Producing fast and accurate mock galaxy catalogues

Albert Izard (UC Riverside / Jet Propulsion Laboratory / Caltech)

Collaborators: Alina Kiessling[1], Pablo Fosalba[2], Martin Crocce[2], Eric Huff[1], Phil Bull[1]

[1]Jet Propulsion Laboratory / Caltech

[2]Institut of Space Sciences (ICE)

March 20, 2017

Jet Propulsion Laboratory
California Institute of Technology

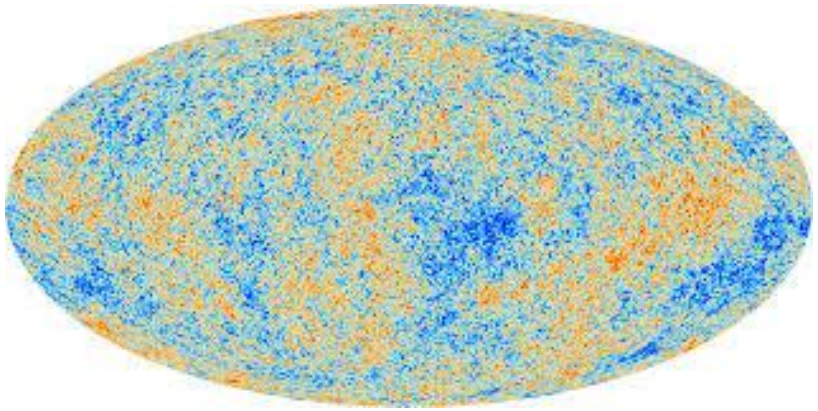
Copyright 2017 California Institute of Technology. U.S government sponsorship acknowledged

Outline

- Why do we need approximate methods?
- Optimization of a simulation method
- Generation of light cone catalogs
- Modeling of weak gravitational lensing
- Mock galaxy catalogs pipeline

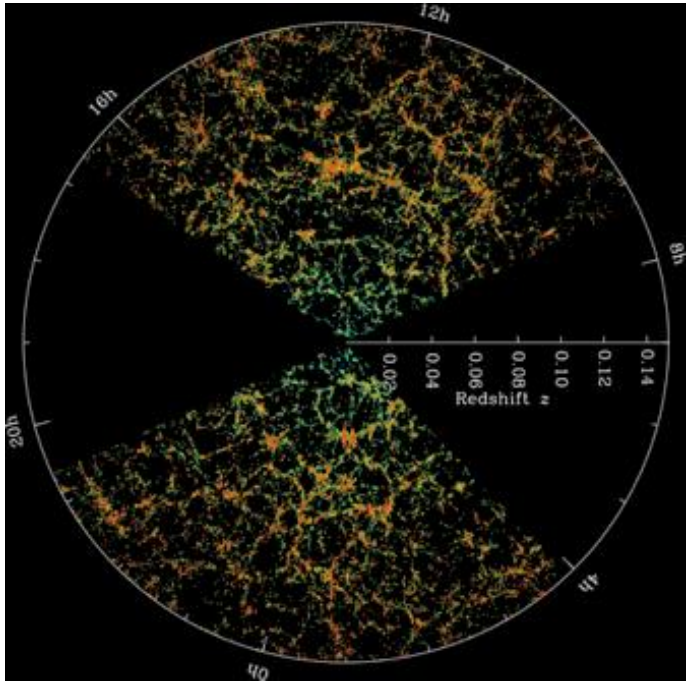
The LCDM model

Large-scale homogeneity and isotropy



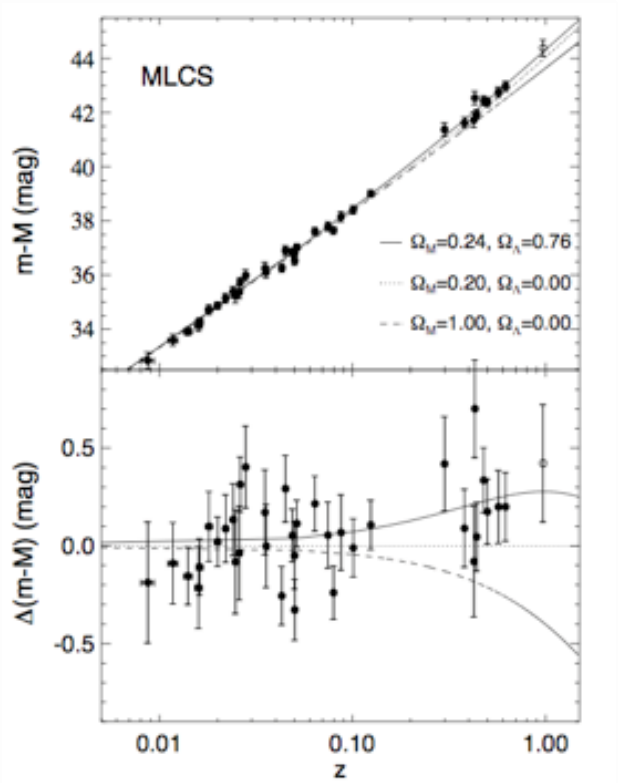
Planck

Gravitational instability

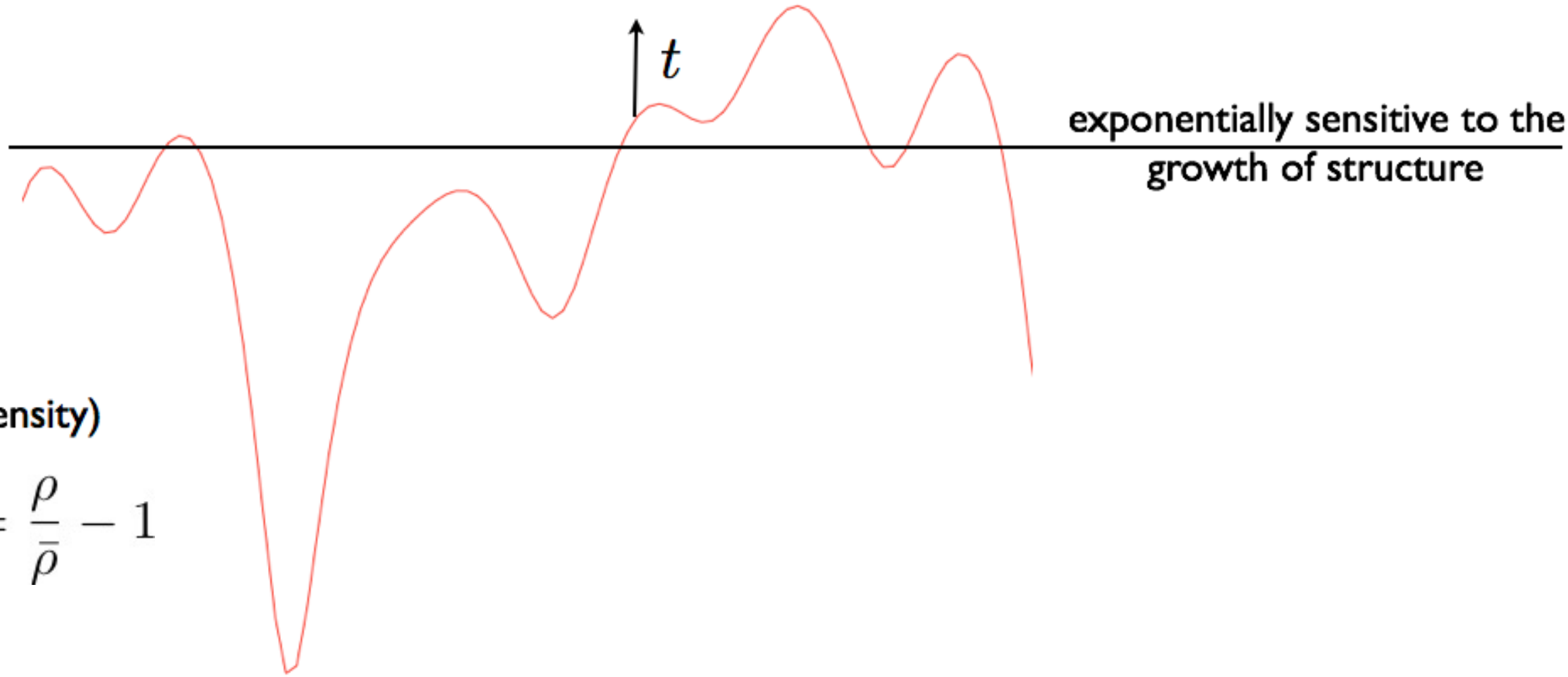


SDSS

Expansion

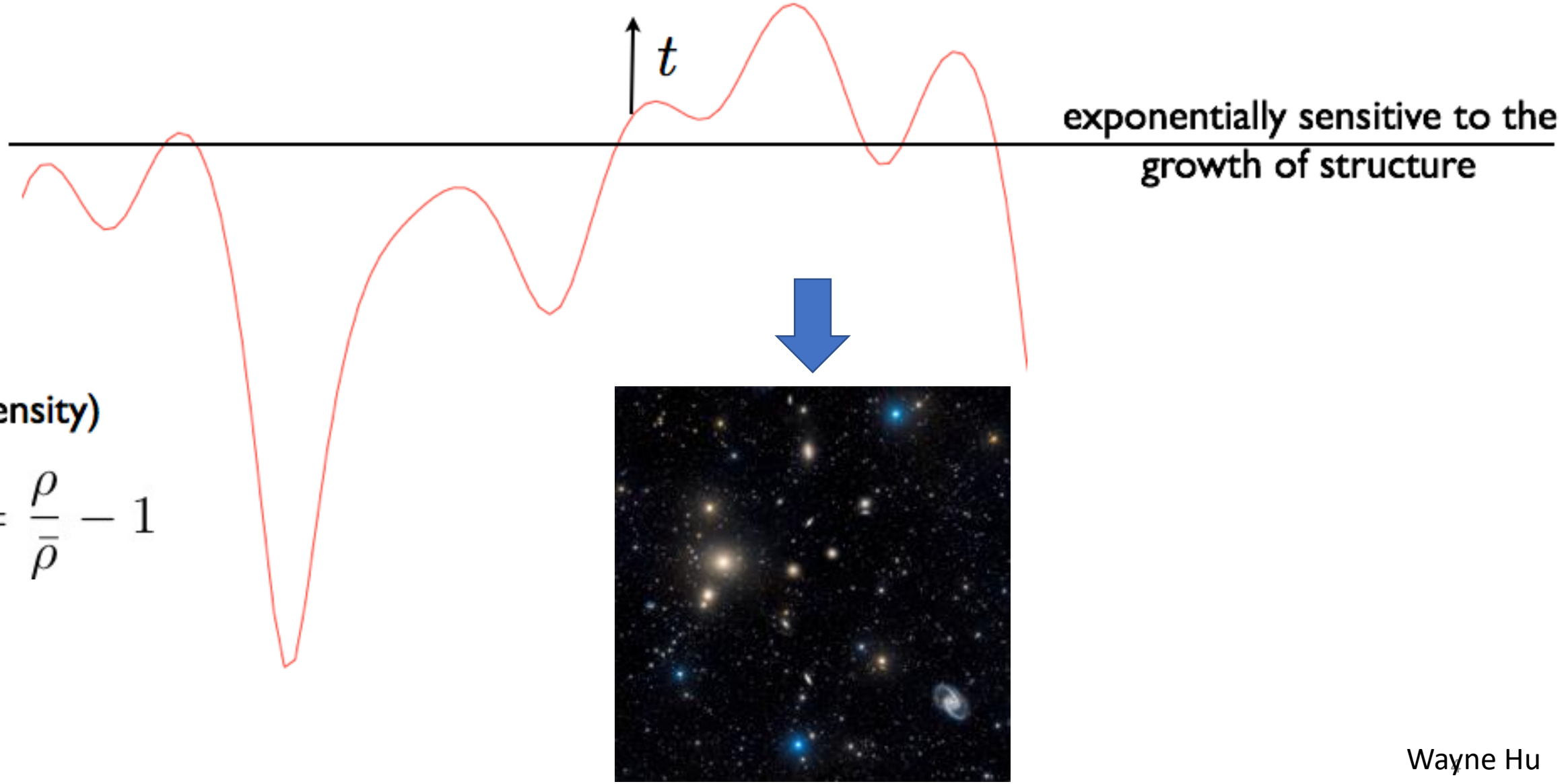


Riess et al. 1998

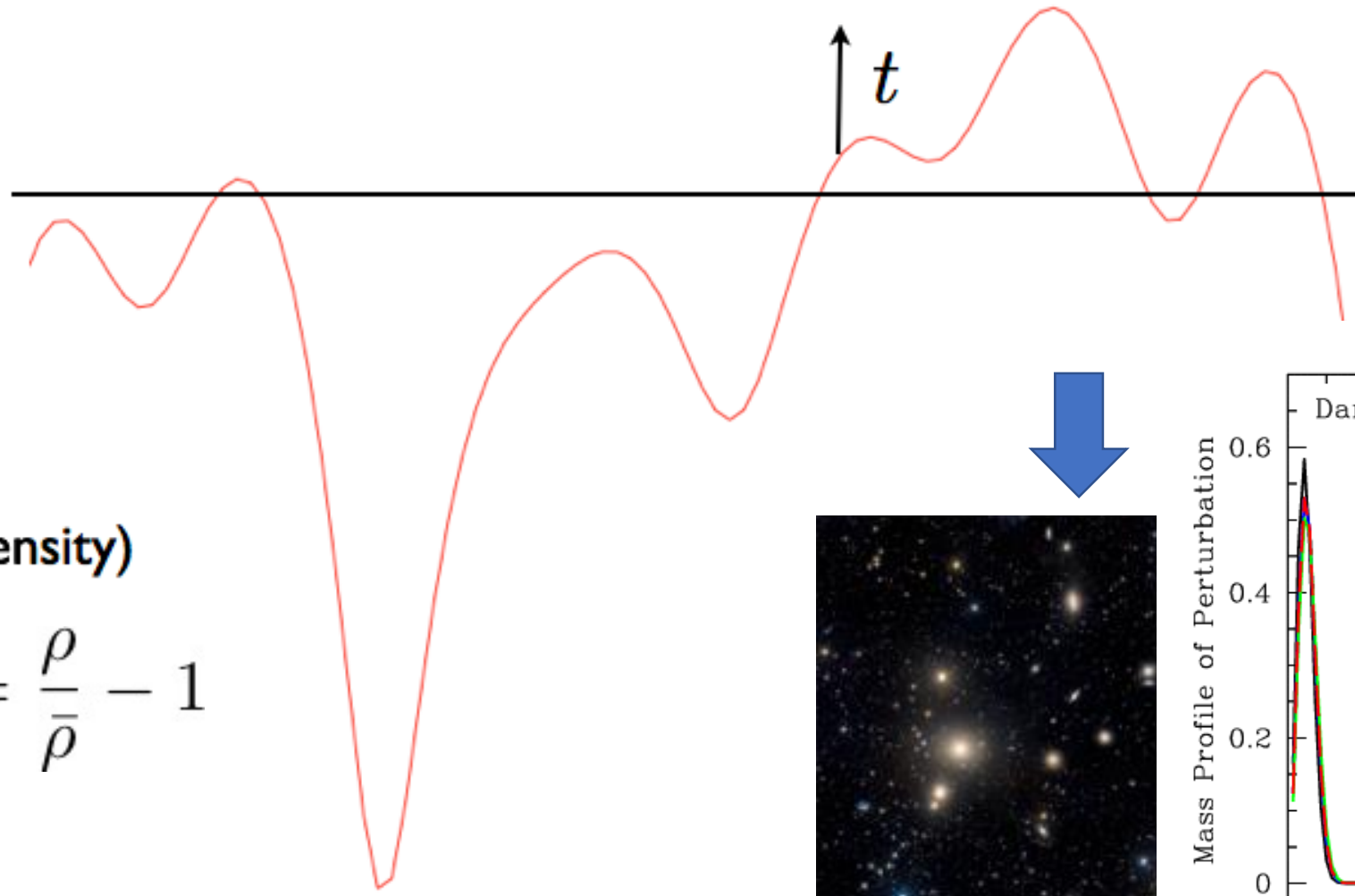


δ
(overdensity)

$$\delta = \frac{\rho}{\bar{\rho}} - 1$$


t
exponentially sensitive to the growth of structure
δ
(overdensity)
$\delta = \frac{\rho}{\bar{\rho}} - 1$

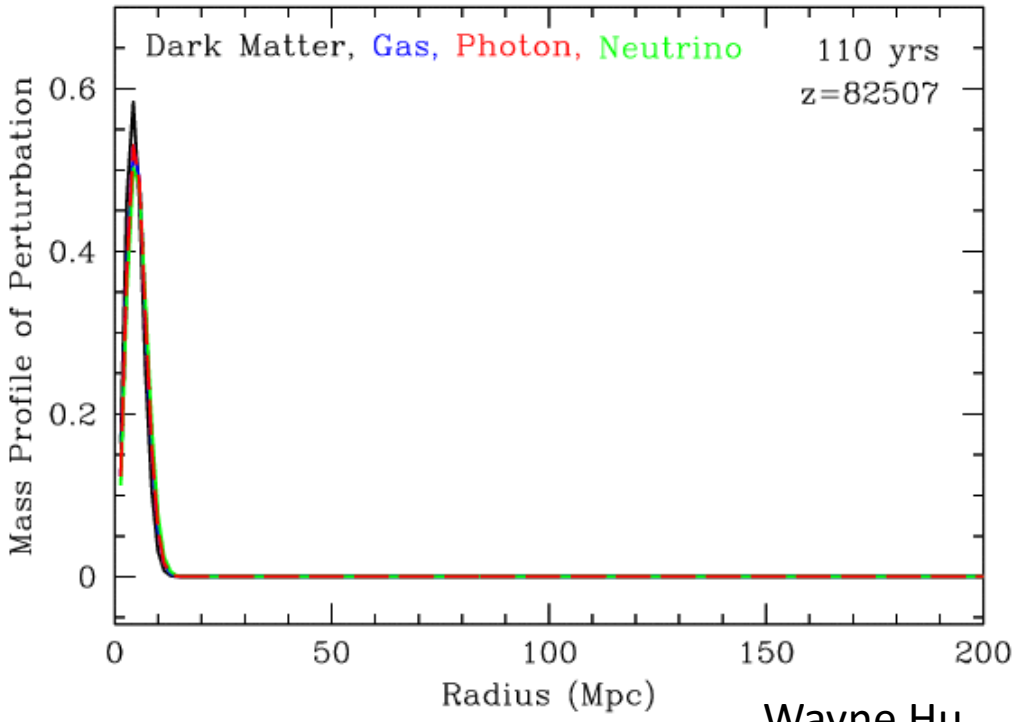
Wayne Hu



exponentially sensitive to the growth of structure

δ
(overdensity)

$$\delta = \frac{\rho}{\bar{\rho}} - 1$$



Wayne Hu

Cosmological probes

- Growth of structure (perturbations)

Poisson

$$\nabla^2\Phi = \frac{3H_0^2}{2a}\Omega_0\delta$$

Linear growth

$$\delta \propto D(z)$$

Growth rate

$$f(z) \equiv \frac{d\ln D}{d\ln a} = \Omega_m(z)^{\gamma}$$

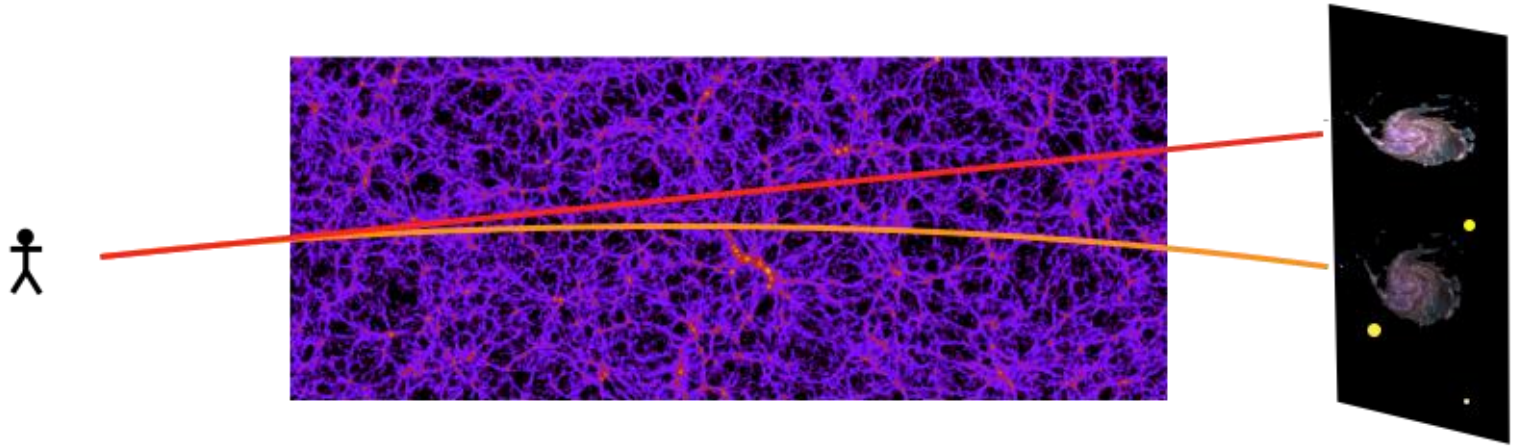
$$\gamma = 0.55$$

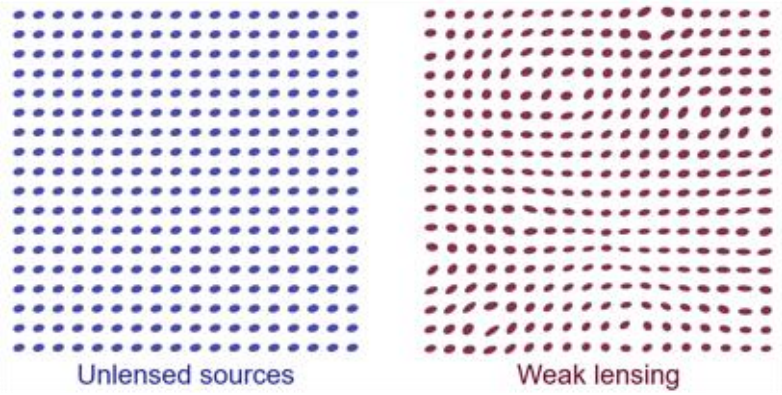
- Geometry (background)

$$H^2(z) = H_0^2\left(\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{DE}(1+z)^{3(1+w_{DE})}\right)$$

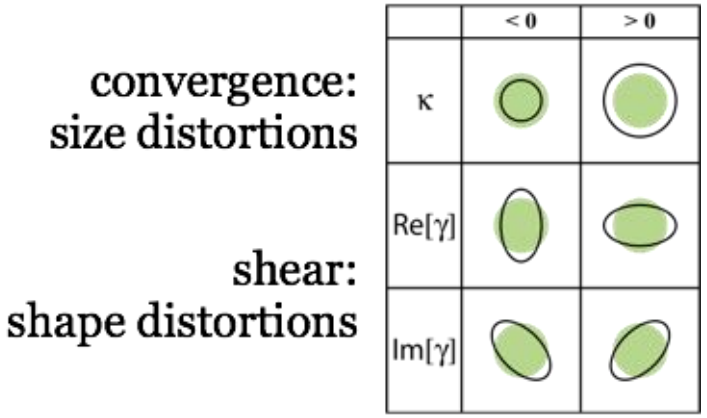
Weak gravitational lensing

Distortion of the light coming from background sources by the foreground matter distribution





(From Chieh-An Lin)

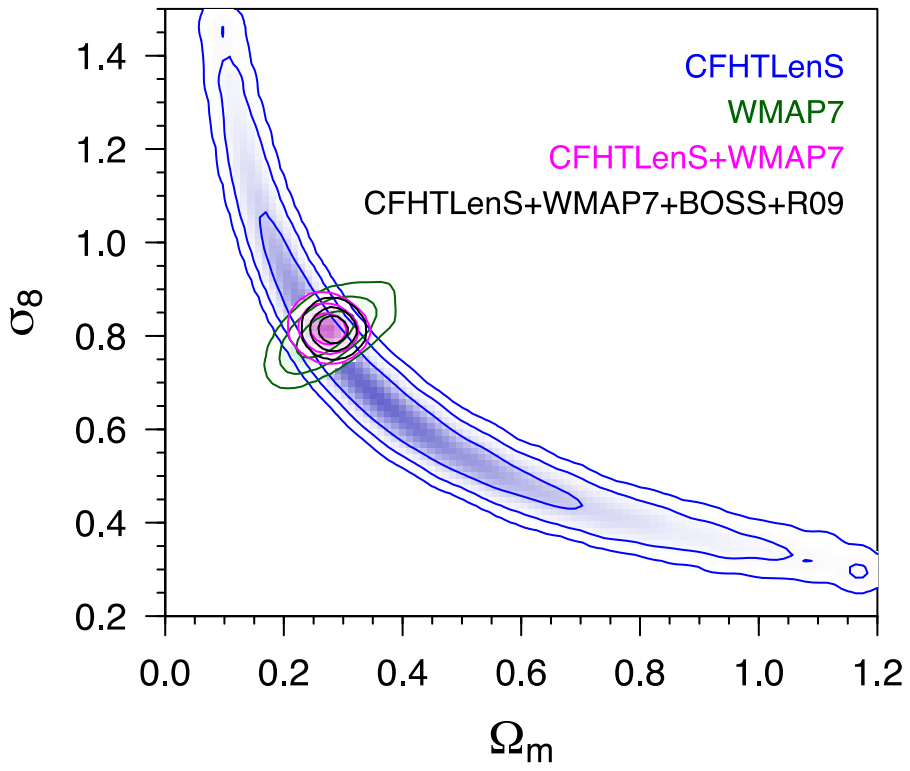


$$A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \partial_i \partial_j \Psi = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}$$

Weak lensing

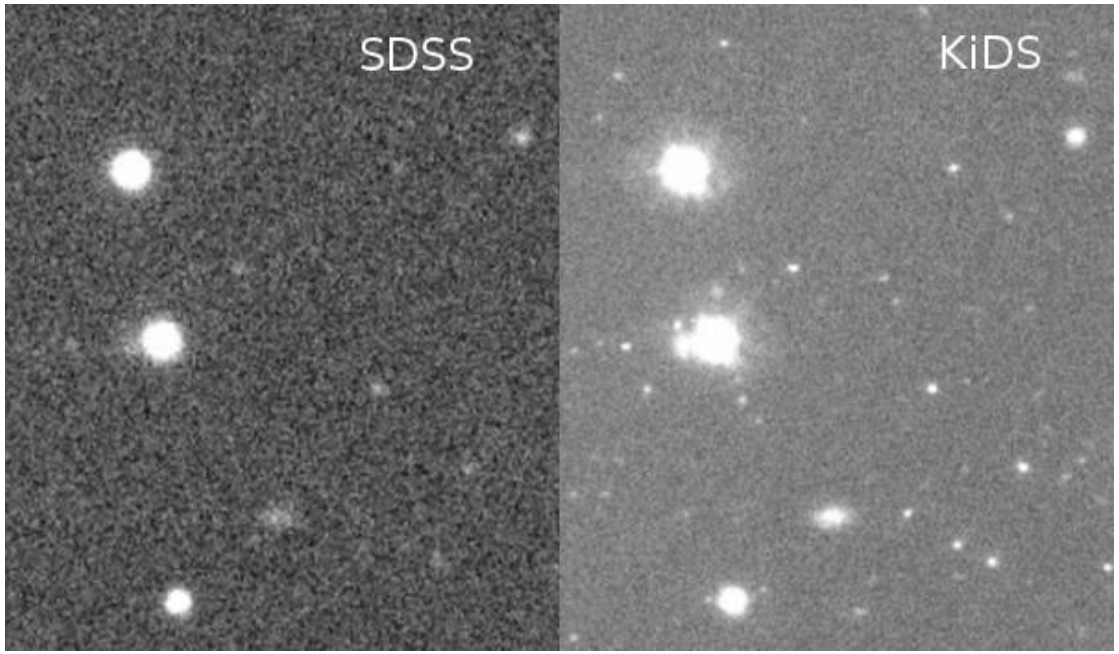
- WL: probes both the cosmic expansion history and the growth of structure

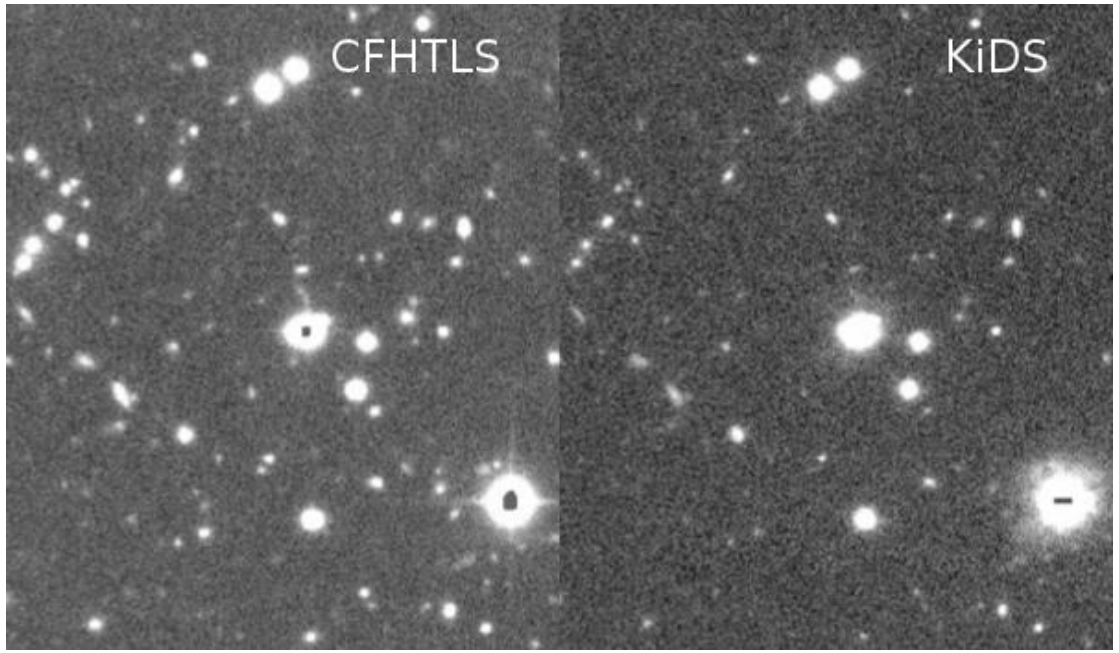
- Cosmic shear: measures the lensing power spectrum in several tomographic bins in different photometric redshifts.



From Kilbinger et al. 2013

Multi-color optical surveys





Sloan Digital Sky Survey

Kilo Degree Survey
2 mag deeper
1500 deg^2
Image quality

Canada-France-Hawaii
Telescope Lensing Survey
154 deg^2

Past or ongoing: CFHTLS, Pan-STARRS, KiDS, DES, HSC

Future: LSST, Euclid, WFIRST-AFTA

WL measurement errors

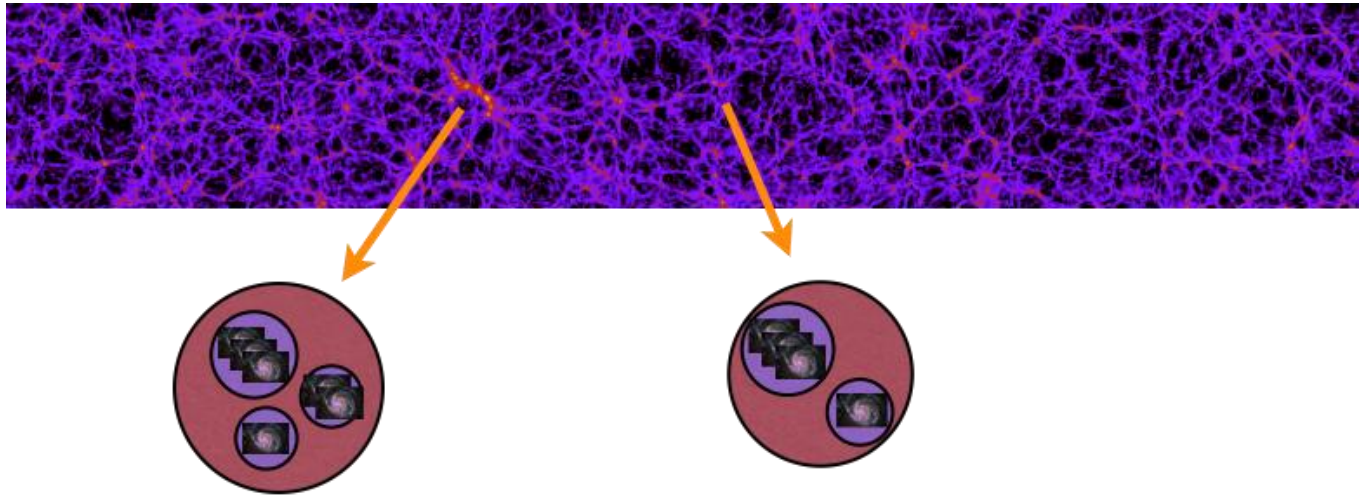
- Shape noise
- PSF calibration
- Sample variance
- Photometric errors
- Blending
- Intrinsic Alignments
- Theoretical predictions of the signal measured

Modeling the galaxy distribution

The dominant force on large enough scales is gravity

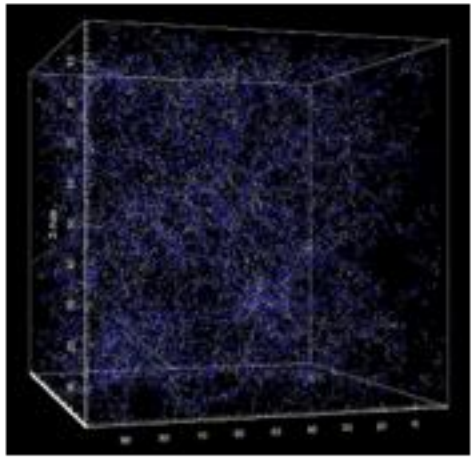
N-body simulations

- Matter evolves and forms collapsed objects, halos.
- Filaments form bridges connecting them
- Galaxies are born in halos

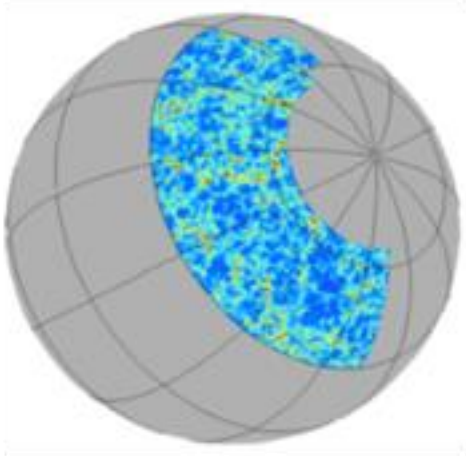


Galaxy catalogs pipeline

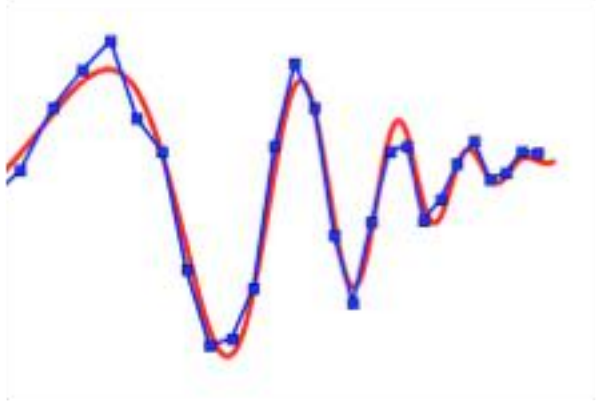
N-body simulation



Realistic catalogues

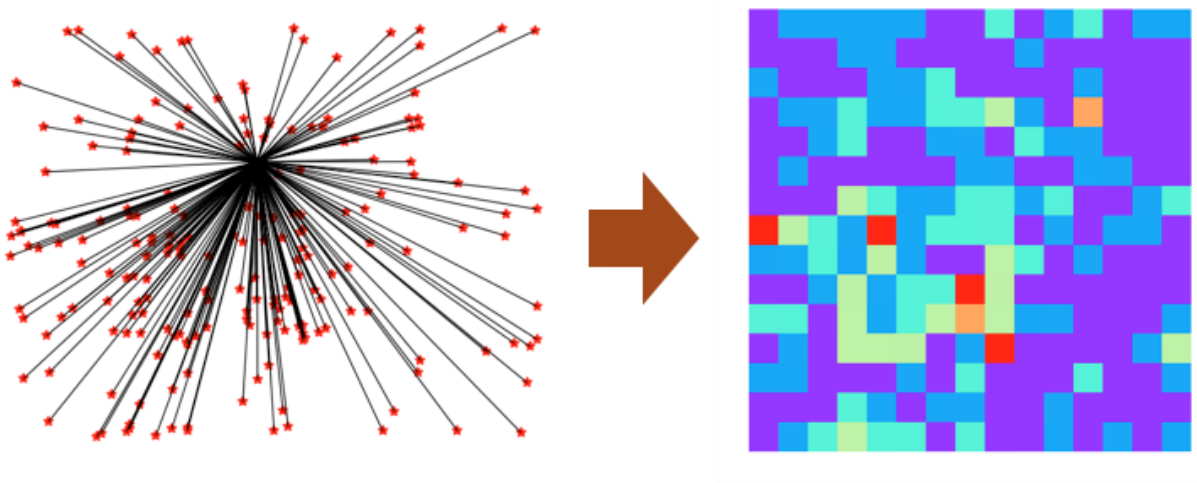


Modelling observables



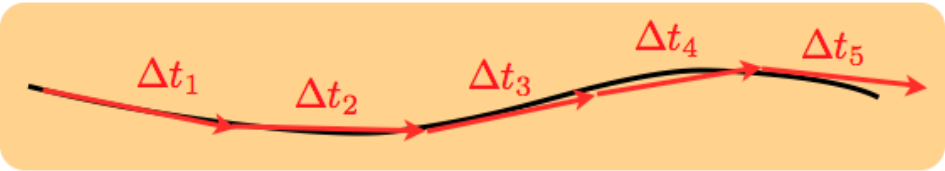
N-body simulations

- Sample discretely the phase space by point-like particles
- Estimate forces



+ FFT

- Integrate equation of motion



The need for approximate methods

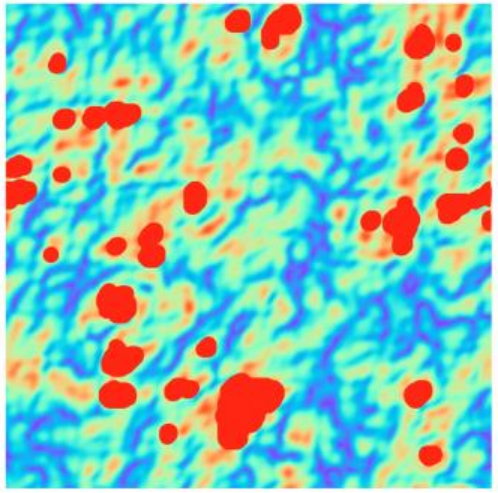
- **Problem**
  - Estimate **covariance matrices** for weak lensing and clustering
  - Provide **mock catalogues** to calibrate and validate analysis pipelines
- **Requirements**
  - Sample large volumes (tens of Gpc^3)
  - Produce massive ensembles of realizations ($>10^3$)
  - Explore different cosmologies or gravitational models
- **Solution**
  - Wait until brute force is possible…
  - Efficient usage of computational resources

Approximate methods

- Avoid solving the most expensive part of a numerical simulation

I. Run a cheap evolution of the density field

II.
- Use a biasing prescription to populate the density field (log-normal, Quick Particle Mesh, PATCHY, EZmocks, HALOGEN)
- Identify collapsed regions as halos (COLA, PINNOCHIO, PTHalos)

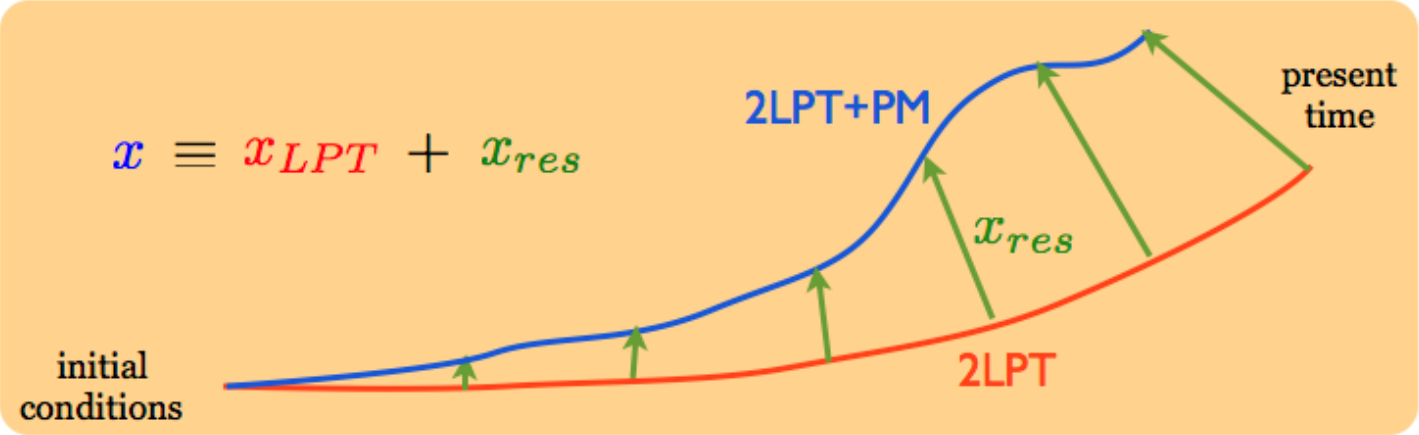


The COLA method

COmoving Lagrangian Acceleration (Tassev et al. 2013):

Use Perturbation Theory for an approximated guess of the solution

Add residual displacements with a cheap numerical N-body solver



$$\partial_t^2 \boldsymbol{x}_{\rm res}(t) = -\nabla\Phi(t) - \partial_t^2 \boldsymbol{x}_{\rm LPT}(t),$$

Particle positions $\xrightarrow[\text{Interpolation}]{\text{Poisson equation}}$ Forces at particle positions $\longrightarrow$ Residual acceleration wrt 2LPT $\xrightarrow{\text{Equation of motion}}$ integrate trajectory (time step)

Particle-Mesh

COLA method

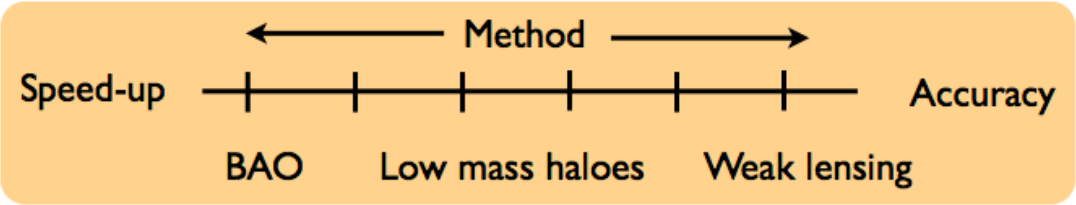
Comparison with other approximate methods

PROS

- ✓ Large scale dynamics is exact
- ✓ Accuracy at small scales is adjustable
- ✓ 2-3 orders of magnitude faster than conventional N-body simulations
- ✓ The dark matter field is available

CONS

- ❖ Large memory consumption
- ❖ Not as fast as fast methods using biasing prescriptions



Starting point

- COLA, Tassev et al. 2013
- Parallel Cola, Koda et al. 2015
  - Gaussian 2LPT initial conditions
  - Parallelized COLA algorithm
  - Halo finder on-the-fly for comoving snapshots
- Fiducial N-body simulation: MICE-GC

MICE

Marenostrum Institut de Ciències de l'Espai Simulations

Optimization of the simulation method

Optimization of code parameters

- Cheap numerical evolution:
  - Particle-Mesh algorithm
  - Time steps

Parameters:

- Mesh size
- Particle mass

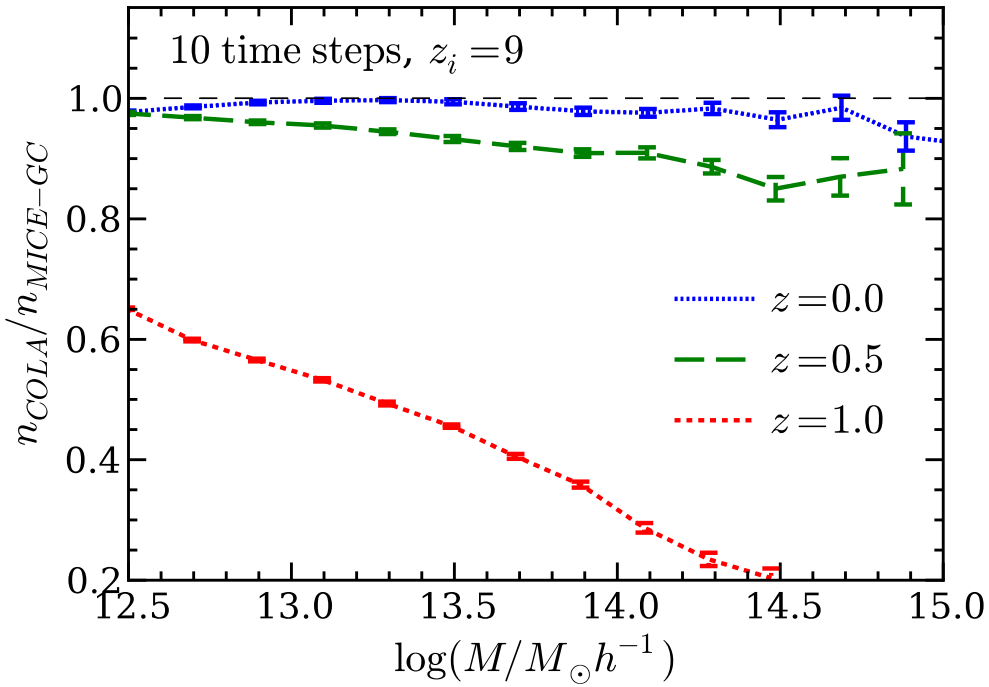
- Number of time steps
- Distribution along time
- Initial redshift

- Target observables
  - Matter power spectra
  - Halo abundance and clustering

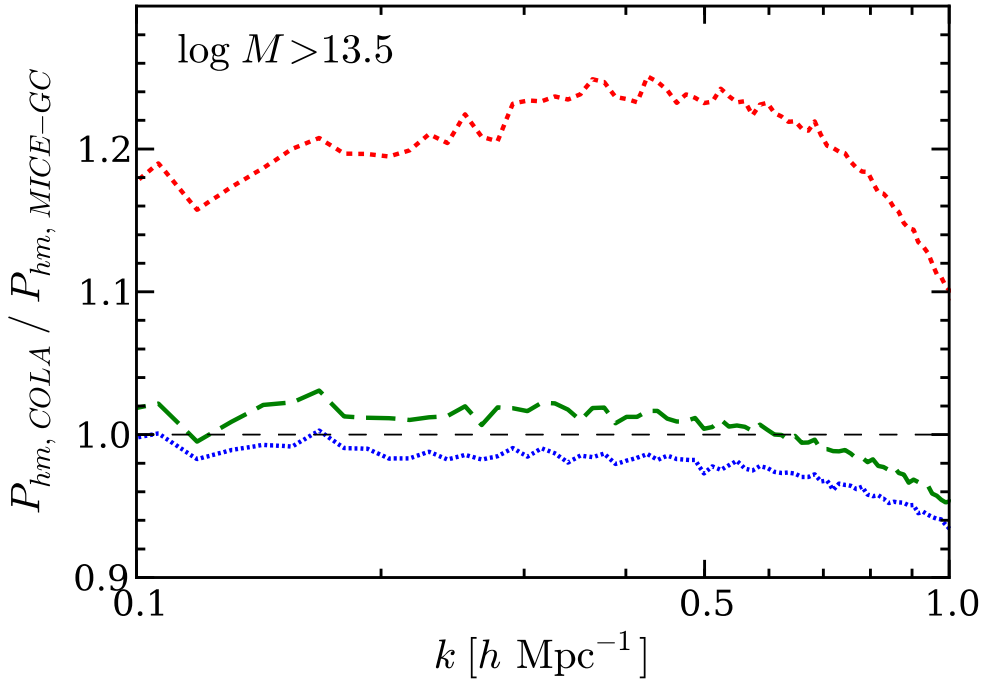
Limitations of 10 time steps

- Halo formation is not accurately captured before 10 time steps

Halo abundance

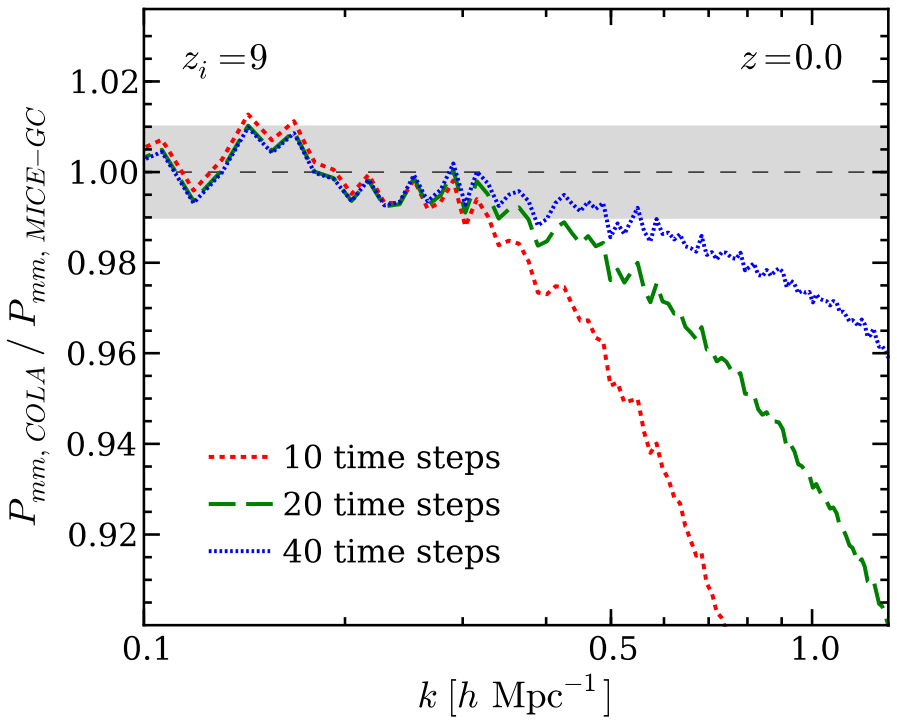


Halo-matter cross-correlation

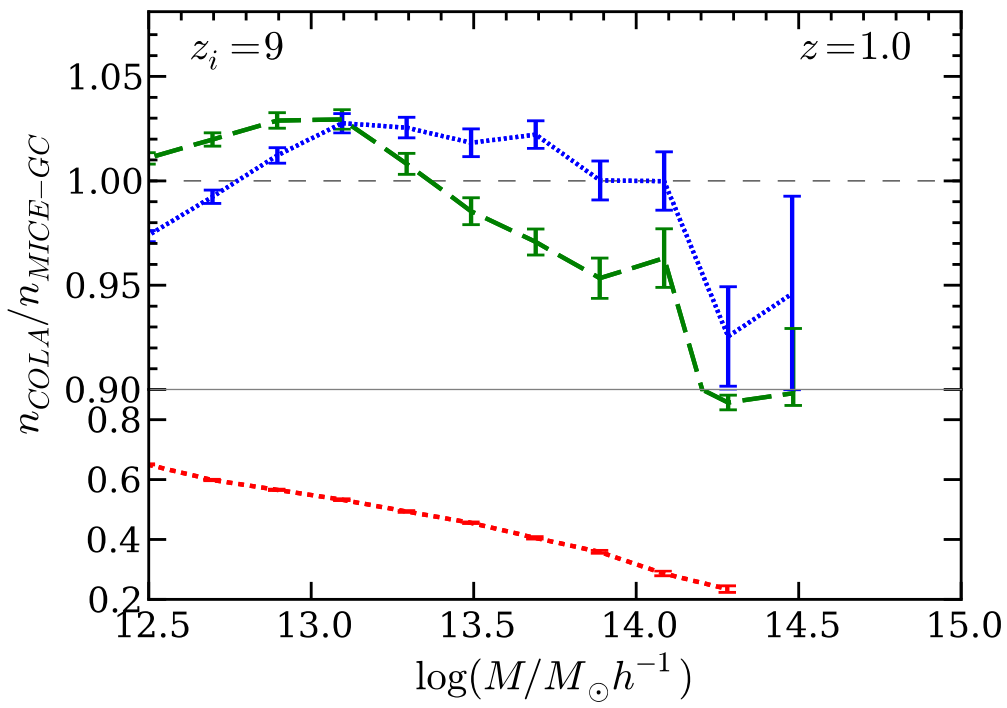


Number of time steps

Matter power spectrum



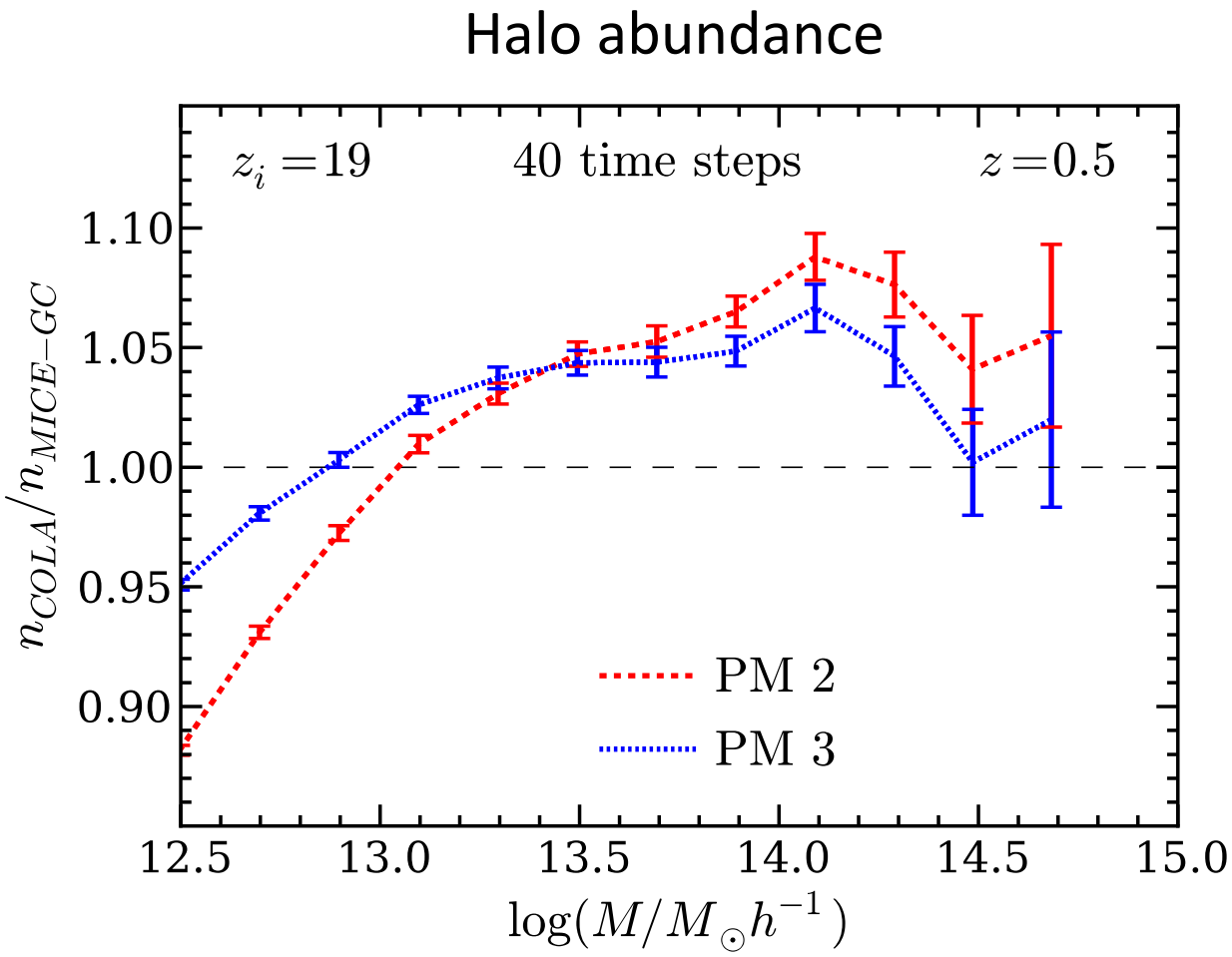
Halo abundance



Trade-off ↕ accuracy at small scales vs simulation time

> 10 time steps before the redshift of interest

Size of the force mesh



- PM grid 3 times finer than particle separation is necessary to resolve ~100 particle halos

Optimal set-up

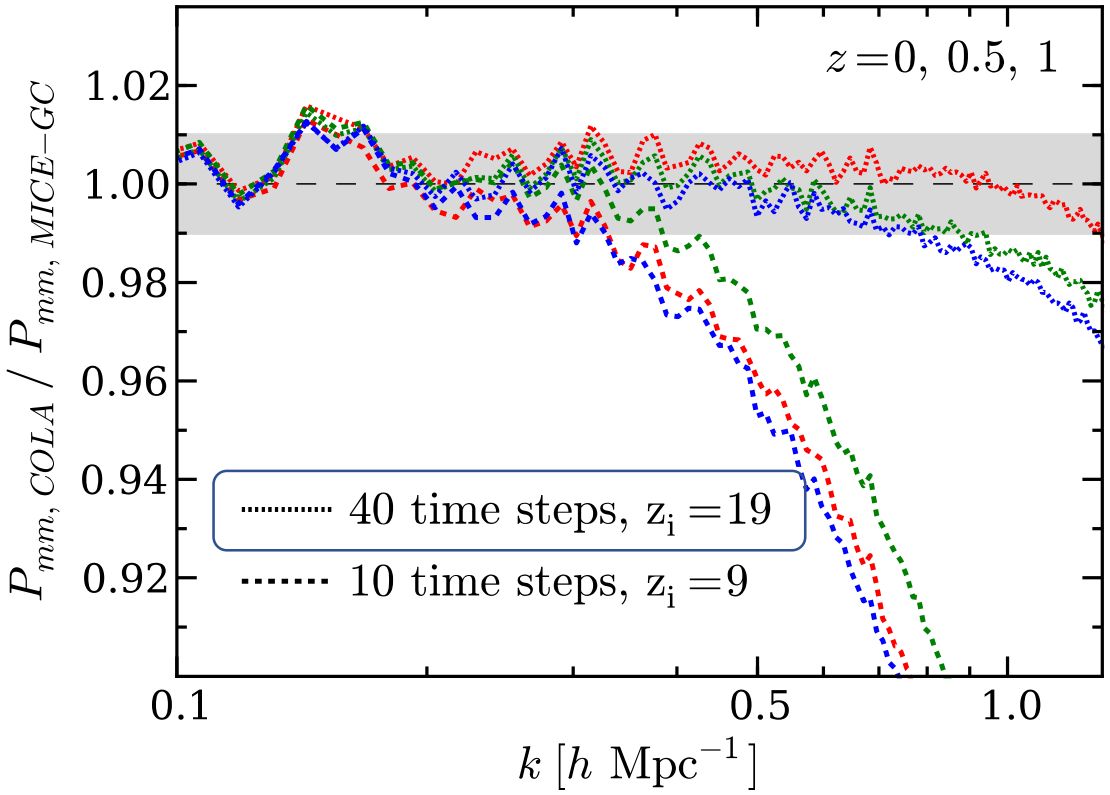
40 time steps

Linear time sampling distribution with the scale factor

$z_i = 19$

$PM_{grid} = 3$

Matter power spectrum

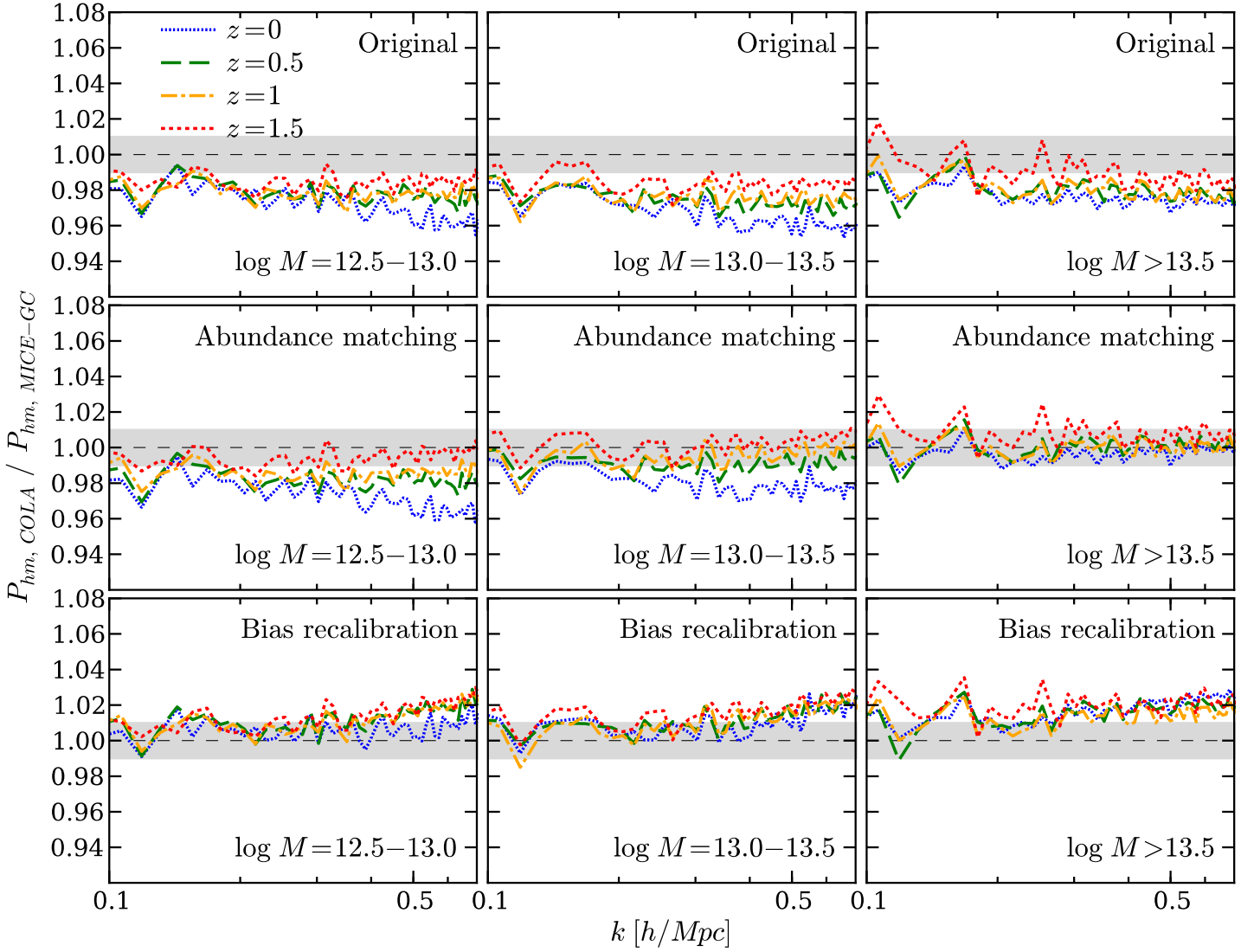


- 2048^3 particles run on 1024 cores in 40 minutes (2.7 Tb memory)
- 1% agreement up to k ~ 1 h/Mpc
- 5% accuracy in the mass function

Halo clustering

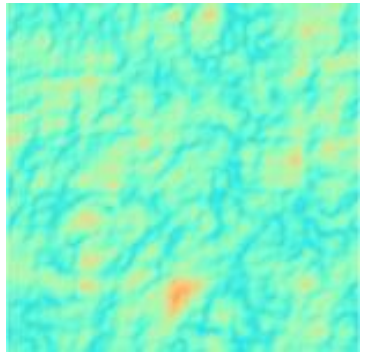
Halo bias is recovered within ~2% without applying any correction

Halo-matter cross-power spectrum

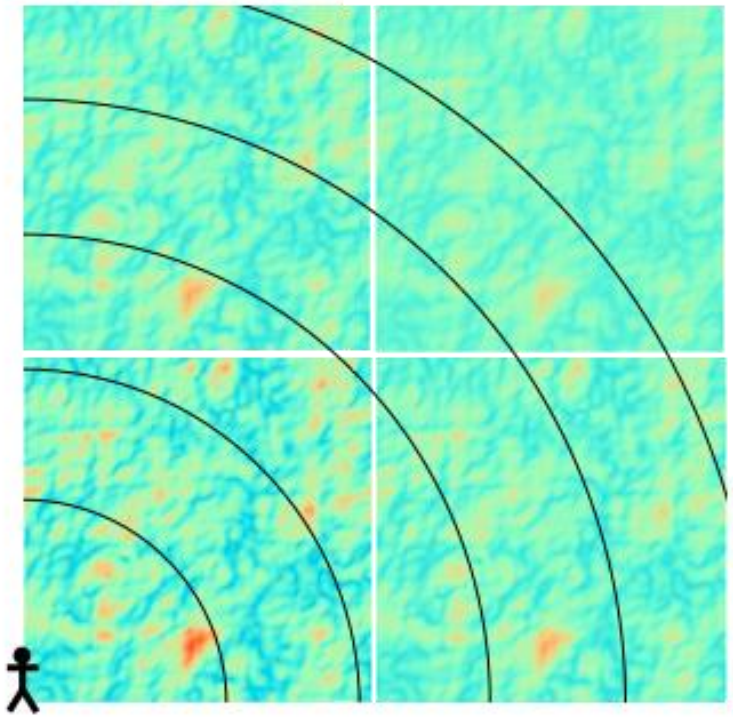


Generation of light cone catalogs

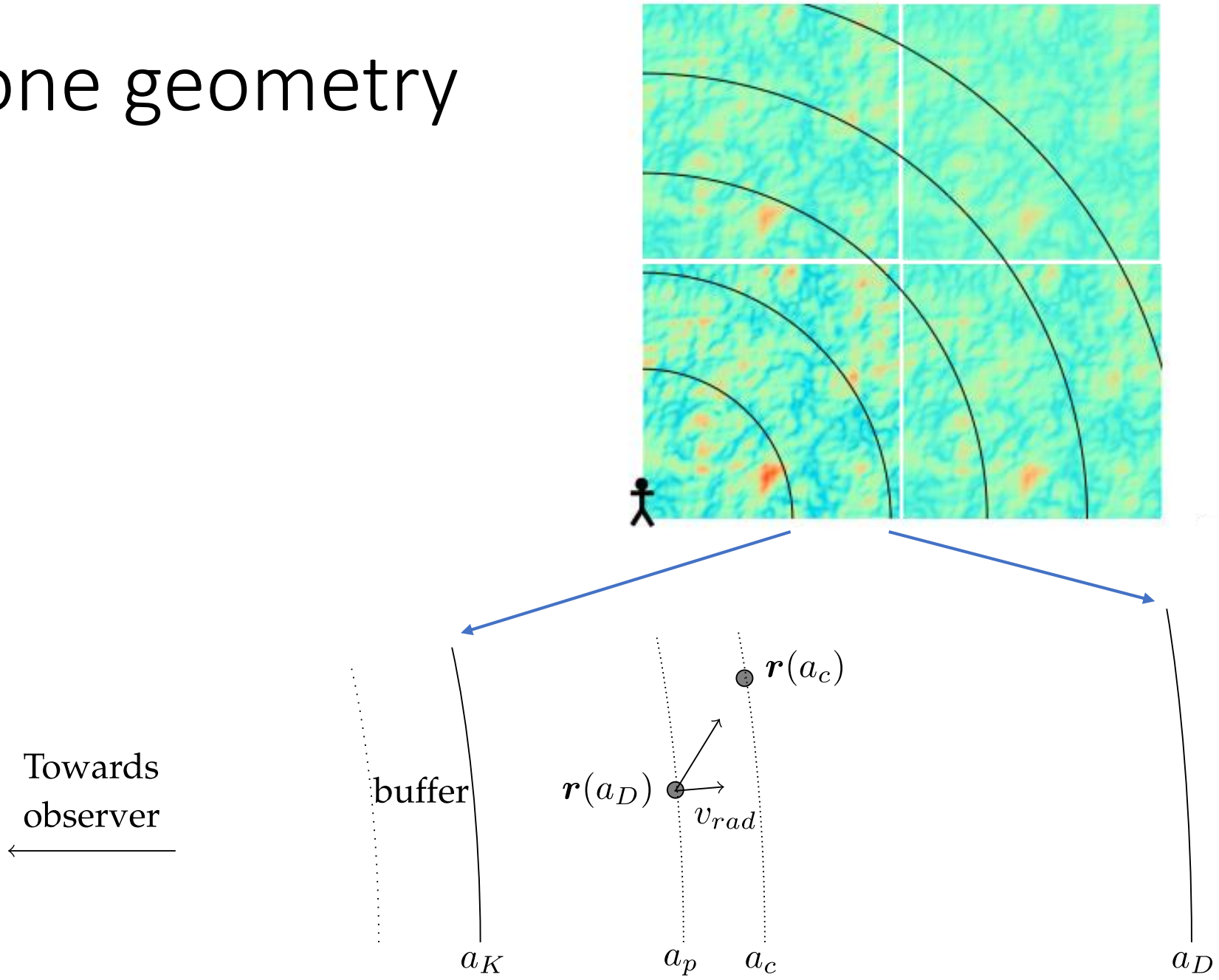
Light cone geometry



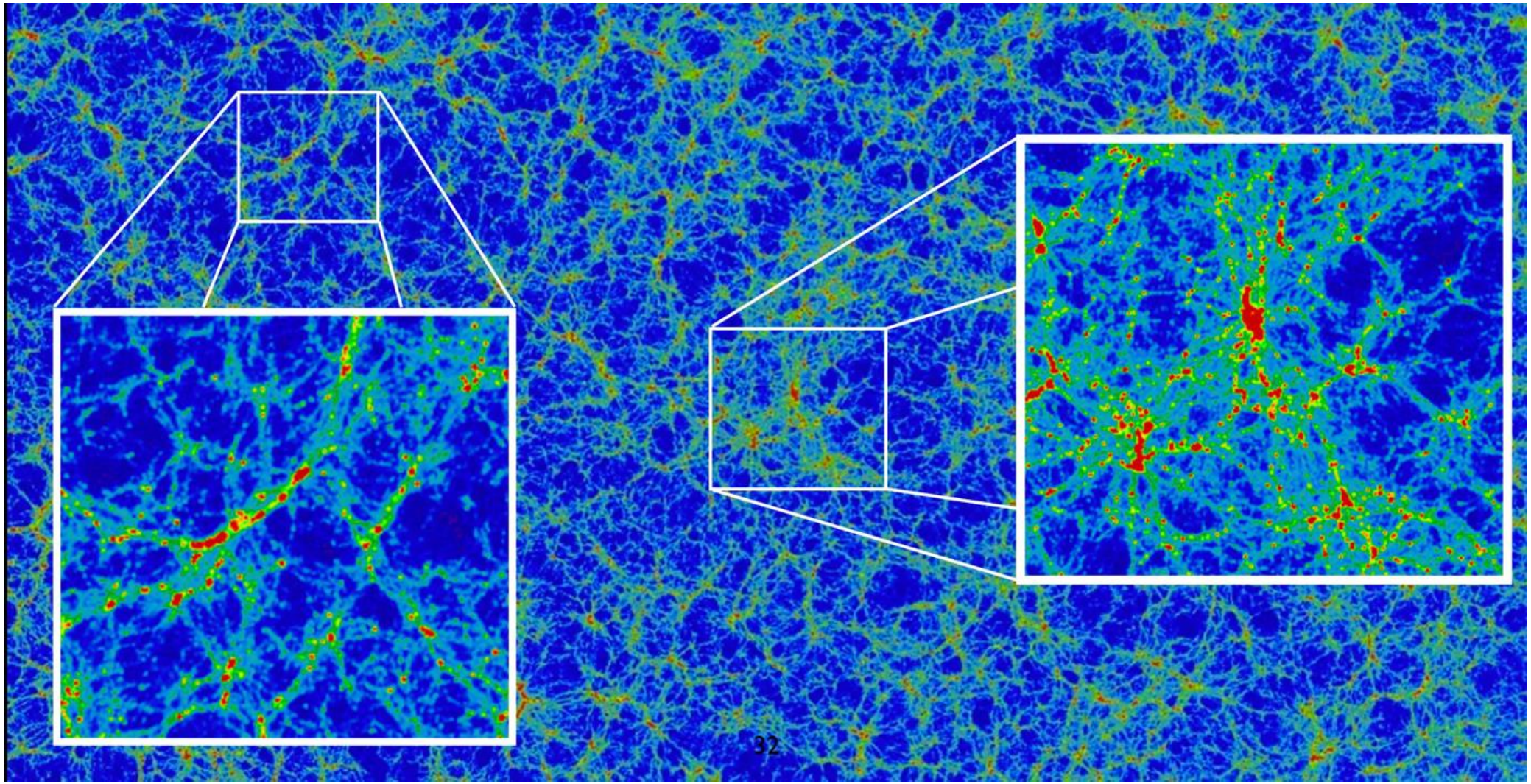
Light cone geometry



Light cone geometry

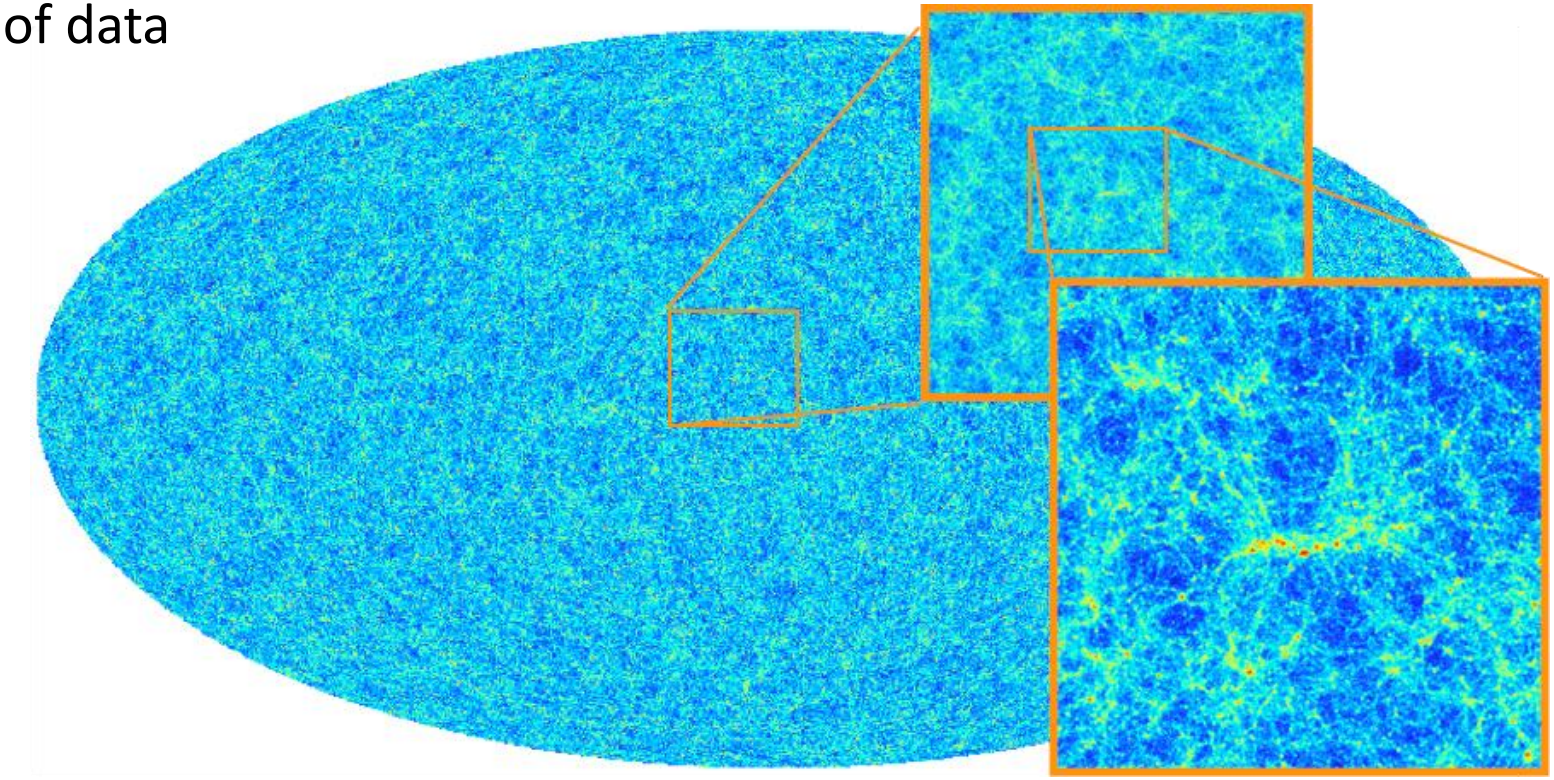


3D particle light cone



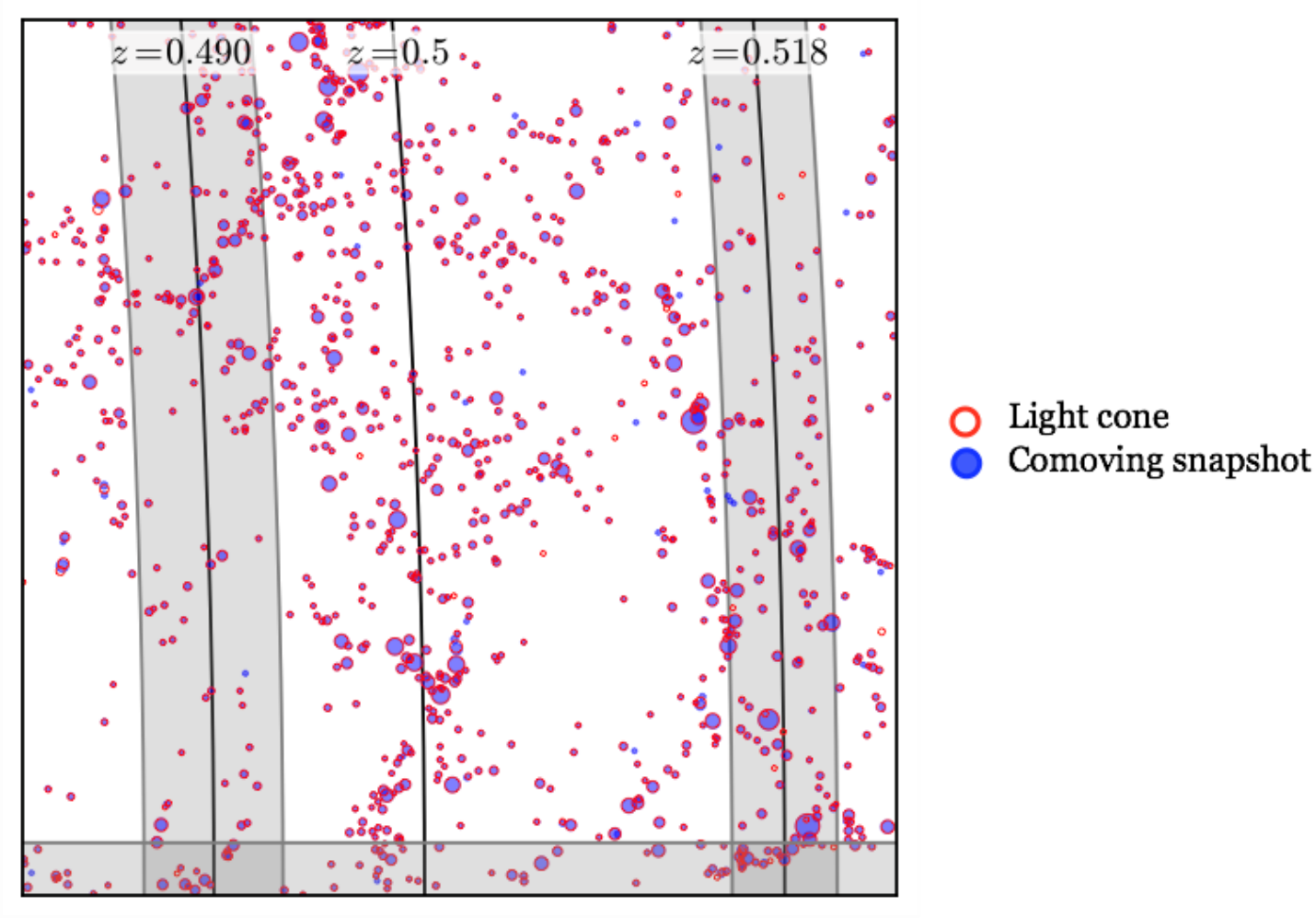
2D projected matter field

- Originally proposed and implemented by Fosalba et al. 2008, using the Healpix discretization. Now adapted to COLA.
  - 265 concentric shells up to z=1.4
  - 50M pixels (1.7 arcmin)
  - 50 Gb of data



Halo catalogues

- Friends-of-Friends objects
- Buffer zones around edges

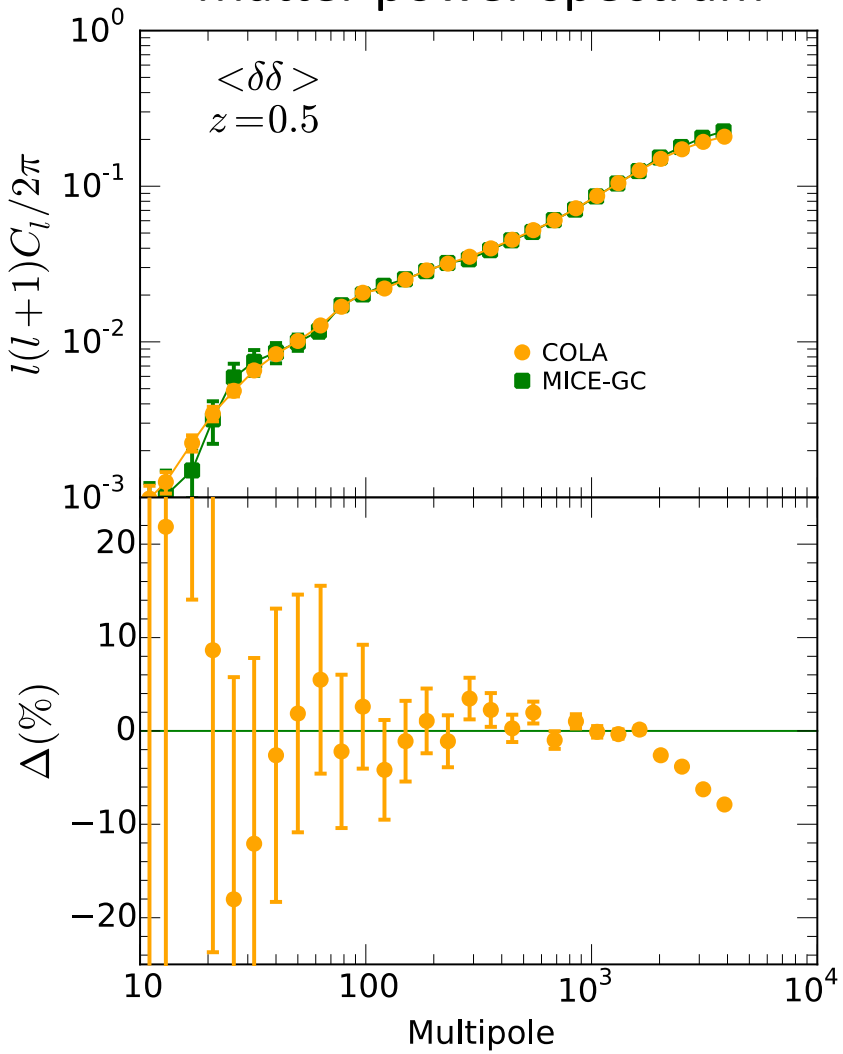


Modeling of weak gravitational lensing

The challenge of simulating weak lensing

- Weak lensing samples:
  - Large volumes
  - The matter distribution at small scales
- Future WL experiments will narrow down error-bars

Matter power spectrum



Weak lensing pipeline

SIMULATION

2048^3 particles
Lbox = 1536 Mpc/h
$m_p = 3x10^{10} M_{sun}$

z. ≤ 1.4
All sky
2 box replicas (64 total)

1024 cores
~2h / run
~300 realizations

Weak lensing pipeline

SIMULATION

2048^3 particles
Lbox = 1536 Mpc/h
$m_p = 3x10^{10} M_{sun}$

z. ≤ 1.4
All sky
2 box replicas (64 total)

1024 cores
~2h / run
~300 realizations

On-the-fly →

Projected matter density field (2D, Healpix format)

Weak lensing pipeline

SIMULATION

2048^3 particles
Lbox = 1536 Mpc/h
$m_p = 3\times10^{10}\ M_{sun}$

z. ≤ 1.4
All sky
2 box replicas (64 total)

1024 cores
~2h / run
~300 realizations

On-the-fly

Projected matter density field (2D, Healpix format)

$$\kappa(\boldsymbol{\theta}_i,\chi) = \frac{3H_0^2\Omega_m}{2c^2}\sum_j^{\chi_j<\chi}\delta(\boldsymbol{\theta}_i,\chi_j)\frac{(\chi-\chi_j)\chi_j}{a_j\chi}\Delta\chi_j$$

Born approximation

Convergence maps

$$\hat{\epsilon}_{lm} = -\sqrt{\frac{(l+2)(l-1)}{l(l+1)}}\hat{\kappa}_{lm}$$

Harmonic space (all sky)

$$\gamma_1 = -\sum_{lm}\hat{\epsilon}_{lm}X_{1,lm}$$

$$\gamma_2 = \sum_{lm} i\hat{\epsilon}_{lm}X_{2,lm},$$

Shear maps

Weak lensing pipeline

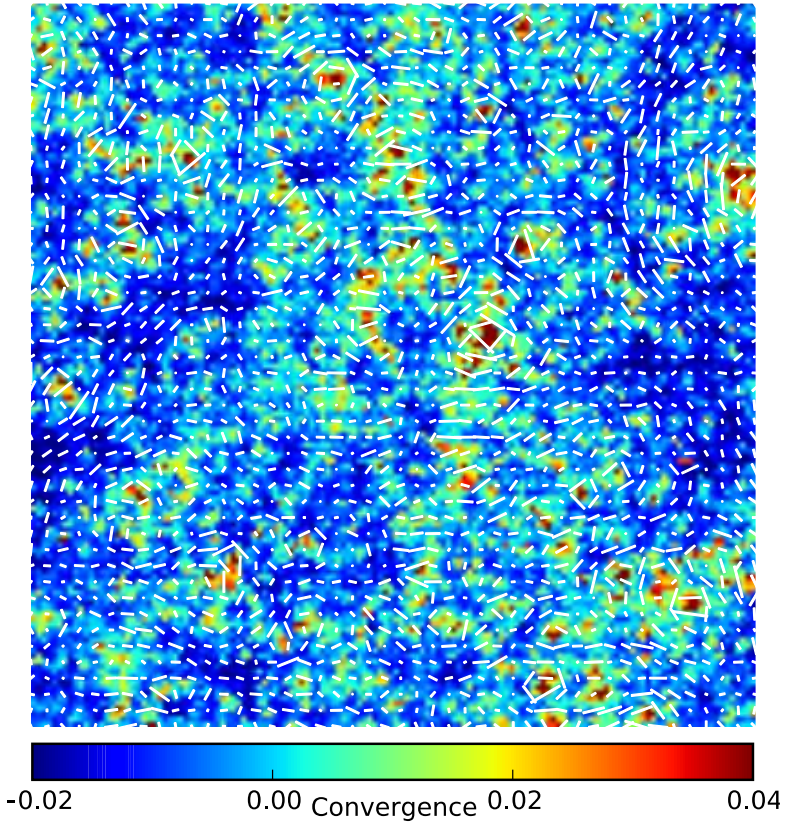
SIMULATION

2048^3 particles
Lbox = 1536 Mpc/h
$m_p = 3x10^{10} M_{sun}$

z. ≤ 1.4
All sky
2 box replicas (64 total)

1024 cores
~2h / run
~300 realizations

On-the-fly



Projected matter density field (2D, Healpix format)

Born approximation

Convergence maps

Harmonic space (all sky)

Shear maps

Weak lensing pipeline

SIMULATION

2048^3 particles
Lbox = 1536 Mpc/h
$m_p = 3x10^{10} M_{sun}$

z. ≤ 1.4
All sky
2 box replicas (64 total)

1024 cores
~2h / run
~300 realizations

On-the-fly → Projected matter density field (2D, Healpix format)

↓ Born approximation

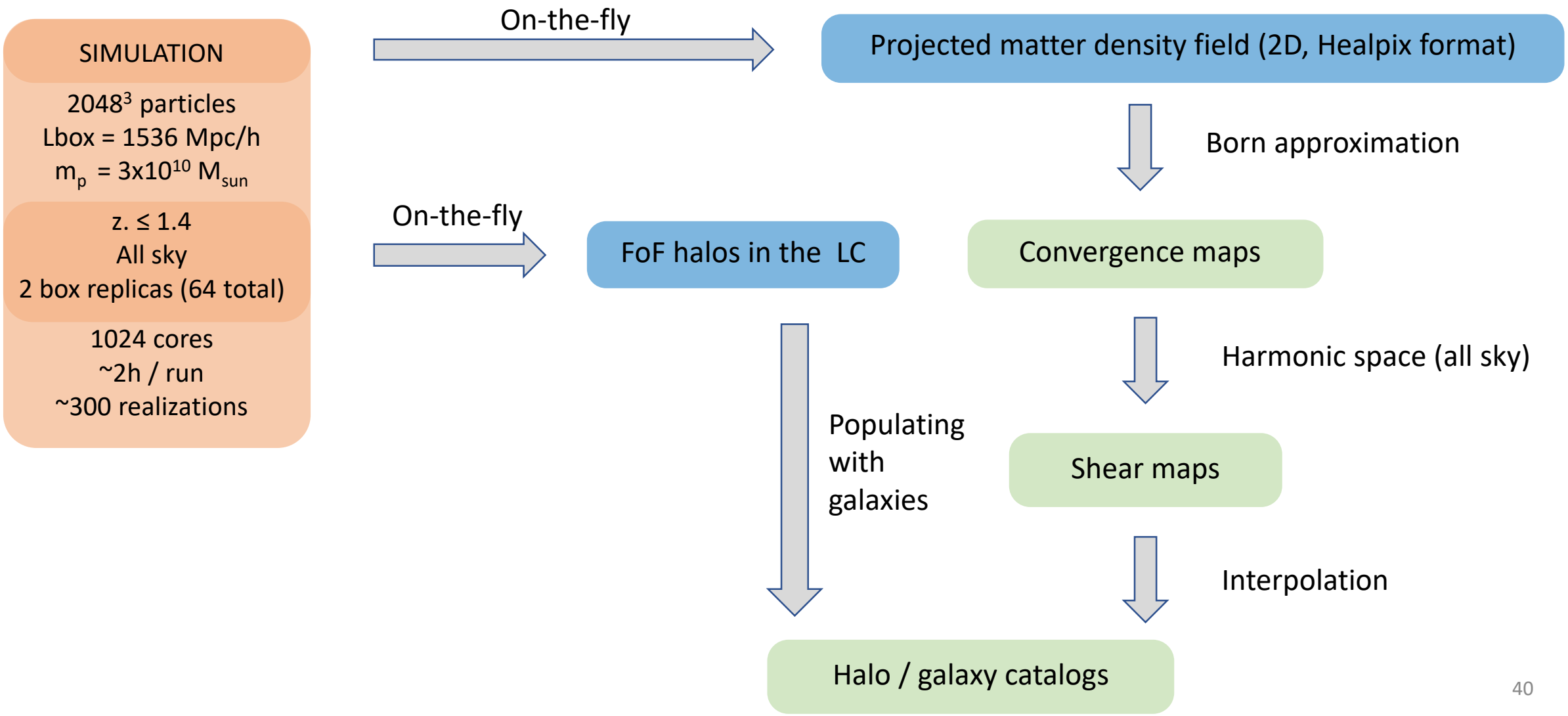
Convergence maps

↓ Harmonic space (all sky)

Shear maps

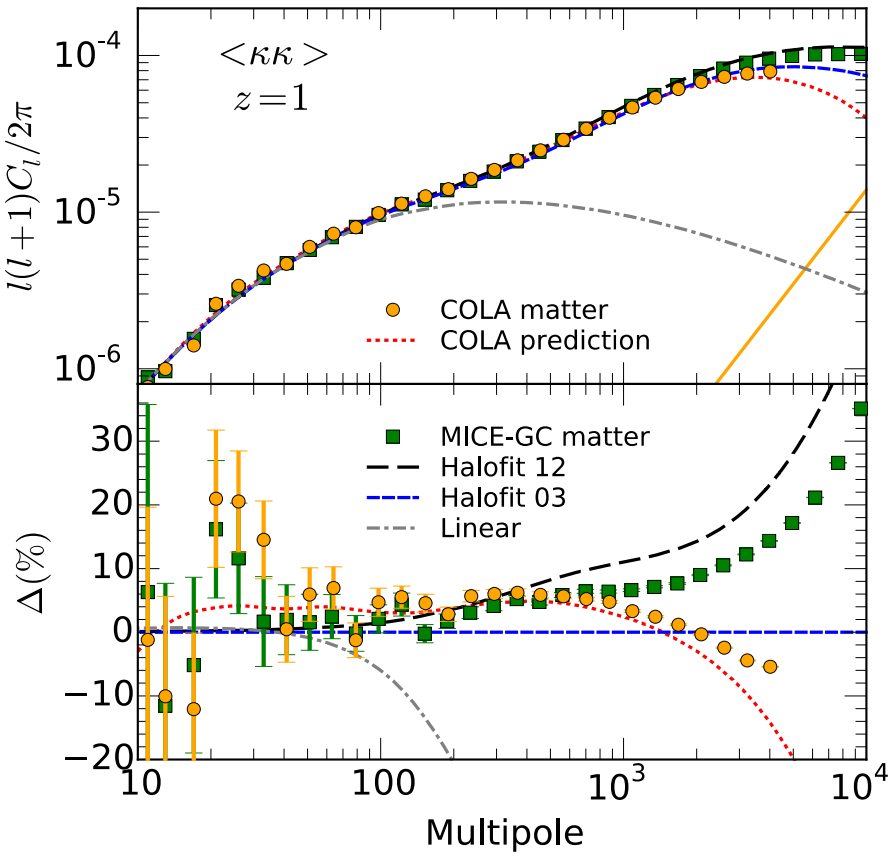
On-the-fly → FoF halos in the LC

Weak lensing pipeline



Convergence maps

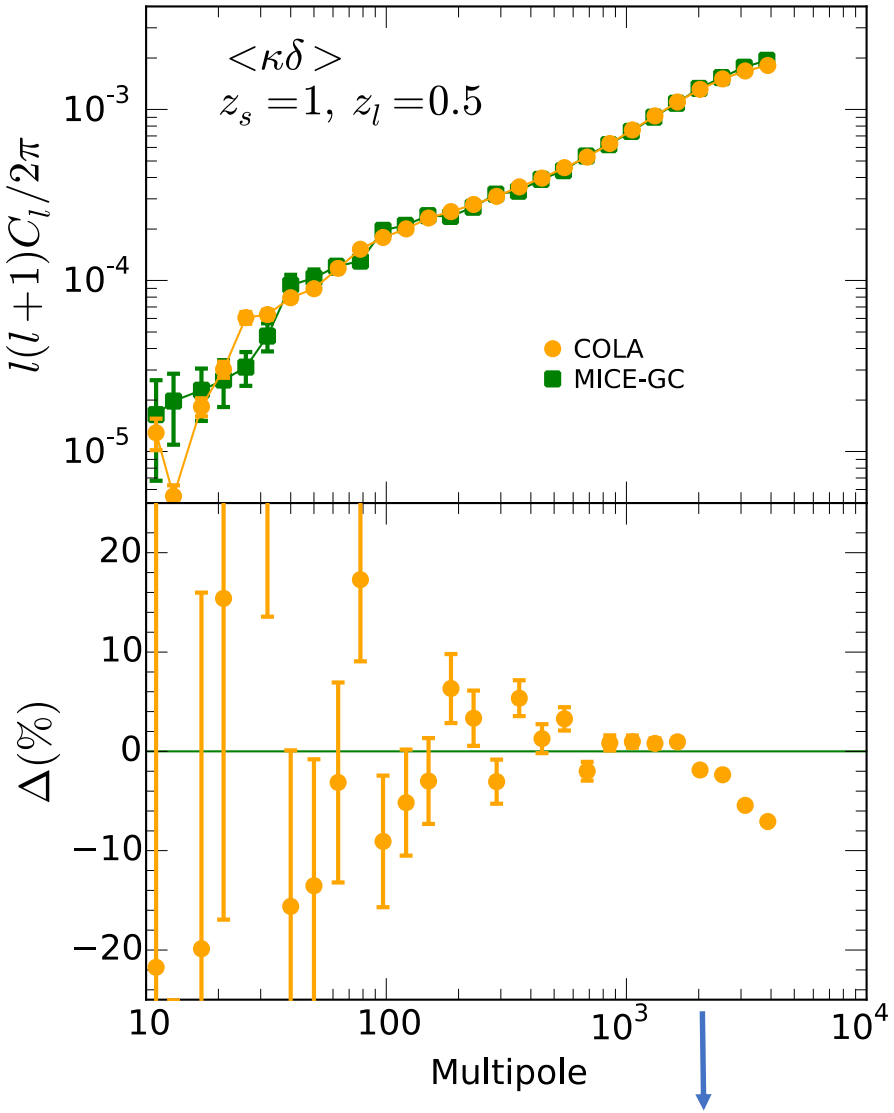
Convergence power spectrum



Prediction: $C_l^\kappa = \frac{9H_0^4\Omega_m^2}{4c^4}\int dr P(k,z)\frac{(r_s-r)^2}{r_s^2 a^2}$

non linear COLA

Convergence-matter cross-power spectrum

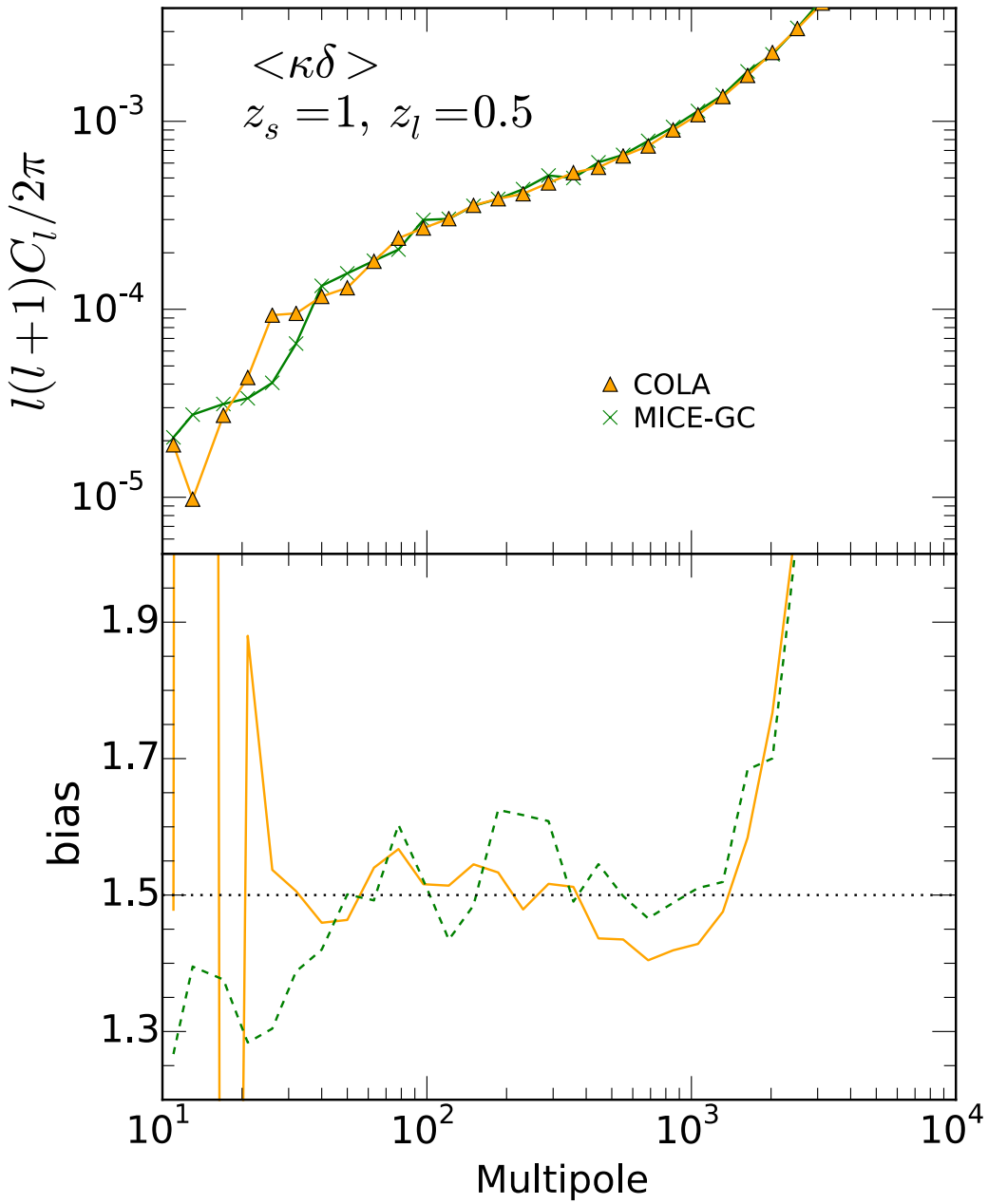


~2 Mpc/h (z=0.5)

Halo power spectra

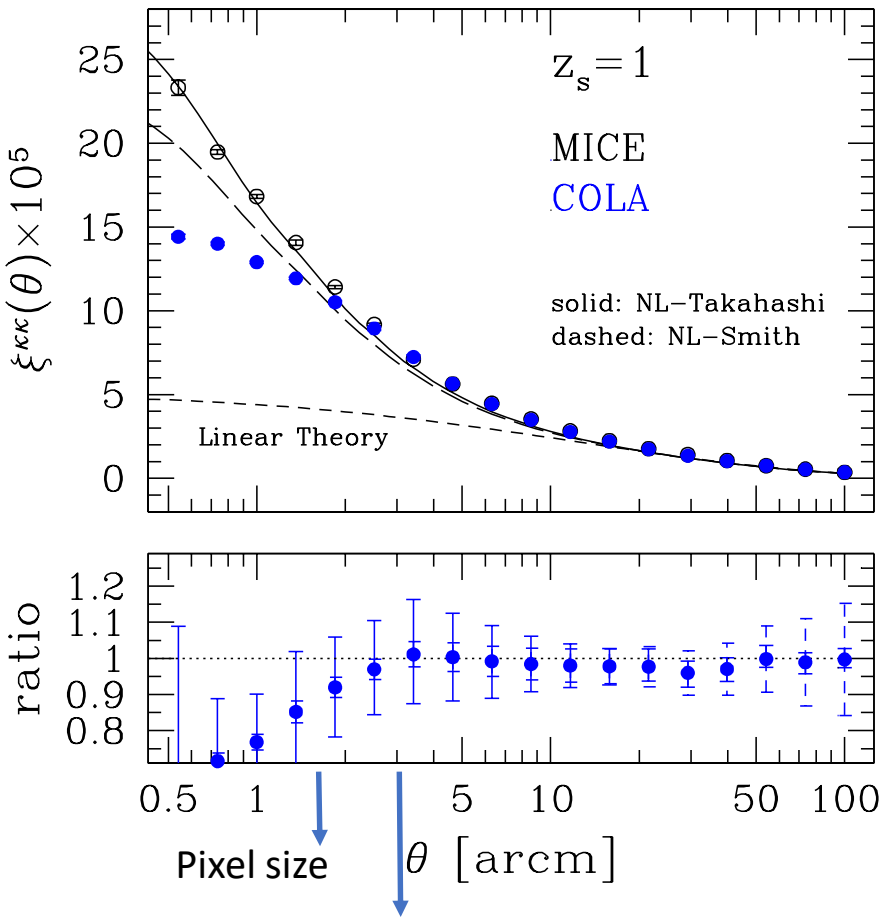
All sky halo sample

M > 3x10^{12} M$_{sun}$/h



Halo WL correlation functions

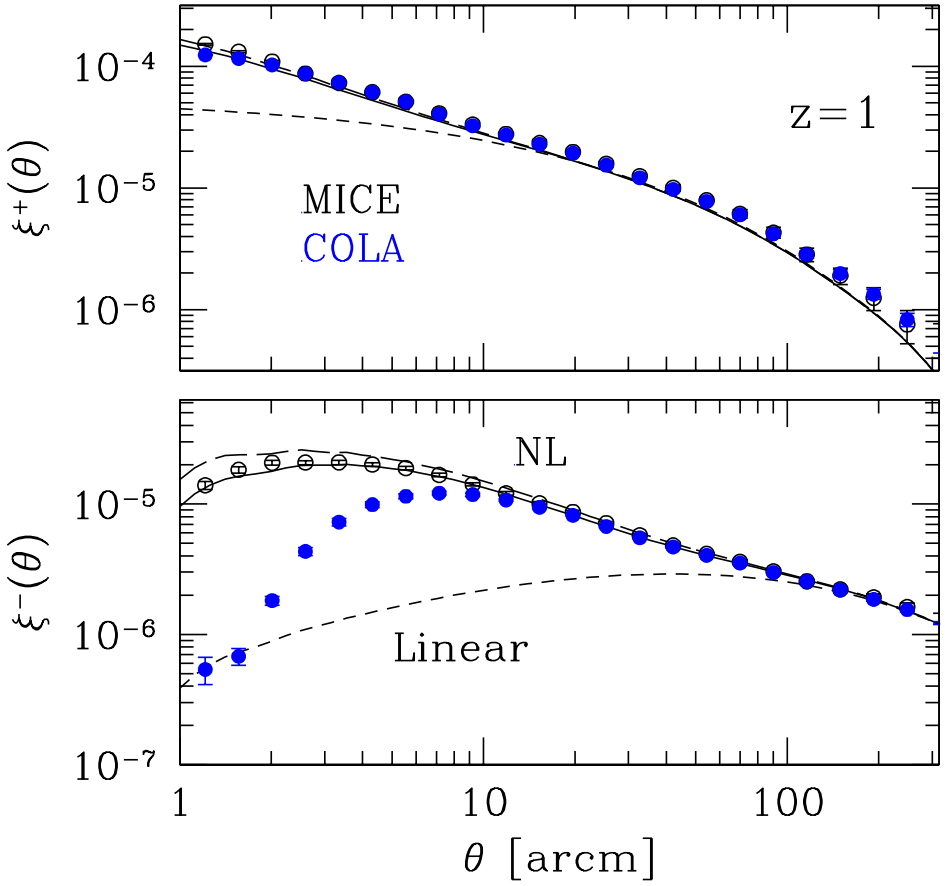
Convergence correlation function



Pixel size

~1 Mpc/h (z=0.5)

Shear correlation function



Implementing systematic effects and estimating covariance matrices

Impact of systematic effects on WL

- Cosmic shear probes non-linear scales and it is affected by:
  - Shape uncertainties
  - Photometric redshift errors
  - Survey selection function
- These may vary across the sky and in a coherent way, affecting both the signal and the covariance matrix
- To understand the error budget of the systematics on the observables we use the mock catalogs and model the conditional probability

$$P(\gamma_{obs}, z_{obs}, detection | \gamma_{true}, z_{true}, \theta)$$

which can take into account spatial correlations between shapes, colors and the selection function

Pipeline to populate halos with galaxies

Halos (x,v,z,M_h) $\xrightarrow{\text{SHMF, Mass-concentration, NFW profile}}$ **Sub-halos** (x,v,z,N_{sh},M_{sh}) $\xrightarrow{\text{Ghost (Phil)}}$ **Galaxies** (x,v,z,M_*,SFR)

Pipeline to populate halos with galaxies

Halos (x,v,z,M_h) $\xrightarrow{\text{SHMF, Mass-concentration, NFW profile}}$ **Sub-halos** (x,v,z,N_{sh},M_{sh}) $\xrightarrow{\text{Ghost (Phil)}}$ **Galaxies** (x,v,z,M_*,SFR) $\Rightarrow$

Ghost (Phil) or Stellar Pop. Synthesis models (SED) $\Rightarrow$ **Galaxies** (x,v,z,m_{obs}) $\xrightarrow{\text{+WL +Photo-z}}$ **Galaxies** $(x,v,z_{phot},m_{obs},\kappa,\gamma_1,\gamma_2)$ $\xrightarrow{\text{+mask +systematics (Eric)}}$ **Samples** $\Rightarrow$

$\Rightarrow$ **Measurements, Covariances**

Summary

- Future galaxy surveys will be dominated by **systematics** unless they are detected and mitigated.
- **Fast methods** are able to generate fast yet reasonably accurate galaxy mock catalogues.
- The methodology presented produces halo **catalogues in the light cone and weak lensing maps** in a consistent way.
- Weak lensing is accurately modeled to **scales one order of magnitude** beyond the transition to the non-linear regime.
- Developing a pipeline to populate **~300 simulations with galaxies** having WL quantities.

Thanks!

Displacement vector $x = q + s$ $s_{res} \equiv s - D_1 s_1 - D_2 s_2$

Equation of motion.

$$T^2[s_{res}] = -\frac{3}{2}\Omega_M a \partial_x \partial_x^{-2} \delta(x,a) - T^2[D_1]s_1 - T^2[D_2]s_2$$

Time operators
Drift and Kick

$$D(a_i, a_f): \; x(a_i) \to x(a_f)$$

$$K(a_i, a_f): \; v(a_i) \to v(a_f)$$

$$L_\pm(a): \; v(a) \to v(a) \pm (T[D_1](a)s_1 + T[D_2](a)s_2)$$

$$L_+(a_{n+1}) \left(\prod^n K(a_{i+\frac{1}{2}}, a_{i+1}) D(a_i, a_{i+1}) K(a_i, a_{i+\frac{1}{2}}) \right) L_-(a_0)$$

residual displacement $\boldsymbol{x}_{\rm res}(t) \equiv \boldsymbol{x}(t) - \boldsymbol{x}_{\rm LPT}(t).$

residual acceleration $\partial_t^2 \boldsymbol{x}_{\rm res}(t) = -\nabla\Phi(t) - \partial_t^2 \boldsymbol{x}_{\rm LPT}(t)$

Particle-Mesh $\boldsymbol{k}^2 \tilde{\phi}(\boldsymbol{k}) = 4\pi G \tilde{\rho}(\boldsymbol{k})$

Temporal operators

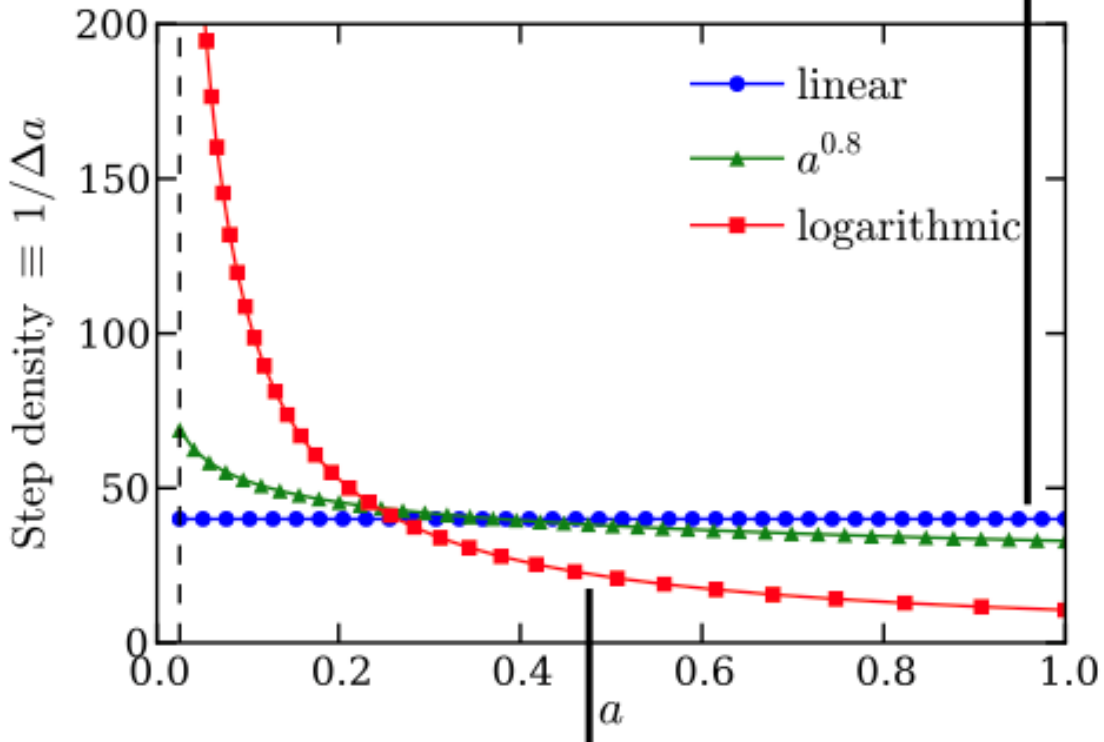
$$D(a_i, a_{i+1}): \quad \boldsymbol{s}(a_i) \mapsto \boldsymbol{s}(a_{i+1}) = \boldsymbol{s}(a_i) + \boldsymbol{v}(a_{i+1/2})\Delta t + [D_1(a_{i+1}) - D_1(a_i)]\boldsymbol{s}_1 + [D_2(a_{i+1}) - D_2(a_i)]\boldsymbol{s}_2$$

$$K(a_{i+1/2}, a_{i+3/2}): \quad \boldsymbol{v}(a_{i+1/2}) \mapsto \boldsymbol{v}(a_{i+3/2}) = \boldsymbol{v}(a_{i+1/2}) + \Delta t \times \left(-\frac{1}{2}\nabla\phi(a_{i+1}) - \partial_t^2 \boldsymbol{s}_1(t) - \partial_t^2 \boldsymbol{s}_2(t)\right).$$

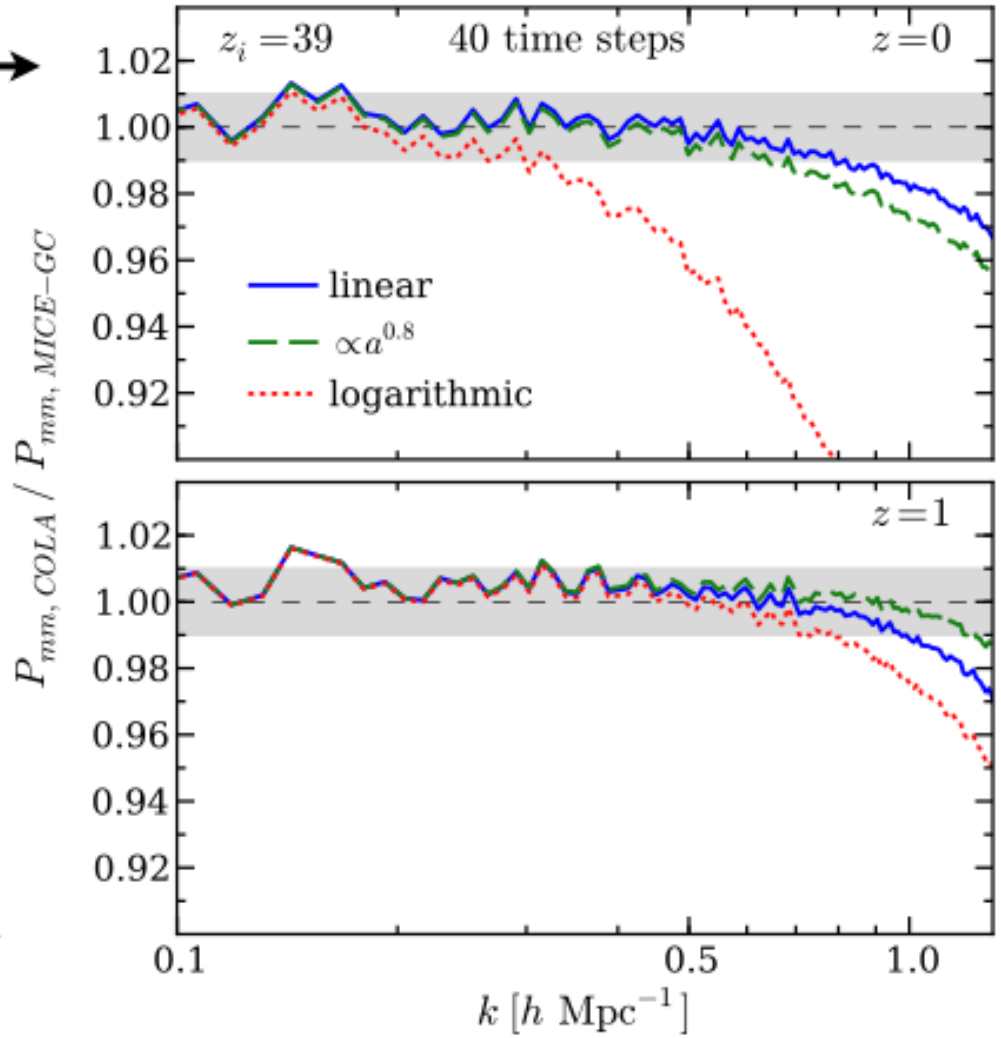
Complete evolution $$L_+(a)\left(\prod_{i=0}^{n} K(a_{i+\frac{1}{2}}, a_{i+\frac{3}{2}}) D(a_i, a_{i+1})\right) K(a_i, a_{i+\frac{1}{2}}) L_-(a)$$

Transform to 2LPT observer $$L_\pm(a): \quad \boldsymbol{v}(a) \mapsto \boldsymbol{v}(a) = \boldsymbol{v}(a) \pm (\partial_t \boldsymbol{s}_1(t) + \partial_t \boldsymbol{s}_2(t))$$

How to optimally distribute 40 time steps between the initial conditions and z=0?



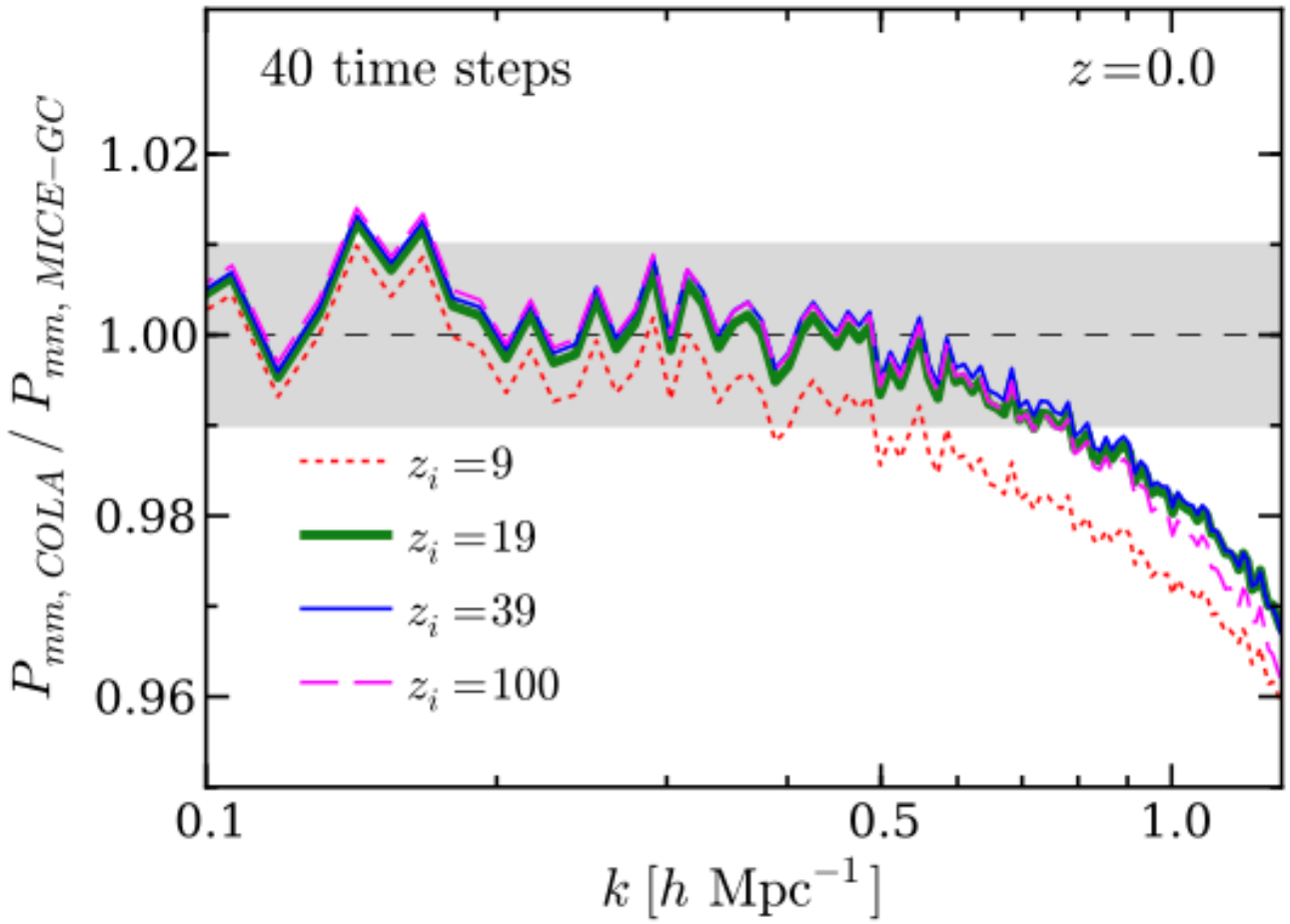
Matter power spectrum



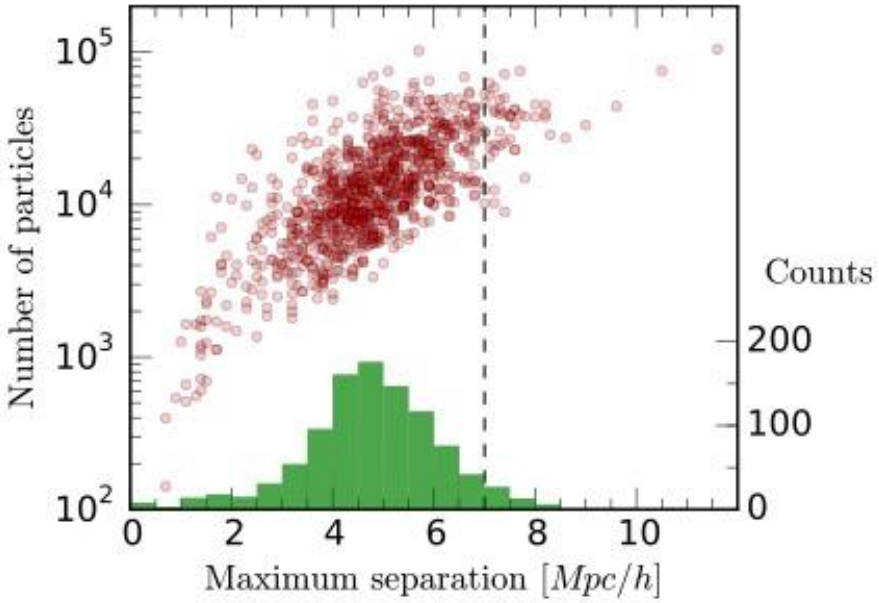
Distributions with varying step density are unbalanced

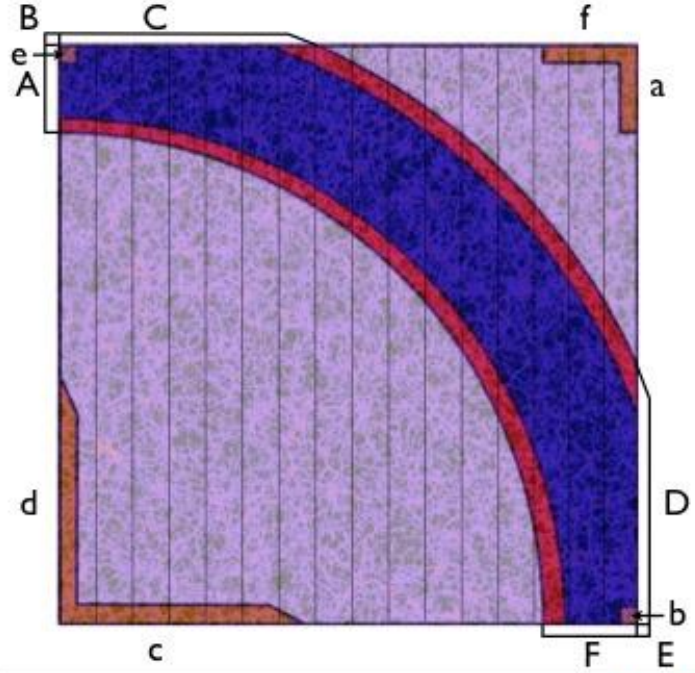
First guess: initial scale factor equal to the step width

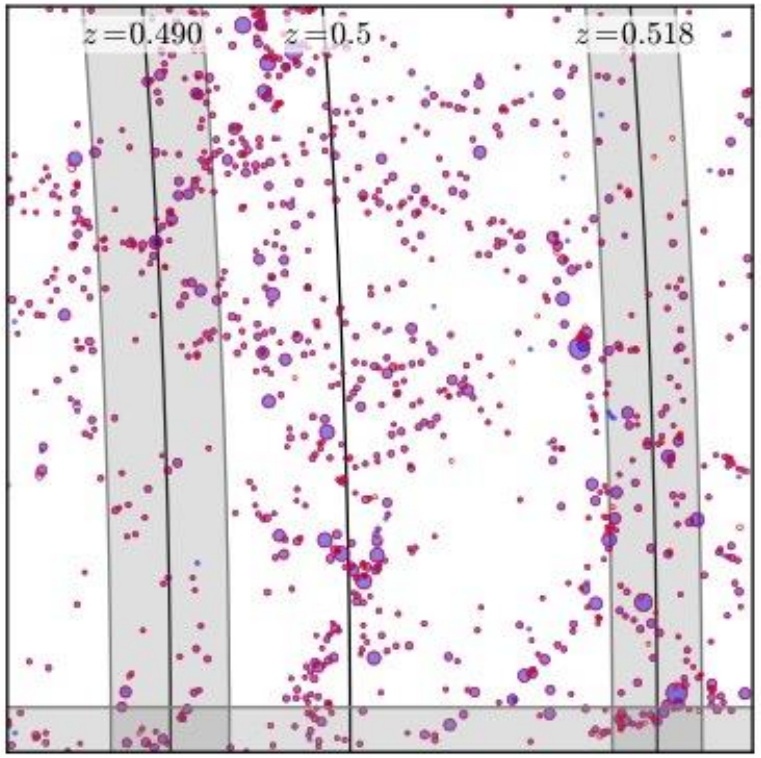
Matter power spectrum

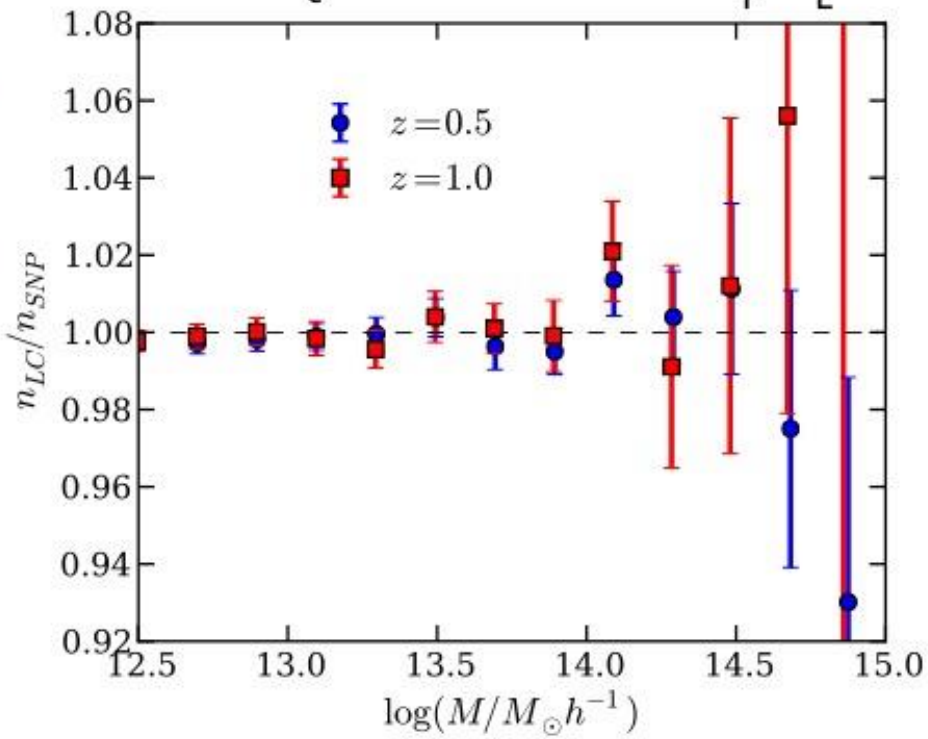


For 40 time steps, z_i=19 gives the best performance (Pk and MF)


Number of particles
Counts
Maximum separation [Mpc/h]


B
C
f
e
A
a
d
D
b
c
F
E


z=0.490
z=0.5
z=0.518


z=0.5
z=1.0
n_{LC}/n_{SNP}
$\log(M/M_\odot h^{-1})$