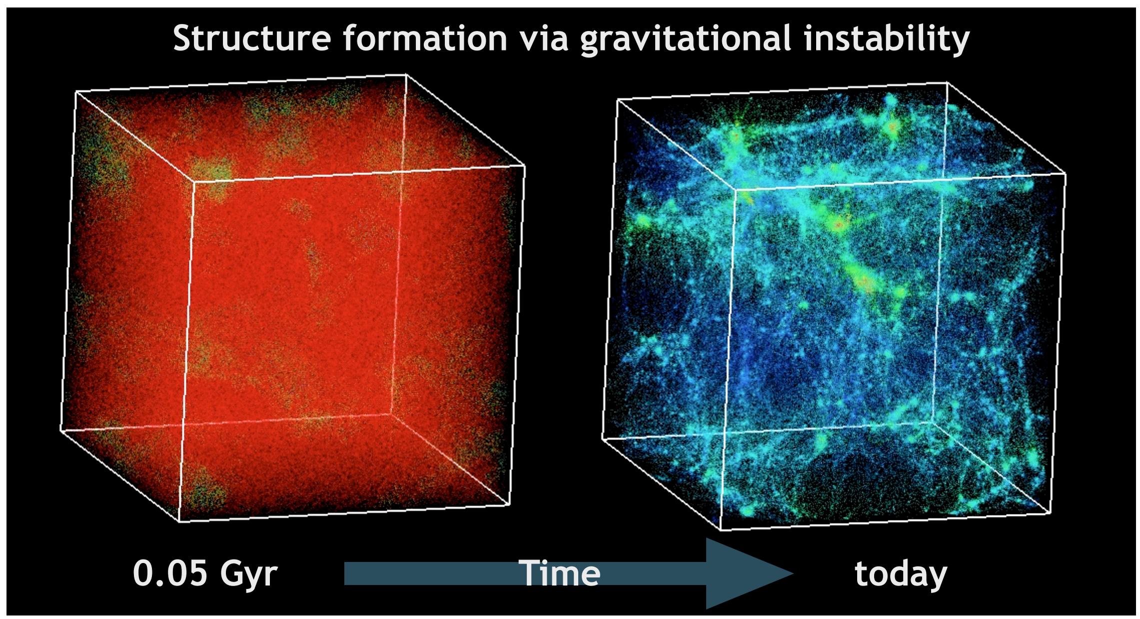


Simulating the Universe

- Gravity dominates at large scales, key task: solve the Vlasov-Poisson equation (VPE)
- VPE is 6-D and cannot be solved as a PDE
- N-body methods; gravity has (i) no shielding but is (ii) naturally Lagrangian
- At smaller scales add gas physics, feedback, etc. (sub-grid modeling inevitable)
- Calibrate simulations against observations

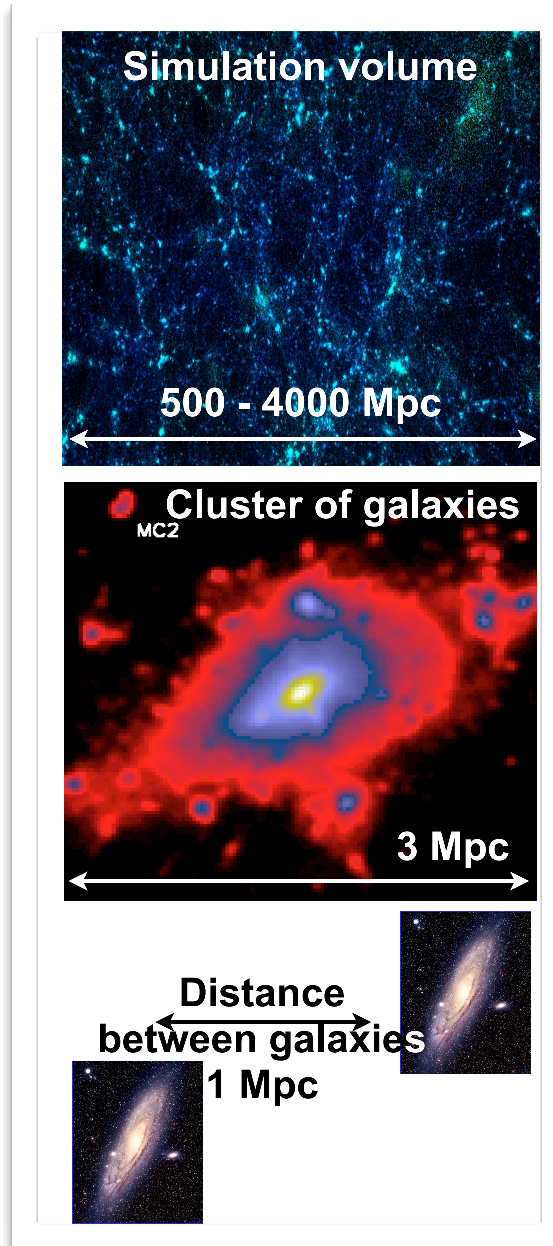
$$
\begin{aligned}
\frac{\partial f_i}{\partial t} + \dot{\mathbf{x}}\frac{\partial f_i}{\partial \mathbf{x}} - \nabla\phi\frac{\partial f_i}{\partial \mathbf{p}} &= 0, \qquad \mathbf{p} = a^2\dot{\mathbf{x}},\\
\nabla^2\phi &= 4\pi G a^2(\rho(\mathbf{x},t) - \langle\rho_{\rm dm}(t)\rangle) = 4\pi G a^2 \Omega_{\rm dm}\delta_{\rm dm}\rho_{\rm cr},\\
\delta_{\rm dm}(\mathbf{x},t) &= (\rho_{\rm dm} - \langle\rho_{\rm dm}\rangle)/\langle\rho_{\rm dm}\rangle),\\
\rho_{\rm dm}(\mathbf{x},t) &= a^{-3}\sum_i m_i \int d^3\mathbf{p}\, f_i(\mathbf{x},\dot{\mathbf{x}},t).
\end{aligned}
$$

Cosmological Vlasov-Poisson Equation

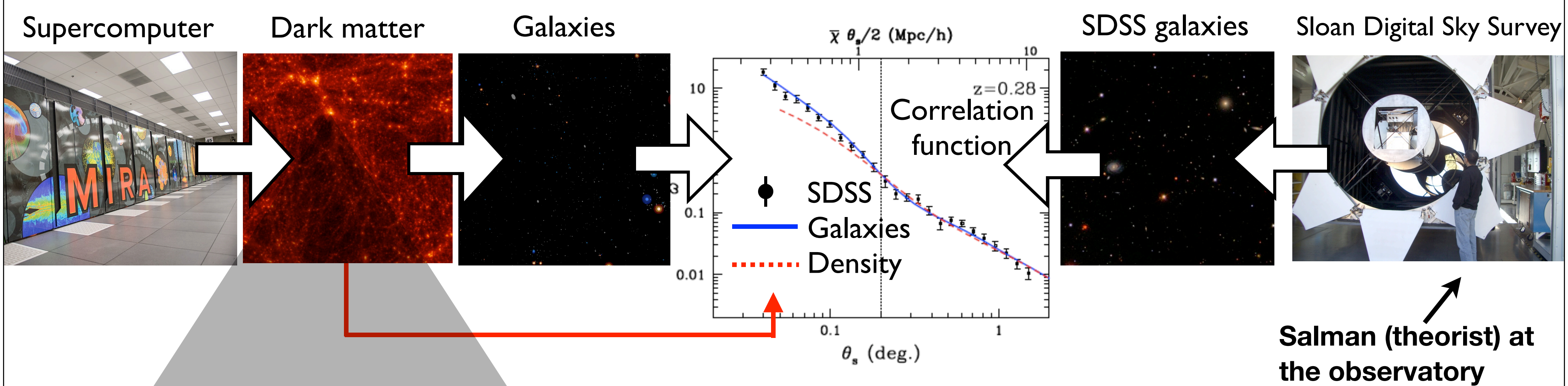


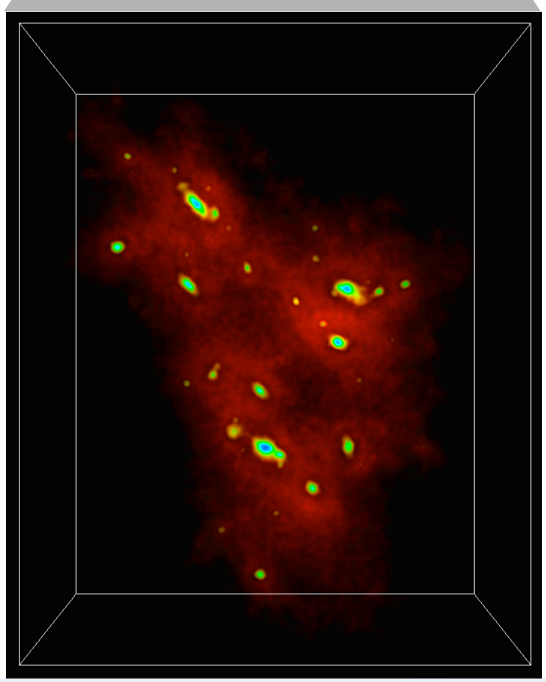
Computing the Universe: Simulating Surveys

- **Simulation Volume:** Large survey sizes impose simulation volumes ~(**3** Gpc)^3, memory required ~**100**TB **- 1**PB
- **Number of Particles:** Mass resolution depends on ultimate object to be resolved, ~**10**^{8}**-10**^{10} solar masses, N~**10**^{11}**-10**^{12} **"Outer Rim Simulation"**
- **Force Resolution:** ~kpc, yields a (global) spatial dynamic range of **10**^{6}
- **Throughput:** Large numbers of simulations required (**100**'s **--1000**'s), development of analysis suites, and emulators; peta-exascale computing exploits, **"Mira Universe"**
- **Computationally very challenging!** HACC is aimed to meet these requirements



Connecting the Simulations to Observations

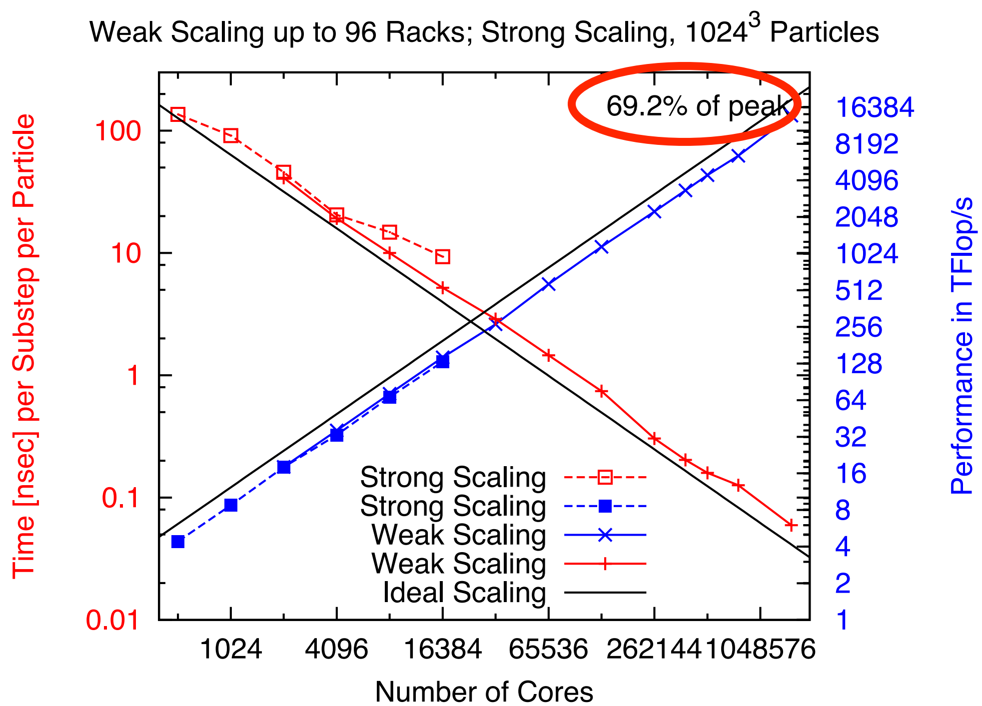




- The Halo Concept: find “dark matter clumps” via neighbor finding or overdensity algorithms
- Abundance of halos and correlation function both very sensitive to cosmology
- Properties of halos, e.g., density profiles, also cosmology dependent
- Assign galaxies to halos depending on their mass and/or formation history to create synthetic skies

HACC Performance on the BG/Q

- HACC: Hardware/Hybrid Accelerated Cosmology Code
- Highly portable, runs on diverse architectures
- On BG/Q: TreeParticleMesh code
- Achieved 13.94PFlops on Sequoia, 90% parallel efficiency on 1,572,864 cores
- 3.6 trillion particle benchmark run



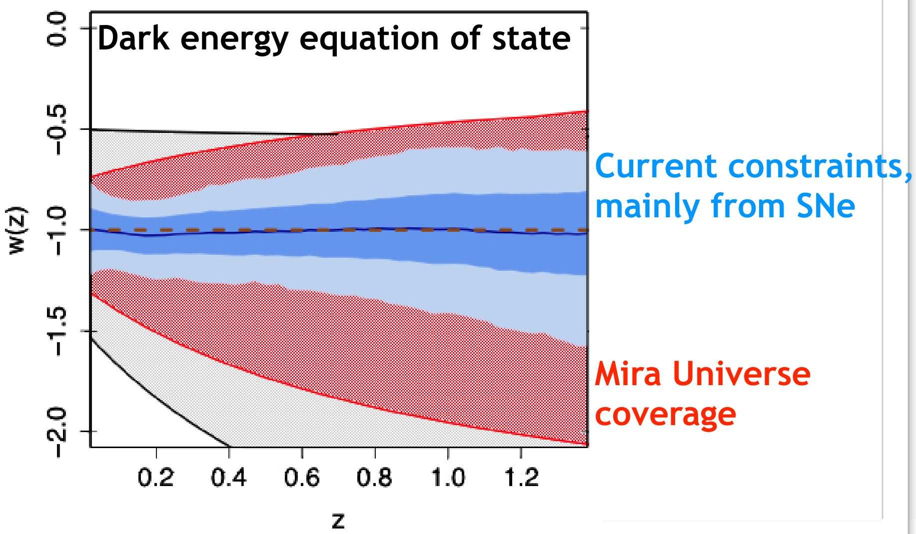
Habib et al. SC12, Gordon Bell Finalist

The HACC Science Program on Mira

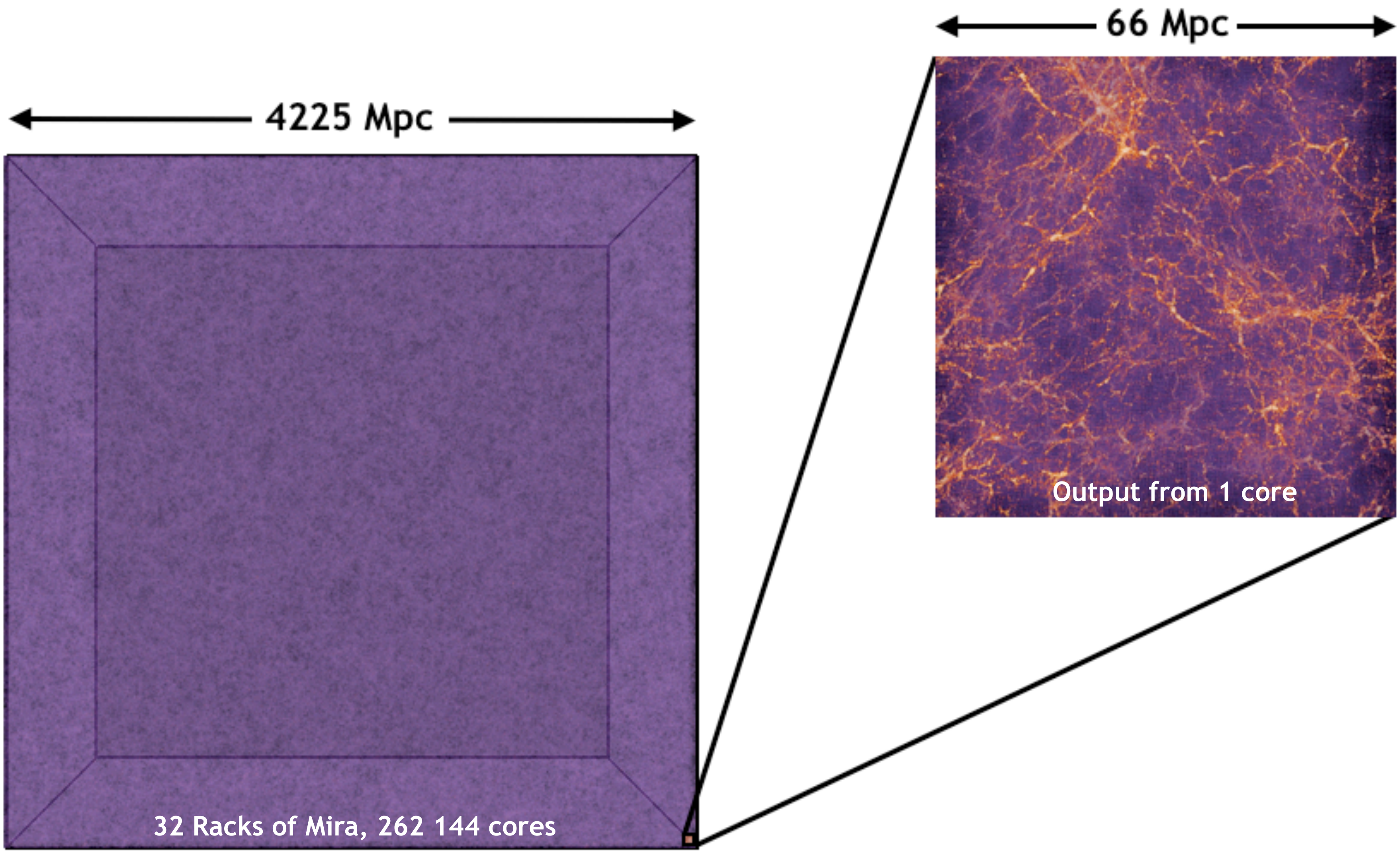
- **The Outer Rim Simulation**
  - Largest high-resolution cosmological simulation ever carried out
  - 1.1 trillion particles, 4.225Gpc volume, ~7kpc force resolution
  - Unprecedented statistics and details leading to new scientific insights
- **The Mira Universe**
  - Large suite of different cosmological models, 2.1Gpc volume, 30 billion particles each
  - Models include dynamical dark energy, neutrino effects
  - Will lead to new prediction tools to be used by ongoing and future surveys

The Outer Rim -- where thoughts, time, and space become one

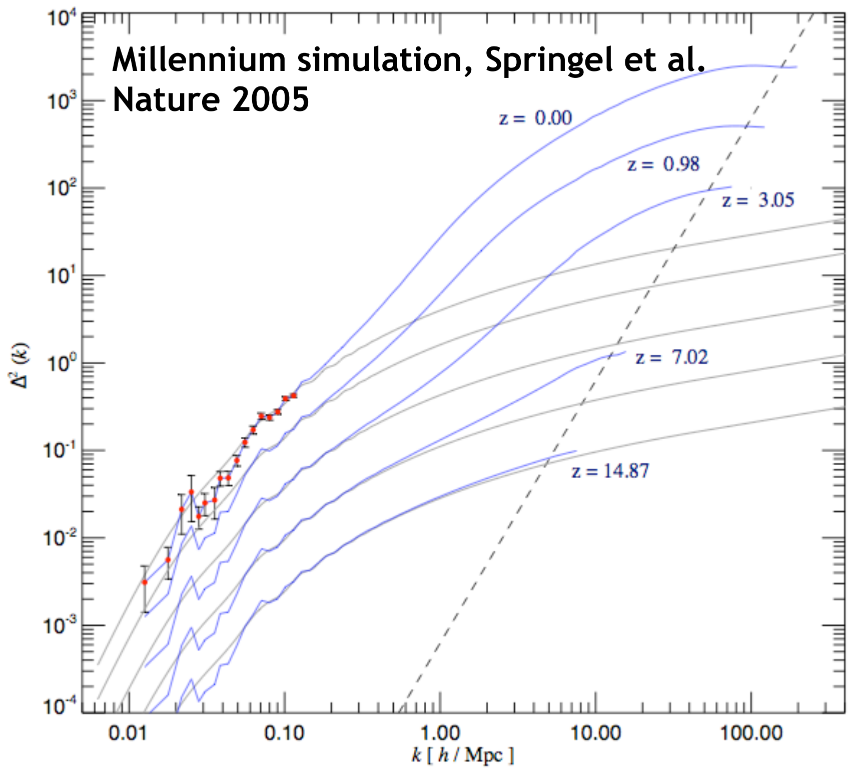
The Mira Universe

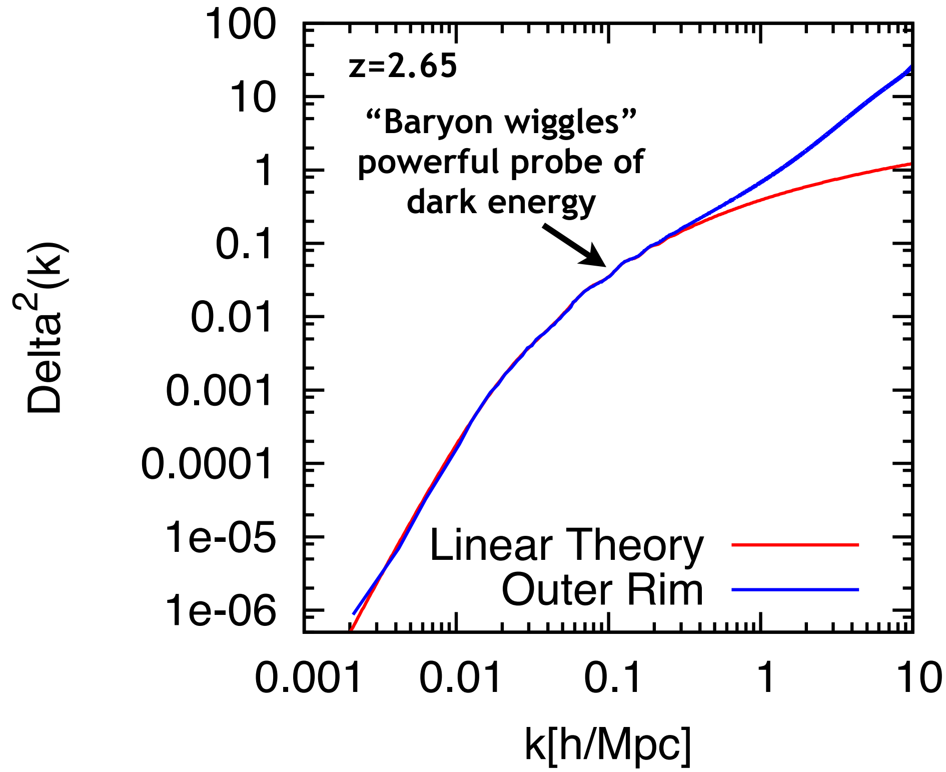


The Outer Rim Simulation



Outer Rim Run -- First Science Results



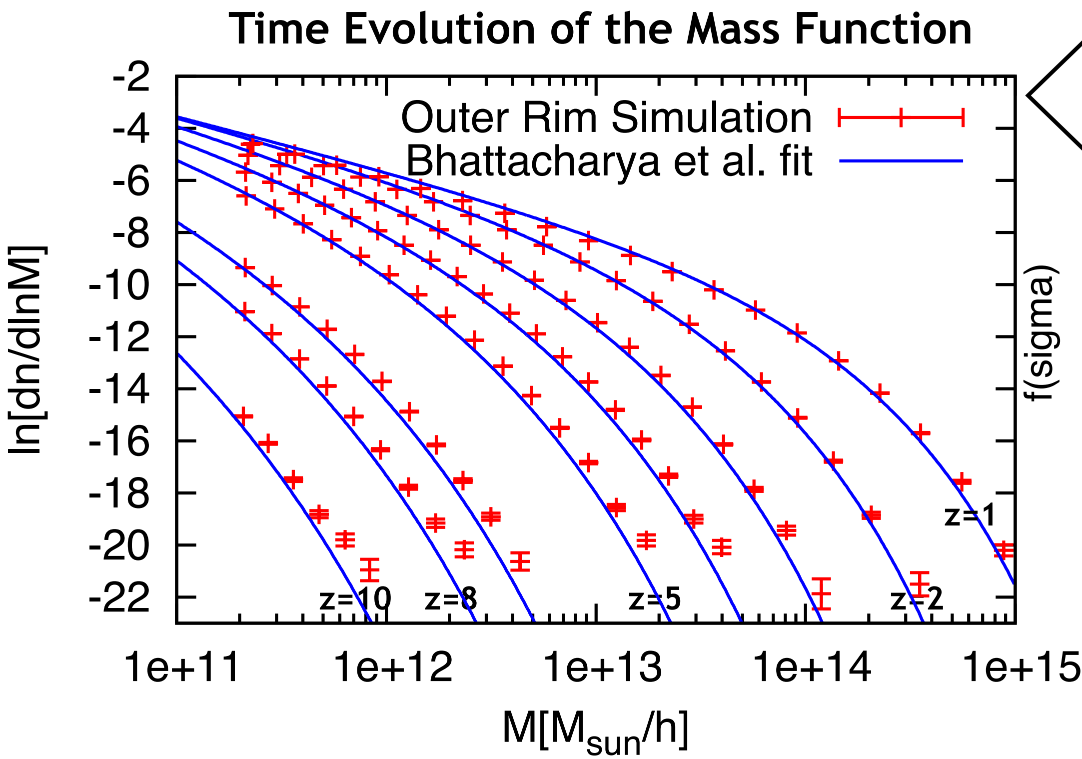


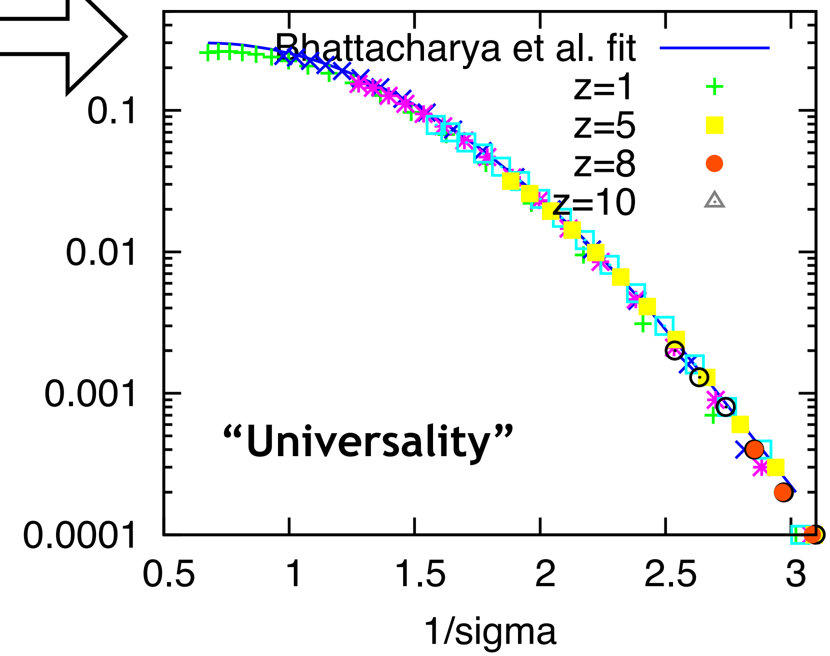
2-point correlation function:

$$\xi(\vec{x}) = \int \frac{d^3\vec{y}}{V}\delta(\vec{y}-\vec{x})\delta(\vec{y}) = \int \frac{d^3\vec{k}}{(2\pi)^3 V}|\delta_k|^2 e^{i\vec{k}\cdot\vec{x}}$$

power spectrum

Outer Rim Run -- First Science Results





$$f(\sigma, z) = \frac{M}{\rho_b(z)} \frac{dn(M,z)}{d\ln[\sigma^{-1}(M,z)]}$$

Halo number density

Background density of the Universe

$$\sigma^2(M,z) = \frac{D^2(z)}{2\pi^2} \int_0^\infty k^2 P(k) W^2(kR(M))\, dk$$

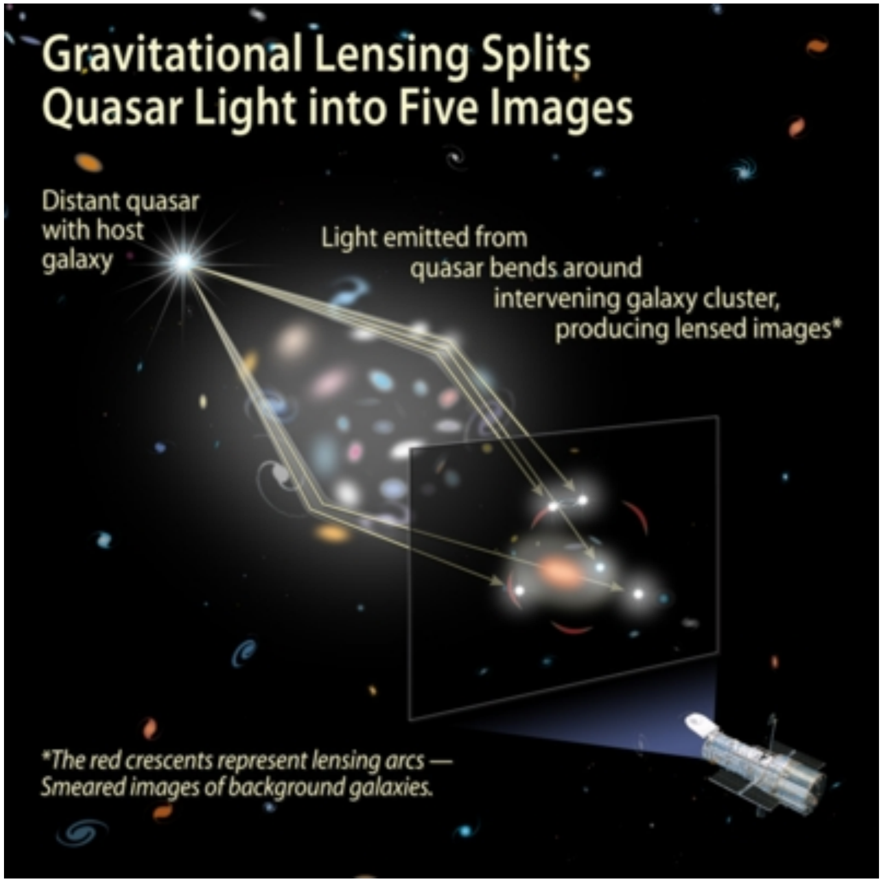
Growth function

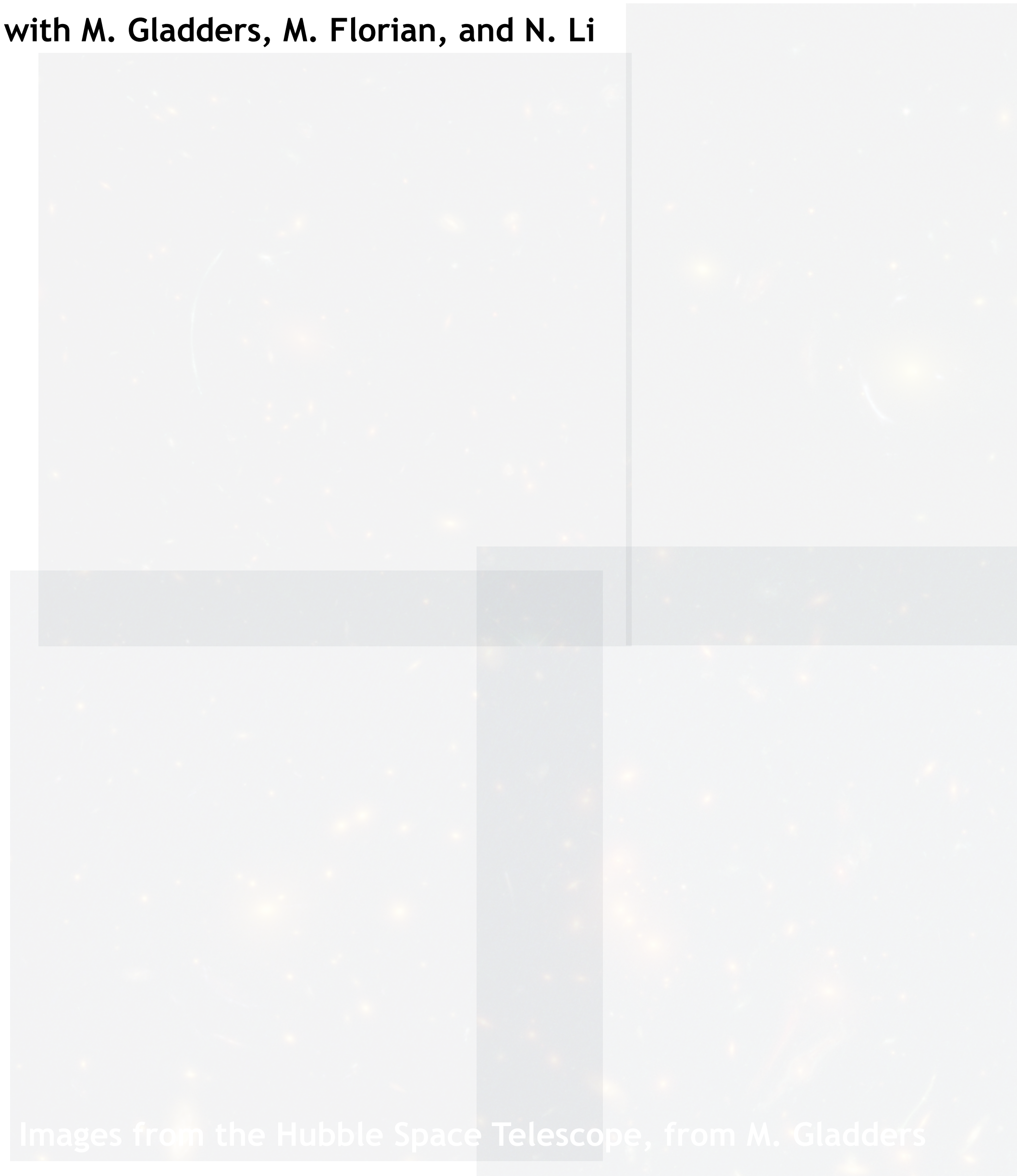
Linear matter power spectrum

Top hat filter

Outer Rim Run -- Strong Lensing

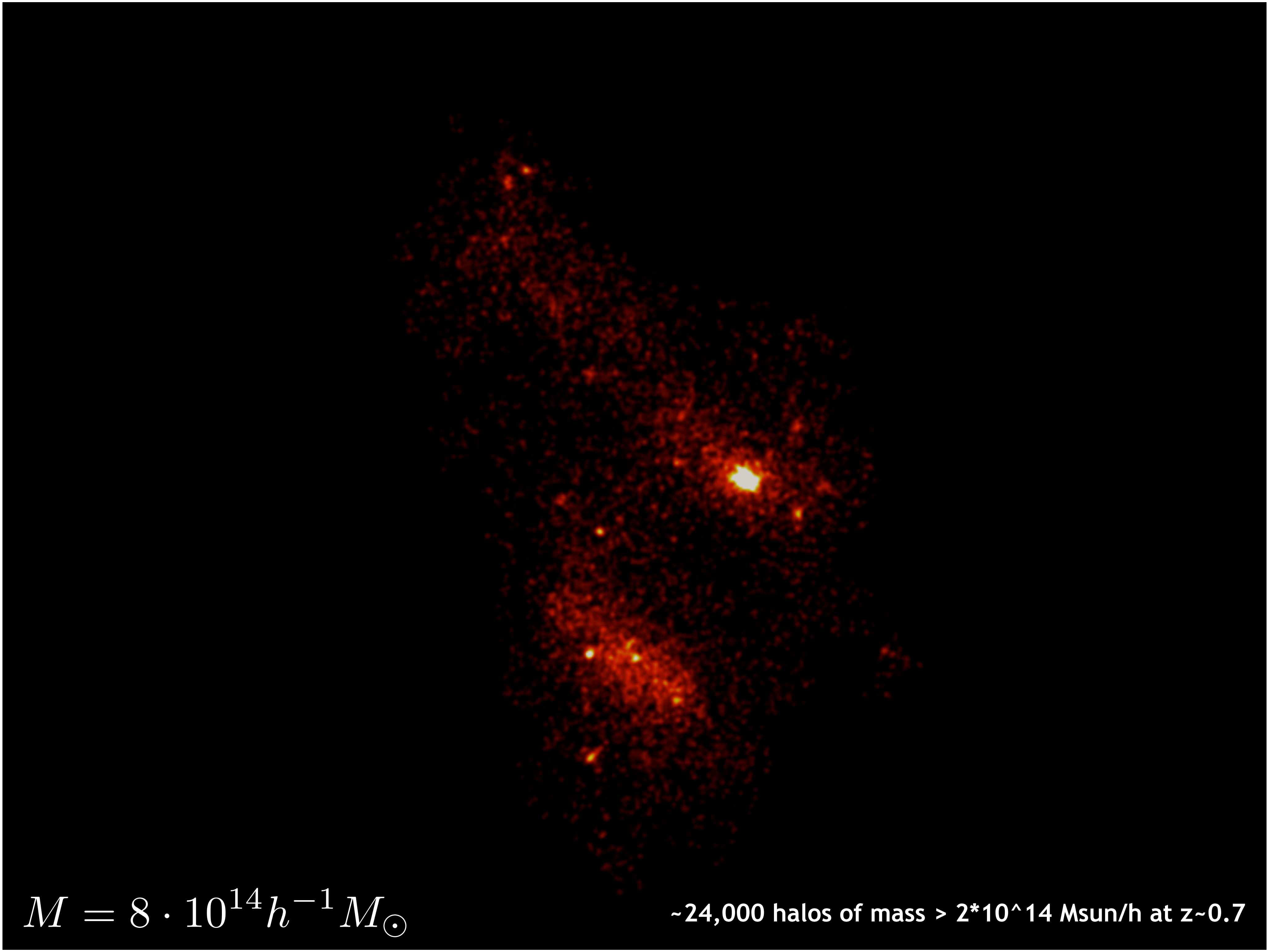
A joint UChicago/ANL project in collaboration with M. Gladders, M. Florian, and N. Li

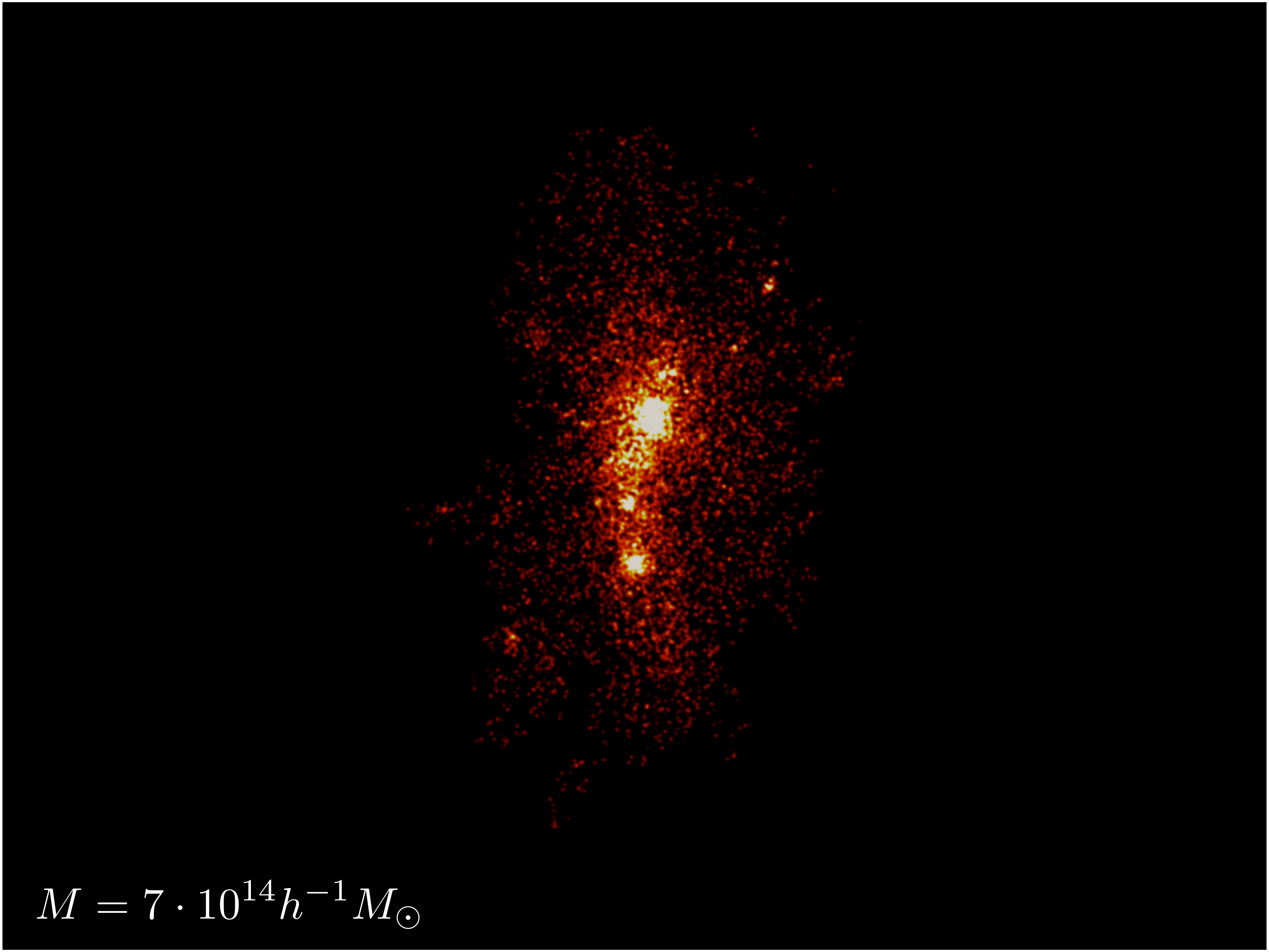


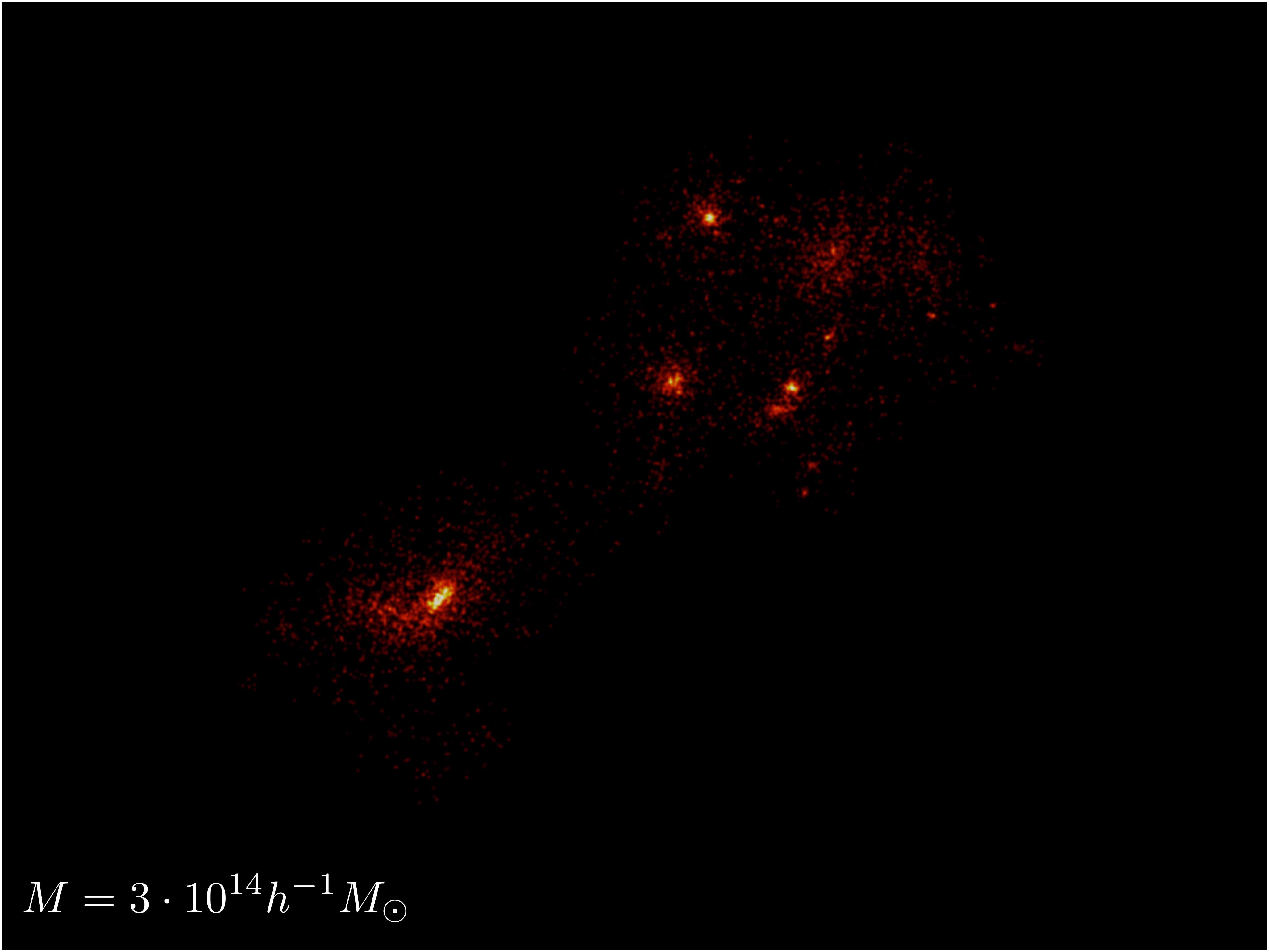


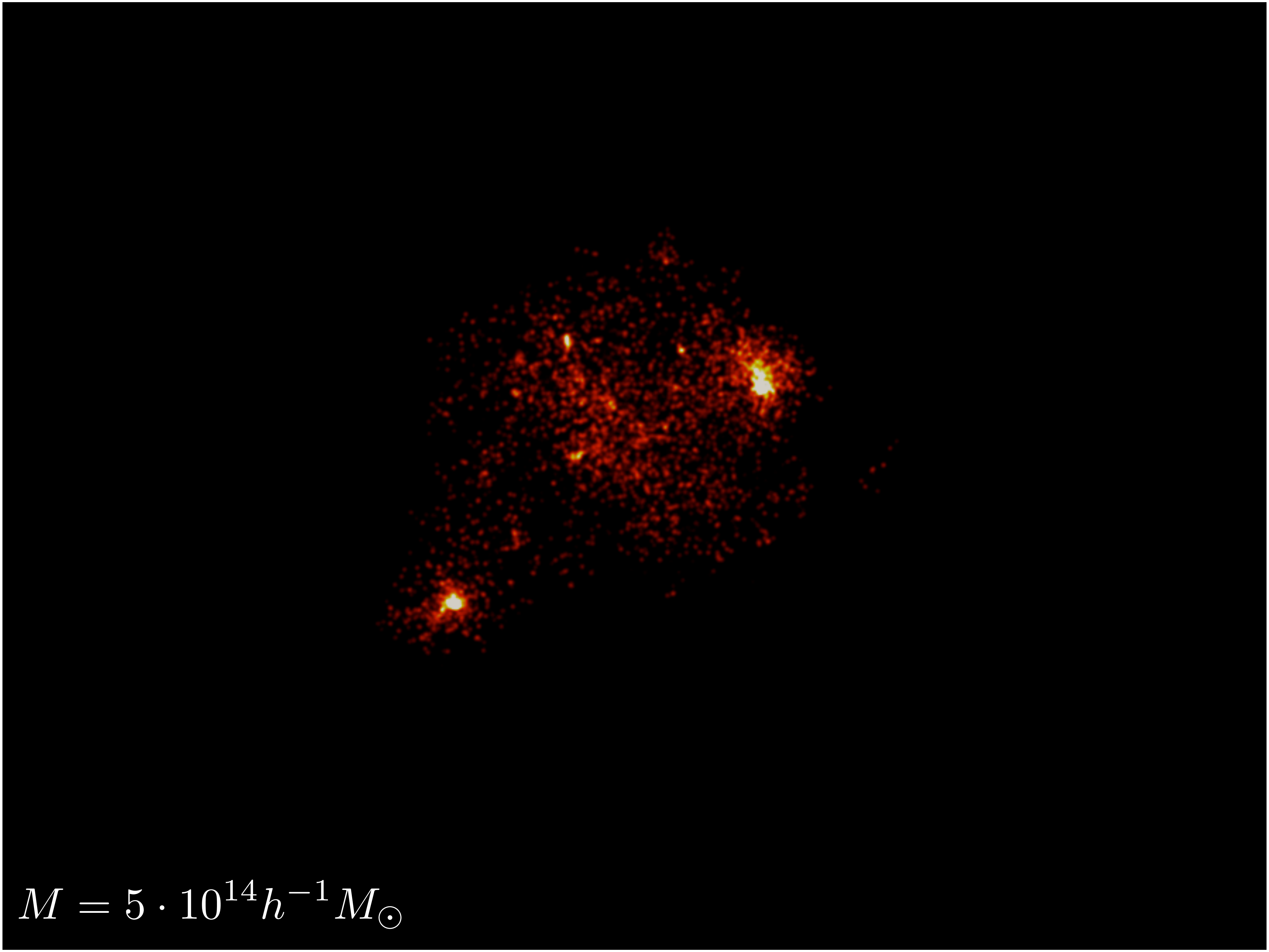
Images from the Hubble Space Telescope, from M. Gladders

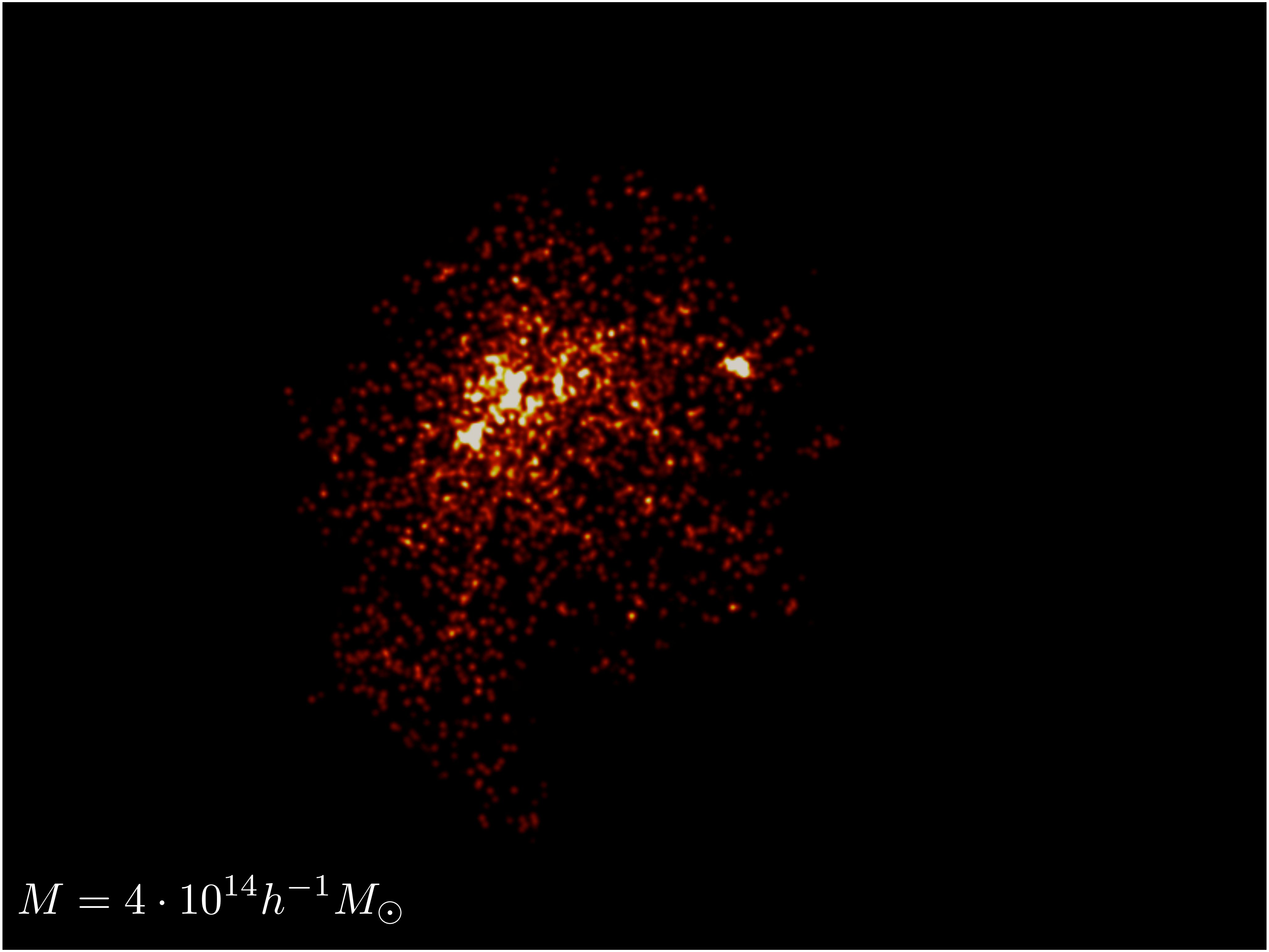
- Outer Rim simulation will allow a full modeling effort of strong lenses with high statistics and realistic source distributions!


$M = 8 \cdot 10^{14} h^{-1} M_{\odot}$
~24,000 halos of mass > 2*10^14 Msun/h at z~0.7


$M = 7 \cdot 10^{14} h^{-1} M_{\odot}$

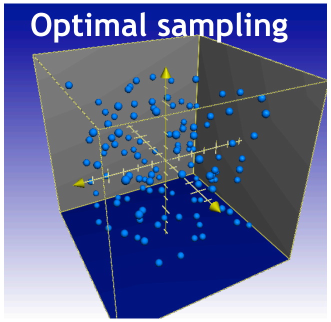

$M = 3 \cdot 10^{14} h^{-1} M_{\odot}$

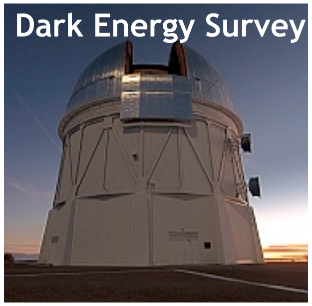

$M = 5 \cdot 10^{14} h^{-1} M_{\odot}$

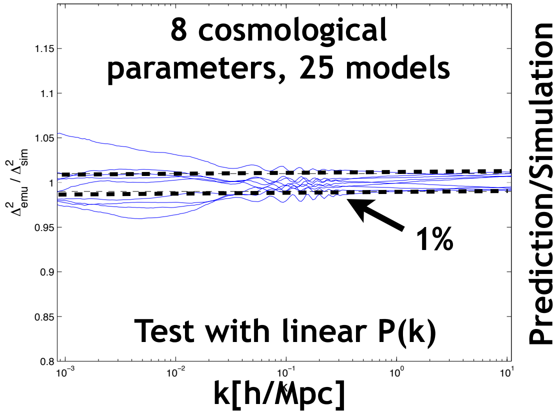

$M = 4 \cdot 10^{14} h^{-1} M_{\odot}$

The Mira Universe Design

- **Challenge:** To extract cosmological constraints from observations in non-linear regime, need to run Marko Chain Monte Carlo code; input: 10,000 - 100,000 different models
- **Current strategy:** Fitting functions , accurate at 10% level, not good enough!
- **Brute force:** Simulations, ~30 years on 2000 processor cluster...
- **Only alternative:** emulators -- from a smaller number of simulations build very accurate prediction schemes
- **"Ingredients":** Optimal sampling methods to decide which models to simulate, efficient representation of simulation outcome, powerful interpolation scheme
- **For Mira Universe:** Nested design -- obtain first set of emulators after 25 models, improvement of accuracy with 50 models, final result with 100 models, augment parameter space with 20 more models



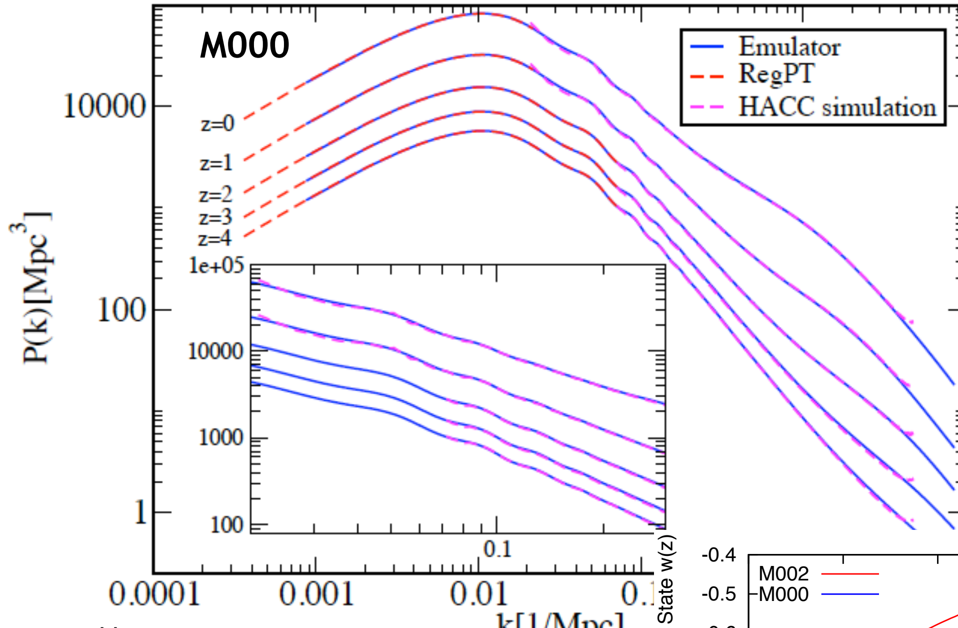




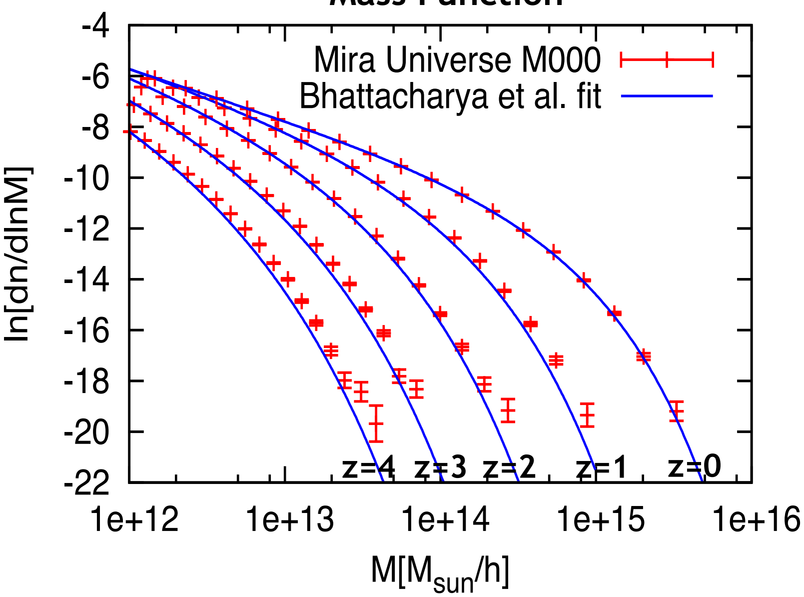
Heitmann et al. 2006, Habib et al. 2007, Lawrence et al. 2010, Kwan et al. 2012, Heitmann et al. 2013

Some results --

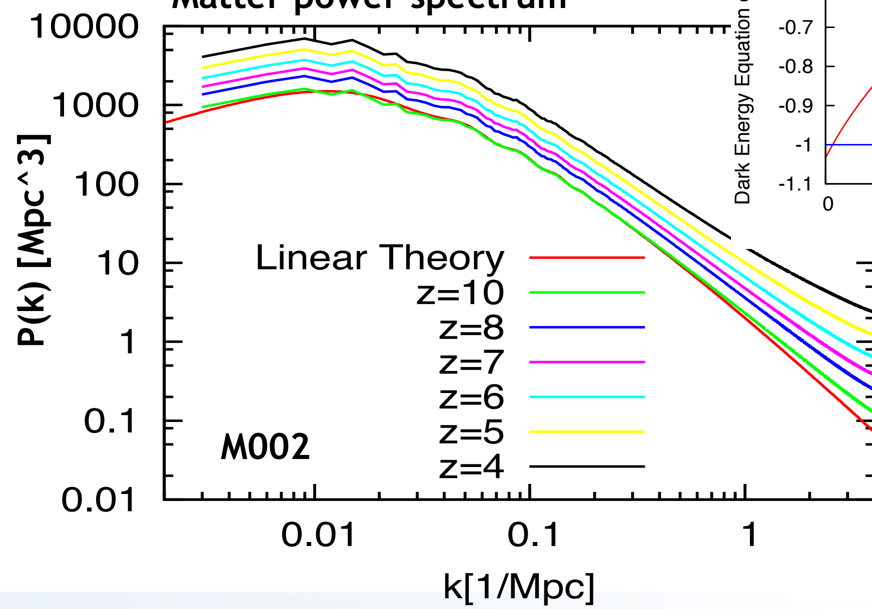
- Simulation of standard cosmological model finished, first set of analysis carried out
- Verified results, finalized in-situ analysis tools
- Second set of analysis ongoing, production of synthetic skies
- Next set of models running

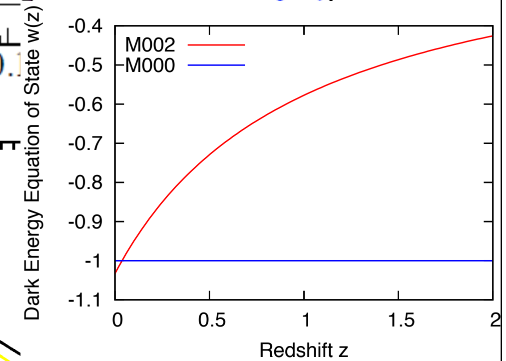


Mass Function



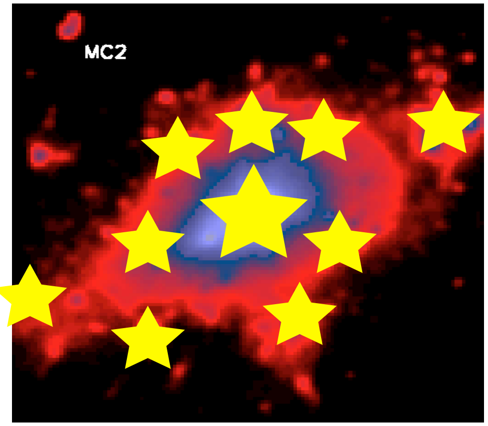
Matter power spectrum





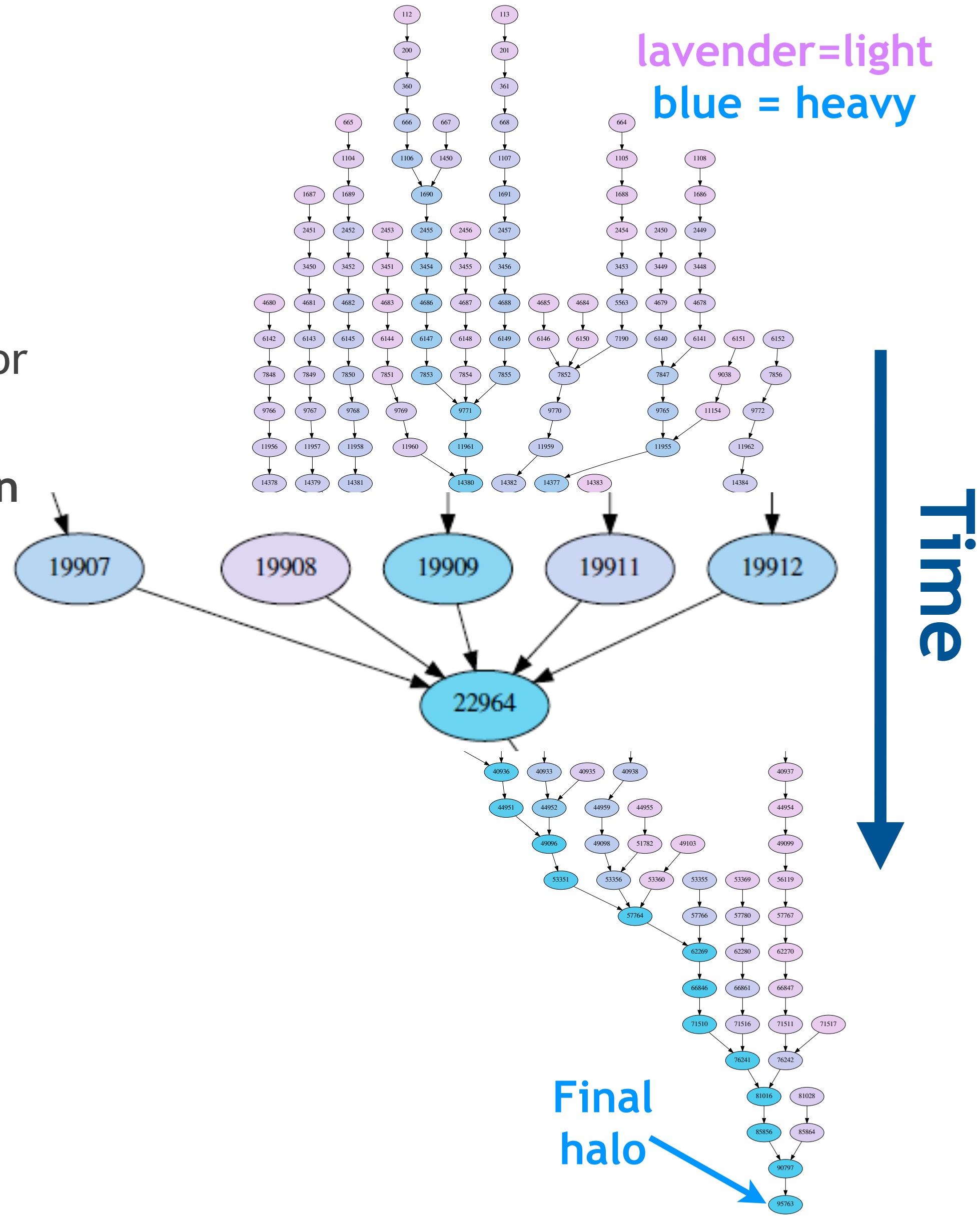
The Halo Occupation Distribution (HOD) -- Where are the Galaxies?

- Idea: find halos in the simulation, depending on their mass, populate them with galaxies
- Above mass threshold: each halo hosts central galaxy; the heavier the halo, the more satellites
- 5 parameter model, parameters adjusted by comparison to observations, different classes of galaxies = different HODs
- Mira Universe: interplay of cosmology, HOD



Semi-Analytic Modeling

- **HOD has many advantages:**
  - Easy implementation, easily tunable to observations, excellent for analysis of observations, covariances etc.
  - But: no physics insights, no predictive power, does not take into account dynamics, no predictions at early times for which we don’t have observations
- **Semi-analytic modeling of galaxy formation closer to “first principle approach”**
  - Track halos through evolution and build “merger trees”
  - On top of merger trees, apply physical principles to determine properties of galaxies, such as mass of stars and gas in each galaxies, luminosity, black hole content,chemical composition etc.
  - We use Galacticus developed by Andrew Benson
  - First test synthetic skies developed



Summary and Outlook

- **HACC is first and currently only cosmology code that runs at scale on BG/Q systems, achieved 69.2% of peak, almost 14 PFlops on Sequoia**
- **Nevertheless, code still undergoing improvement with respect to algorithmic implementations of the short range solver, time stepper etc.**
- **In addition, continue work on in-situ analysis tools, refine models for generating synthetic skies, analyze simulations**
- **Largest ever high-resolution run, *The Outer Rim Simulation*, currently running, analysis ongoing**
  - Many hurdles had to be overcome to make this happen, including memory management, I/O, and analysis tools
- ***The Mira Universe* will lead to an unprecedented set of simulations, spanning 8 cosmological parameters, including different dark energy models and neutrinos**
  - Runs have been started after carefully mapping out model space
  - Analysis ongoing
- **Sharing data with community: large effort making data publicly available, including analysis tools and workflow, "PDACS"**