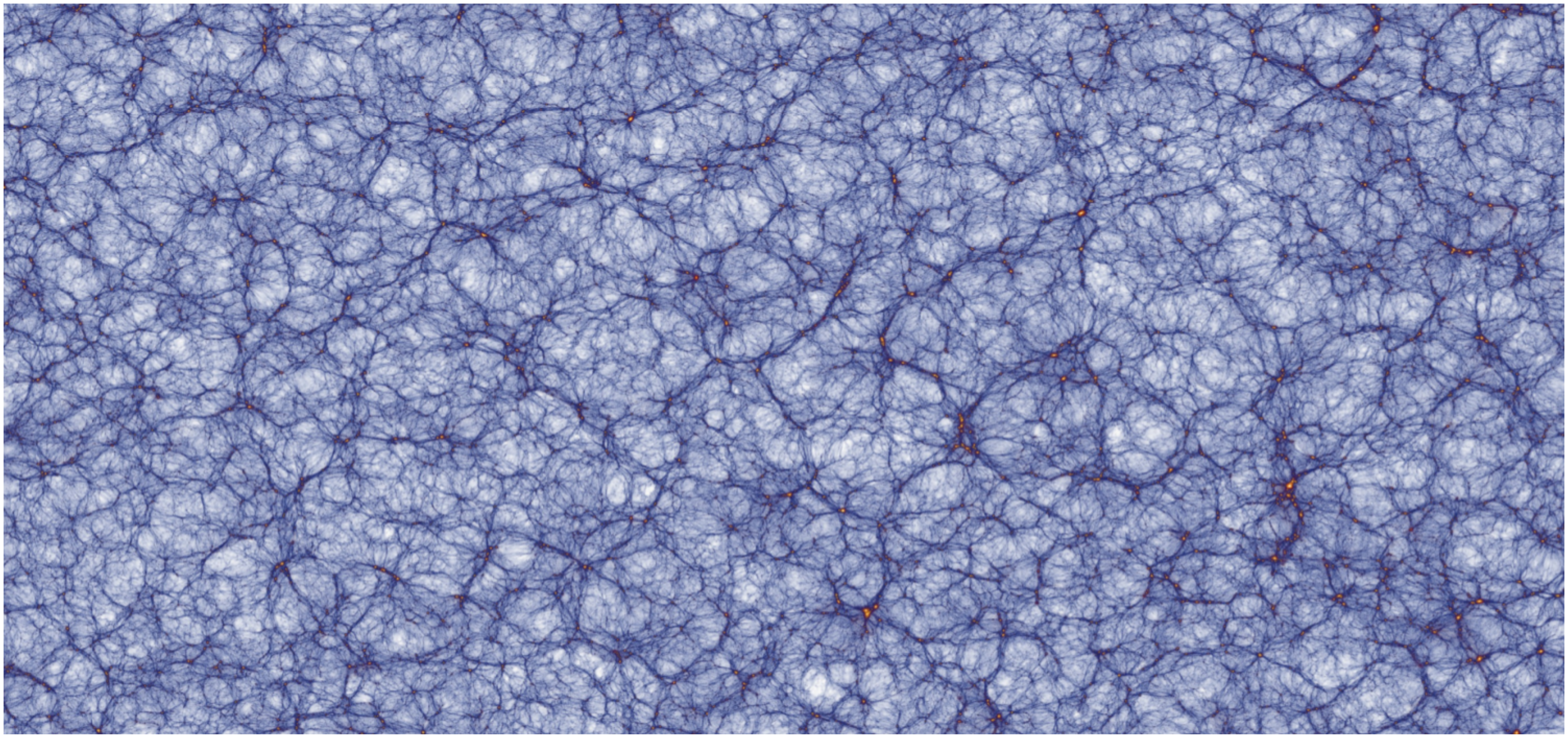
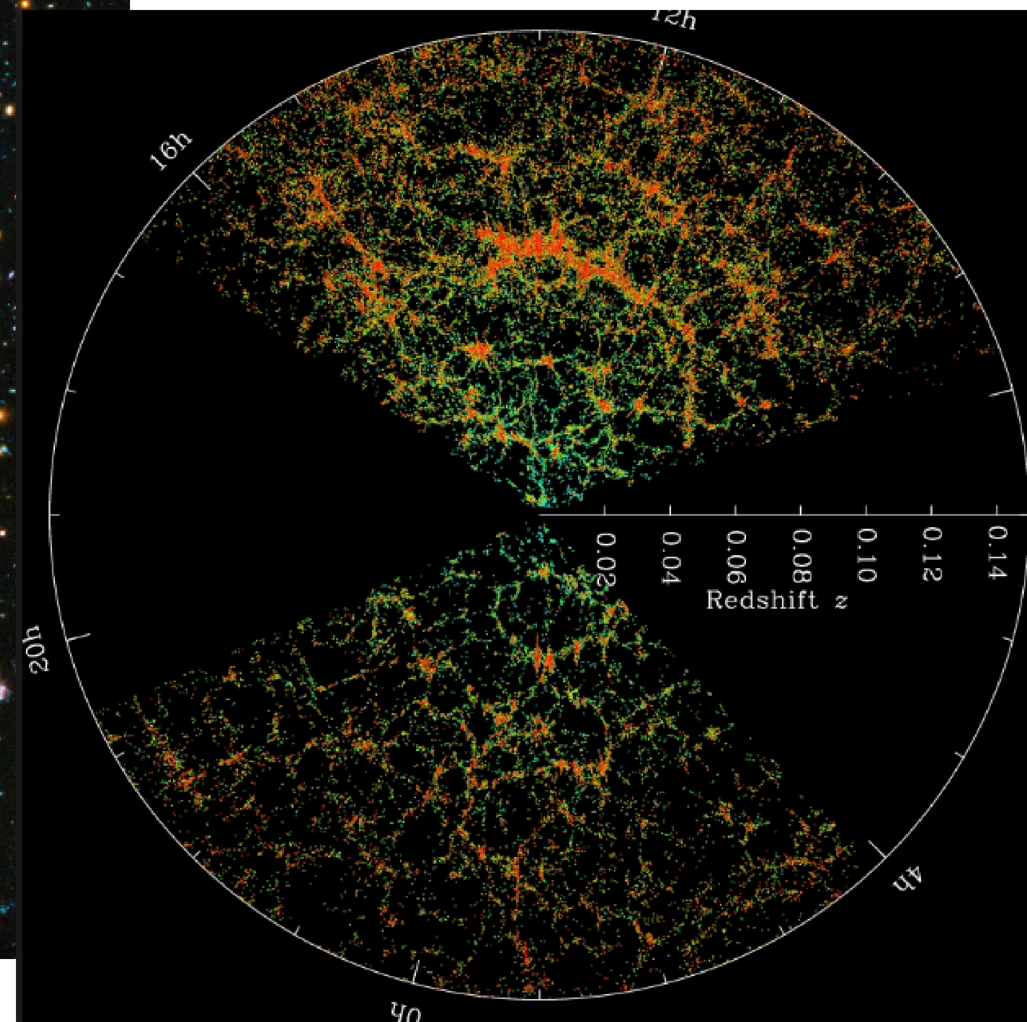
Simulating the formation of structure in the Universe

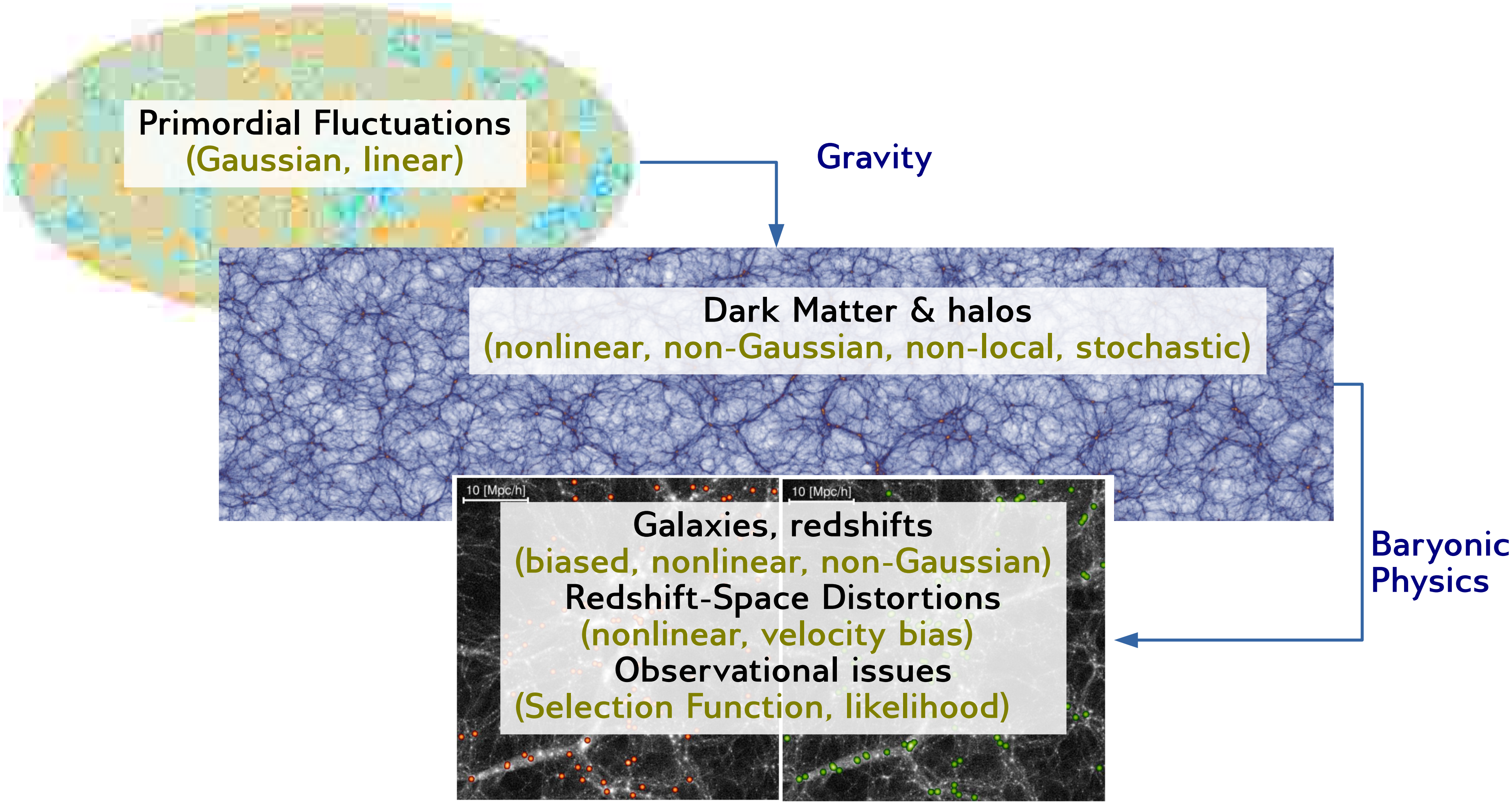


Raul E. Angulo

How can we explain the very diverse universe we observe and use it to infer fundamental physics



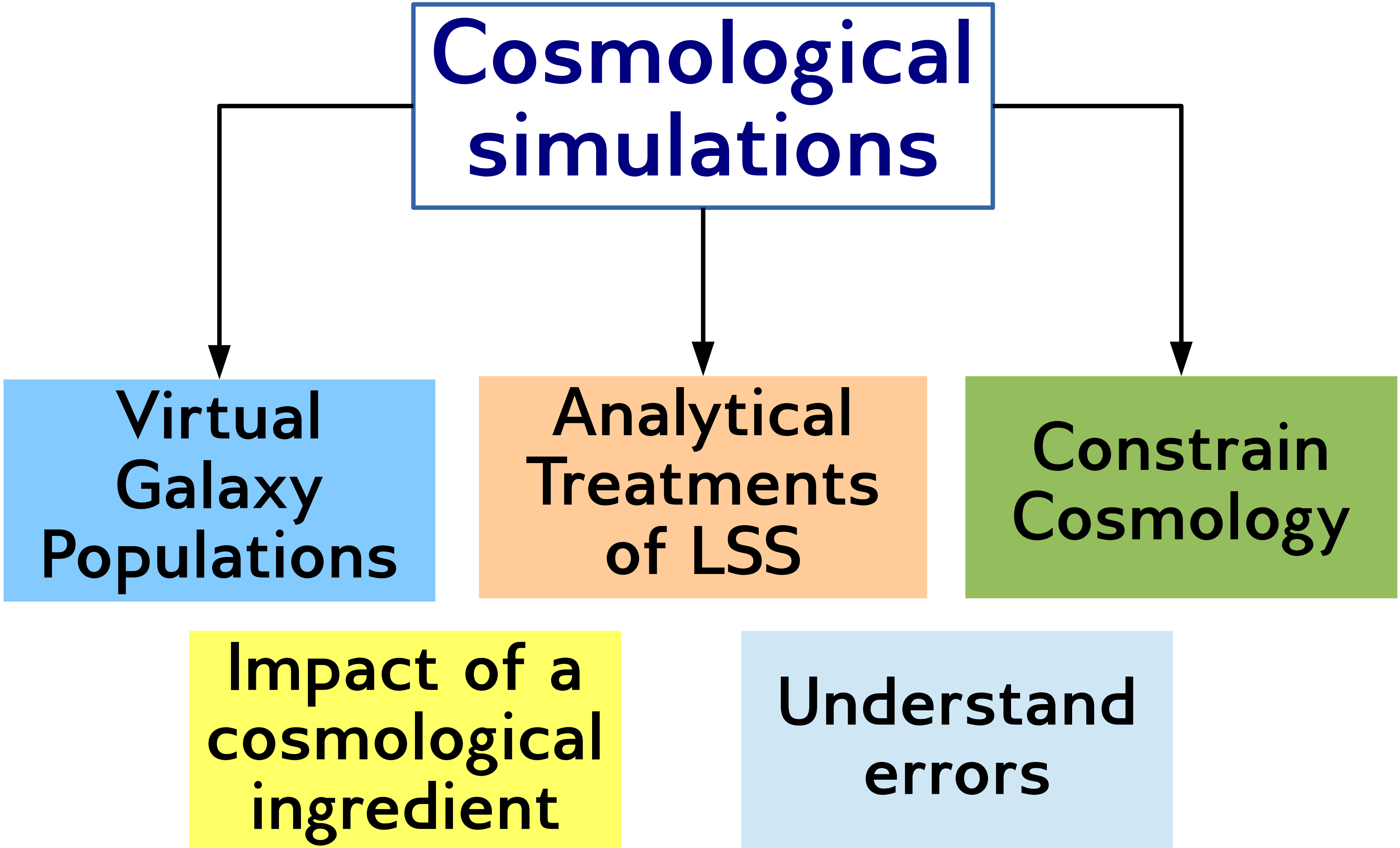
Explaining the structure in the Universe is a solvable (but very hard) problem.



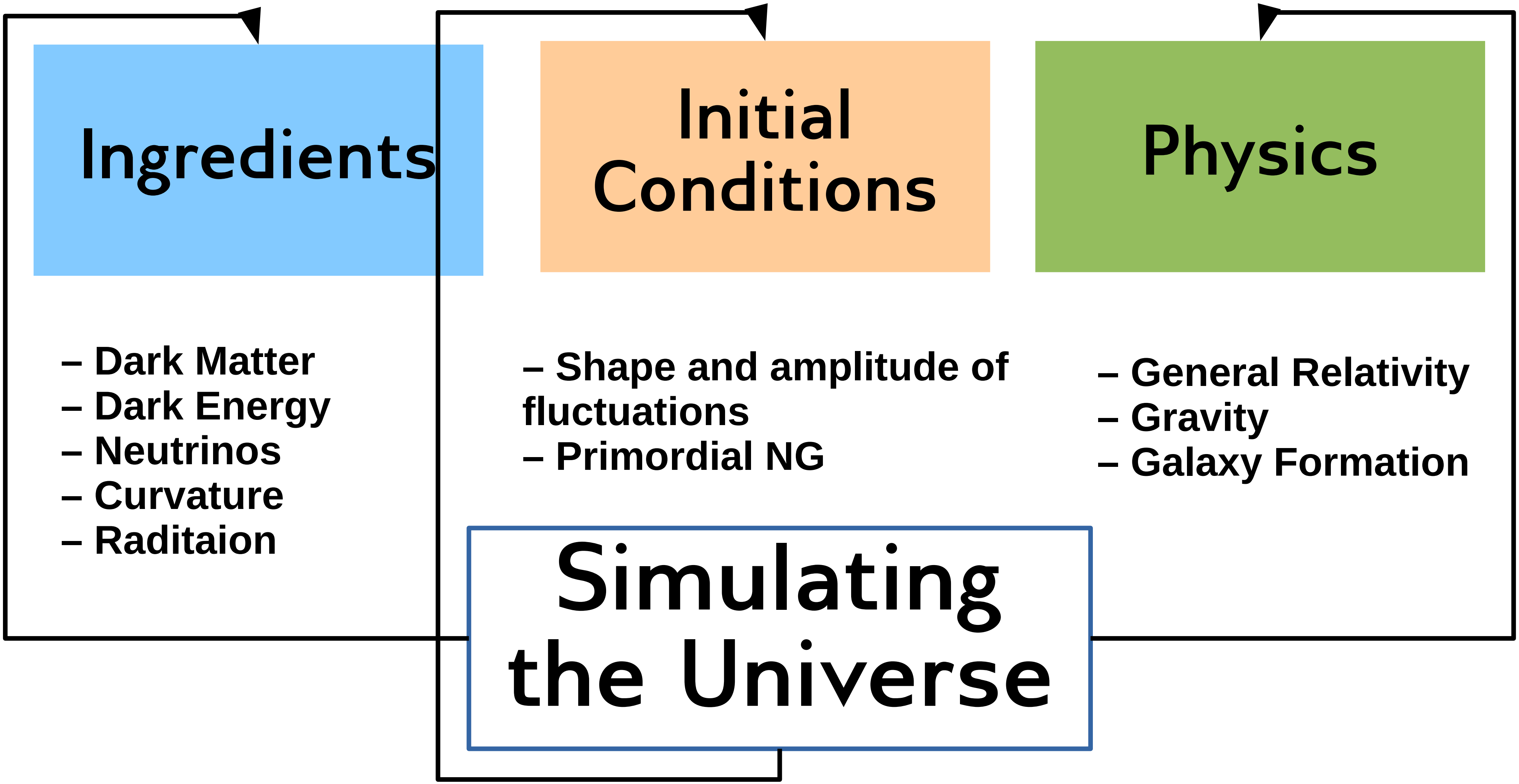
Numerical simulations are the most accurate way to bridge 13.7 billion years of nonlinear evolution

Simulations have been essential in the establishment of the "cosmology standard model"

Numerical simulations are an essential tool for precision cosmology



- → Background
- → Methods
- → Current State of the Art
- → The next decade
- → Open questions & challenges


Ingredients
Initial Conditions
Physics
– Dark Matter
– Dark Energy
– Neutrinos
– Curvature
– Raditaion
– Shape and amplitude of fluctuations
– Primordial NG
– General Relativity
– Gravity
– Galaxy Formation
Simulating the Universe

Assumptions made in the simplest case

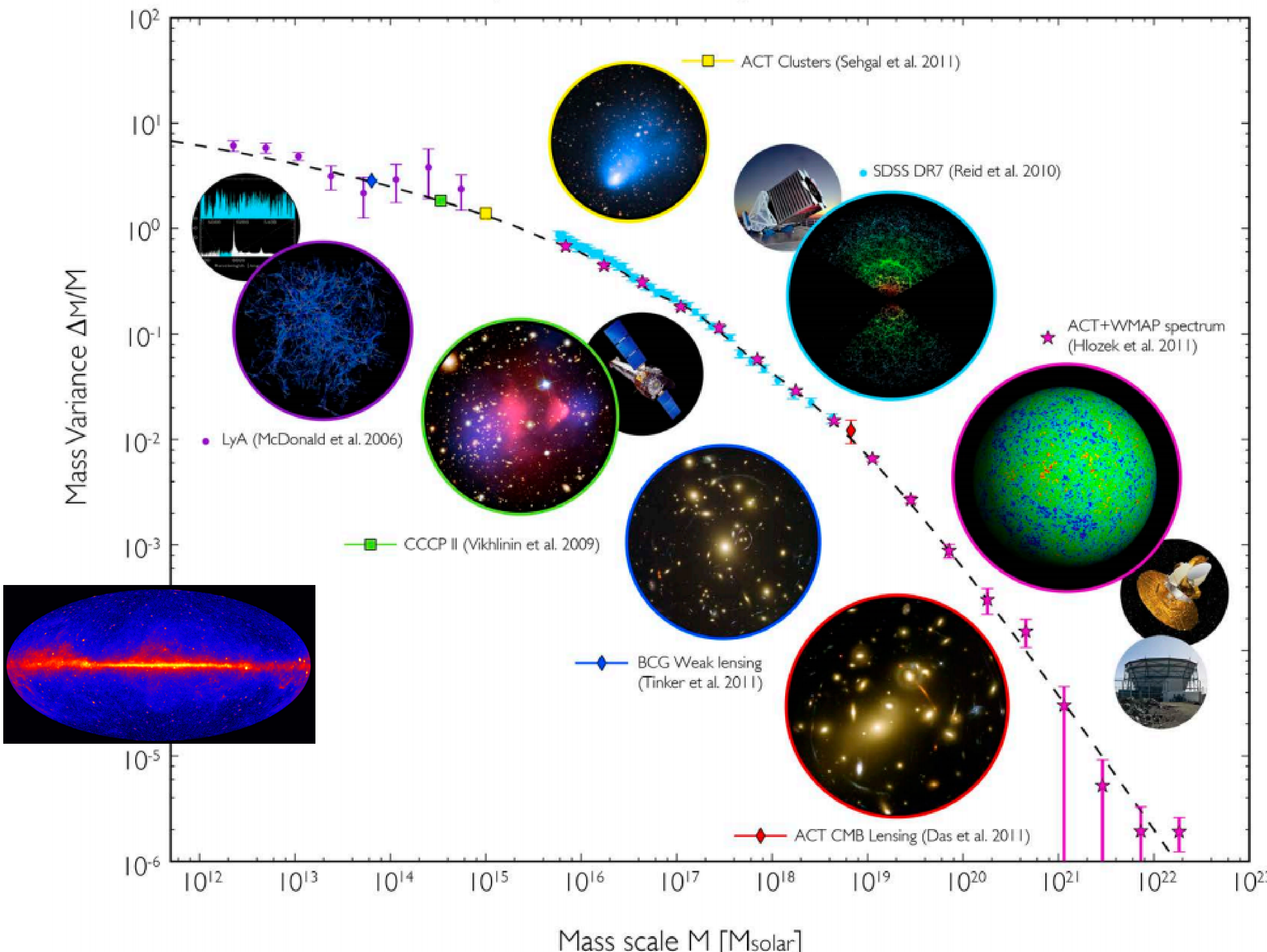
Usually referred to as dark-matter (gravity) only simulations

GR at the background level

Dark Matter as the main gravitating ingredient

Newtonian Gravity as the only force to consider

Evidence supporting the simplest case



Primack 2015

Our problem is reduced to simulating the evolution of a

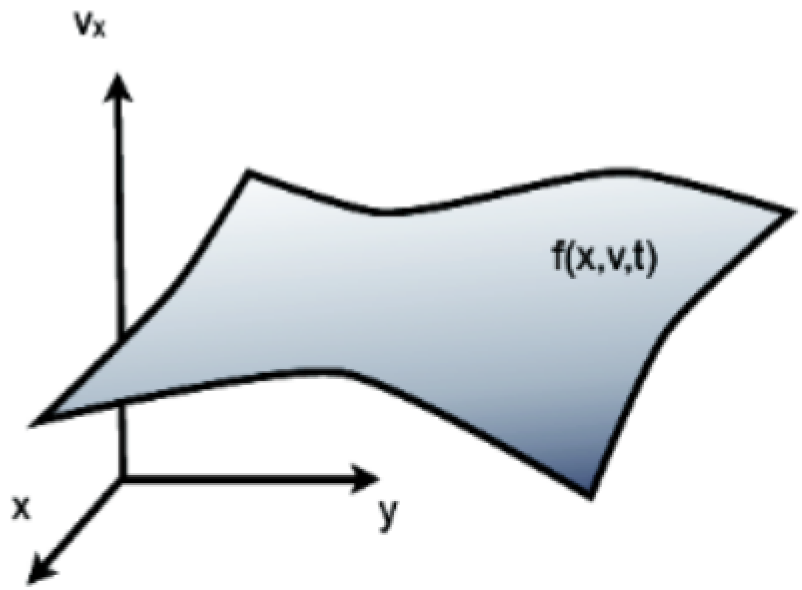
initially smooth,
cold,
with zero cross-section,
collisionless

fluid under the effect of self-gravity in an expanding Universe.

Simply solve newton's law for many googolplexian particles

Simulating structure formation in the Universe

Most of the mass in the Universe is in the form of an unknown elementary particle: the Cold Dark Matter



dark matter properties

- Cold
- Collisionless
- Gravity-only
- Smooth

...but simulating trillions of micro-physical CDM particles is impossible

CDM forms a "sheet": A continuous 3D surface embedded in a 6D space

The Vlassov-Poisson Equation

$$0 = \frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \frac{\mathbf{v}}{a^2} \cdot \frac{\partial f}{\partial \mathbf{v}} - \frac{\partial f}{\partial \mathbf{x}} \frac{\partial \Phi}{\partial \mathbf{x}}$$

$$\nabla^2 \Phi = \frac{4\pi G}{a} \int f \mathrm{d}^3 v.$$

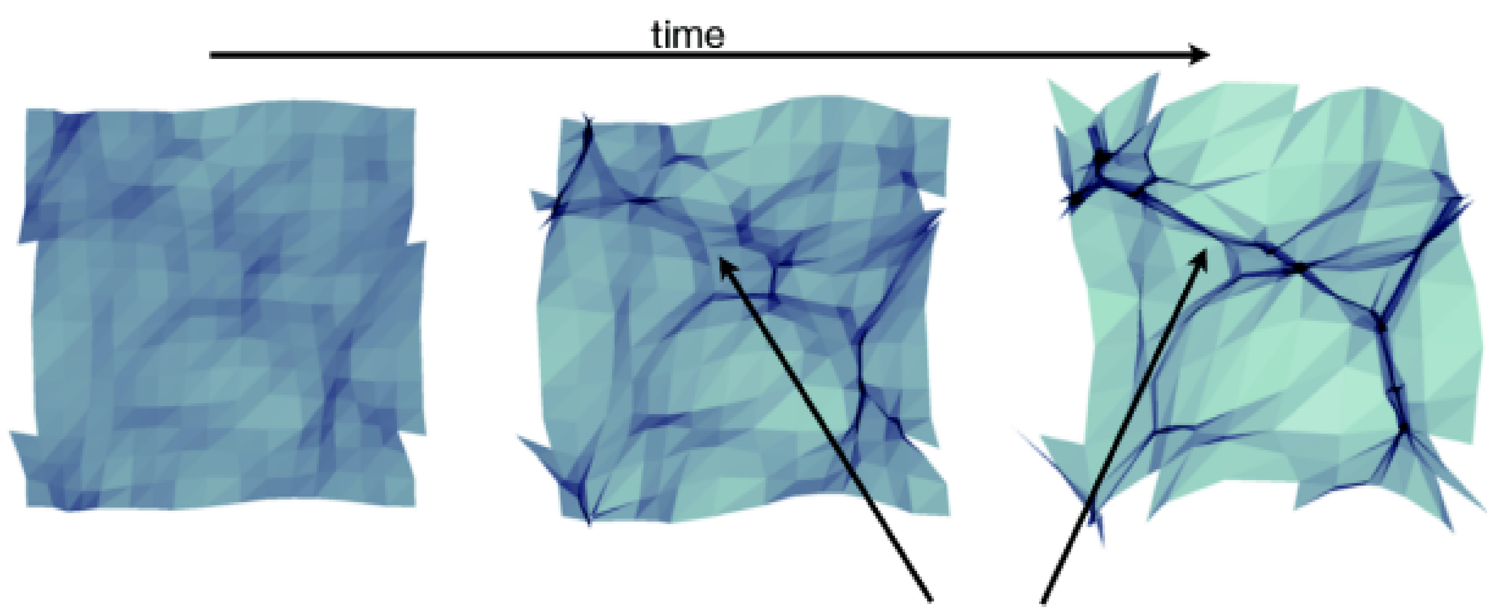
CDM Sheet Properties

→ phase-space is conserved along characteristics

→ It can never tear

→ It can never intersect

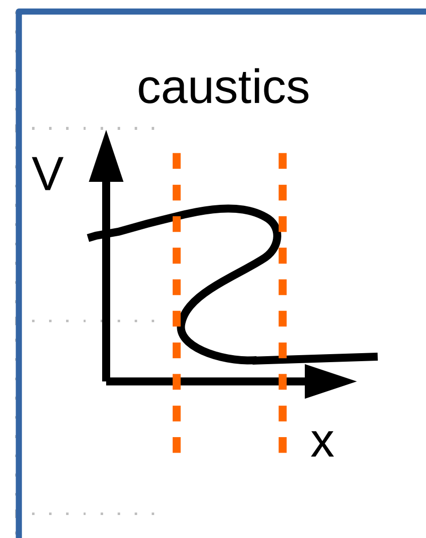
Credit: Oliver Hahn

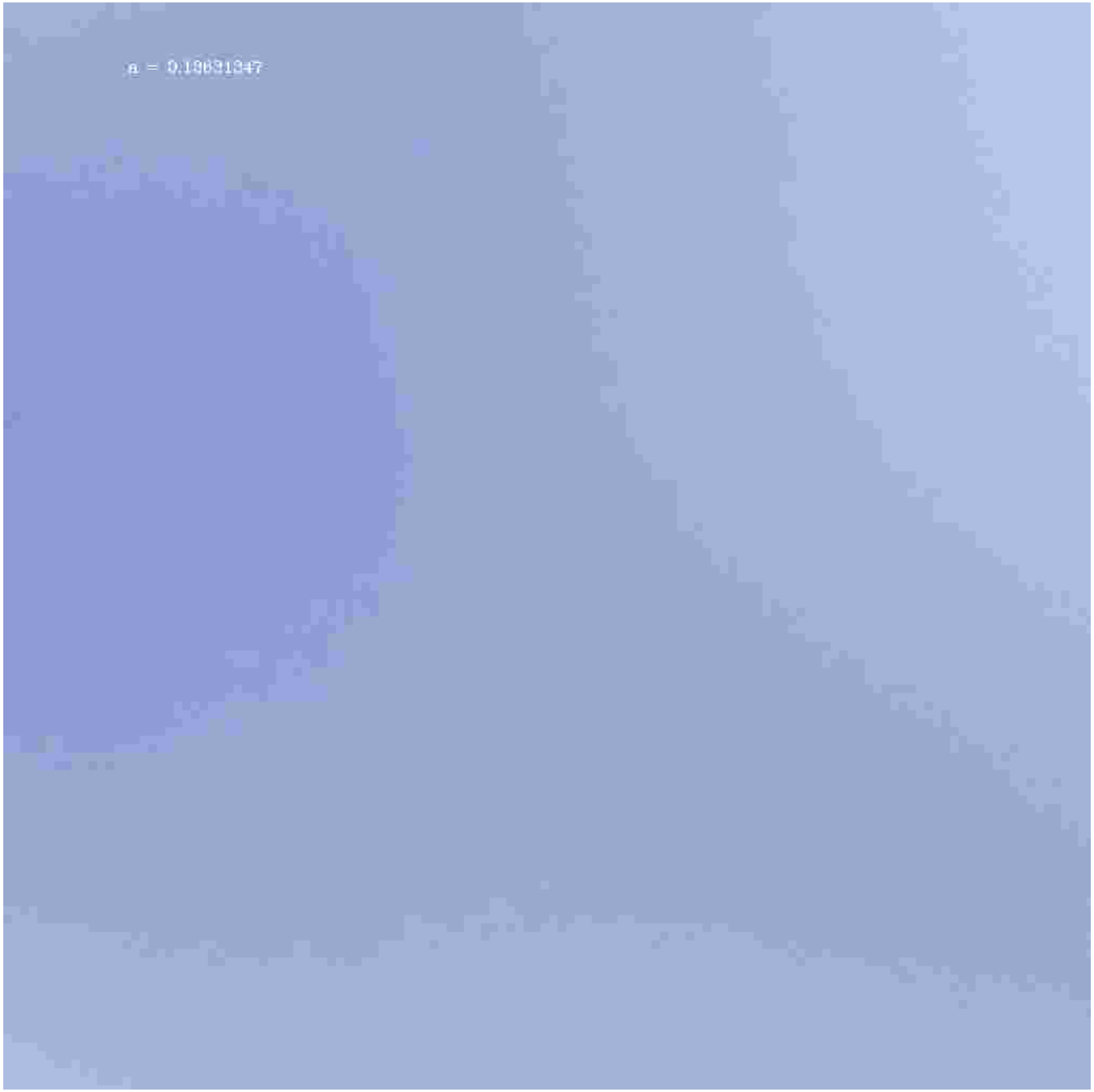


shell crossing

multi-stream regions appear

density, velocity = sum over many cells




a = 0.18631347

- → Background
- → **Methods**
- → Current State of the Art
- → The next decade
- → Open questions & challenges

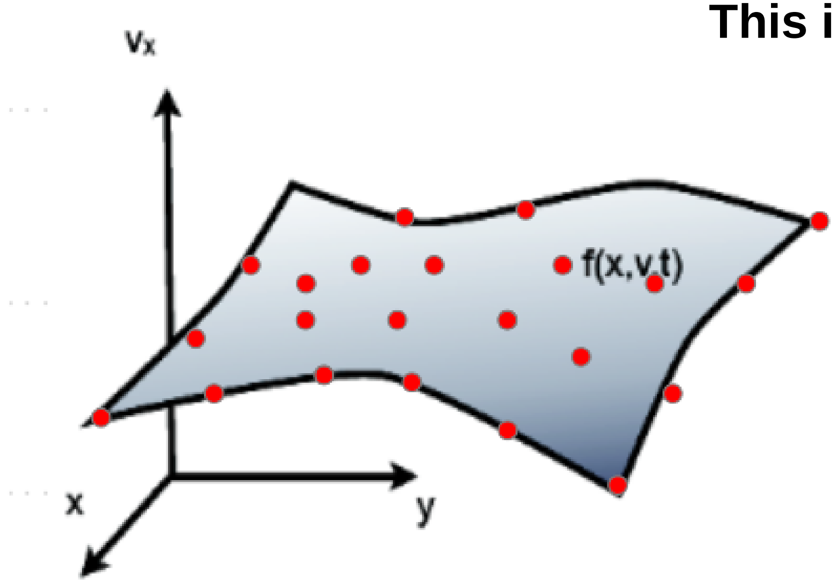
Solving Vlassov-Poisson via a Montecarlo sampling and coarse-graining

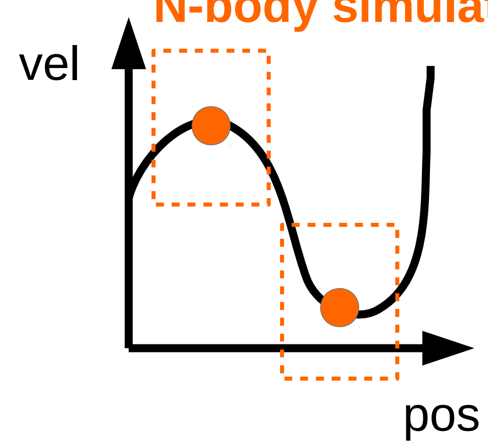
The "method of characteristics" is used to solve the partial differential equation that the VP is.

$$H(p,x) = \frac{p^2}{2m} + \Phi(x)$$

The solution yields the equation of motions of the Hamiltonian of classical mechanics

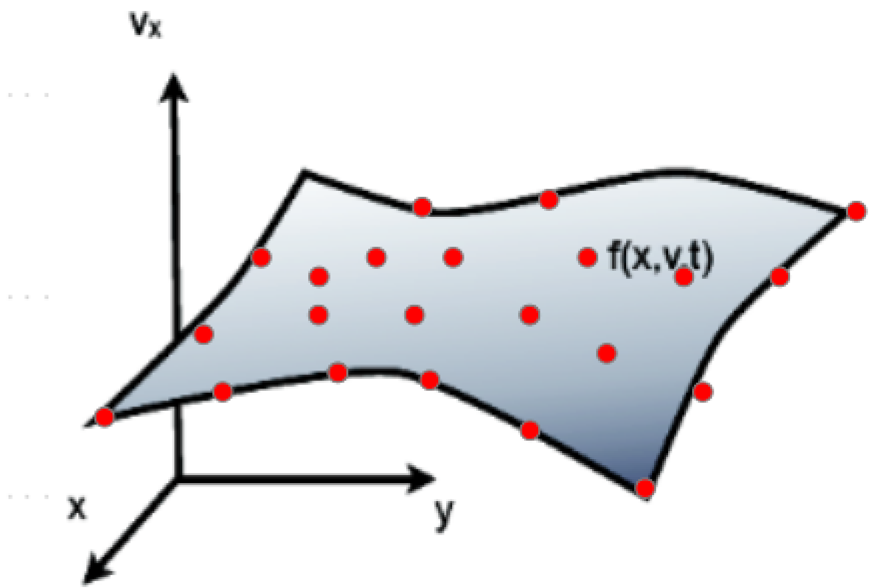
This is the correct solution as N goes to infinity





Standard approach to solving the VP equation:

Montecarlo Sampling and coarse graining the CDM distribution function



$$\frac{d^2\mathbf{x}_i}{d^2t} = \nabla_i \Phi(\mathbf{x}_i),$$

$$\Phi(\mathbf{x}) = -G \sum_i \frac{m_i}{[(\mathbf{x}_i - \mathbf{x})^2 + \varepsilon^2]}$$

An N-body code

→ Compute ICs

Loop over N timesteps

→ Kick velocities dt/2

→ Drift particles dt

→ Compute forces

→ Kick velocities dt/2

Leap frog symplectic integrator

The computational challenge

Modern cosmological simulations pose hard problems in terms of execution time, RAM consumption, and data handling

CPU and load Imbalances

Quadrillion force calculations with large anisotropies
and very different dynamical timescales

RAM

Above hundreds of Tb of RAM necessary to hold basic information
Additional requirements memory imbalances and data analyses

I/O & Disk Space

Data products can be en excess of dozens of Petabytes.

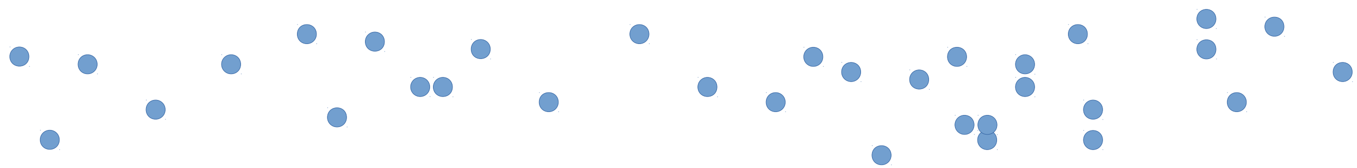
We require a combination of extremely efficient and scalable algorithms, and a large Supercomputer!

MXXL

- L = 3000 Mpc/h
- N = 6720^3 particles
- Є = 10 kpc/h
- M = 6.18×10^9 Msun/h

Force calculation

The problem is to estimate the gravitational interaction of
A set of N discrete particles

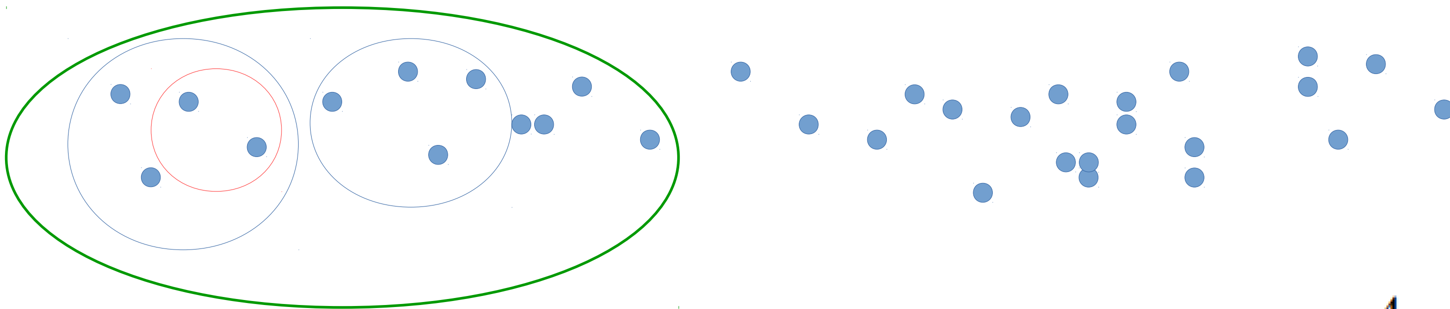


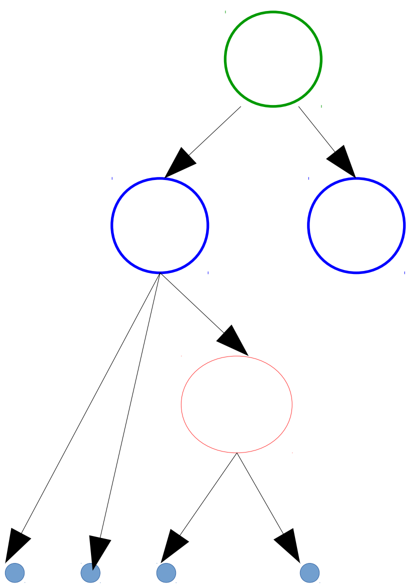
$$\phi(\mathbf{x}_i) = -\sum_{j=1\ldots N} \frac{4\pi G}{a|\mathbf{x}_i - \mathbf{x}_j|}$$

For each particle, we need the add up the contribution of
N-1 particles. Thus, this is a NxN problem!

Force calculation

The problem is to estimate the gravitational interaction of
A set of N discrete particles





$$\phi(\mathbf{x}_i) = - \sum_{j=1\ldots N} \frac{4\pi G}{a|\mathbf{x}_i - \mathbf{x}_j|}$$

$$\frac{1}{|x - x_j|} = \frac{1}{|(x-\lambda) - (x_j - \lambda)|}$$

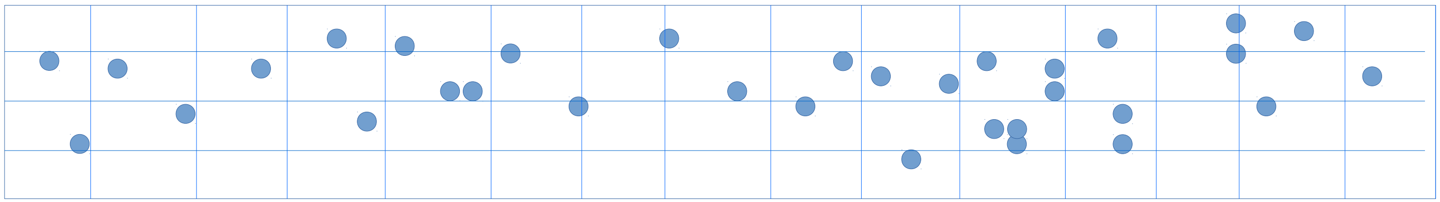
$$\frac{1}{|\mathbf{y} + \boldsymbol{\lambda} - \mathbf{x_j}|} \simeq \frac{1}{|y|} - \mathbf{y} \cdot \frac{\boldsymbol{\lambda} - \mathbf{x}_j}{|\mathbf{y}|^3} + \ldots$$

monopole dipole

The decision to open a node is given by a desired accuracy.
The efficiency depends on the clustering but ~ N log(N), allows
Individual timesteps, good load/cpu balances.

Force calculation

The problem is to estimate the gravitational interaction of
A set of N discrete particles



$$\nabla^2\phi = \frac{4\pi G}{a}(\rho - \bar{\rho})$$

Interpolation Methods

1) Nearest Grid Point (0^{th} order)

2) Clouds-in-Cells (1^{st} order)

3) Triangular shaped cloud (2^{nd} order)

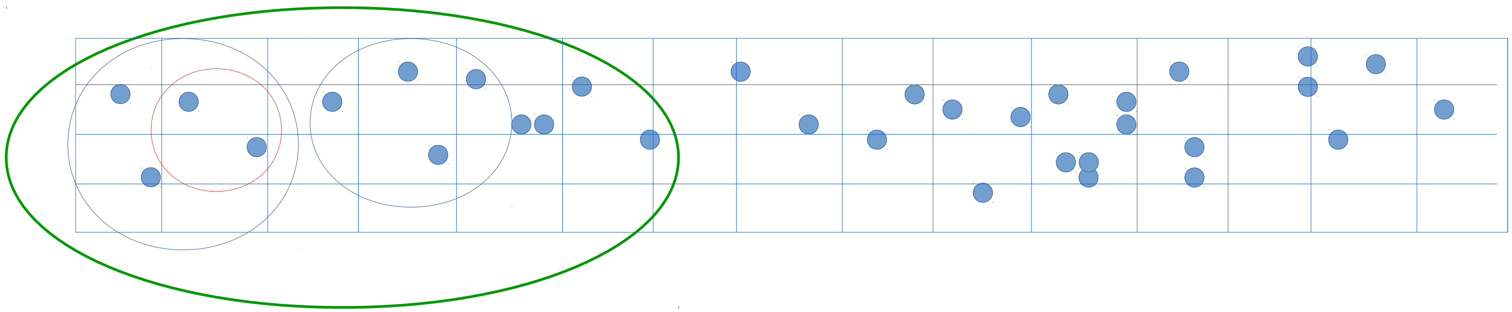
$$\nabla^2\phi \propto \delta \quad \Leftrightarrow \quad \tilde{\phi} \propto -\tilde{\delta}/k^2$$

$$\mathbf{a} = -\boldsymbol{\nabla}\phi \quad \Leftrightarrow \quad \tilde{\mathbf{a}} \propto -\frac{i\mathbf{k}}{k^2}\tilde{\delta}$$

Fast, easy to parallelise, portable FFT libraries, scales as N;
but bad load balance, limited spatial resolution, global timesteps

Force calculation

The problem is to estimate the gravitational interaction of
A set of N discrete particles



Alternatives

1) Adaptive Mesh refinement

2) Ewald summation for trees

3) Direct Summation

4) Fast Multipole methods

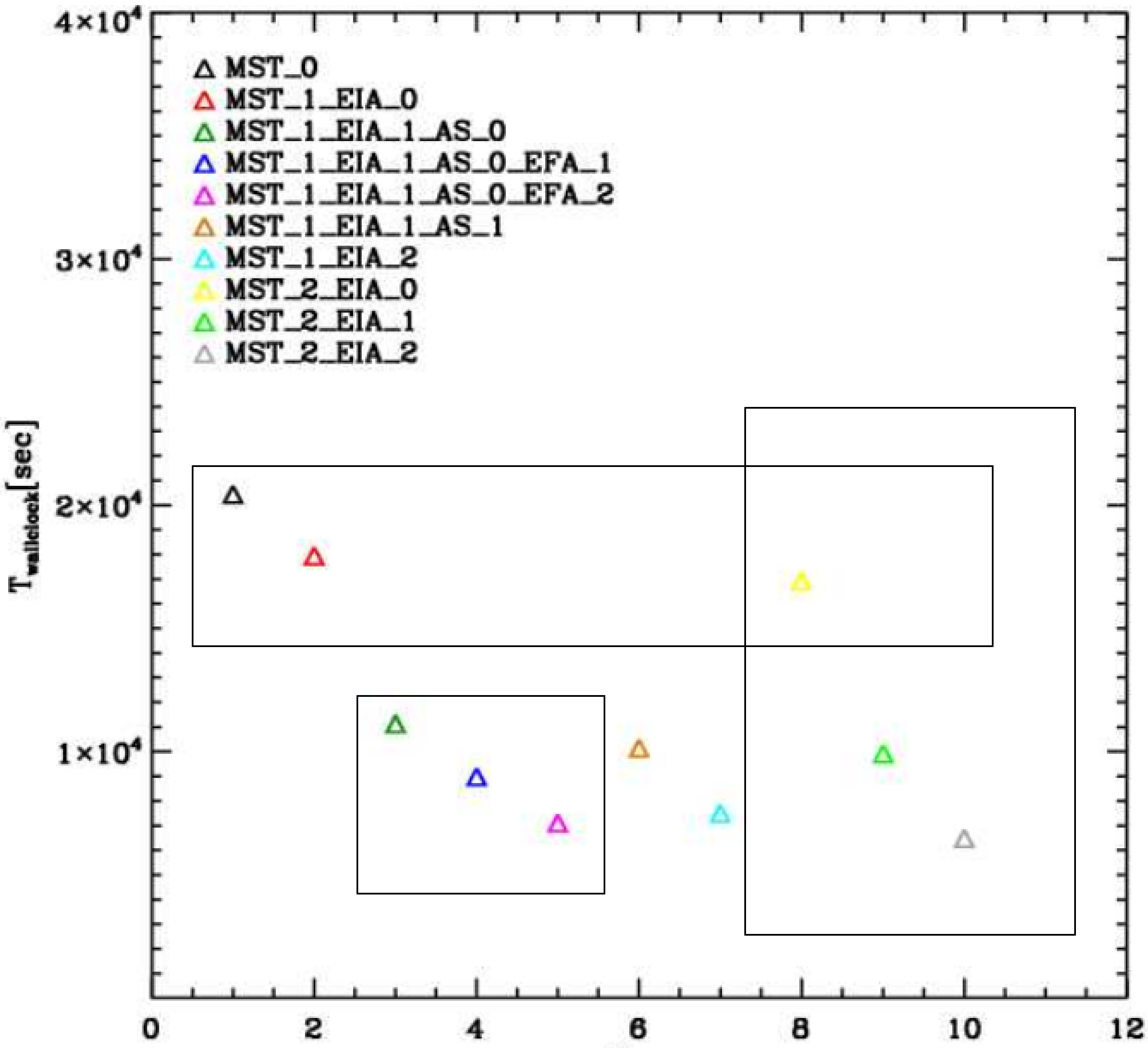
$$\phi_{\boldsymbol{k}} = \phi_{\boldsymbol{k}}^{\text{long}} + \phi_{\boldsymbol{k}}^{\text{short}}$$

$$\phi^{\text{short}}(\boldsymbol{x}) = -G \sum_i \frac{m_i}{r_i} \operatorname{erfc}\left(\frac{r_i}{2r_s}\right)$$

$$\phi_{\boldsymbol{k}}^{\text{long}} = \phi_{\boldsymbol{k}} \exp(-\boldsymbol{k}^2 r_s^2)$$

Number of particles is not precision

Force and time integration parameters can change execution times by factors of a few



MST: Maximum allowed timestep

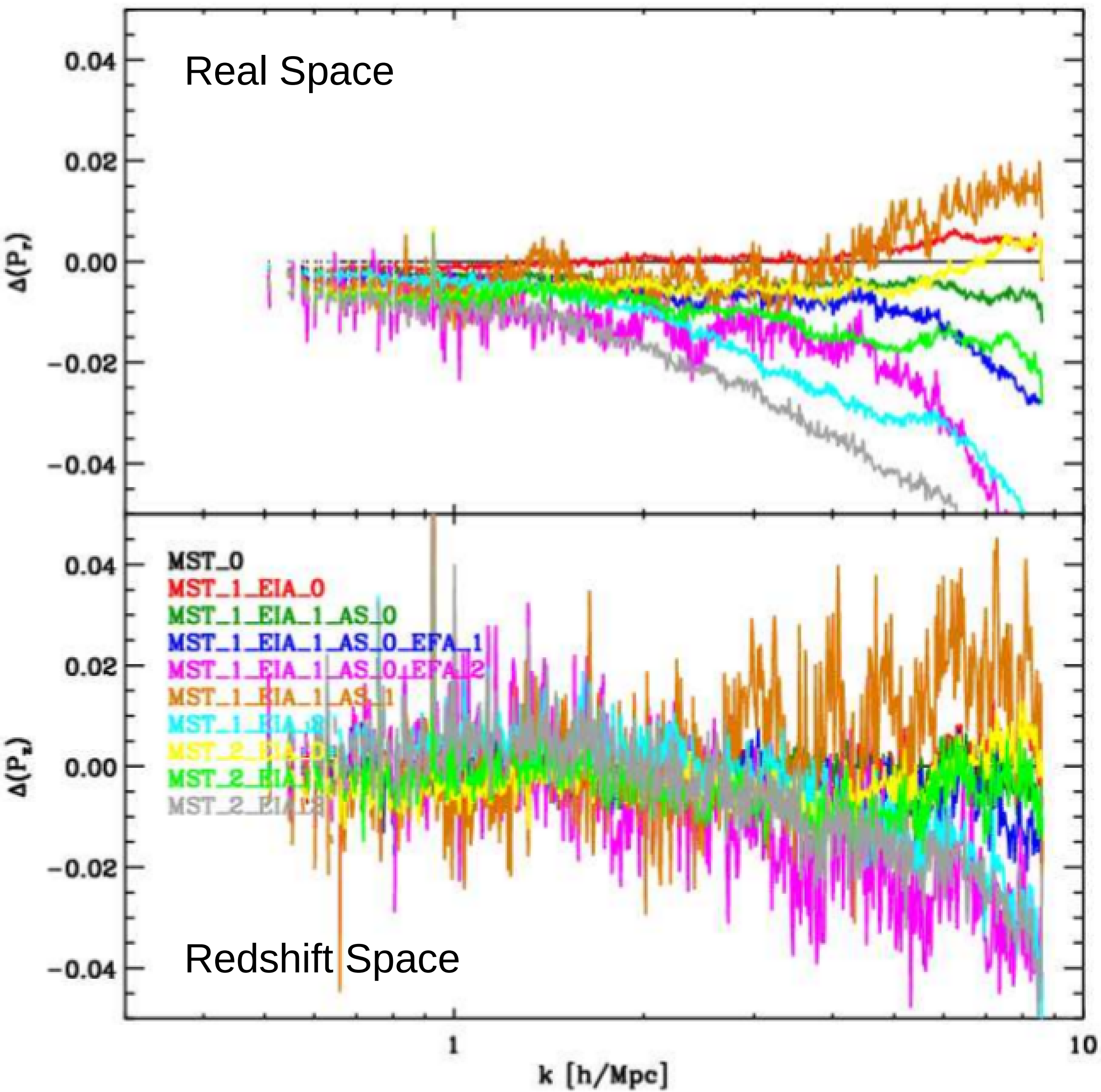
EIA: Error in time Integration

AS: Smoothing applied to mesh force

EFA: Error in Force Calculation

Number of particles is not precision

Errors on the power spectra induced by numerical errors



MST: Maximum allowed timestep

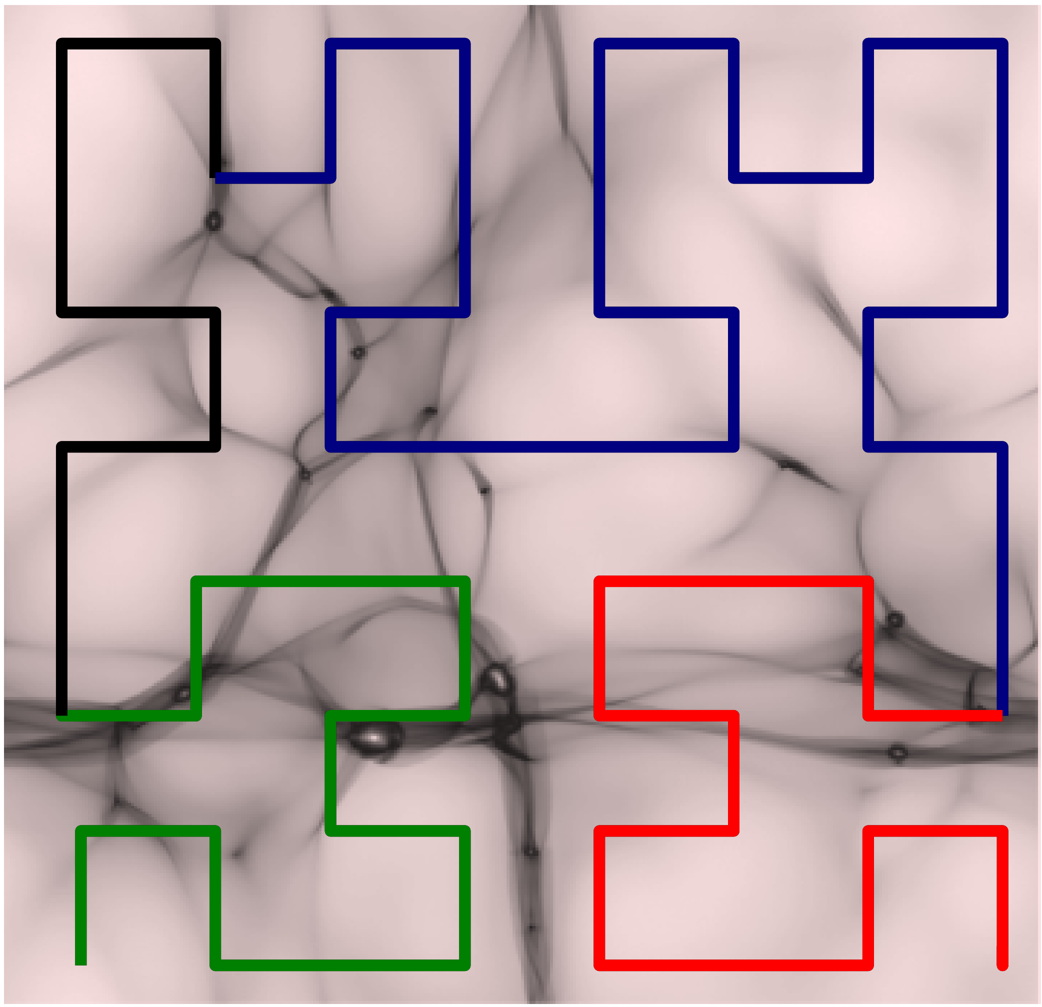
EIA: Error in time Integration

AS: Smoothing applied to mesh force

EFA: Error in Force Calculation

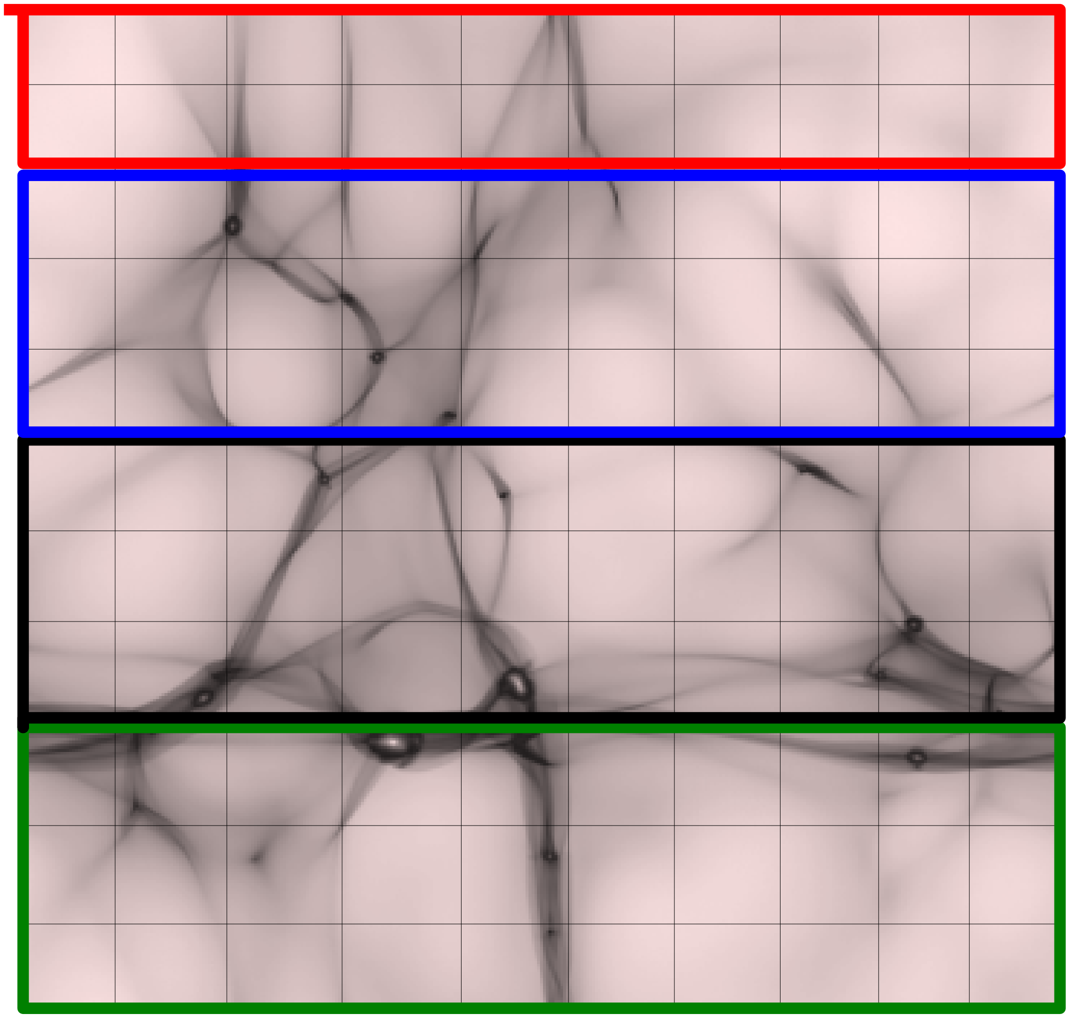
Not including:
1) Starting redshift
2) Transients from the ICs
3) Softening Length

Computational domain decomposition

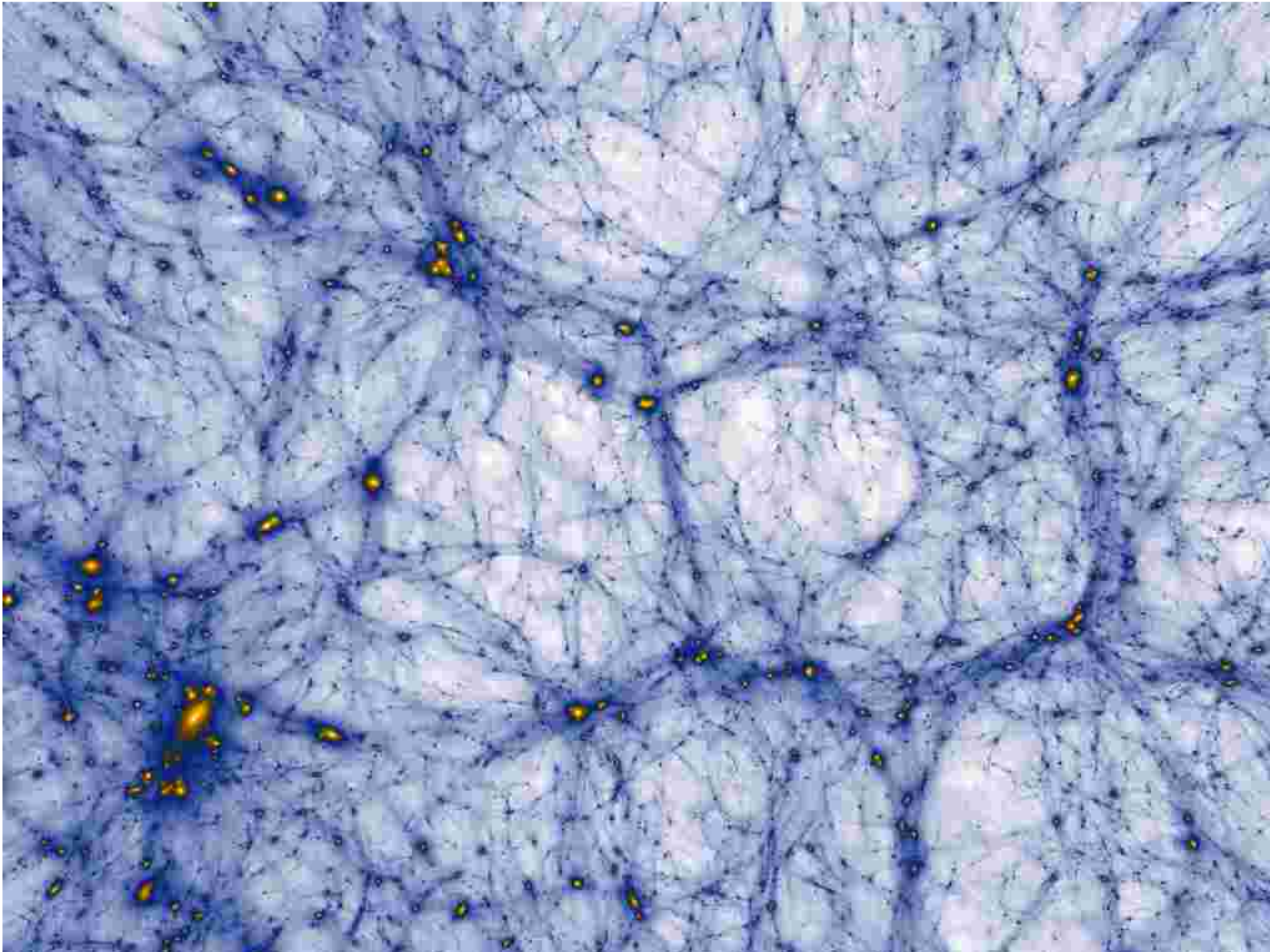


MPI Task #1
MPI Task #2
MPI Task #3
MPI Task #4

Force Calculation



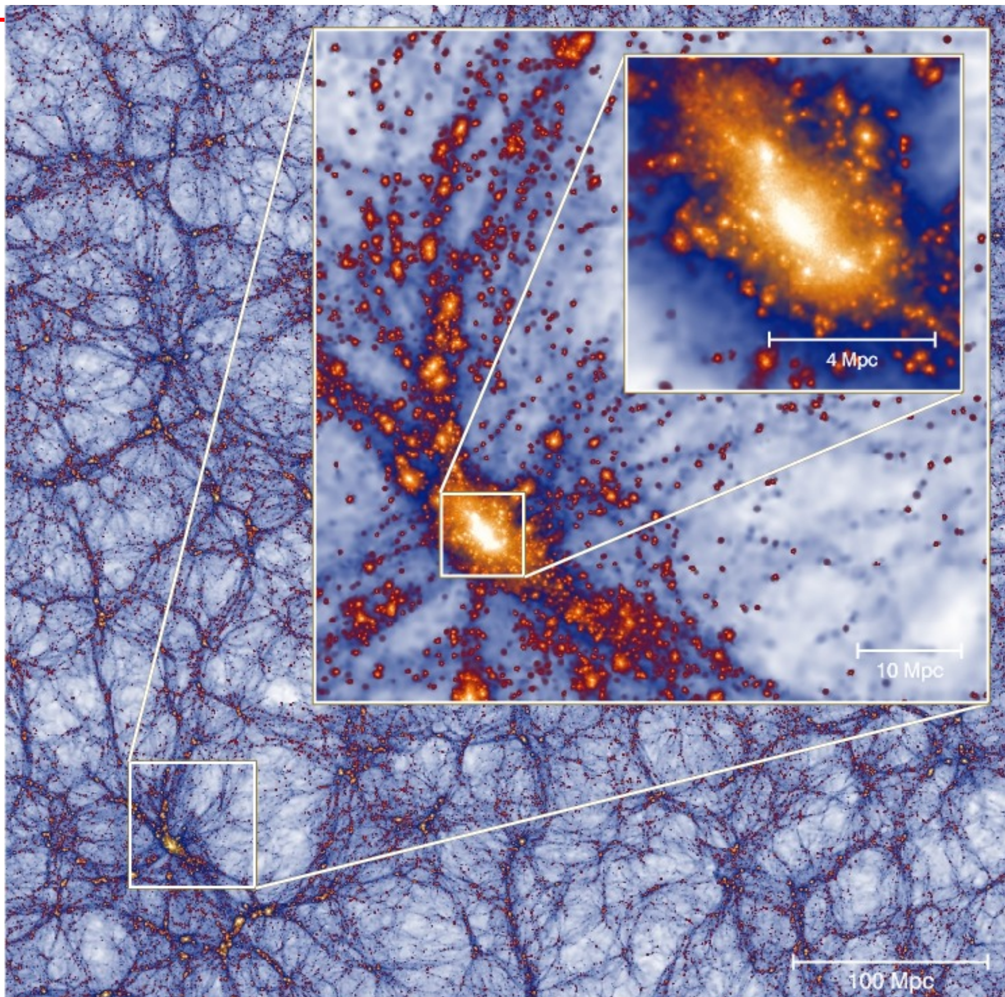
MPI Task #1
MPI Task #2
MPI Task #3
MPI Task #4



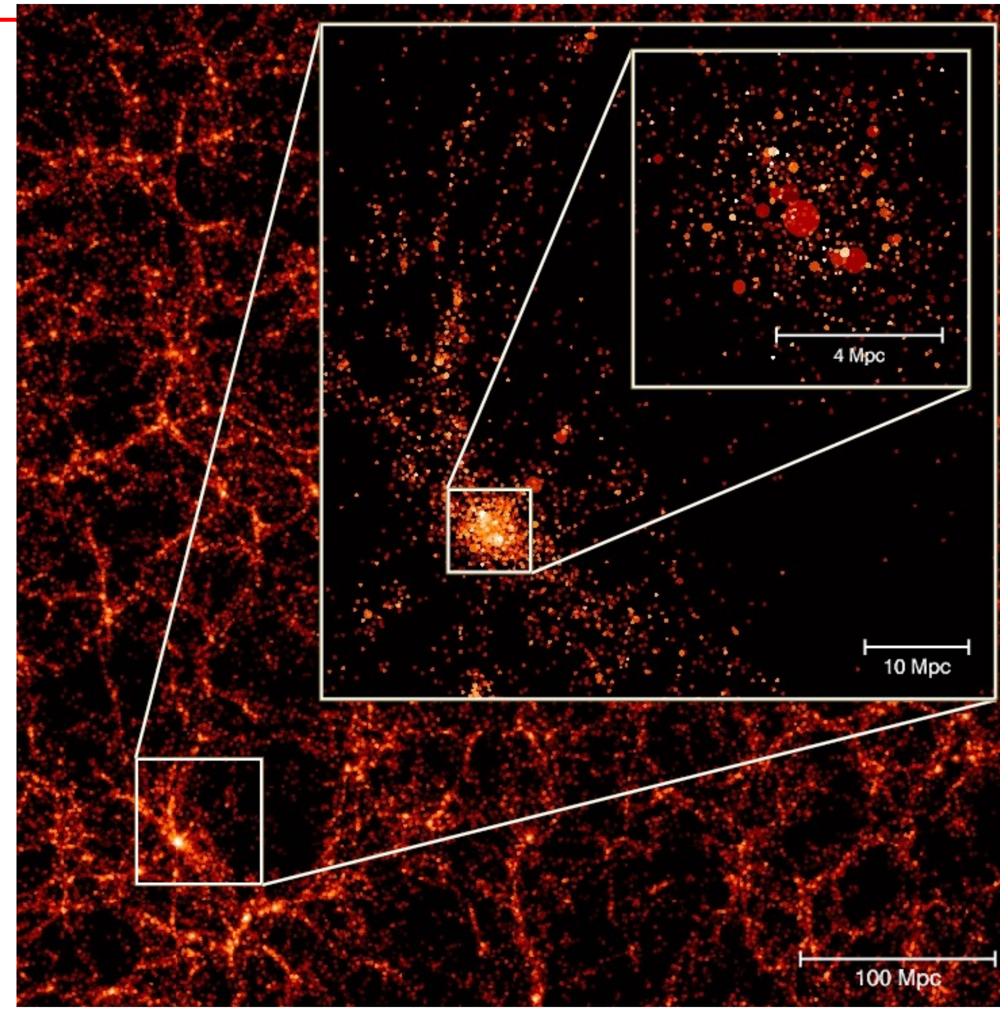
Dark Matter and galaxies

A realistic galaxy formation modelling on a 3Gpc/h simulated box at z=0

DARK MATTER

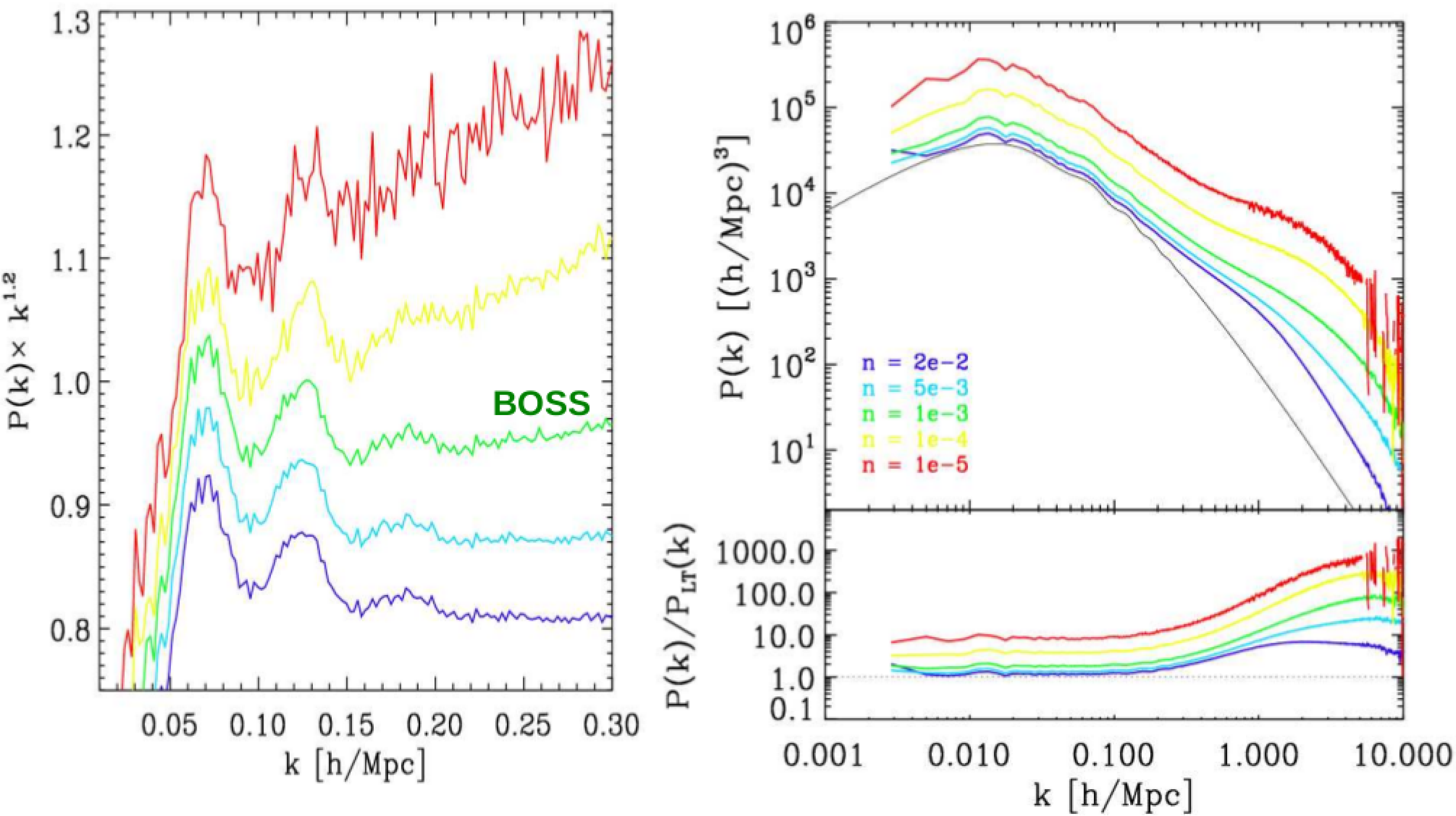


GALAXIES



Different galaxy catalogues in the MXXL simulation trace the BAO features with a scale-dependent bias

POWER SPECTRA OF THE GALAXY DISTRIBUTION AT Z=0 FOR DIFFERENT SPACE DENSITIES



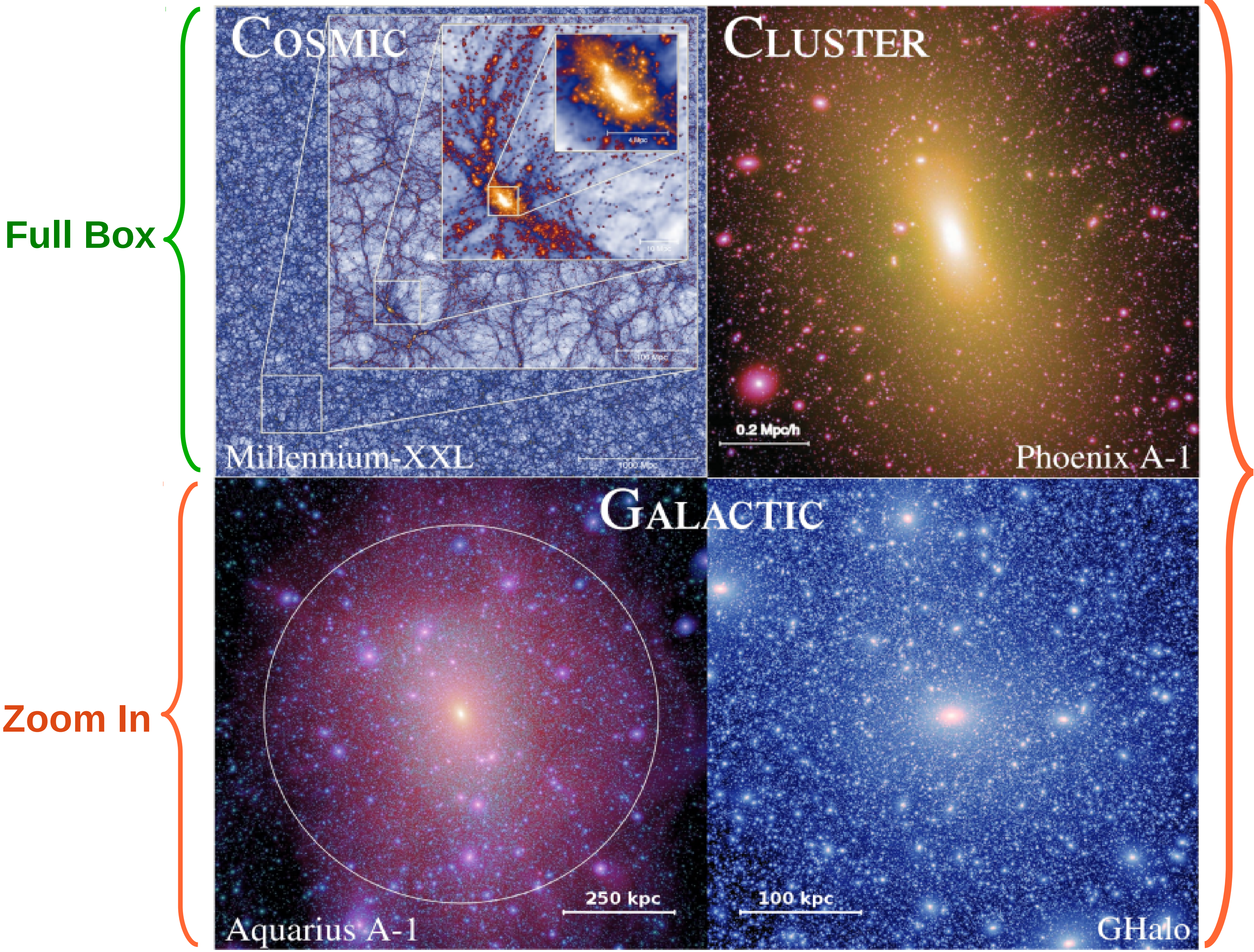
→ Background

→ Methods

→ **Current State of the Art**

→ The next decade

→ Open questions & challenges



Full Box
Zoom In
Cosmic
Cluster
Galactic
Millennium-XXL
Phoenix A-1
0.2 Mpc/h
Aquarius A-1
250 kpc
100 kpc
GHalo

DM-only simulations

> 1 trillion particles

Cosmic

Name	Code	L_{box} $[h^{-1}\mathrm{Mpc}]$	N_p $[10^9]$	m_p $[h^{-1}\,M_\odot]$	ϵ_{soft} $[h^{-1}\mathrm{kpc}]$	$N_{halo}^{>100p}$ $[10^6]$	ref.
DEUS FUR	Ramses-Deus	21000	550	1.2×10^{12}	$40.0^\dagger$	145	[259]
Horizon Run 3	GotPM	10815	370	2.5×10^{11}	150.0	~190	[260]
Millennium-XXL	Gadget-3	3000	300	6.2×10^{9}	10.0	170	[220]
Horizon-4Π	Ramses	2000	69	7.8×10^{9}	$7.6^\dagger$	~40	[261]
Millennium	Gadget-2	500	10	8.6×10^{8}	5.0	4.5	[181]
Millennium-II	Gadget-3	100	10	6.9×10^{6}	1.0	2.3	[87]
MultiDark Run1	Art	1000	8.6	8.7×10^{9}	$7.6^\dagger$	3.3	[36]
Bolshoi	Art	250	8.6	1.4×10^{8}	$1.0^\dagger$	2.4	[262]

† For AMR simulations (Ramses, Art) ϵ_{soft} refers to the highest resolution cell width.

> 1 billion particles

Cluster

Name	Code	L_{hires} $[h^{-1}\mathrm{Mpc}]$	$N_{p,hires}$ $[10^9]$	$m_{p,hires}$ $[h^{-1}\,M_\odot]$	ϵ_{soft} $[h^{-1}\mathrm{kpc}]$	$N_{sub}^{>100p}$ $[10^3]$	ref.
Phoenix A-1	Gadget-3	41.2	4.1	6.4×10^{5}	0.15	60	[263]

Galactic

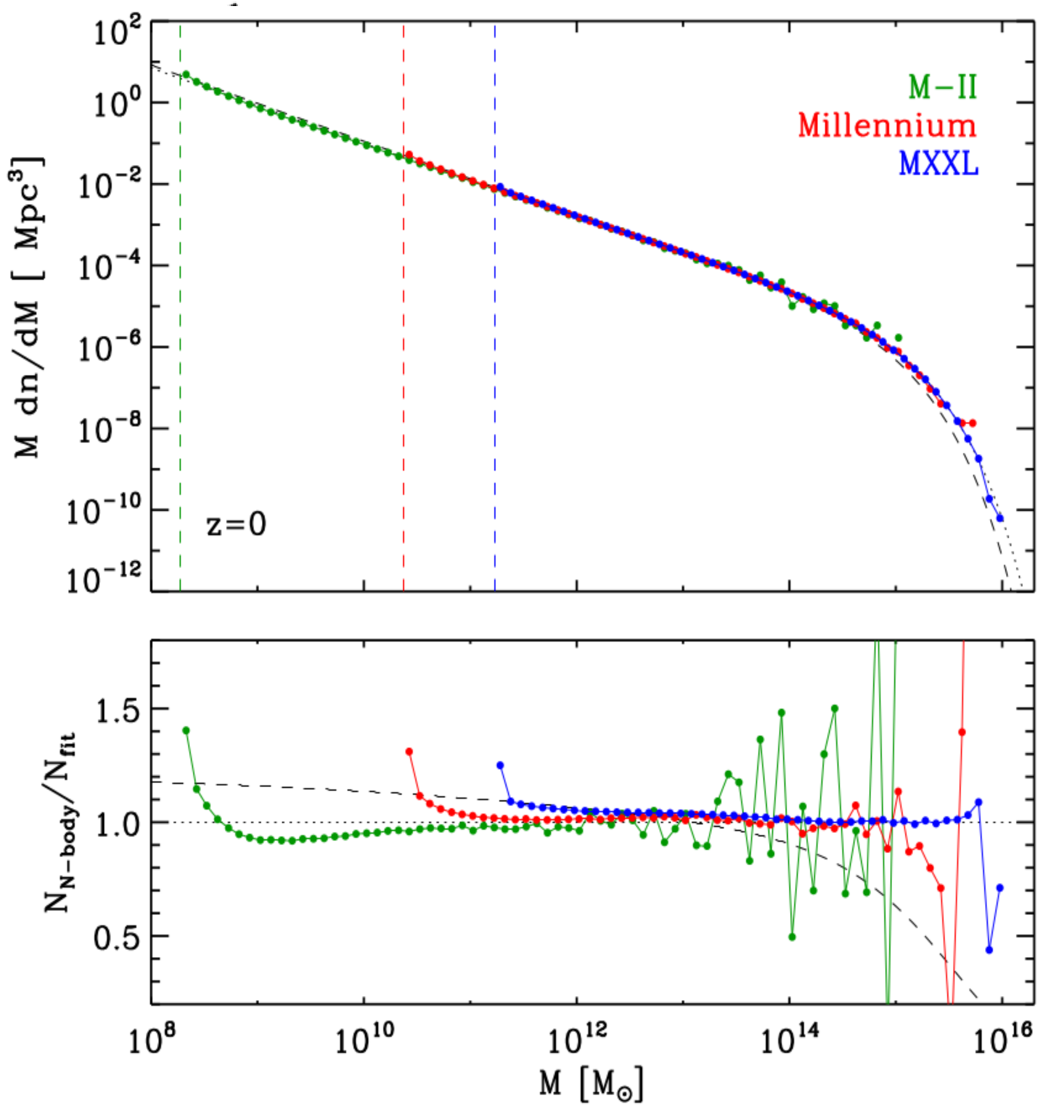
Name	Code	L_{hires} $[\mathrm{Mpc}]$	$N_{p,hires}$ $[10^9]$	$m_{p,hires}$ $[M_\odot]$	ϵ_{soft} $[\mathrm{pc}]$	$N_{sub}^{>100p}$ $[10^3]$	ref.
Aquarius A-1	Gadget-3	5.9	4.3×10^{9}	1.7×10^{3}	20.5	82	[45]
GHalo	Pkdgrav2	3.89	2.1×10^{9}	1.0×10^{3}	61.0	43	[32]
Via Lactea II	Pkdgrav2	4.86	1.0×10^{9}	4.1×10^{3}	40.0	13	[44]

Kuhlen, Vogelsberger, Angulo 2012

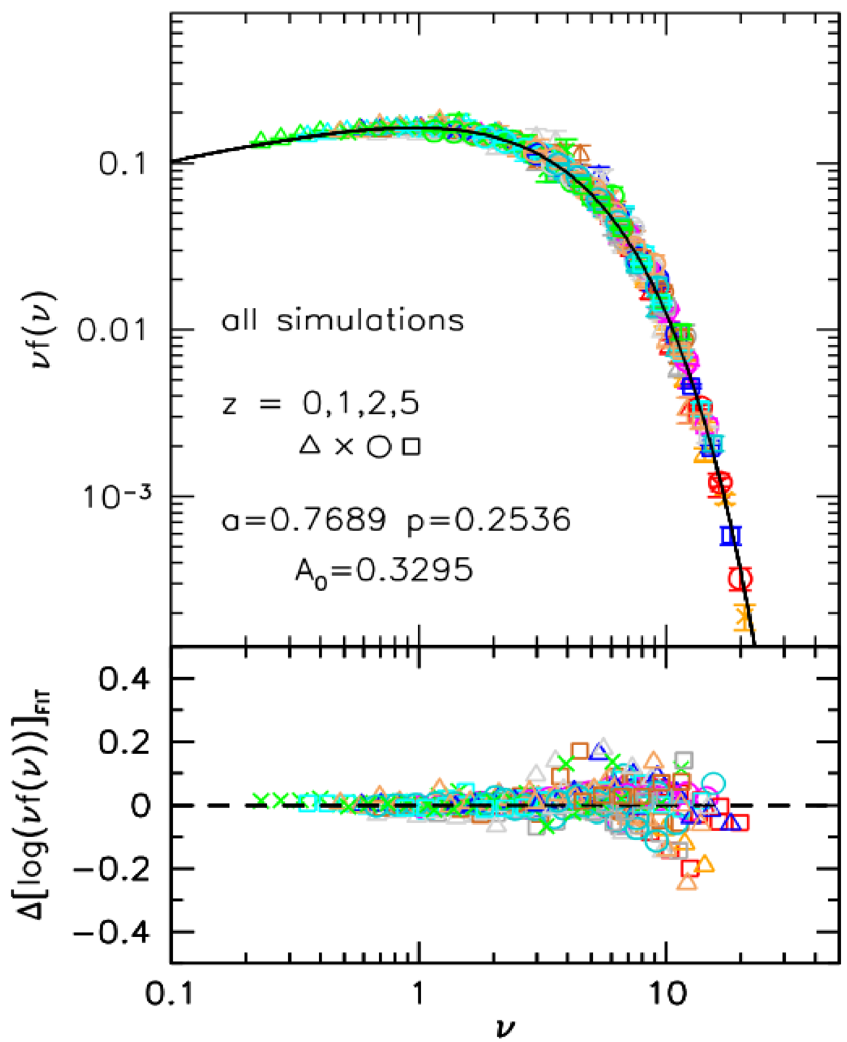
The abundance of CDM collapsed structures

Simulations resolve the mass range relevant for galaxy formation
If written in the adequate variables, the abundance is universal

dn/dM ~ $M^{-1.9}$ (at small masses)

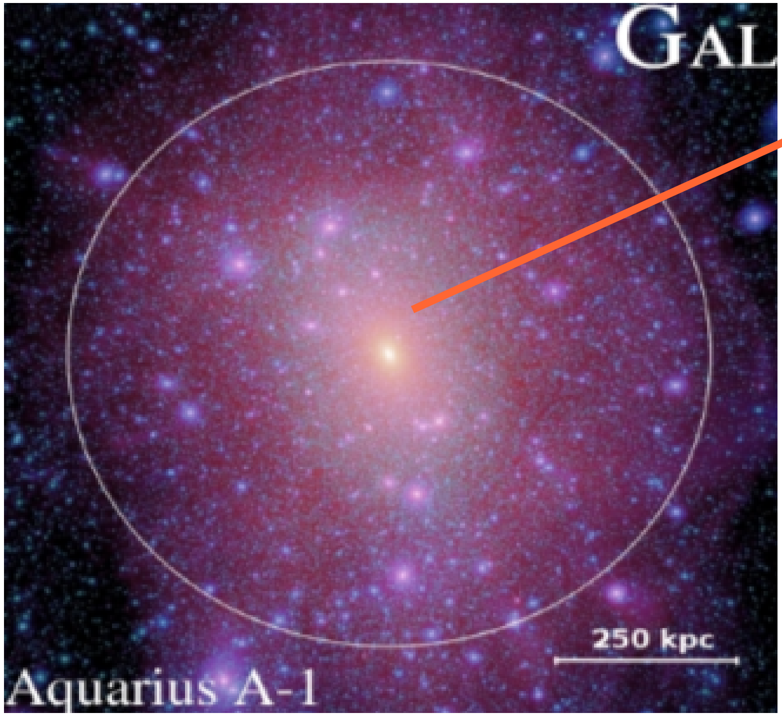


Angulo et al 2012



Despali et al 2015

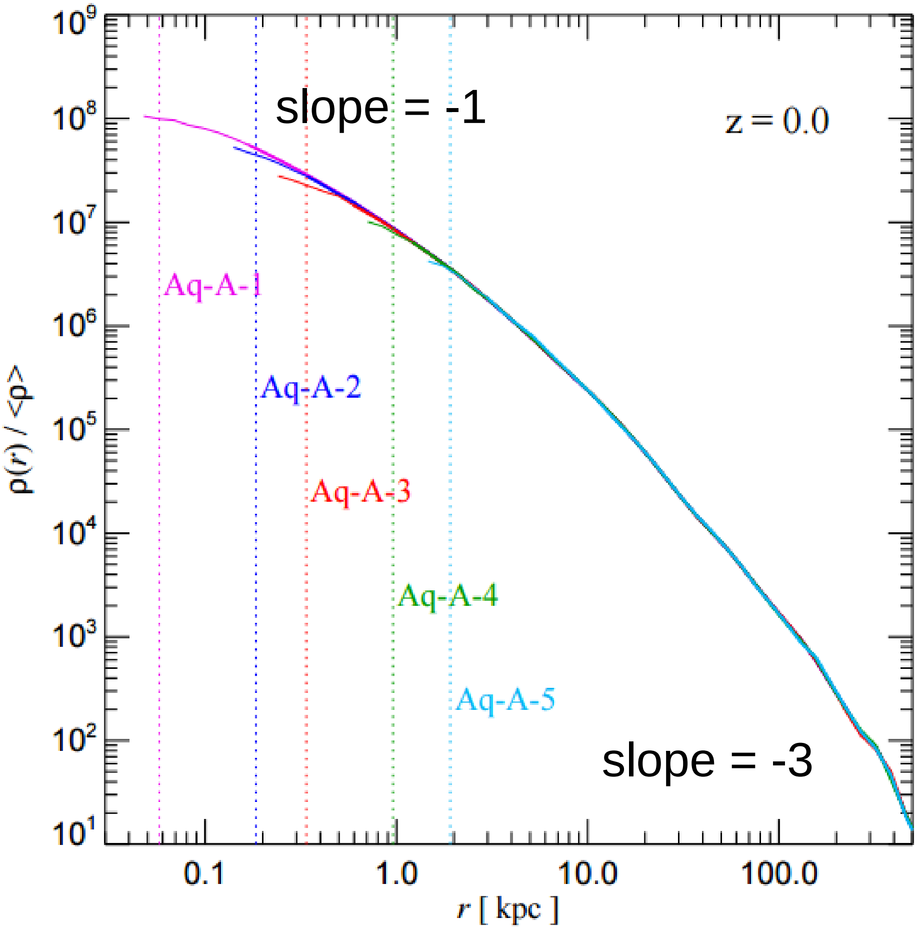
The inner structure of Dark Matter halos



Springel et al 2008

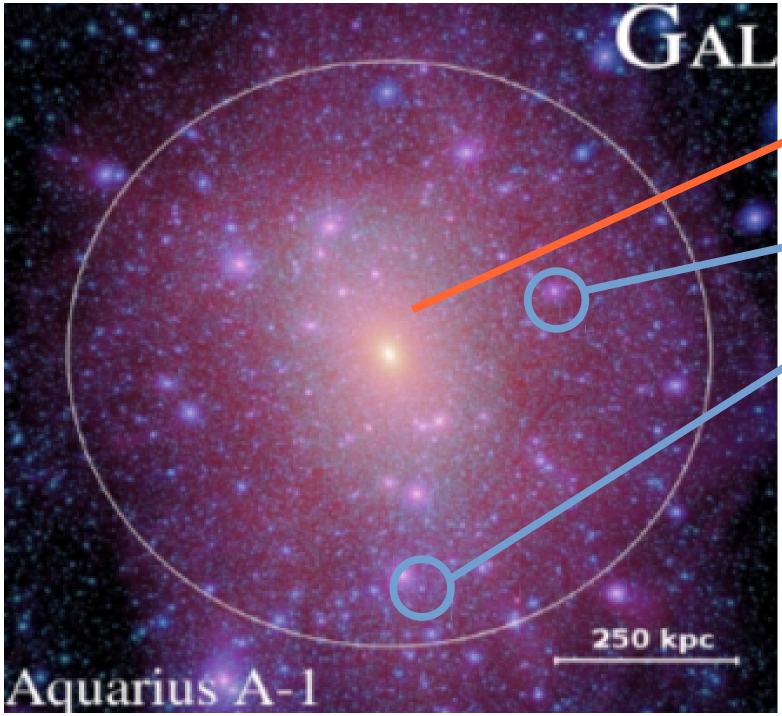
Smooth distribution

Density profile is described by NFW/Einasto functional form, independent of mass



Springel et al 2008

The inner structure of Dark Matter halos

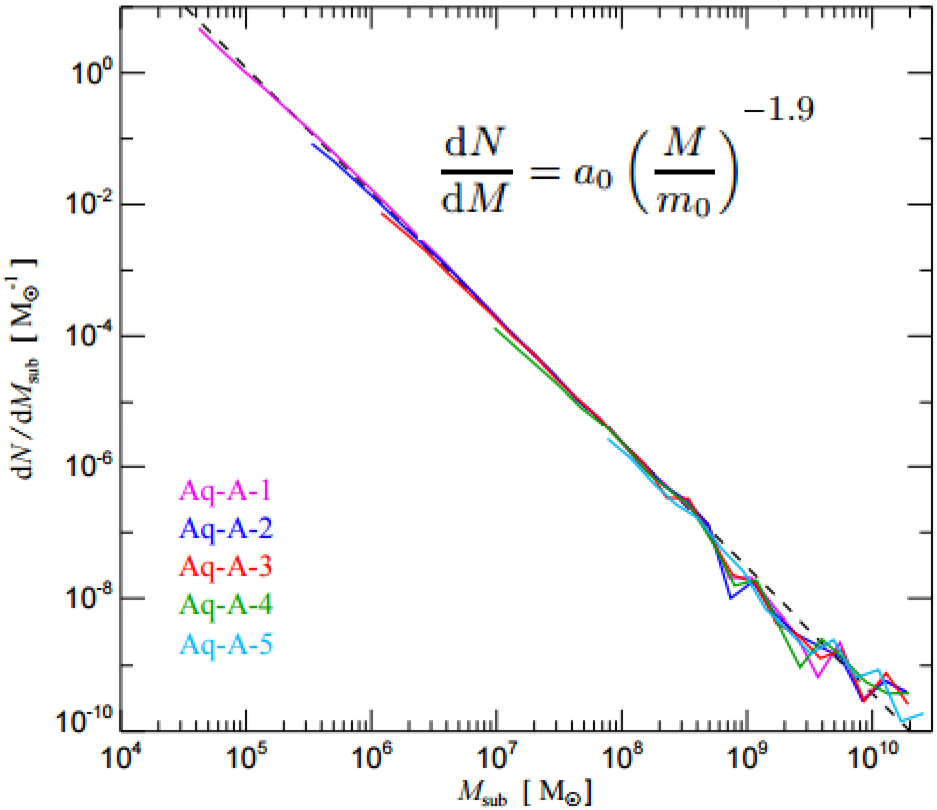


Springel et al 2008

Smooth distribution

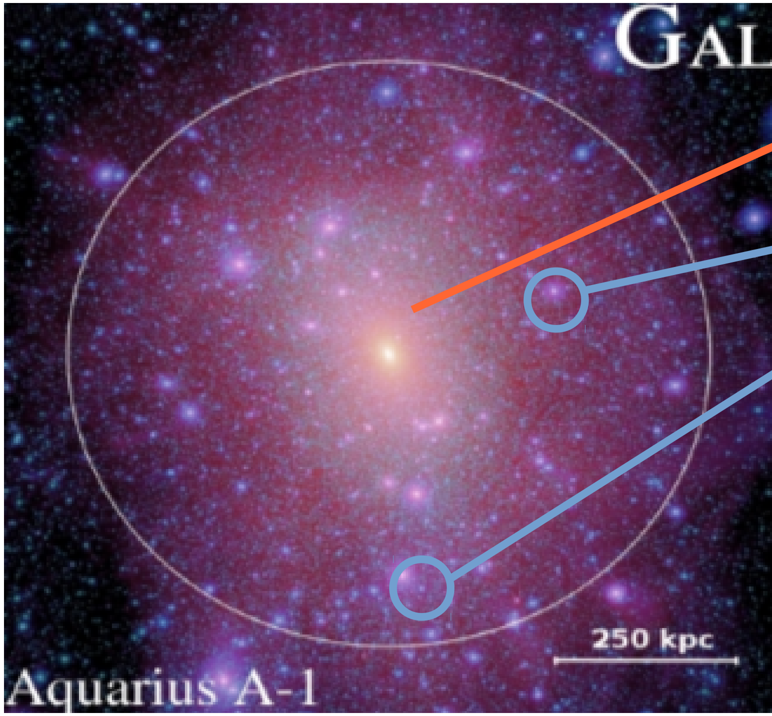
Hierarchy of substructures

→ **Abundance**



Springel et al 2008

The inner structure of Dark Matter halos

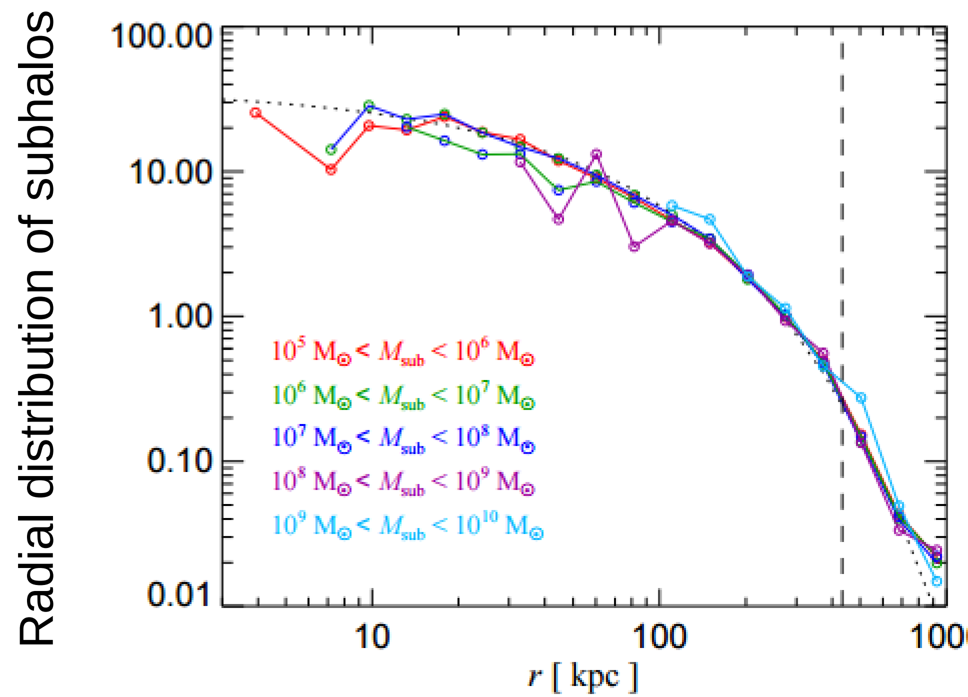


Springel et al 2008

Smooth distribution

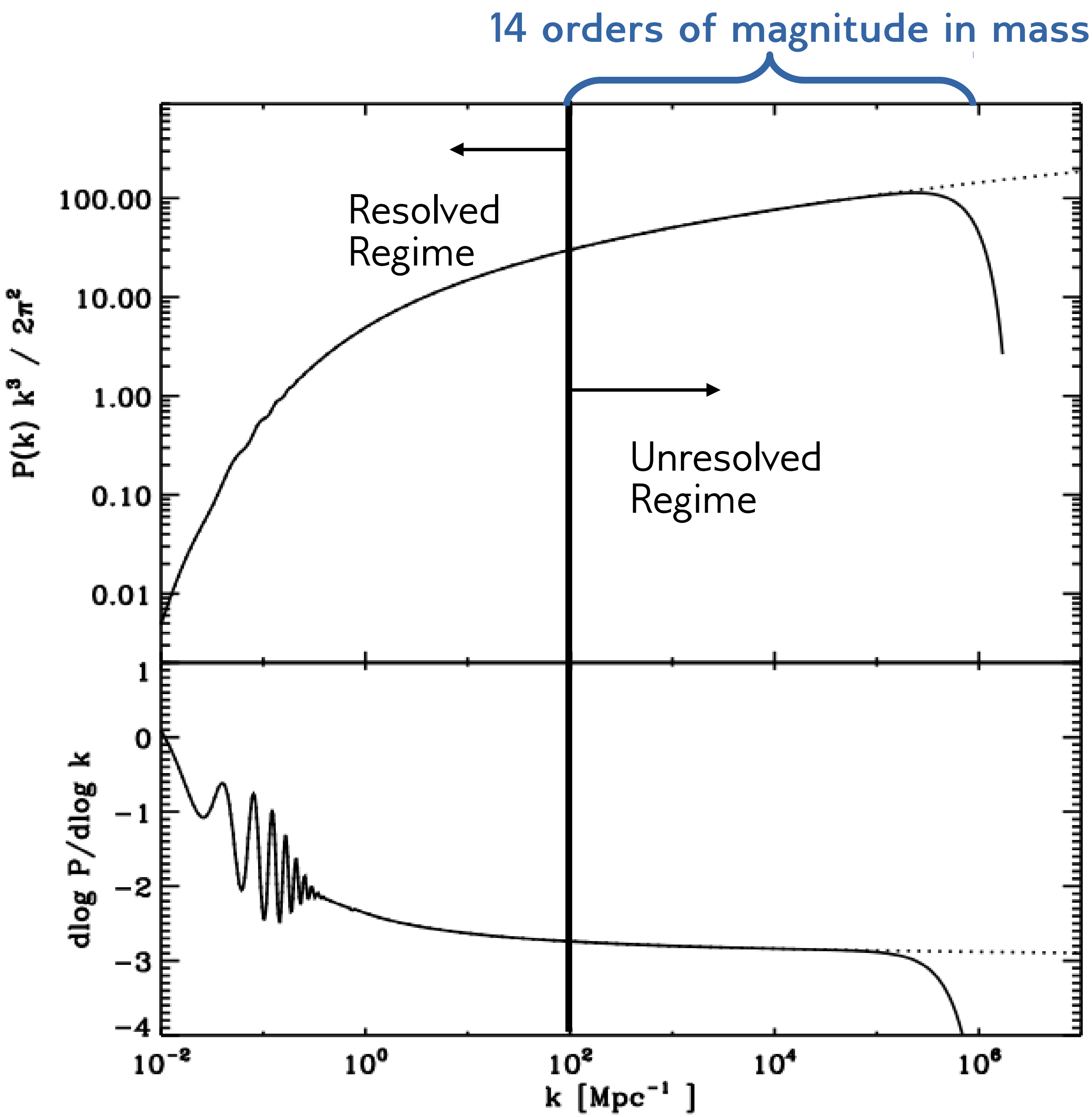
Hierarchy of substructures

→ **Abundance**
→ **Radial distribution**



Springel et al 2008

Structure formation for a 100GeV DM particle



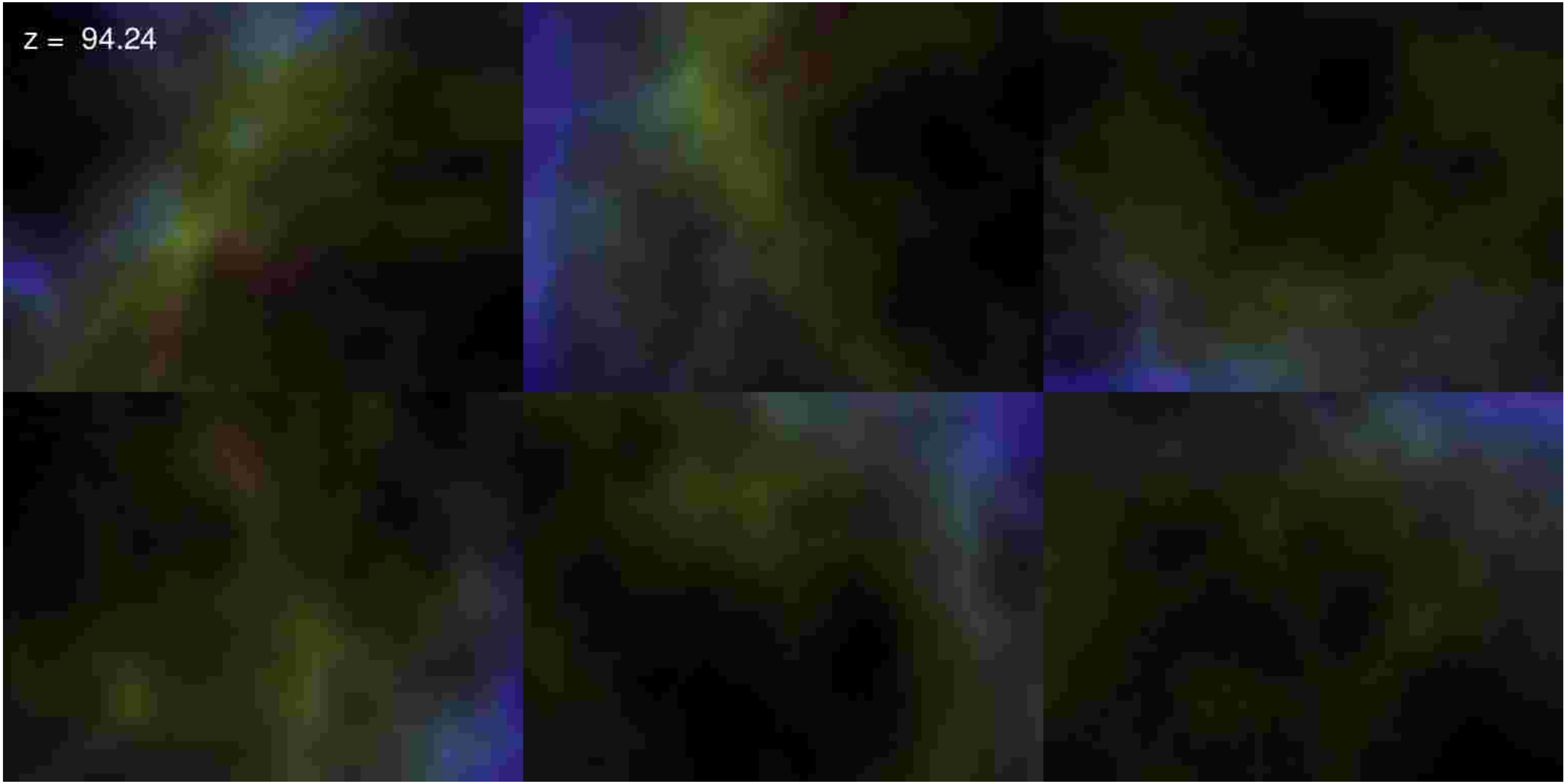
Angulo & White 2010

A simulation of the full DM hierarchy would require 10^{21} particles

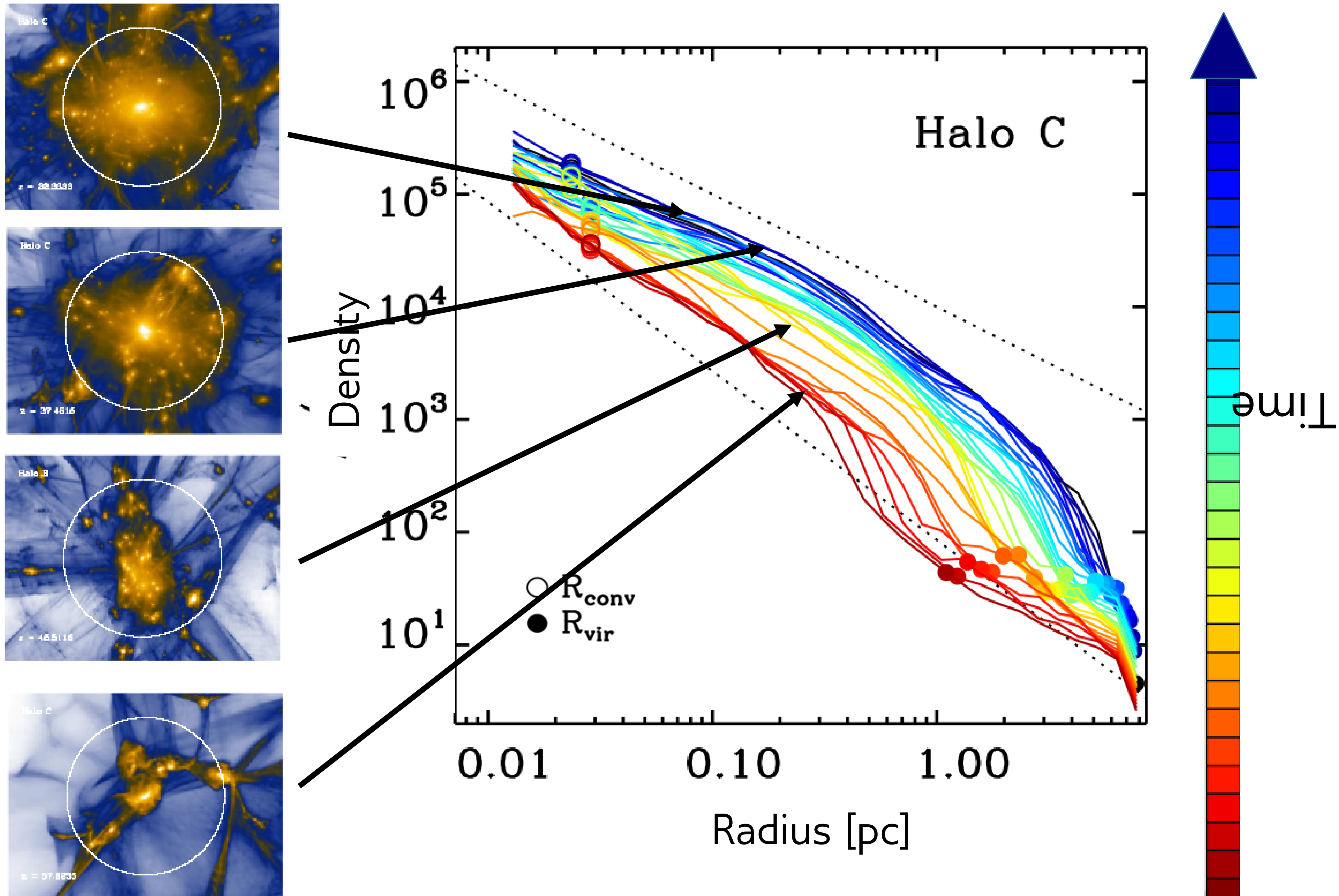
Current resolution studies *can not* be regarded as a proof of convergence

Structure formation at the free streaming mass

(Angulo, Hahn, Ludlow, Bonoli 2016)


z = 94.24

Structure formation at the free streaming mass



→ Background

→ Methods

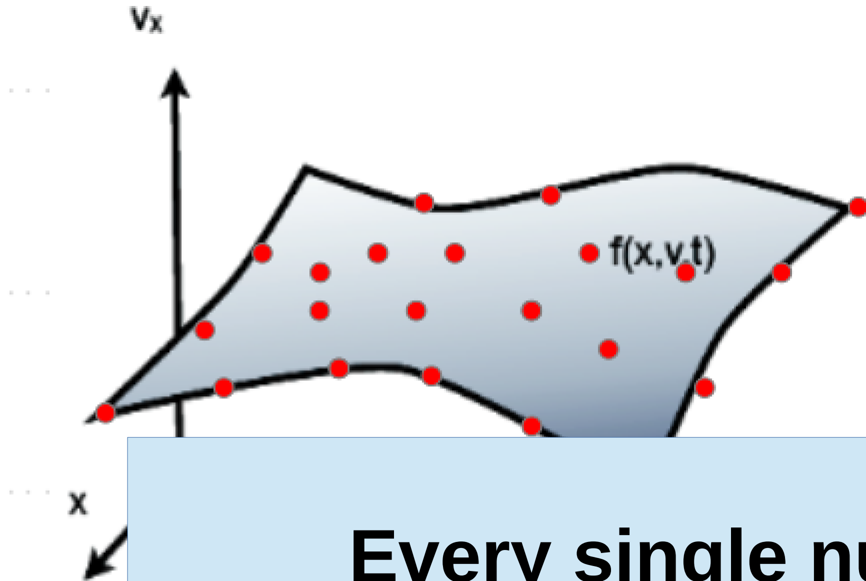
→ Current State of the Art

→ **The next decade:**
i) New VP solvers &
ii) Cosmological Parameters

→ Open questions & challenges

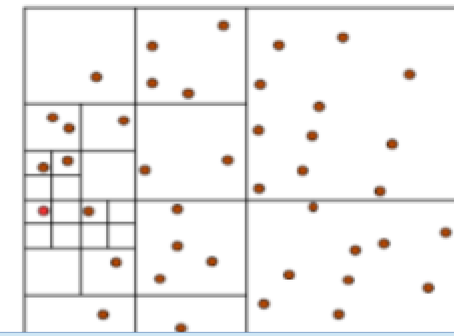
Standard approach to solving the VP equation:

Montecarlo Sampling and coarse graining the CDM distribution function



Tree Algorithms

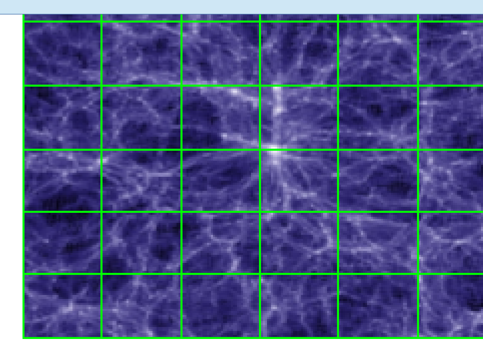
Multipole decomposition



Every single numerical simulation out there (even SPH/AMR) relies on the same assumption

$$\frac{d^2\mathbf{x}}{d^2t}$$

$$\Phi(\mathbf{x}) = -G\sum_i \frac{m_i}{[(\mathbf{x}_i - \mathbf{x})^2 + \varepsilon^2]}$$



Two examples where the N-body fails:

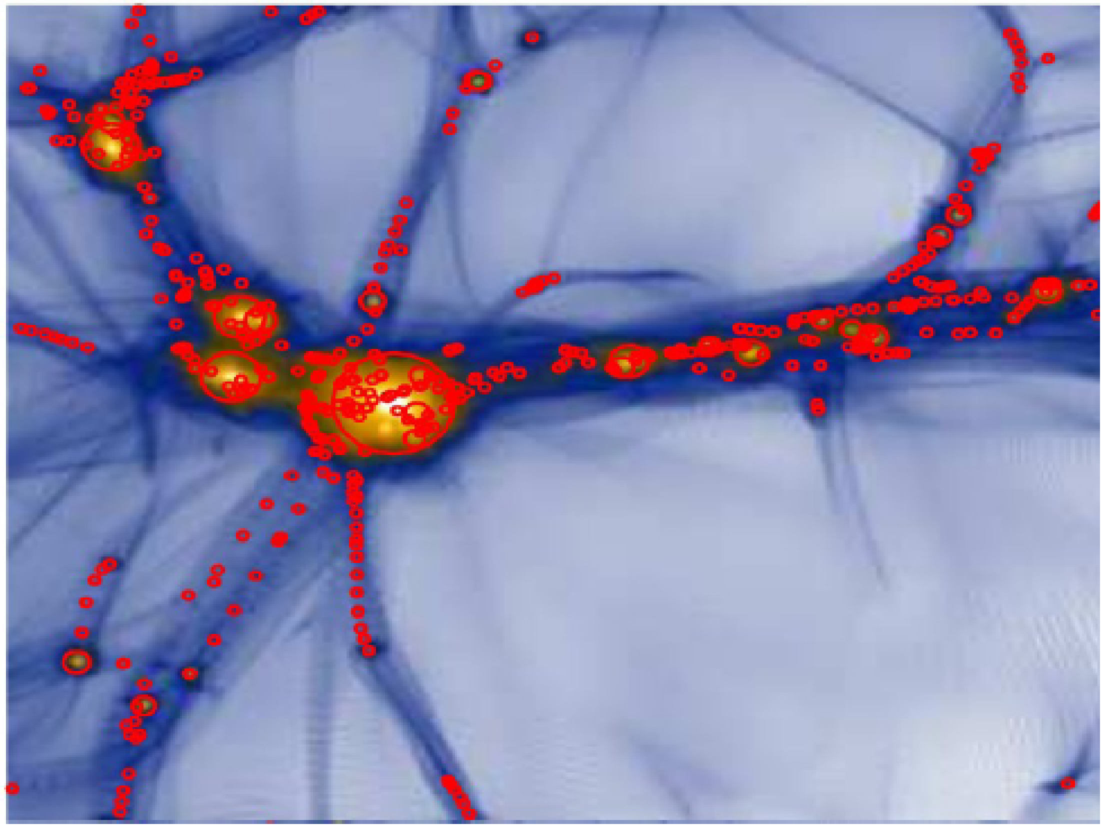
1) Two fluids with distinct primordial power spectra
2) Artificial fragmentation of filaments

$$\Phi(\mathbf{x}) = -G \sum_i \frac{m_i}{[(\mathbf{x}_i - \mathbf{x})^2 + \varepsilon^2]}$$

Two competing requirements For setting epsilon

i) A *large* ε value to reduce noise.

ii) A *small* ε value to resolve structures



The evolution of the fine and coarse grained distribution functions are NOT equivalent.

1) Two fluids with distinct primordial power spectra
2) Artificial fragmentation of filaments

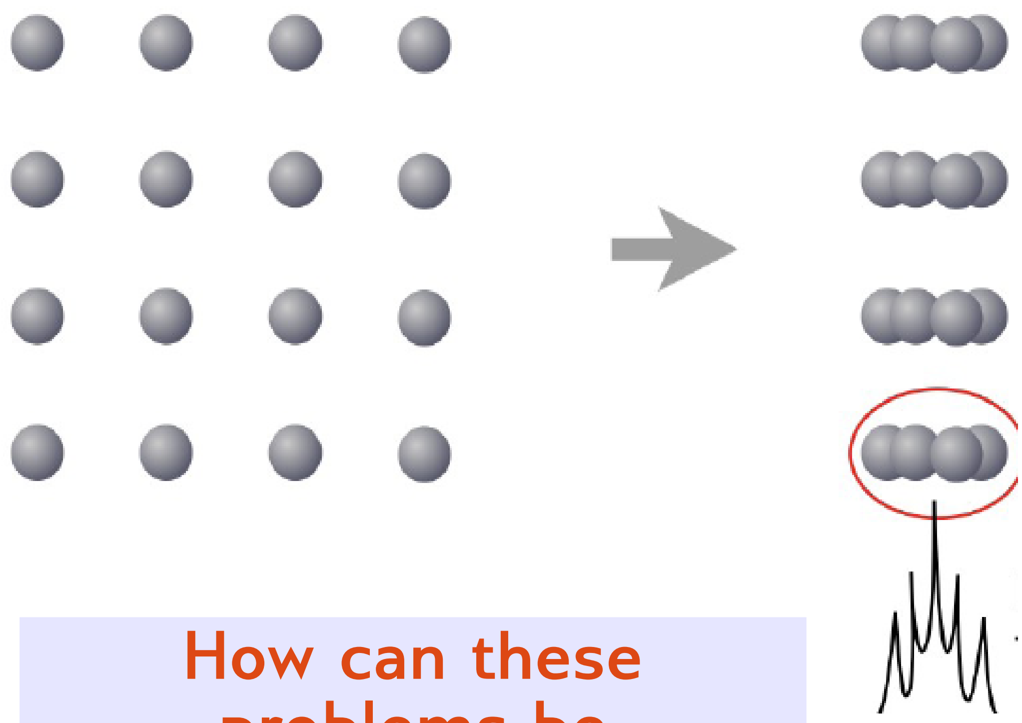
$$\Phi(\mathbf{x}) = -G \sum_i \frac{m_i}{[(\mathbf{x}_i - \mathbf{x})^2 + \varepsilon^2]}$$

Two competing requirements
For setting epsilon

i) A *large* ε value
to reduce noise.

ii) A *small* ε value to
resolve structures

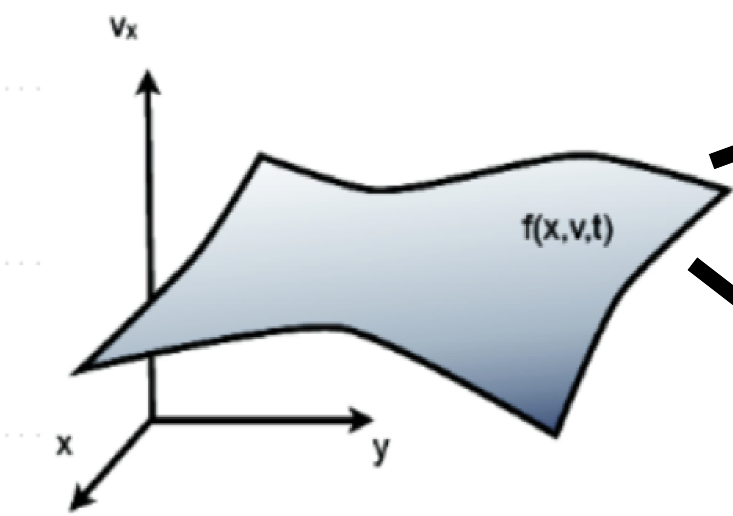
Anisotropic compression in triaxial collapse

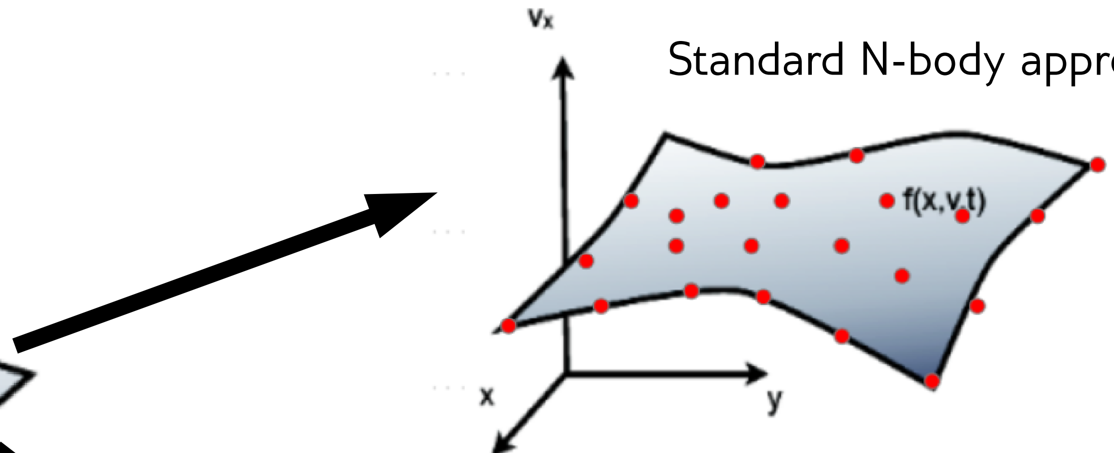


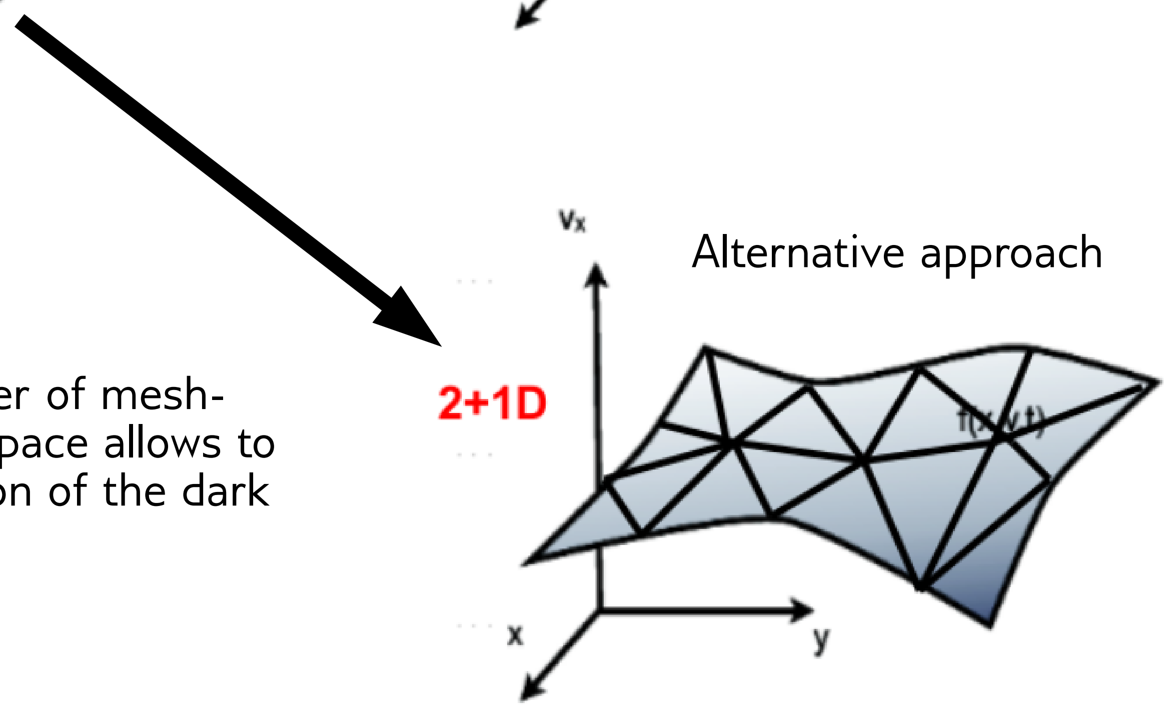
How can these problems be cured/tested?

Tessellation of the DM fluid with phase-space Lagrangian elements

(Abel+ 2012, Shandarin+ 2012, Kaehler+ 2013, Hahn+ 2013, Angulo+ 2013, Hahn & Angulo 2014)



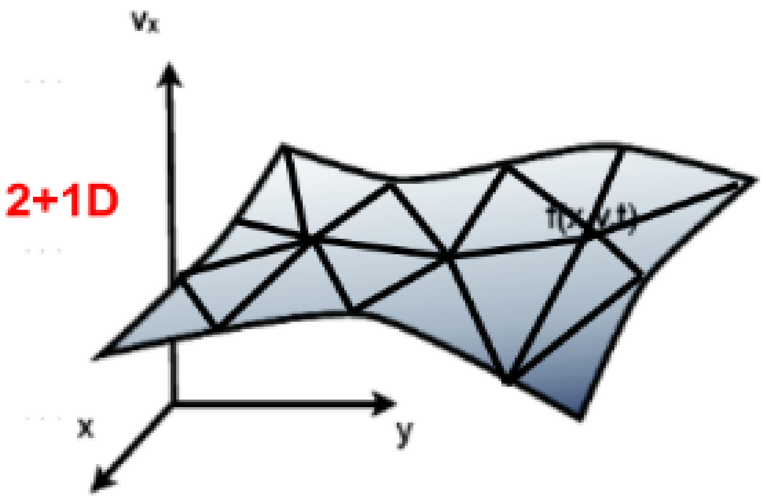




A tessellation of a finite number of mesh-generating points in Lagrangian space allows to continuously map the deformation of the dark matter sheet

Warm Dark Matter structure formation without noise (Angulo, Hahn, Abel 2013b)

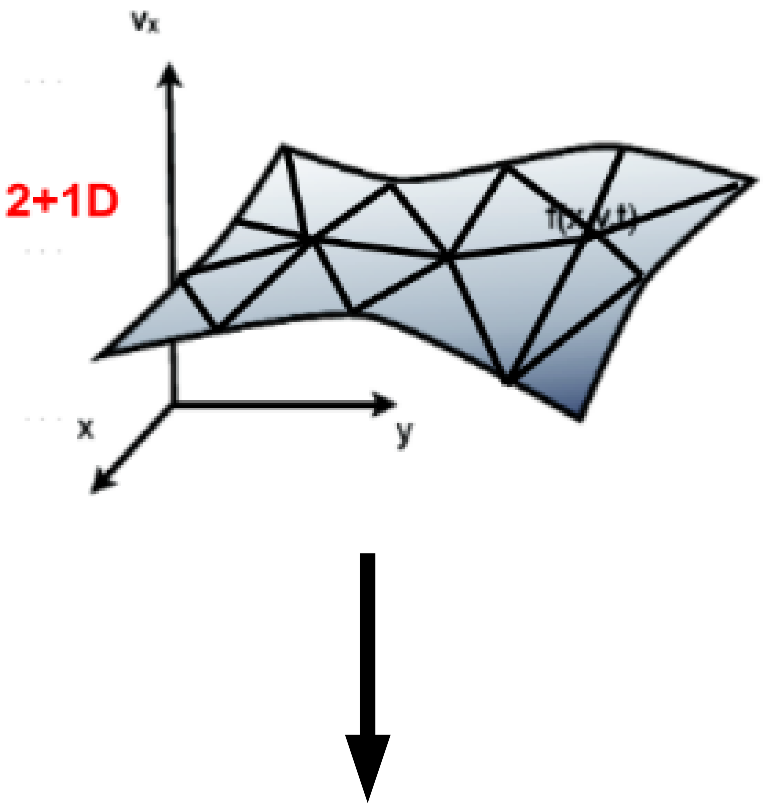
New sheet-based simulation code with reduced collisionality and noise



$$\nabla^2\Phi = \frac{4\pi G}{a}\int f \mathrm{d}^3 v$$

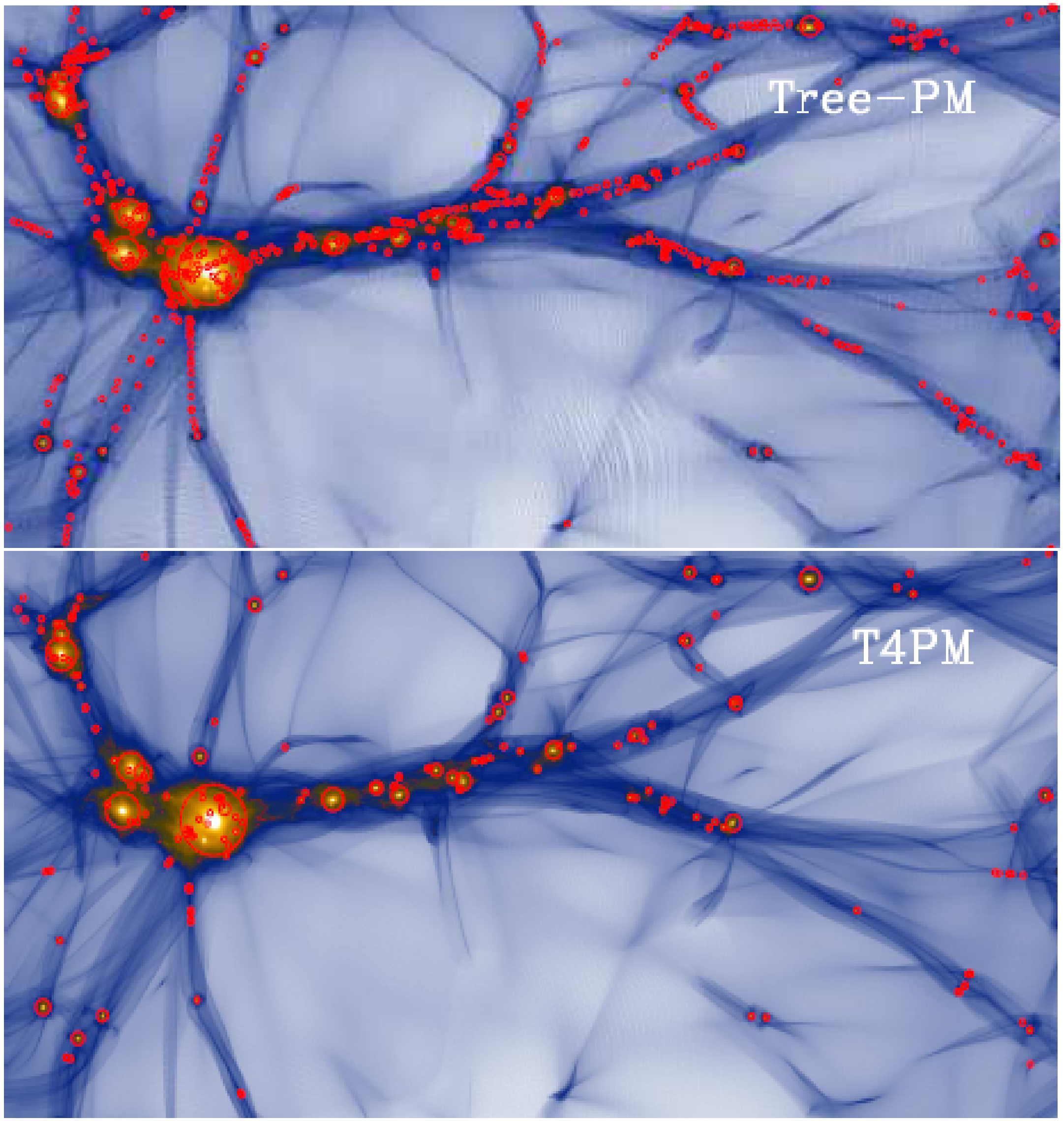
Warm Dark Matter structure formation without noise (Angulo, Hahn, Abel 2013b)

New sheet-based simulation code with reduced collisionality and noise



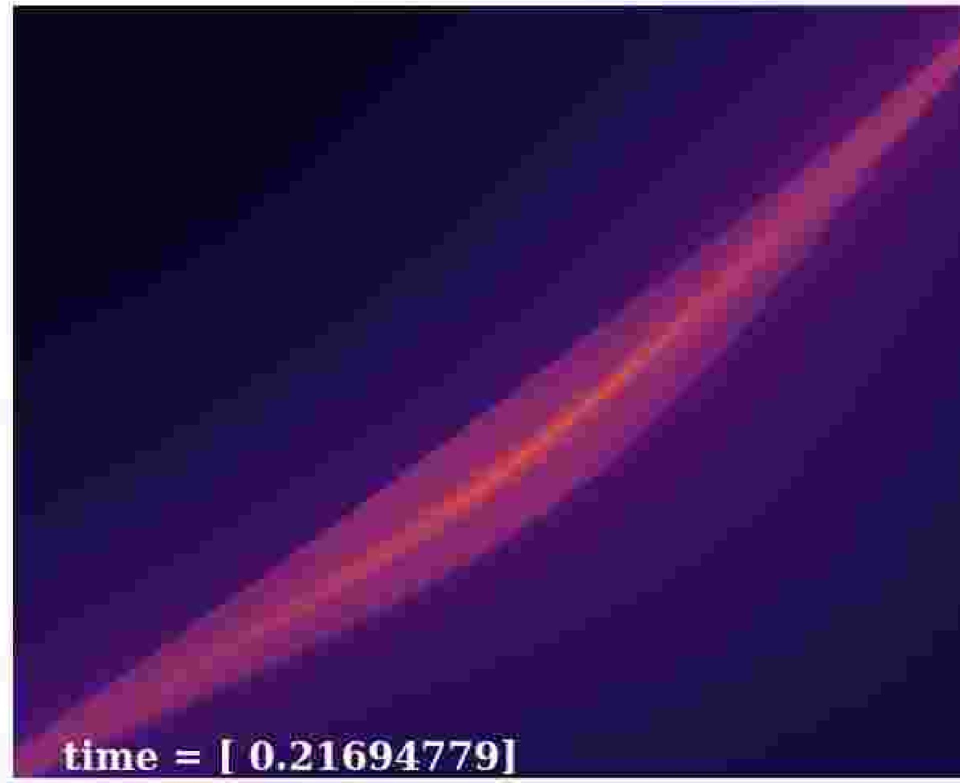
$$\nabla^2\Phi = \frac{4\pi G}{a}\int f\,\mathrm{d}^3v$$

(No need for a “softening length”)



Self-gravitating filament plus spherically-symmetric top-hat perturbation

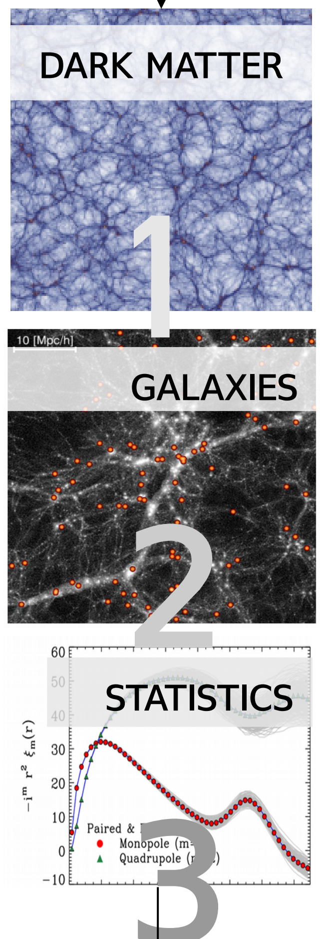
Standard N-body Simulation



Adaptively refined Lagrangian maps

The problem of optimally exploiting future and current surveys

Input Cosmology



Perturbation theory

– PT breaks quickly
– Higher order expansions loose predictive power

Analytic function

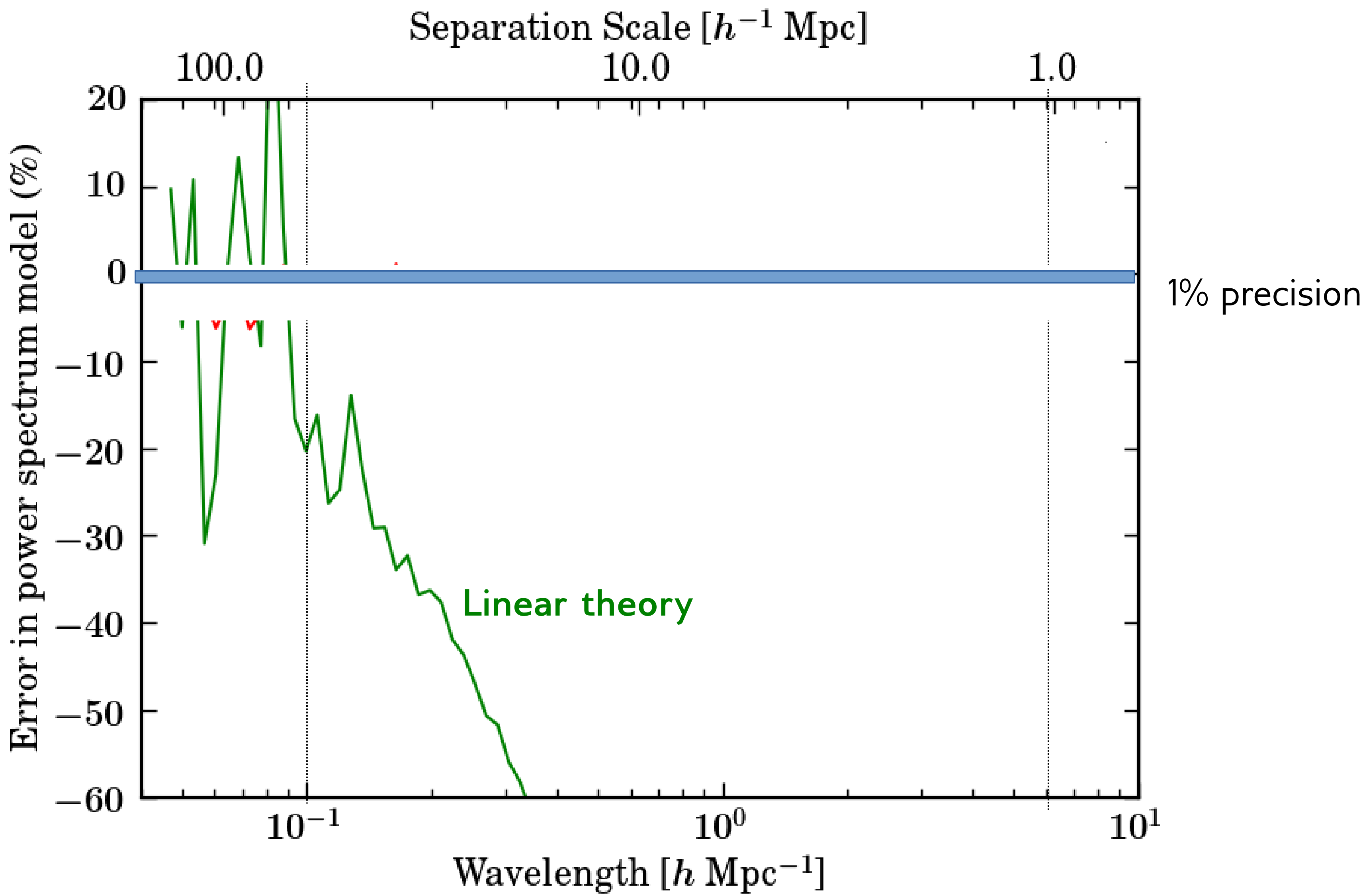
– Galaxy formation physics cannot be fully captured

Correlation functions

– Limited set of observables
– Hard to model survey setup
– Unknown likelihoods

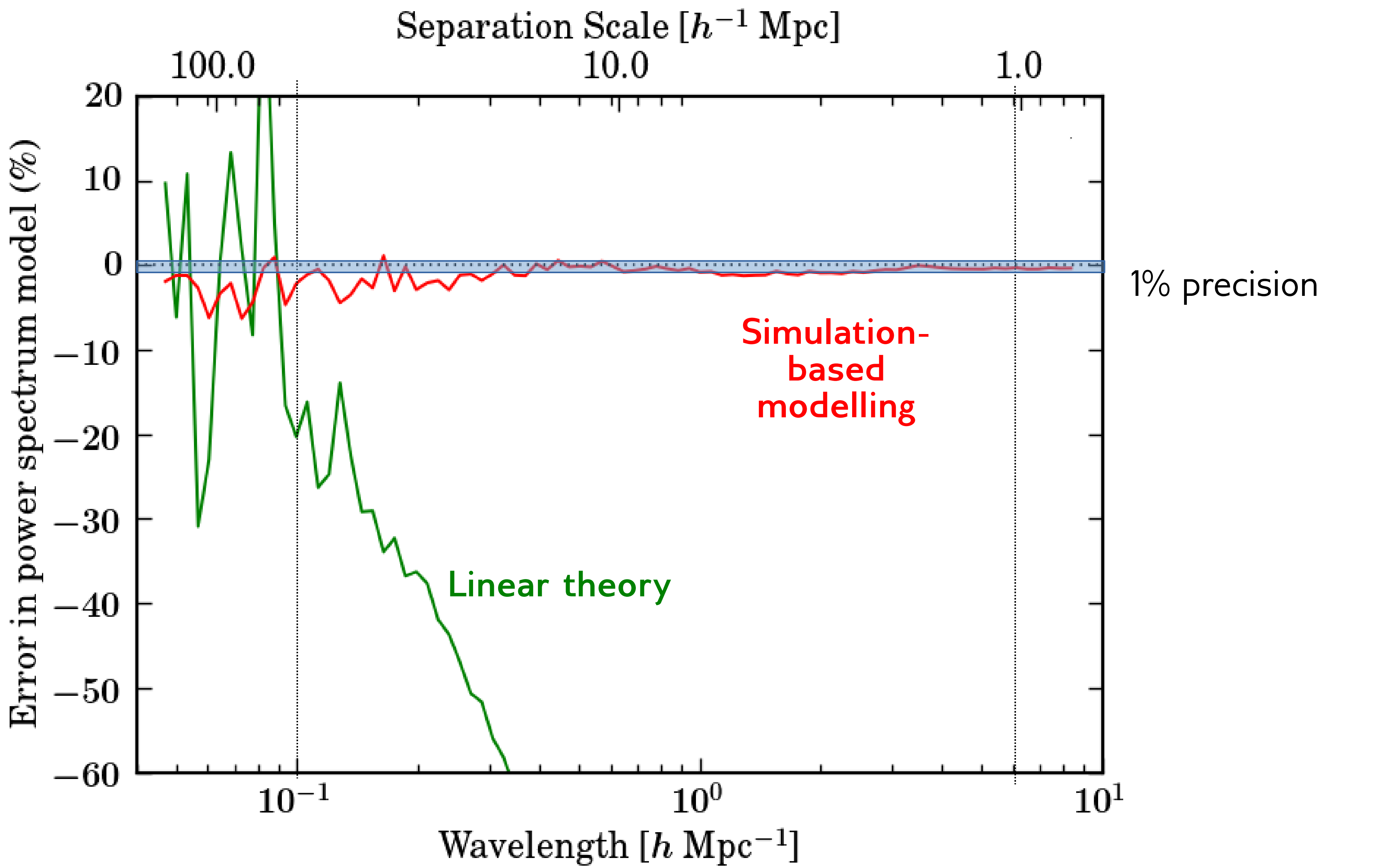
Cosmology

SHAM galaxies with $n=10^{-3}\,(h/\mathrm{Mpc})^3$



SHAM galaxies with n=10^{-3} (h/Mpc)^3

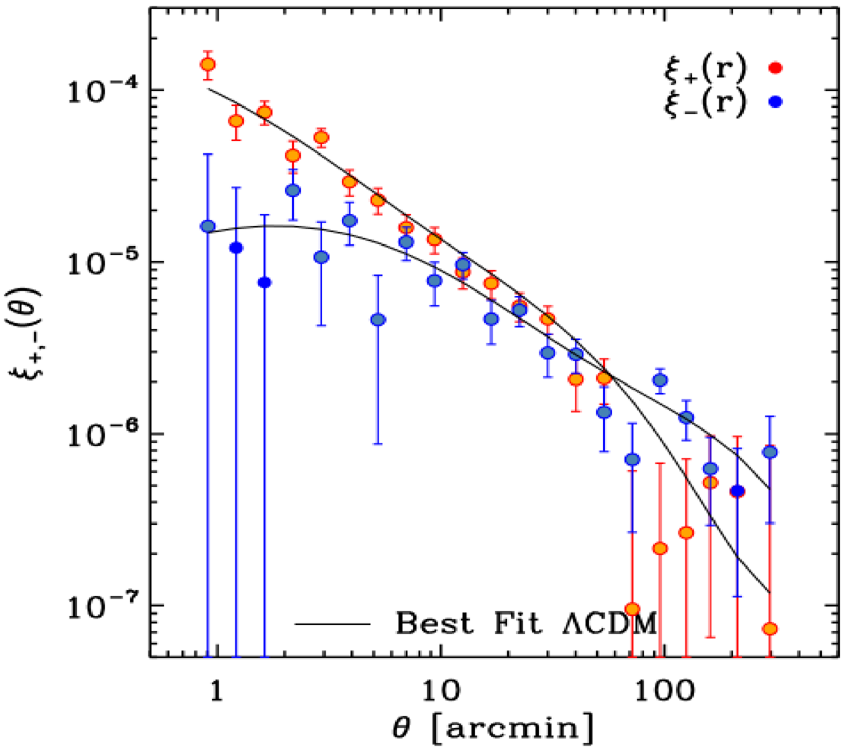
The power spectrum of SHAM galaxies with n=10^{-3} Mpc/h (including nonlinear RSD and nonlinear bias) is predicted at percent level, down to k = 10 h/Mpc



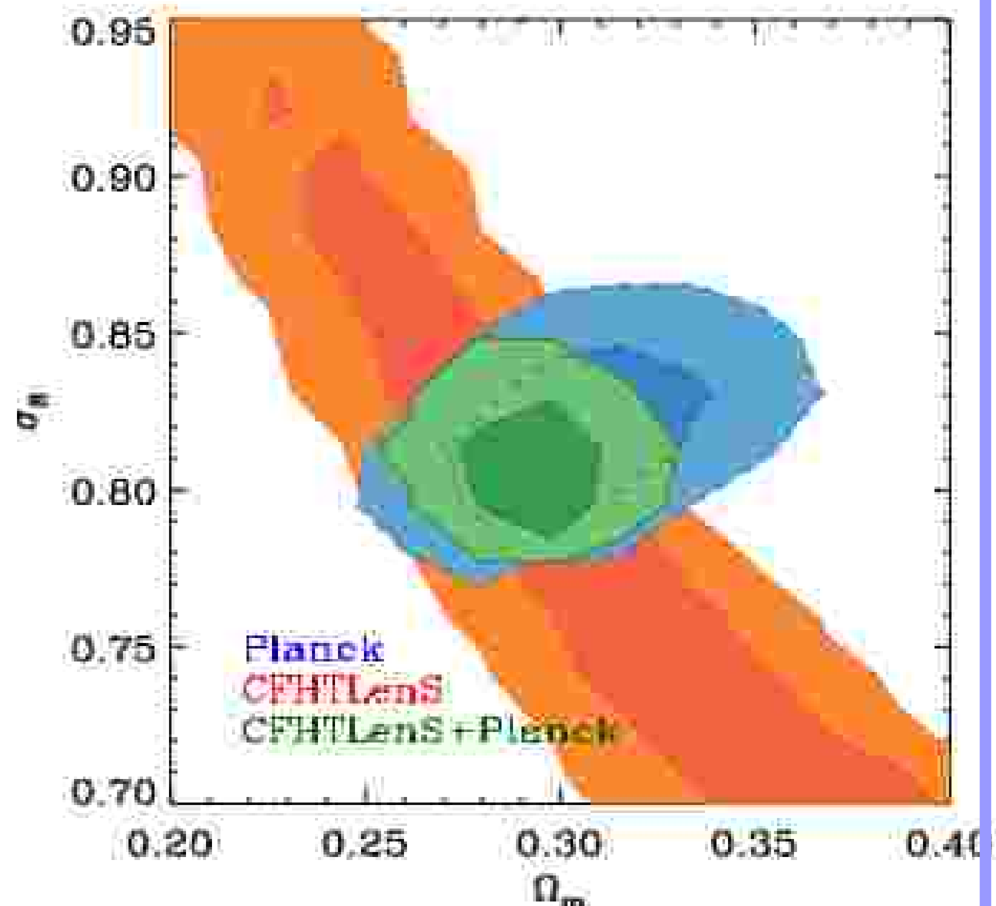
LSS forward modelling applied to lensing

ANALYSIS OF CFHTLenS USING MILLIONS OF SIMULATED UNIVERSES

Shear Correlation measurements



$$\Omega_m = 0.29 \pm 0.01$$
$$\sigma_8 = 0.81 \pm 0.01$$



Angulo & Hilbert 2015

Open Problems & Challenges

Observations are way ahead of theory, how can we catch up?

→ No simulation, even gravity only, can simultaneously resolve the volume and host halos of current surveys.

→ How can we increase the accuracy and precision of N-body Simulations?

→ We have a reasonably accurate theory of galaxy formation and nonlinear structures, but it is computationally slow… How do we take advantage of this in cosmological inferences?

Open Problems and Challenges

Can we resolve the full hierarchy of structures?

→ **Maybe, after 2050...**

→ Resolve the kinematic of stars in the smallest dwarf galaxies

→ What is the origin of nonlinear density profiles?

→ Improved predictions for the phase-space structure

→ Improved modelling of the microphysical properties of DM (and neutrinos).

Open Problems and Challenges

The impact of hydrodynamics/galaxy formation

→ What are the degeneracies between galaxy formation and Cosmology? How can we break those?

→ Under what conditions do baryons affect the central density of galaxies, and the orbits/dynamical friction of galaxies? (i.e. when gravity-only break?)

→ How realistic are current implementations of stellar/AGN feedback (hydrodynamical decoupling, energy injection) of what happens in molecular clouds? (Better treatment of radiation/non-thermal pressure support, non resolved turbulence, etc.)

Open Problems and Challenges

How to efficiently use the next generation of supercomputer facilities?

Future supercomputers will have ~1e5-1e6 CPUs, little memory per node, and enhanced by co-processors. Future codes will need different parallelisation strategies, have some redundancy, and mixed algorithms.

→ Analysis will be impossible in postprocessing. We need to inline everything in runtime.

→ Data products will be huge... how to best handle and distribute it?