MAX-PLANCK-INSTITUT FÜR ASTROPHYSIK

ORIGINS Excellence Cluster

# Large-Scale Structure Observables: Overview

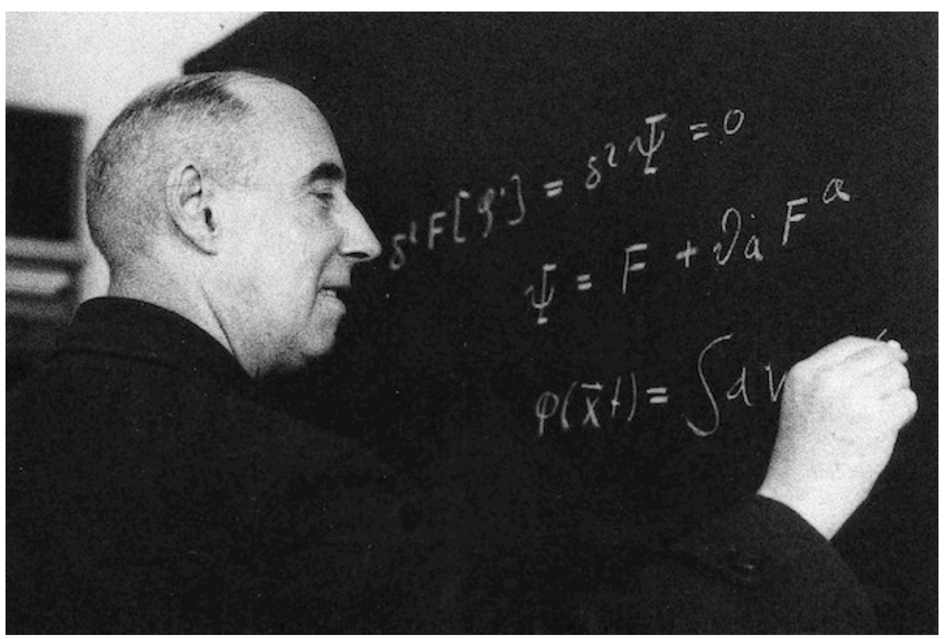

Fabian Schmidt
MPA


COSMOFondue

COSMOFondue, Geneva, June 12 2025

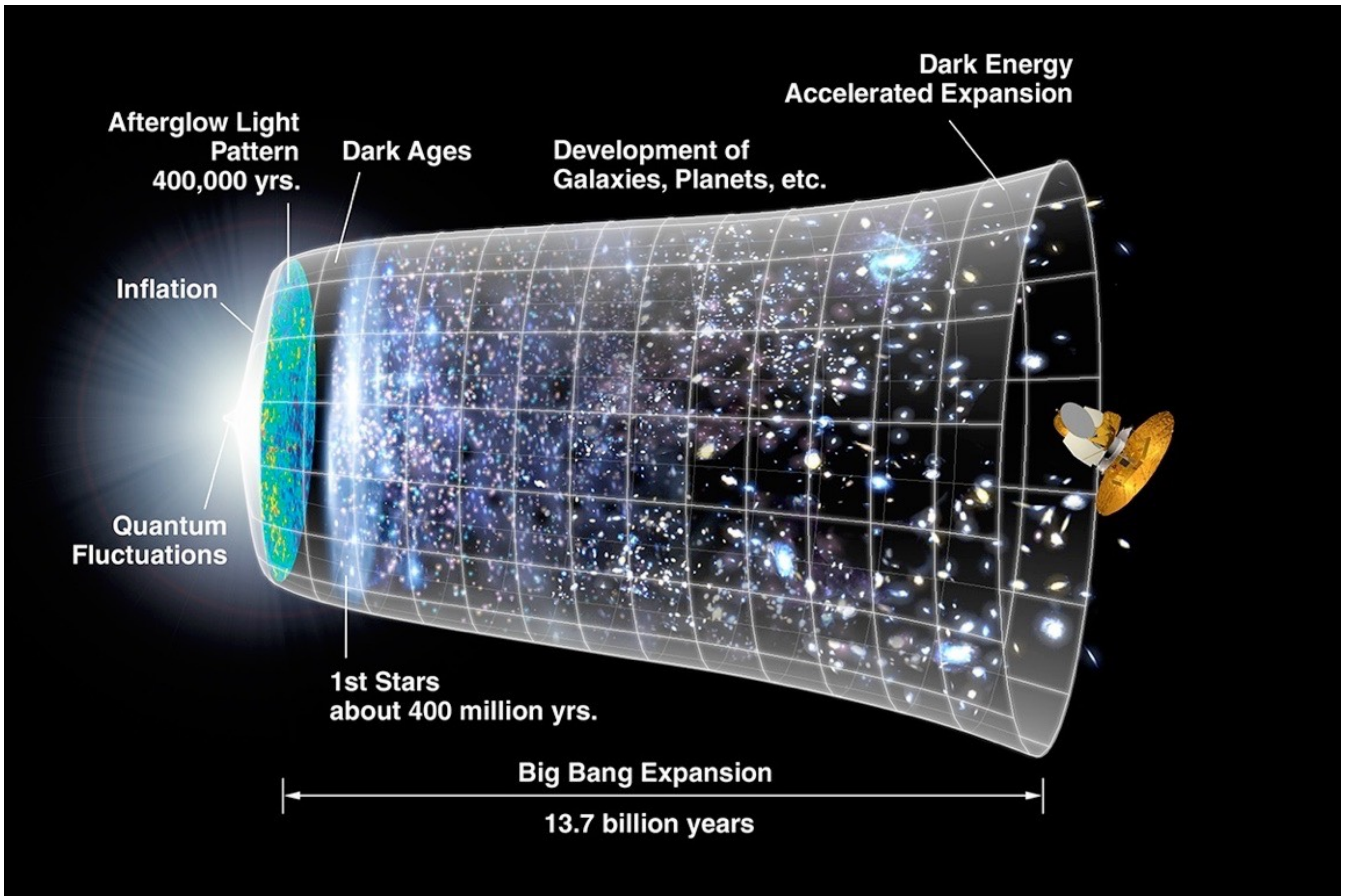
# CosmoFONDUE in Stückelberg's footsteps

- I couldn't figure out how to incorporate the *Stückelberg trick* into my talk…



# Motivation

- Using large-scale structure, we can learn about



# Motivation

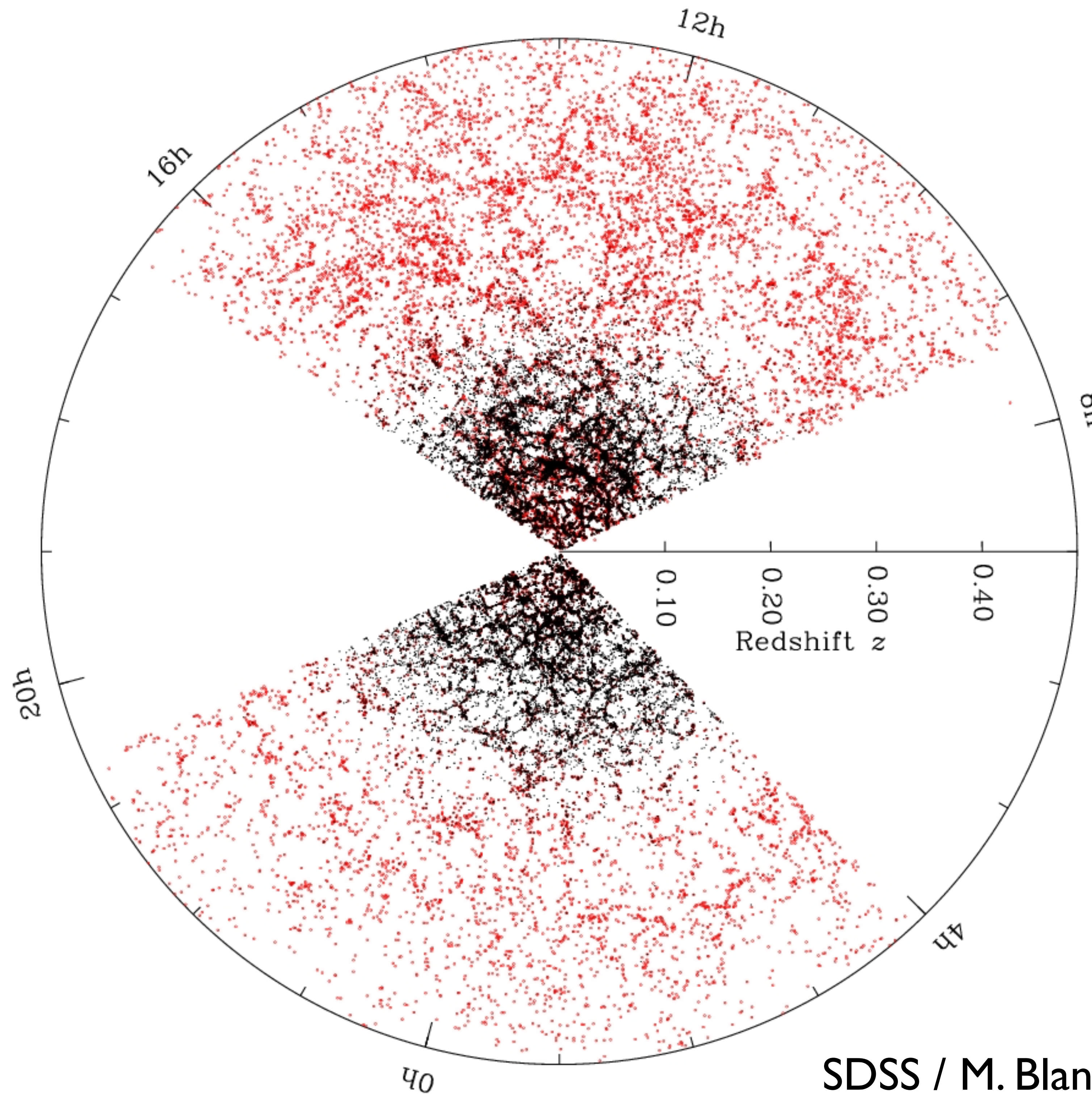
- Inflation: reconstruct the properties of the initial conditions, and look for gravitational waves $f_{\rm NL}$ $n_s,\ r$
- Dark Energy and Gravity: the growth of structure depends sensitively on the expansion history of the Universe, and the nature of gravity $w_0,\ w_a,\ f \propto D'/D$

  Growth equation: $D'' + aH\,D' = 4\pi\, G\, \bar{\rho}\, D$

- Dark Matter: how "cold" is cold dark matter ? What is the sum of neutrino masses ? $\sum m_\nu$

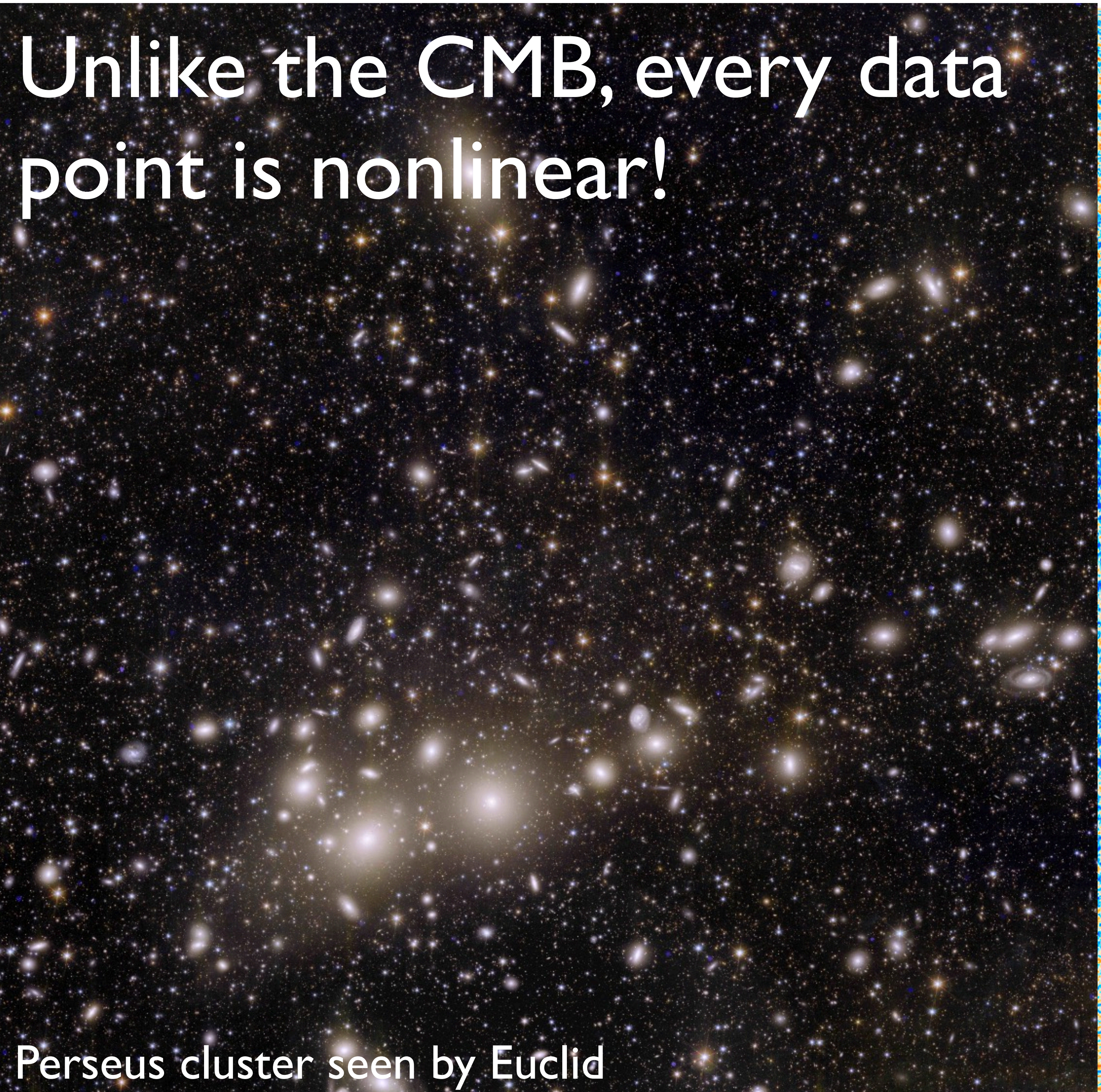
# Challenge: massive, highly nonlinear data set

- State of the art galaxy redshift surveys collect sky positions and redshifts for *millions* of galaxies
- Covering a substantial fraction of the observable universe

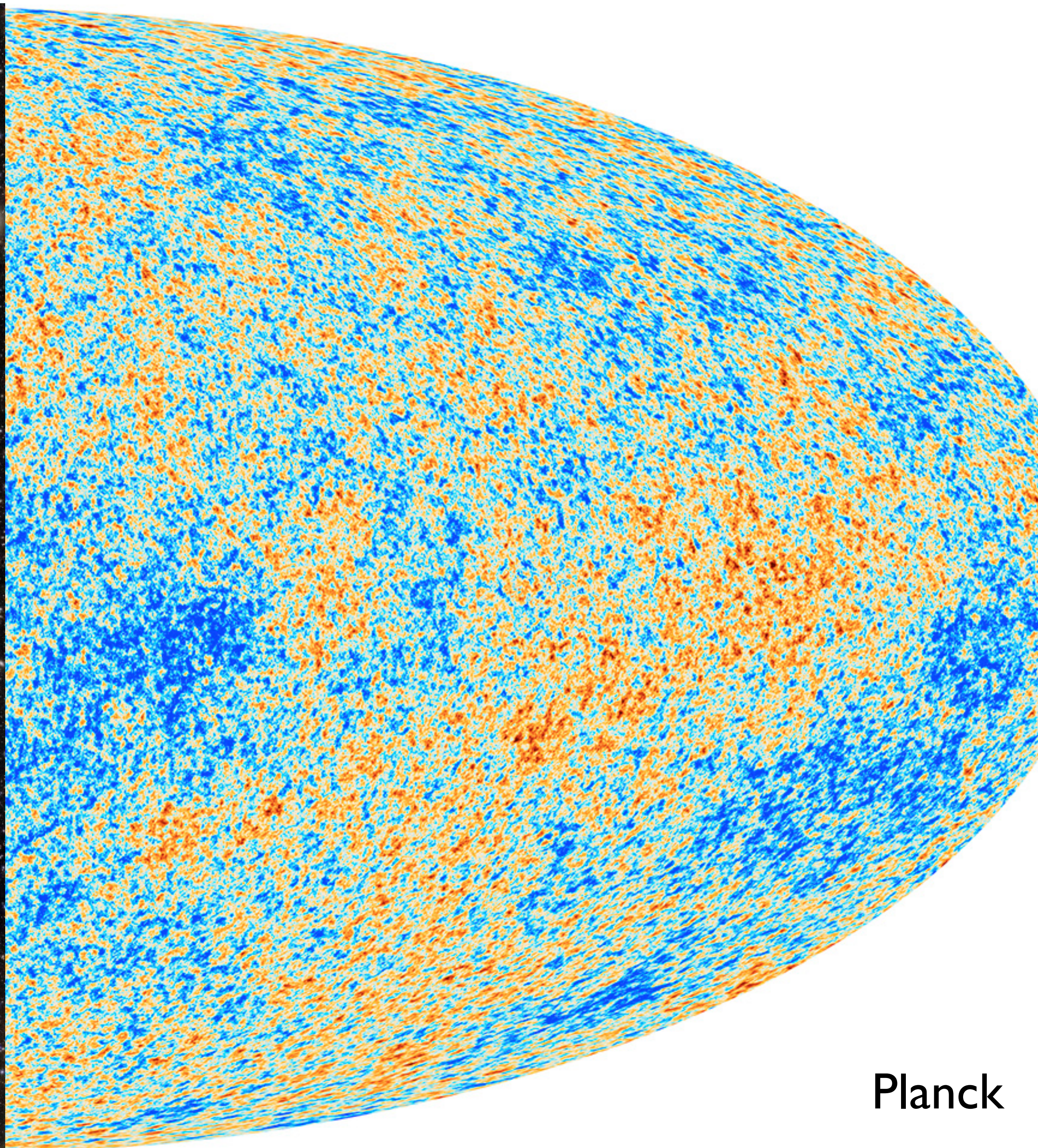



SDSS / M. Blanton

# Challenge: massive, highly nonlinear data set



Perseus cluster seen by Euclid


Planck

# Outline

I. Probes

II. Signals

III. Methods

# I. Probes

- Galaxies:

  Precise sky position
  Precise or poor redshift

  - Number counts (clustering)
  - Shapes (weak gravitational lensing & intrinsic alignments)

- Diffuse gas, line intensity mapping
  => Matteo Viel's talk

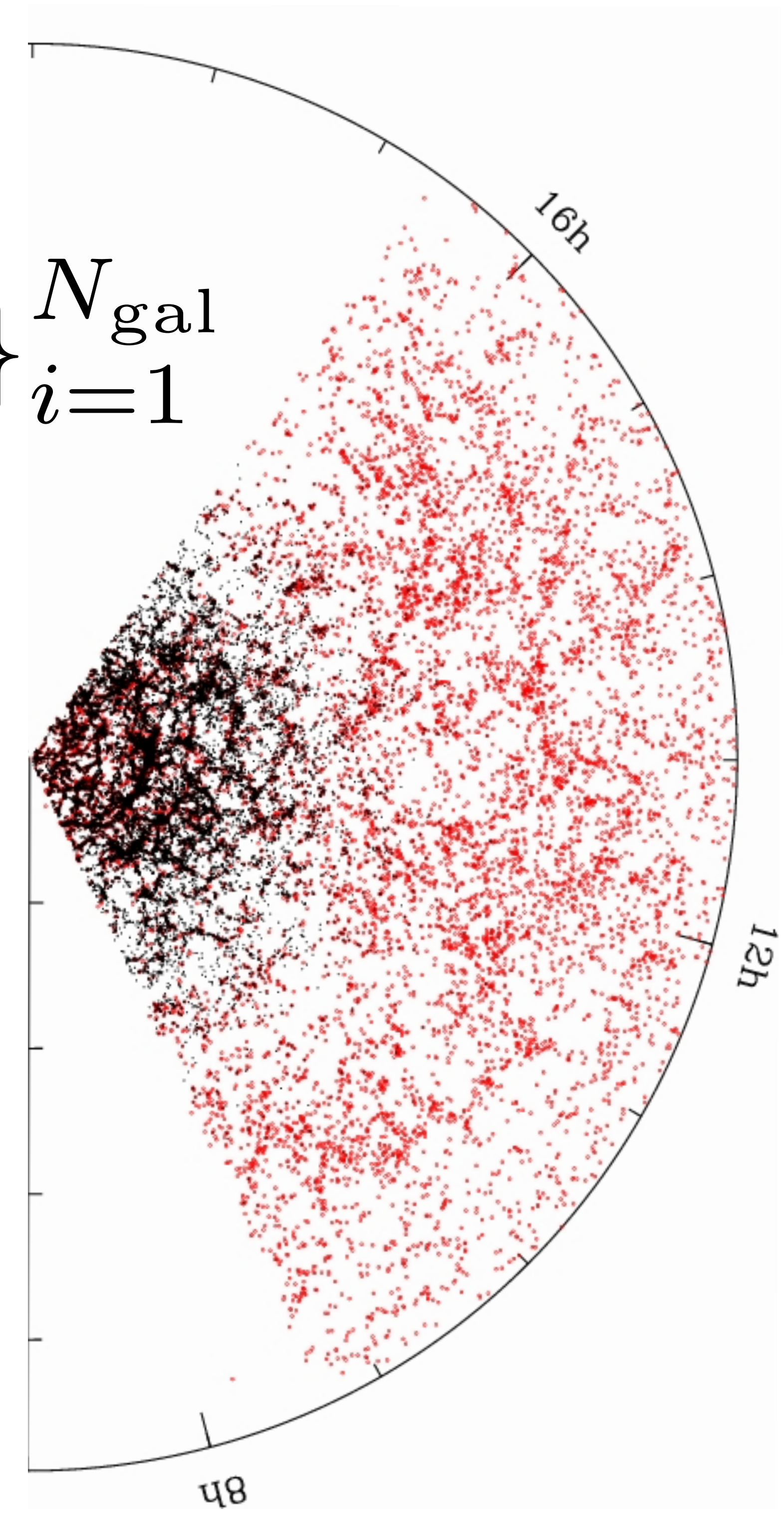
  (Often) poor sky position
  Precise redshift

- GW sources => Michele Maggiore's talk

  (Often) poor sky position
  Precise *distance*

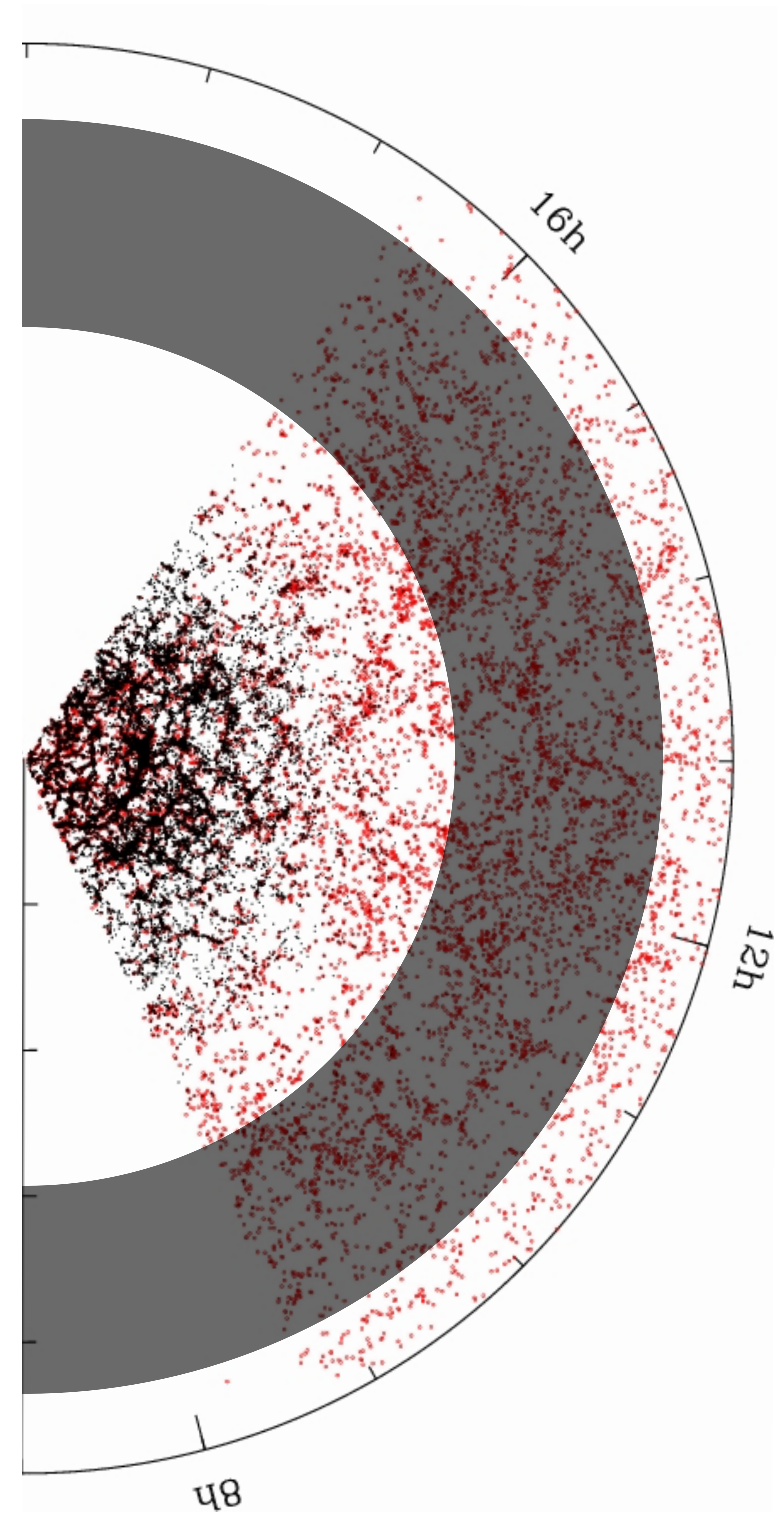
# Galaxy clustering

$$\{\hat{\boldsymbol{n}}_i,\ z_i\}_{i=1}^{N_{\rm gal}}$$



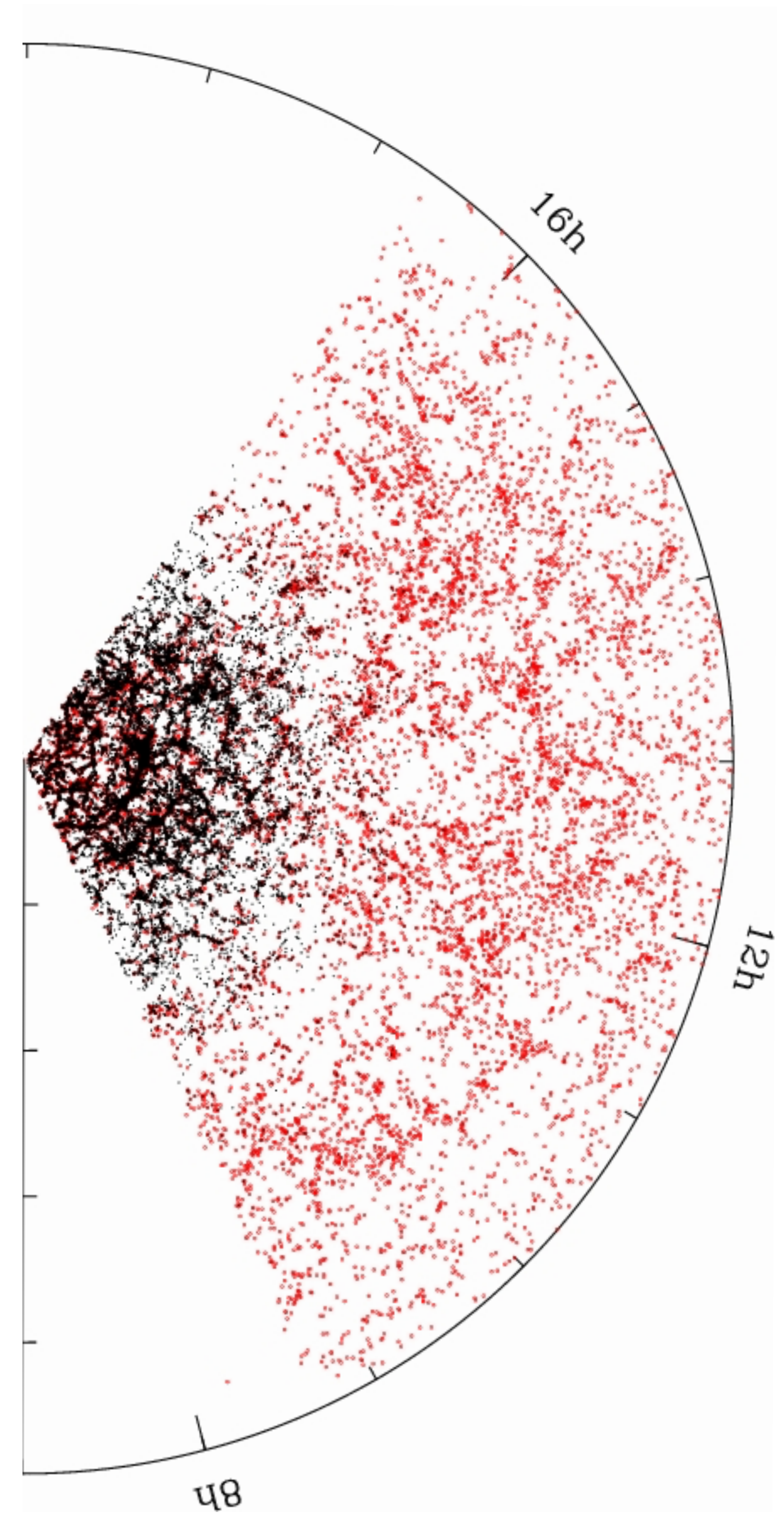
# Galaxy clustering

- Galaxy surveys come in two variants:
- *Photometric* surveys perform imaging only; redshift estimate using "colors" (different filters), accurate at $\Delta z \sim 0.03 - 0.1$
  - Lose many line-of-sight modes
  - Lose higher-order statistics (central limit theorem)



# Galaxy clustering

- Galaxy surveys come in two variants:
- *Photometric* surveys perform imaging only; redshift estimate using "colors" (different filters), accurate at $\Delta z \sim 0.03 - 0.1$
  - Lose many line-of-sight modes
  - Lose higher-order statistics (central limit theorem)
- *Spectroscopic* surveys additionally obtain a spectrum for each galaxy => precise redshift
  - Significantly higher cost per galaxy

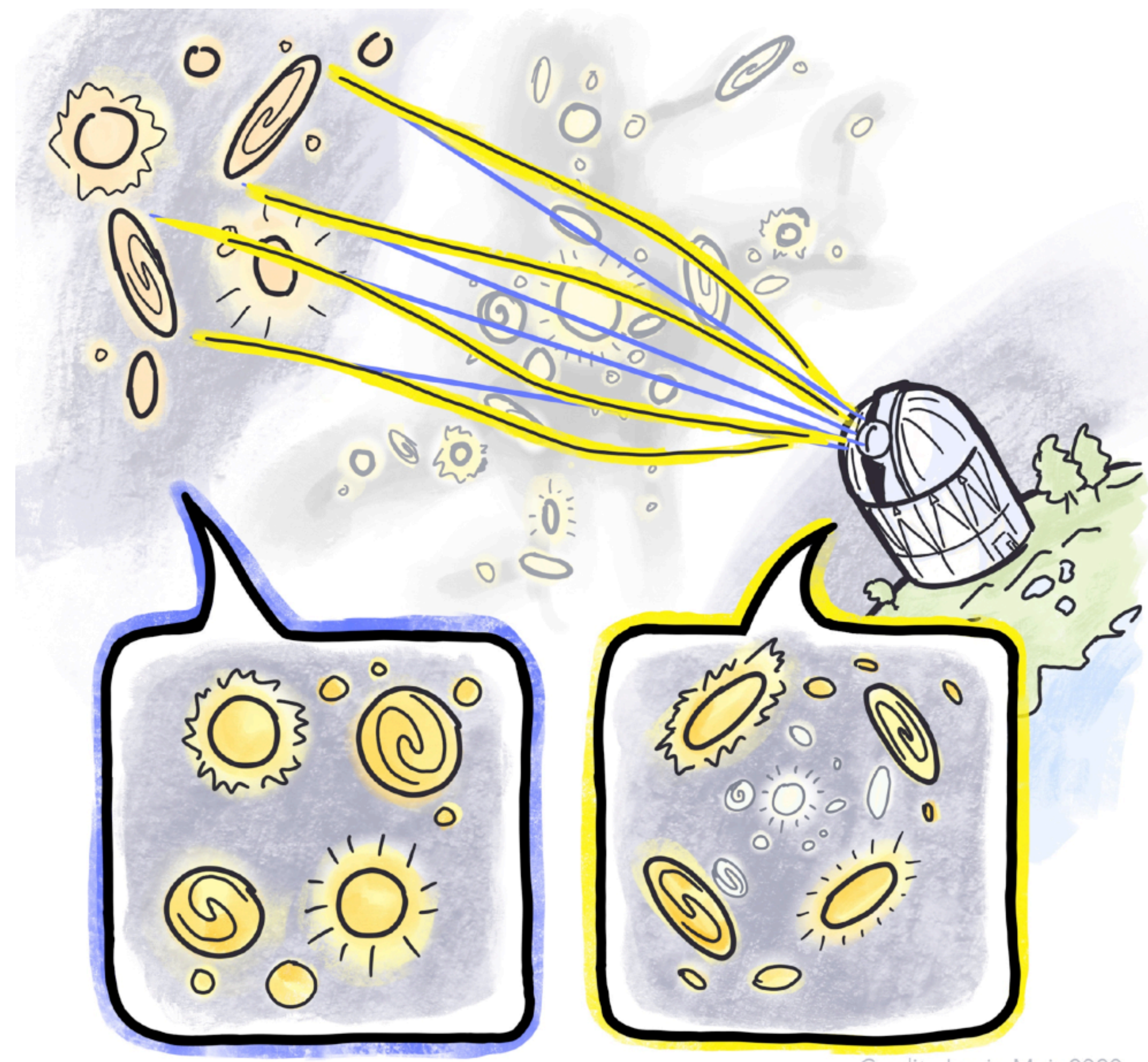


# Galaxy shapes

- Beyond the point source limit: measure second moments of galaxy image on the sky
- Now have a set of symmetric 2-tensors on the sky:

$$\{(T_{kl})_i,\ z_i\}_{i=1}^{N_{\rm gal}}$$

- Tensor observable: E/B decomposition, …
- Again, can have projected (more common) or 3D cases



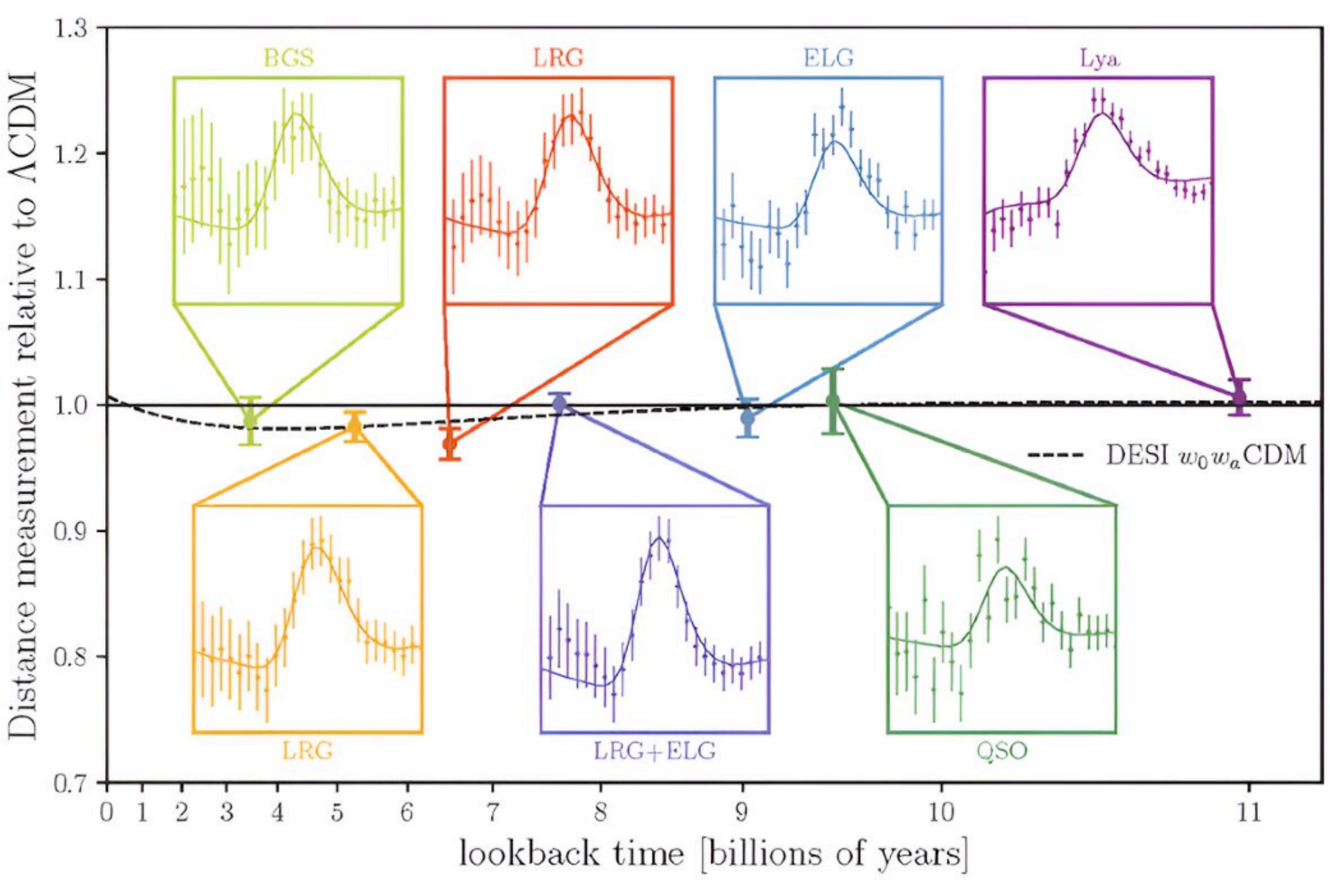

Jessie Muir / DES

# II. Signals

# Galaxy clustering classic probes: BAO

- Standing acoustic wave in pre-recombination baryon-photon fluid, with known period: sound horizon $r_s$
- Measure angular and redshift-space scale in 3D galaxy clustering -> distance $d_A$ and Hubble parameter relative to $r_s$
- BAO reconstruction approach to increase contrast of BAO feature (undo nonlinear damping)

*Baryon acoustic oscillation*




DESI collaboration

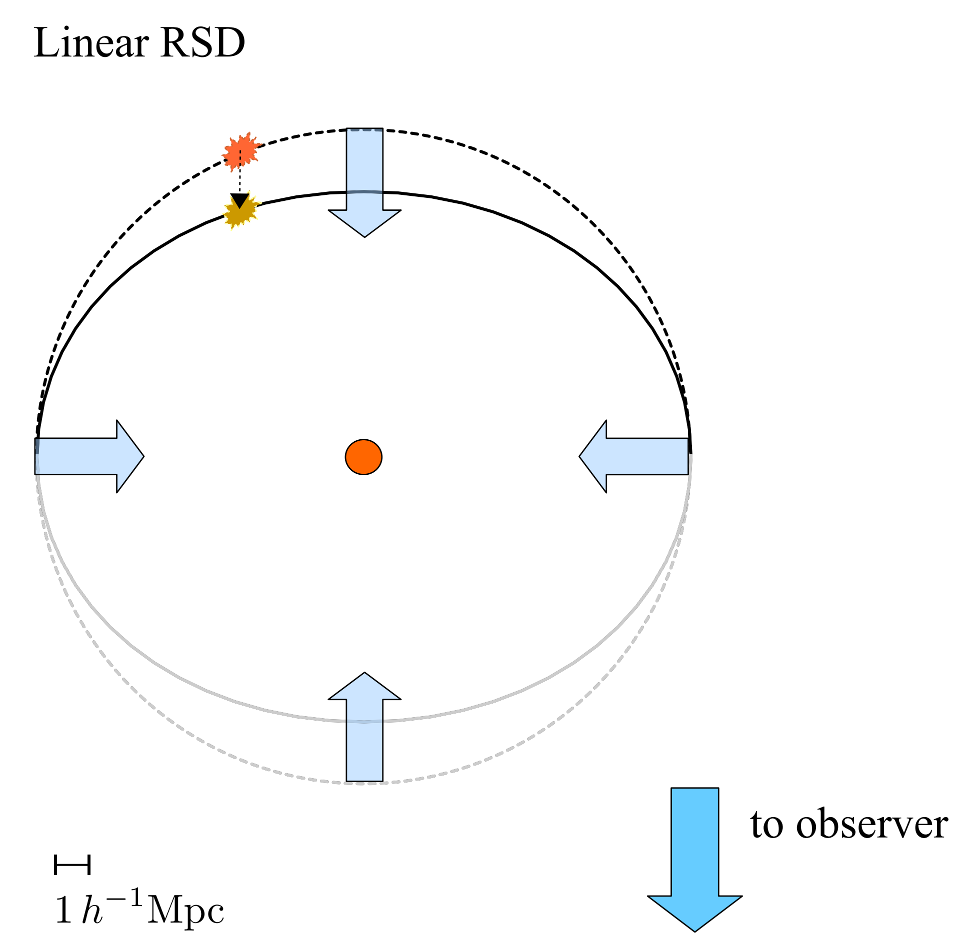
# Galaxy clustering classic probes: RSD

- Observed galaxy positions $x_{obs}$ are given by position on the sky and measured redshift — not "true" positions
- Main effect: Doppler shift to redshift due to peculiar velocity of galaxy:

$$\boldsymbol{x}_{\rm obs} = \boldsymbol{x} + \frac{\boldsymbol{u}_{\rm g} \cdot \hat{\boldsymbol{n}}}{aH} \hat{\boldsymbol{n}}$$

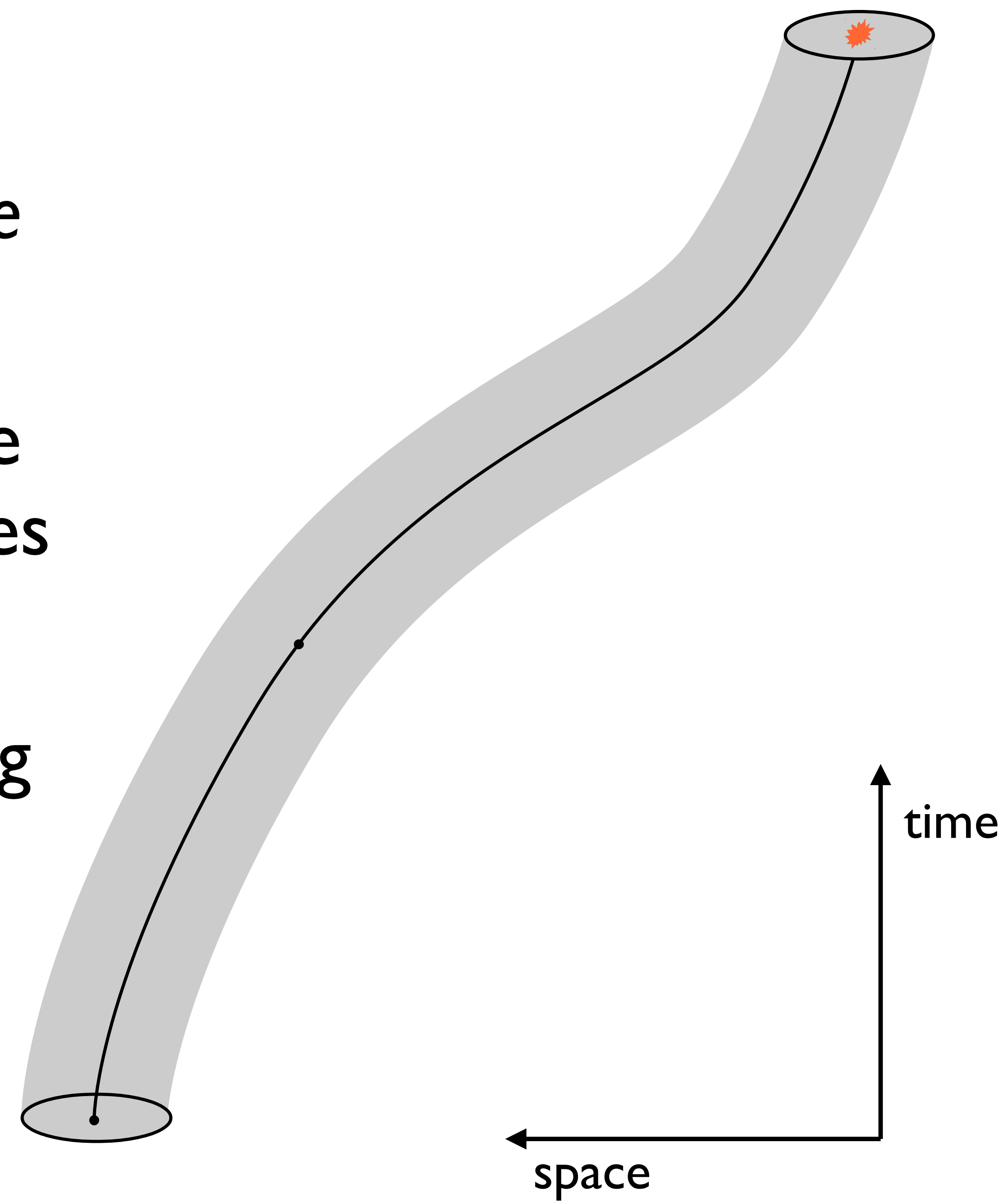
- Equivalence principle (EP): galaxy velocity = matter fluid velocity on large scales
- Probes *time derivative of growth factor*

*Redshift-space distortions*



# Beyond linear theory: making use of EP

- We cannot predict galaxy positions from first principles; capture uncertainties in *effective bias coefficients* (EFT)
- Leading gravitational observable is tidal field $\partial_i\partial_j\Phi$ which includes density $\delta \propto \nabla^2\Psi$
- Along *entire trajectory* of forming galaxy
- Many free parameters, but also have EP-protected terms

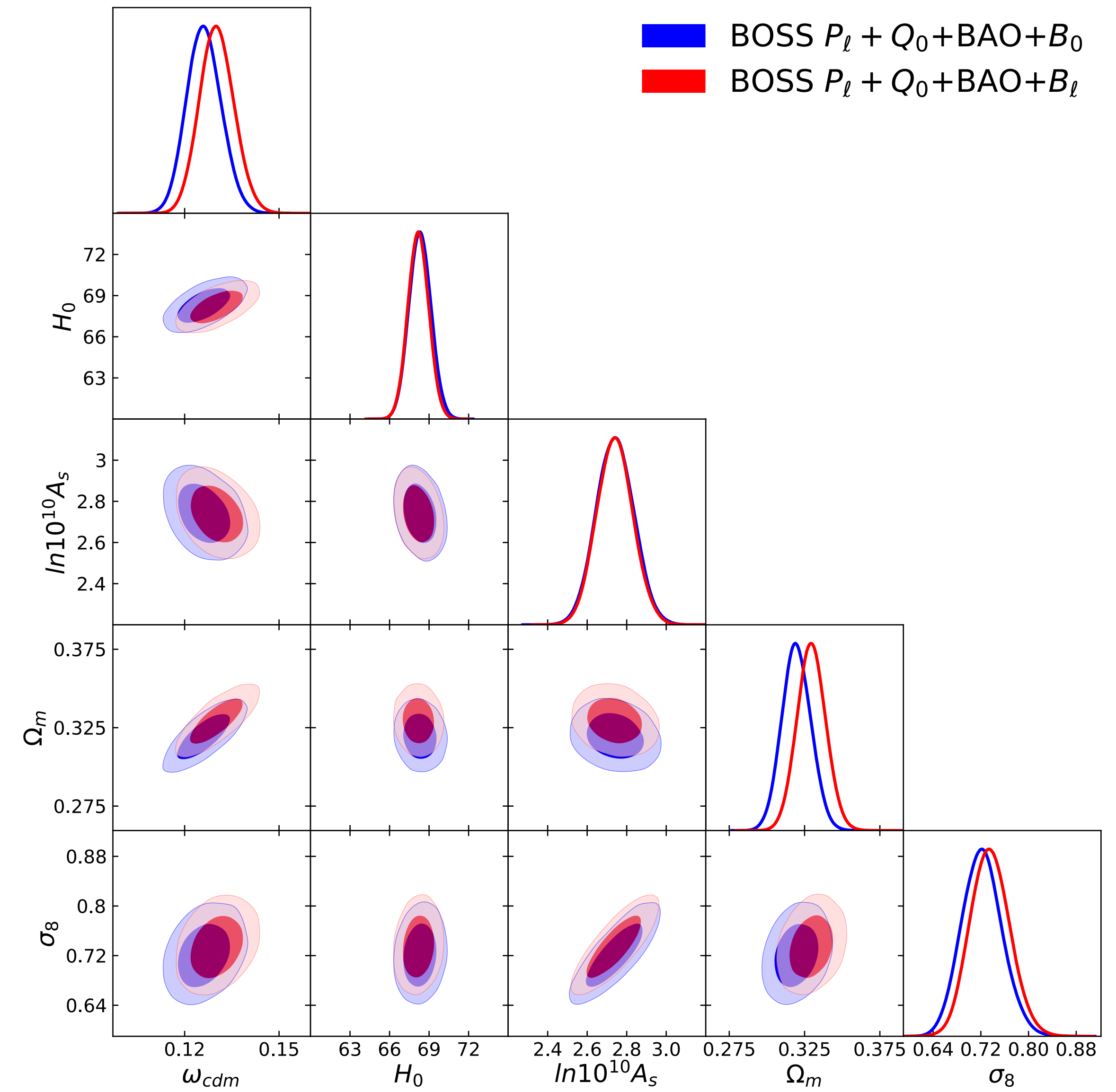



Desjacques, Jeong, FS (2017)

# Current state: power spectrum + bispectrum

- Protected *displacement* terms in galaxy density start at second order
- These probe growth factor (or $\sigma_8$)
- Appear at leading order in galaxy 3-pt function = bispectrum
- Current SOTA 1-loop Pk+Bk (up to 4th order in perturbations)

$$\sigma(H_0)/H_0 \approx 1.2\%; \quad \sigma(\sigma_8)/\sigma_8 \approx 4.5\%$$

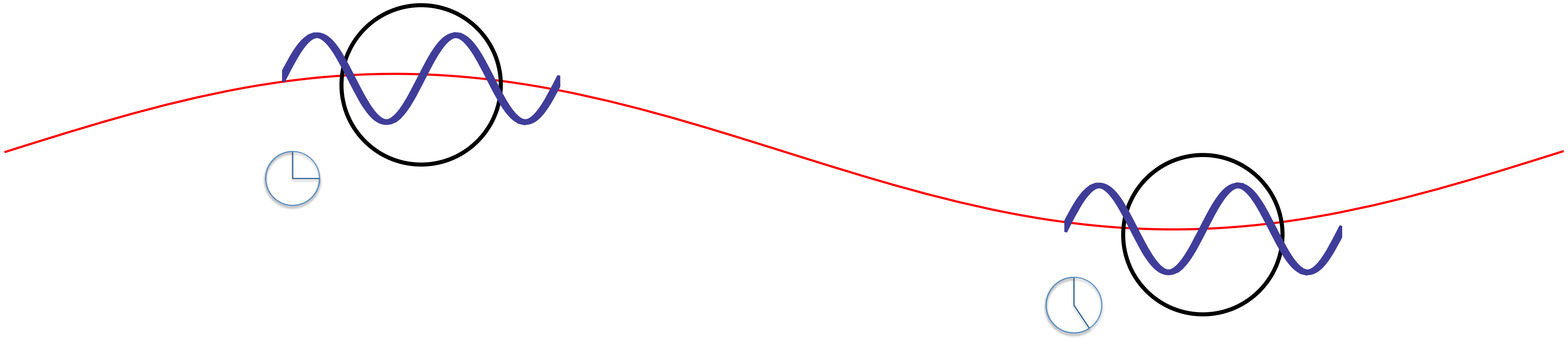



Ivanov, Philcox, et al, 2023
D'Amico et al 2023

# Probing inflation

# Single-field inflation

- Fluctuations generated at some point know nothing about larger-scale perturbations that left the horizon long ago

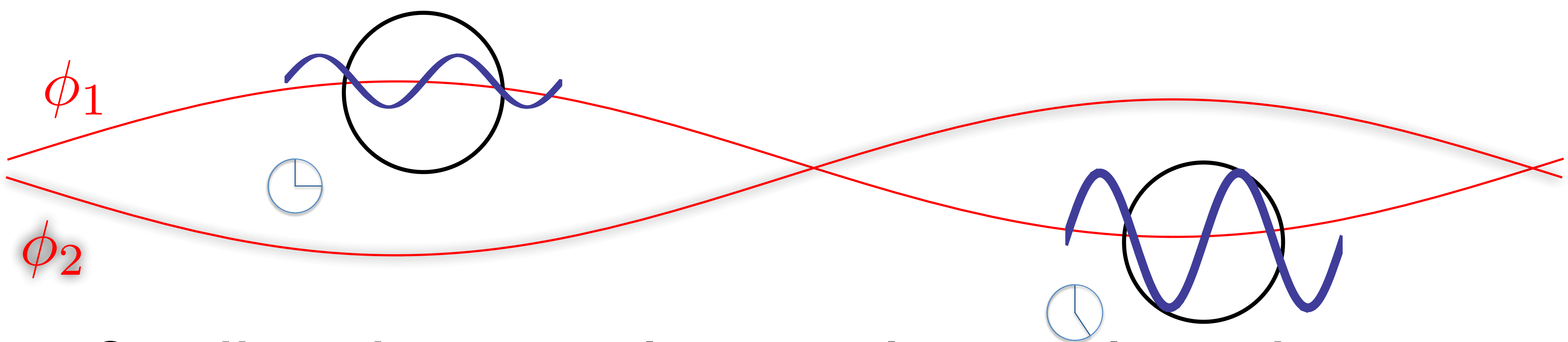


- long-wavelength perturbation just corresponds to different "wall time" - age of the local patch

Pajer, FS, Zaldarriaga (2013)

# Multi-field inflation

- Values of other fields distinguish different patches



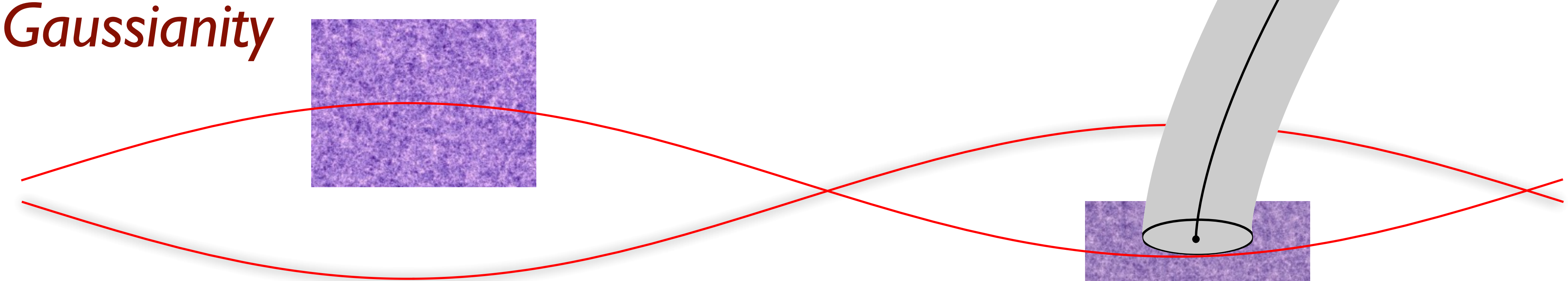

- Small-scale perturbations know about long-wavelength perturbations
- <=> Primordial non-Gaussianity of local type is generated

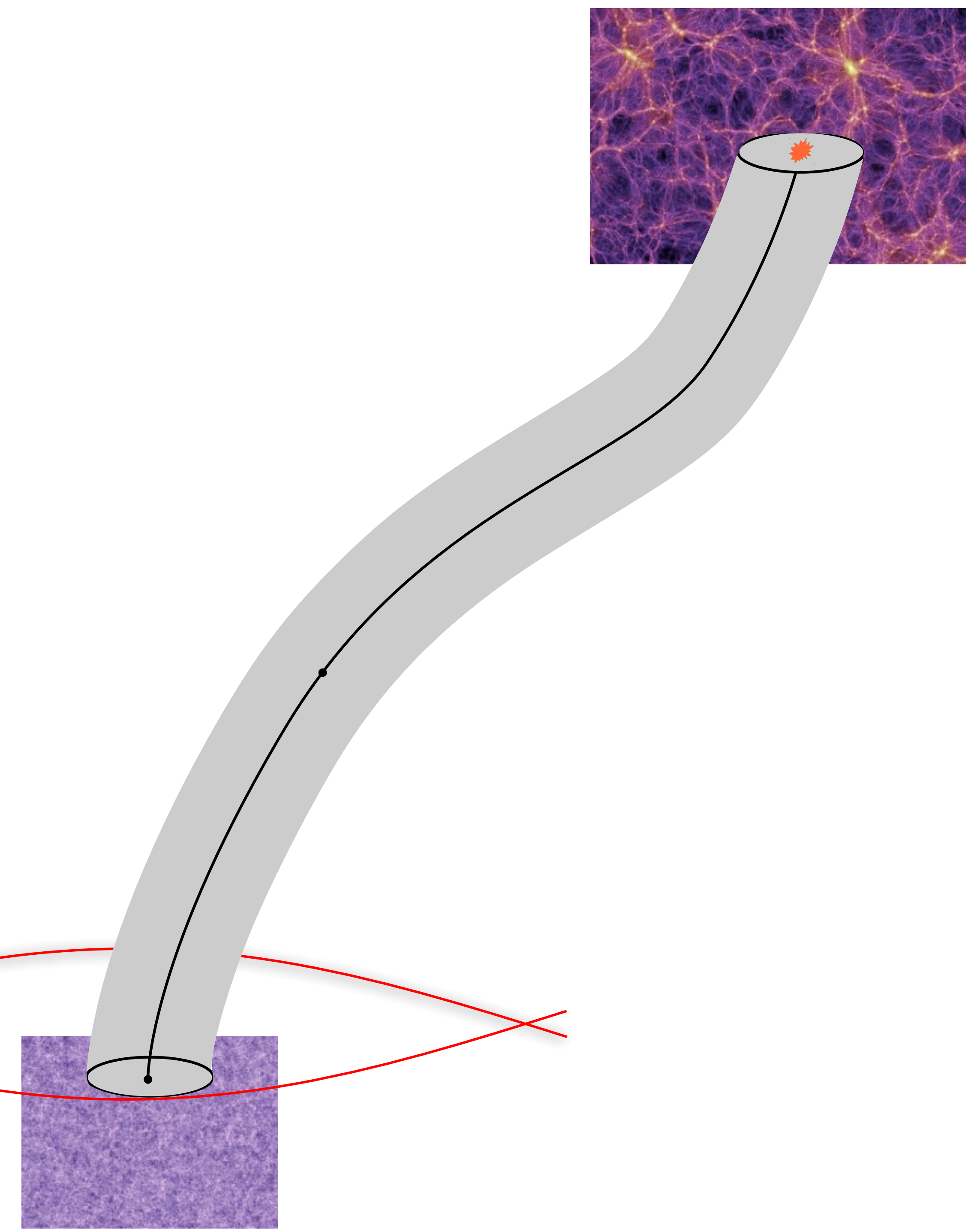
Pajer, FS, Zaldarriaga (2013)

# From initial conditions to galaxies

- In multi-field inflation, amplitude of initial conditions depends on large-scale ***potential***

*local primordial non-Gaussianity*





# Galaxy clustering in multi-field inflation

- Galaxy density now follows the potential, in addition to matter:

$$n_g(\mathbf{x}) = \bar{n}_g \left[ b\, \delta_m(\mathbf{x}) + b_{\rm NG}\, \phi(\mathbf{x}) \right]$$

- Coefficient quantifies the change in galaxy density when changing amplitude of initial fluctuations:

$$b_{\rm NG} = \frac{\partial \ln \bar{n}_g}{\partial \ln \sigma_8}$$

Also, for the standard local bias: $b = \dfrac{\partial \ln \bar{n}_g}{\partial \ln \bar{\rho}_m}$

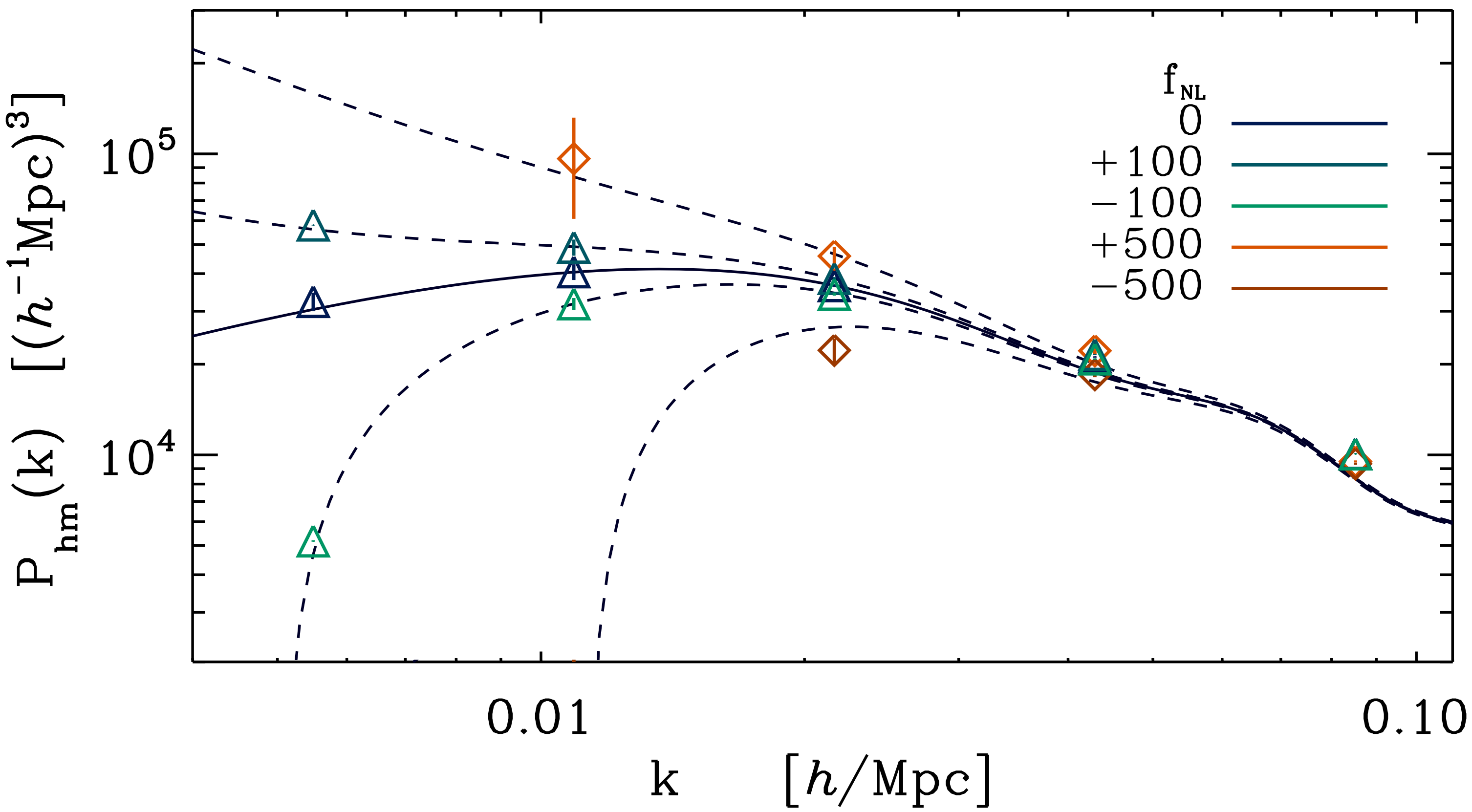
# Galaxy clustering in multi-field inflation

- Smoking-gun signature: clustering increases towards large scales
- Probing highest energy physics with galaxies on the largest scales
- Current constraints:

$$\Delta f_{\rm NL}({\rm CMB}) \sim 3$$

$$\Delta f_{\rm NL}({\rm LSS}) \sim 15$$

$$P_g(k) = \left[ b^2 + 2b\, b_{\rm NG} f_{\rm NL} \frac{A}{k^2} \right] P_m(k)$$



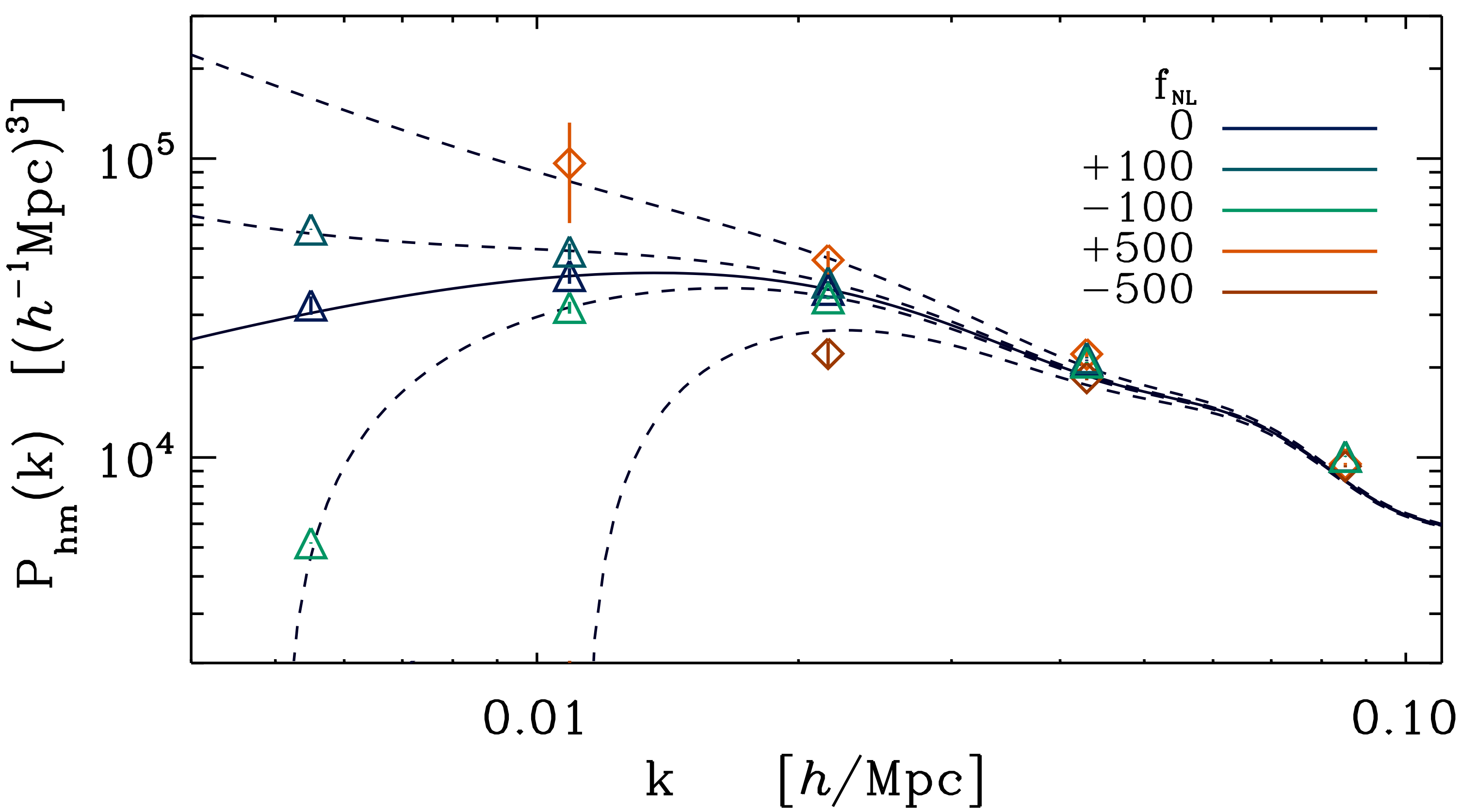

Dalal et al., 2008

FS, Jeong, Desjacques, 2012

# Galaxy clustering in multi-field inflation

- More generally: galaxy clustering probes squeezed-limit of n-point functions of primordial curvature perturbations
- Cf. *cosmological colliders*

$$P_g(k) = \left[ b^2 + 2b\, b_{\rm NG} f_{\rm NL} \frac{A}{k^2} \right] P_m(k)$$

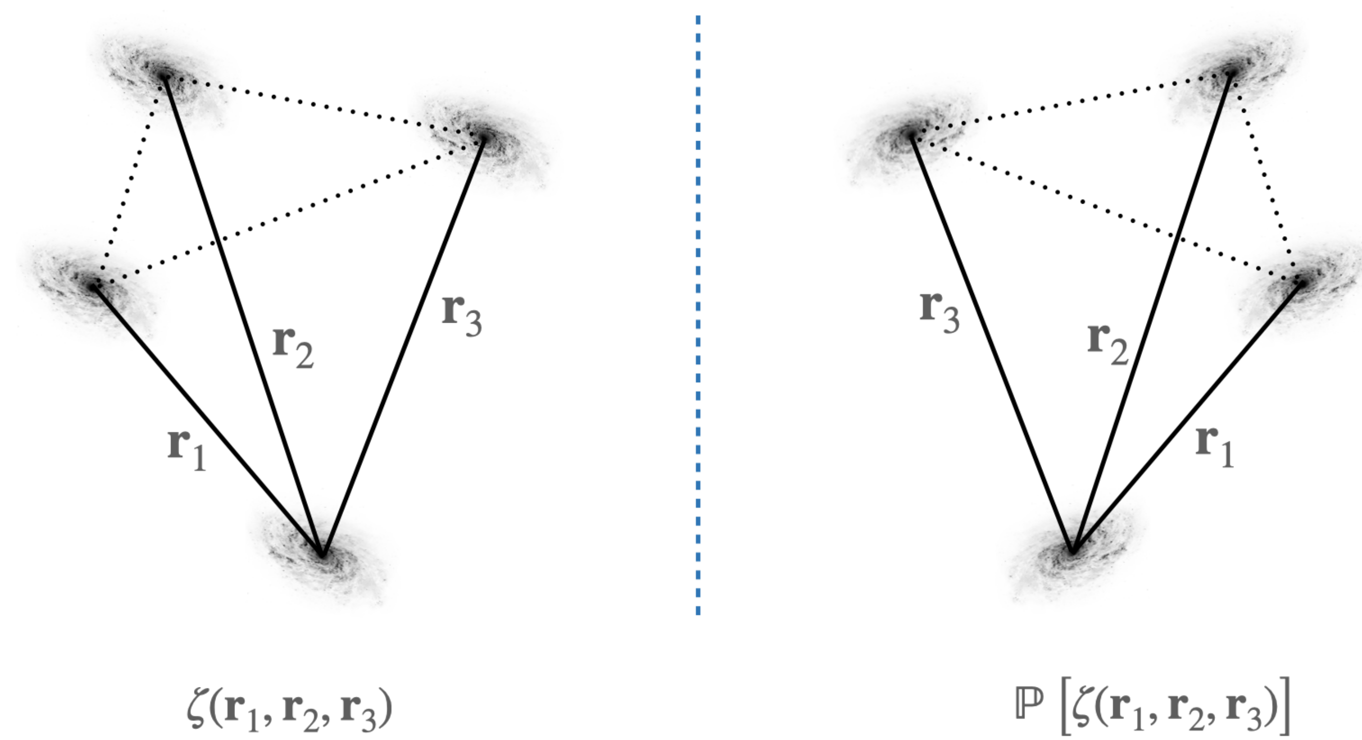



Dalal et al., 2008

FS, Jeong, Desjacques, 2012

# Galaxy clustering and parity violation

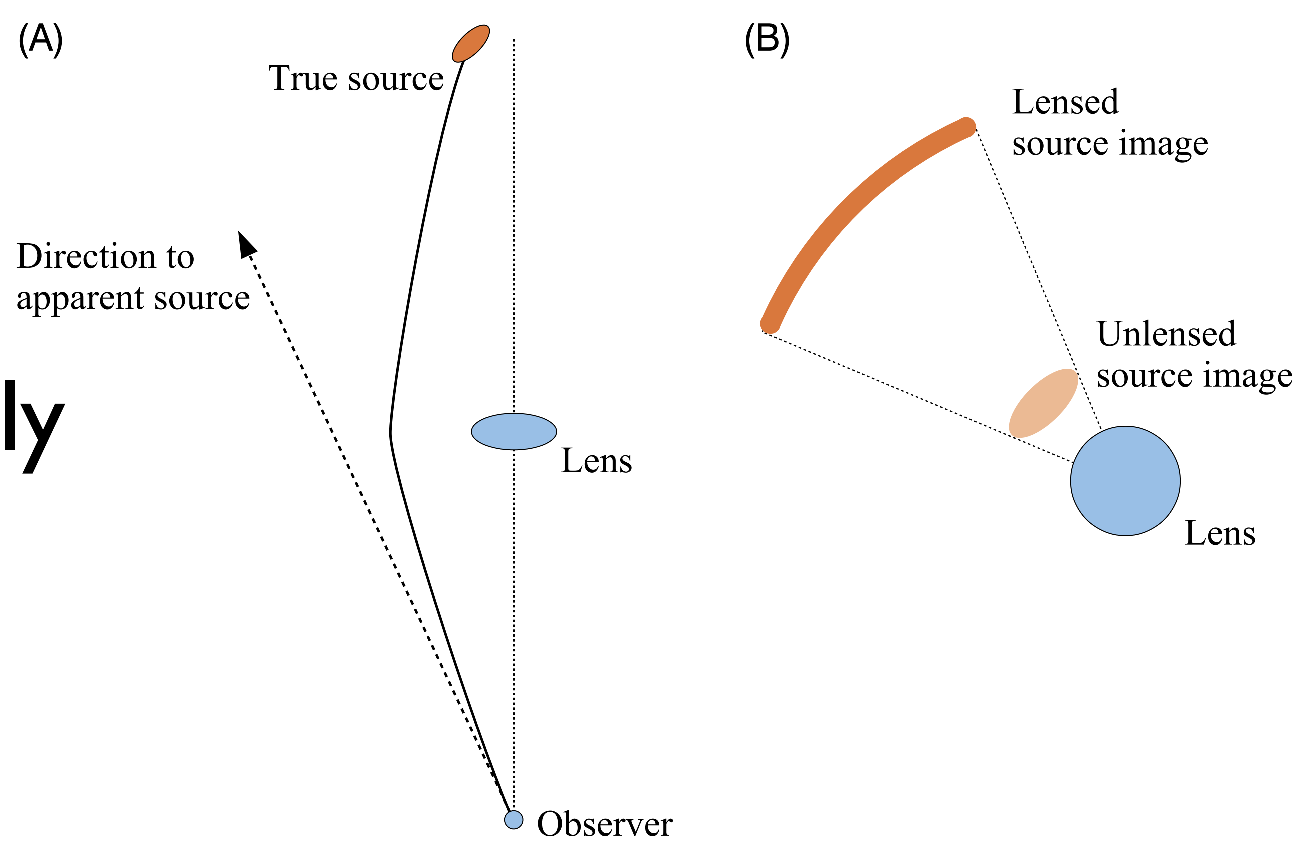
- Was there (large-scale) parity-violating physics during inflation?
- CMB temperature not (very) sensitive to parity violation: 2D
- Galaxy 4-pt function is!




Hou, Slepian, Kahn (2023); Philcox (2023)

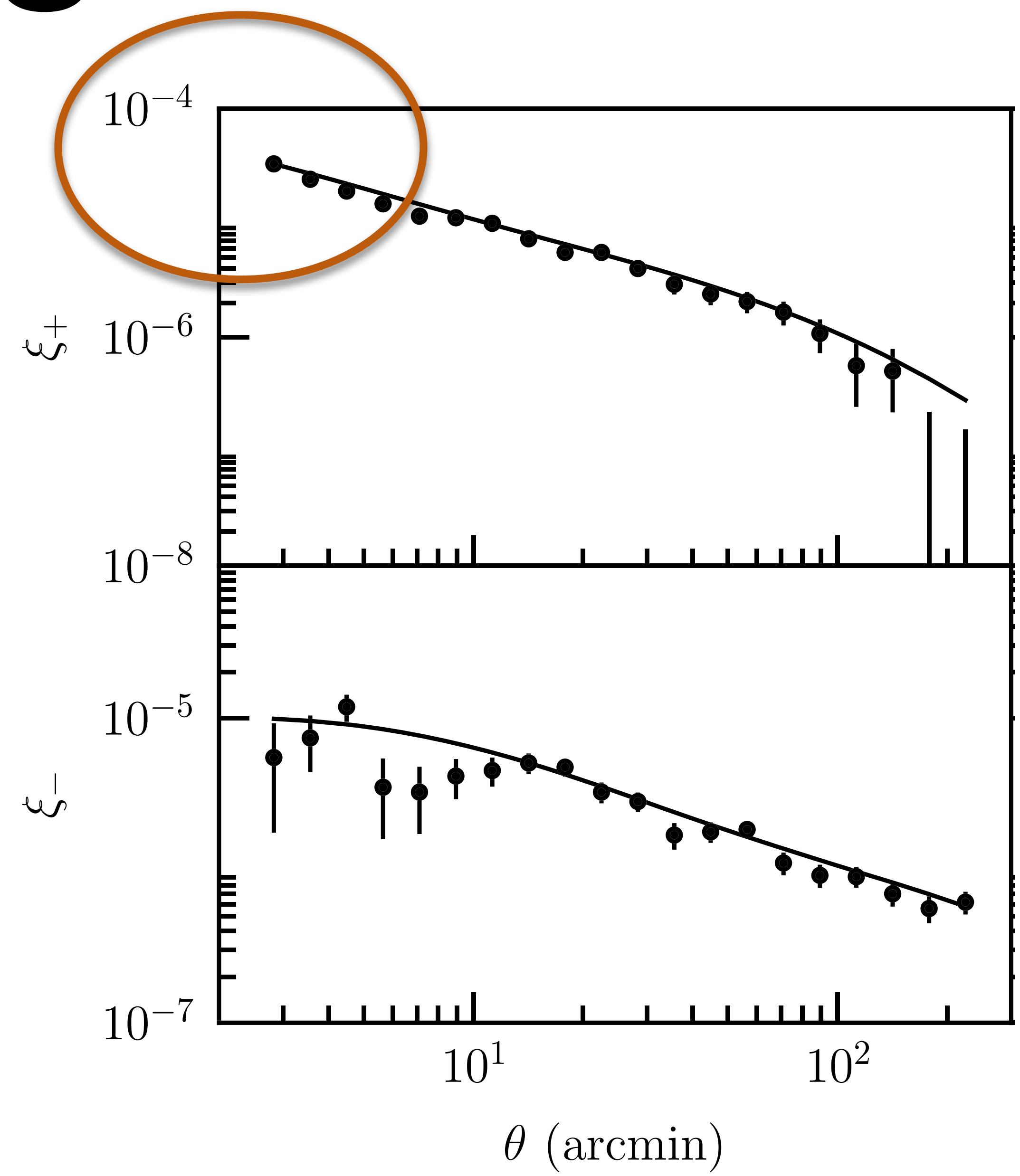
# Galaxy shapes: weak lensing

- Gravitational lensing distorts galaxy shapes
- Signal controlled by gradient of deflection angle: *shear*
- Can detect lensing statistically by using shape correlations
  - Assuming that *intrinsic* galaxy shapes are uncorrelated on large scales!



# Galaxy shapes: weak lensing

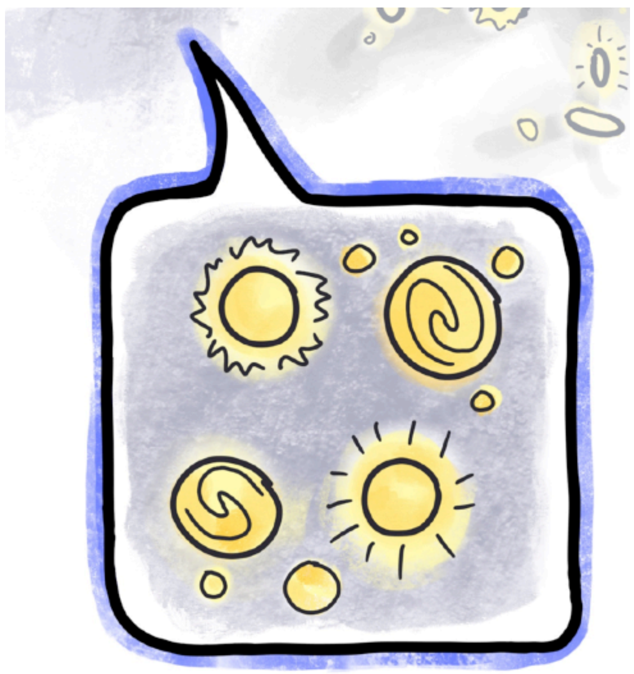
- Weak lensing probes second derivatives of spacetime potentials
- Probes all clustering stress-energy components (Poisson equation)
- Constraint on growth factor ($\sigma_8$)
- Small signal — need to go to very small scales




DES Collaboration (Y1)

# Galaxy shapes: intrinsic alignments

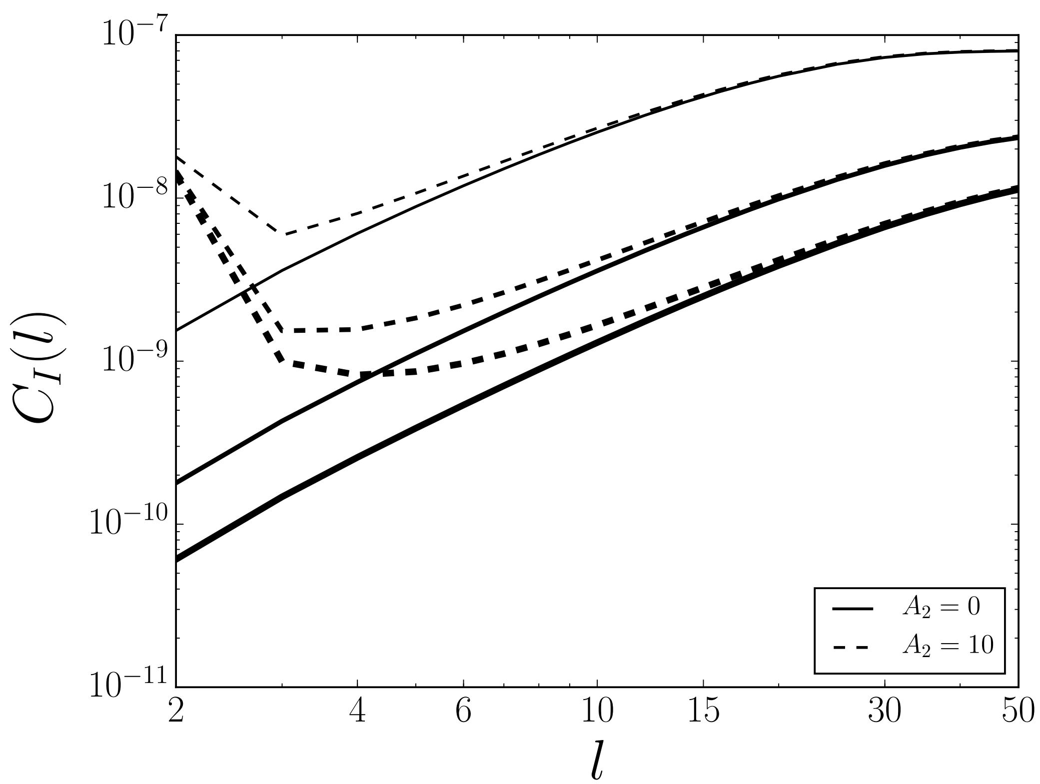
- Essentially a biased tracer, but a 3-tensor instead of 3-scalar
- Can write a similar EFT as for galaxy clustering, using appropriate symmetries
- In principle, offers similar but complementary information to galaxy clustering



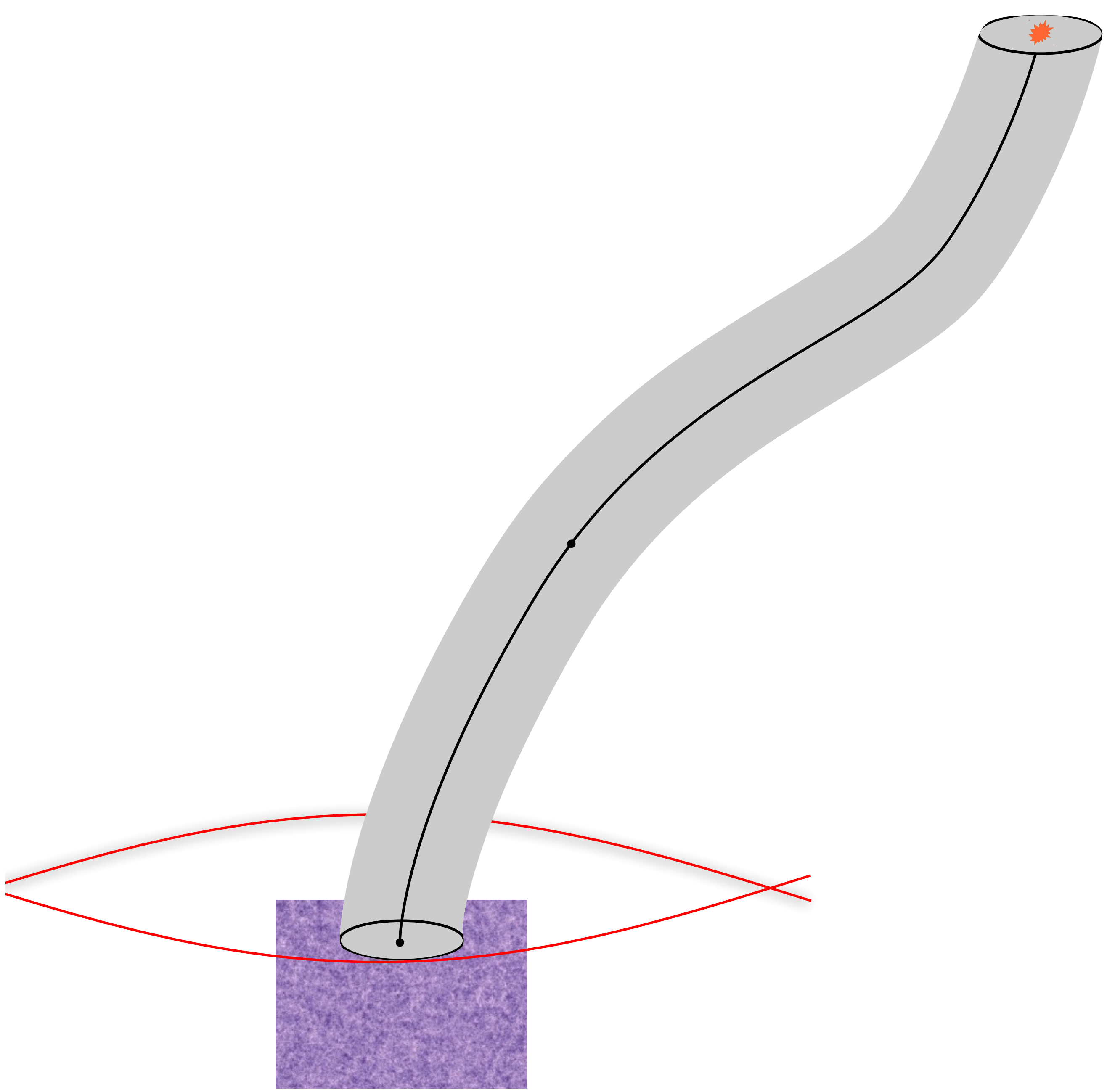
Jessie Muir / DES

# Galaxy shapes: intrinsic alignments

- Example A: probe *anisotropic squeezed limit* of curvature bispectrum => spinning particles during inflation



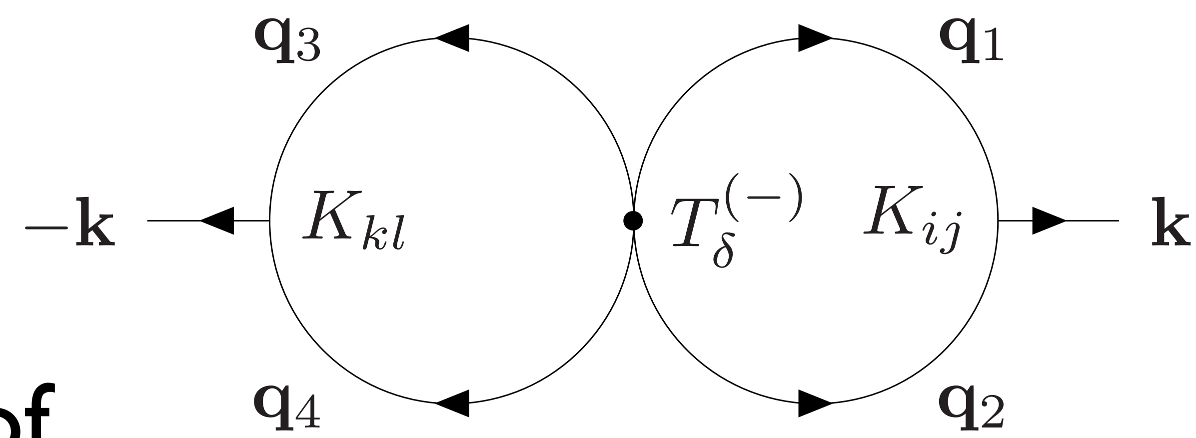
(a) $A_2 = 10.$



Schmidt, Chisari, Dvorkin 2015; Chisari et al 2016

# Galaxy shapes: intrinsic alignments

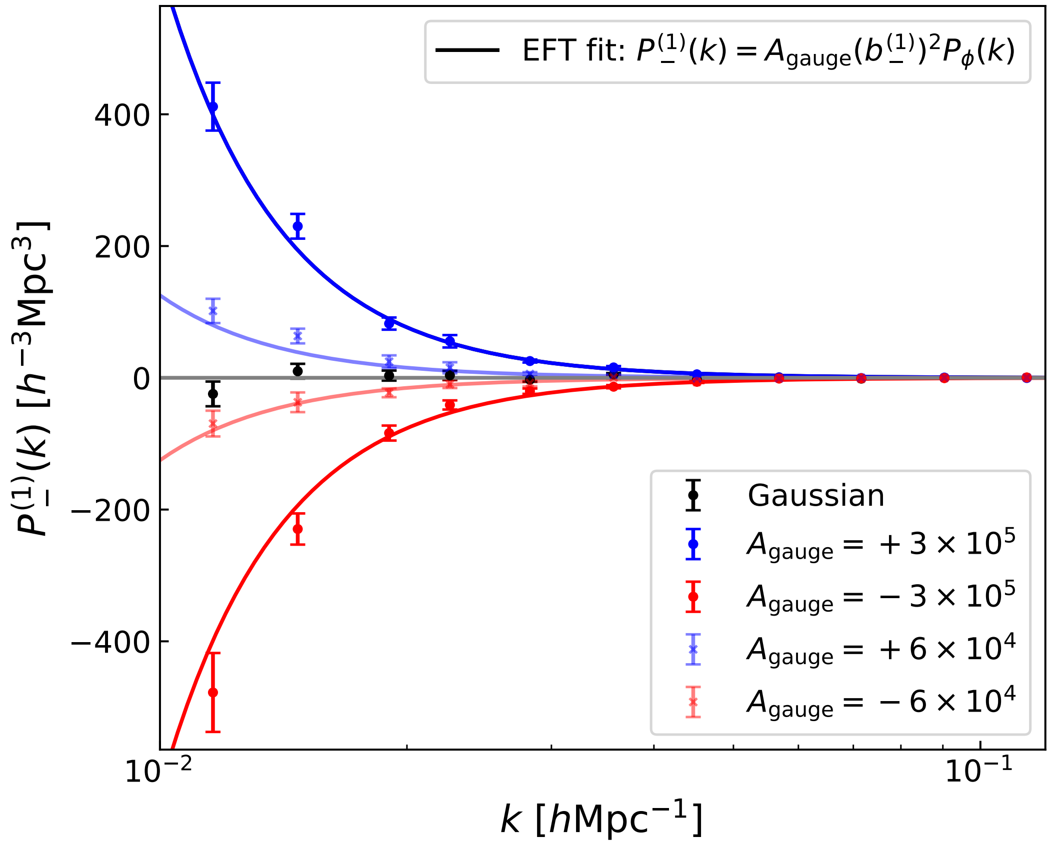
- Example B: probe parity violation in primordial perturbations
- Enhanced large-scale correlation induced in case of enhanced collapsed limit of primordial trispectrum




Kurita et al, in prep.

# Galaxy shapes: intrinsic alignments

- Example B: probe parity violation in primordial perturbations
- Enhanced large-scale correlation induced in case of enhanced collapsed limit of primordial trispectrum
- Parity-odd statistic *EB*: smoking gun of parity violation




Kurita et al, in prep.

# Galaxies as grav. wave detectors

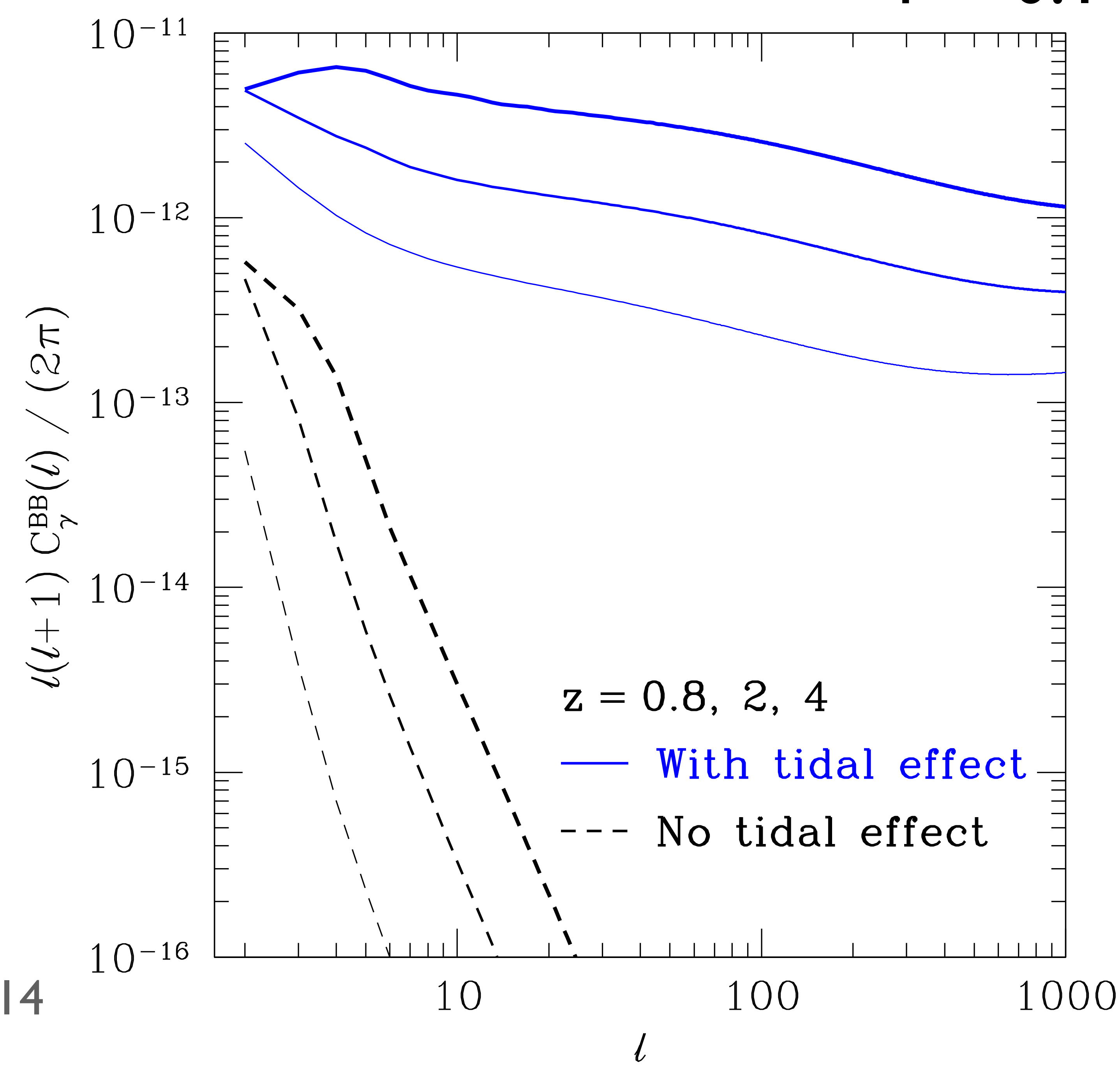
- Intrinsic alignment ~ galaxies aligning with large-scale tidal fields
- GW induce a local tidal field, in addition to acting as gravitational lenses
- Large galaxy surveys can be used to search for inflationary GW, via odd-parity (*B-mode*) component of shape correlations

FS, Pajer, Zaldarriaga, 2013
FS, Jeong, 2012b

# Galaxies as grav. wave detectors

- Tidal effect dominates over lensing
- Very small signal - difficult to measure even for Euclid / LSST
- Still, one of the few possible ways to independently confirm CMB searches

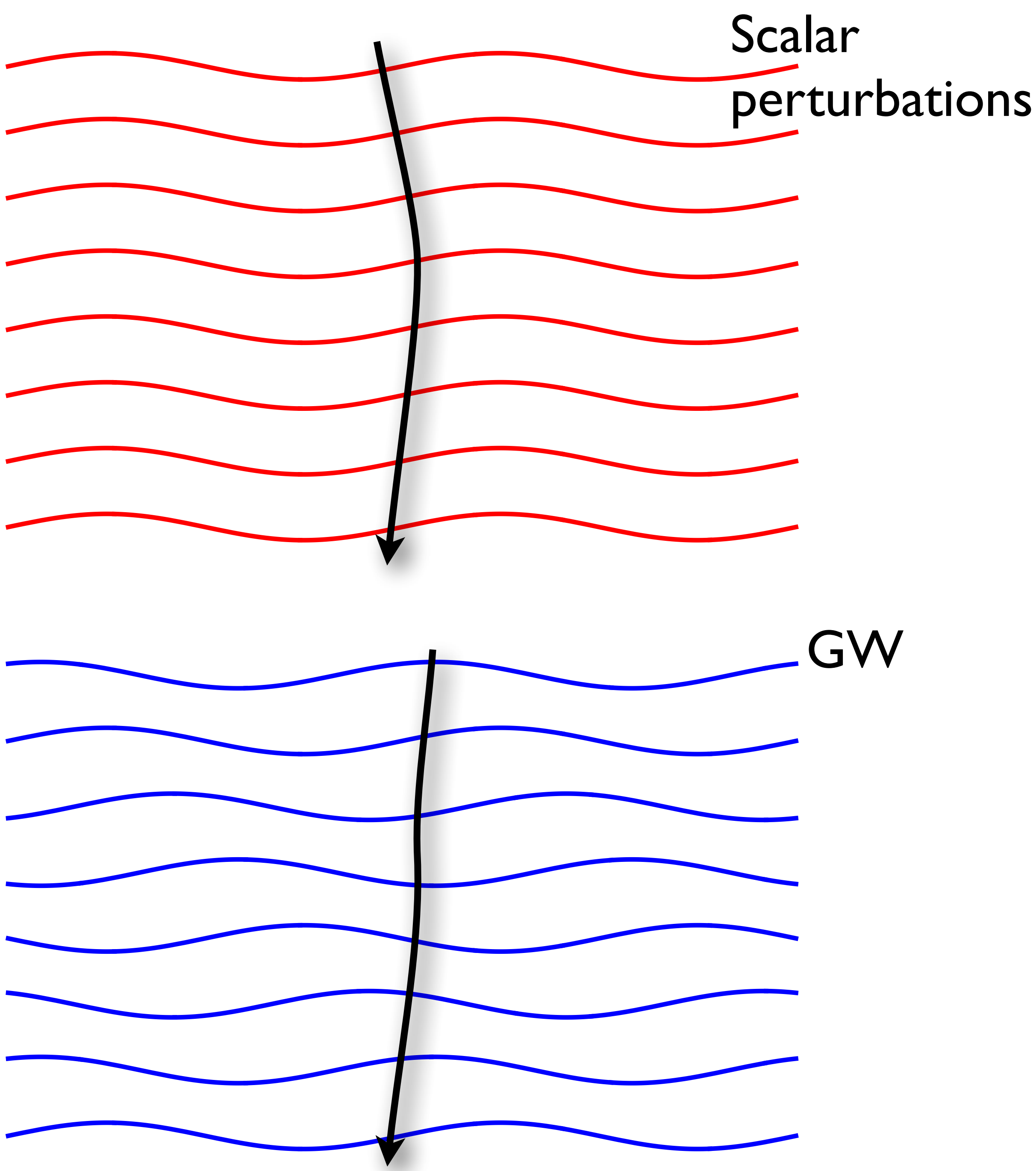
r = 0.1




Chisari, Dvorkin, Schmidt '14

# Why is intrinsic alignment so large for GW ?

- Actually, the correct question is: *why is the GW lensing contribution so small ?*
- Cancelation of lensing effect along the line of sight because GW propagate



# Galaxy shapes: intrinsic alignments

Tensor observable => non-degenerate signals!

Especially when combined with galaxy clustering

| Shape statistic: | Spinning particles | Gravitational waves | Parity violation |
|---|---|---|---|
| *EE* | ✅ | ✅ | ❌ |
| *EB* | ❌ | ❌ | ✅ |
| *BB* | ❌ | ✅ | ❌ |

# III. Methods

(I realize I am running out of time at this point)

# Beyond classical n-point functions

- Much excitement in LSS about exploring information beyond 2- and 3-pt statistics, e.g.
  - *Machine-learned compressions*, coupled with simulation-based inference or emulators
  - *Field-level inference:* strictly optimal Bayesian inference, explicitly inferring initial conditions of observed universe

# Field-level inference

$$P(\theta) \propto \int \mathcal{D}\boldsymbol{\delta}_{\rm in}\, P\left(\boldsymbol{\delta}_g \middle| \boldsymbol{\delta}_{\rm fwd}[\boldsymbol{\delta}_{\rm in}, \theta]\right) P_{\rm prior}\left(\boldsymbol{\delta}_{\rm in}, \theta\right)$$

- Scheme:
  - Discretize field on grid/lattice
  - Draw initial conditions from prior
  - Forward-evolve using gravity
  - Evaluate likelihood on data and repeat
- Results in samples from the *joint posterior of initial conditions and cosmological parameters*

Pioneered by Jasche, Kitaura, Ensslin; Mo et al

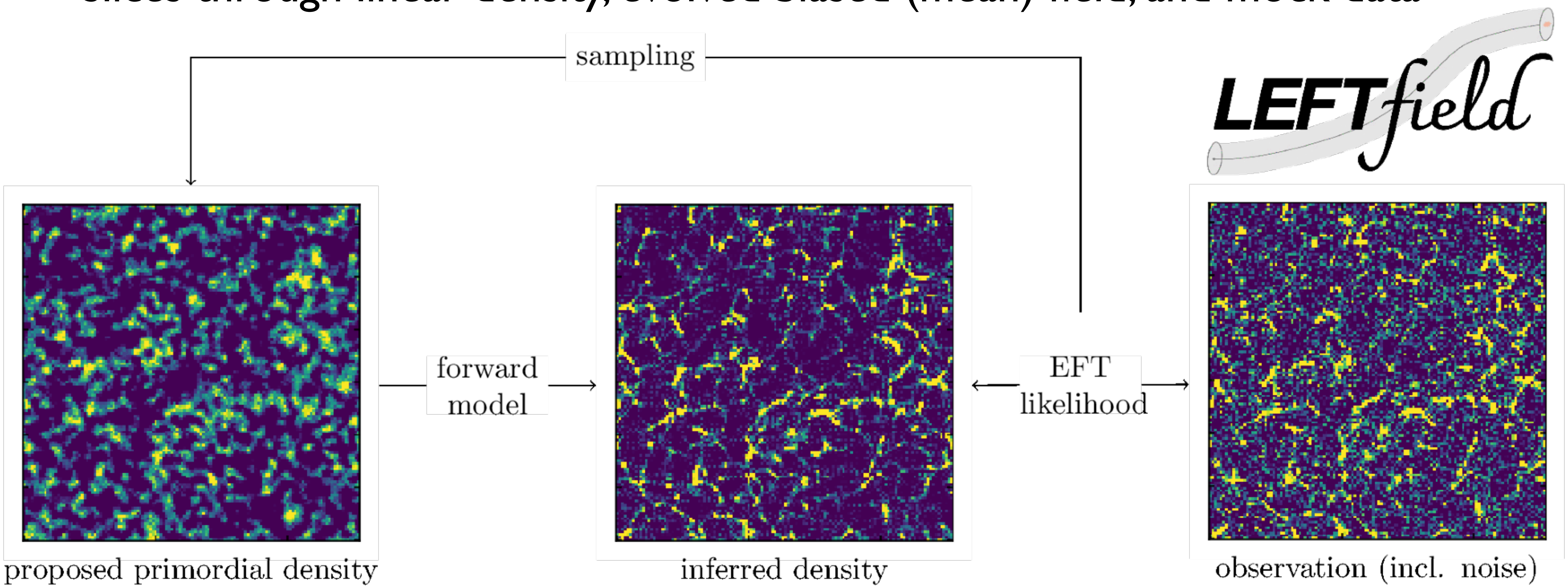
# Field-level inference

$$P(\theta) \propto \int \mathcal{D}\boldsymbol{\delta}_{\rm in}\, P\left(\boldsymbol{\delta}_g \middle| \boldsymbol{\delta}_{\rm fwd}[\boldsymbol{\delta}_{\rm in}, \theta]\right) P_{\rm prior}\left(\boldsymbol{\delta}_{\rm in}, \theta\right)$$

- Scheme:
  - Discretize field on grid/lattice (Nyquist frequency = cutoff Λ)
  - Draw initial conditions from prior
  - Forward-evolve using gravity
  - Evaluate likelihood on data and repeat
- Challenge: even with fairly coarse resolution, have to sample million(s) of parameters
  - Key: Hamiltonian Monte Carlo

# Visualization: results from field-level inference on mock data

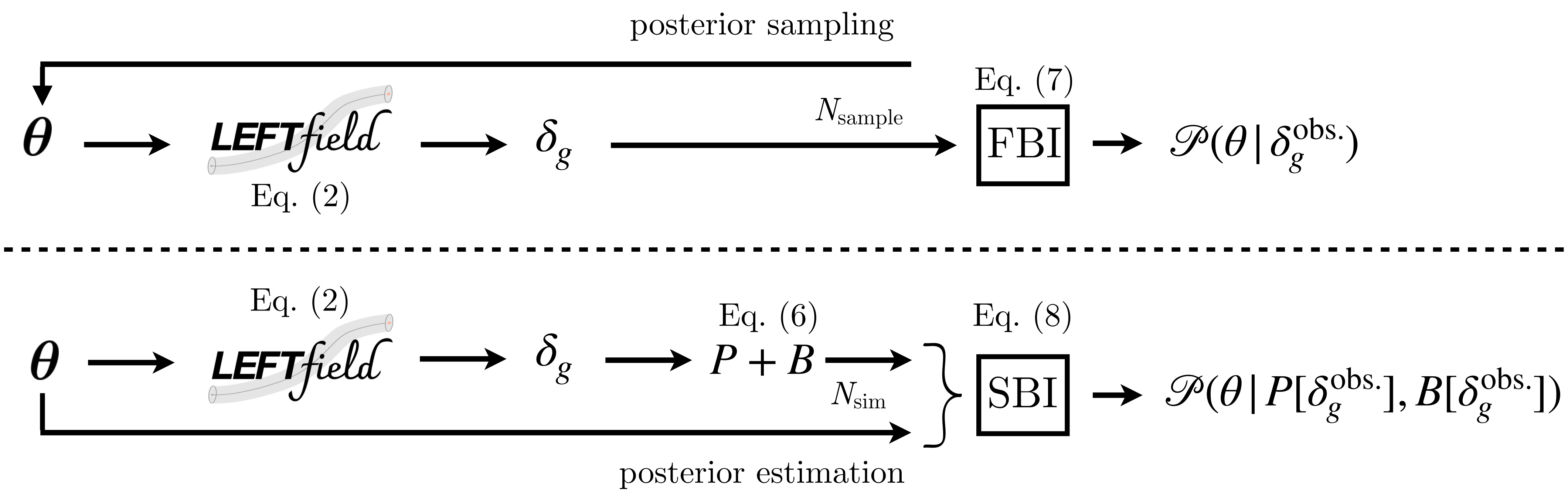
- Slices through linear density, evolved biased (mean) field, and mock data




Julia Stadler

https://www.mpa-garching.mpg.de/1095991/hl202405

# Field-level inference: Inferring $\sigma_8$ from rest-frame tracers



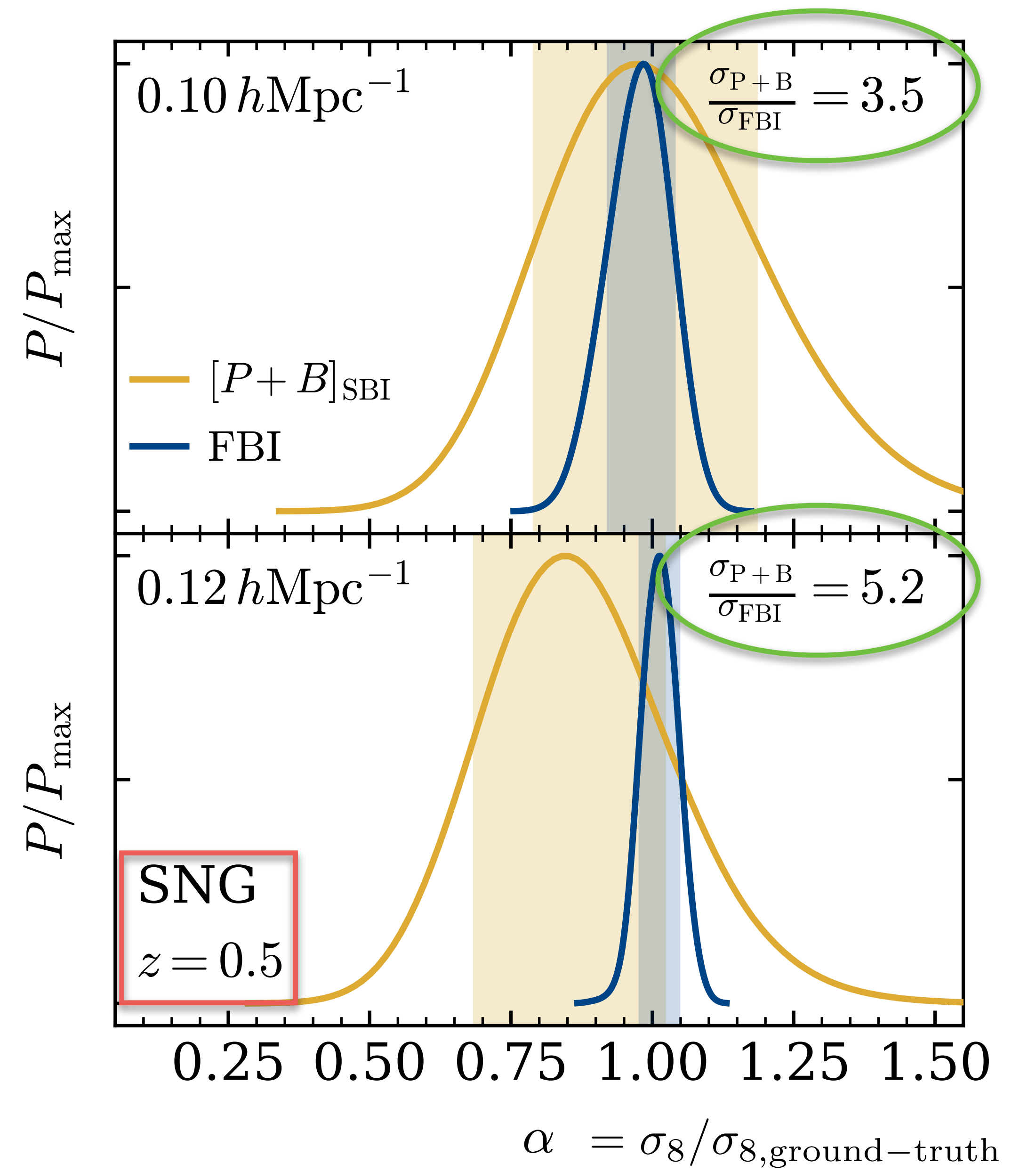

- Idea: compare field-level result with power spectrum + bispectrum using the same forward model and modes of the data

  - Via simulation-based inference (SBI) using the same forward model as in the field-level analysis

Nguyen, FS, Tucci, Reinecke, Kostić PRL 2024, arXiv:2403.03220

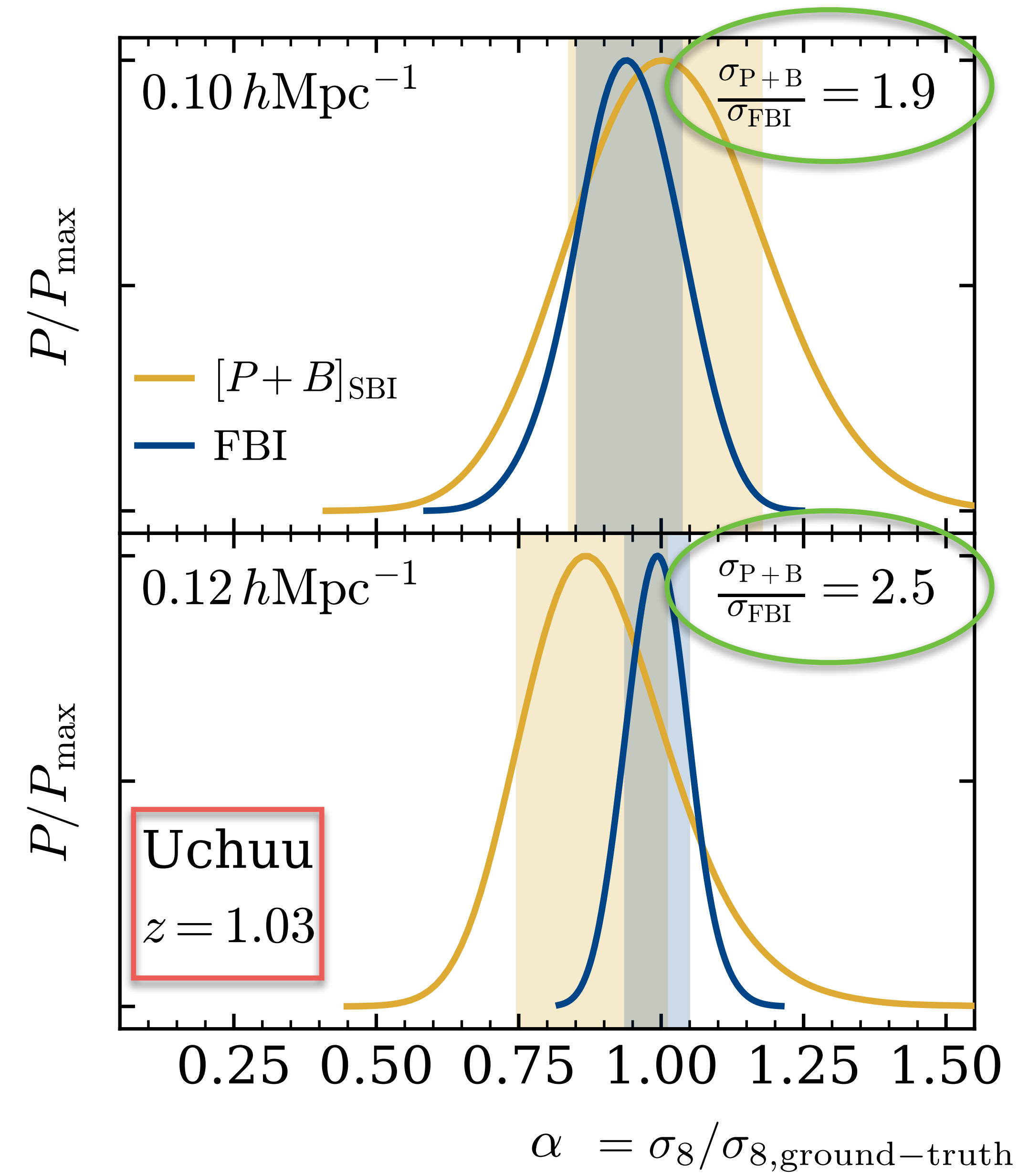
# Field-level inference: Inferring σ8 from rest-frame tracers

- First results on field-level σ8 inference from dark matter halos in real space
    - Marginalizing over bias and stochastic terms
- Field-level inference vs power spectrum + bispectrum using the same forward model and modes of the data



# Field-level inference: Inferring $\sigma_8$ from rest-frame tracers

- First results on field-level $\sigma_8$ inference from dark matter halos in real space
  - Marginalizing over bias and stochastic terms
- Field-level inference vs power spectrum + bispectrum using the same forward model and modes of the data



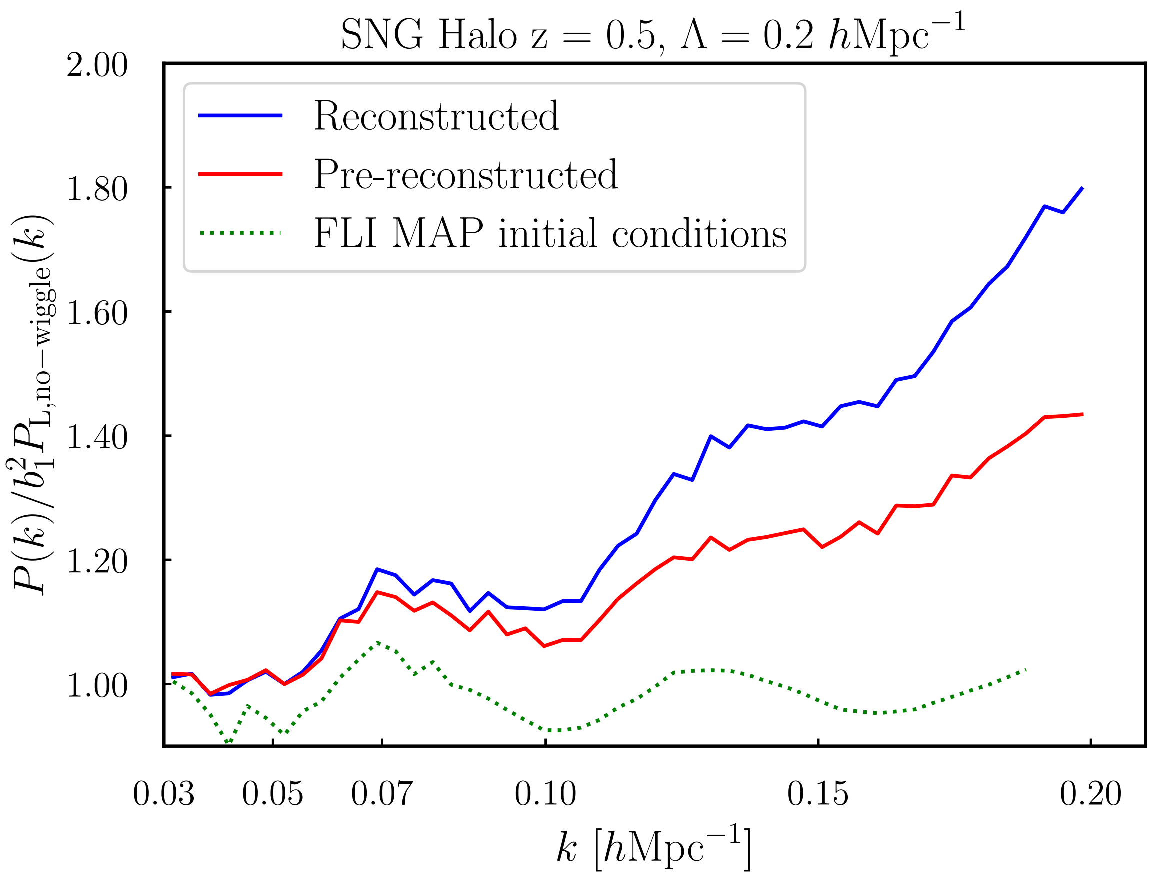

Nguyen, FS, Tucci, Reinecke, Kostić PRL 2024, arXiv:2403.03220

# Field-level inference of BAO scale

- Reconstruction idea: estimate large-scale displacements from galaxy density field, then move galaxies *back* to inferred initial positions

- Improves error bar on BAO scale by up to 50%

- Can we also do this in a *forward* approach by performing joint field-level inference of initial density field and BAO scale?




Babić, FS, Tucci (2025), arXiv:2505.13588

# Field-level inference of BAO scale

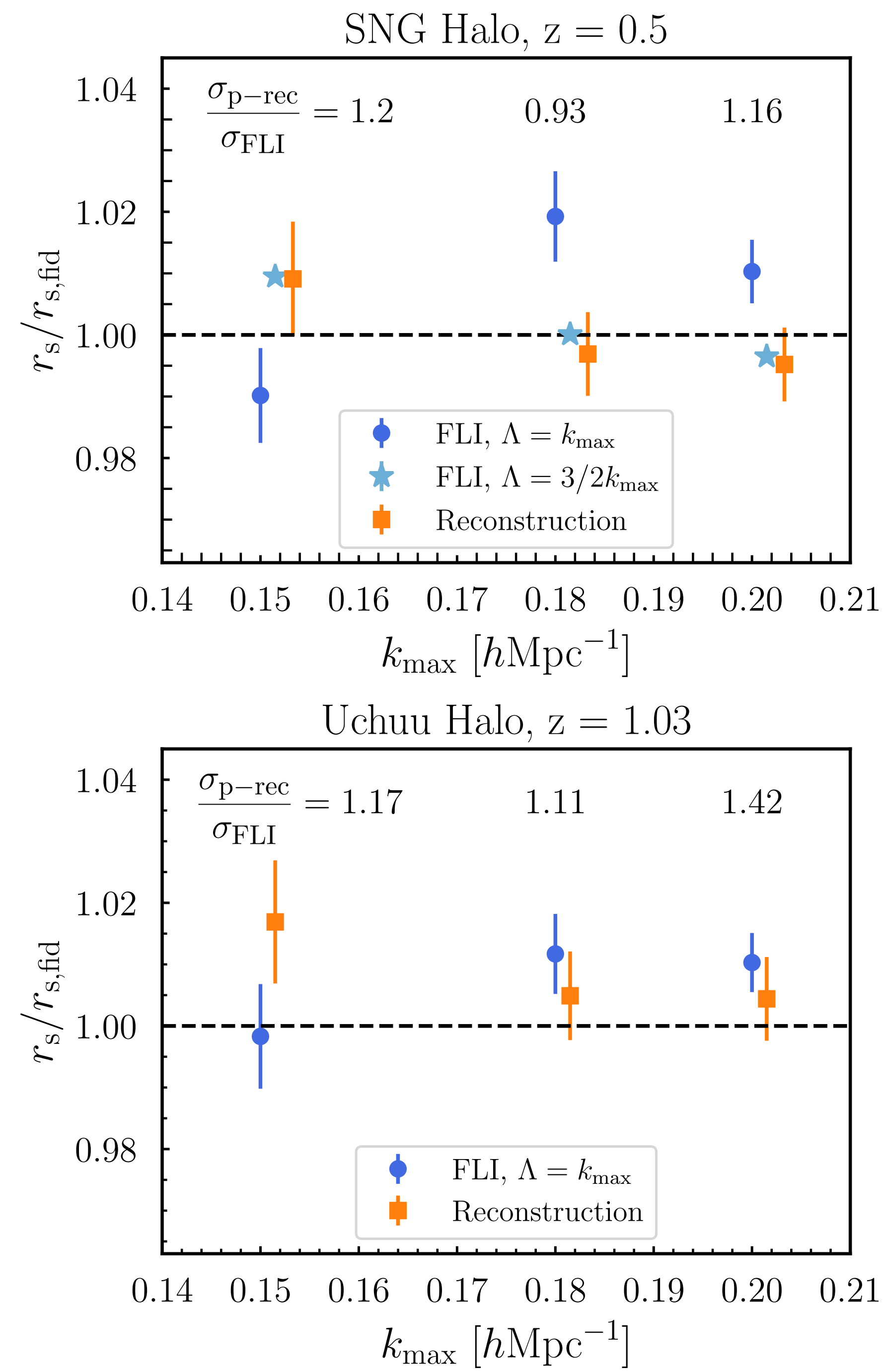
- Field-level inference of BAO scale using a trick: moving BAO feature in linear (initial) density field:

$$f(k, r_s) = \frac{T^2_{\rm BAO}(k|r_s)}{T^2_{\rm BAO}(k|r_{s,{\rm fid}})}\,,$$

$$T^2_{\rm BAO}(k|r_s) = 1 + A\sin(k\,r_s + \phi)\exp(-k/k_{\rm D})$$

- Compare with reconstruction analysis applied to the same scales of the data

- Note: reconstruction uses fixed linear bias, field-level inference infers all bias coefficients jointly with BAO scale




Babić, FS, Tucci (2025), arXiv:2505.13588

# Field-level inference of BAO scale

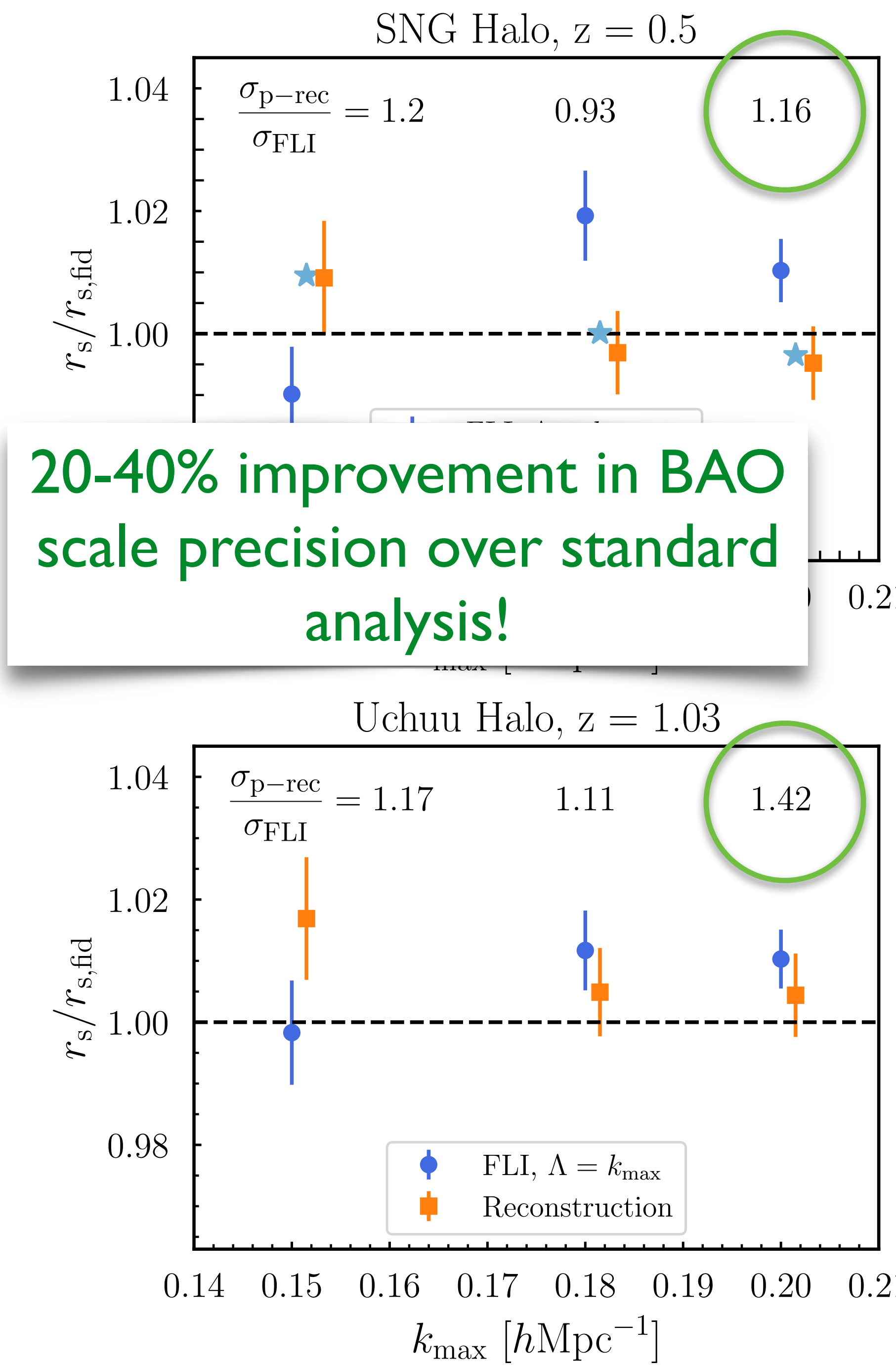
- Field-level inference of BAO scale using a trick: moving BAO feature in linear (initial) density field:

$$f(k, r_s) = \frac{T^2_{\rm BAO}(k|r_s)}{T^2_{\rm BAO}(k|r_{s,\rm fid})}\,,$$

$$T^2_{\rm BAO}(k|r_s) = 1 + A\sin(k\,r_s + \phi)\exp(-k/k_{\rm D})$$

- Compare with reconstruction analysis applied to the same scales of the data

- Note: reconstruction uses fixed linear bias, field-level inference infers all bias coefficients jointly with BAO scale




20-40% improvement in BAO scale precision over standard analysis!

Babić, FS, Tucci (2025), arXiv:2505.13588

# Conclusions

- After all these years we are, currently, still stuck with ΛCDM… but:
- We are continuing to find new signals to search for
- Inference/analysis methods have made tremendous progress — expect to extract much more from the data in coming few years