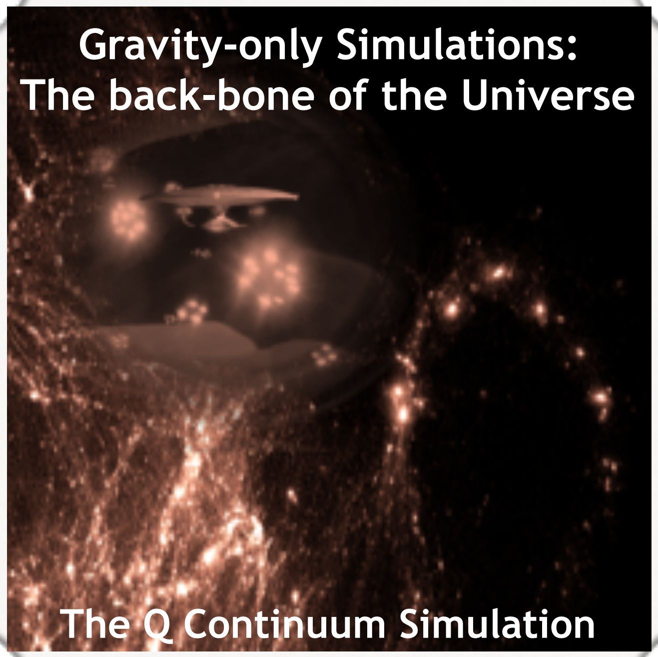
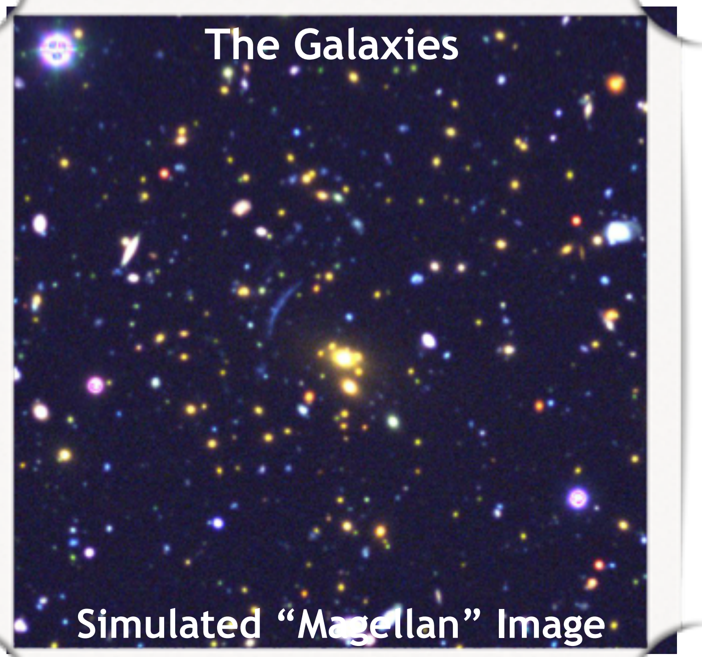
DESchool: How to Simulate the Universe

Katrin Heitmann

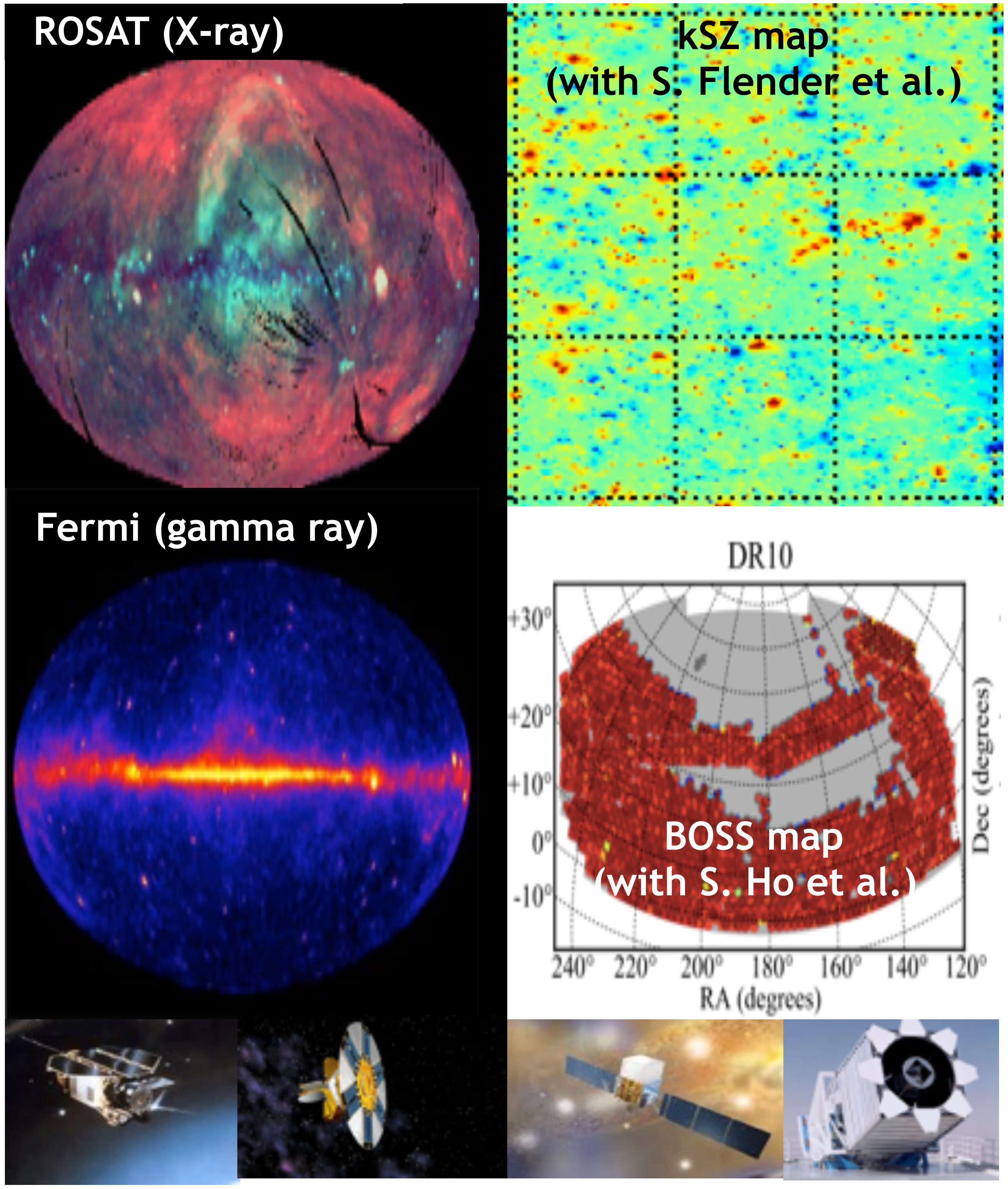
DE School, October 26, 2015





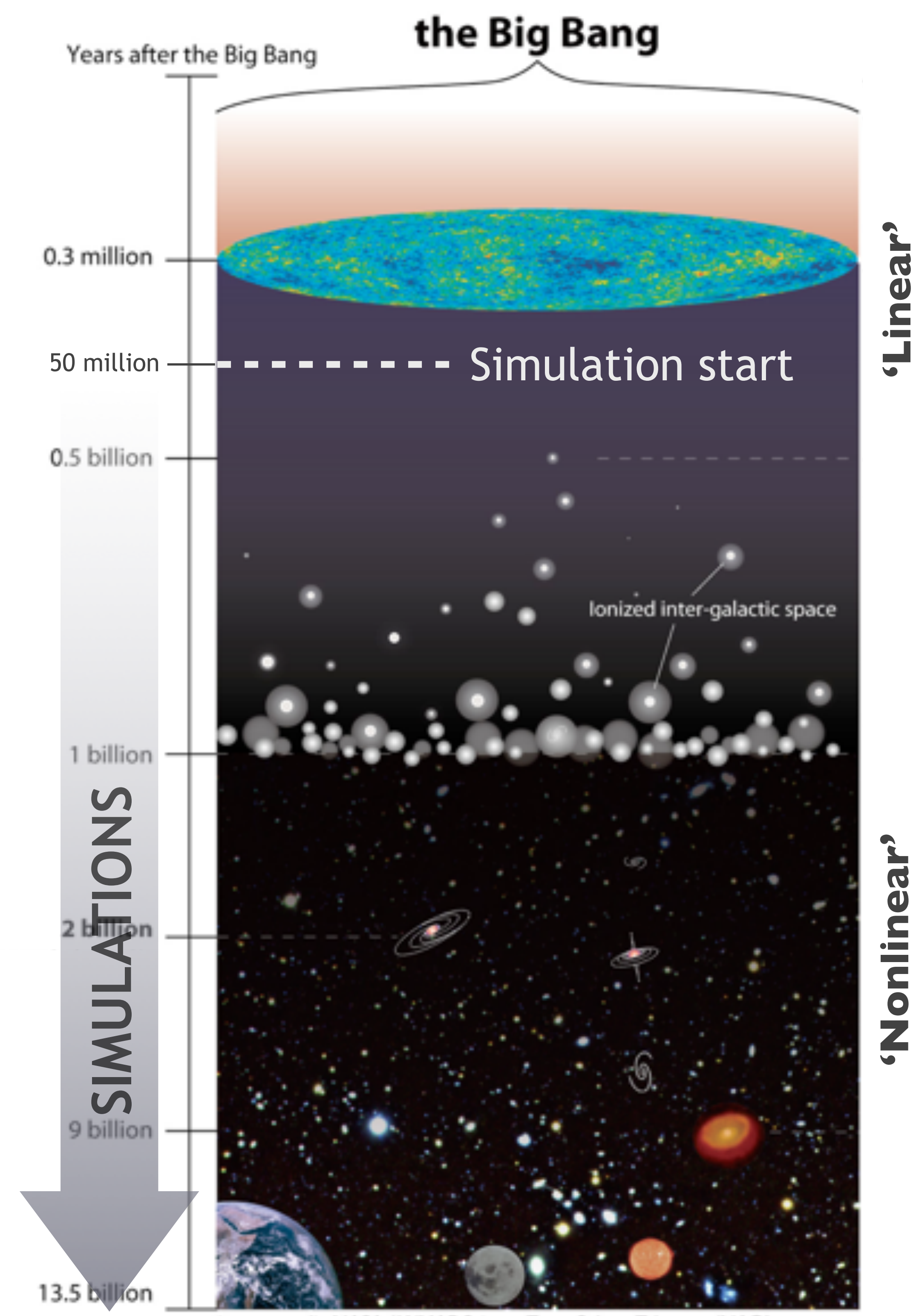
Modern Cosmology and Sky Maps

- Modern cosmology is the story of mapping the sky in multiple wavebands
- Maps cover measurements of objects (stars, galaxies) and fields (temperature)
- Maps can be large (Sloan Digital Sky Survey has~200 million galaxies, many billions for planned surveys)
- Statistical analysis of sky maps
- All precision cosmological analyses constitute a statistical inverse problem: **from sky maps to scientific inference**
- Therefore: *No* cosmology without (large-scale) computing

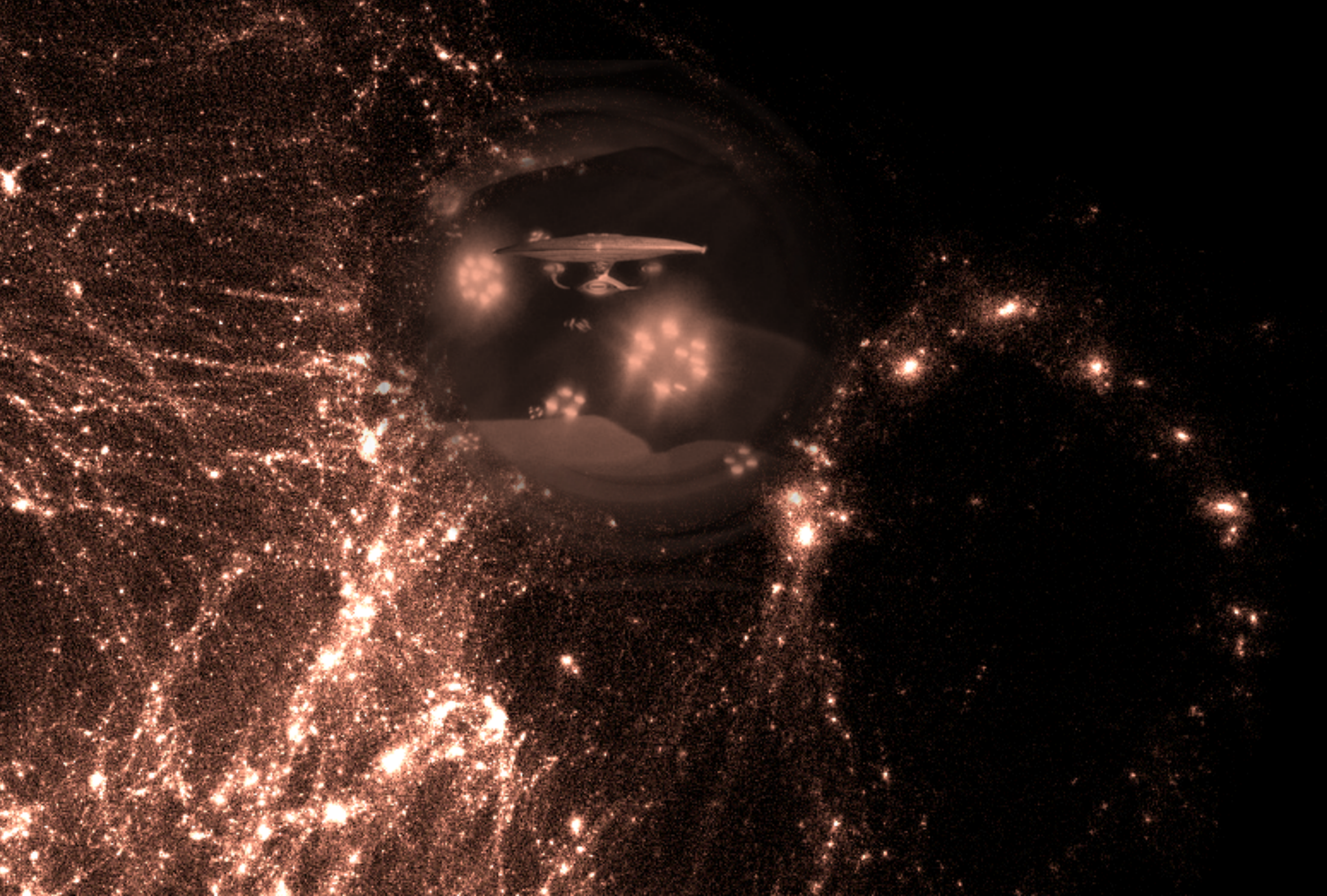


The Evolution of the Universe: Structure Formation

- **Solid understanding of structure formation; success underpins most cosmic discovery**
  - Initial conditions determined by primordial fluctuations, measured from the cosmic microwave background
  - 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



Gravity-only Simulations, the Backbone of the Universe



Computing the Universe

- Gravity dominates at large scales, key task: solve the Vlasov-Poisson equation (VPE)
- VPE is 6-D and cannot be solved as PDE, therefore N-body methods
- Particles are tracers of the dark matter in the Universe, mass typically at least ~10^9 $M_\odot$

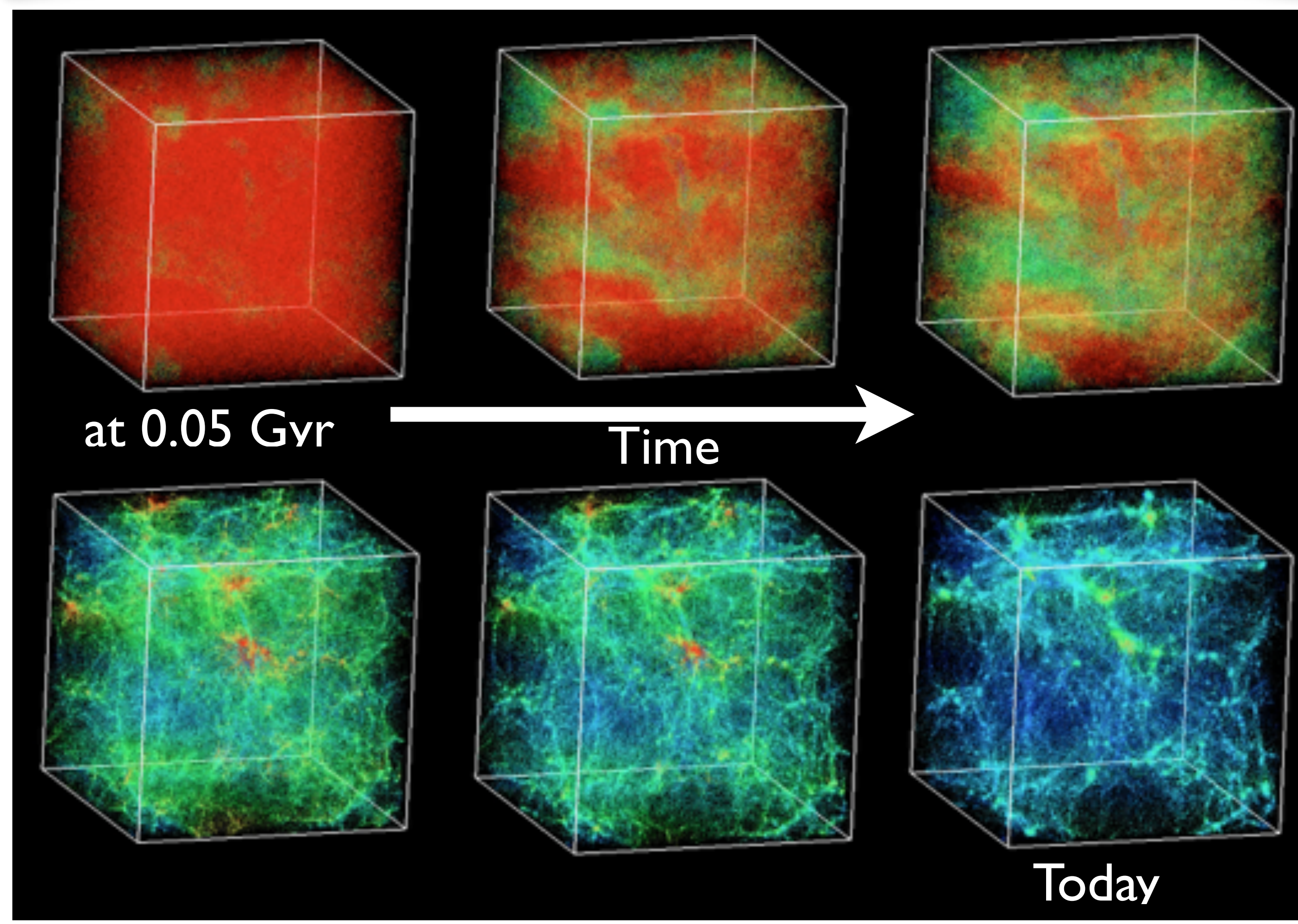
$$m_p \sim V/n_p$$

- At smaller scales, add gas physics, feedback etc., sub-grid modeling inevitable

> **"The Universe is far too complicated a structure to be studied deductively, starting from initial conditions and solving the equations of motion."**
>
> *Robert Dicke (Jayne Lectures, 1969)*

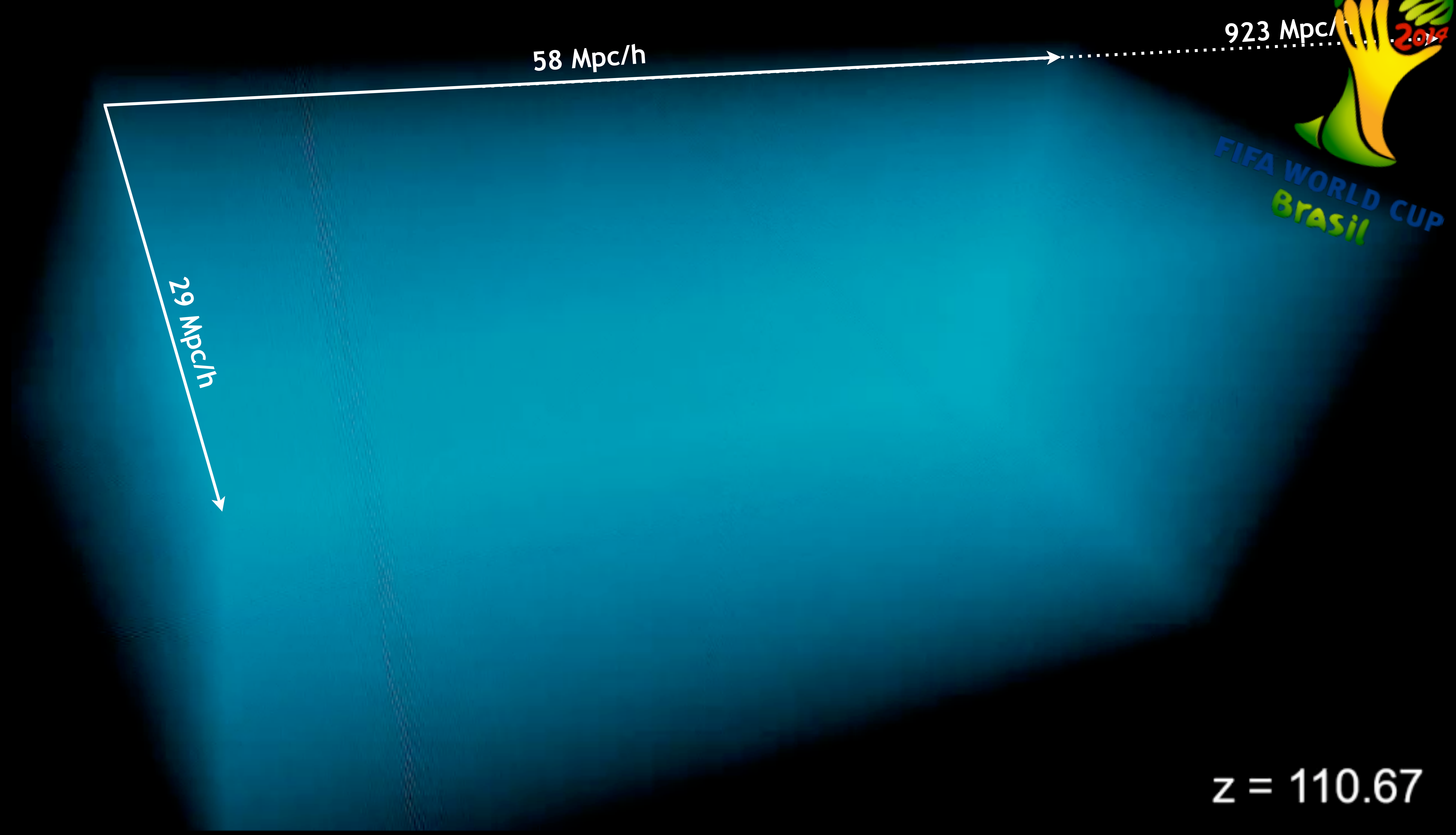
$$
\begin{aligned}
\frac{\partial f_i}{\partial t} + \dot{\mathbf{x}}\frac{\partial f_i}{\partial \mathbf{x}} - \nabla\phi\frac{\partial f_i}{\partial \mathbf{p}} = 0, &\qquad \mathbf{p} = a^2\dot{\mathbf{x}}, \\
\nabla^2\phi &= 4\pi G a^2(\rho(\mathbf{x},t) - \langle\rho_{\mathrm{dm}}(t)\rangle) = 4\pi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

Cosmological Vlasov-Poisson Equation



The Q Continuum* Simulation

Finished Sunday, July 13, 2014 on ~90% of Titan under INCITE, evolving more than half a trillion particles. Shown is the output from one node (~33 million particles), 1/16384 of the full simulation, generated 2.5PB of data, still being analyzed

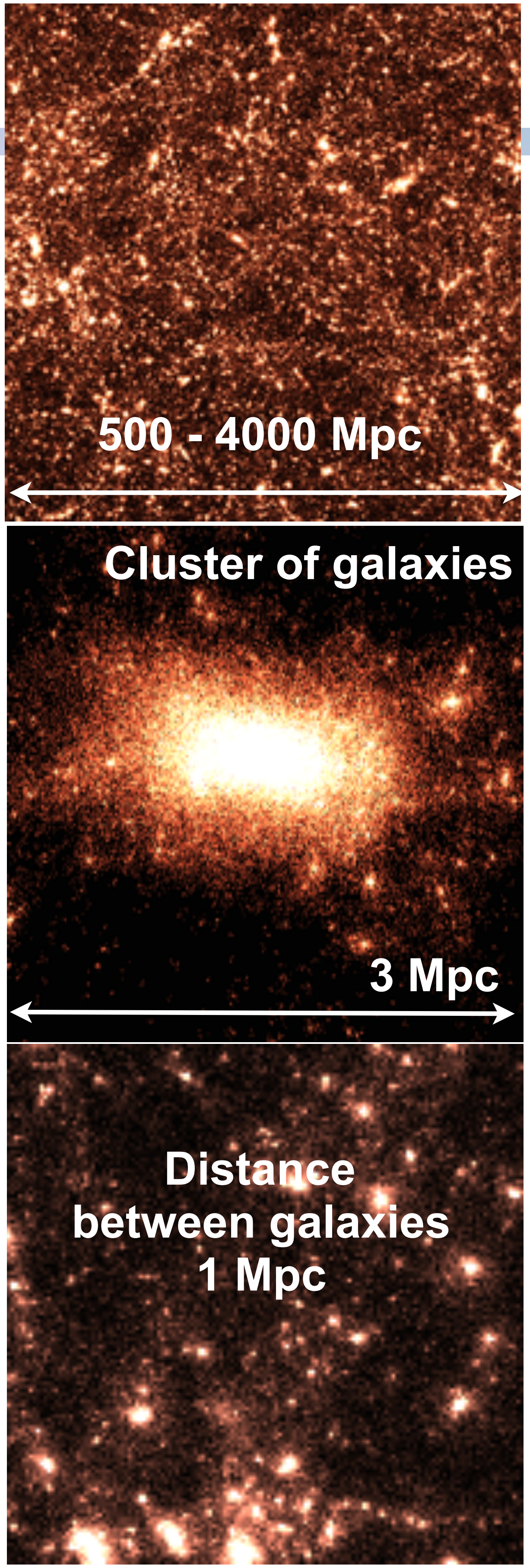


*Q Continuum: Extradimensional plane of existence

Visualization: Silvio Rizzi, Joe Insley et. al., Argonne

Simulating Surveys

- **Simulation Volume:** Large survey sizes impose simulation volumes ~(4 Gpc)^3, memory required ~**100**TB **- 1**PB
- **Number of Particles:** Mass resolution depends on ultimate object to be resolved, $\sim 10^8 \mathrm{M}_\odot - 10^{10} \mathrm{M}_\odot$ (subhalos to halos), N~$10^{11} - 10^{12}$
- **Easy to remember:** ~(1000 Mpc)^3 and ~(1000)^3 particles lead to a mass resolution of $\sim 10^{11} \mathrm{M}_\odot$
- **Therefore:** ~(10,000)^3 particles in ~(1000 Mpc)^3 leads to $\sim 10^8 \mathrm{M}_\odot$, (5Gpc)^3 at this mass resolution would require 125 trillion particles = 5PB per snapshot! (Mira has 0.75PB overall, biggest DOE open machine)
- **Force Resolution:** ~kpc, yields a (global) spatial dynamic range of ~**10^6**



An Early Simulation

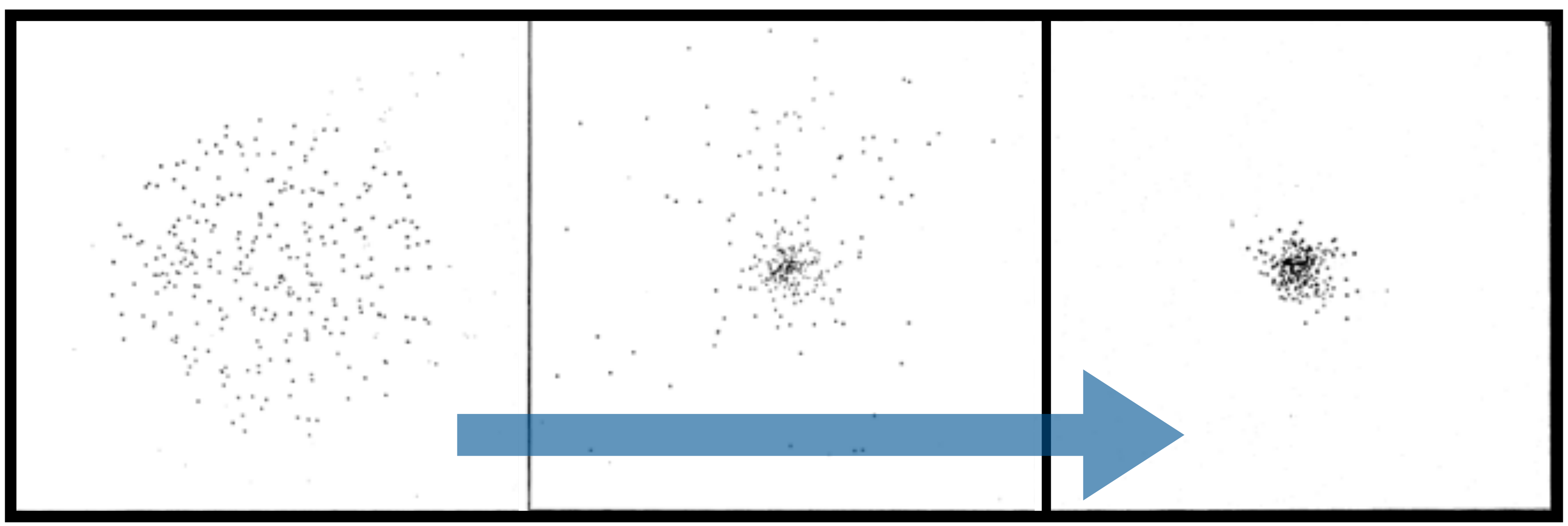
ASTRONOMICAL JOURNAL VOLUME 75, NUMBER 1 FEBRUARY 1970

Structure of the Coma Cluster of Galaxies*

P. J. E. Peebles†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 7 October 1969)

In some cosmologies, a cluster of galaxies is imagined to be a gravitationally bound system which, in analogy with the formation of the Galaxy, originated as a collapsing protocluster. It is shown that a numerical model based on this picture is consistent with the observed features of the Coma Cluster of galaxies. The cluster mass derived from this model agrees with previous values; however, an analysis of the observational uncertainty within the framework of the model shows that the derived mass could be consistent with the estimated total mass provided by the galaxies in the cluster.

- Suite of 300 particle simulations
- Run on a CDC 3600, ~1Mflops*, 32KB+ memory at Los Alamos
- Is 9 orders of magnitude in improvement in both performance and memory enough for precision cosmology?



His Acknowledgement: The lengthy numerical computations were made possible through the generous hospitality of the Los Alamos Scientific Laboratory.

* Today's laptop is ~1000 faster

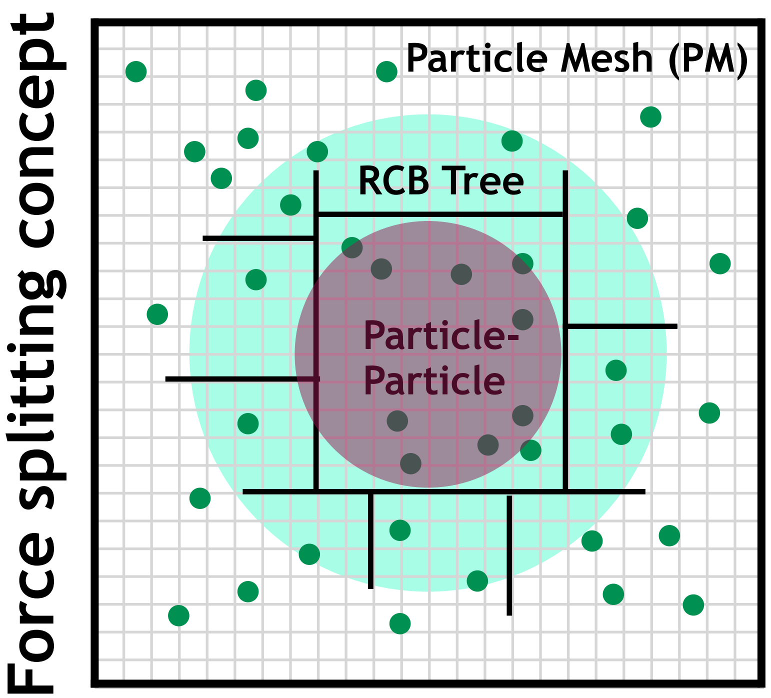
Computing the Universe: HACC

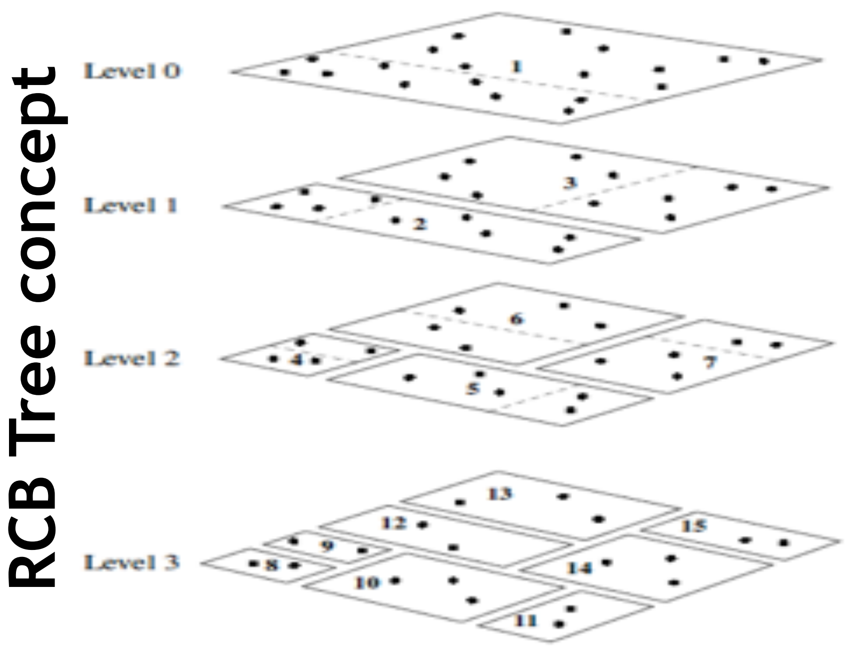
Habib et al. J. Phys. Conf. Ser. 2009; Pope et al. Comp. Sci. Eng. 2010;
Habib et al. SC12, arXiv:1211.4864, Habib et al. SC13, Habib et al. New Astronomy 2016

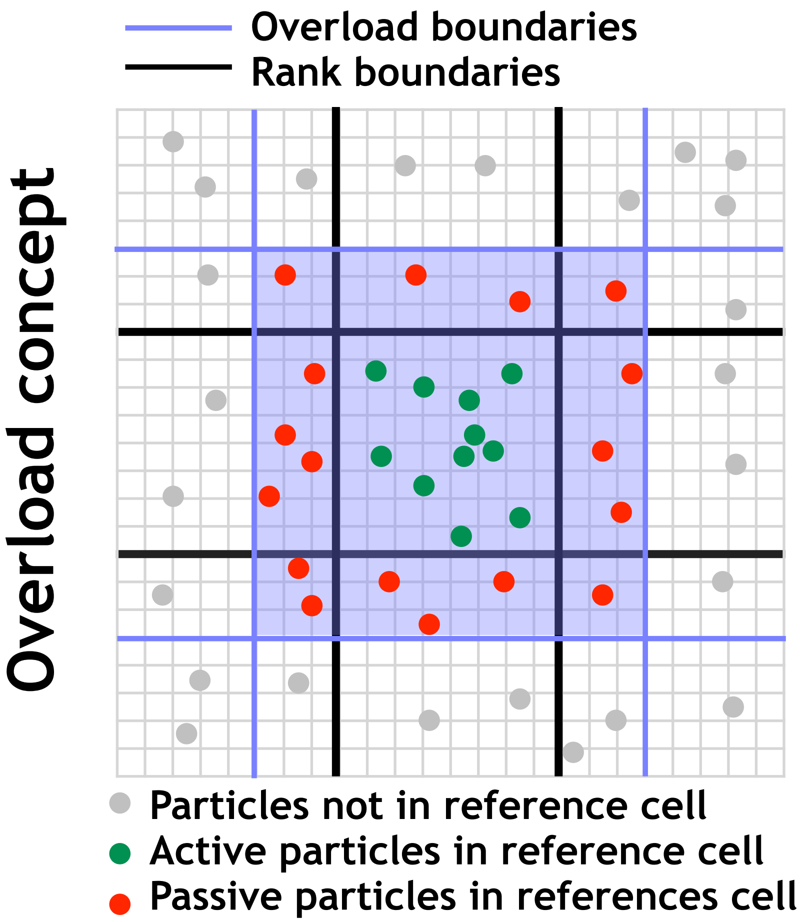
- **HACC:** **H**ardware/**H**ybrid **A**ccelerated **C**osmology **C**ode for N-body simulations at extreme scales
- **The HACC Story**
  - **MC2:** Friendly **M**esh-based **C**osmology **C**ode (particle-mesh code), written in High-Performance Fortran (2002-2008)
  - **MC3:** In 2008, Roadrunner arrives at Los Alamos, first supercomputer to break the Petaflop barrier, **C**ell-accelerated, brand new code, C/C++/MPI and Cell intrinsics
  - **HACC:** Starting in 2010: Code optimized for GPU systems, standard X86, BG/Q
- **Why did we write a new N-body code?**
  - **Portability:** Ever changing landscape of supercomputers poses huge challenge
  - **Scalability:** How to use more than 1 Million cores efficiently?
  - **Flexibility:** It’s always easier to modify our own code!
- **HACC has run on all available architectures successfully at scale!**

HACC in a Nutshell

- **Long-range/short range force splitting:**
  - **Long-range:** Spectral Particle-Mesh solver, C/C++/MPI, **unchanged for different architectures**, FFT performance dictates scaling (custom pencil decomposed FFT)
  - **Short-range:** **Depending on node architecture** switch between tree and particle-particle algorithm; tree needs "thinking" (building, walking) but computationally less demanding (BG/Q, X86), PP easier but computationally more expensive (Cell, GPU)
- **Overload concept to allow for easy swap of short-range solver and minimization of communication (reassignment of passive/active in regular intervals)**



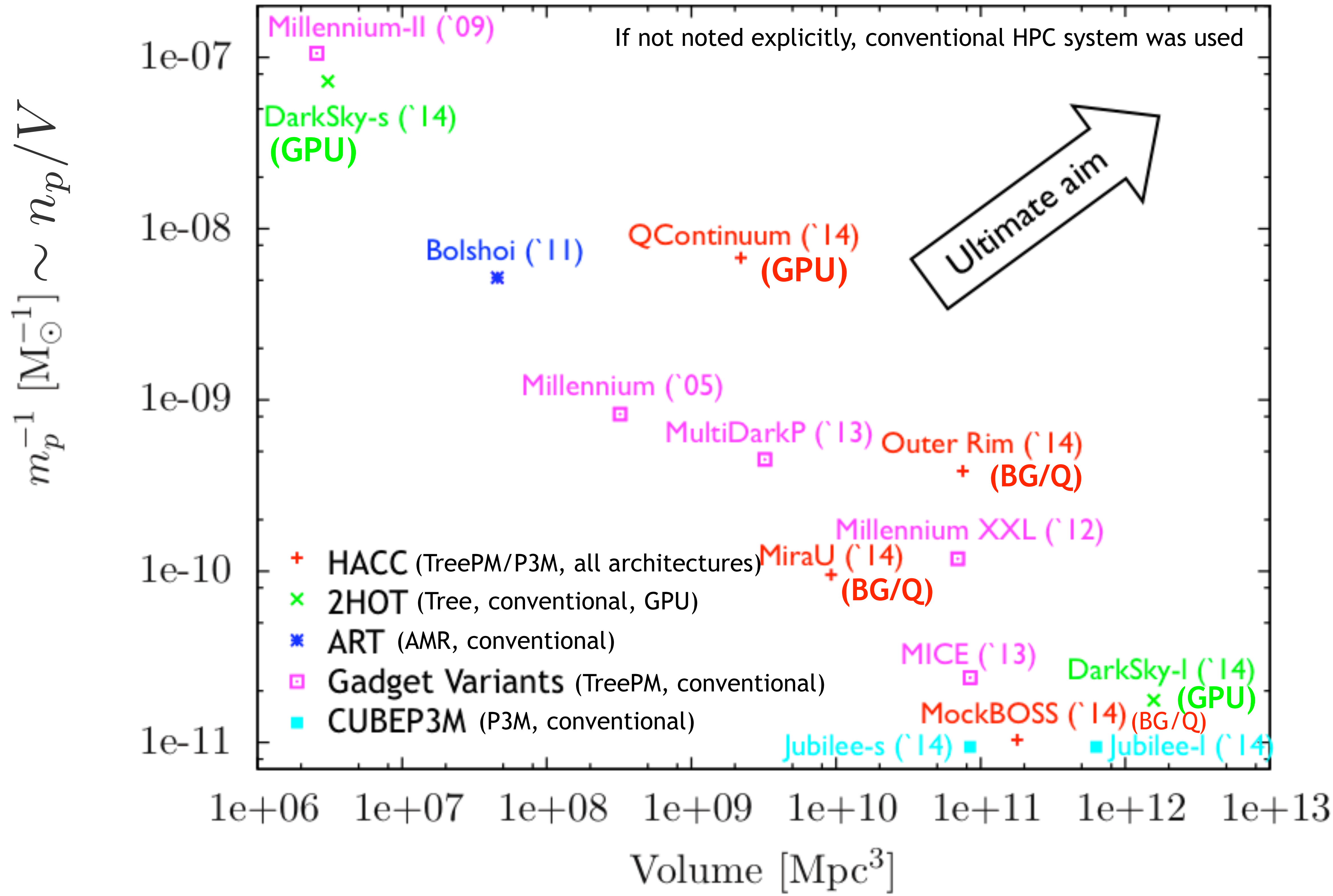




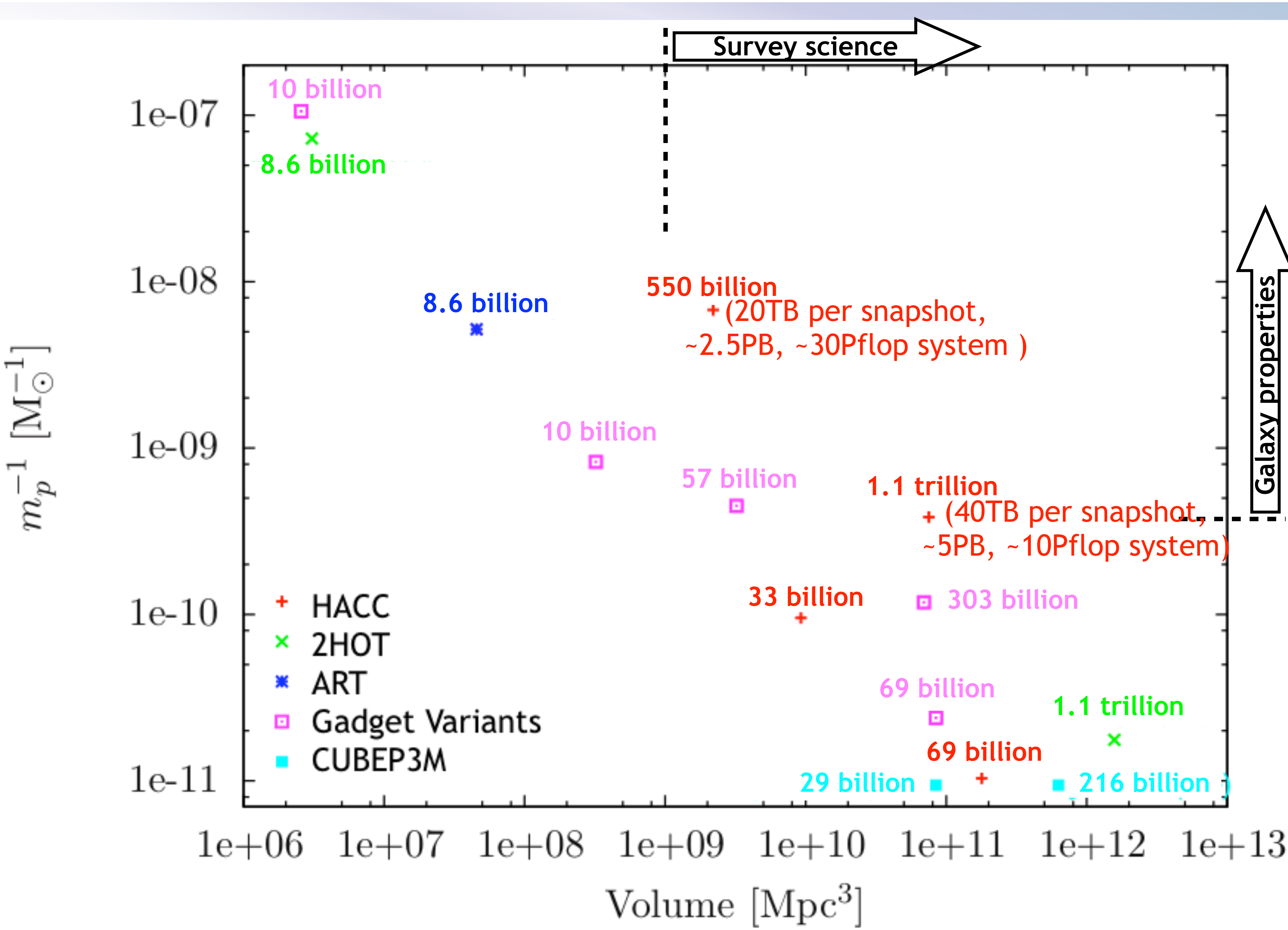
MC2
Auror
SUMMIT
Edison
CRAY
TITAN
MIRA

Current State-of-the-Art Simulations, Gravity Only

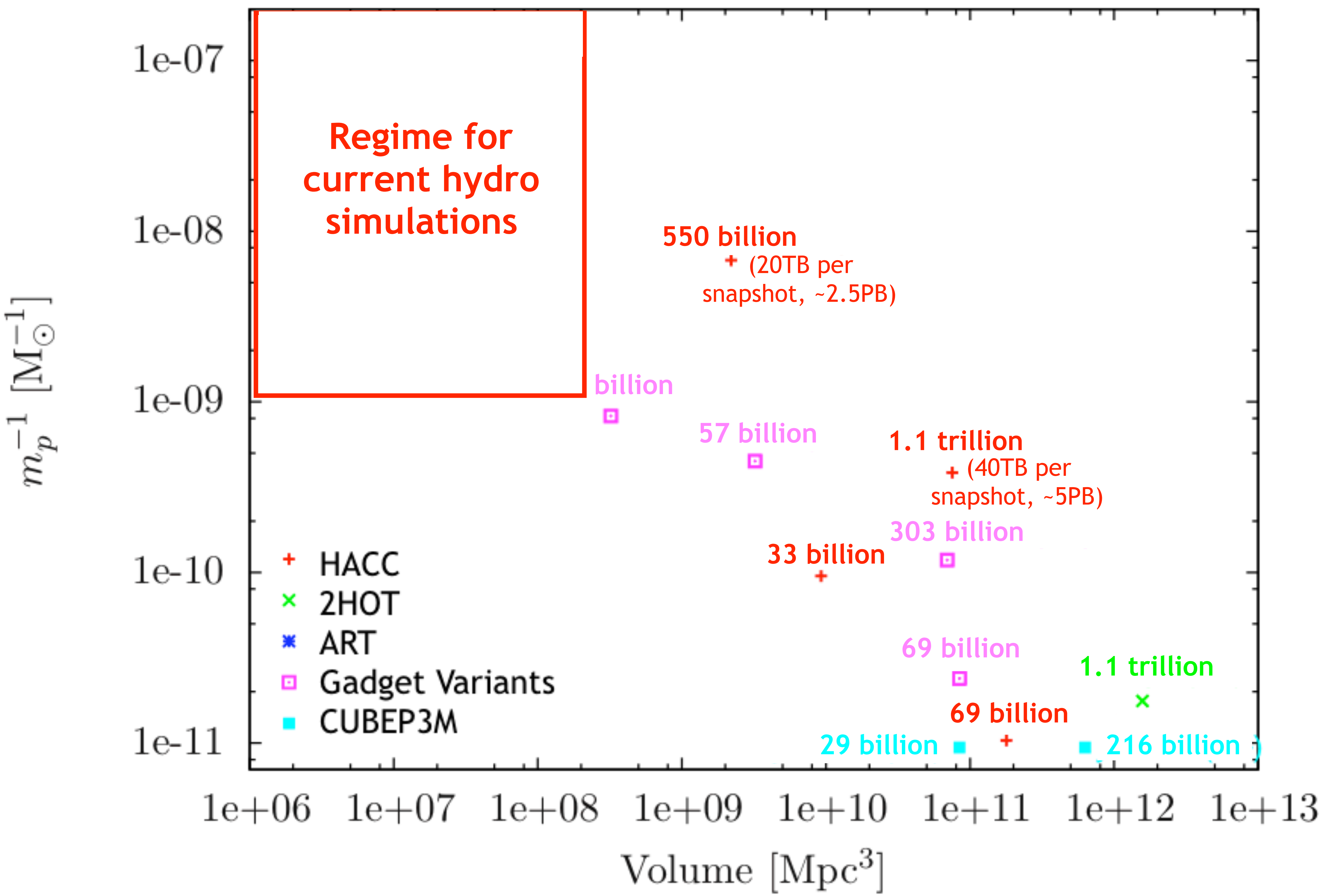
Focus on simulations important for Survey Science



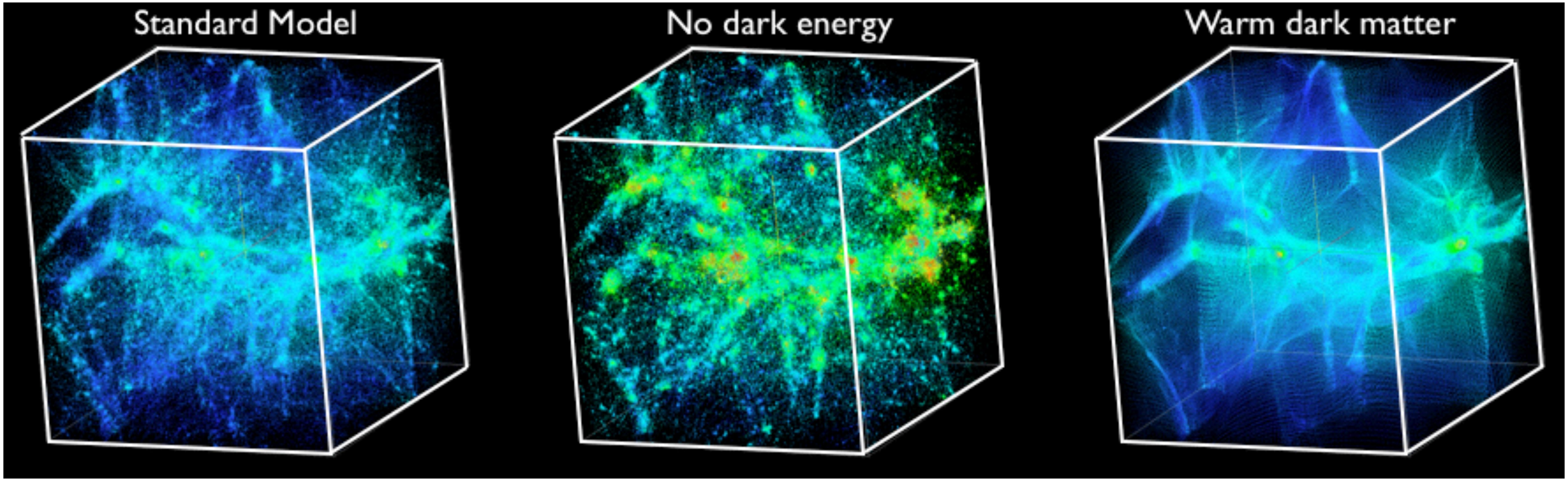
Gravity Only, Number of Particles



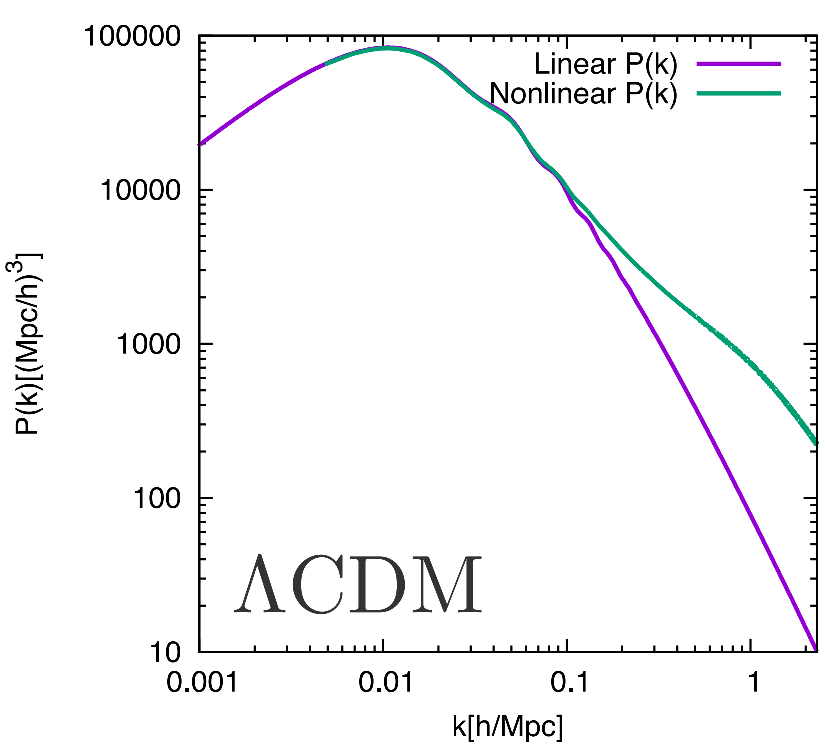
Current State-of-the-Art Hydro Simulations



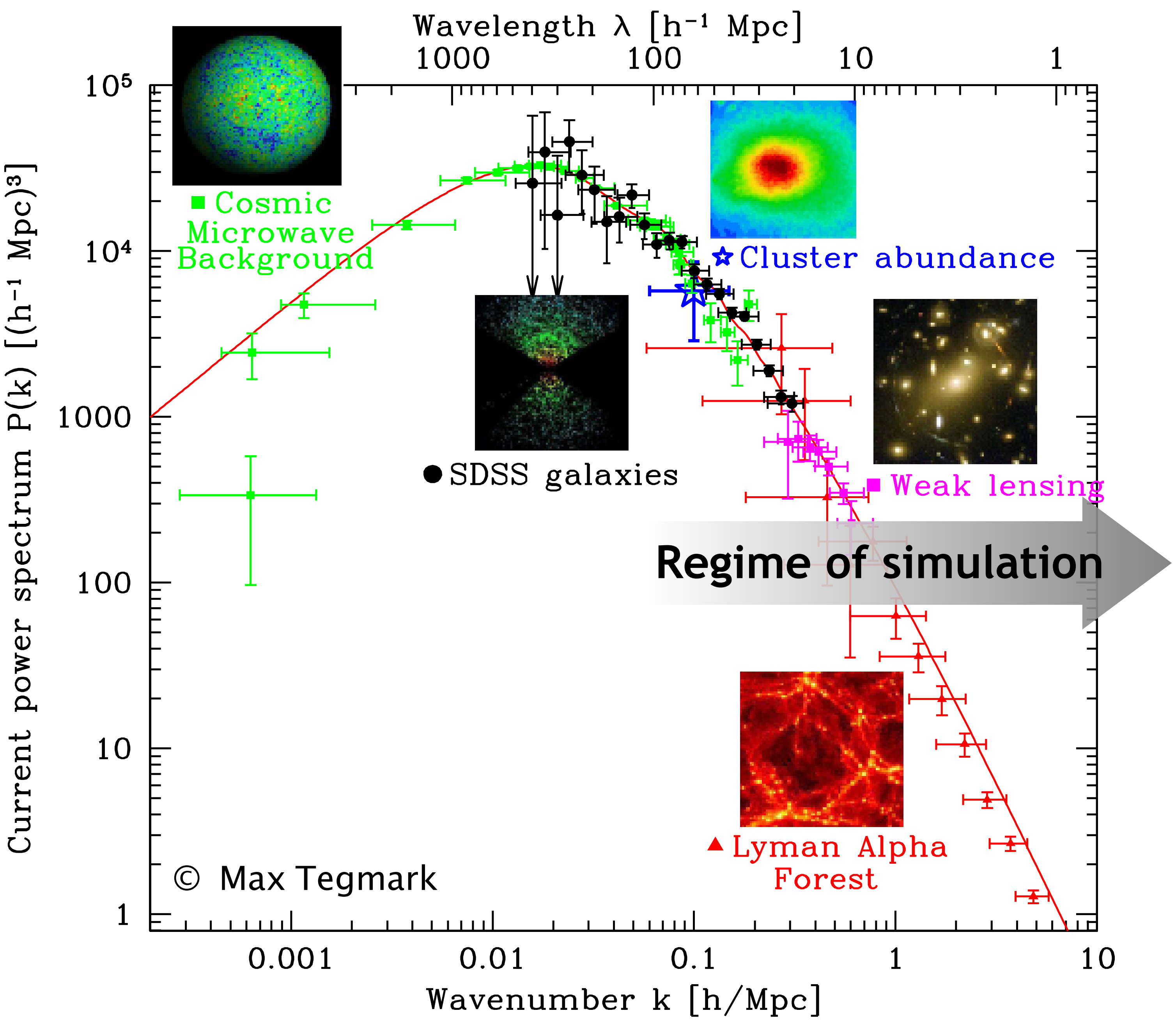
Exploring Structure Formation in the Dark Universe



- **Exploration of different dark energy models:** How would a dark energy model beyond Einstein's cosmological constant alter the distribution of matter (and galaxies) in the Universe?
- **Exploration of dark matter and neutrinos in the Universe:** What can cosmology tell us about different matter components in the Universe?
- **In the simulation:** Adopt the correct initial condition, change the background equations appropriately, adjust the gravitational interaction, maybe add new species, …



The Matter Power Spectrum



2-point correlation function:

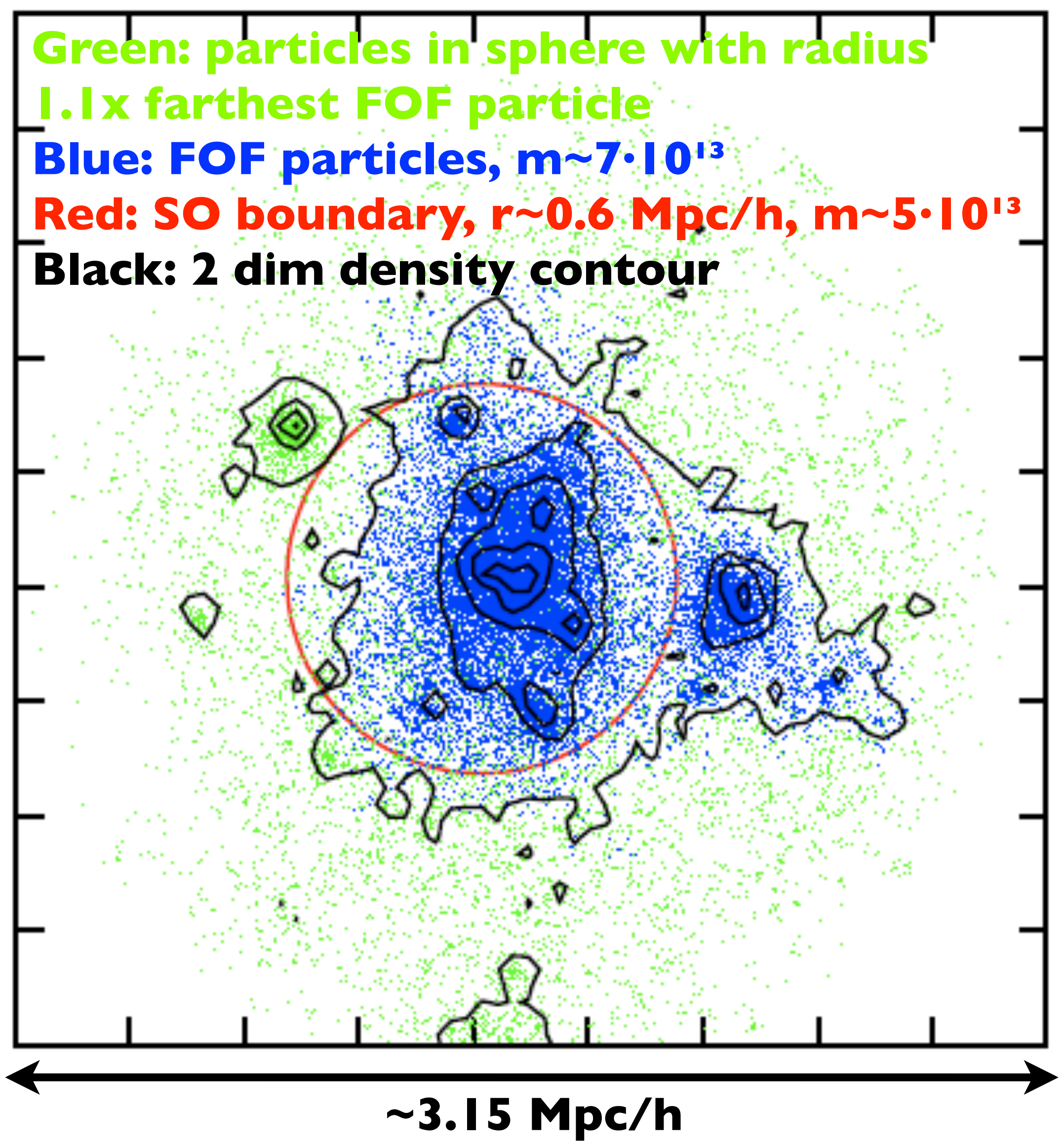
$$\xi(\vec{x}) = \int \frac{d^3\vec{y}}{V}\delta(\vec{y}-\vec{x})\delta(\vec{y}) = \int \frac{d^3\vec{k}}{(2\pi)^3 V}\underbrace{|\delta_k|^2}_{\text{power spectrum}} e^{i\vec{k}\cdot\vec{x}}$$

- 2-point correlation function: excess probability of finding an object pair separated by a distance r_{12} compared to that of a random distribution
- P(k): power spectrum, Fourier transform of correlation function

$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}$$

- Power spectrum very sensitive to physics of interest: amount and properties of dark matter, dark energy, neutrino mass, ...
- Many different probes for measuring P(k)

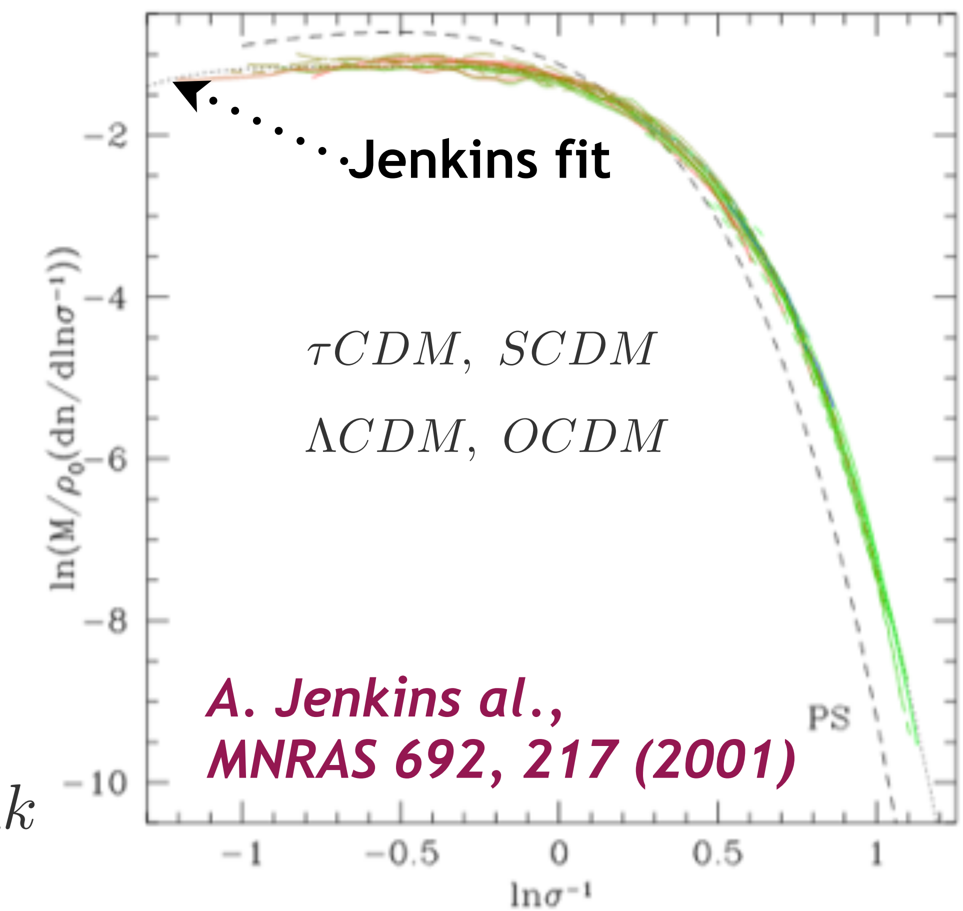
Halos



- HACC delivers large-volume simulations with very high mass resolution, therefore excellent halo statistics
- Halo statistics (e.g., mass function) as well as halo properties (e.g., profile) hold wealth of cosmology information as well as insights about structure formation
- How to define a halo?
  - **Friends-of-friends:** Follows iso-density contours, hence tracks shapes of bound objects faithfully, strong/weak lensing masses
  - **Overdensity:** Spheres around density peaks, easy to relate to e.g X-ray gas, but halos are not really spheres ...
  - **Choose halo definition with observational probe in mind, none is perfect!**

Halo Profiles and Mass Function

- **Two remarkable results derived from simulations:**
- **"Universality" of the halo mass function for FoF halos, b=0.2 halos**
  - When expressed via σ_i^{-1} a single formula fits all the mass functions for different cosmologies and over different epochs at the 10-20% level
  - $\sigma^2(M,z) = \frac{D_+^2(z)}{2\pi^2}\int_0^\infty k^2 P_{lin}(k) W^2(k;M)\mathrm{d}k$
- **"Universality" of the halo density profile**
  - Navarro, Frenk, White: Relaxed halos in simulations can be described by remarkably simple radial profile
  - Concentration c and mass m characterize halo



NFW profile:

$$\rho(r) = \frac{\delta\rho_{\rm crit}}{(r/r_s)(1+r/r_s)^2}$$

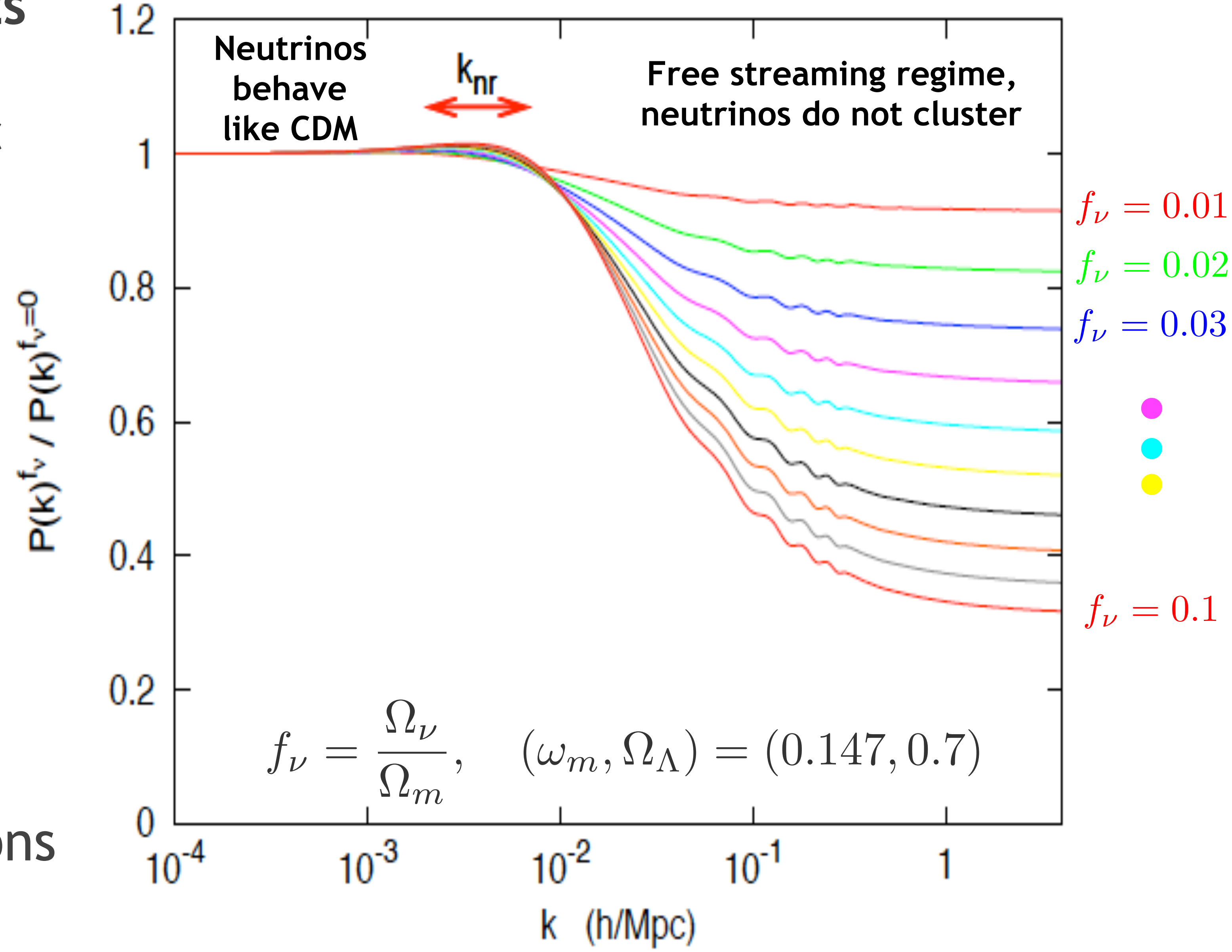
$$c_\Delta = r_\Delta/r_s$$

NFW, ApJ, 490, 493 (1997)

Neutrinos and Large Scale Structure

- Cosmology (CMB and large scale structure) provides constraints on the mass sum of neutrinos and the number of relativistic species
- Massive neutrinos have **scale dependent** effect on power spectrum
  - **On large scales:** Neutrino fluctuations grow like dark matter fluctuations
  - **On small scales:** Free-streaming suppresses the growth of neutrino fluctuations

Influence of three massive neutrinos on the linear matter power spectrum

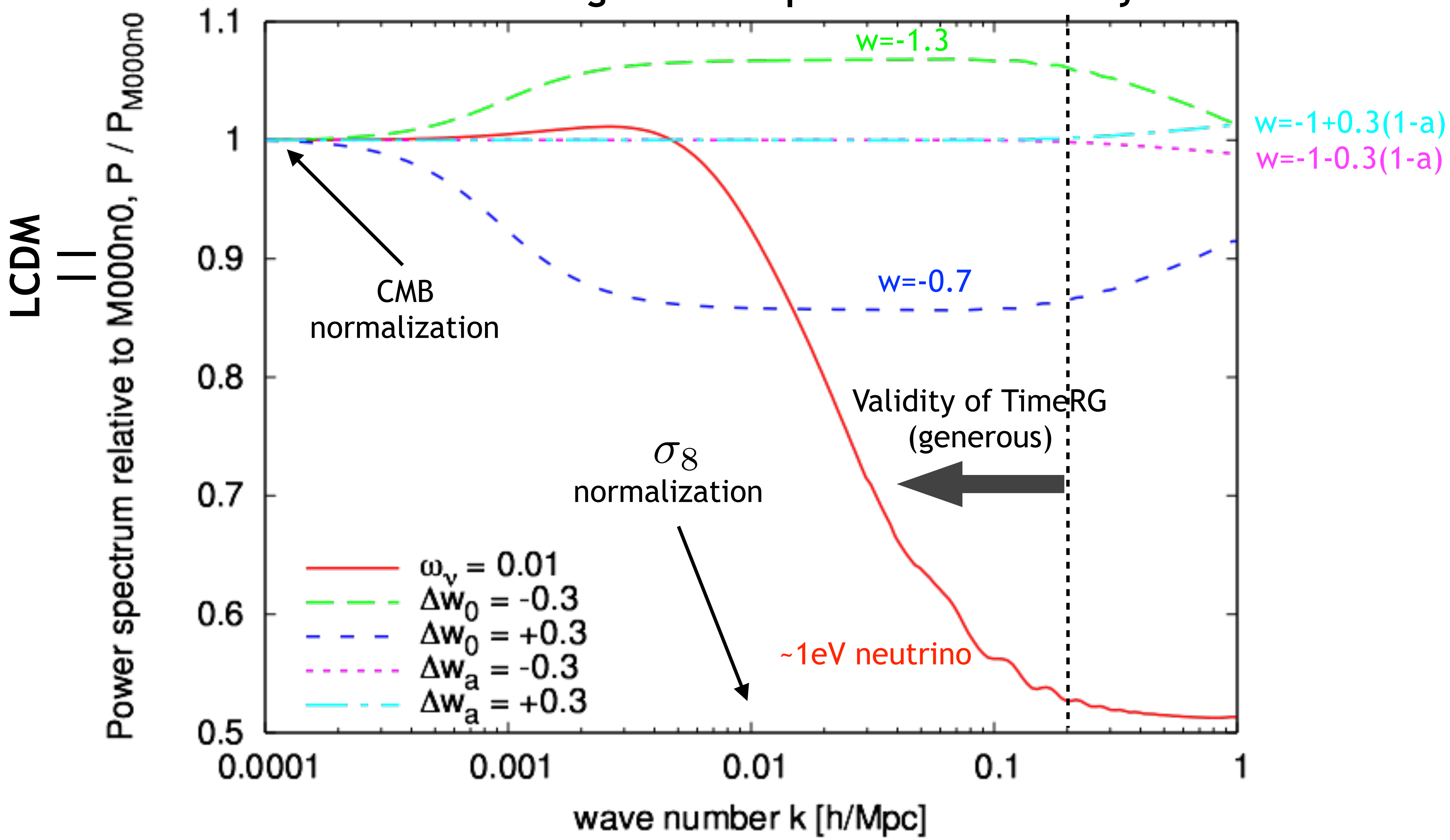


Lesgourgues & Pastor, Phys. Rept. 429, 307, 2006

Neutrinos and Dark Energy: The Power Spectrum

- Need to study neutrinos and dynamical dark energy simultaneously due to degenerate effects on e.g. the power spectrum (see, e.g., Santos et al. 2013 for Euclid forecasts)

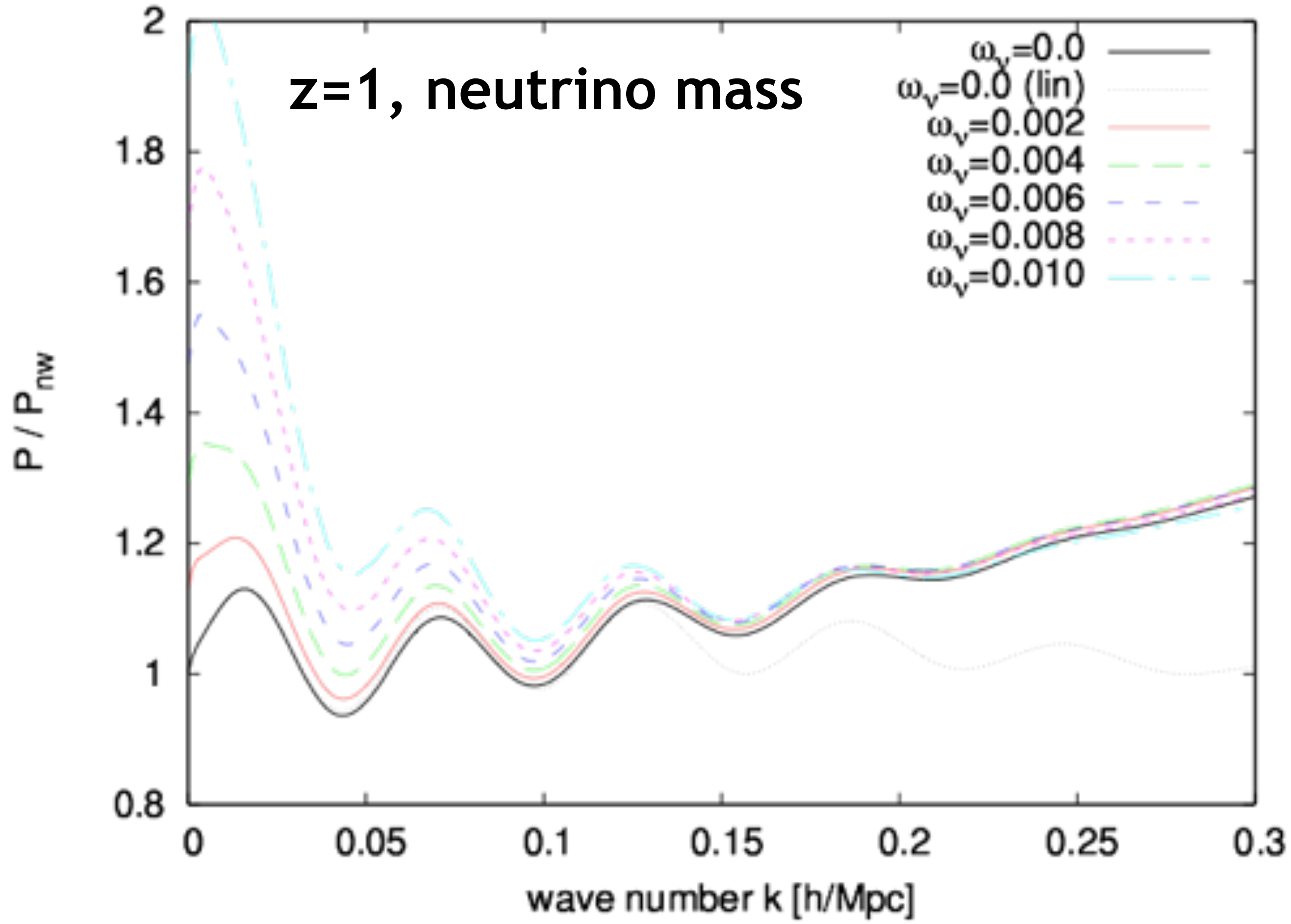
Results from higher order perturbation theory



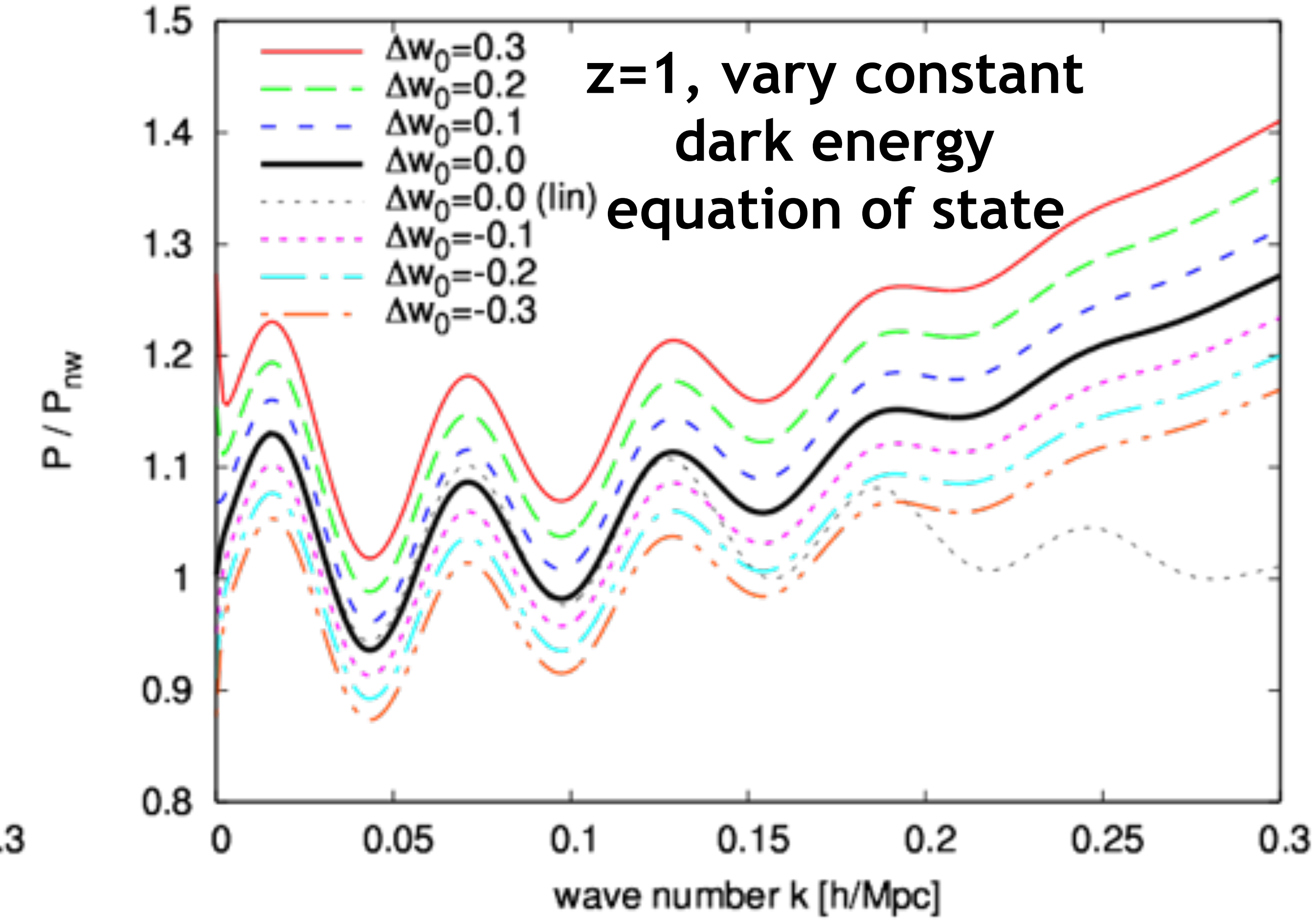
Upadhye et al., arXiv:1309.5872, Phys. Rev. D

Some more thoughts on P(k)

LCDM + massive neutrinos



Early dark energy + massive neutrinos



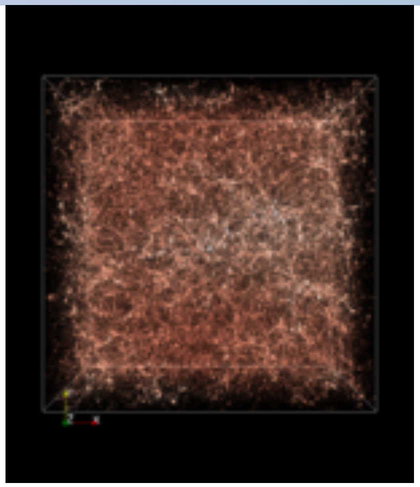
- Investigate different physical effects on e.g. the matter power spectrum at high accuracy
- Here: 2100 Mpc simulations with ~32.8 billion particles, $10^{10}\mathrm{M}_\odot$ particle mass to study massive neutrinos as well as time varying dark energy EOS
- Develop "higher order" perturbation theory approach

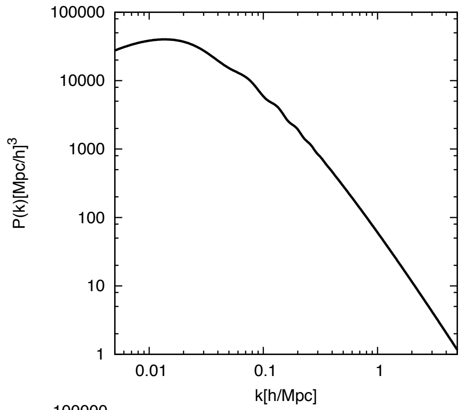
A. Upadhye et al. 2013

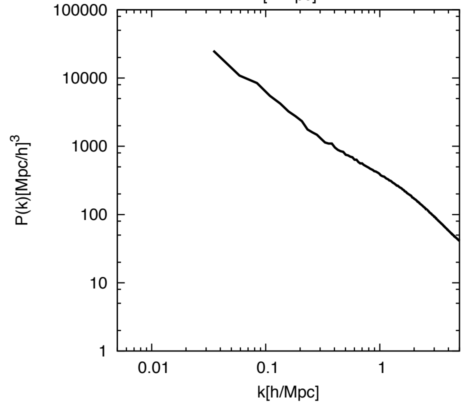
A Quiz –

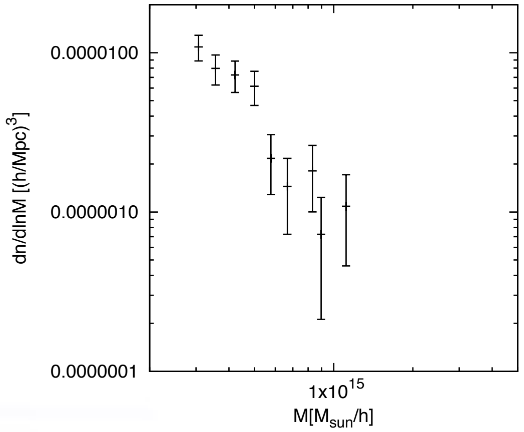
- We have carried out four simulations, following four different cosmologies given to you
  - 256 Mpc/h, 128^3 particles
  - For each of them, you have (i) an image of the dark matter particles colored with respect to their potential (coarse), (ii) a linear P(k), (iii) a nonlinear P(k), (iv) a mass function

1. Quickly determine the differences in the cosmologies by examining the parameter sheet given to you
2. Discuss the effects on structure formation you would expect from the cosmological parameters that differ from a vanilla LCDM model
3. Four each cosmology, find the matching particle distribution, linear P(k), nonlinear P(k), and mass function
4. Which was your favorite "observable", which one was the easiest to assign and which one the most difficult?



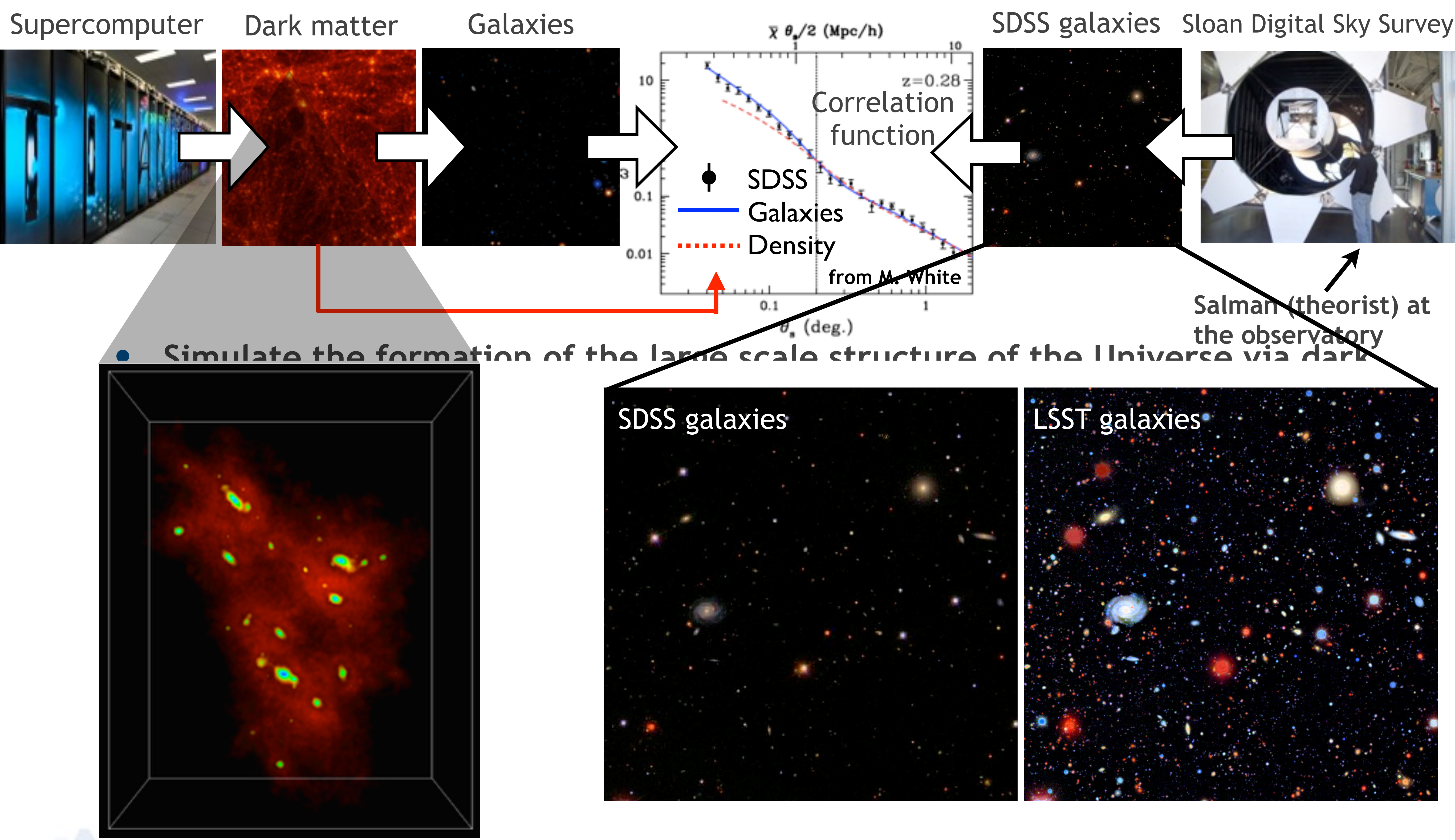






Connecting Theory and Observations

Connecting Theory and Observations



- Simulate the formation of the large scale structure of the Universe via dark

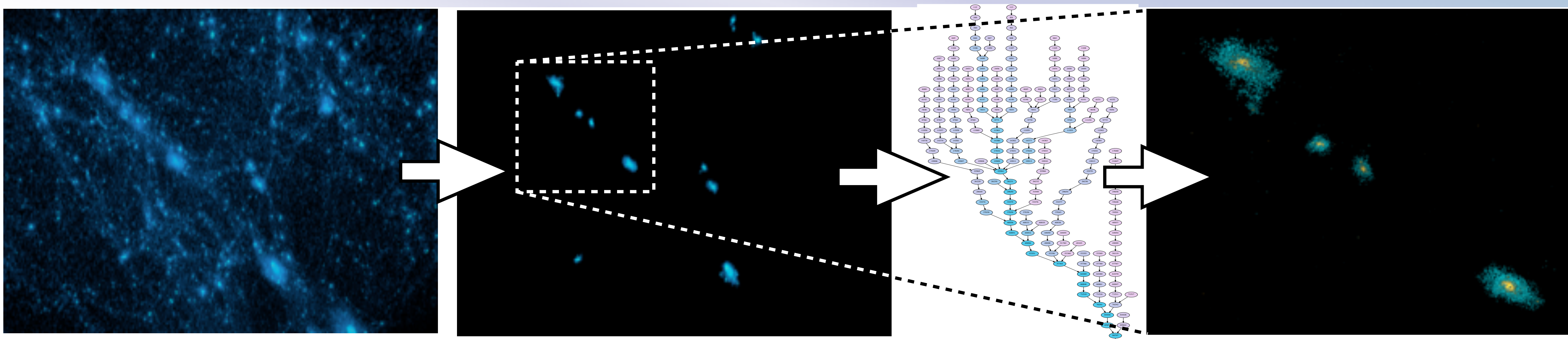
Questions –

- What halo properties can inform you about the galaxies that reside in them?
- What would be the connection of the properties to the galaxies?
- What physics processes are important for galaxy formation that you would want to incorporate in your simulation?
- What do you think are the biggest challenges for generating synthetics maps for LSST DESC?

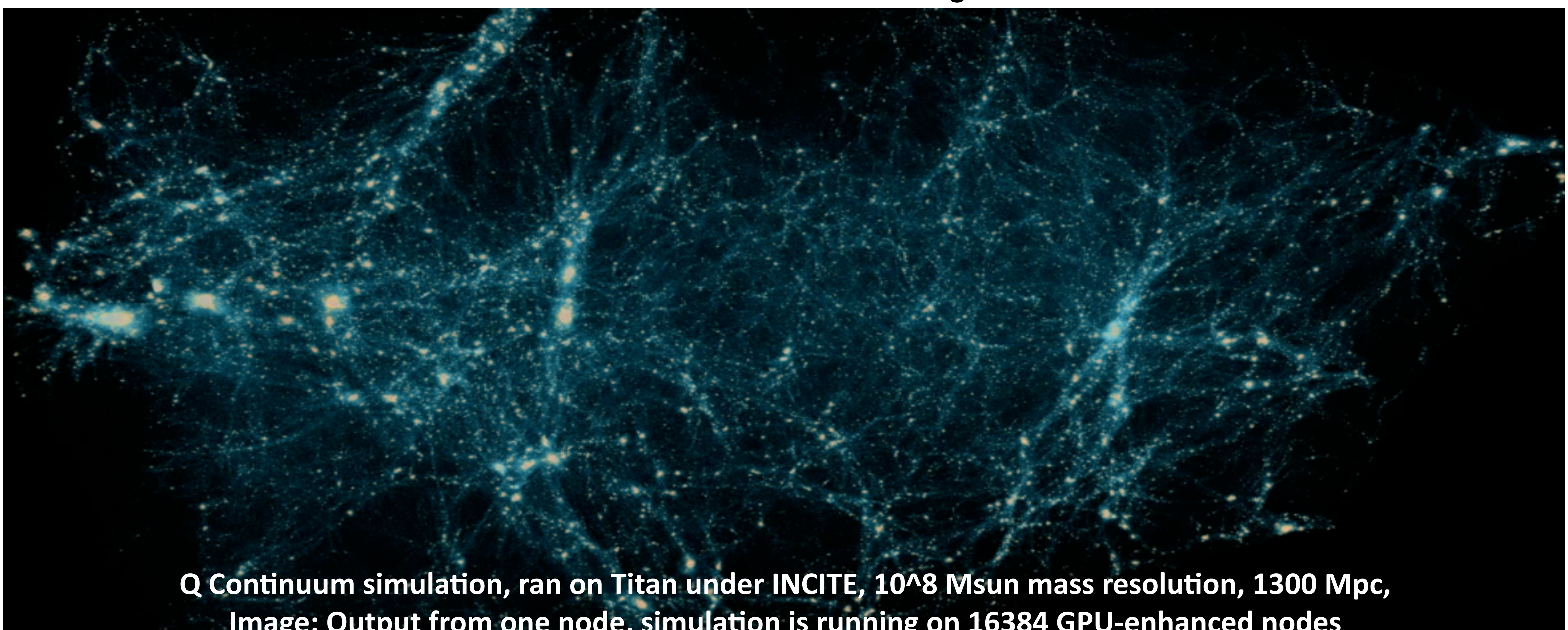
Computing the Universe: SHAM, HAM, CAM, SAM ...

- Let's return to the problem of modeling the galaxies on top of high-quality gravity-only simulations
- **Advantage of Post-processing:** Flexibility in how we model the galaxies, can be tailored to galaxies of interest and the resolution of the simulations
- **Halo Occupation Distribution (HOD):** Very popular for building mocks for LRGs (SDSS, BOSS); populate halos with central and satellite galaxies depending on their mass (for SDSS LRG HOD see, e.g. Zheng et al. 2009), moderate mass resolution required, parameters set by matching to observations
- **Sub-Halo Abundance Matching (SHAM):** Takes into account information about subhalo population as well as formation history of halos (see e.g. Conroy et al. 2006), requires rather high mass resolution
- **Conditional Abundance Matching (CAM):** Include merger tree information, Hearin et al. 2013++ "*The Dark Side of Galaxy Color*"
- **Semi-Analytic Modeling (SAM):** Extract merger trees (including subhalos) from high-resolution simulations and model star formation, feedback, gas cooling on top of those, zoo of parameters (~250), but: diverse set of galaxy properties will be predicted, in principle, will give answers at unobserved redshifts

SAM in Action



HACC simulation | Insitu halo finder | Merger tree code | Galacticus



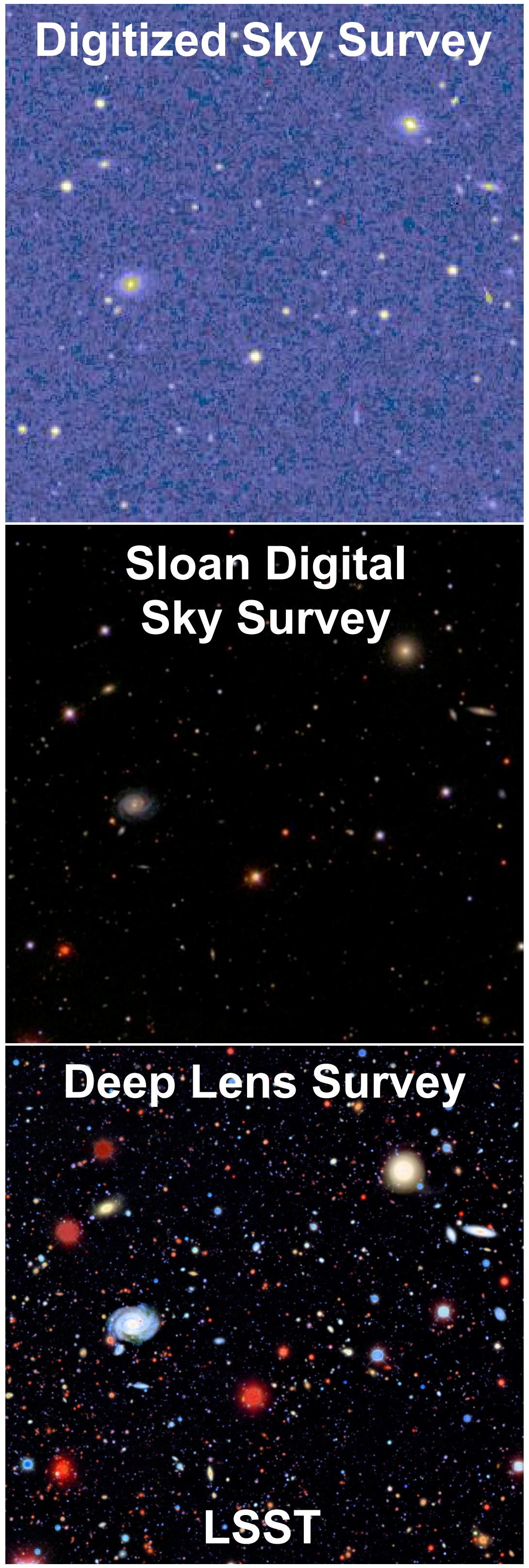
The Role of Cosmological Simulations in LSST DESC

Roles of Cosmological Simulations in Survey Science

Past

Present

Future

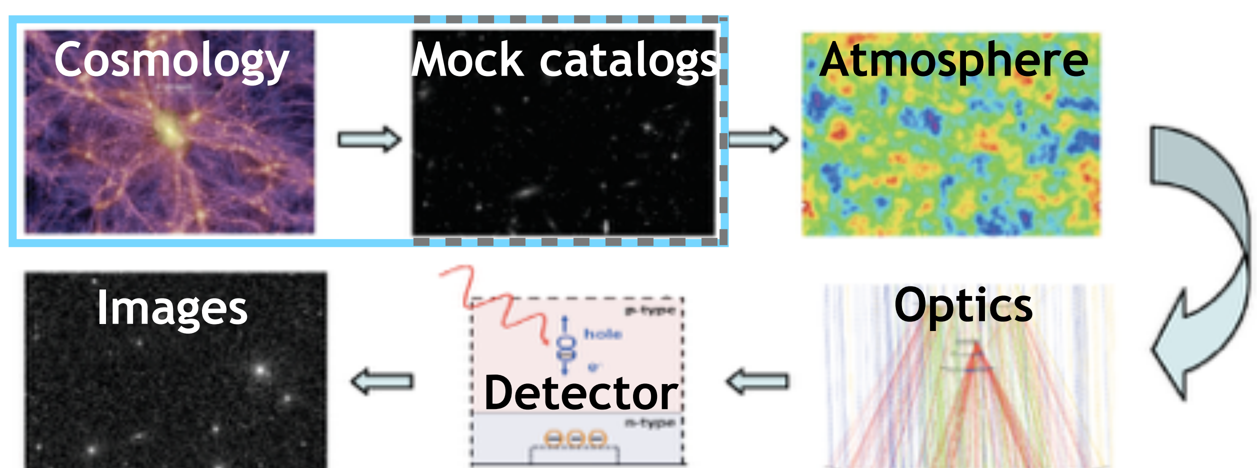


(1) Solving the Inverse Problem

- Exploring fundamental physics
- Fast, very accurate predictions tools (emulators) for physics and observables of interest
- Astrophysical systematics
- Predictions for covariances

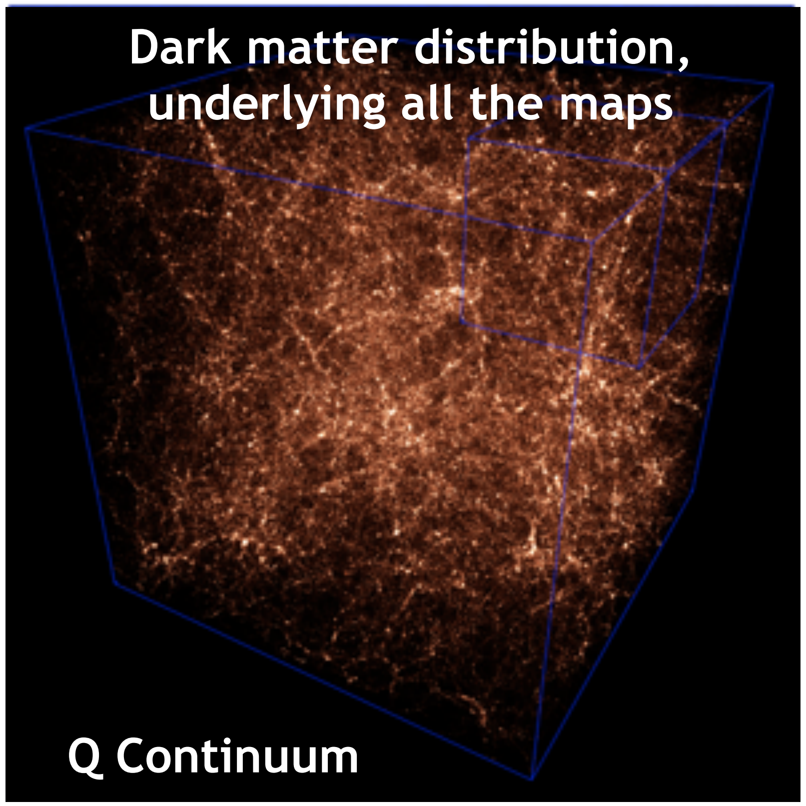
(2) Cosmology simulations and the survey

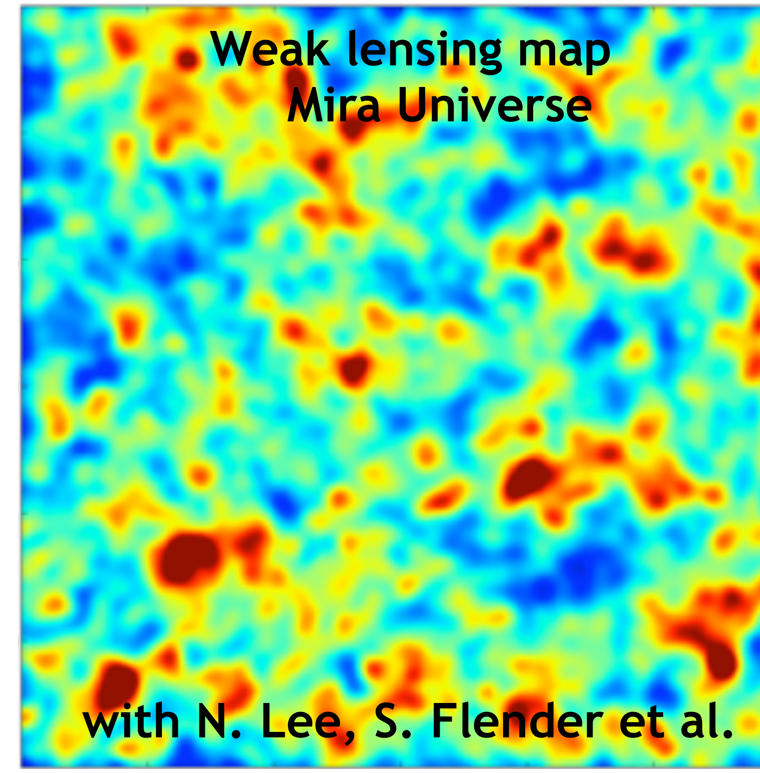
- First part of end-to-end simulation
- Control of systematics

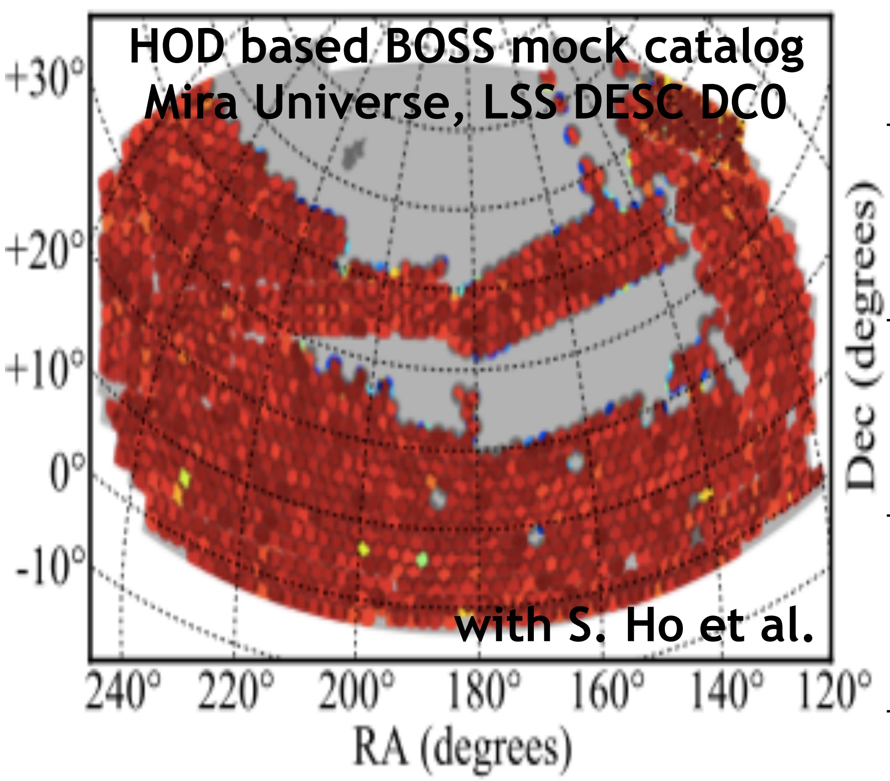


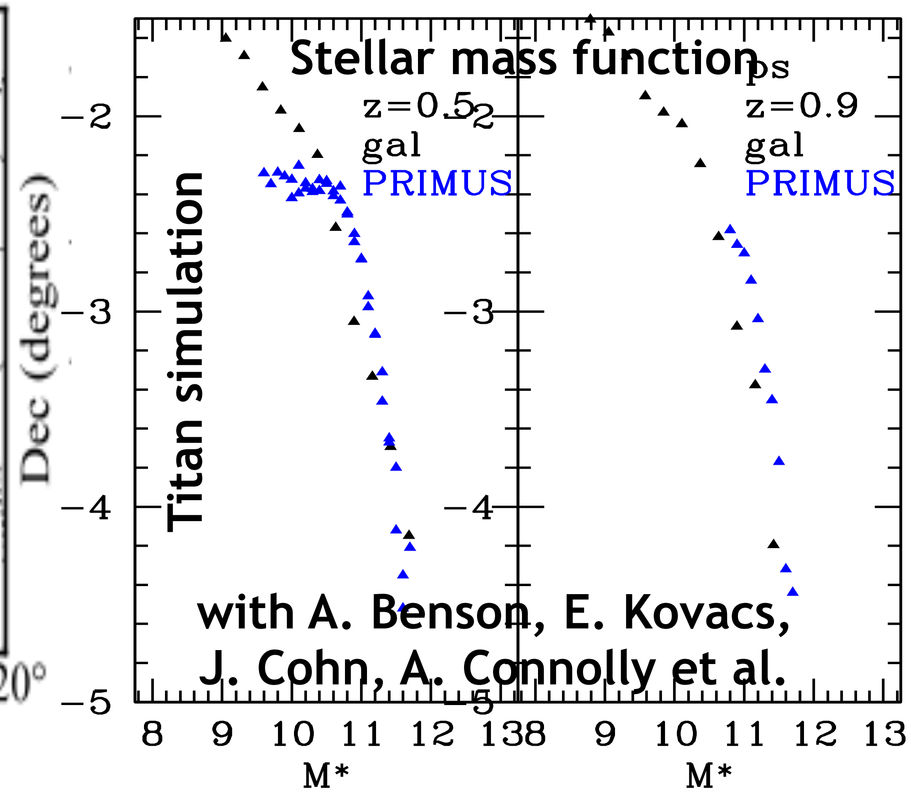
from the LSST Science Book

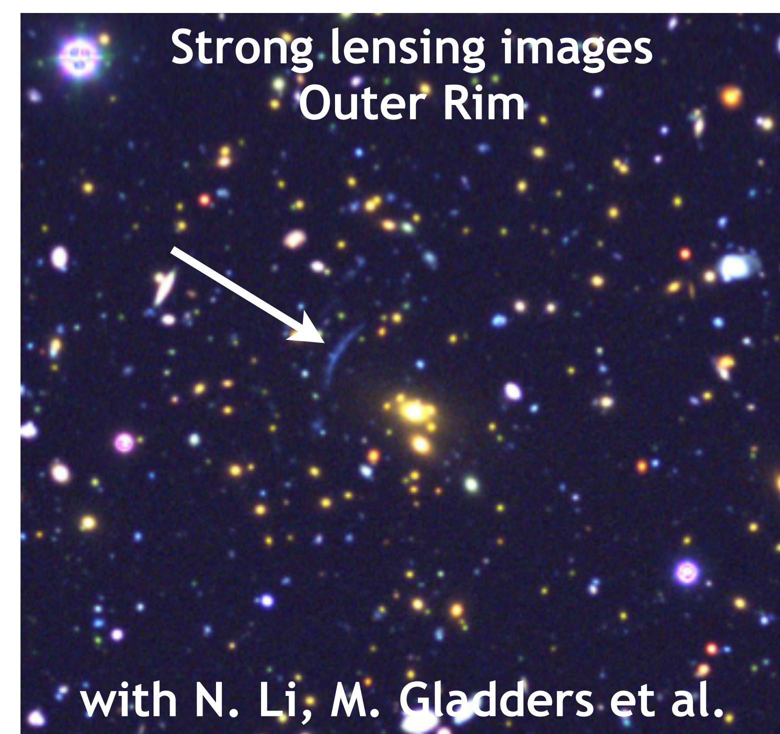
Examples for DESC Synthetic Sky Maps









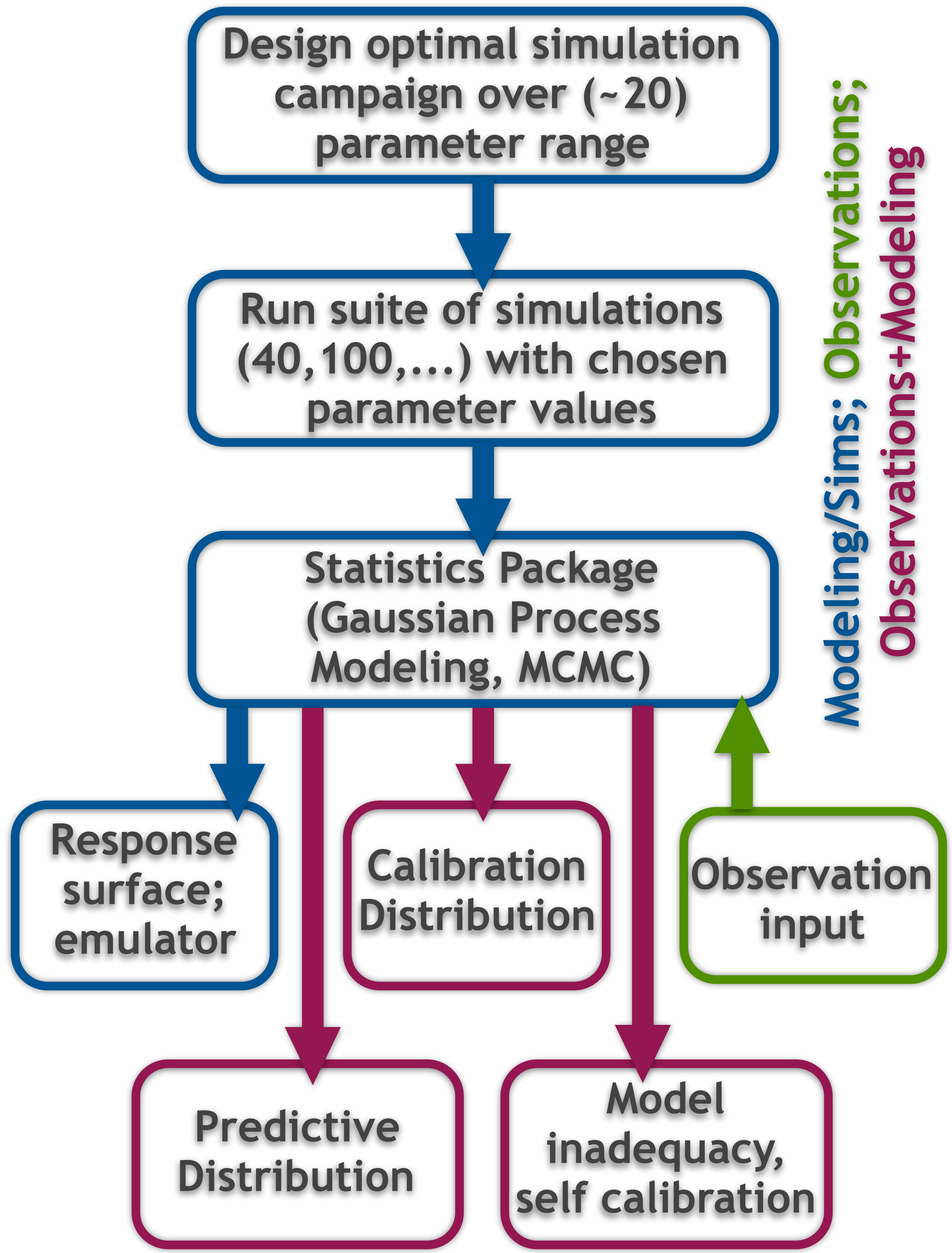


Discussion

- What are you most excited about with regard to LSST?
- Which is the dark energy probe within DESC you will be focusing mostly on?
- What is your biggest "concern" with regard to this probe that you think you will spend most of your time on?
- How do you think simulations could help you with your work to tackle those concerns?
- What kind of simulations would you need for your work and what scales do they have to cover?
- Do you think it will be easy/hard/impossible to use simulations to help with your problem?
- Would you be willing to work with the simulation groups directly to "enhance" the simulations in such a way that they will be useful to solve your problem?

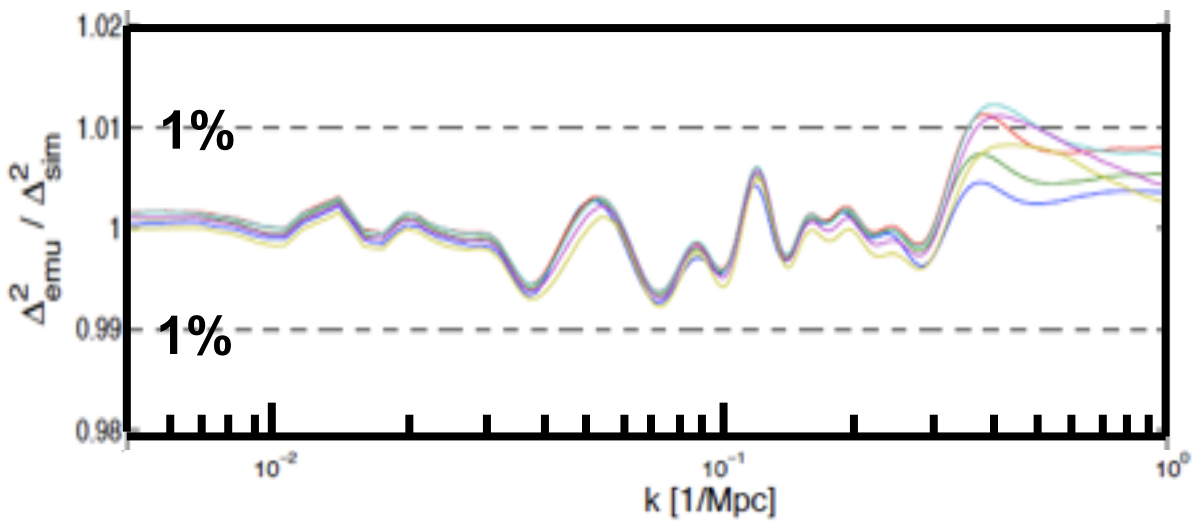
Cosmic Emulators

- **Challenge:** Extract cosmological constraints from observations in the nonlinear regime, need to run Markov Chain Monte Carlo code, input: 10,000 - 100,000 different models
- **Current Strategy:** Fitting functions for, e.g., P(k), accurate at the 10% level, this is not good enough! This will lead to biases in the dark energy equation of state estimates that could escape notice
- **Our alternative:** Emulators, fast prediction schemes build from a limited set of high-accuracy simulations

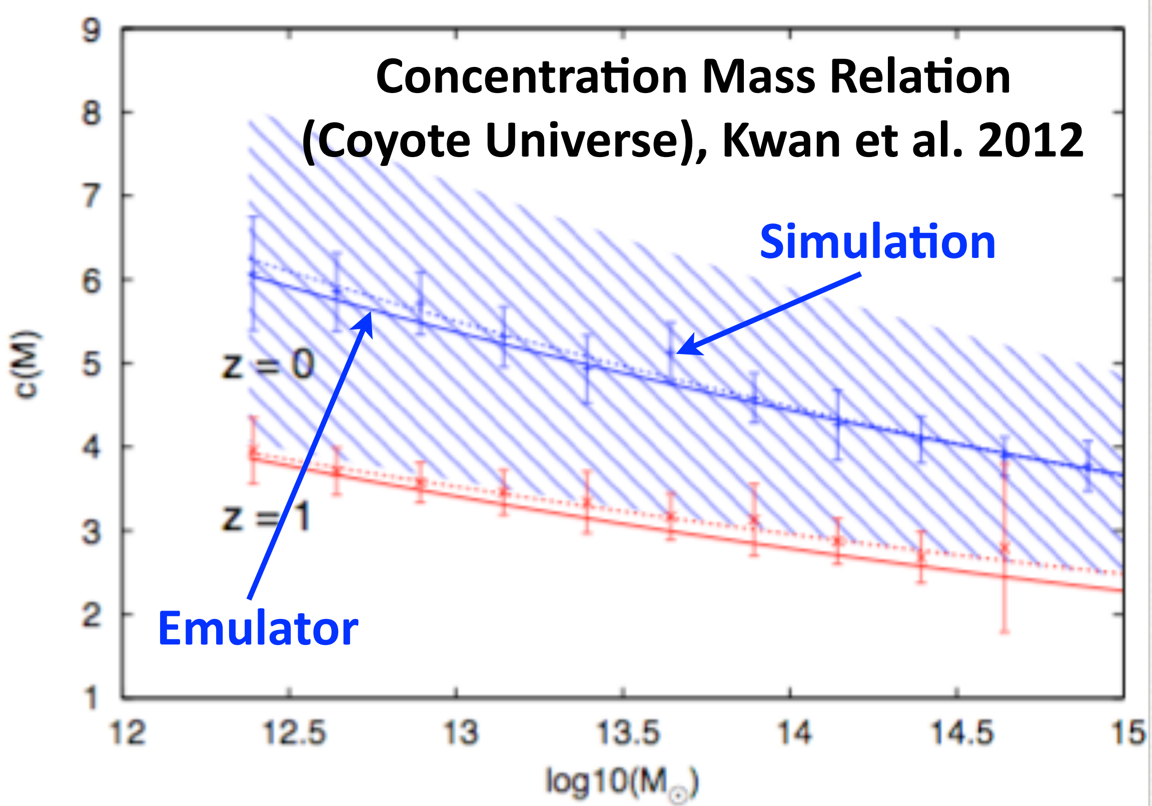


Cosmic Emulator Gallery

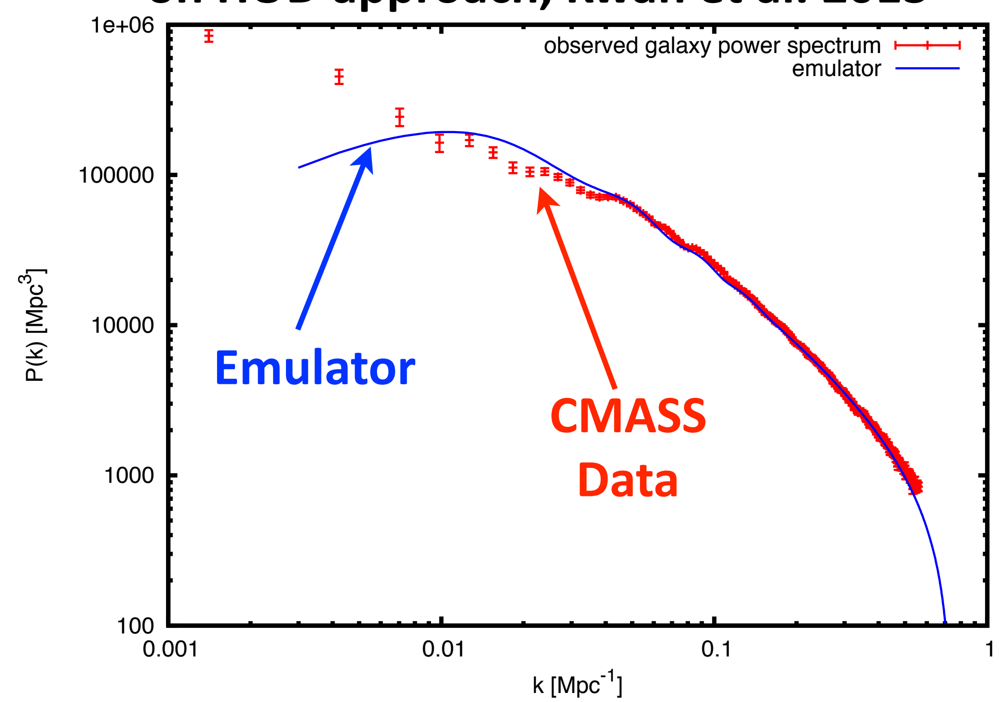
Matter power spectrum (Coyote Universe), Lawrence et al. 2010, Heitmann et al. 2013



Matter power spectrum prediction for wCDM cosmology out to k~1/Mpc at ~1% accuracy, out to k~10/Mpc at ~3% accuracy



Galaxy power spectrum (Mira Universe) based on HOD approach, Kwan et al. 2013



https://hep.anl.gov/cosmology/CosmicEmu

DESC Science: Strong Lensing

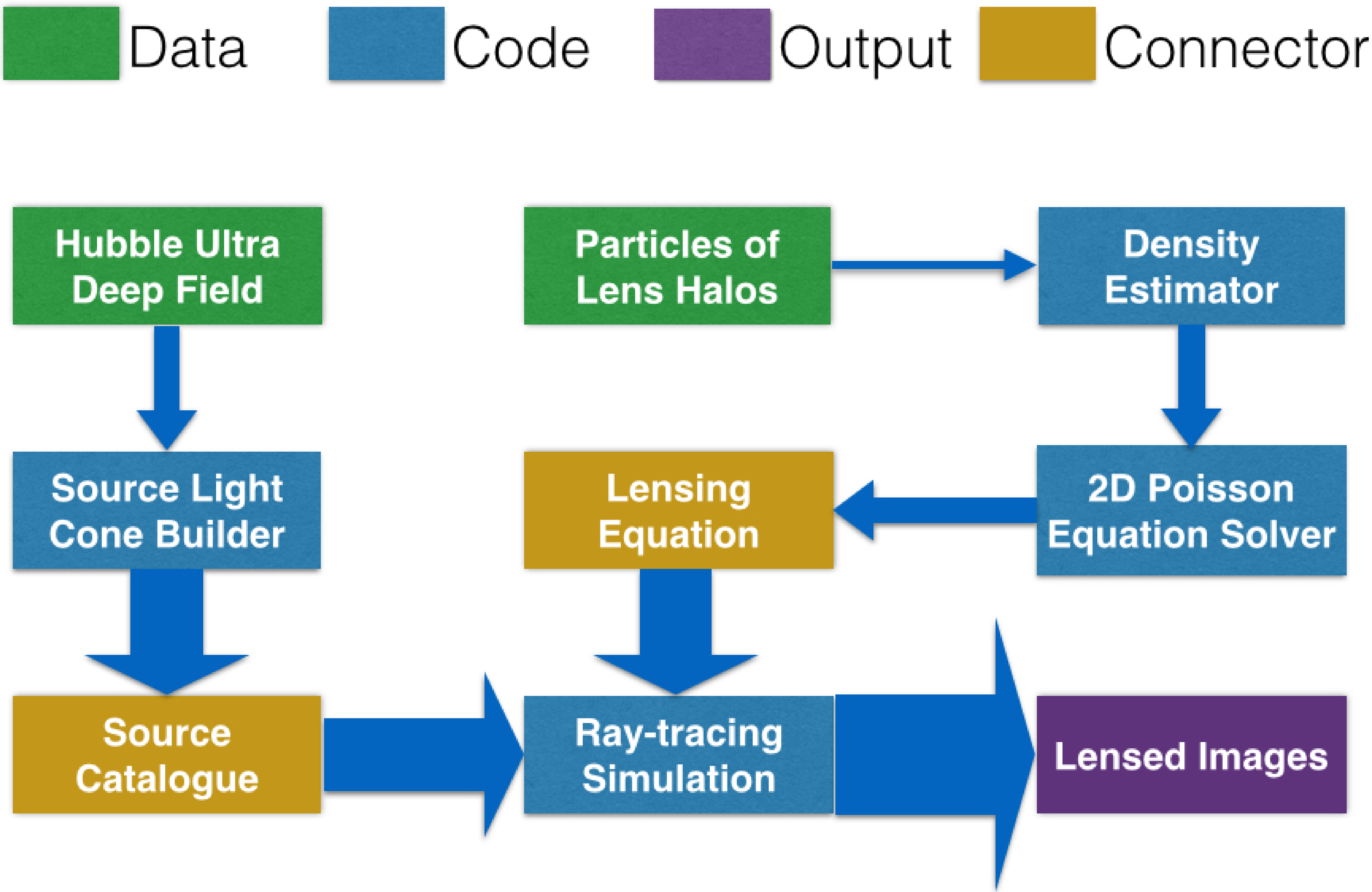
Simulated

Real

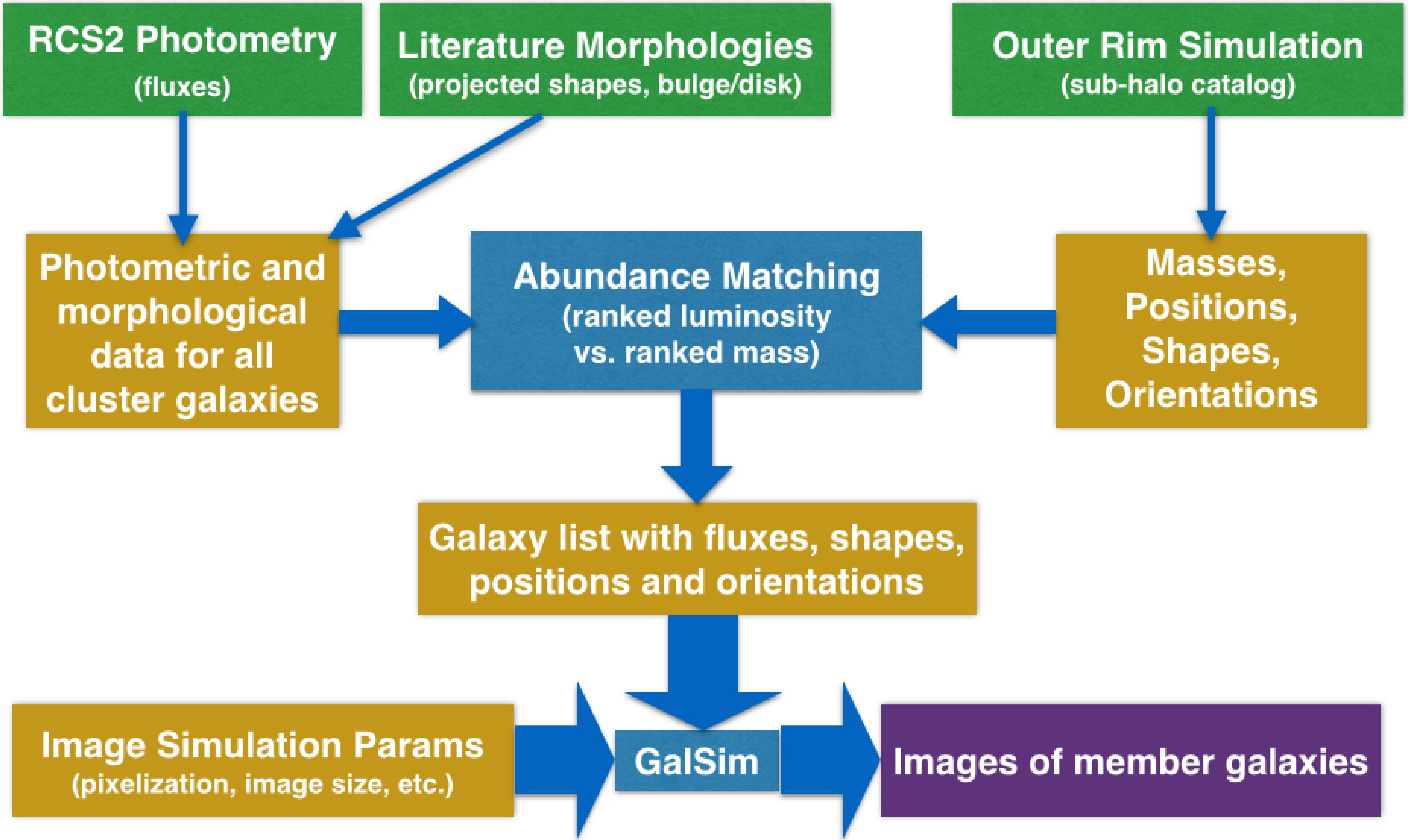
Simulated strong lens image to match SPT cluster observations taken with the MegaCAM camera on Magellan, in collaboration L. Bleem, M. Florian, S. Habib, M. Gladders, N. Li, S. Rangel

N. Li et al., in preparation

Producing Lensed Images



Producing Images of Member Galaxies



Producing Images of L.O.S. Galaxies