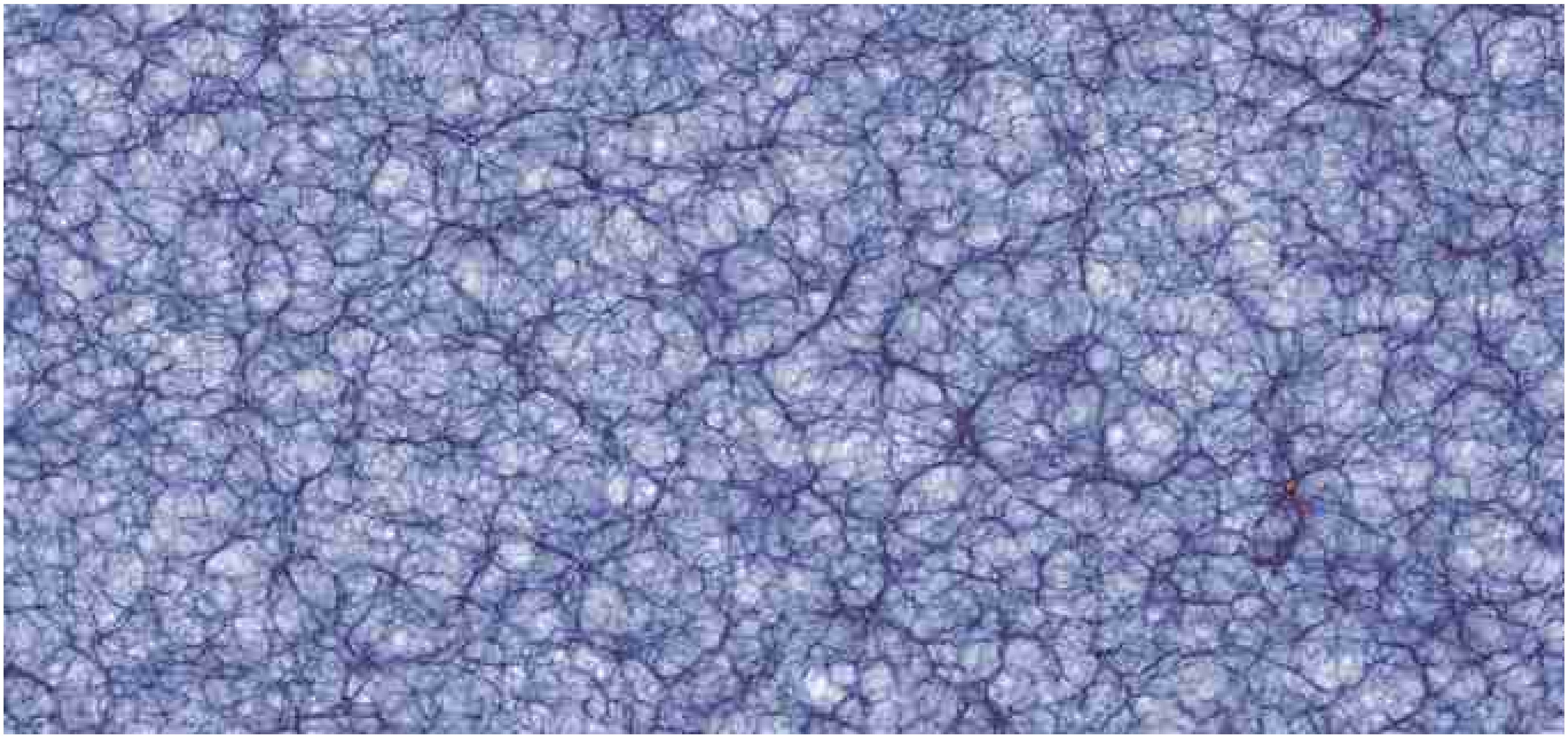
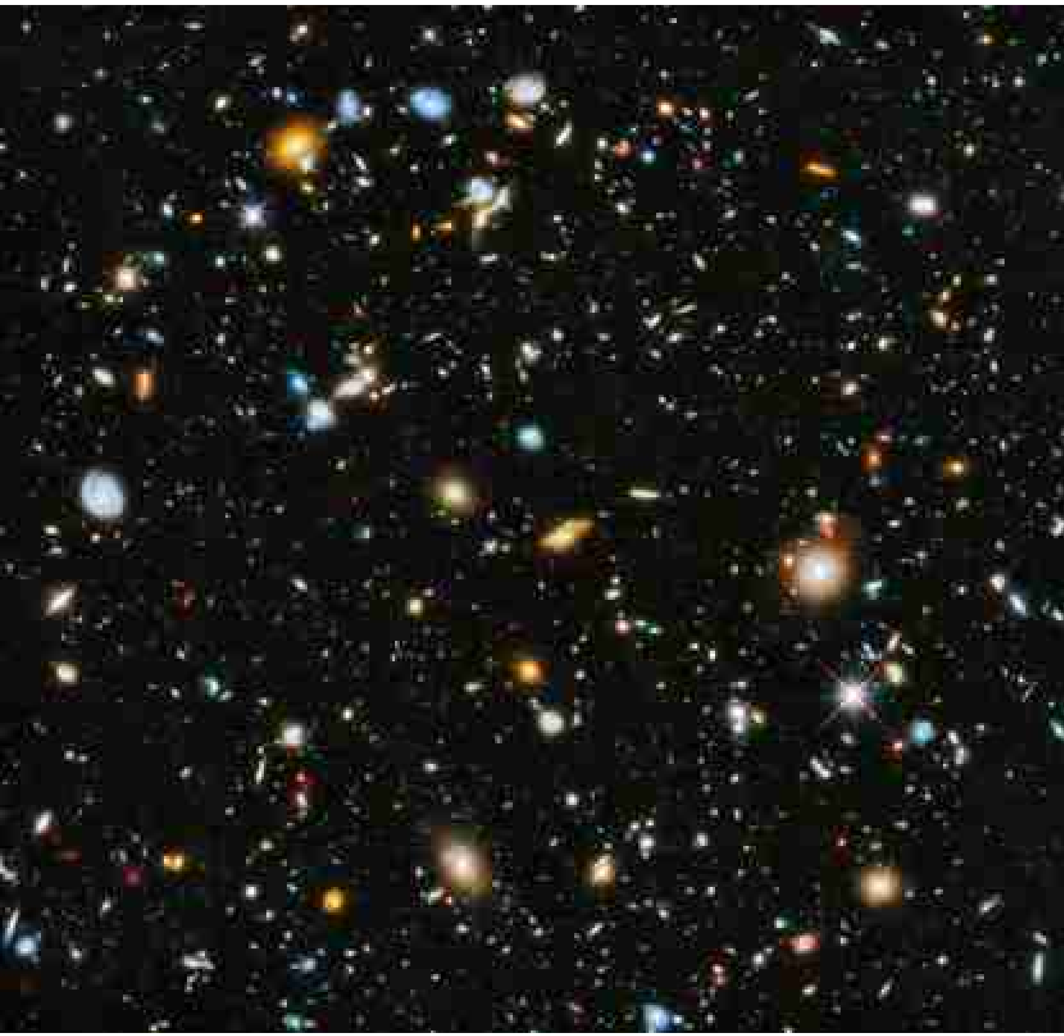
Simulating the formation of structure in the Universe

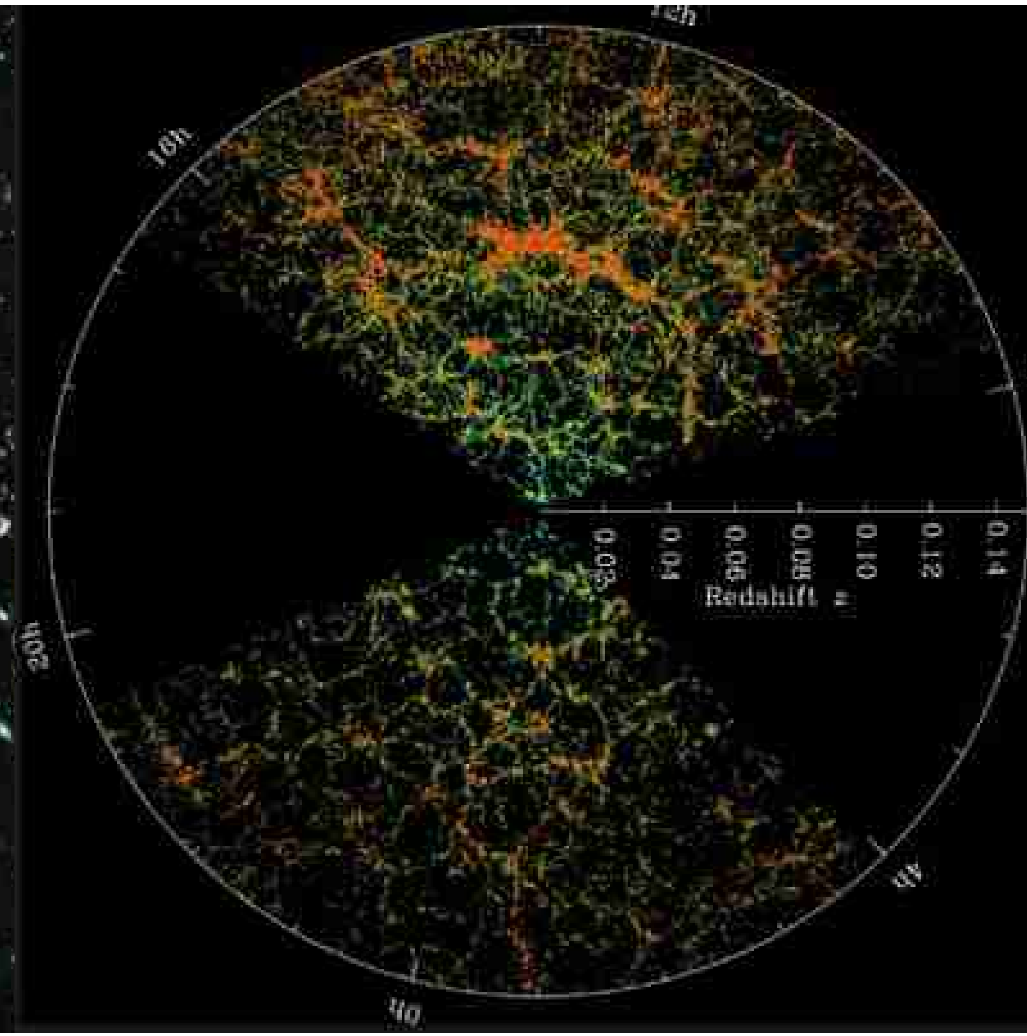


Raul E. Angulo

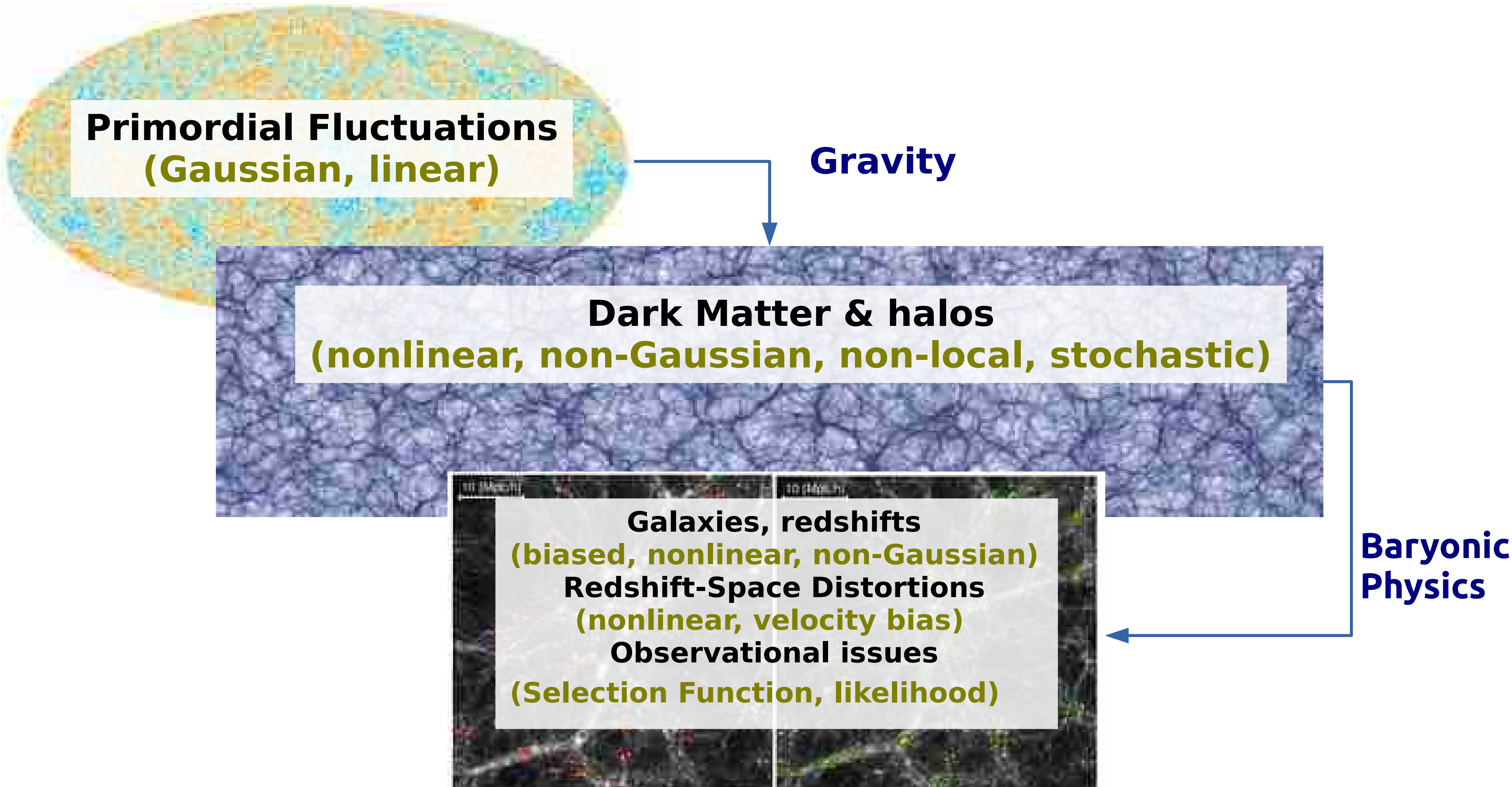
Alexander von Humboldt Stiftung/Foundation
ikerbasque Basque Foundation for Science
DIPC
erc

Is it possible to understand the very diverse universe we observe and even use it to infer fundamental physics?





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

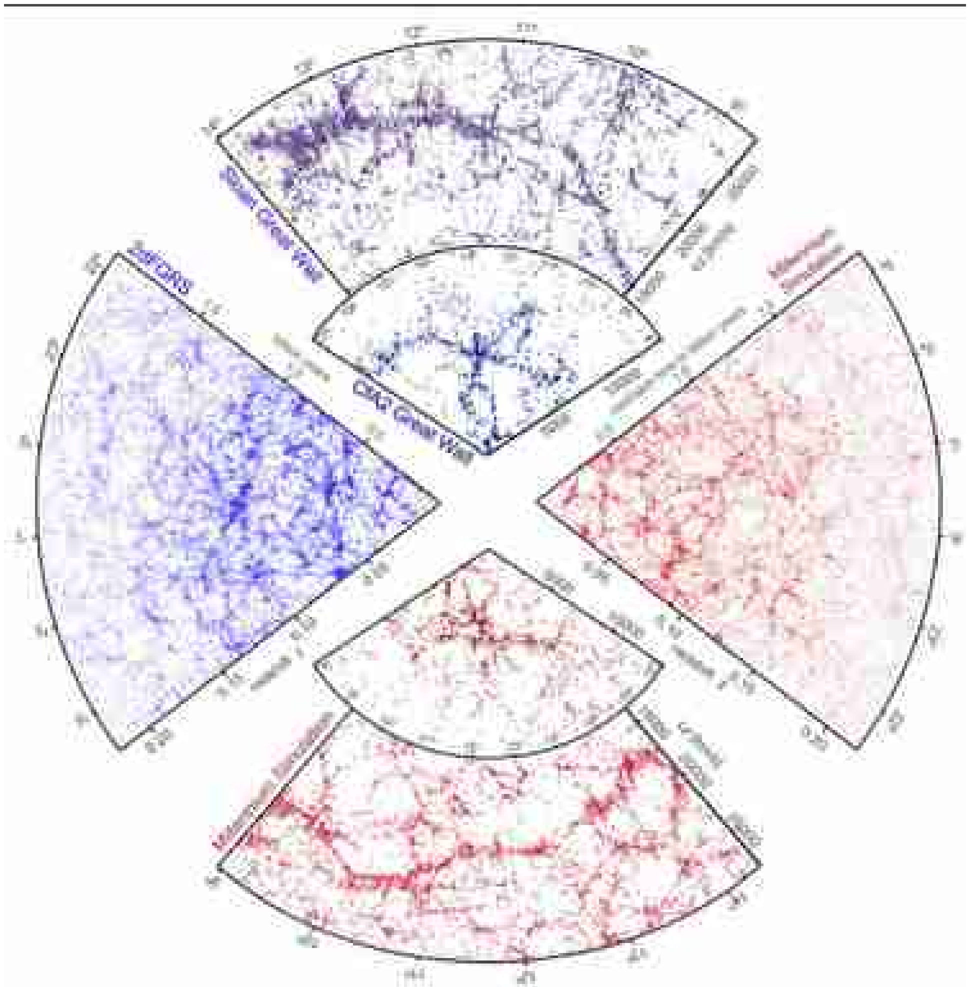


Numerical simulations are the most accurate way to understand nonlinear evolution in the Universe

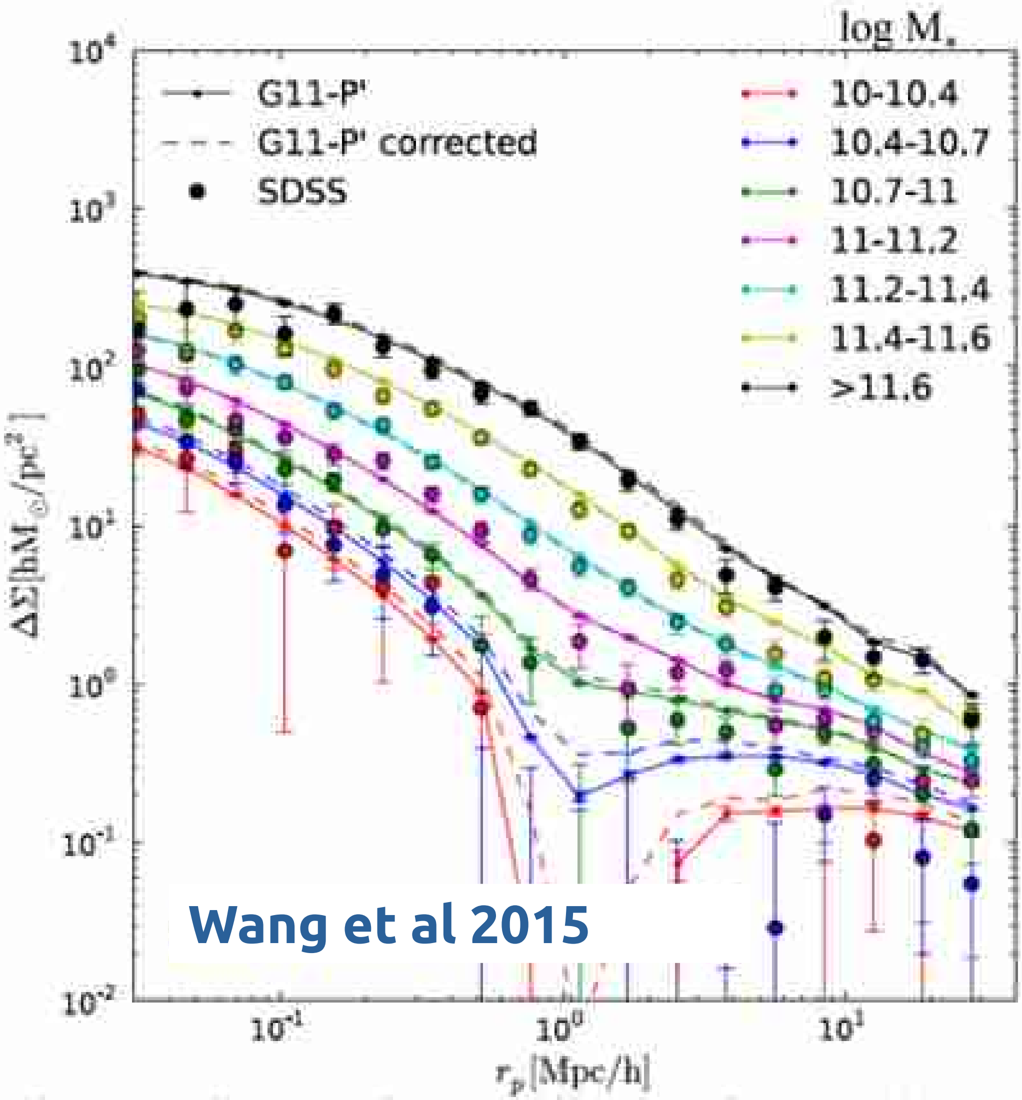
For a given set of components, evolution model and initial conditions, we can "predict" the low redshift Universe.

The formation and evolution of structure in the universe predicted by a state-of-the-art cosmological hydrodynamic simulation

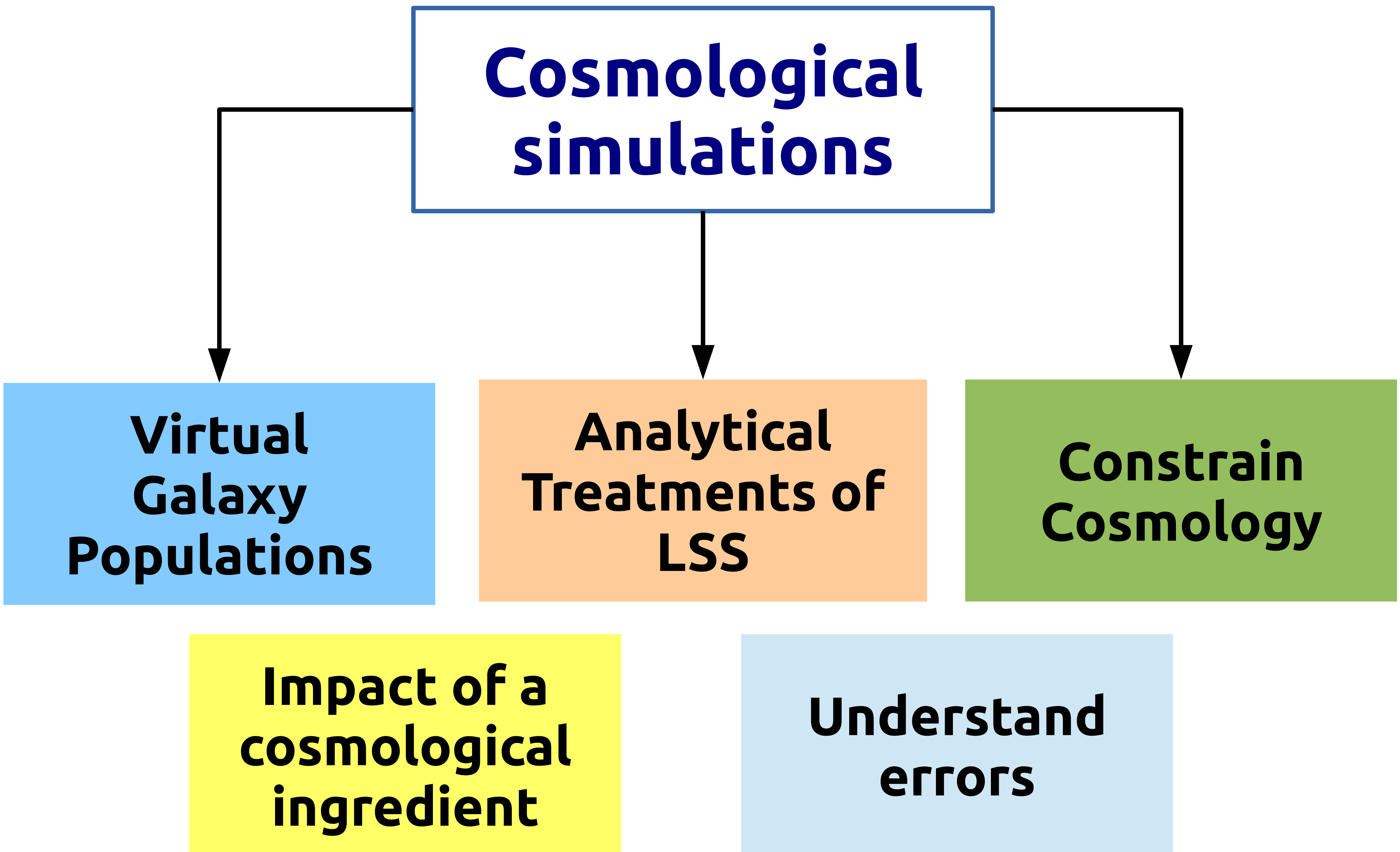
The distribution of galaxies and their surrounding dark matter halos as predicted by a dark matter only simulation



Springel et al 2006



Numerical simulations are an essential tool for precision cosmology



Lectures on simulating the formation of structure in the Universe

→ Basic Methods & Algorithms

→ Beyond the simplest models

→ State of the Art

→ The future

Living Rev Comput Astrophys (2022)8:1
https://doi.org/10.1007/s41115-021-00013-z

REVIEW ARTICLE

Check for updates

Large-scale dark matter simulations

Raul E. Angulo[1,2] · Oliver Hahn[3,4,5]

Received: 24 March 2021 / Accepted: 11 November 2021

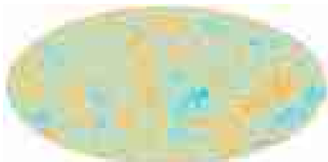
https://arxiv.org/abs/2112.05165

Abstract

We review the field of collisionless numerical simulations for the large-scale structure of the Universe. We start by providing the main set of equations solved by these simulations and their connection with General Relativity. We then recap the relevant numerical approaches: discretization of the phase-space distribution (focusing on N-body but including alternatives, e.g., Lagrangian submanifold and Schrödinger–Poisson) and the respective techniques for their time evolution and force calculation (direct summation, mesh techniques, and hierarchical tree methods). We pay attention to the creation of initial conditions and the connection with Lagrangian Perturbation Theory. We then discuss the possible alternatives in terms of the micro-physical properties of dark matter (e.g., neutralinos, warm dark matter, QCD axions, Bose–Einstein condensates, and primordial black holes), and extensions to account for multiple fluids (baryons and neutrinos), primordial non-Gaussianity and modified gravity. We continue by discussing challenges involved in achieving highly accurate predictions. A key aspect of cosmological simulations is the connection to cosmological observables, we discuss various techniques in this regard: structure finding, galaxy formation and baryonic modelling, the creation of emulators and light-cones, and the role of machine learning. We finalise with a recount of state-of-the-art large-scale simulations and conclude with an outlook for the next decade.

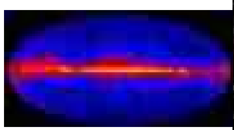
Basic Methods & Algorithms

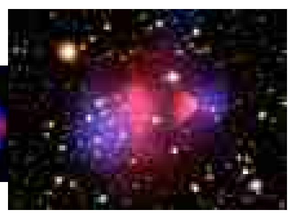
Cold-collisionless fluid under the effect of self-gravity in an expanding Universe



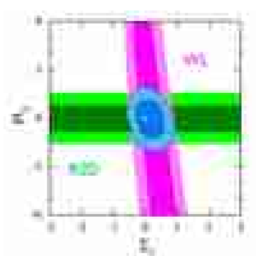
Dark matter is the dominant component

Dark matter is cold (thermal velocity << coherent motions)





Dark matter is non-interacting (collisionless)



No detection of modified GR, nor departures from a cosmological constant

$$-3\mathcal{H}\left(\phi' + \mathcal{H}\psi\right) + \nabla^2\phi + \frac{3}{2}\mathcal{H}^2 = 4\pi G a^2 \rho,$$

Newtonian limit of GR and no backreaction

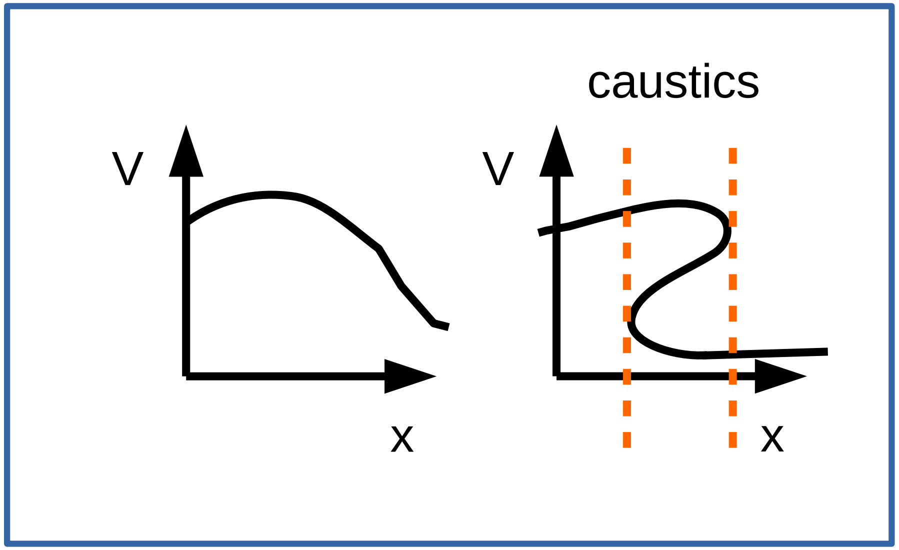
The Vlassov-Poisson Equation

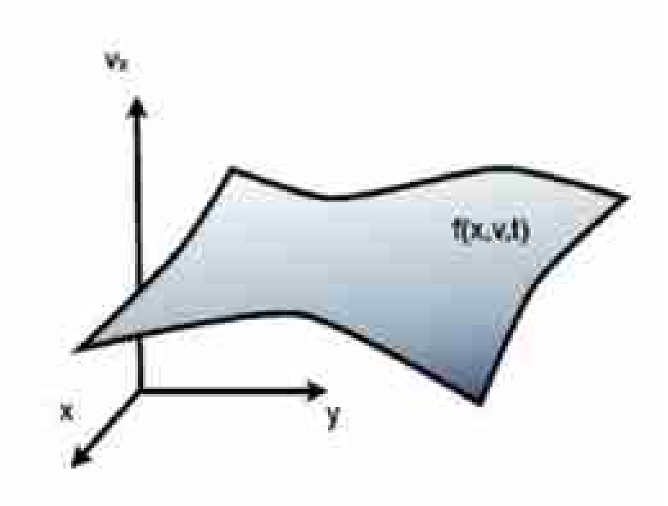
$$0 = \frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla_{\mathbf{x}} f + \frac{\mathbf{F}(\mathbf{x})}{m} \cdot \nabla_{\mathbf{u}} f$$

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

CDM Sheet Properties

- → phase-space is conserved along characteristics
- → It can never tear
- → It can never intersect





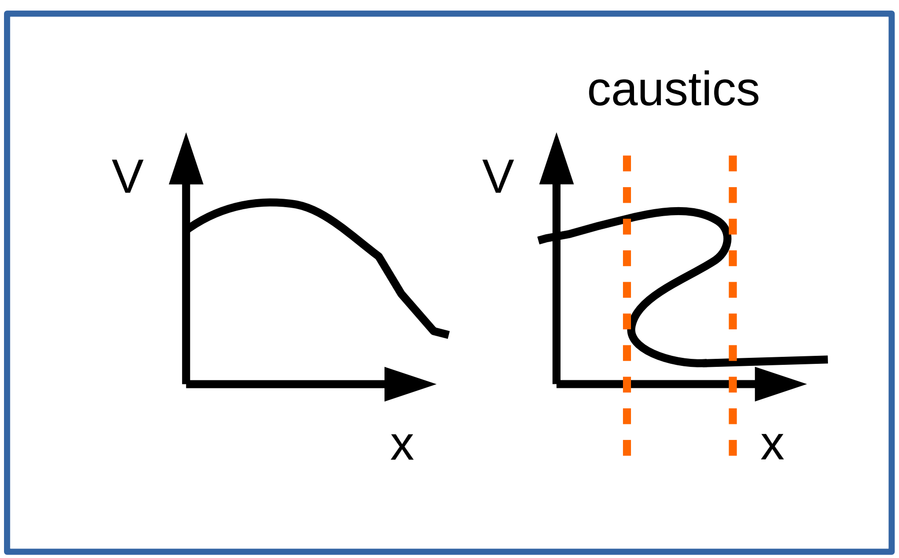
The Vlassov-Poisson Equation

$$0 = \frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla_{\mathbf{x}} f + \frac{\mathbf{F}(\mathbf{x})}{m} \cdot \nabla_{\mathbf{u}} f$$

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

CDM Sheet Properties

- → phase-space is conserved along characteristics
- → It can never tear
- → It can never intersect



Basic Methods

→ Discretization method

→ Initial Conditions

→ Force Calculation

→ Time-stepping

Solving Vlassov-Poisson via MonteCarlo sampling and coarse-graining

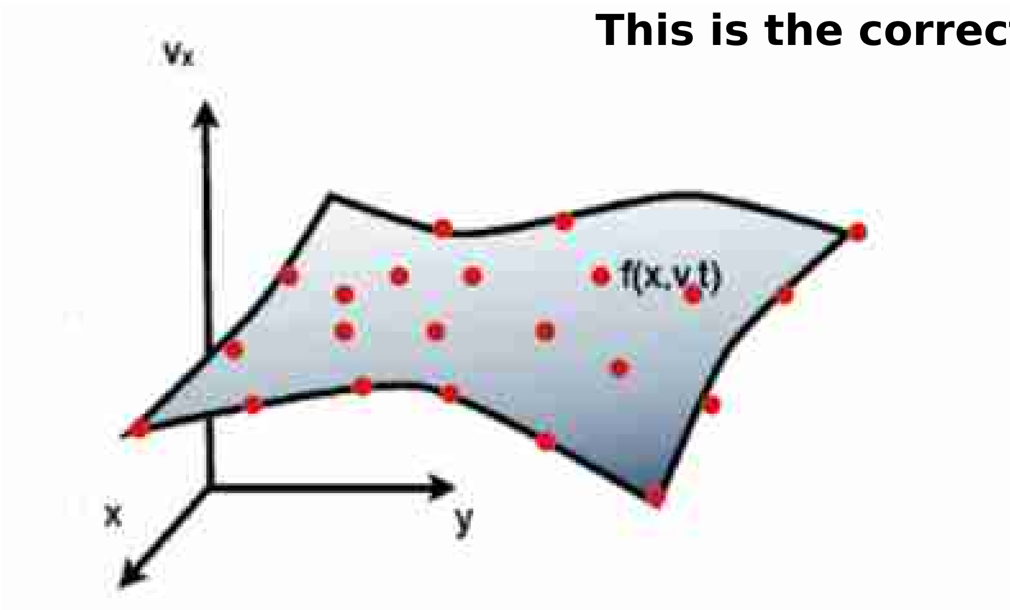
The "method of characteristics" is used to solve the Vlassov-Poisson partial differential equation.

$$\dot{\mathbf{x}}_c = \frac{\mathbf{v}_c}{a^2}$$

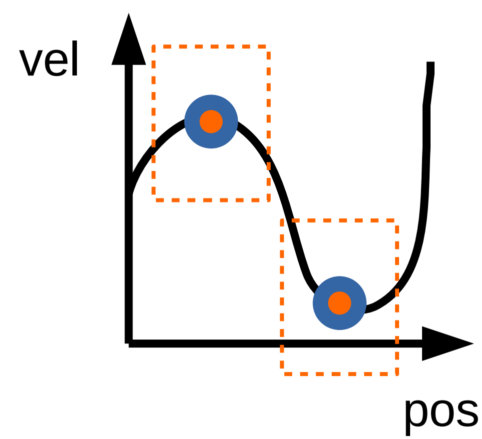
$$\dot{\mathbf{v}}_c = -\nabla_{\mathbf{x}} \phi \big|_{\mathbf{x}_c}$$

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

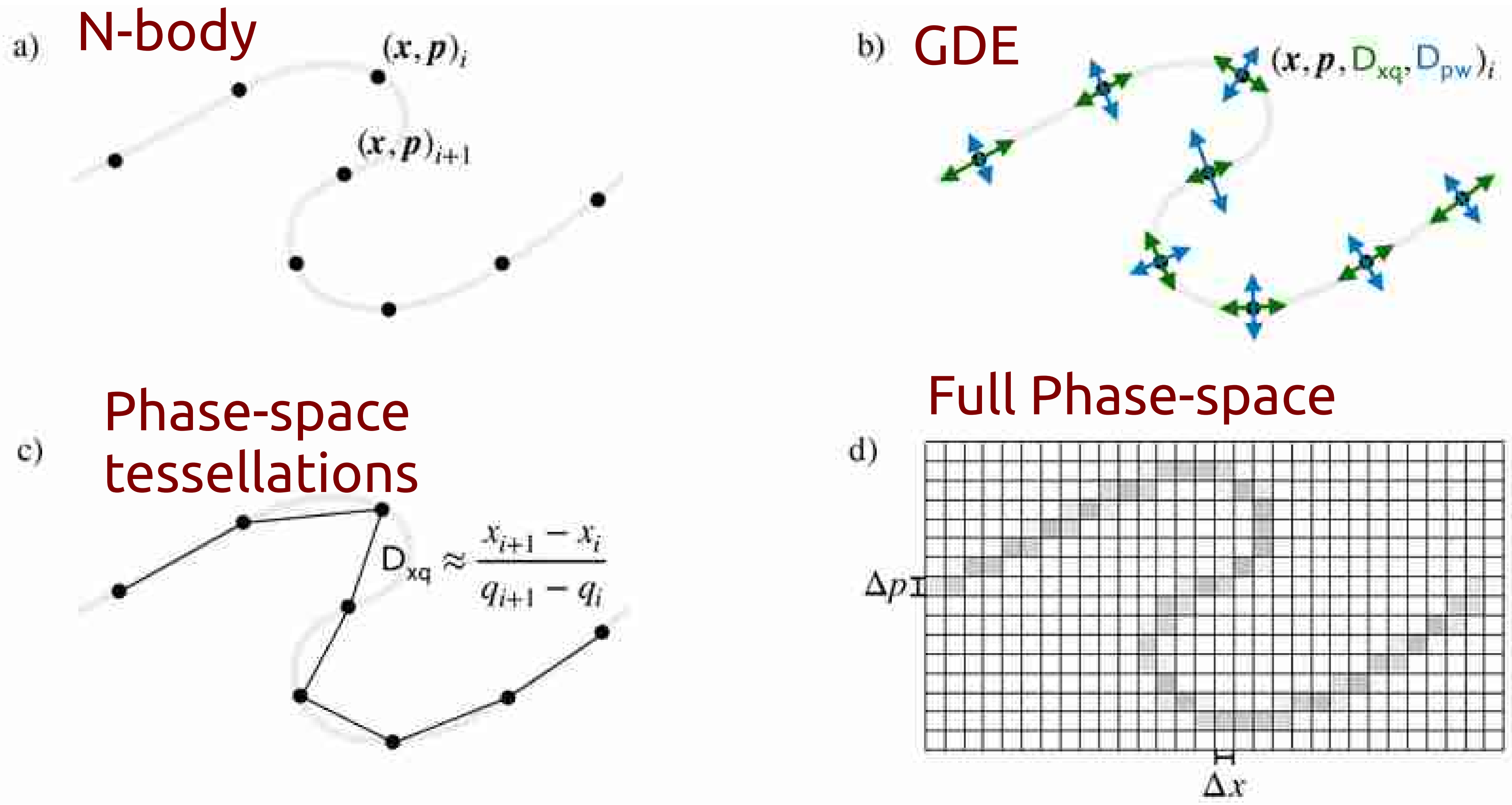
This is the correct solution as N goes to infinity



N-body simulation particle



Alternative discretization methods



The N-body equation

An apparently simple equation, is in fact quite hard to solve and very quickly becomes nonlinear

$$\ddot{\mathbf{x}} = -G \sum_{i}^{N} \frac{m_i \left(\mathbf{x} - \mathbf{x}_i\right)}{\left|\mathbf{x} - \mathbf{x}_i\right|^3}$$

Relevant in many areas of Astrophysics:

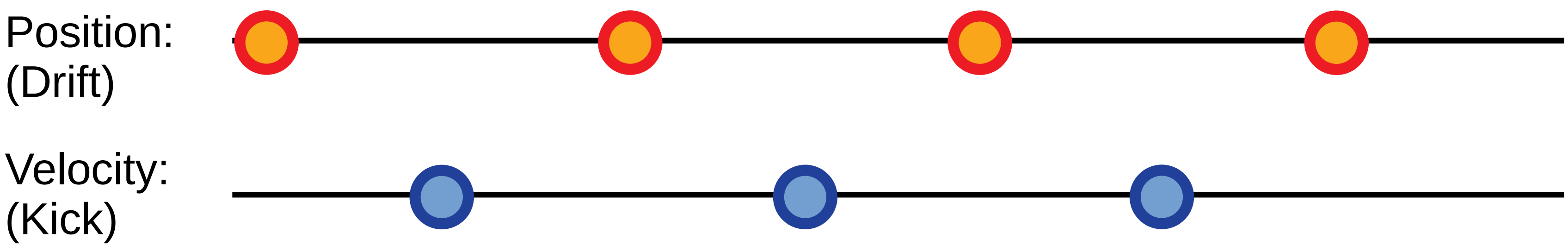
1) Motions of planets
2) Globular clusters
3) Stars in galaxies
4) Galaxies in clusters
5) Large-scale structure

Time-stepping

A leapfrog scheme (aka position/velocity Verlet) is 2nd order accurate but preserves the simplicity of 1st order integrators

$$\ddot{\mathbf{x}} = -G \sum_{i}^{N} \frac{m_i (\mathbf{x} - \mathbf{x}_i)}{|\mathbf{x} - \mathbf{x}_i|^3}$$

$$\mathbf{x}^{l+1} = \mathbf{x}^{l} + \Delta t\, \mathbf{v}^{l+1/2} + \mathcal{O}(\Delta t^3)$$
$$\mathbf{v}^{l+1/2} = \mathbf{v}^{l-1/2} + \Delta t\, \mathbf{a}^{l} + \mathcal{O}(\Delta t^3)$$



Advantages:

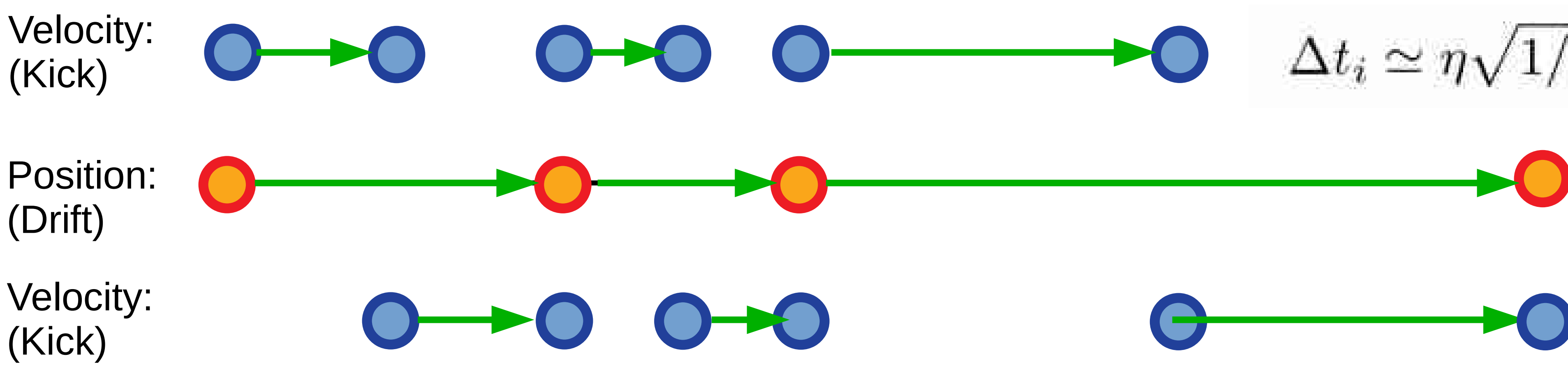
- **Time-reversible**
- **Conserves angular momentum exactly (not energy though) in a spherically symmetric potential.**
- **Symplectic (preserves phase-space volume: det(J)=1)**

Time-stepping

A leapfrog scheme (aka position/velocity Verlet) is 2nd order accurate but preserves the simplicity of 1st order integrators

$$\ddot{\mathbf{x}} = -G \sum_{i}^{N} \frac{m_i (\mathbf{x} - \mathbf{x}_i)}{|\mathbf{x} - \mathbf{x}_i|^3}$$

$$\mathbf{x}^{l+1} = \mathbf{x}^{l} + \Delta t\, \mathbf{v}^{l+1/2} + \mathcal{O}(\Delta t^3)$$
$$\mathbf{v}^{l+1/2} = \mathbf{v}^{l-1/2} + \Delta t\, \mathbf{a}^{l} + \mathcal{O}(\Delta t^3)$$



$$\Delta t_i \simeq \eta \sqrt{1/a_i}$$

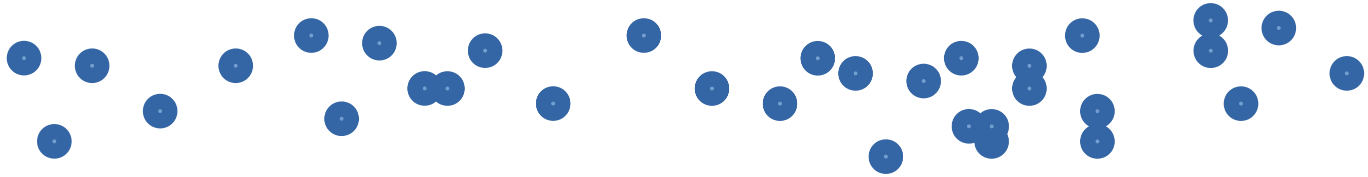
dT3 terms can be important, specially for long steps:**

1) Comoving Lagrangian Accelerator

2) FastPM

Force calculation

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

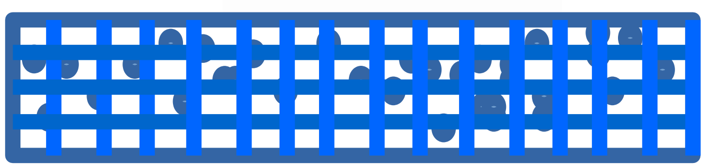


$$\ddot{\mathbf{x}} = -G\sum_{i}^{N} \frac{m_i\,(\mathbf{x}-\mathbf{x}_i)}{|\mathbf{x}-\mathbf{x}_i|^3}$$

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

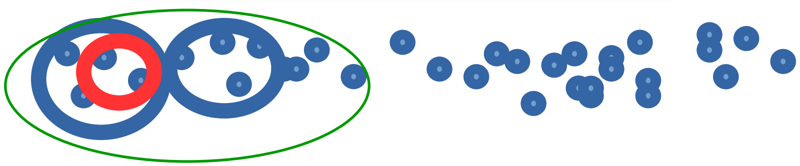
Particle-Mesh

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



Hierarchical-Trees

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



Force calculation: softening length

A regularization of force calculations is needed to avoid unrealistic close-encounters and large-angle scatterings among particles

$$\ddot{\mathbf{x}} = -G \sum_{i}^{N} \frac{m_i (\mathbf{x} - \mathbf{x}_i)}{|\mathbf{x} - \mathbf{x}_i|^3} \quad \Rightarrow \quad \ddot{\mathbf{x}} = -G \sum_{i}^{N} \frac{m_i (\mathbf{x} - \mathbf{x}_i)}{(|\mathbf{x} - \mathbf{x}_i|^2 + \epsilon^2)^{3/2}}$$

Forces are limited to:

$$\max[|\ddot{\mathbf{x}}|] = \frac{2Gm_i}{3^{3/2}\epsilon^2}$$

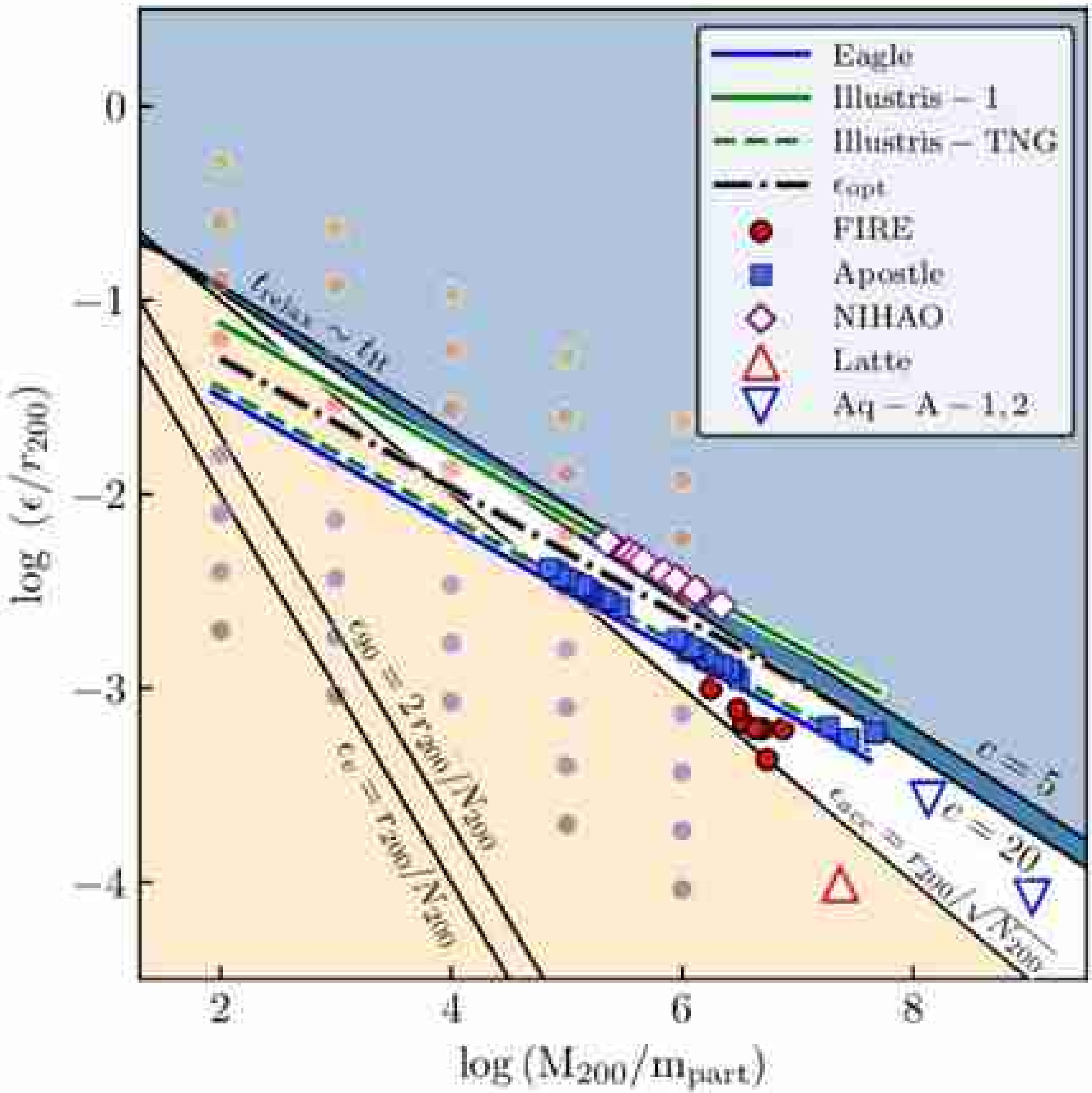
Collisionless Relaxation

- Phase Mixing
- Chaotic Mixing
- Violent Relaxation
- Landau Damping

Wang&White 2007

Force calculation: softening length

Various requirements for an "optimal" choice



Ludlow et al (2019)

$$\epsilon \sim (V/N)^{1/3}$$

$$\epsilon_{\rm cm}(z) \lesssim \left(\frac{3\,\Omega_{\rm DM}}{8\,\pi}\right)^{1/3} \left(\frac{N_{200}}{100}\right)^{1/3} l$$

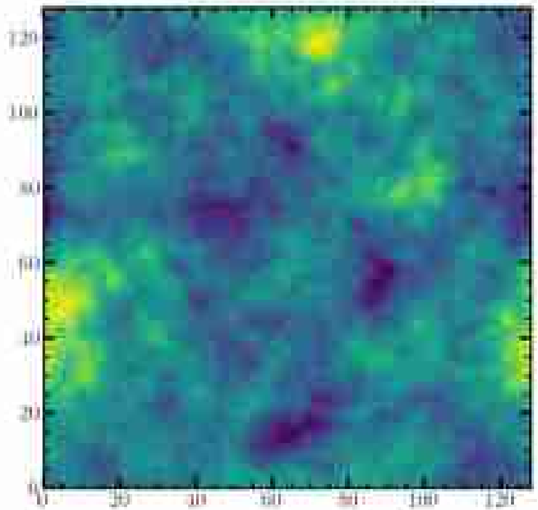
$$\epsilon_{\rm cm} \gg l \left(\frac{3\,\Omega_{\rm DM}}{800\,\pi}\,\frac{1}{N_{200}^2}\right)^{1/3}$$

$$\epsilon_{\rm acc} \gtrsim r_{200}/\sqrt{N_{200}}$$

$$\frac{t_{\rm relax}}{t_{\rm cross}} \approx \frac{N}{8\ln(R/\epsilon)}$$

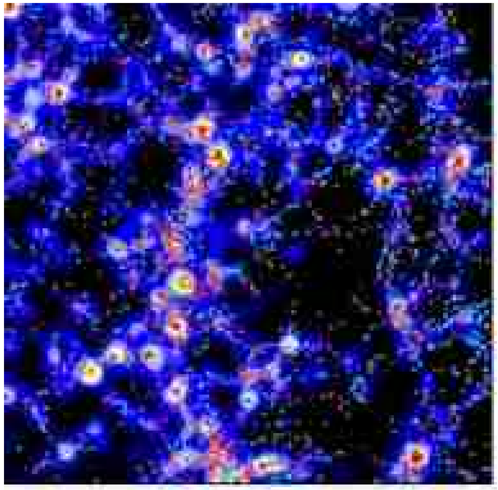
Initial Conditions

Gaussian random field

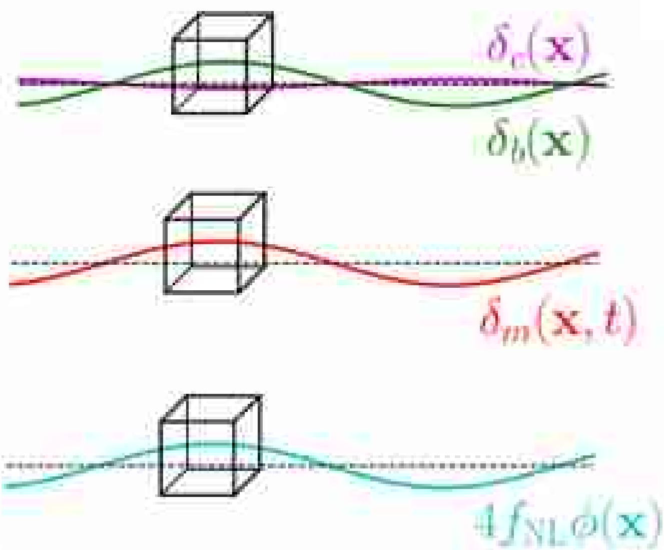


Zoom-in

Constrained Realization

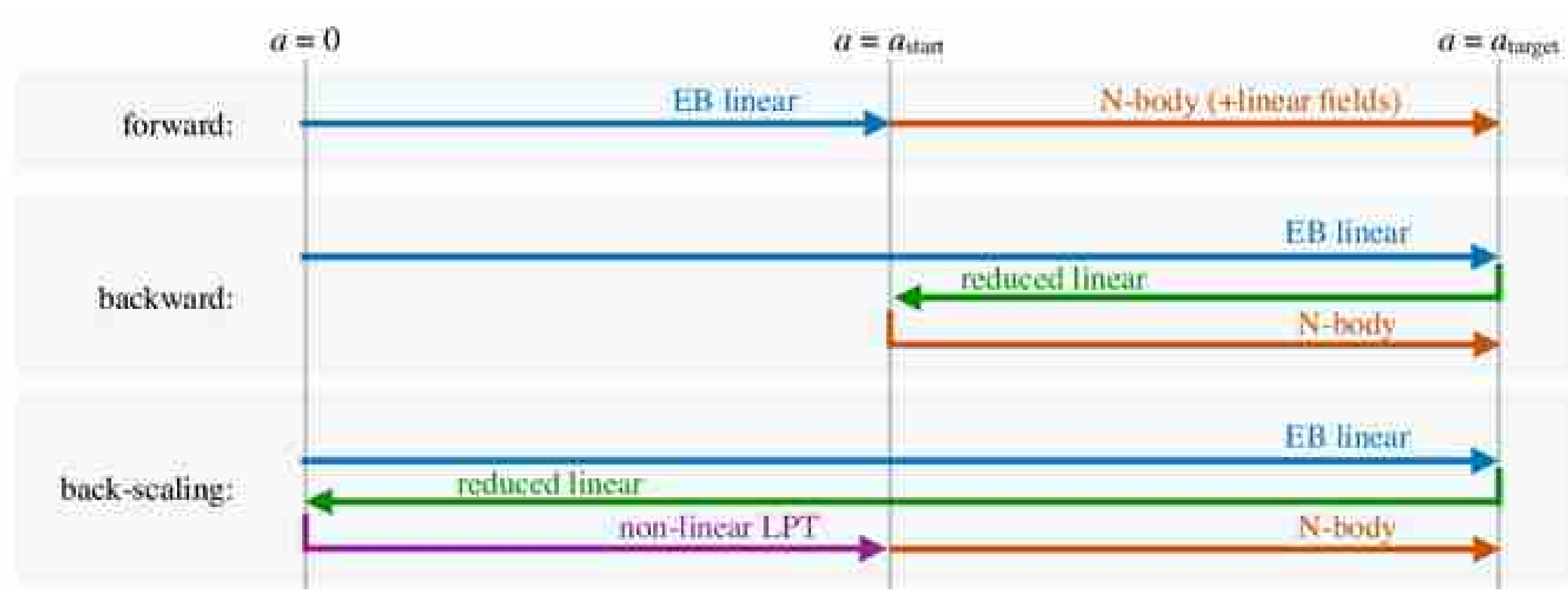


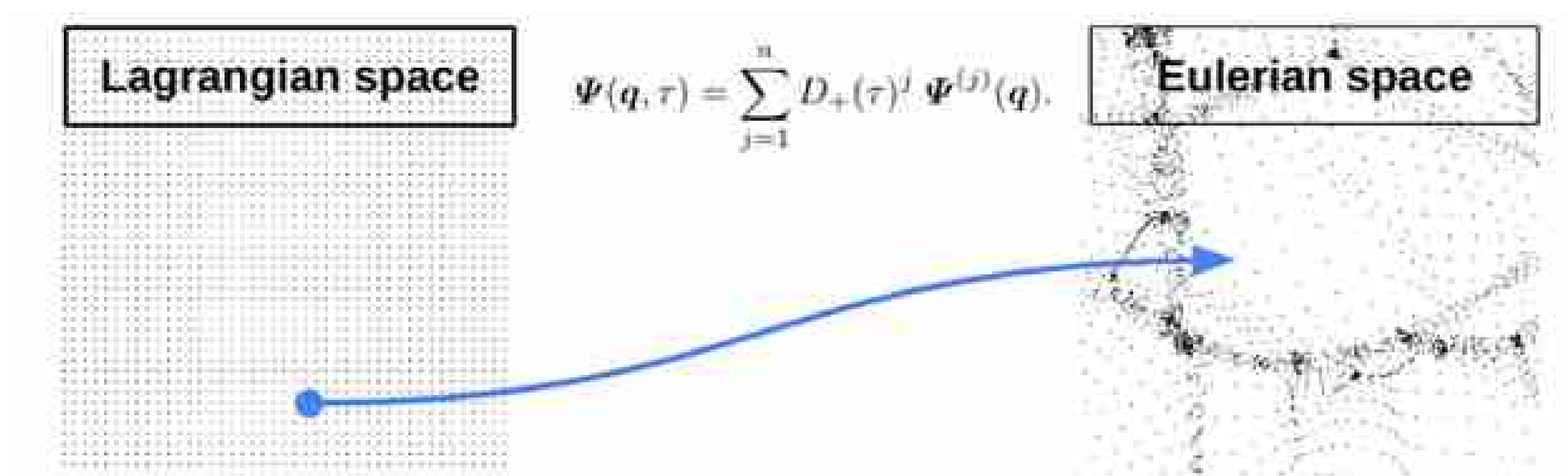
Separate Universe



Initial Conditions

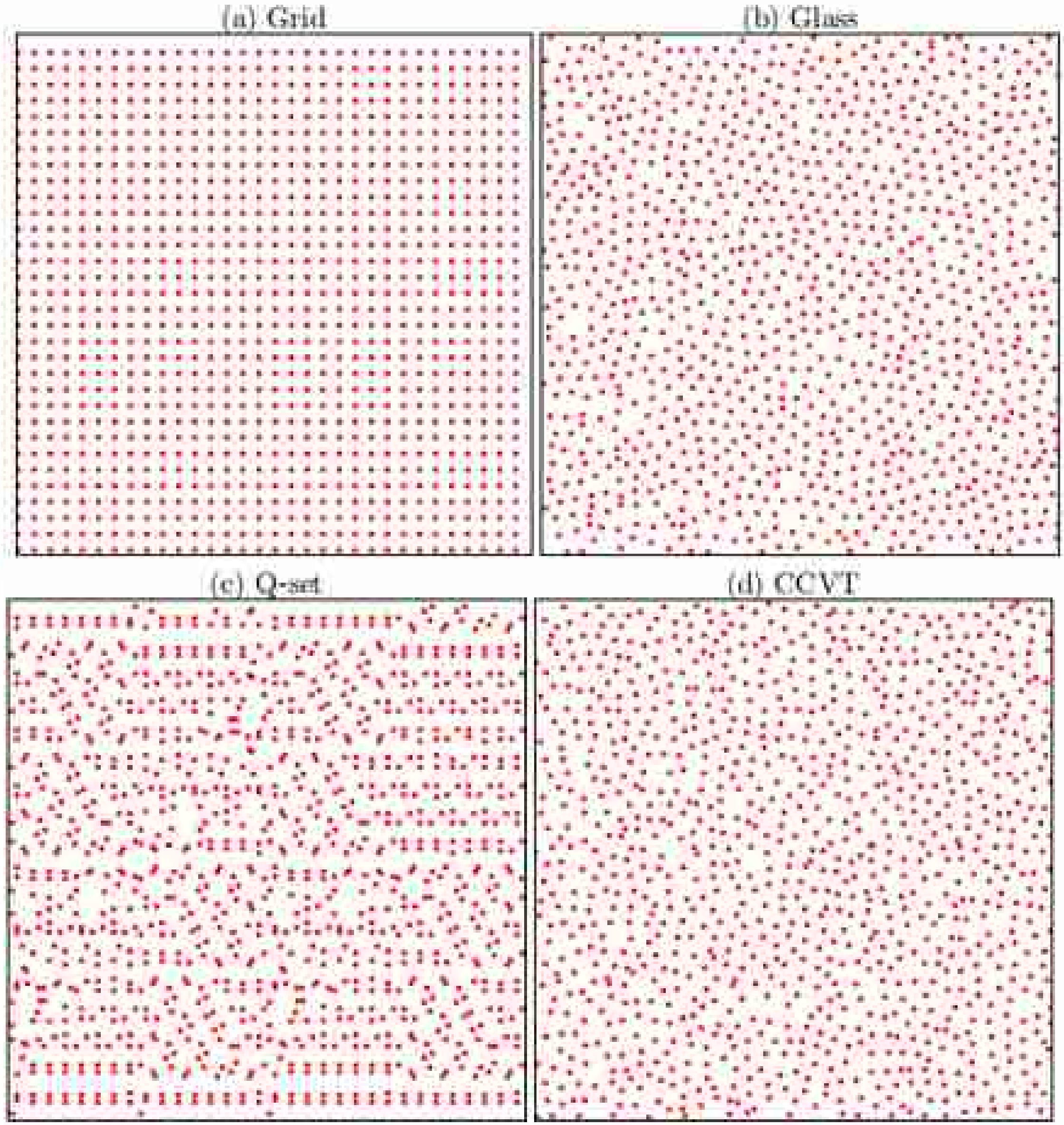
We need to create a field realisation compatible with the statistical properties of the CMB





Initial Conditions

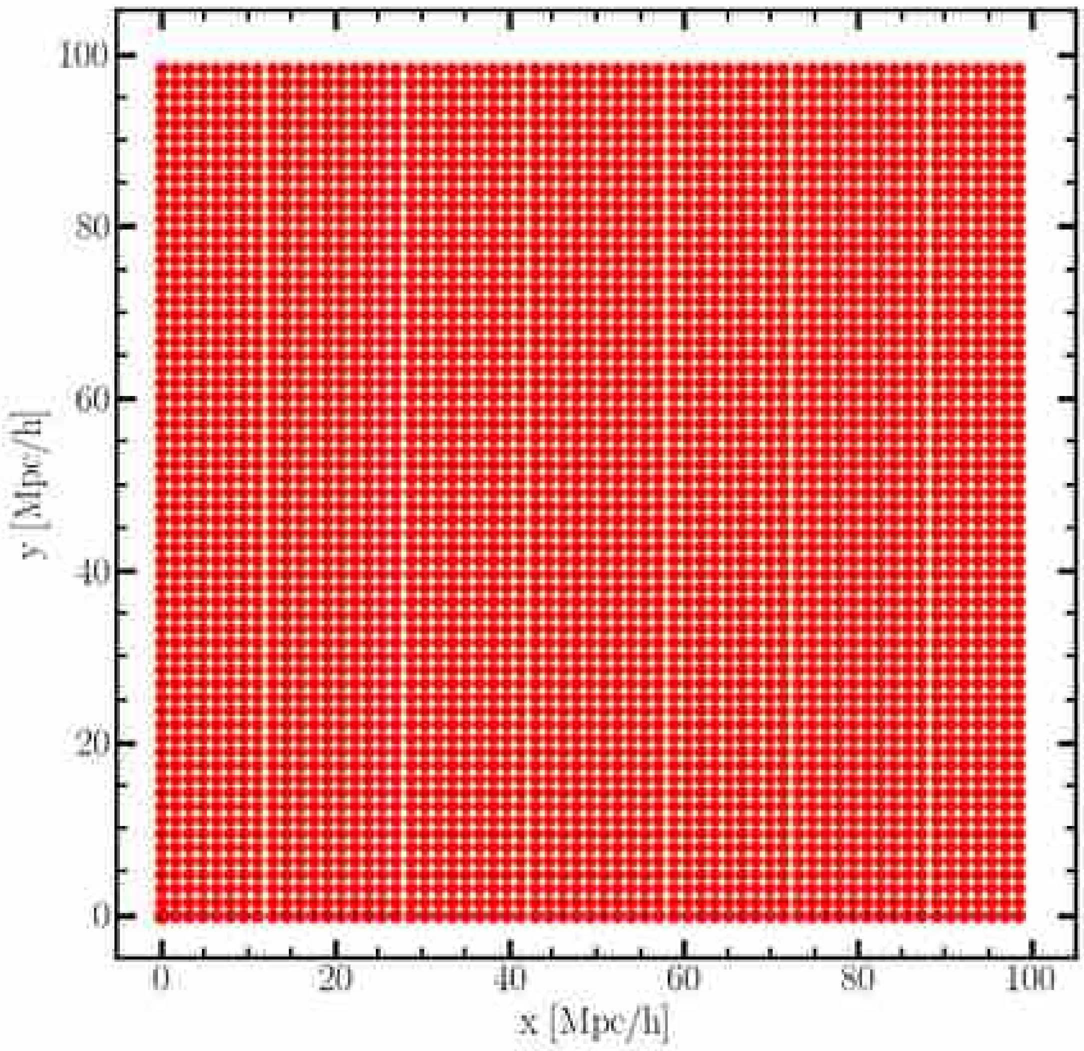
An initially homogeneous (grid or glass usually) are initialized with position and velocity from (L)PT



Globally homogeneous, stable and isotropic particle distribution

Initial Conditions

An initially homogeneous (grid or glass usually) are initialized with position and velocity from (L)PT

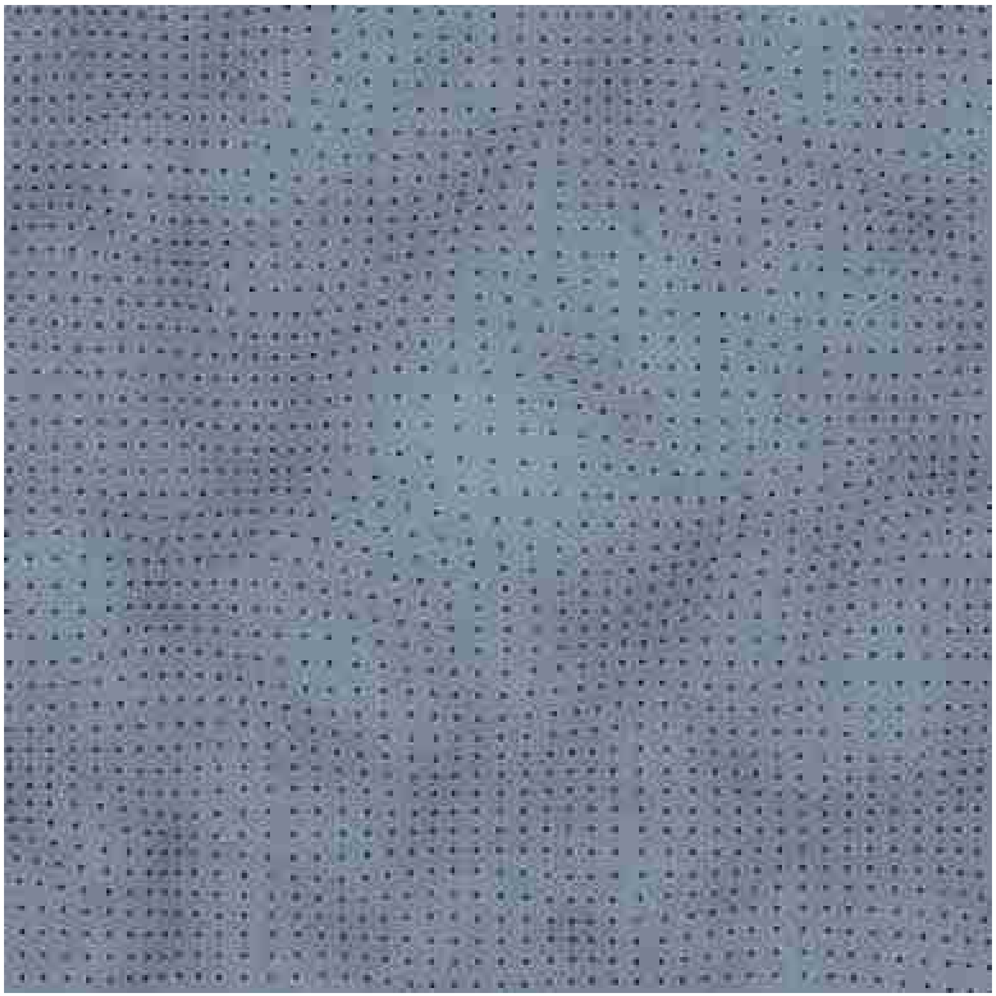


The Zeldovich Approximation

$$\mathbf{v} = -\frac{2a\dot{D}}{3H_0^2\Omega_m}\nabla\phi$$

$$\mathbf{x} = \mathbf{q} - \frac{2aD}{3H_0^2\Omega_m}\nabla\phi$$

The structure of an N-body code



An N-body code

→ Discretise space
→ Compute ICs

Loop over N timesteps

→ Kick velocities dt/2

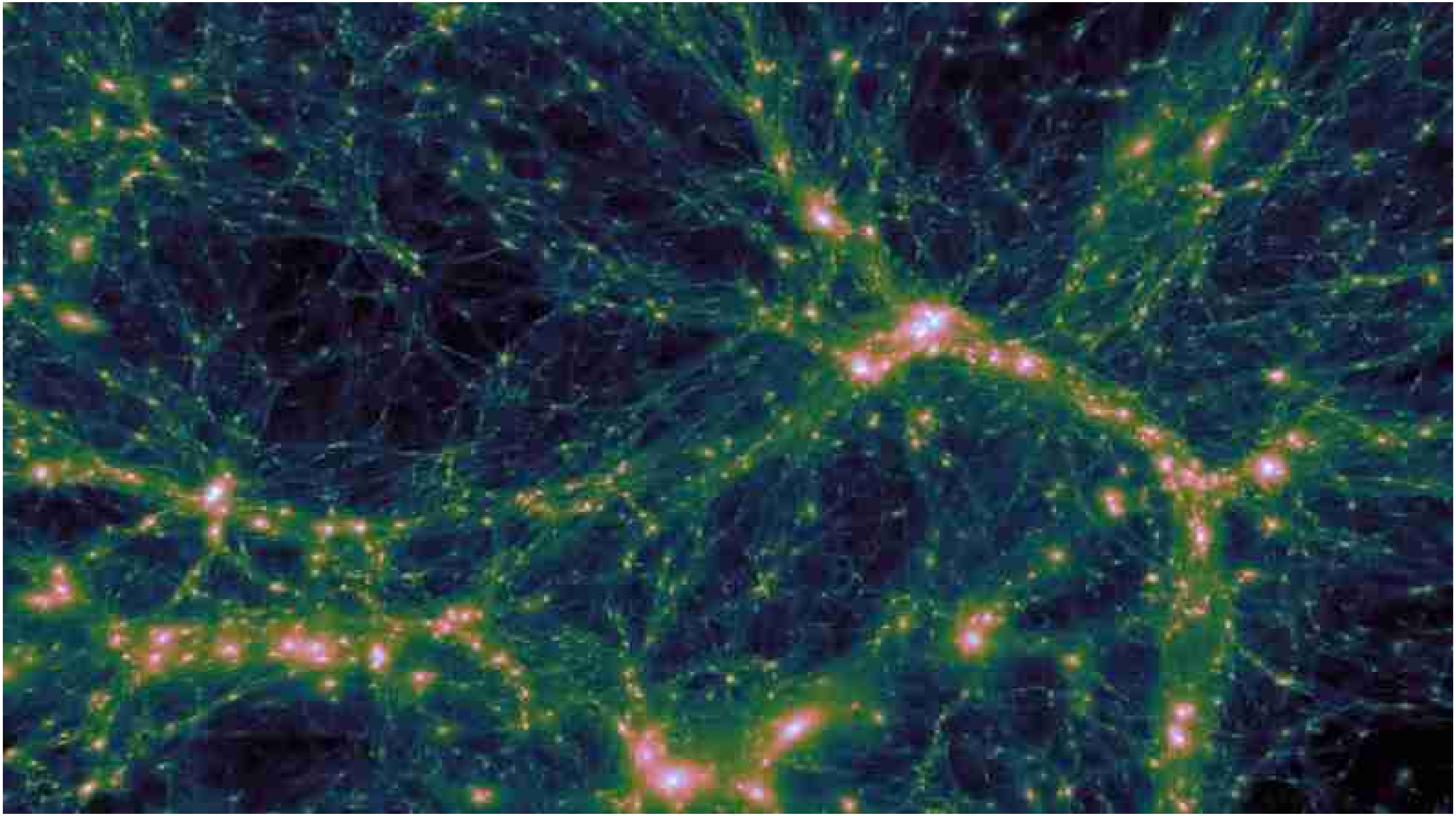
→ Drift particles dt

→ Compute forces

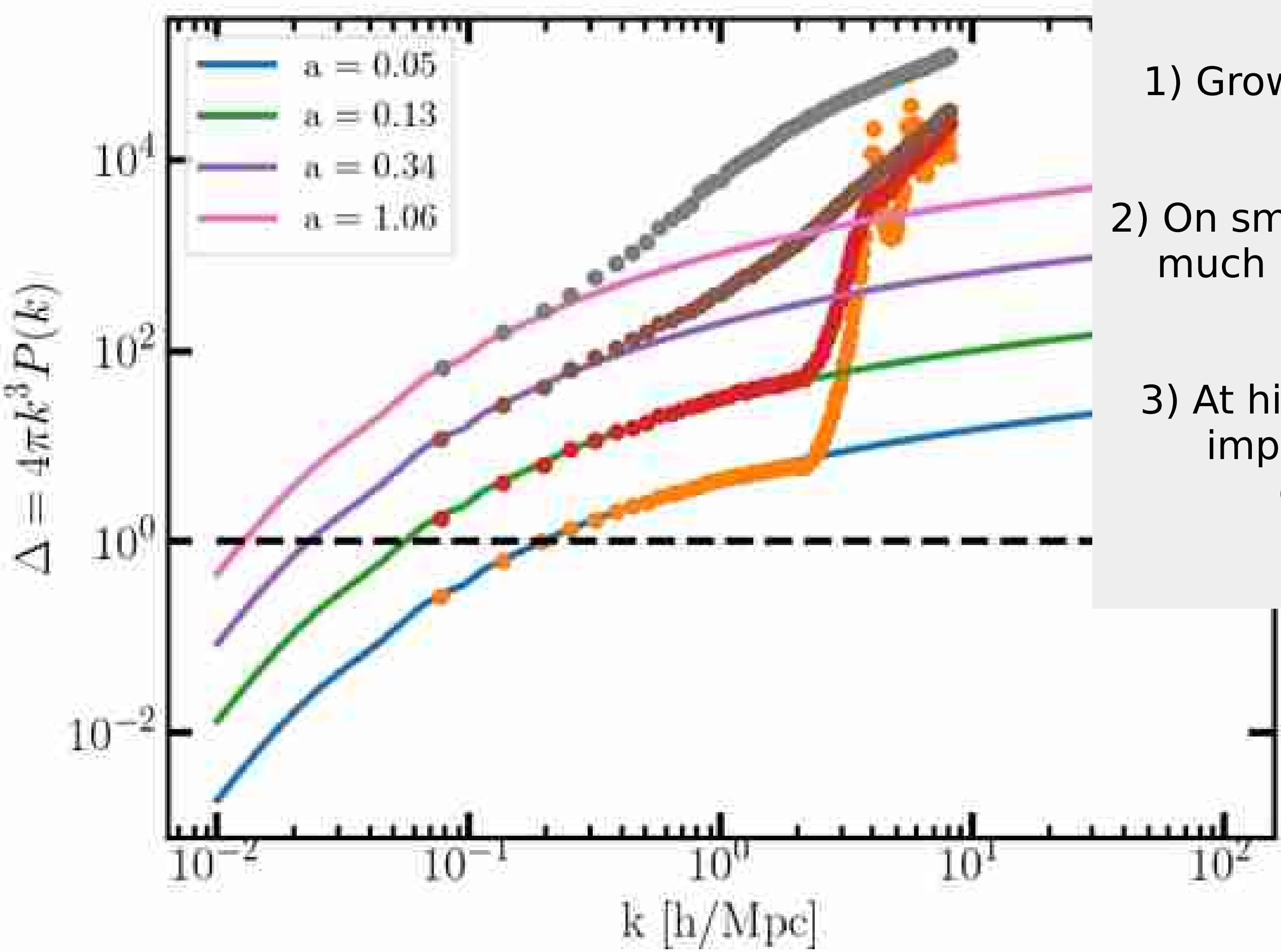
→ Kick velocities dt/2

→ estimate a new dt

Leap frog symplectic integrator



The nonlinear matter power spectrum



1) Growth is scale-independent on large scales

2) On smaller scales, growth happens much more quickly than in linear theory

3) At high k and high redshifts, the imprint of the initial particle distribution is visible

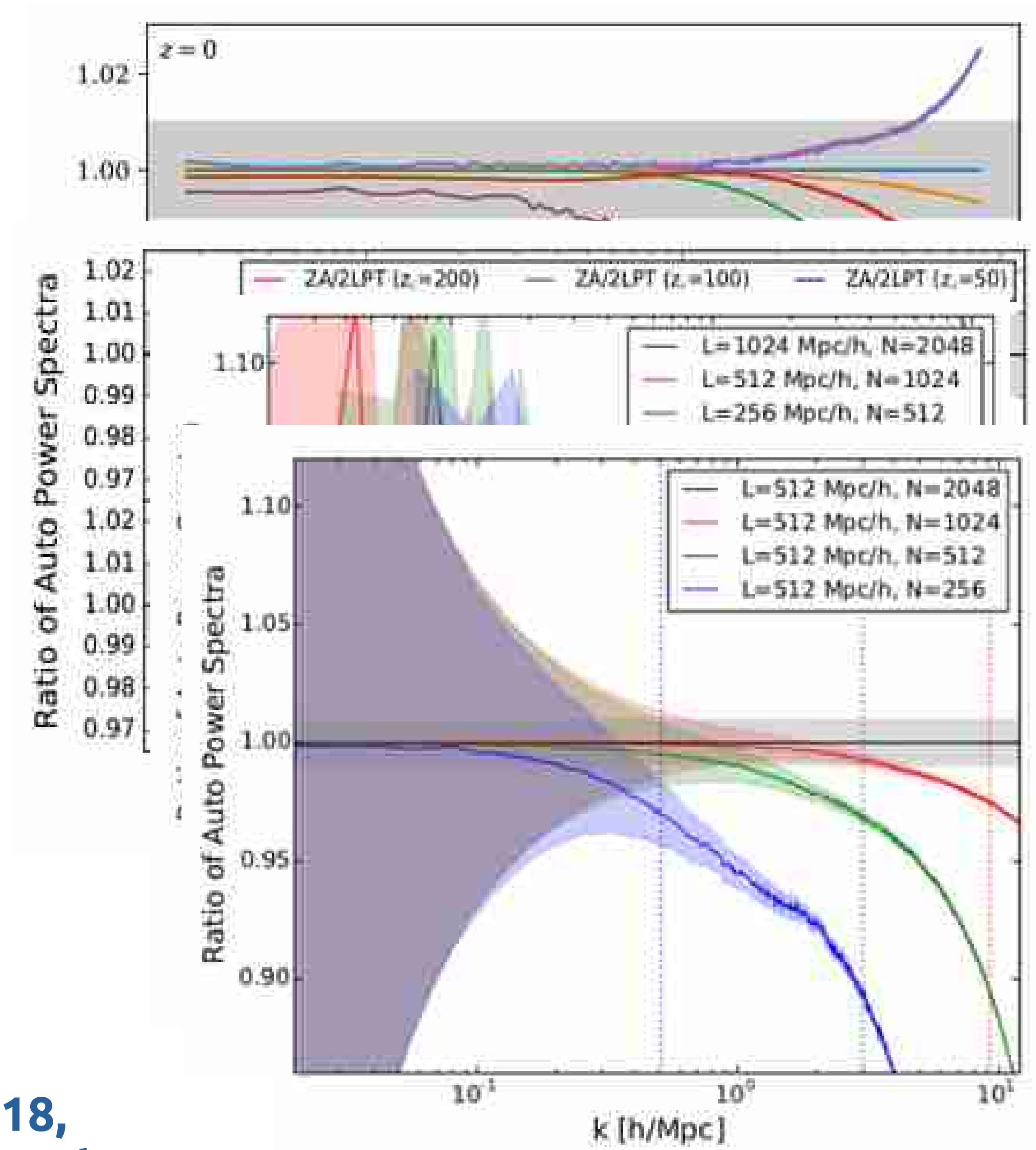
The challenge of high accuracy

i) Numerical errors

ii) Initial Conditions

iii) Box size

iv) Mass resolution



→ Future surveys (LSST, DESI, EUCLID) require simulations L ~ 4Gpc/h and 4+ trillion particles

Schneider+ 2016, Garrison et al 2018, Angulo+2021, Michaux+2021, Springel+ 2022

Lectures on simulating the formation of structure in the Universe

→ Basic Methods & Algorithms

→ **Beyond the simplest models**

→ State of the Art

→ The future

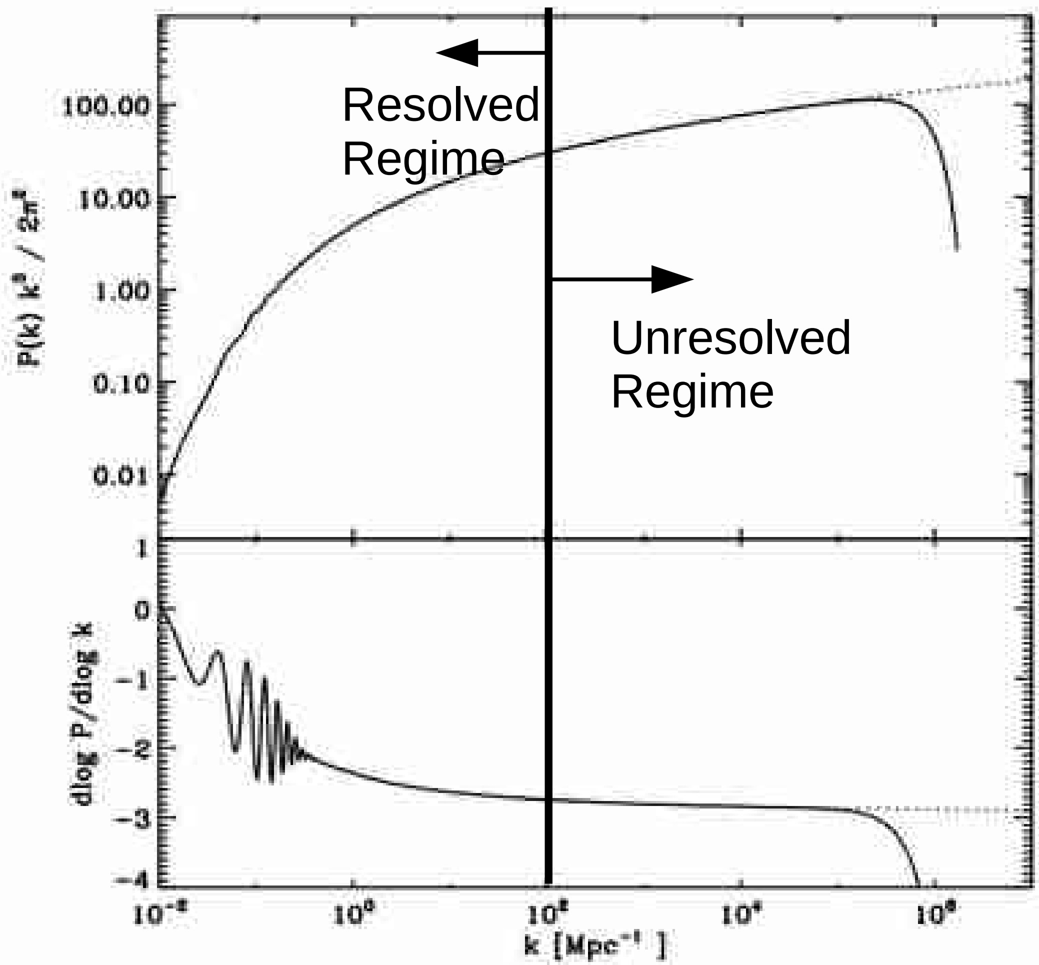
Simulating departures from LCDM

Many flavors of cosmological simulations dropping one or more of the assumptions of the simplest case

Cold	→ Warm
Collisionless	→ Self-Interacting
DM only	→ +baryons, neutrinos
Gaussian IC	→ Primordial NG
GR	→ Modified gravity
Classical particles	→ PBH /axions/wave DM

CDM at the free-streaming scale

Full hierarchy spans 21 orders of magnitude in mass

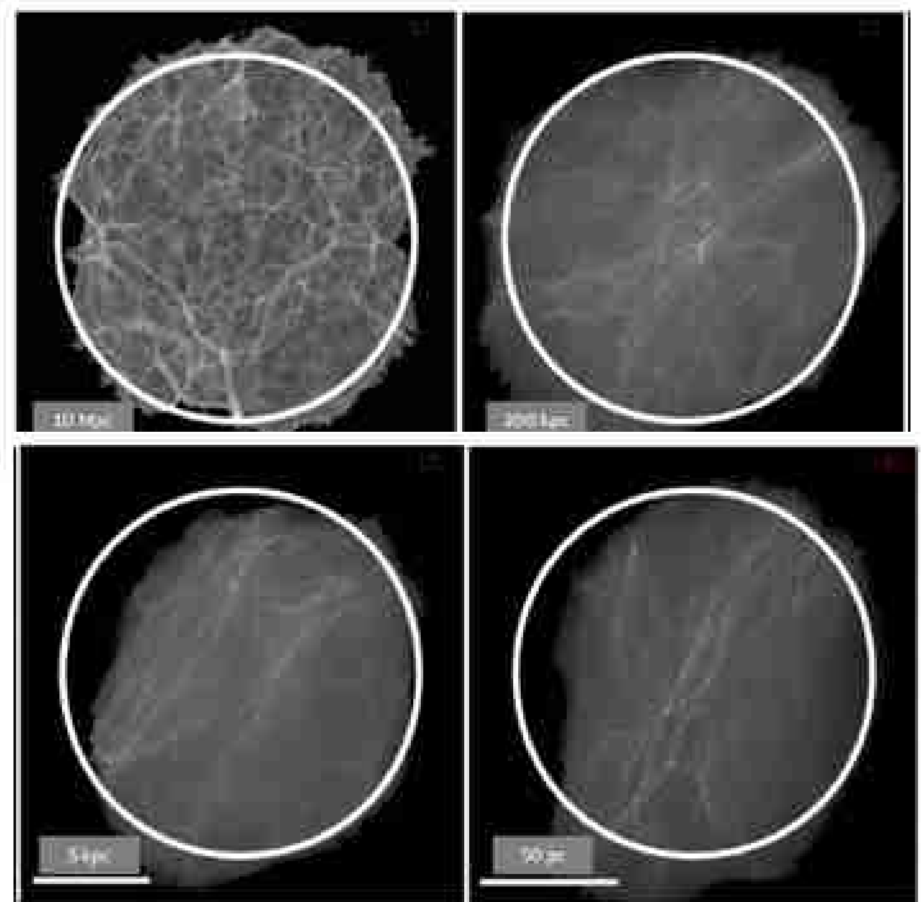


Angulo & White 2010

The 100GeV neutralino-DM power spectrum

1) Free streaming suppresses fluctuations below 1 earth mass

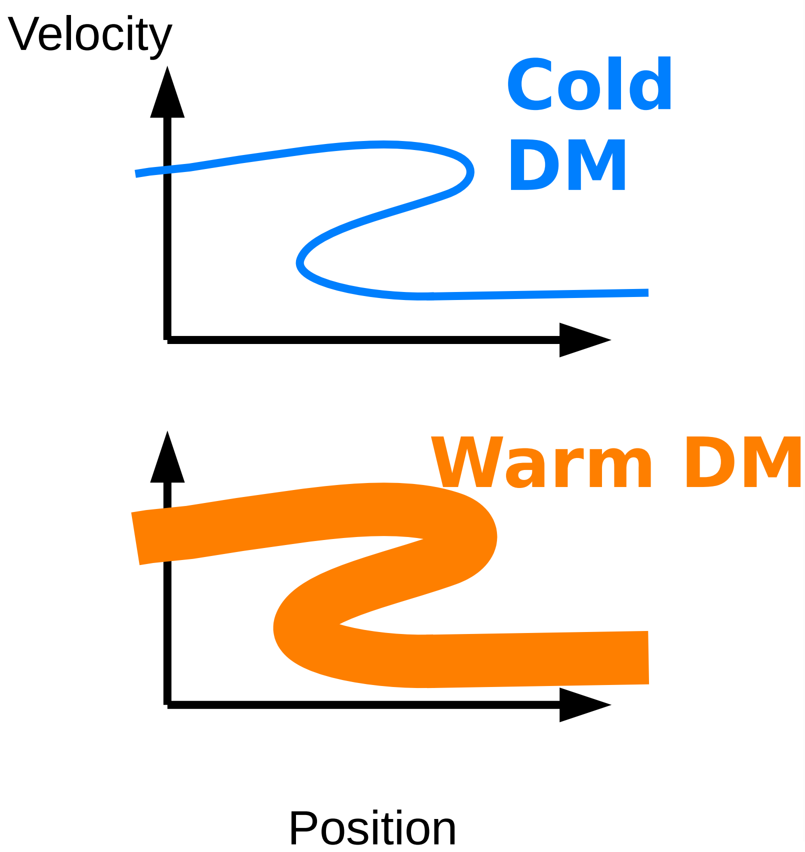
2) A full CDM simulation requires 1e21 particles.

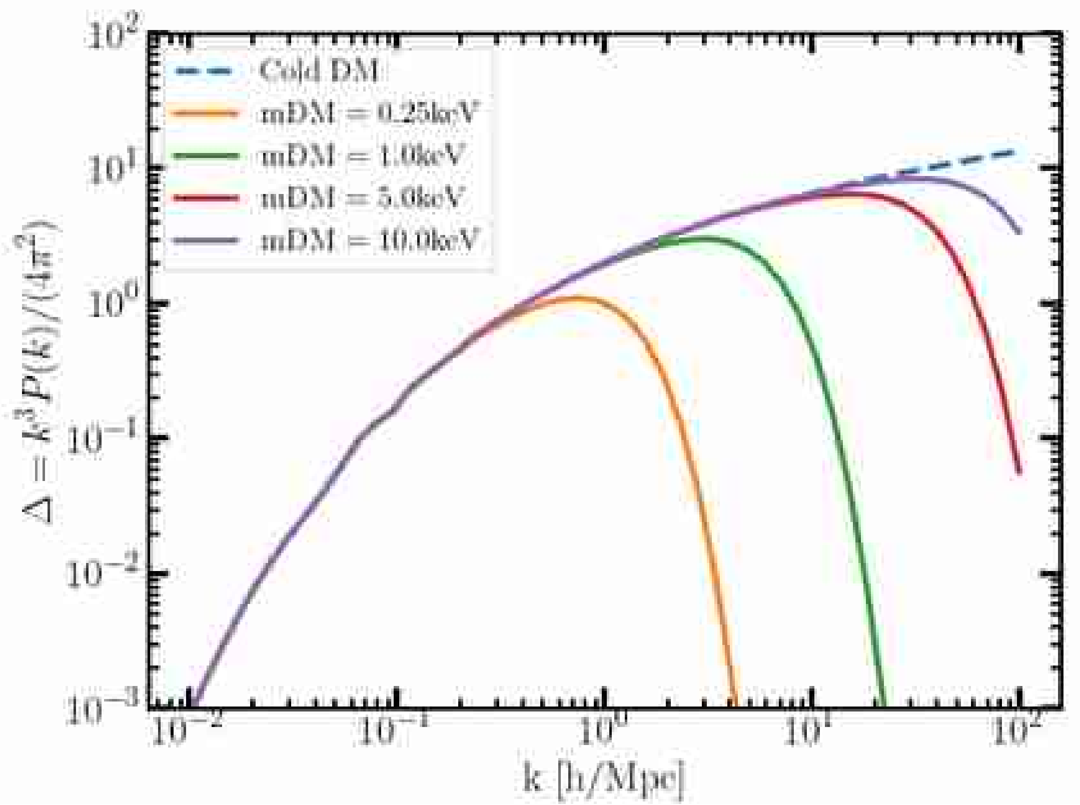


VVV; Wang et al 2021

Warm Dark Matter

Free streaming of particles out of overdensities erases primordial fluctuations on small scales





Since the extent in velocity space is quite small, typical N-body simulations assume WDM as Cold with a modified initial power spectrum

Warm Dark Matter

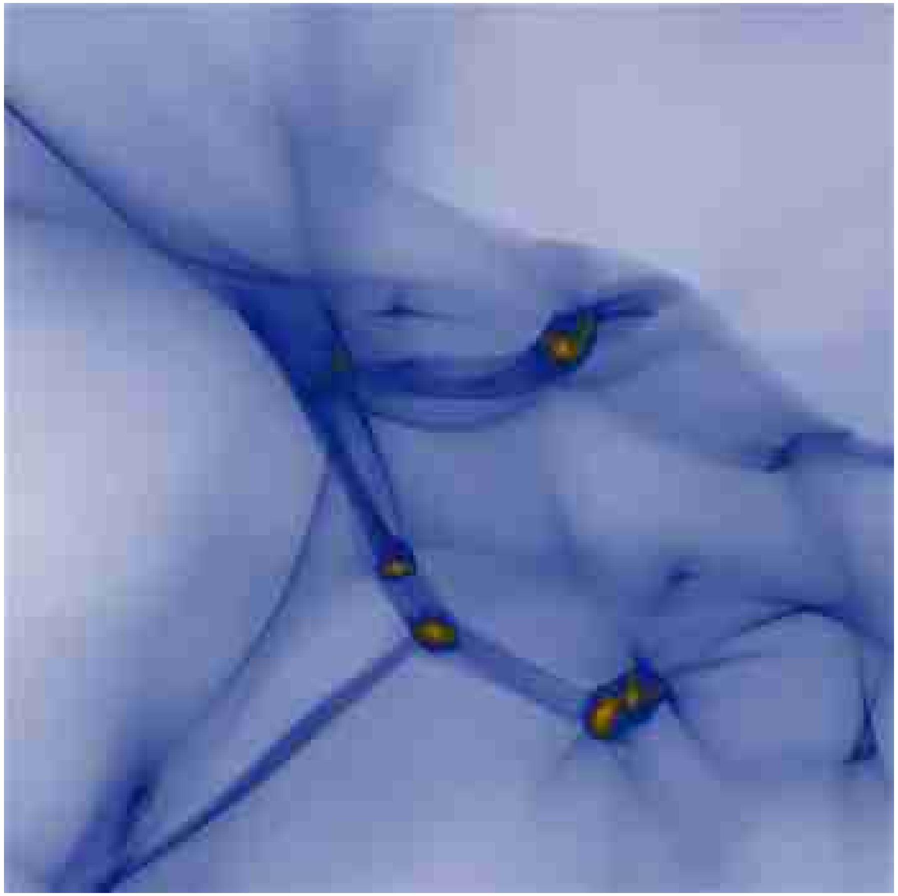
Free streaming of particles out of overdensities erases primordial fluctuations on small scales

No bottom-up formation

Genuine 1D and 2D structures

Halos have no progenitors beyond some time

Very rapid mass growth during formation



Angulo et al (2012)

Warm Dark Matter

Halos are less concentrated, have less substructures, and are expected to have (very small) cores

Lovell et al (2014)

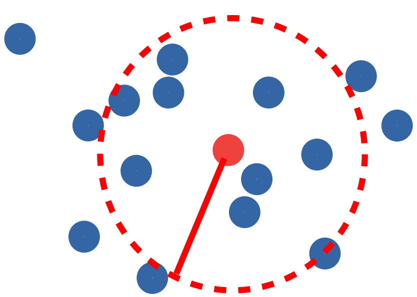
Self-interacting Dark Matter

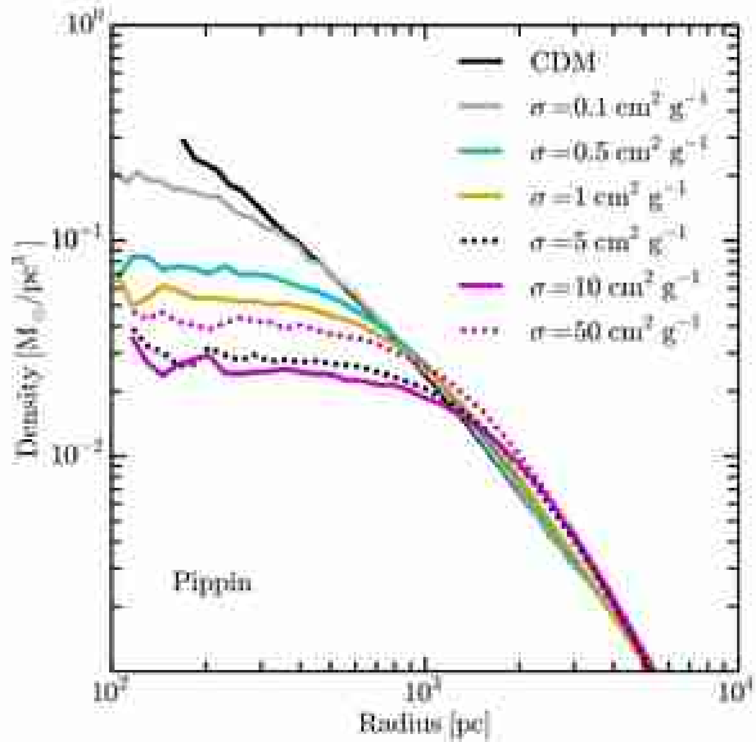
A large cross-section reduces the central density in dark matter halos and make them rounder

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{ma^2} \cdot \nabla_x f - m \nabla_x \phi \cdot \nabla_p f = \Gamma_{\rm in} - \Gamma_{\rm out},$$

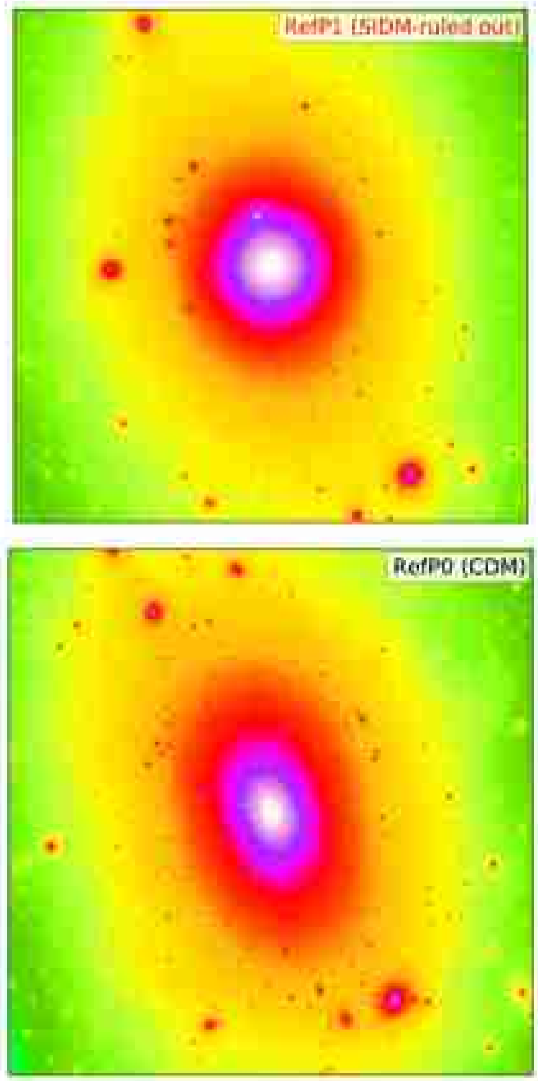
$$\Gamma_{out} = f(\mathbf{x}, \mathbf{u}) \int \mathrm{d}\mathbf{u}\, f_v(\mathbf{x}, \mathbf{u}, t) \rho(\mathbf{x}) \frac{\sigma_\chi}{m} |\mathbf{u} - \mathbf{u}_i|$$

$$P_{ij} = \rho_{ij} |\mathbf{v}_i - \mathbf{v}_j| (\sigma/m) \Delta t$$





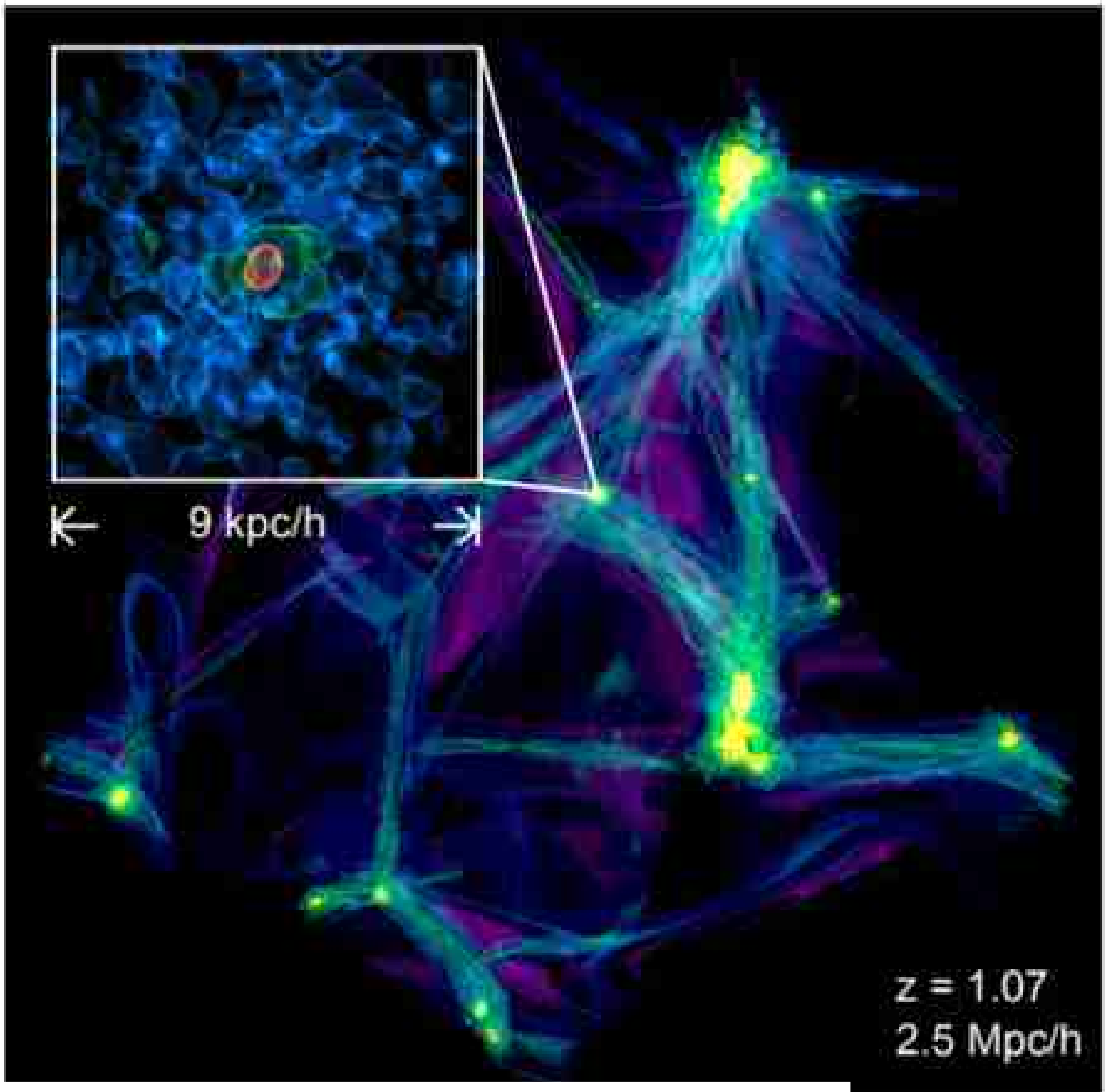
Elbert et al (2015)

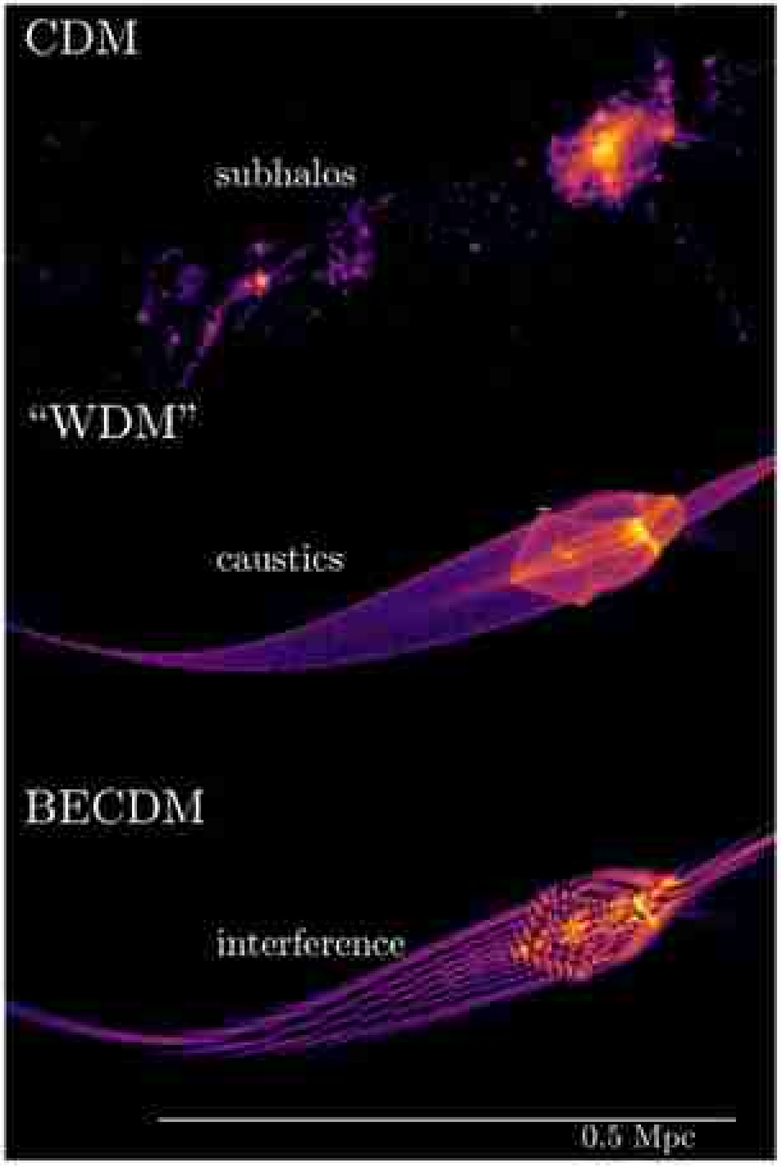


Vogelsberger et al (2012)

Axion Dark Matter

Indistinguishable from CDM on large scales, but small-mass haloes are erased and a "soliton" at the center of halos





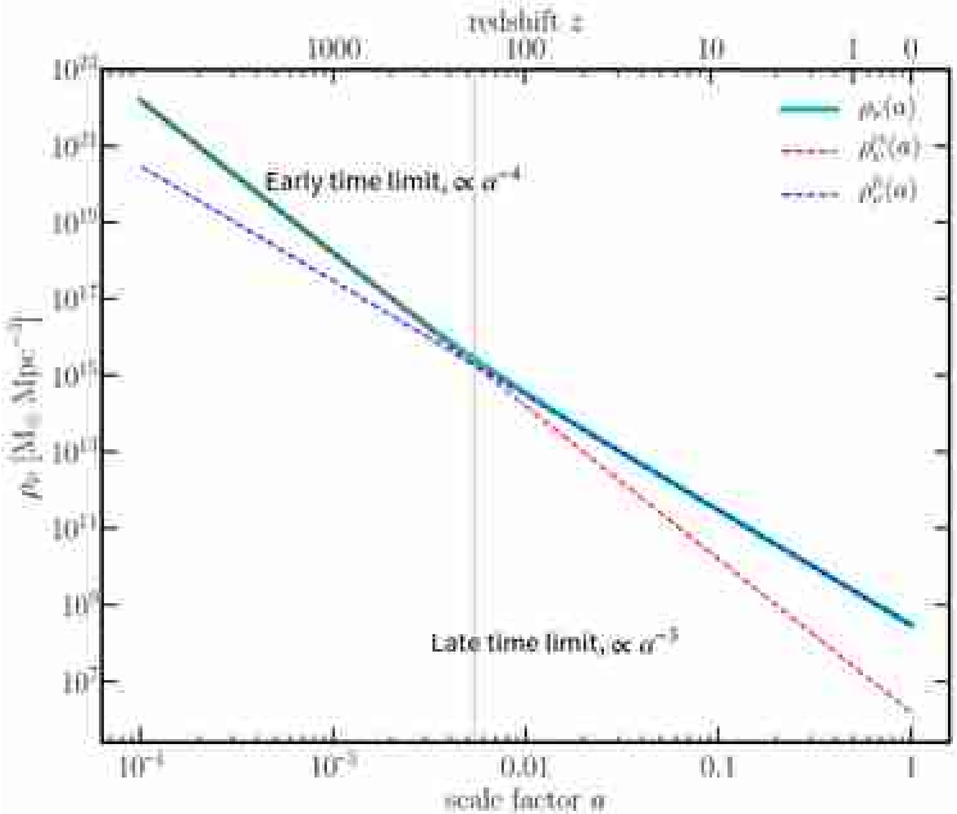
Mocz+ 2020, Schive+ 2016

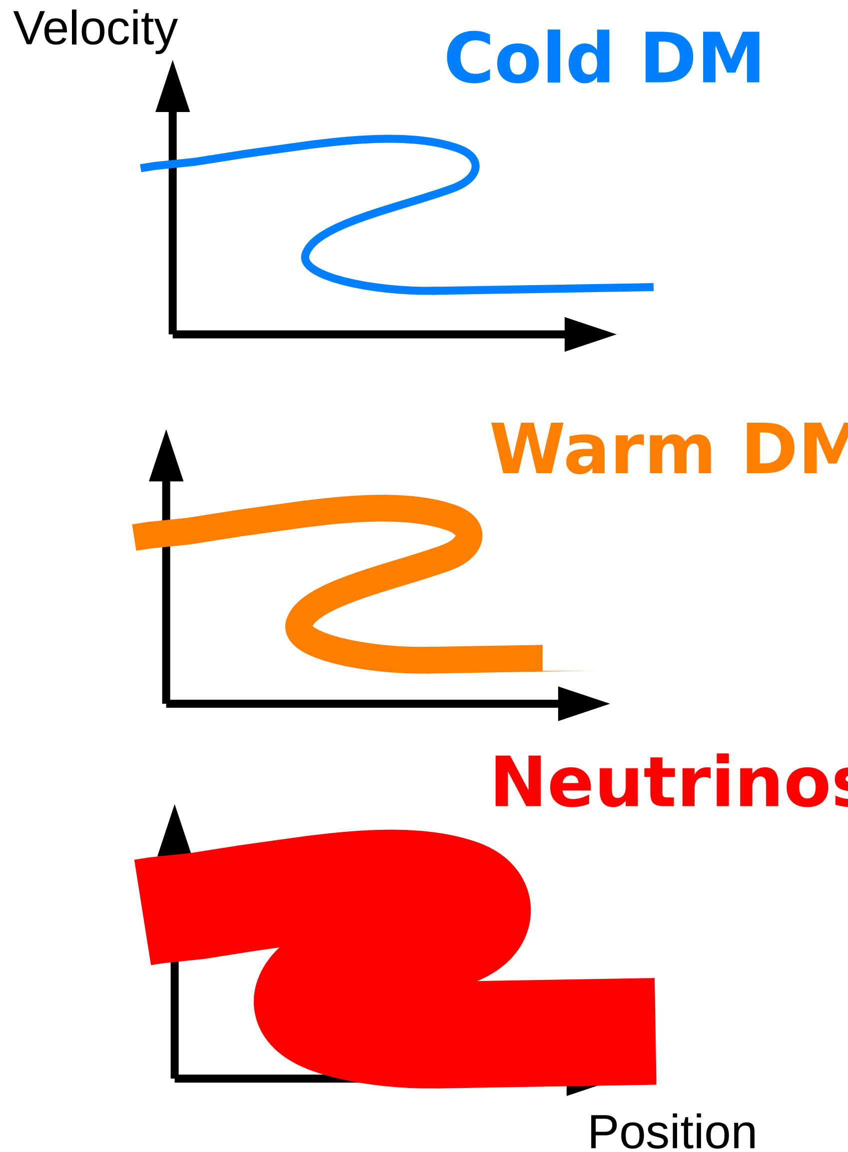
Massive Neutrinos

Current constraints indicate masses between 0.05 and 0.15eV, even with hints of detection

$$f_\nu(\mathbf{u}, t) = (\exp(\mathbf{u}c/k_b T_\nu) + 1)^{-1}$$

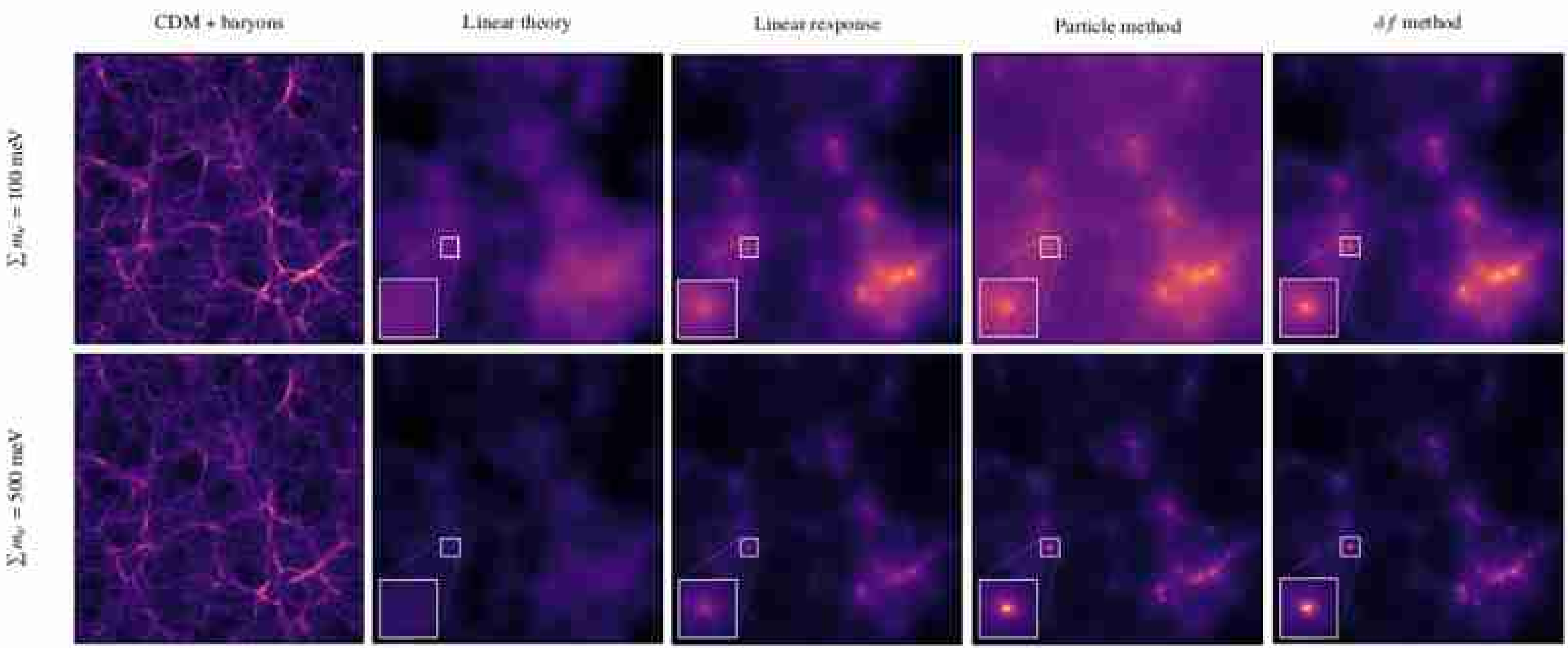
$$\Omega_\nu h^2 = \frac{1}{93.14eV} \sum_i m_{\nu,i}$$





Massive Neutrinos

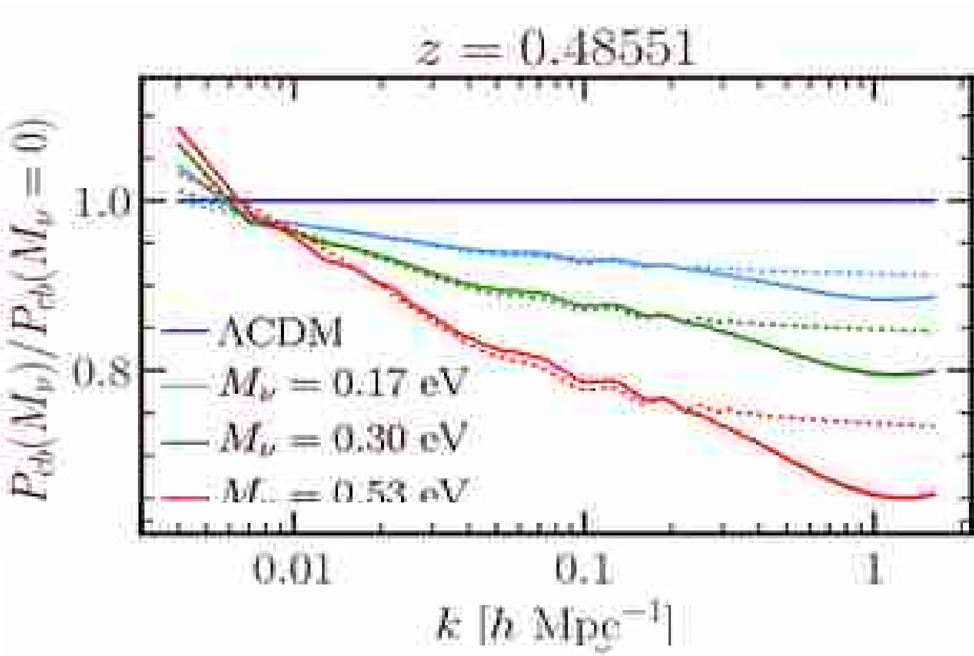
Neutrinos suppress structure growth on small scales

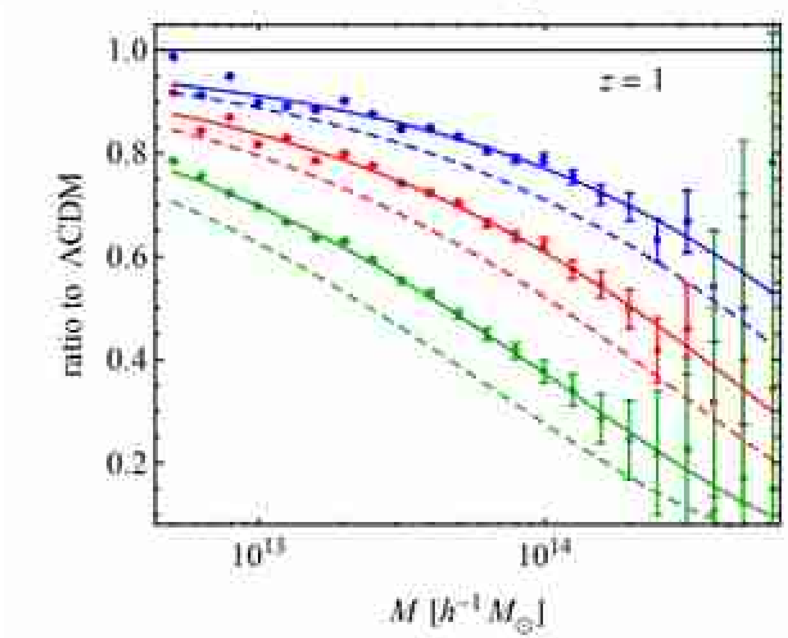


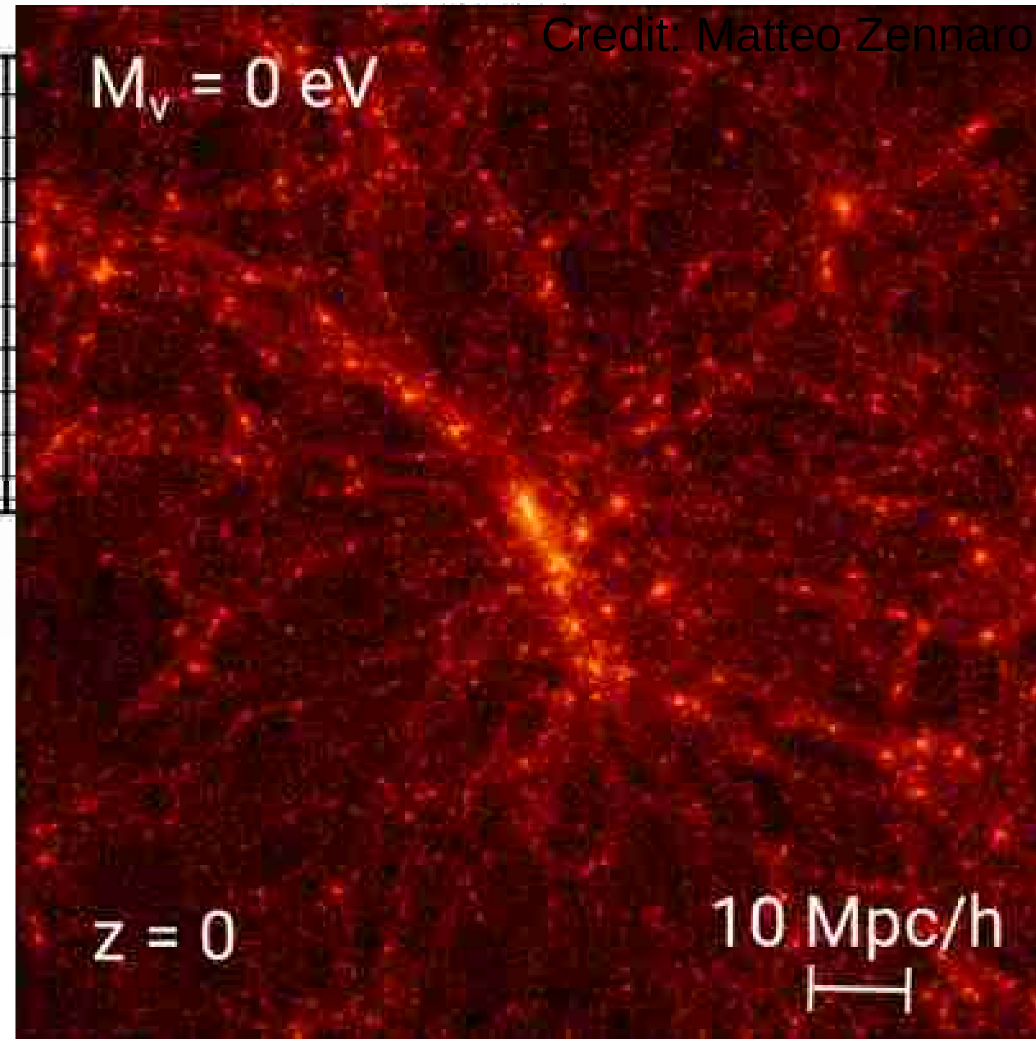
Different implementations agree at the subpercent level for light neutrinos
Adamek, Angulo et al (2022) – EUCLID massive neutrino comparison project

Massive Neutrinos

Neutrinos suppress structure growth on small scales







Credit: M. Zennaro

Two-fluid simulations (baryons & DM)

Unlike dark matter, baryons couple to the radiation field which breaks an initial symmetry causing distinct linear P(k)

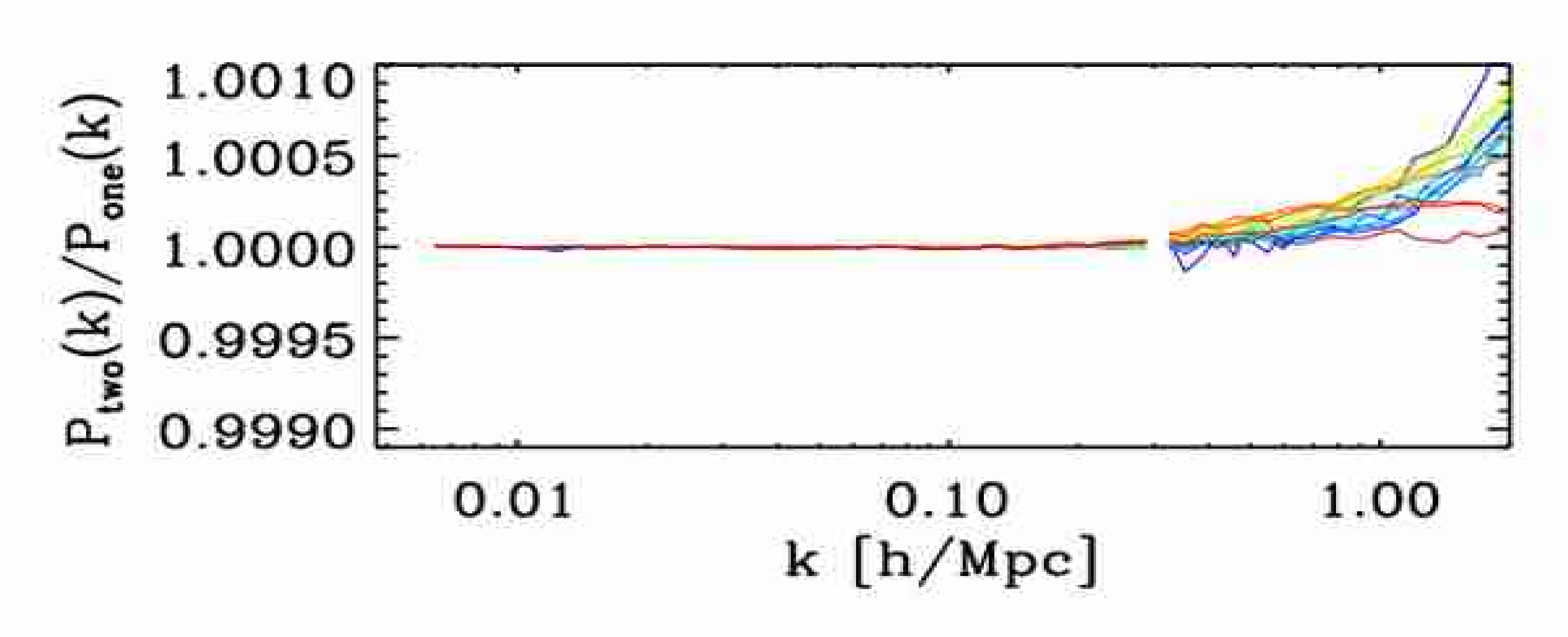
1) The similarity between baryons and DM Is broken by photons.

2) Prior recombination and on subhorizon scales, baryons and photons are coupled, creating oscillations and opposing the growth of fluctuations.

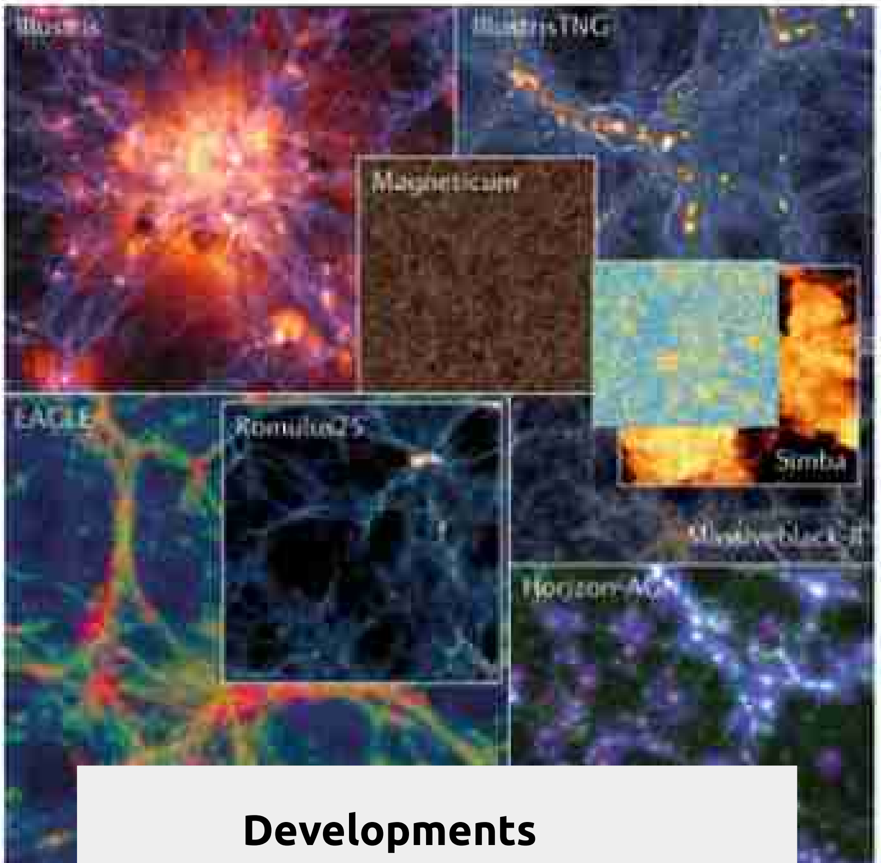
3) In contrast, DM grows mostly unimpeded.

4) After recombination, baryons and photons decouple.

5) At this point the growth is dominated by gravity.



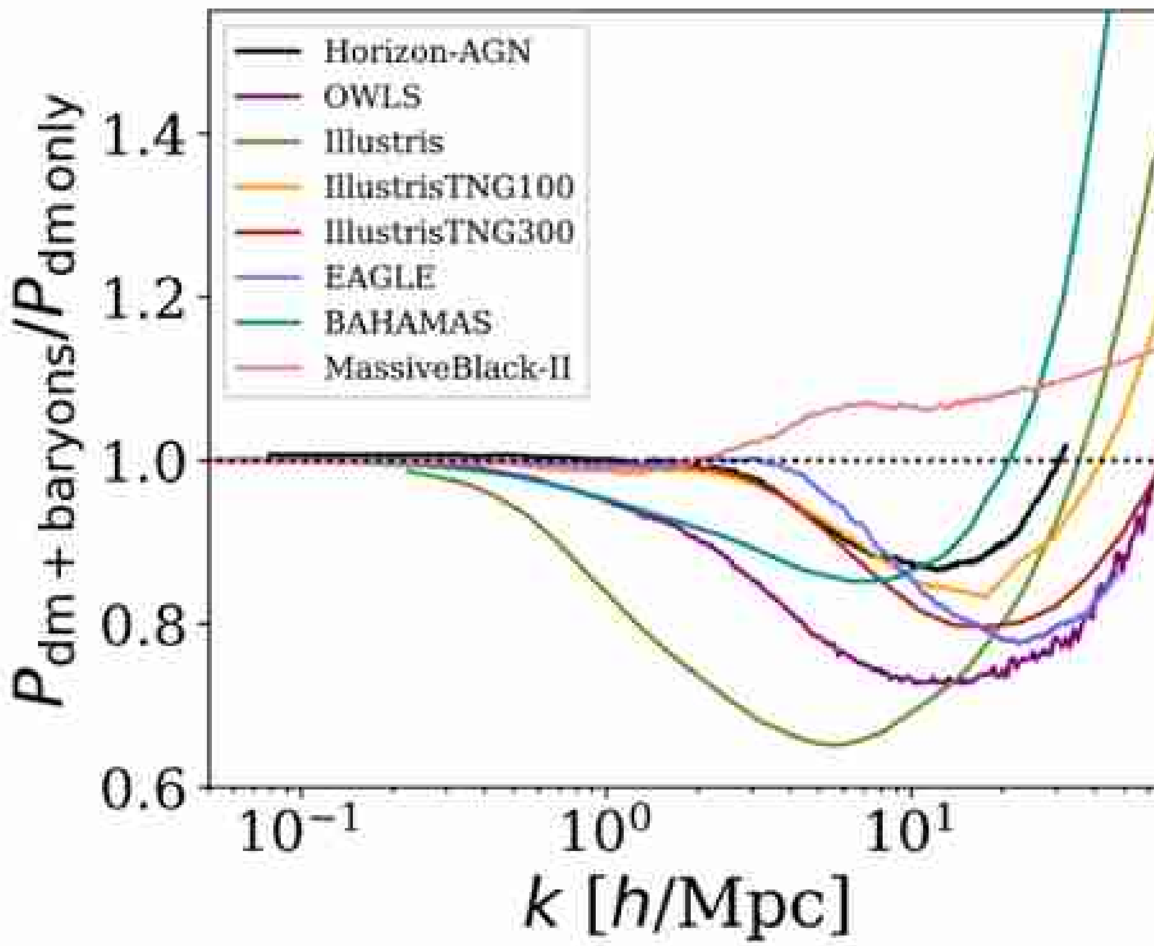
Hydrodynamical forces and galaxy formation is important to predict the large-scale structure of a given cosmology



Developments

1) Machine-learn hydro effects

2) Baryonification techniques



Chisari et al (2019)

Lectures on simulating the formation of structure in the Universe

→ Methods

→ Beyond the simplest models

→ **Current State of the Art**

→ The future

State of the art N-body simulations

The rapid development of LSS surveys have pushed for similar advances in computer simulations

BigData

We produce and analyse huge datasets.

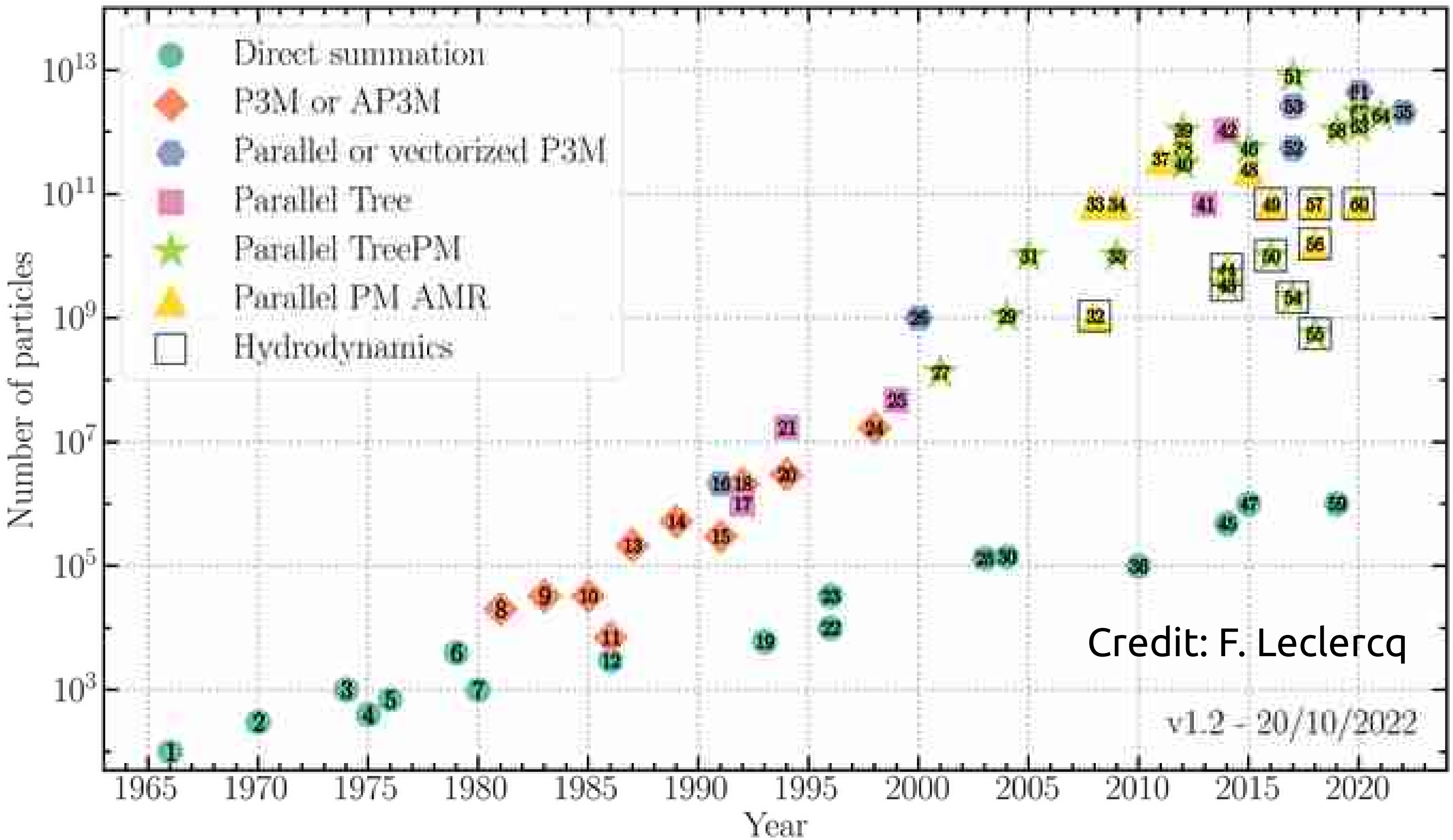
BigScience

We want to use new available resources to tackle our problems with more accuracy and sophistication.

NewScience

We are able to pose and solve problems unreachable in the past.

Cosmological simulations have continuously grown in their mass resolution and volume



Simulations roughly double in size every 17 months.

Cosmological simulations have continuously grown in their mass resolution and volume

Year	Simulation	Code, Algorithm	Supercomputer, Location	Cores $[10^3]$	N_P $[10^{12}]$	Box $[h^{-1}\mathrm{Gpc}]$	ϵ $[h^{-1}\mathrm{kpc}]$
2014	Dark Sky (Skillman et al. 2014)	2HOT FMM	Titan USA	20	1.1	8	36.8
2017	TianNu (Emberson et al. 2017)	CUBEP^{3}M PM-PM-PP	Tianhe-2 China	331	2.97	1.2	13
2017	Euclid Flagship (Potter et al. 2017)	PKDGRAV3 Tree-FMM	PizDaint Switzerland	4	2.0	3.	4.8
2019	Outer Rim (Heitmann et al. 2019)	HACC Tree-PM	Mira USA	524	1.07	3.0	2.84
2019	Cosmo-π (Cheng et al. 2020)	CUBE PM-PM	π 2.0 China	20	4.39	3.2	195
2020	Uchuu (Ishiyama et al. 2021)	GreeM Tree-PM	ATERUI-II Japan	<40	2.0	2.0	4.3
2020	Last Journey (Heitmann et al. 2021)	HACC Tree-PM	Mira USA	524	1.24	3.4	3.14
2021	Far Point (Frontiere et al. 2021)	HACC Tree-PM	Summit USA	?	1.86	1	0.8

Table 1 List of cosmological simulations with a particle number in excess of 1 trillion (10^{12})

The computational challenge

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

Force calculation and load imbalances --> GPUs/FFM

Quadrillion force calculations with large anisotropies and very different dynamical timescales

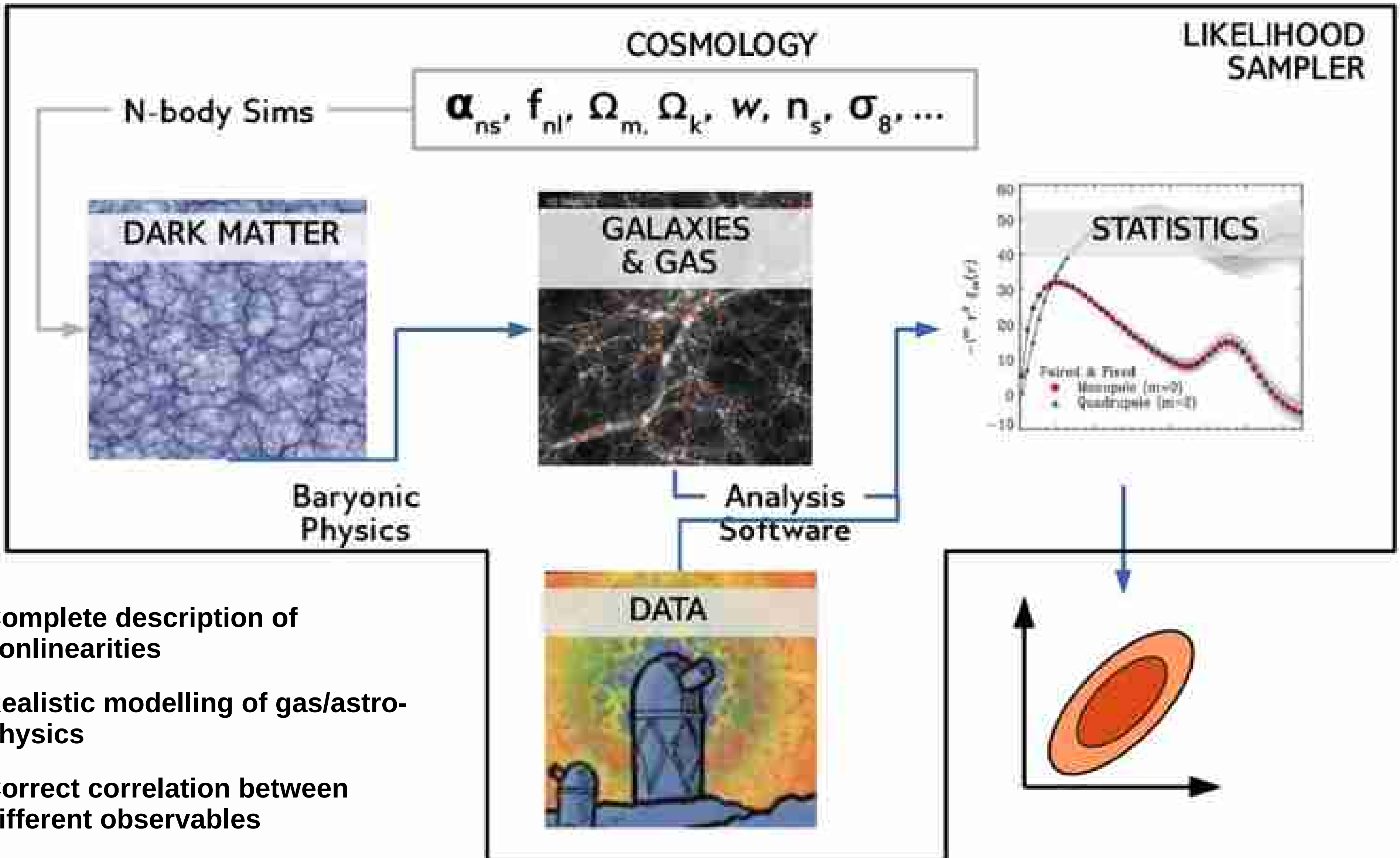
RAM –> Data compression

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

I/O & Disk Space –> on-the-fly compression

Data products can be en excess of dozens of Petabytes. Dissemination is a key issue.

The cosmology challenge



Complete description of nonlinearities

Realistic modelling of gas/astro-physics

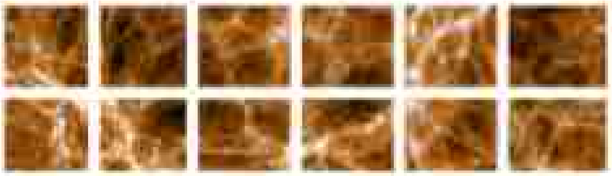
Correct correlation between different observables

Any summary statistics is possible

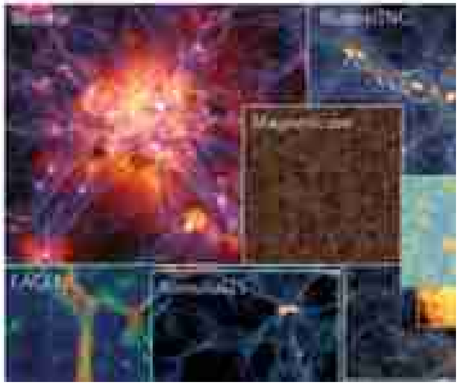
The cosmology challenge

One simulation is challenging, but we need hundreds to Thousands to interpret LSS observations

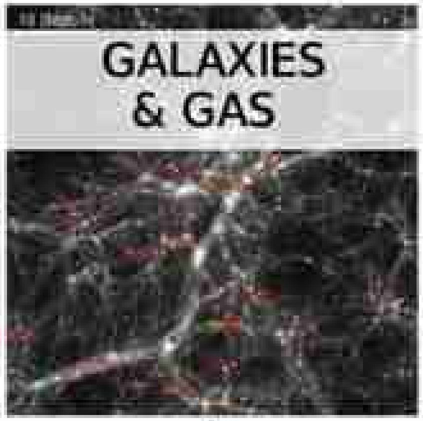
Emulators



Baryons



Robust astrophysics



Machine learning

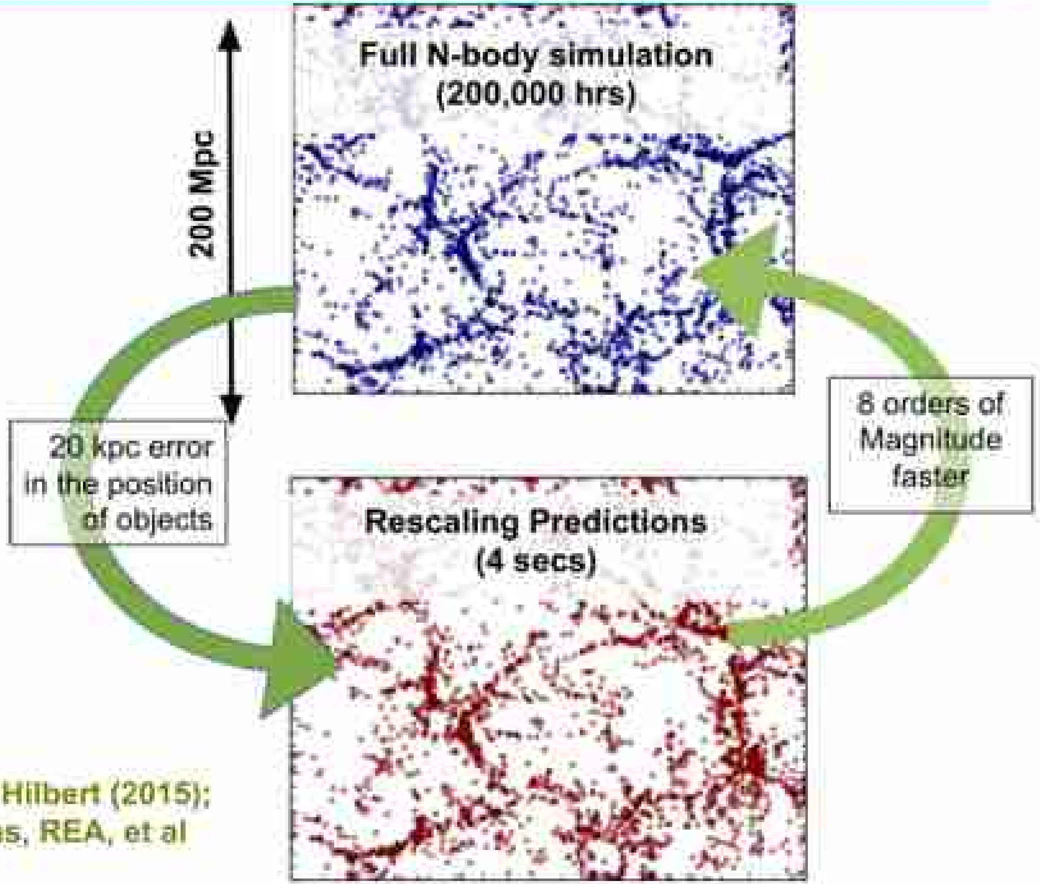
The cosmology-rescaling algorithm

Thousands of simulations are not needed: a handful is enough!

Validated for:

- real/redshift-space
- correlation function/power spectra
- 3-point correlation functions
- (sub)halo mass function
- abundance of voids
- different redshifts/cosmologies

Including massive neutrinos and dynamical dark energy!

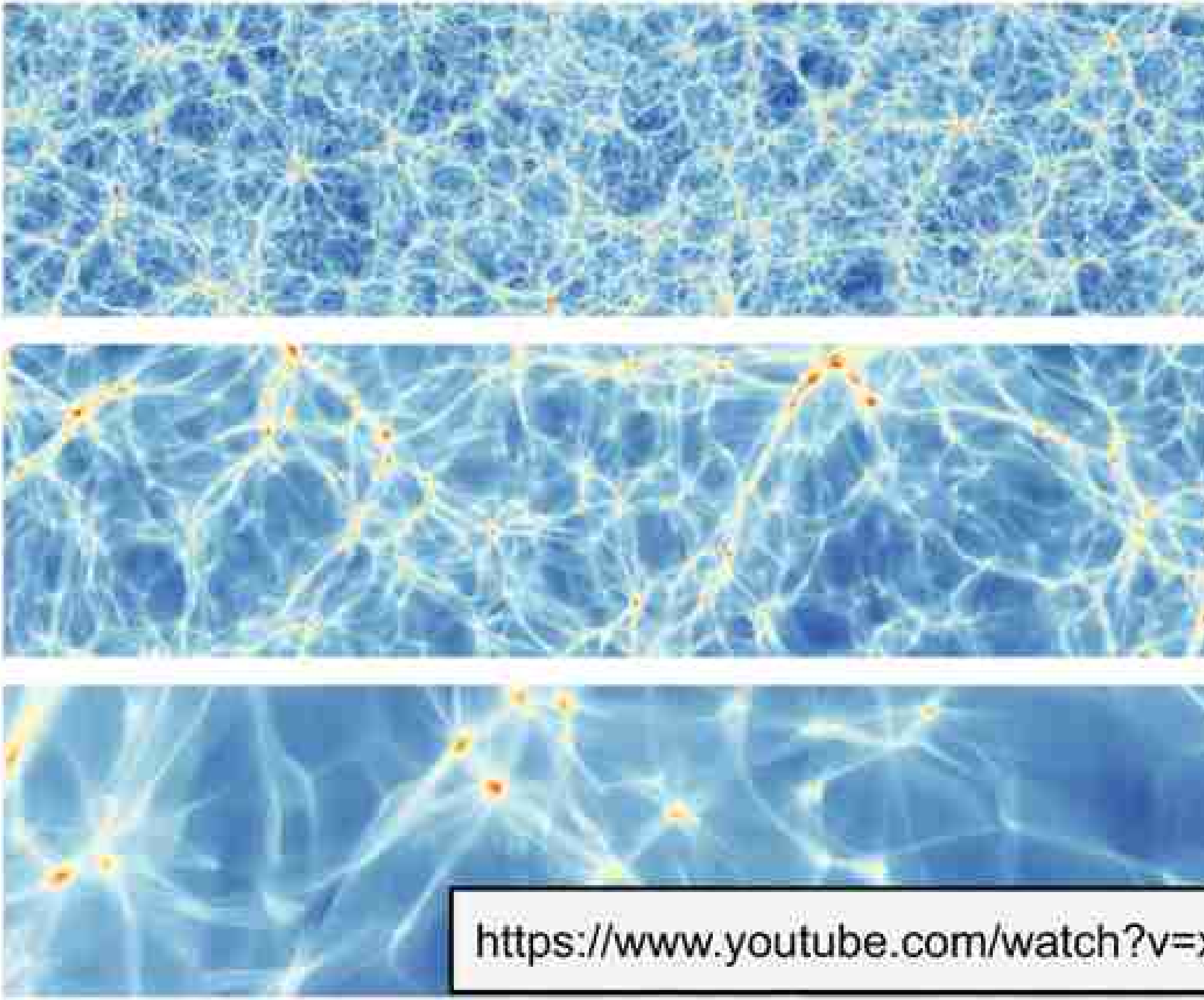


REA & White (2010); Ruiz et al (2011); Mead & Peacock 2013; REA & Hilbert (2015); Renneby, Hilbert, Angulo (2018); Zennaro, REA et al (2019); Contreras, REA, et al (2020); Ondaro, REA, et al (2021); Lopez-Cano, REA+ (2022)

The BACCO Simulation suite

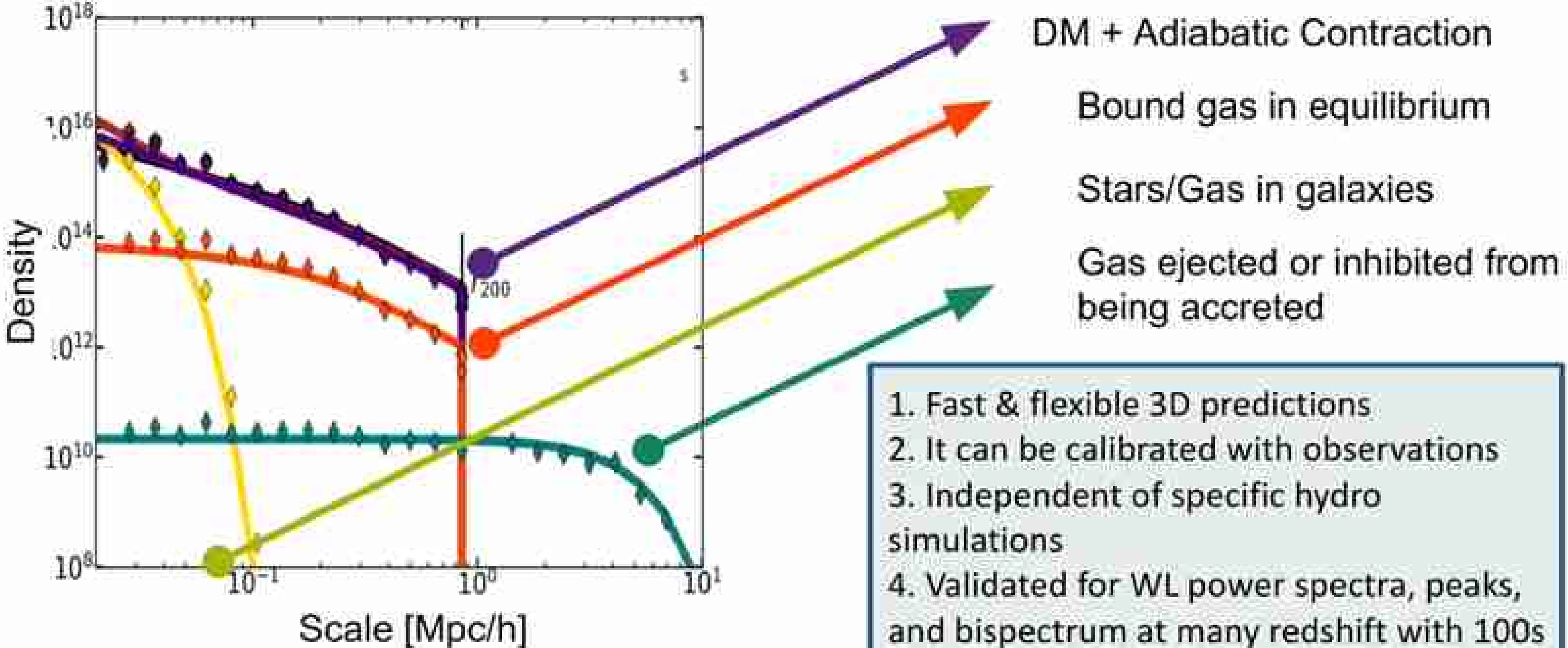
- 5 sims with various cosmologies
- Fixed-paired ICs
- L=2000 Mpc
- N=4320**3
- New group finder
- Phase-space tess
- Orphan tracking
- Merger trees
- 1% accurate

15 Million CPU hours



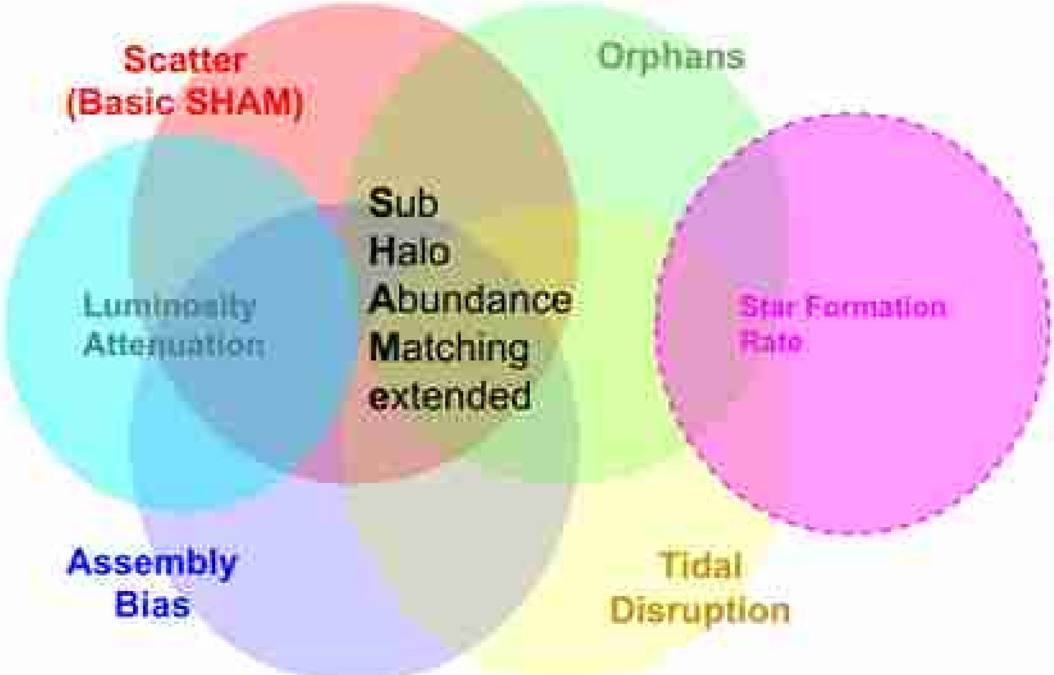
Modelling the baryonic effects on the matter clustering

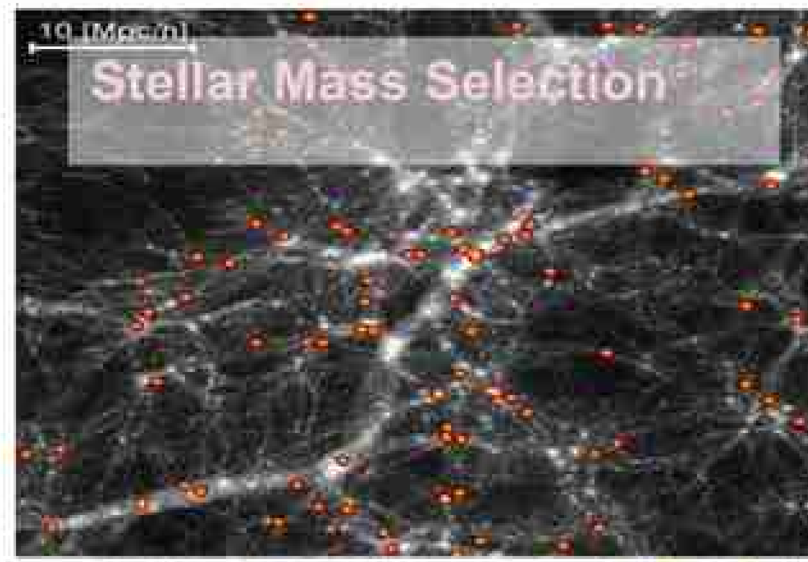
Arico, REA+ (2020,2021,2022); Schneider+(2017, 2019)

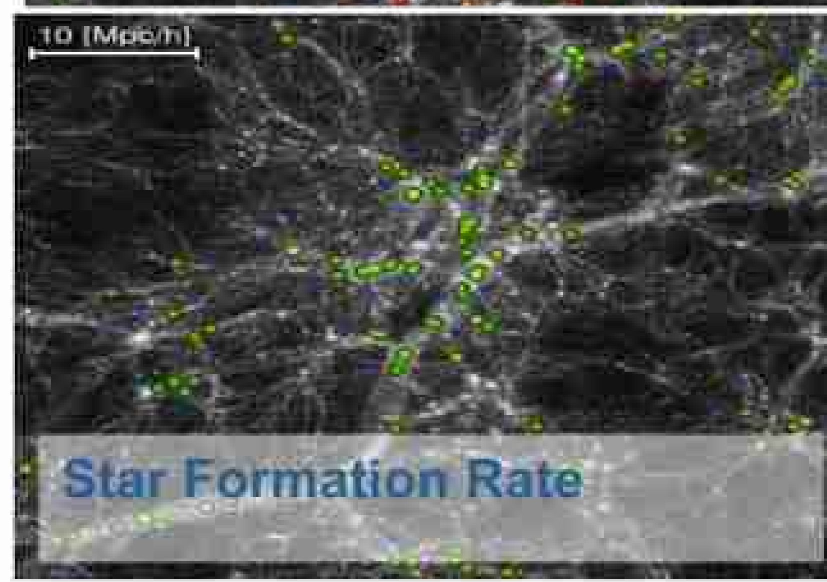


1. Fast & flexible 3D predictions
2. It can be calibrated with observations
3. Independent of specific hydro simulations
4. Validated for WL power spectra, peaks, and bispectrum at many redshift with 100s sims

How to efficiently model galaxies in DM sims?







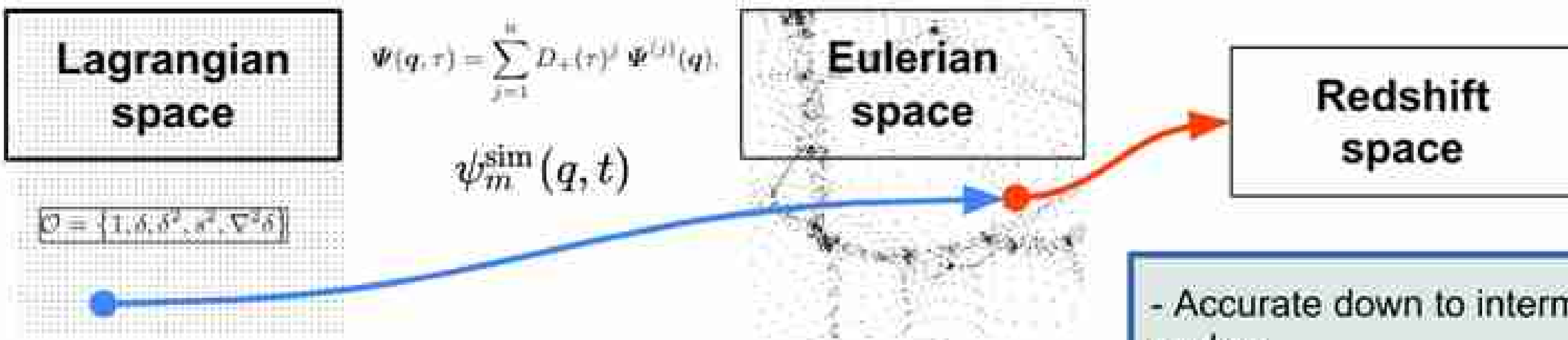
1. Unbiased constraints on cosmology after marginalization over galaxy formation

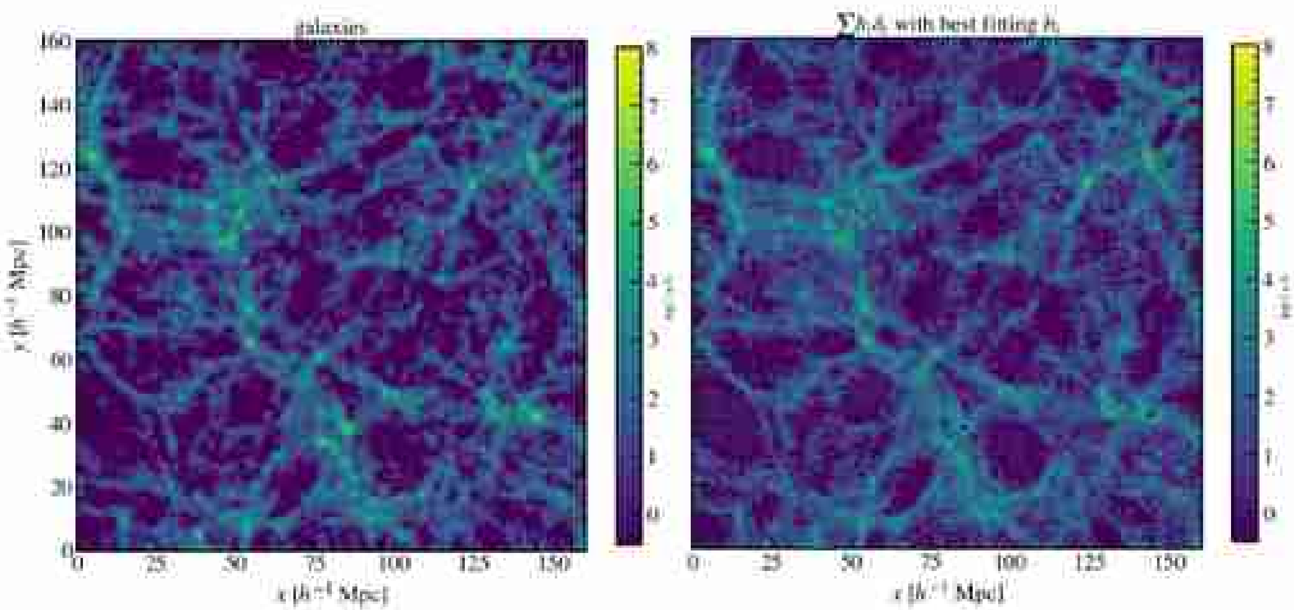
2. Validated against multiple SAM and Hydro sims (inc. M-TNG) down to R ~ 500 kpc/h

Contreras, REA, Zennaro (2020); Contreras, REA, et al (2021); Contreras, REA, MTNG (2022), Ortega, Contreras, REA (2023)

Hybrid bias expansion

Combining numerical and analytical worlds

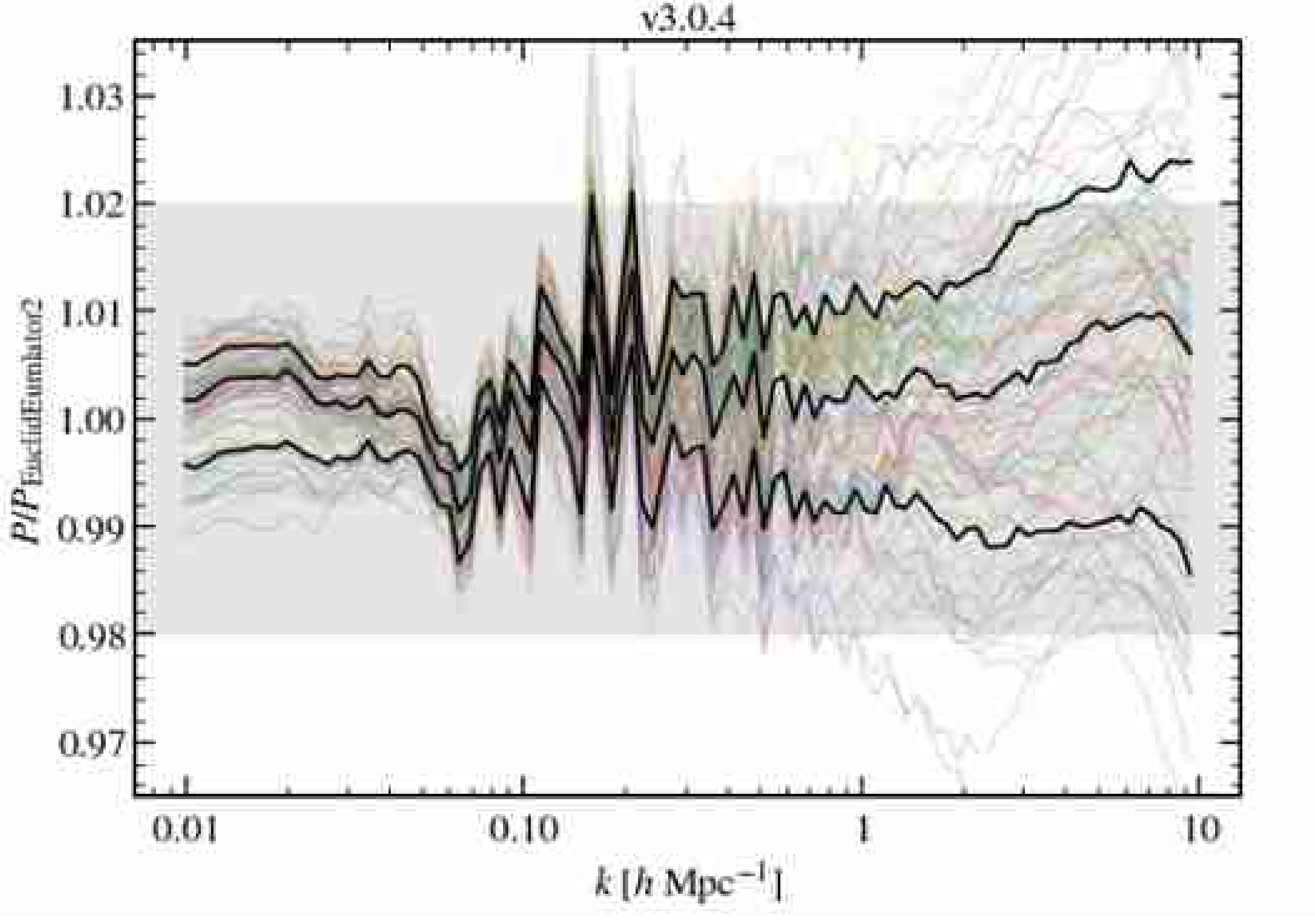




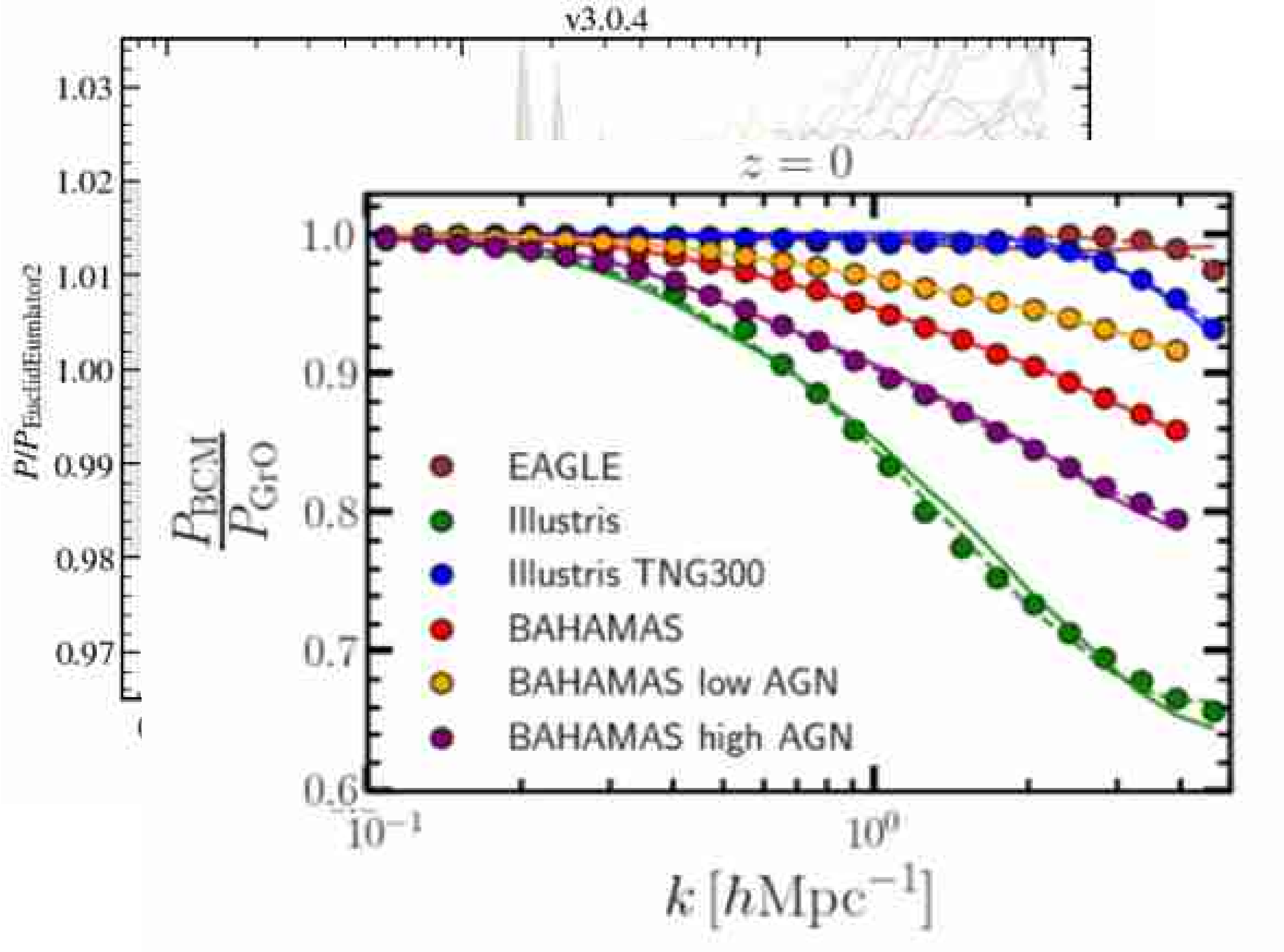
- Accurate down to intermediate scales
- Field-level & multiple summary statistics
- More cosmological information
- Coupled with displacement emulators
- Covariance matrices & Simulation-based inference

Zennaro, REA,+ (2021,2022),
Krokon+(2022,3)
Pellejero-Ibañez, REA+ (2021,2022)
Pellejero-Ibañez, REA, Jamieson (2023)

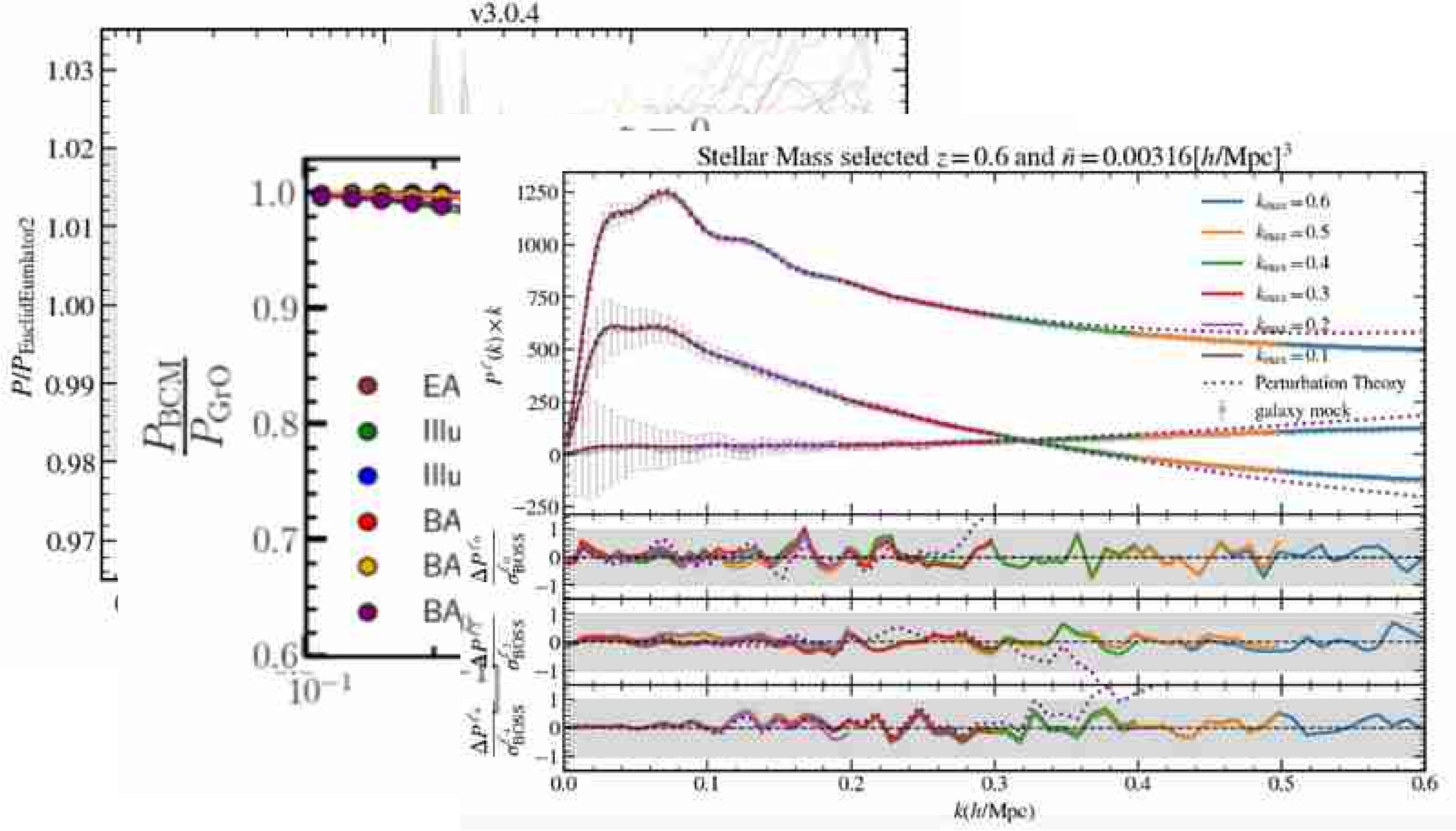
Accuracy in predicting the nonlinear distribution of matter, baryons and galaxies



Accuracy in predicting the nonlinear distribution of matter, baryons and galaxies

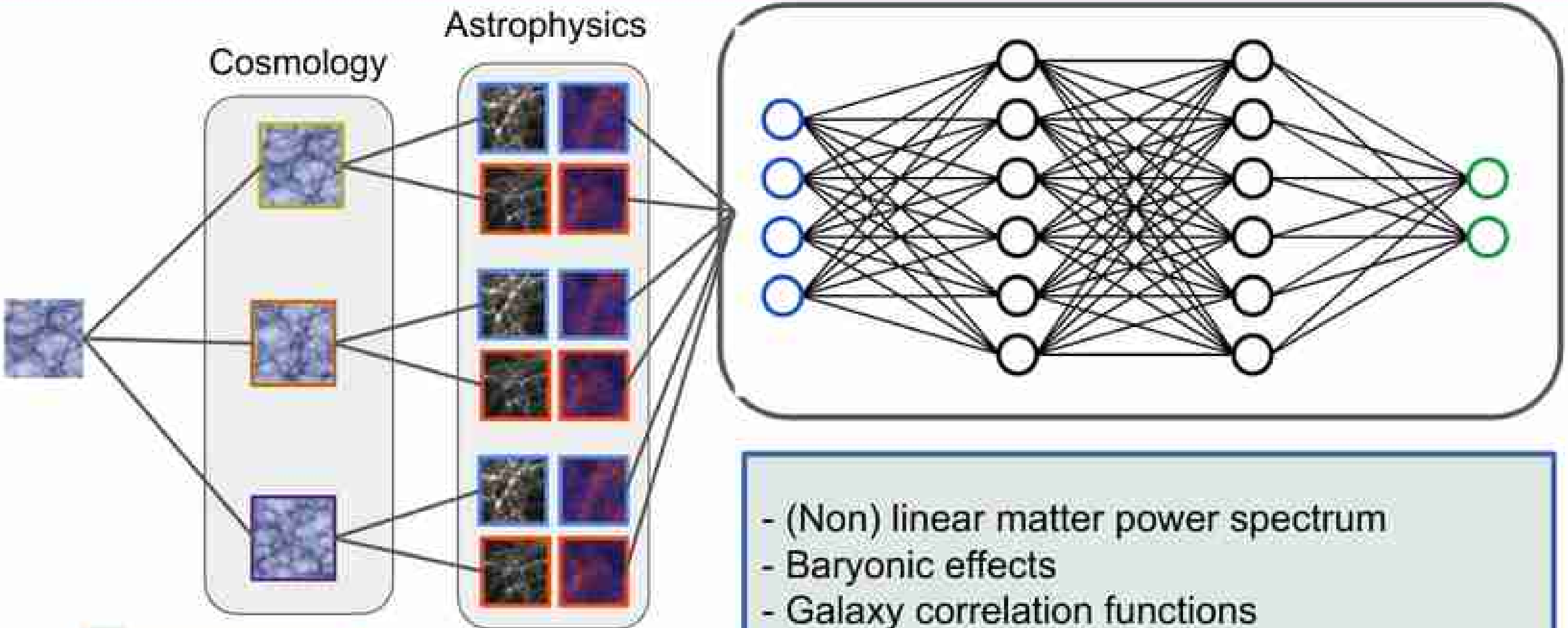


Accuracy in predicting the nonlinear distribution of matter, baryons and galaxies



Emulating LSS with neural networks

Statistics computed at 10k combinations of cosmology and astrophysics



baccoemu 0.3

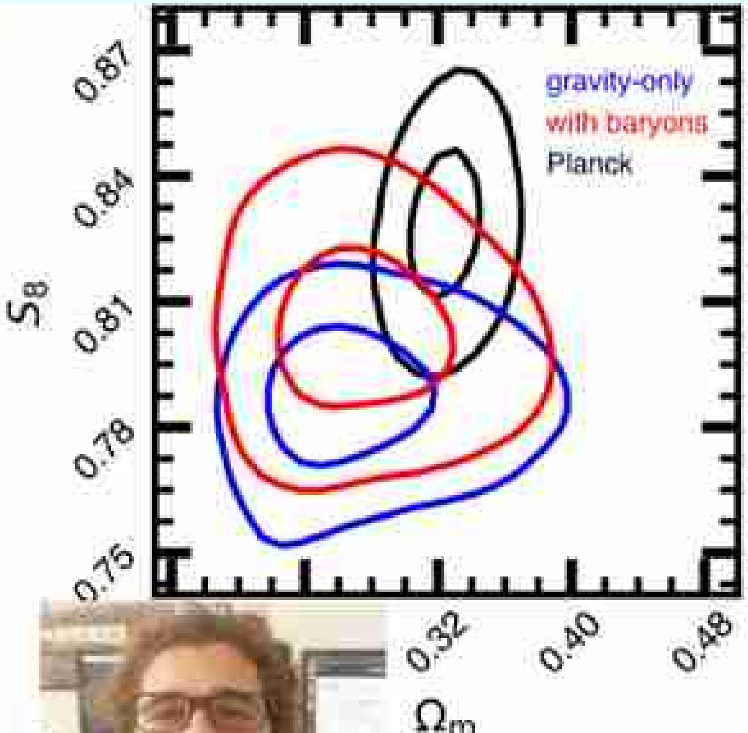
`pip install baccoemu`

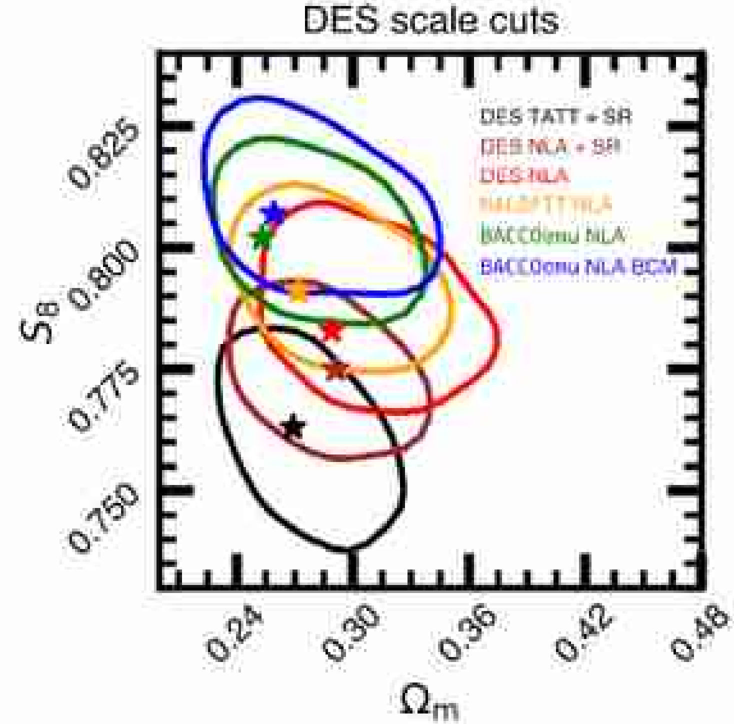
- (Non) linear matter power spectrum
- Baryonic effects
- Galaxy correlation functions
- Biased tracers in real & redshift space
- Intrinsic alignments, tSZ, halo mass function

Exploring the S8 tension

Constraints on cosmology and baryons from small scales

- Non linear matter power spectrum
- No shear ratios
- Baryonic effects
- Intrinsic Alignments





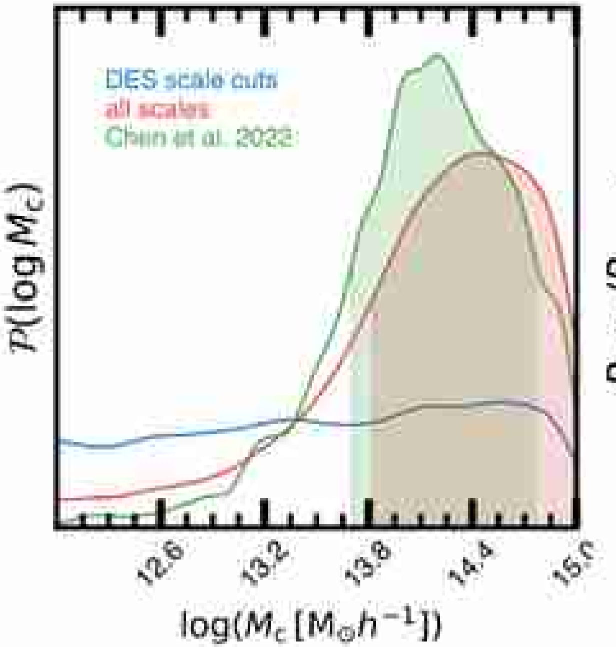
DARK ENERGY SURVEY

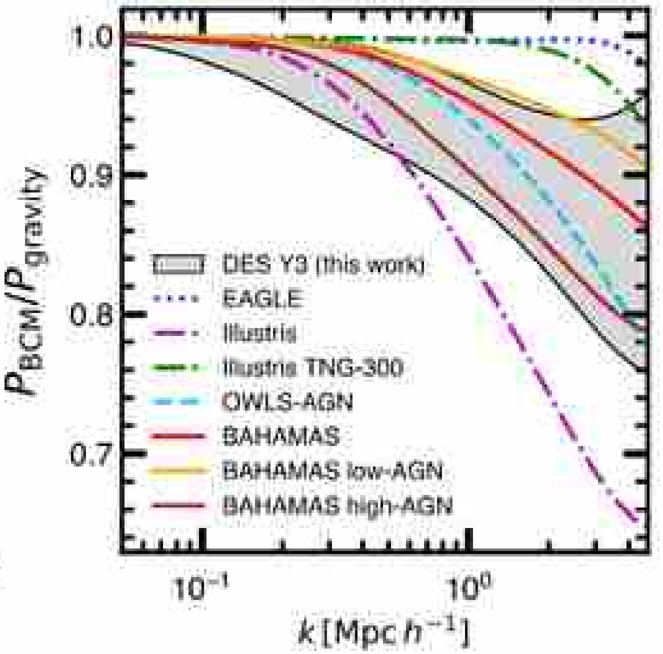
Arico, REA+ (2023)

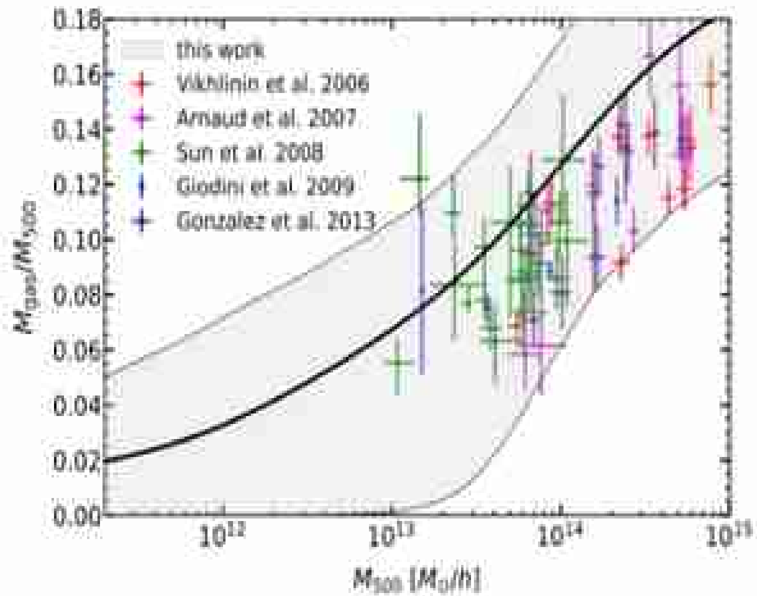
Exploring the S8 tension

Halo mass for which half of the gas is lost due to feedback:

$$\log M_c = 14.38^{+0.60}_{-0.56} \log(h^{-1}M_\odot)$$

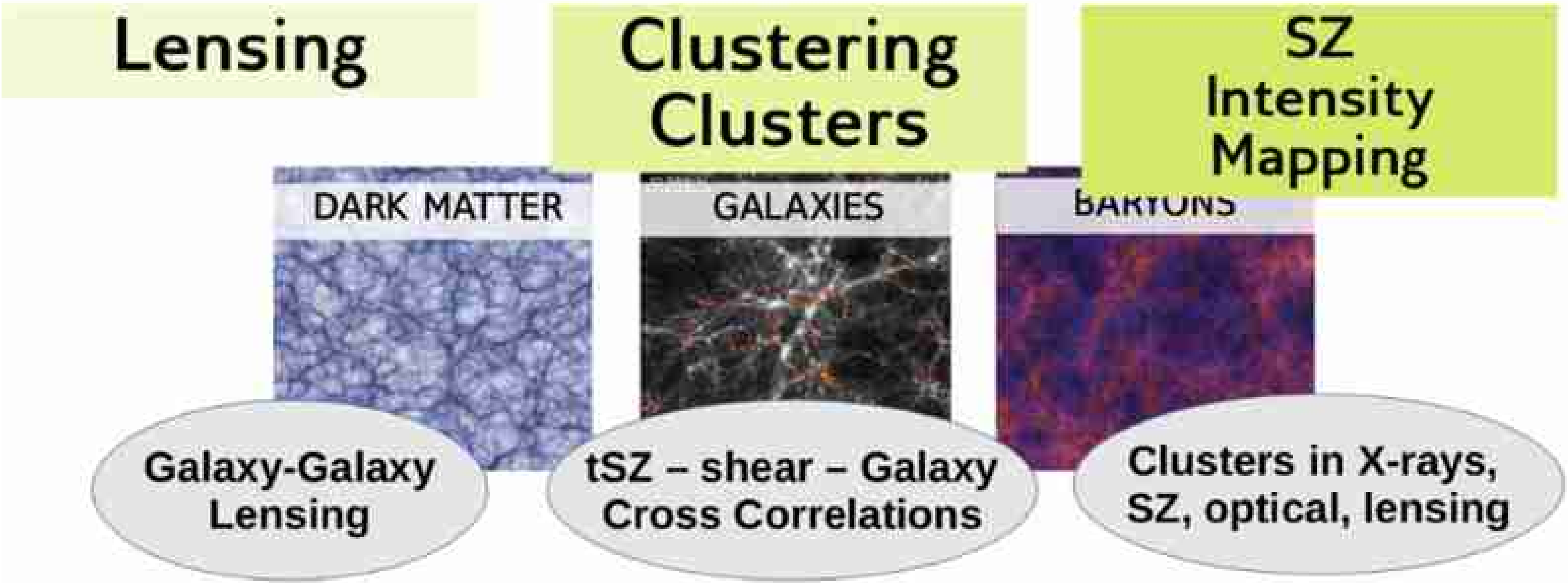






Arico, REA + (2023), Chen et al (2022)

Simulations enable a simultaneous interpretation of LSS data



Outlook

LSS simulations are mature and essential in understanding the Universe

→ This required progress in:

a) Testing of basic assumptions:

DM discretization, use of Newtonian physics, etc

b) Algorithms:

ICs, time integration, single fluids, etc

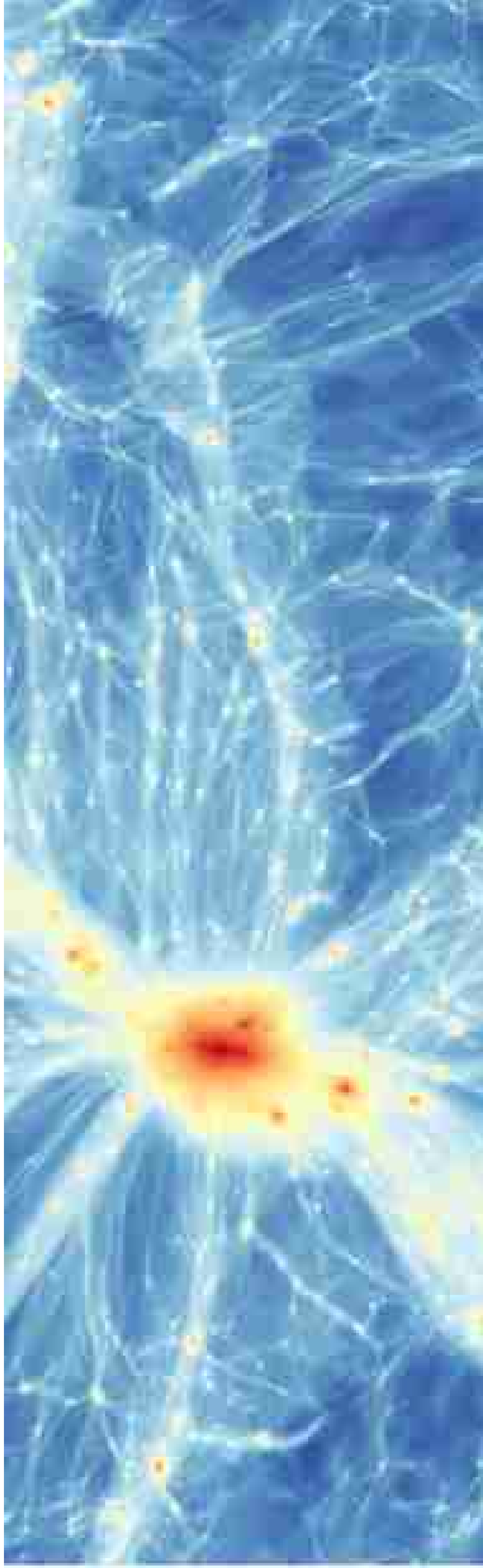
c) Extension of models beyond simple LCDM:

Modified gravity, DM models, neutrinos, etc

d) Reliability:

Systematic tests and comparisons, Lagrangian tessellation

All this with the continuous support of HPC.



Outlook

Getting ready for the next big challenge: directly interpreting data

→ Connection with visible universe is more uncertain but progress is made: galaxies and baryonic effects

→ Efficiently use the next generation of HPC facilities

→ Cosmology variations:

a) Fast methods for gravity and astrophysics

b) Interpolators: emulators & machine learning

N-body simulations could become a key piece to elucidate the nature of dark matter, dark energy, and gravity.