MAX-PLANCK-INSTITUT
FÜR ASTROPHYSIK

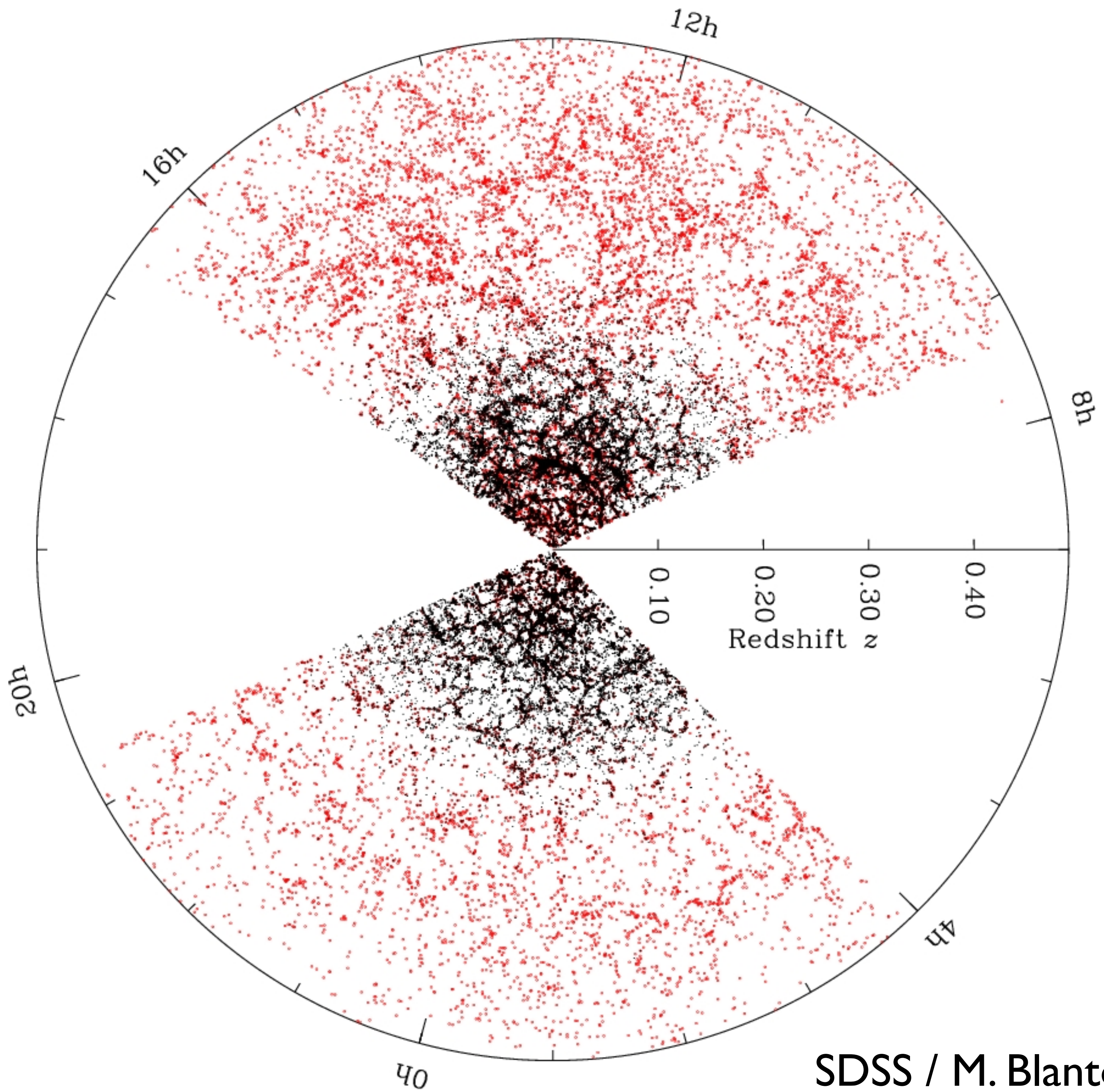
ORIGINS
Excellence Cluster

Non-Gaussianity in Large-Scale Structure

Fabian Schmidt
MPA

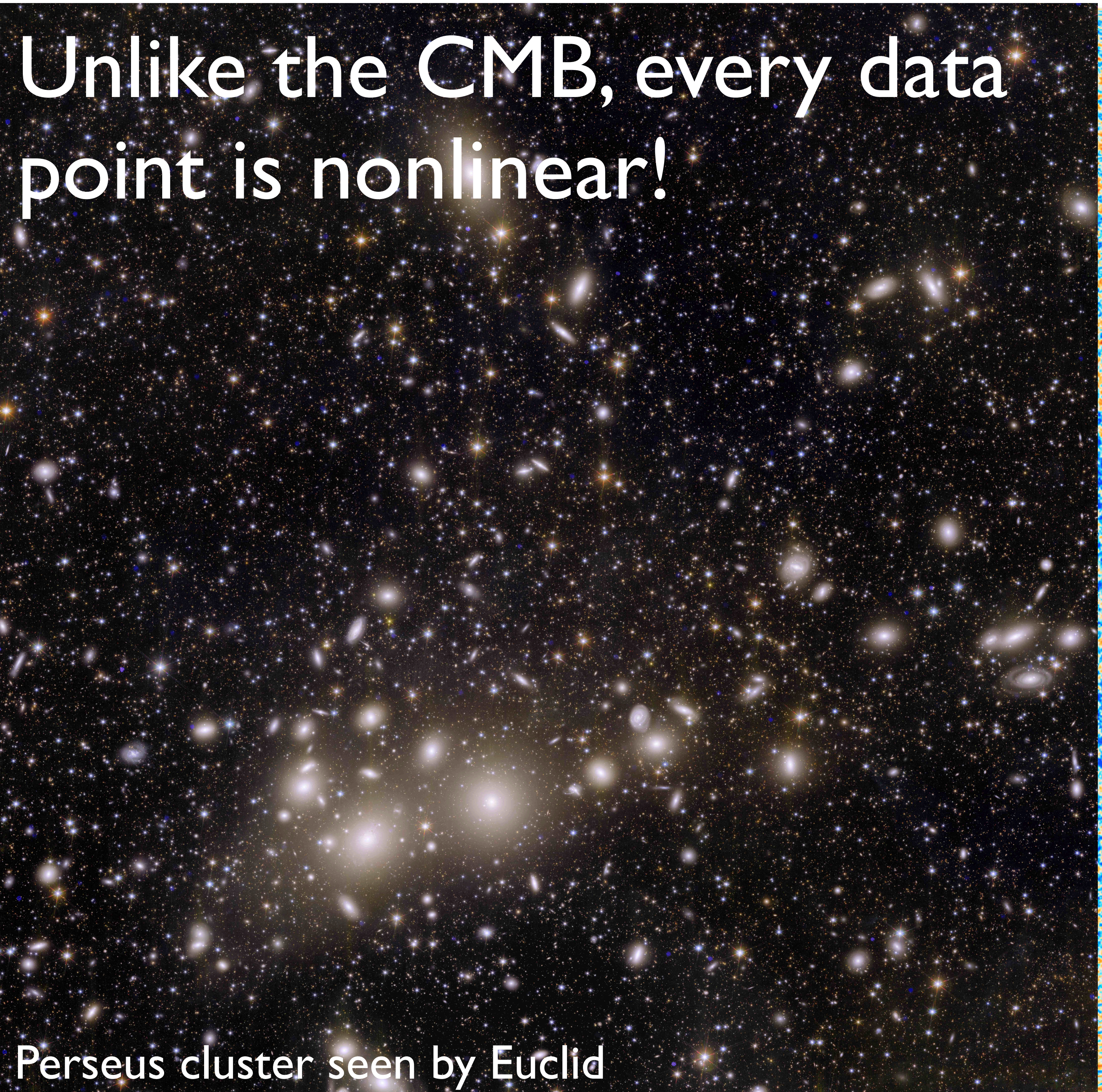
Learning the Deep Mysteries of Nature with Cosmology,
Sep 11, 2025

Non-Gaussianity in Large-Scale Structure

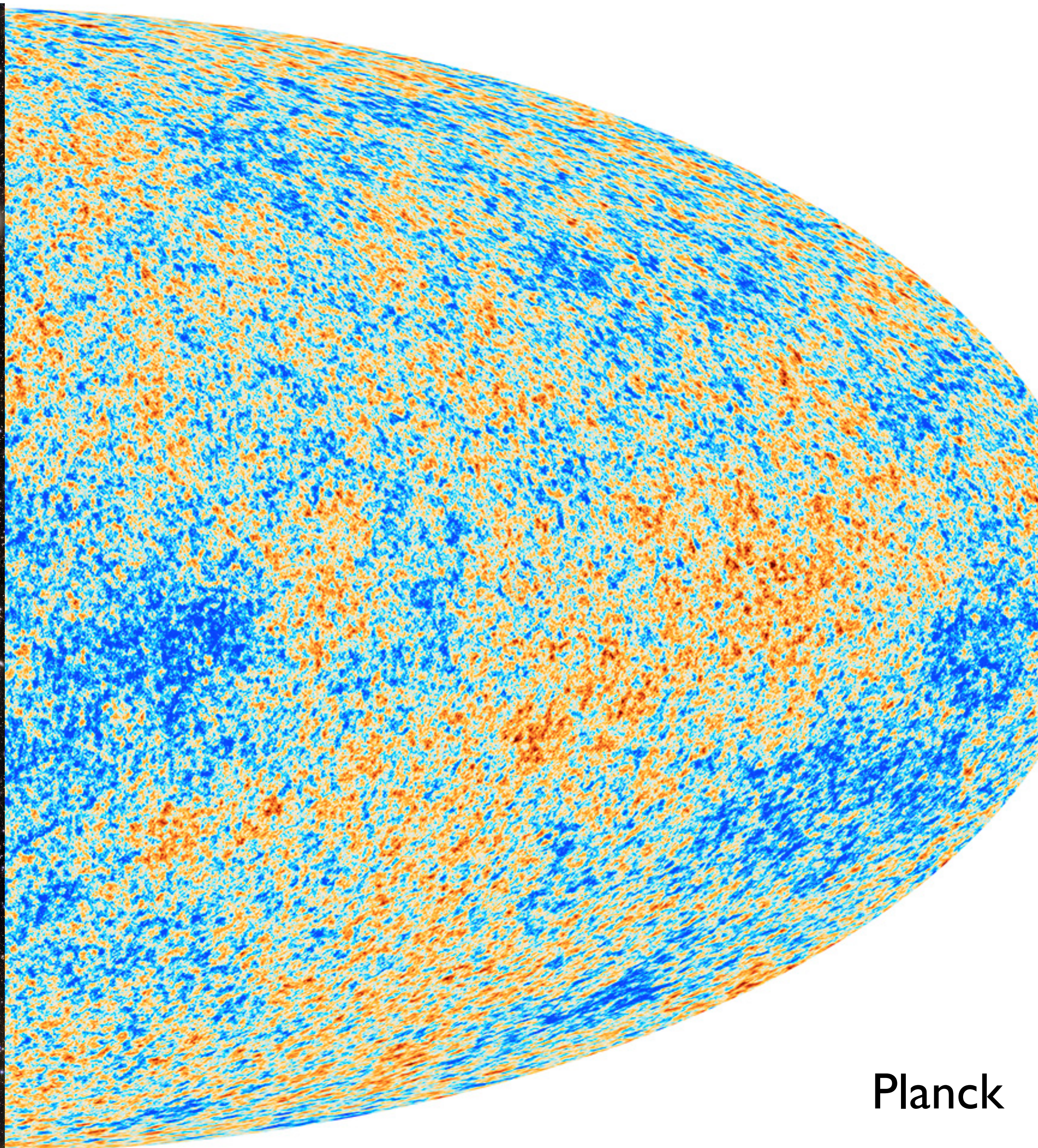


SDSS / M. Blanton

Non-Gaussianity in Large-Scale Structure

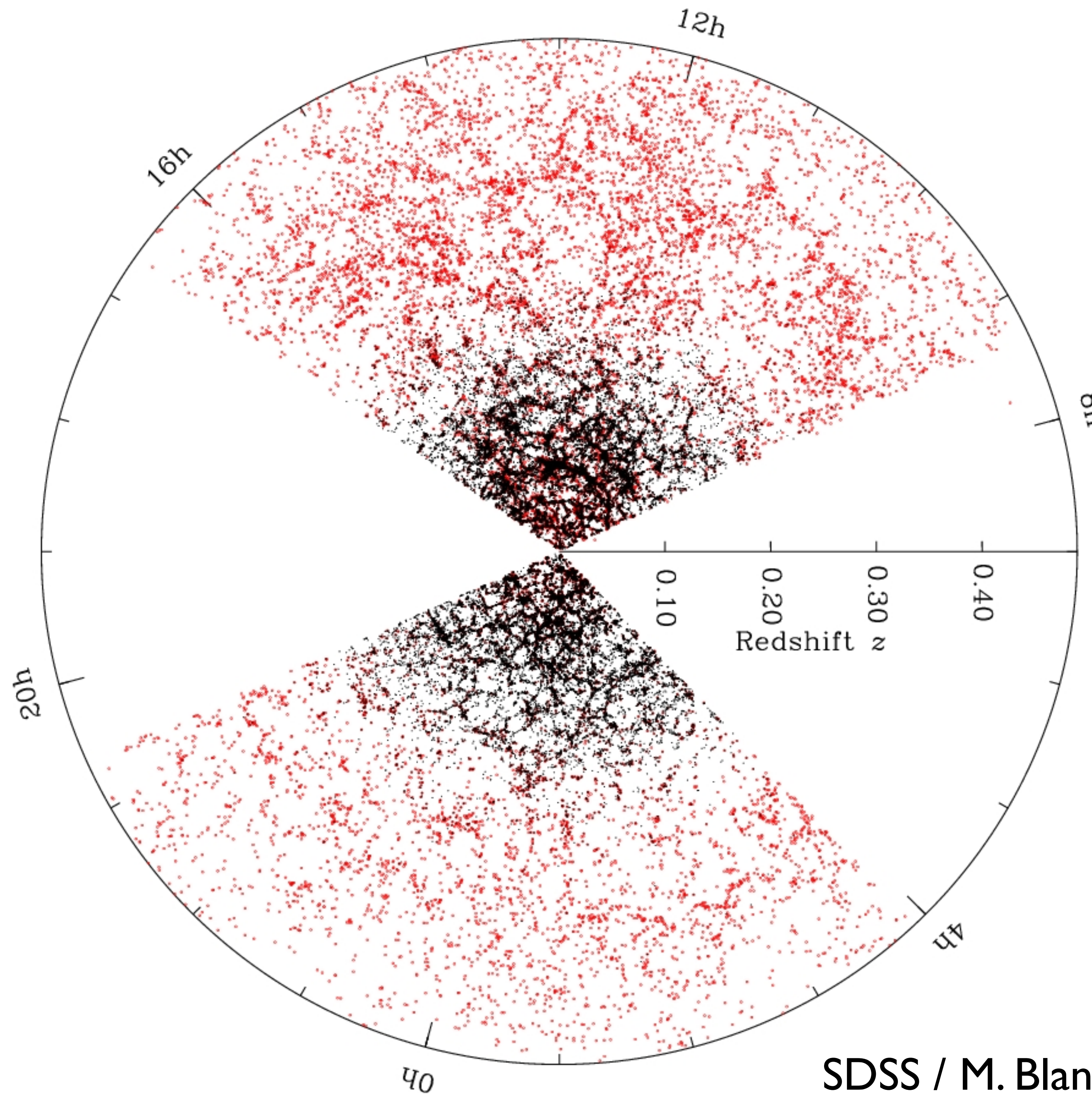


Perseus cluster seen by Euclid


Planck

Non-Gaussianity in Large-Scale Structure

- Prominent non-Gaussianity from nonlinear structure formation (part II)
- But can use this as sensitive of primordial non-Gaussianity as well (part I)



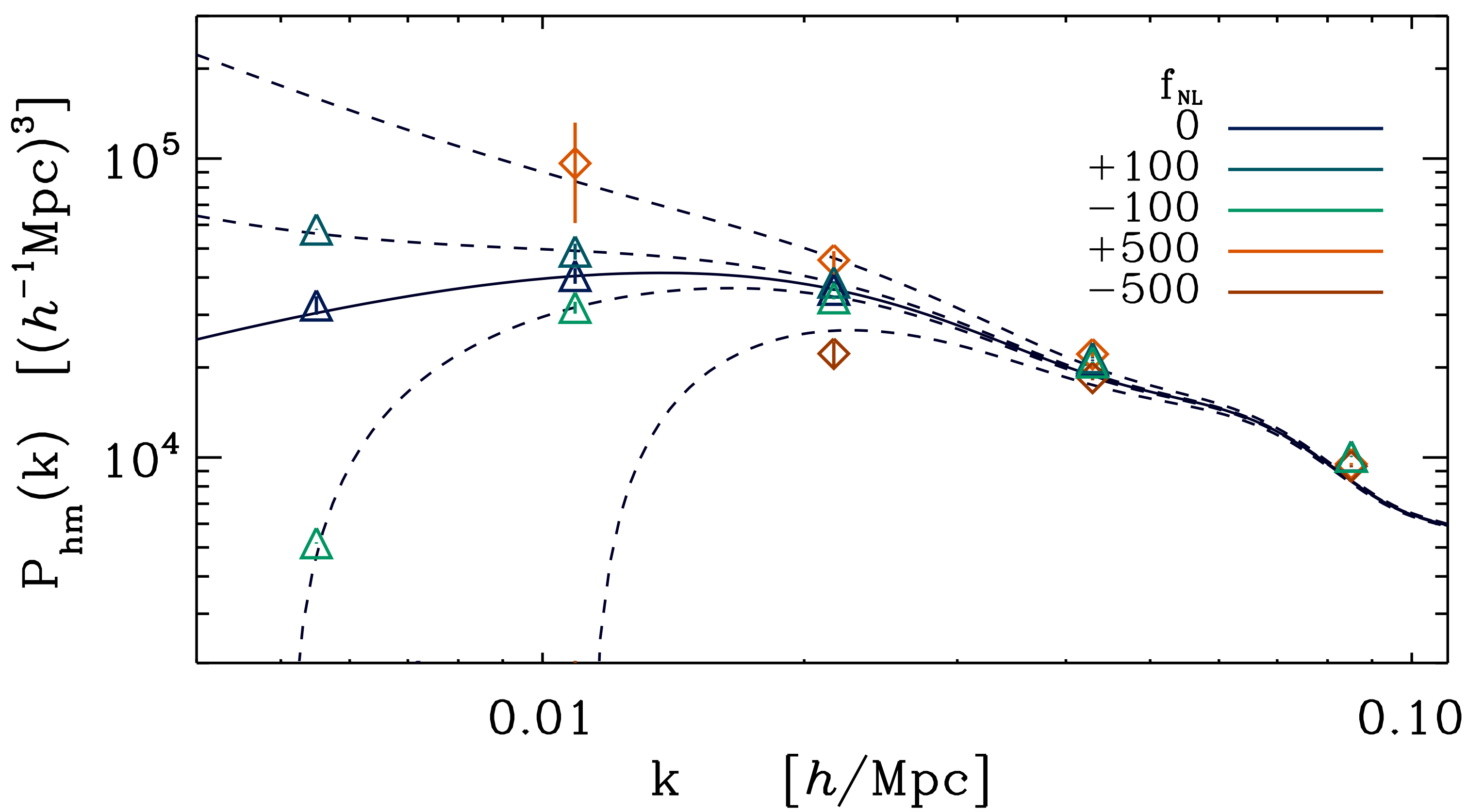
SDSS / M. Blanton

I. Early non-Gaussianities: Probing inflation

Inflation and Galaxy Clustering

- Most prominent signature of inflationary physics in LSS: scale-dependent bias induced by local-type primordial non-Gaussianity
- Discovered in simulations, but it was hiding in theory calculations all along…

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



Dalal et al., 2008
Matarrese/Verde 2008

Inflation and Galaxy Clustering

- How does this generalize to other types of non-Gaussianity, i.e. other types of inflationary physics?
  - Different scalings in squeezed limit
  - Beyond primordial 3-point function

Inflation and Galaxy Clustering

- How does this generalize to other types of non-Gaussianity, i.e. other types of inflationary physics?
  - Different scalings in squeezed limit
  - Beyond primordial 3-point function

A PATH-INTEGRAL APPROACH TO LARGE-SCALE MATTER DISTRIBUTION ORIGINATED BY NON-GAUSSIAN FLUCTUATIONS

SABINO MATARRESE
International School for Advanced Studies, Trieste, Italy
FRANCESCO LUCCHIN
Dipartimento di Fisica G. Galilei, Padova, Italy
AND
SILVIO A. BONOMETTO
International School for Advanced Studies, Trieste, Italy; Dipartimento di Fisica G. Galilei, Padova, Italy; and INFN, Sezione di Padova
Received 1986 July 7; accepted 1986 August 1

Inflation and Galaxy Clustering

- How does this generalize to other types of non-Gaussianity, i.e. other types of inflationary physics?
  - Different scalings in squeezed limit
  - Beyond primordial 3-point function

$$\xi^{(2)}_{R_0;\,\nu,\,R}(x_1, x_2) = \langle \varepsilon_{R_0}(x_1)\,\rho_{\nu,\,R}(x_2)\rangle / \langle \rho_{\nu,\,R}\rangle$$

$$= \left\{ \sum_{L=1}^{\infty} \sum_{[m_L]} \left[\prod_{n=1}^{L} \left(w_R^{(n)}/n! \right)^{m_n} / m_n! \right] \sum_{n=1}^{L} n m_n \left(w^{(n)}_{[R_0:1;\,R:n-1]} \sigma_{R_0} / w_R^{(n)} \right) a_{L-1}(2^{-1/2}\nu) \right\} \Big/ \Pi^{(1)}_{\nu,\,R}$$

Inflation and Galaxy Clustering

- How does this generalize to other types of non-Gaussianity, i.e. other types of inflationary physics?
  - Different scalings in squeezed limit
  - Beyond primordial 3-point function
- Density peaks as a well-defined model, but we were seeking a general result for any physical LSS tracer

Inflation and Galaxy Clustering

- How does this generalize to other types of non-Gaussianity, i.e. other types of inflationary physics?
  - Different scalings in squeezed limit
  - Beyond primordial 3-point function
- Density peaks as a well-defined model, but we were seeking a general result for any physical LSS tracer
- Key ingredient: scaling of primordial correlators in squeezed limit
  - Interesting limit to constrain new particles: *cosmological collider*
- We can predict the scale-dependence of LSS statistics, but amplitude is controlled by tracer-dependent bias parameter

FS, Kamionkowski (2010)
Desjacques, Jeong, FS (2011+)

Inflation and Galaxy Clustering

- How does thi … on-Gaussianity, i.e. other types o
    - Different s
    - Beyond pri
- Density peaks … e were seeking a general result
- Key ingredien … rs in squeezed limit
    - Interesting … : *cosmological collider*
- We can predi … statistics, but amplitude is c … bias parameter

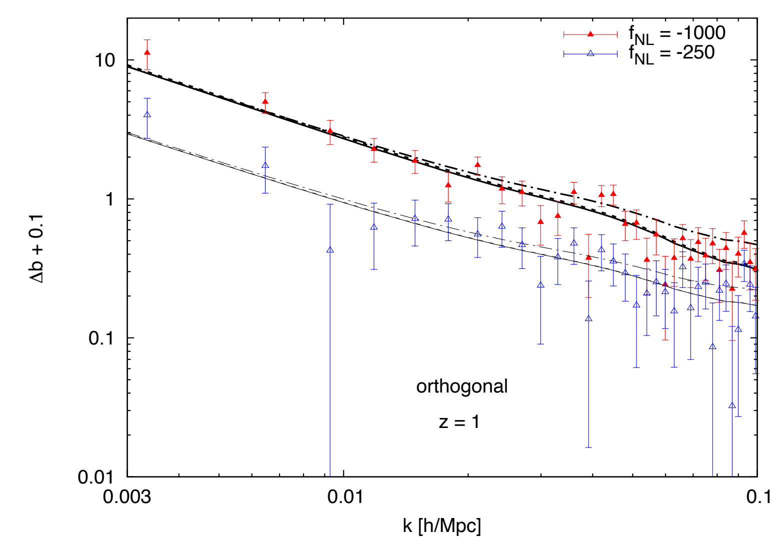


Figure 9. Same as in Fig. 8, but for the orthogonal shape of non-Gaussianity. Note, however, that here the redshift of the halos with mass 1.2×10^{14} Mpc/$h < M < 2.4 \times 10^{14}$ Mpc/h is $z = 1$.

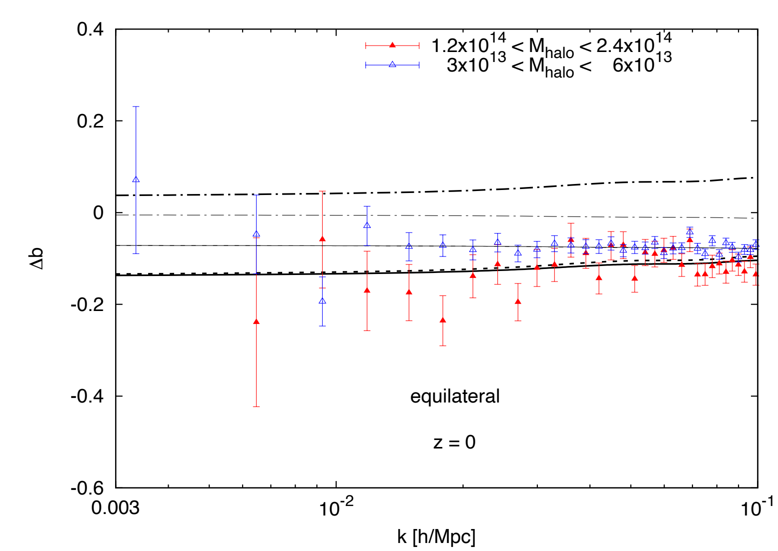


Figure 10. Same as in Fig. 8, but for the equilateral shape of non-Gaussianity. Note the linear scale of the y-axis.

Wagner & Verde (2011)

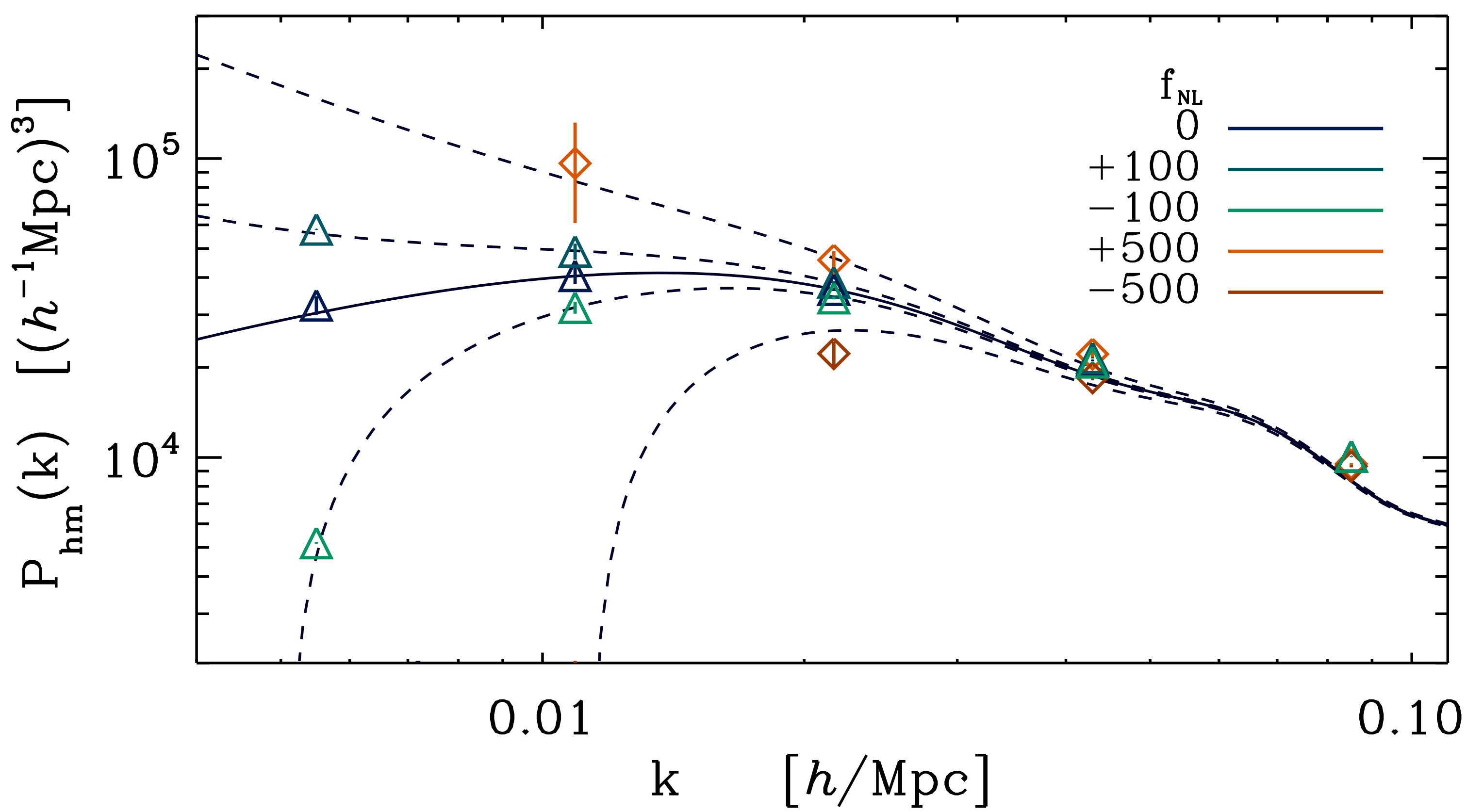
Connecting inflation with LSS in GR

- Smoking-gun signature of *multi-field inflation*
- Probing highest energy physics with galaxies on the largest scales
- Current constraints:

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

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

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

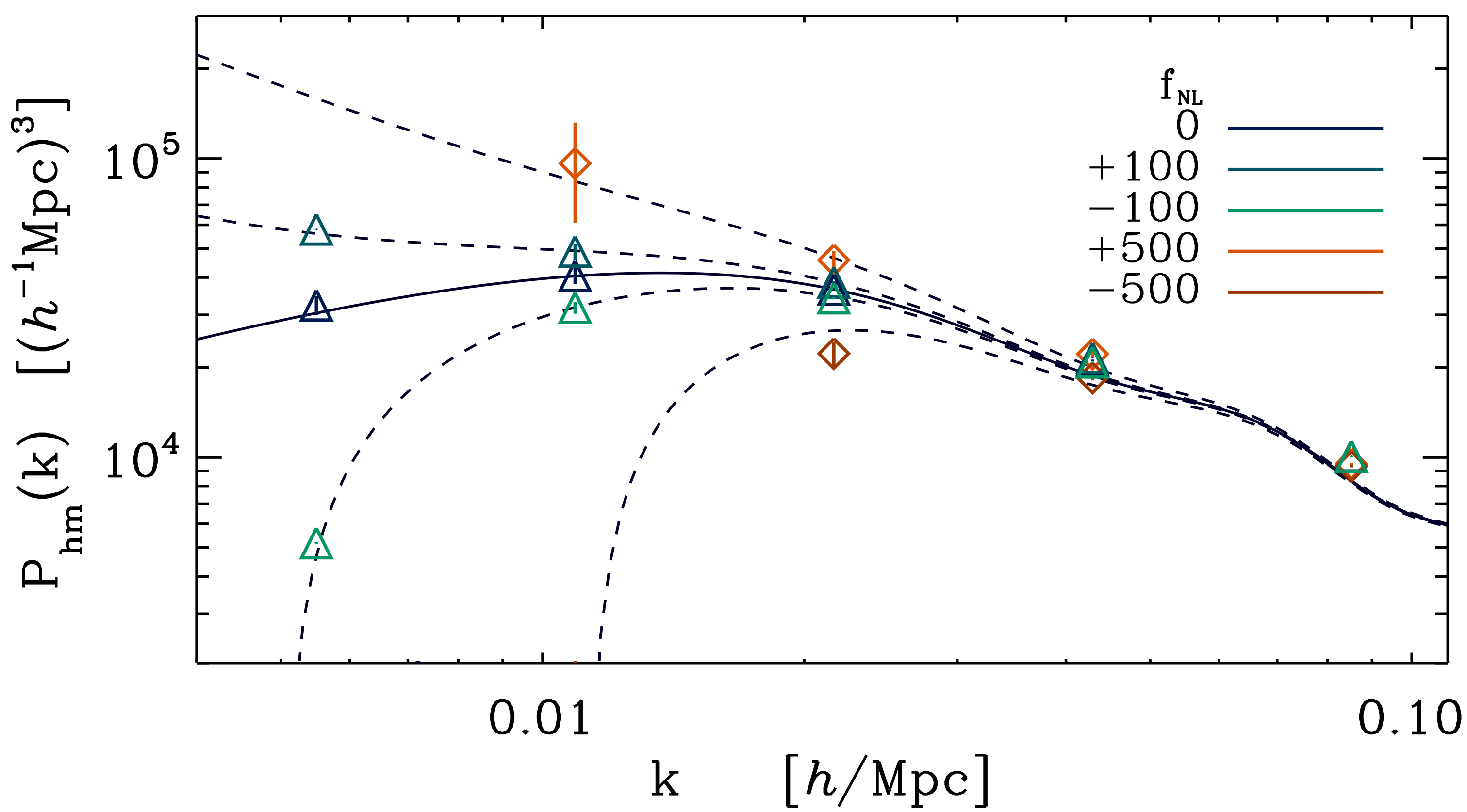


Connecting inflation with LSS in GR

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

- f_{NL}~1 corresponds to a contribution of order Φ to observed galaxy density
- So we need to worry about relativistic corrections to usual quasi-Newtonian treatment of galaxy clustering!

$$\Phi(\boldsymbol{k}) \sim \left(\frac{aH}{k}\right)^2 \delta(\boldsymbol{k})$$

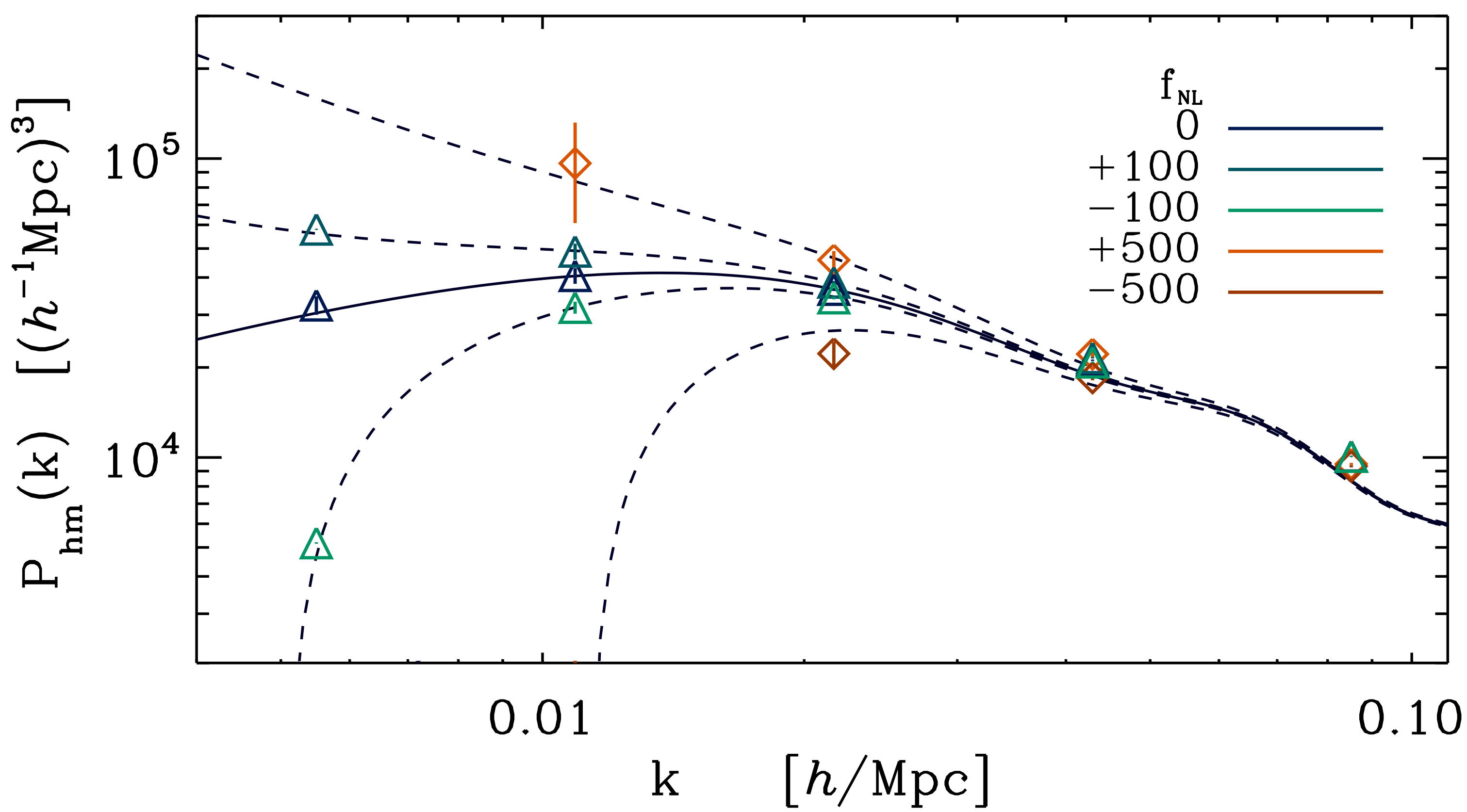


Connecting inflation with LSS in GR

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

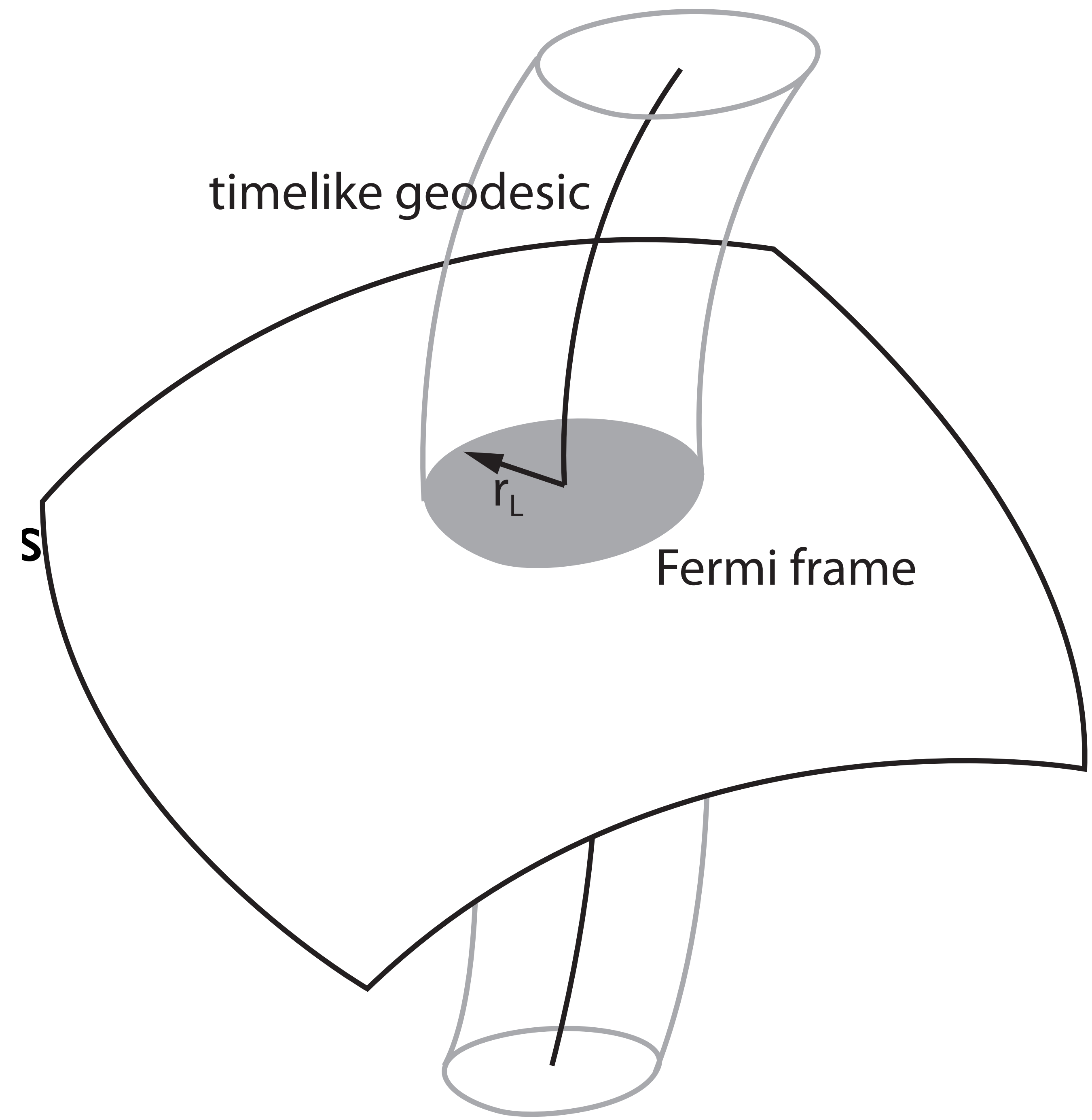
- f_{NL}~1 corresponds to a contribution of order Φ to observed galaxy density
- So we need to worry about relativistic corrections to usual quasi-Newtonian treatment of galaxy clustering!
- In fact, since PNG is a second-order effect, it seems we need a second-order GR calculation
- Goal: get around this by focusing on squeezed limit: coupling of long- with short modes

$$\Phi(\boldsymbol{k}) \sim \left(\frac{aH}{k}\right)^2 \delta(\boldsymbol{k})$$



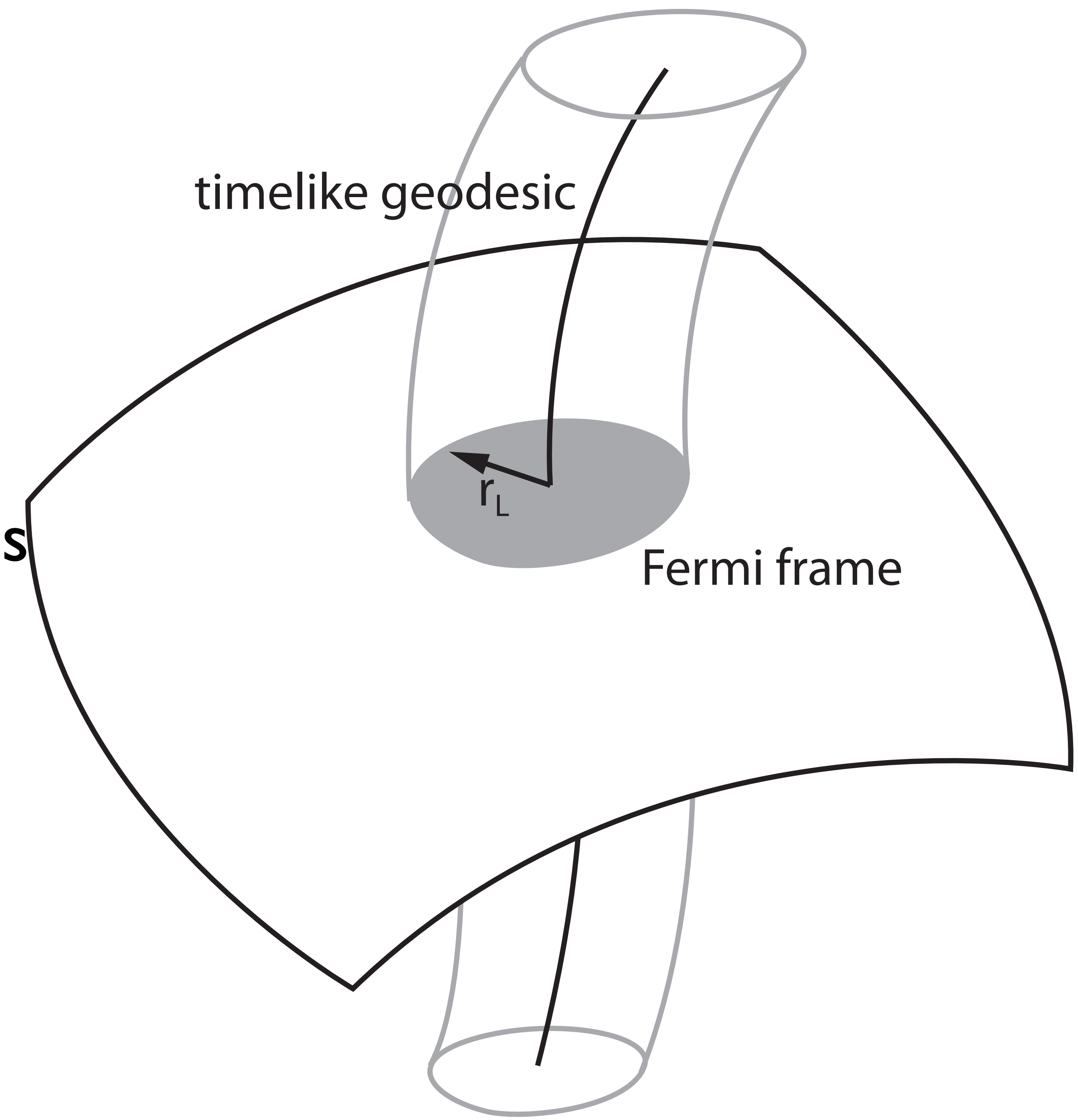
From initial conditions to galaxies

- Consider wordline of a small patch within the Universe



From initial conditions to galaxies

- Consider wordline of a small patch within the Universe
- We can go to a frame so that close to this wordline, the spacetime looks flat at all times
  - Time coordinate is proper time along geodesic
- ***Fermi frame:*** natural frame to describe local gravitational experiments

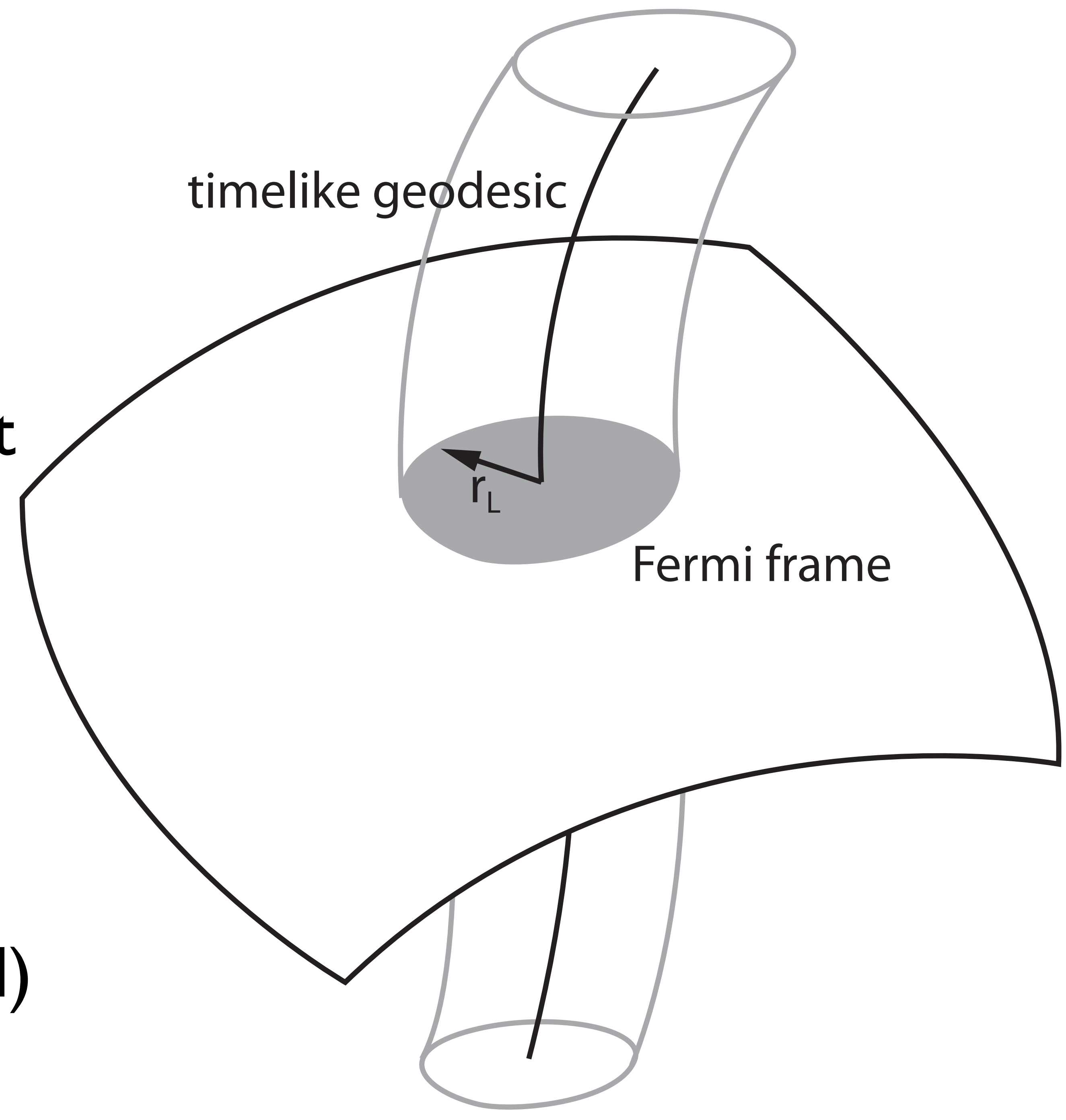


From initial conditions to galaxies

- Generalize to ***Conformal Fermi frame:***
- Valid in region (even outside horizon) around geodesic so that at all times

$$g_{\mu\nu} = a^2(\tau_F)\left[\eta_{\mu\nu} + \mathcal{O}(\mathbf{x}^2)\right]$$

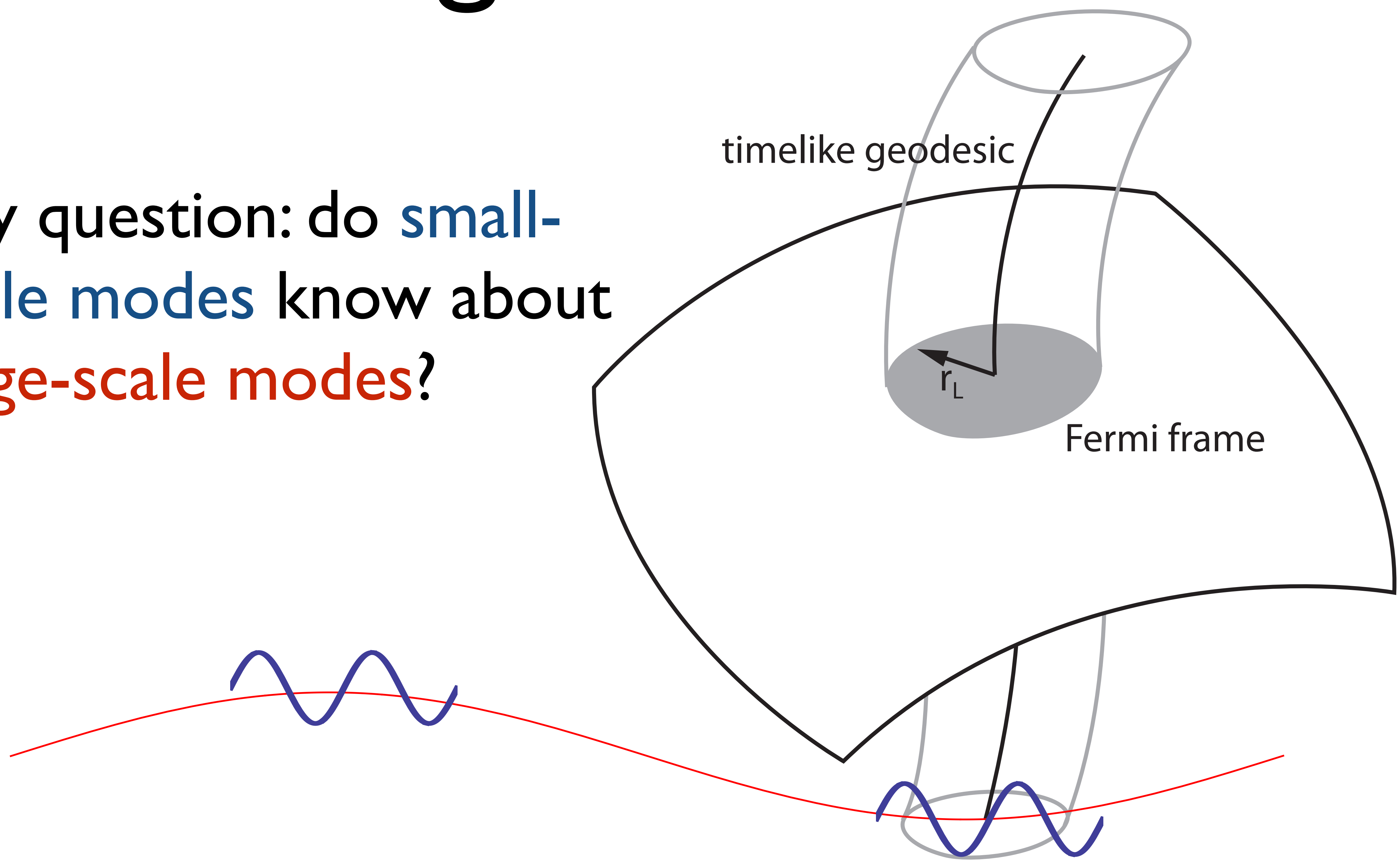
  - Spatial origin is on central geodesic
  - Time coordinate is (conformal) proper time along geodesic



Pajer, FS, Zaldarriaga (2013)

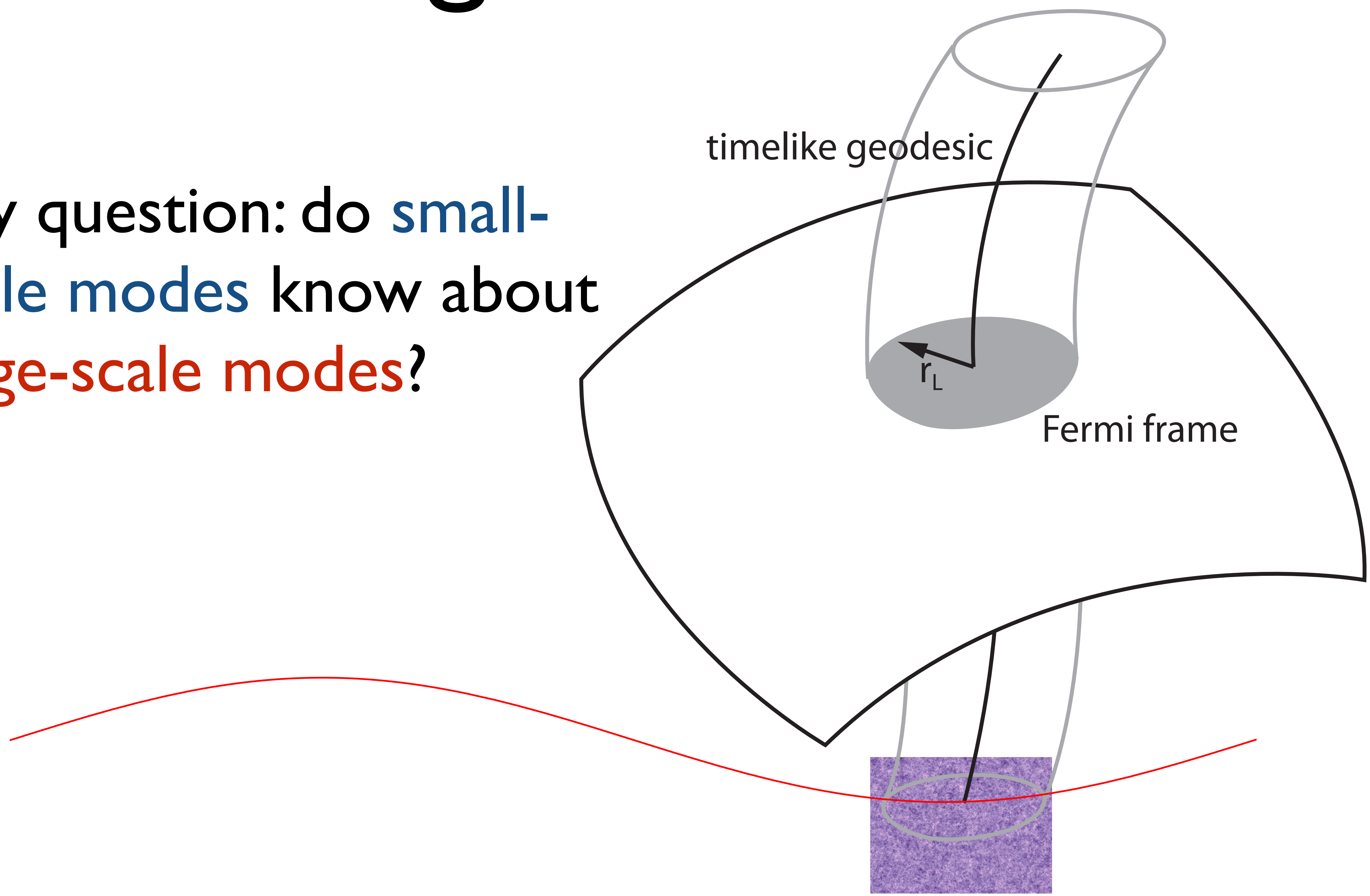
From initial conditions to galaxies

- Key question: do small-scale modes know about large-scale modes?



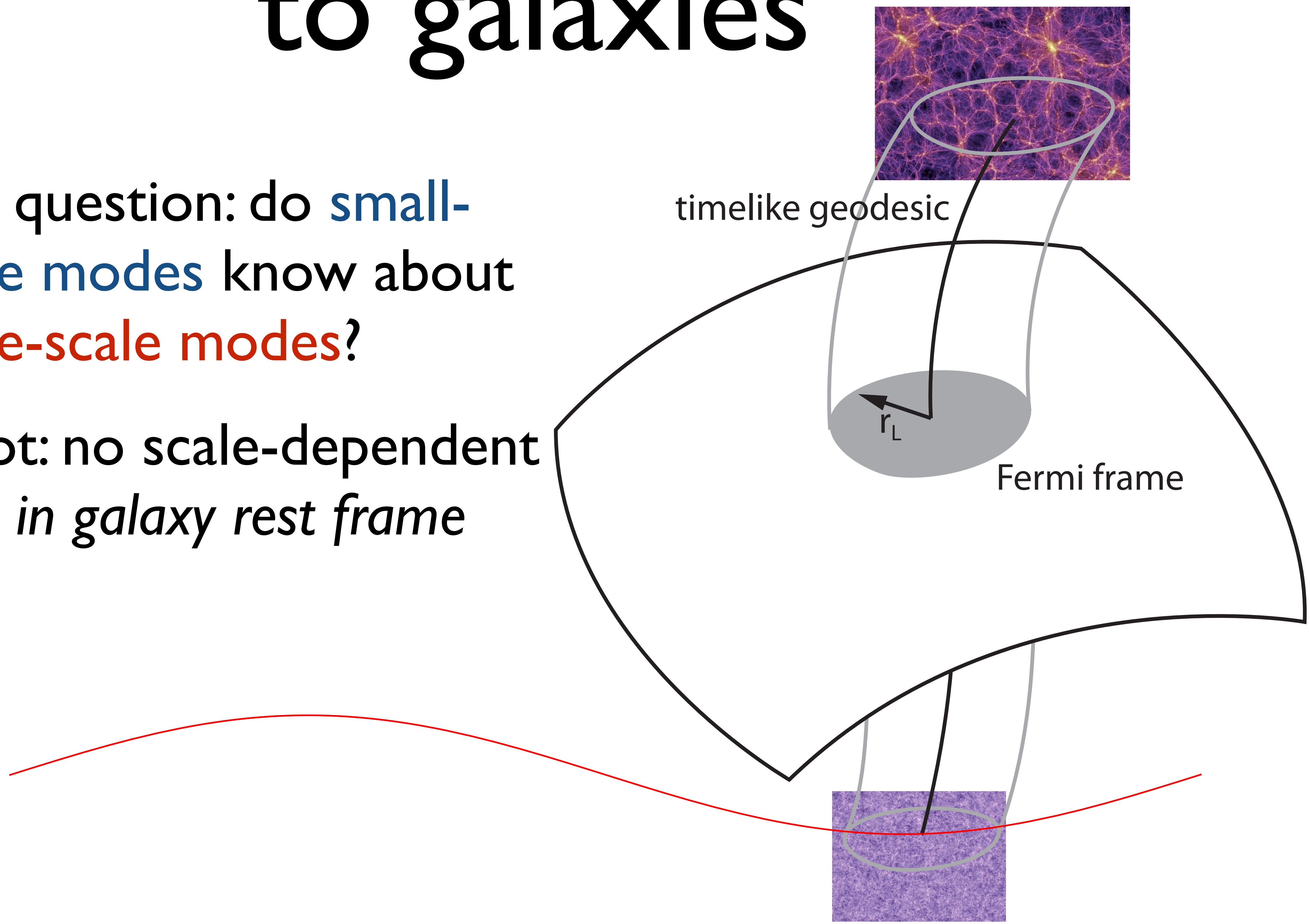
From initial conditions to galaxies

- Key question: do small-scale modes know about large-scale modes?

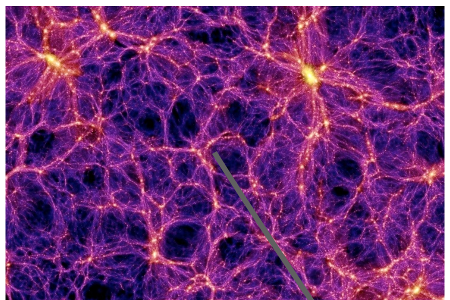


From initial conditions to galaxies

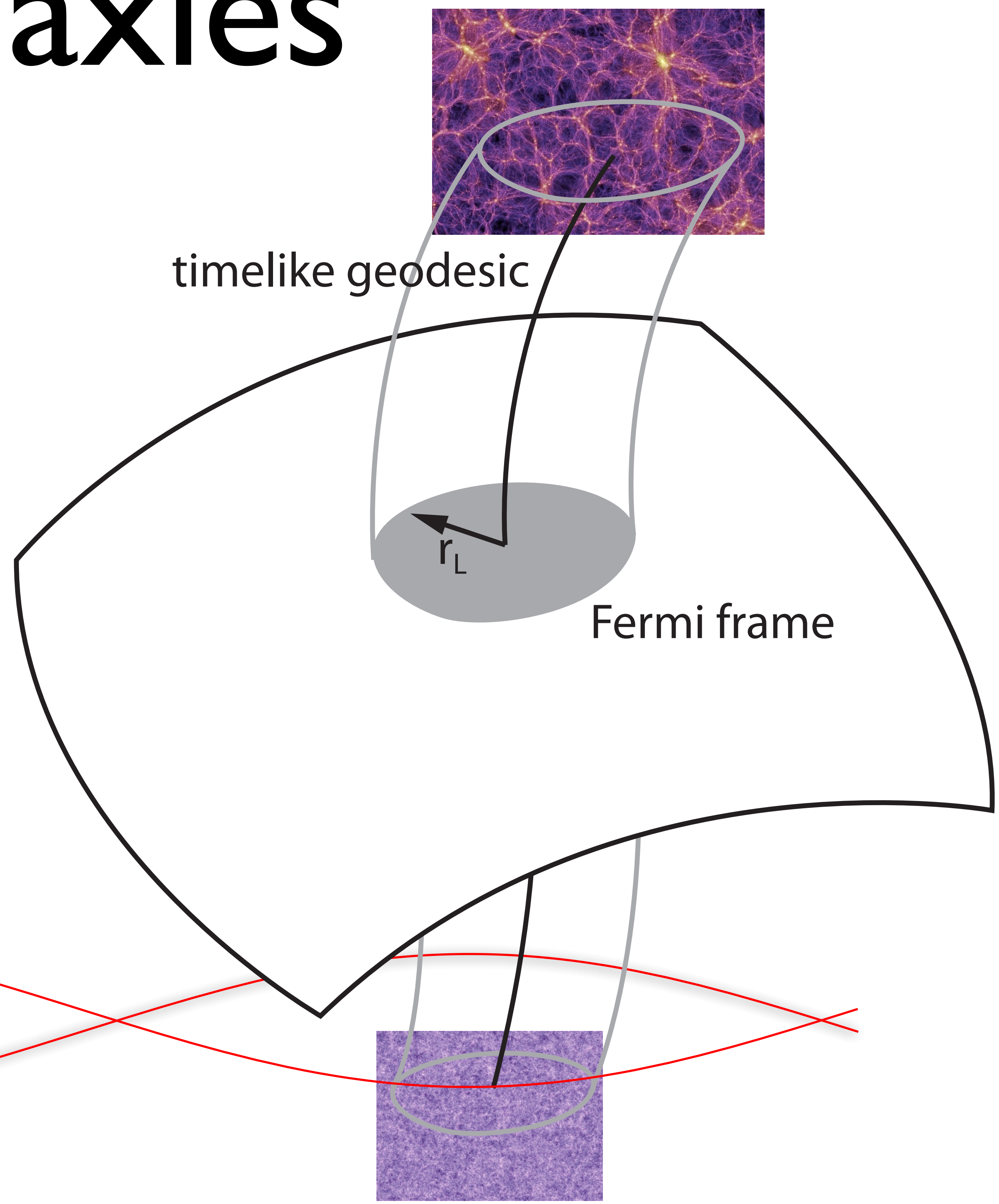
- Key question: do small-scale modes know about large-scale modes?
- If not: no scale-dependent bias *in galaxy rest frame*

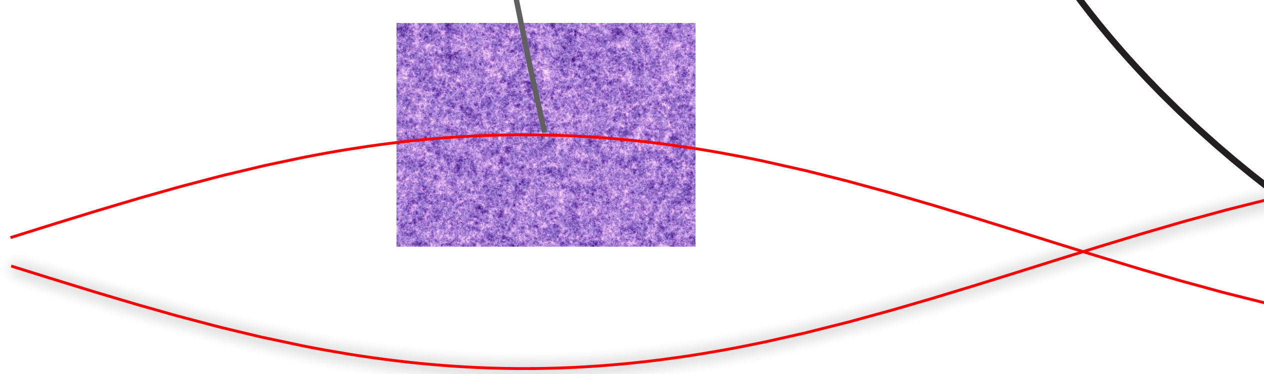


From initial conditions to galaxies



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





$$\langle \delta^2 \rangle = [1 + 4 f_{\rm NL} \phi(\mathbf{x})] \sigma^2$$

Consistency relation and Fermi frame

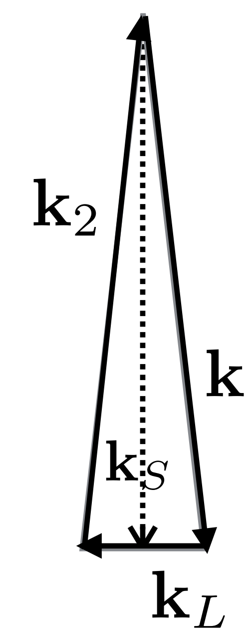
- Metric in comoving gauge (neglecting shift and lapse): $ds^2 = a^2(\tau)[-d\tau^2 + e^{2\zeta} d\mathbf{x}^2]$
- Transform to conformal Fermi frame:

$$x'^i = (1-\zeta)x^i - \frac{1}{2}\partial_j \zeta x^j x^i$$

Consistency relation and Fermi frame

- Metric in comoving gauge (neglecting shift and lapse): $ds^2 = a^2(\tau)[-d\tau^2 + e^{2\zeta} d\mathbf{x}^2]$
- Transform to conformal Fermi frame:

$$x'^i = (1-\zeta)x^i - \frac{1}{2}\partial_j \zeta x^j x^i$$



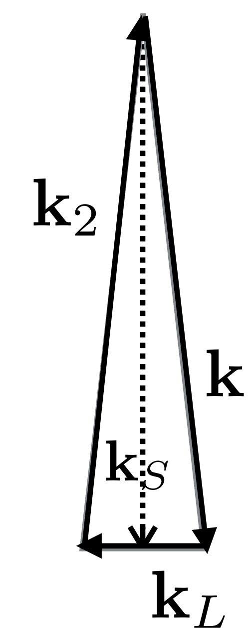
$\mathbf{k}_S = (\mathbf{k}_1 - \mathbf{k}_2)/2$

- Bispectrum in squeezed limit transforms as:

Consistency relation and Fermi frame

- Metric in comoving gauge (neglecting shift and lapse): $ds^2 = a^2(\tau)[-d\tau^2 + e^{2\zeta}d\mathbf{x}^2]$
- Transform to conformal Fermi frame:

$$x'^i = (1-\zeta)x^i - \frac{1}{2}\partial_j\zeta x^j x^i$$

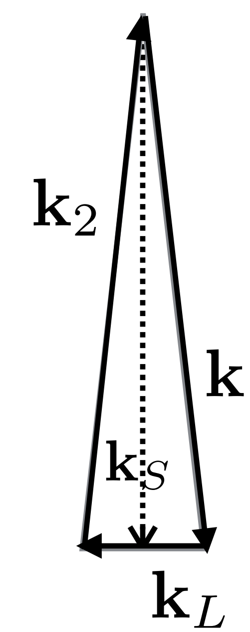


$\mathbf{k}_S = (\mathbf{k}_1 - \mathbf{k}_2)/2$

- Bispectrum in squeezed limit transforms as:

$$B'_\zeta(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = B_\zeta(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) + P_\zeta(k_L)P_\zeta(k_S)\frac{d\ln k_S^3 P_\zeta(k_S)}{d\ln k_S}$$

Consistency relation and Fermi frame



$$\mathbf{k}_S = (\mathbf{k}_1 - \mathbf{k}_2)/2$$

- Metric in comoving gauge (neglecting shift and lapse): $ds^2 = a^2(\tau)[-d\tau^2 + e^{2\zeta} d\mathbf{x}^2]$
- Transform to conformal Fermi frame:

$$x'^i = (1-\zeta)x^i - \frac{1}{2}\partial_j \zeta x^j x^i$$

- Bispectrum in squeezed limit transforms as:

$$B'_\zeta(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = B_\zeta(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) + P_\zeta(k_L) P_\zeta(k_S) \frac{d \ln k_S^3 P_\zeta(k_S)}{d \ln k_S}$$

$$= \mathcal{O}\left(\frac{k_L^2}{k_S^2}\right)$$

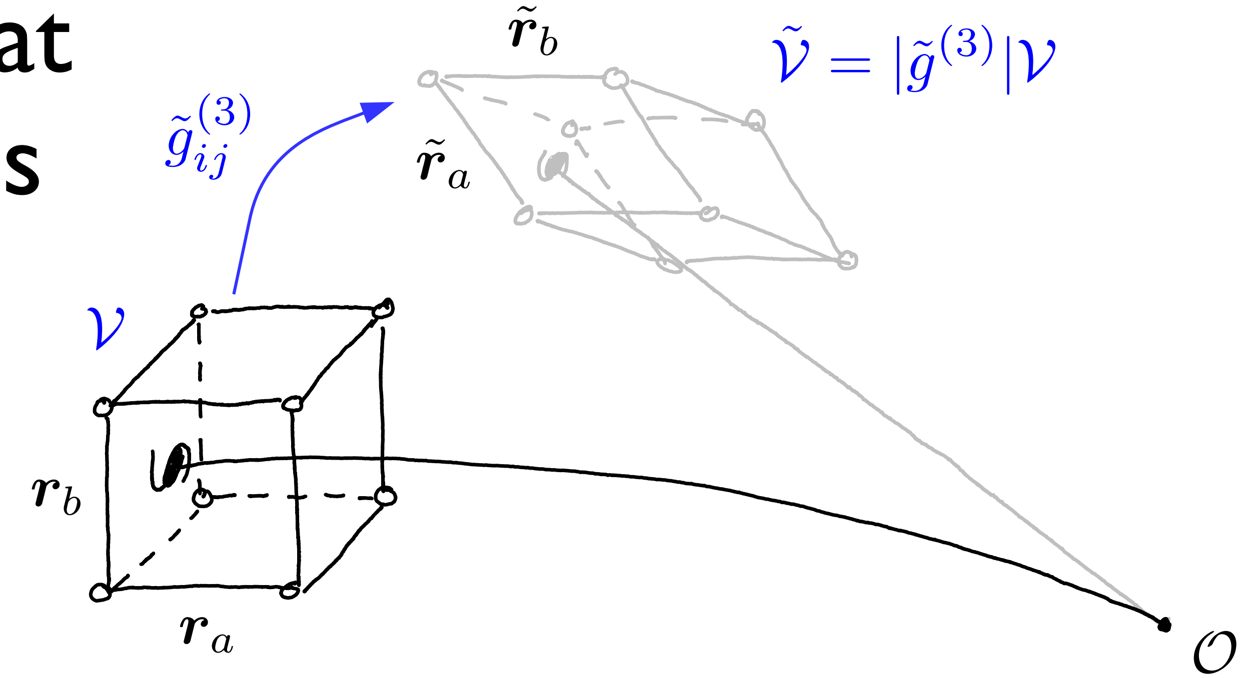
Consistency relation

Observed galaxy clustering

- No coupling of large- and small-scale fluctuations in single field inflation (attractor regime) *in local rest frame* (Fermi frame)
- But still need to map to observations of distant observer
- Nontrivial relativistic corrections at late times

Observed galaxy clustering

- Mapping local Fermi frame to distant observer's measurements understood at linear order in perturbations
  - sufficient for large-scale galaxy P(k)
- Second-order relativistic effects: still work in progress…



Bartolo, Bertacca, Bruni, Matarrese; Bonvin, Clarkson, Durrer, Maartens, Umeh; Ginat (incl. FS) …

II. Late non-Gaussianities: Probing the growth of structure

II. Late non-Gaussianities: Probing the growth of structure

Mon. Not. R. Astron. Soc. **290**, 651–662 (1997)

Large-scale bias in the Universe: bispectrum method

S. Matarrese,[1] L. Verde[1,2] and A. F. Heavens[2]

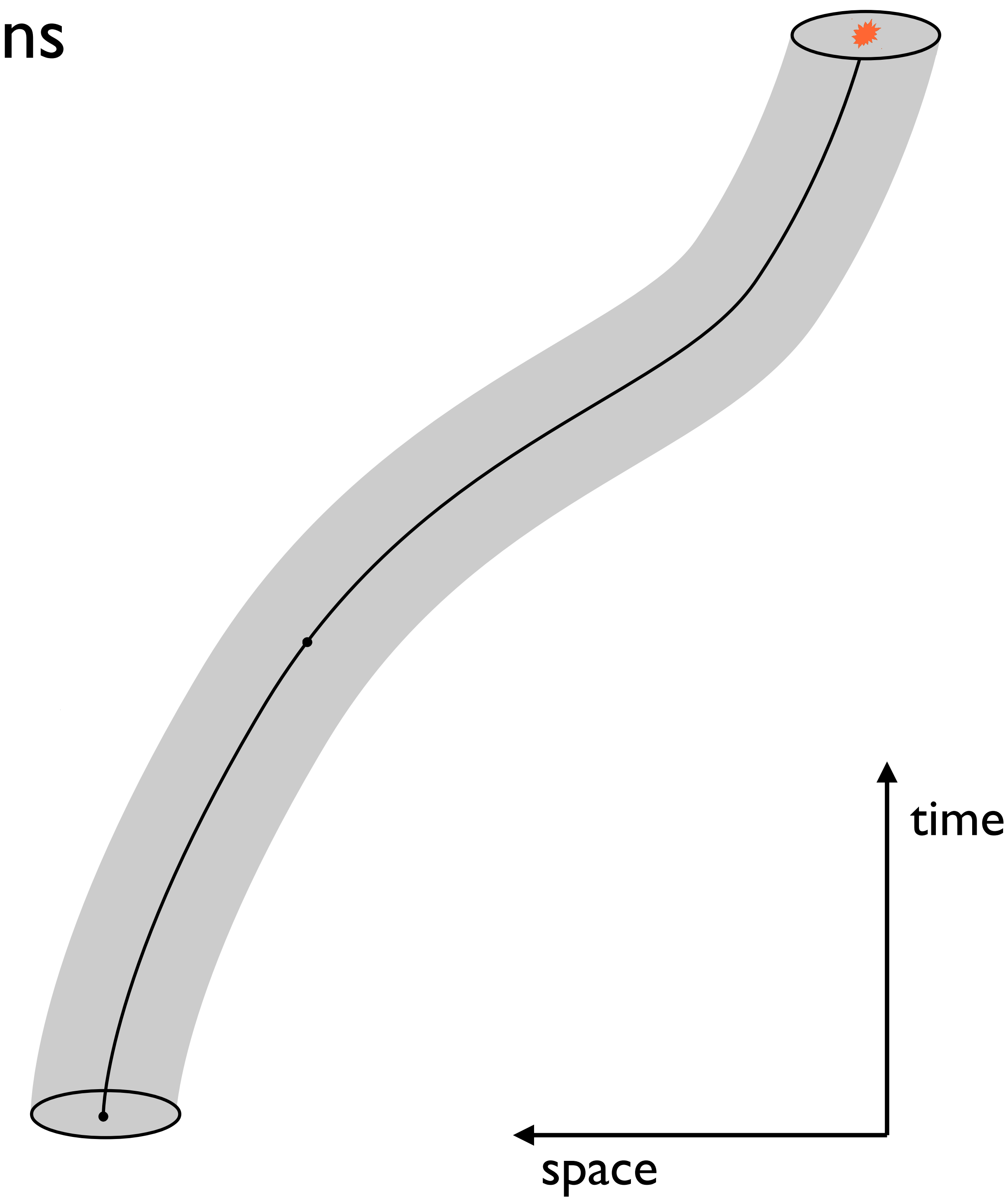
[1]*Dipartimento di Fisica Galileo Galilei, Università di Padova, via Marzolo 8, I-35131 Padova, Italy*
[2]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*

In this paper, we develop an idea of Fry, using second-order perturbation theory to investigate how to measure the bias parameter on large scales. The use of higher order statistics allows the degeneracy between b and Ω_0 to be lifted, and an unambiguous determination of Ω_0 then becomes possible. We apply a likelihood approach to the bispectrum, the three-point function in Fourier space. This paper is

$$\delta_g(\boldsymbol{x}) \simeq b_1 \delta^{(1)}(\boldsymbol{x}) + b_1 \delta^{(2)}(\boldsymbol{x}) + \frac{1}{2} b_2 \delta^{(1)2}(\boldsymbol{x}). \tag{9}$$

Bispectrum breaks bias degeneracies thanks to equivalence principle

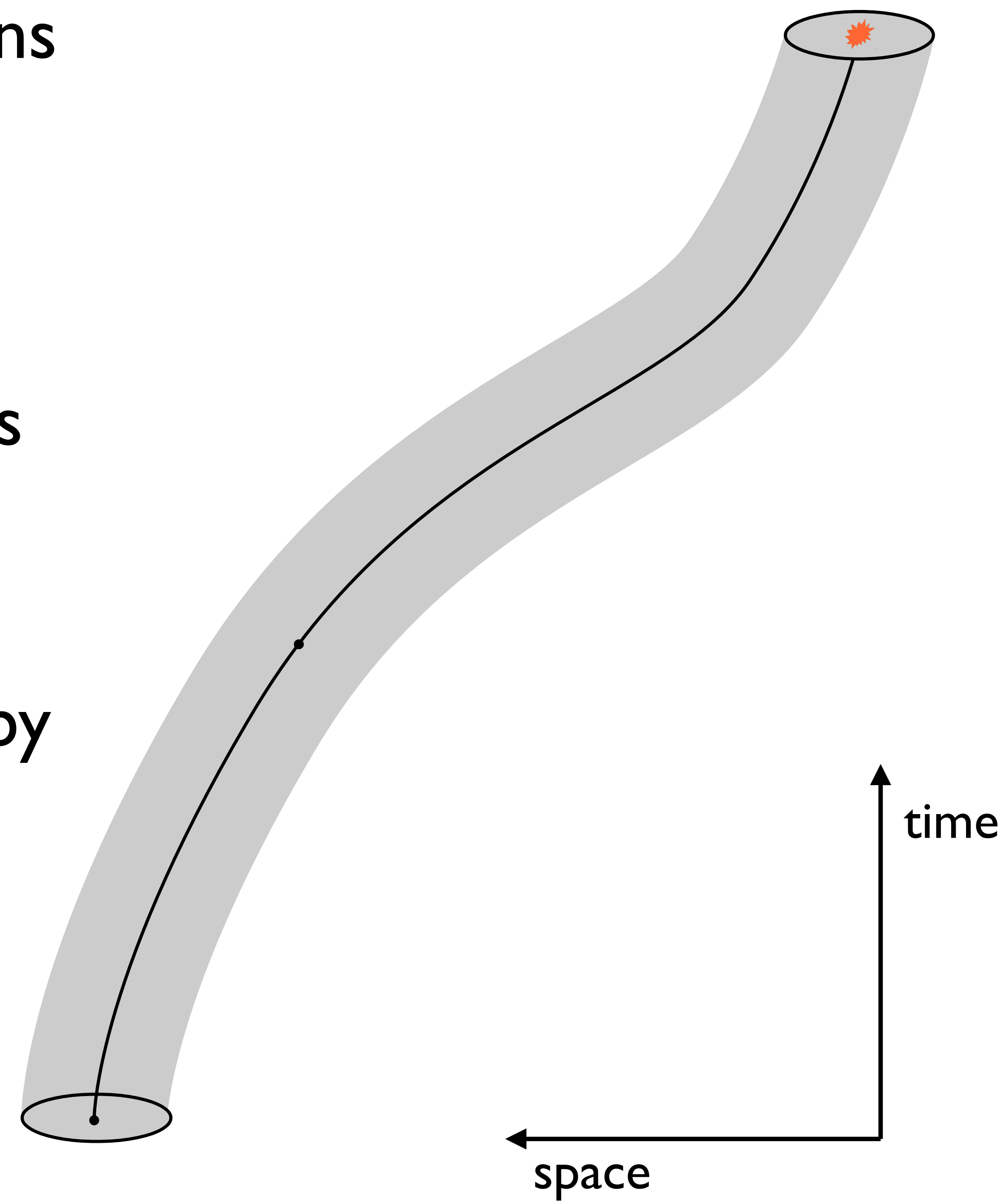
- We cannot predict galaxy positions from first principles; capture uncertainties in effective bias coefficients



Desjacques, Jeong, FS (2017)

Bispectrum breaks bias degeneracies thanks to equivalence principle

- We cannot predict galaxy positions from first principles; capture uncertainties in effective bias coefficients
- Leading gravitational observable is tidal field $\partial_i \partial_j \Phi$ which includes density $\delta \propto \nabla^2 \Psi$
- Some coefficients are protected by equivalence principle - precisely the ones that allow to break degeneracy of bias and amplitude

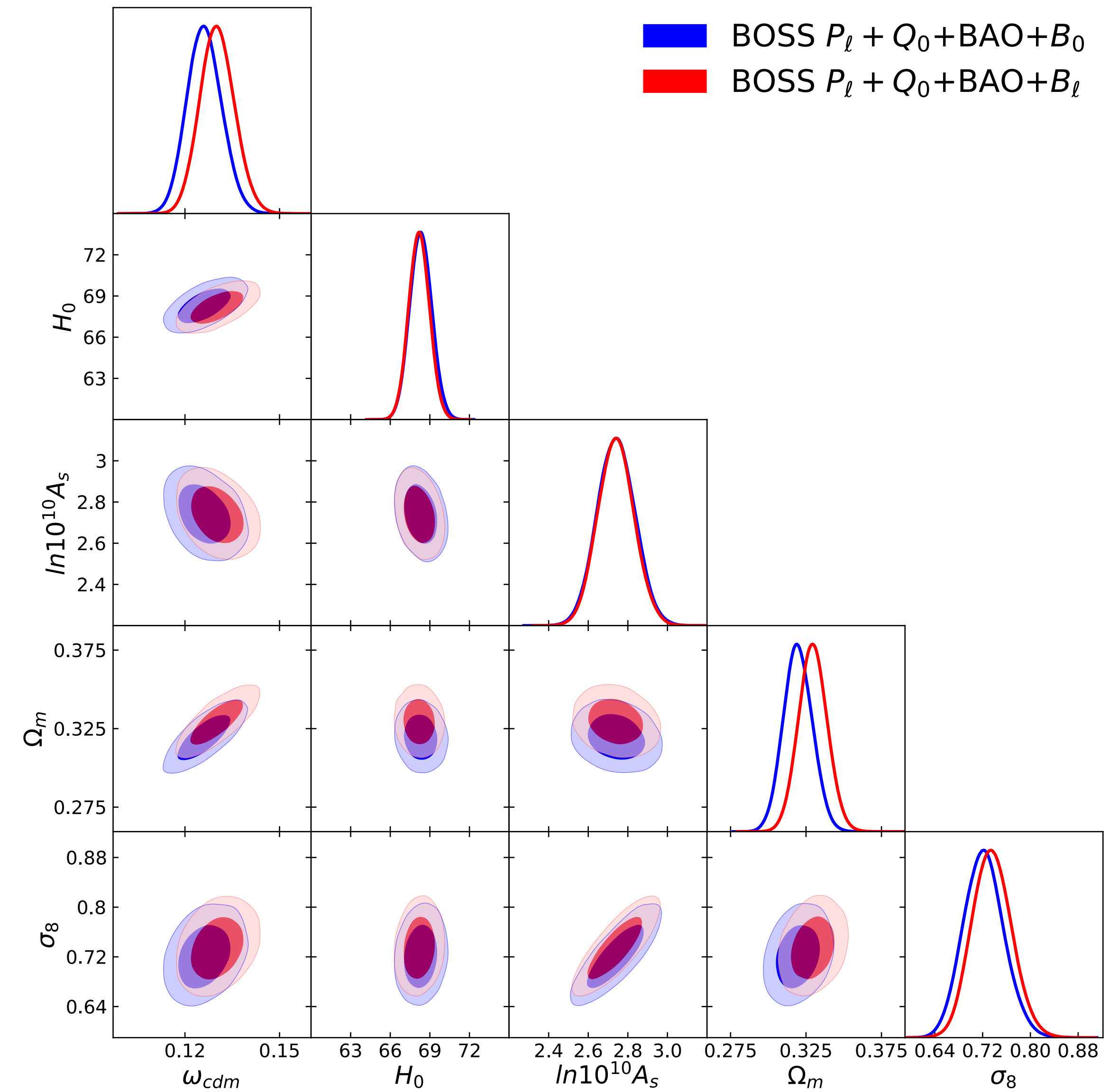


Desjacques, Jeong, FS (2017)

Current state: power spectrum + bispectrum

- Protected *displacement* terms in galaxy density start at second order
- These probe growth factor (or σ_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

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

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

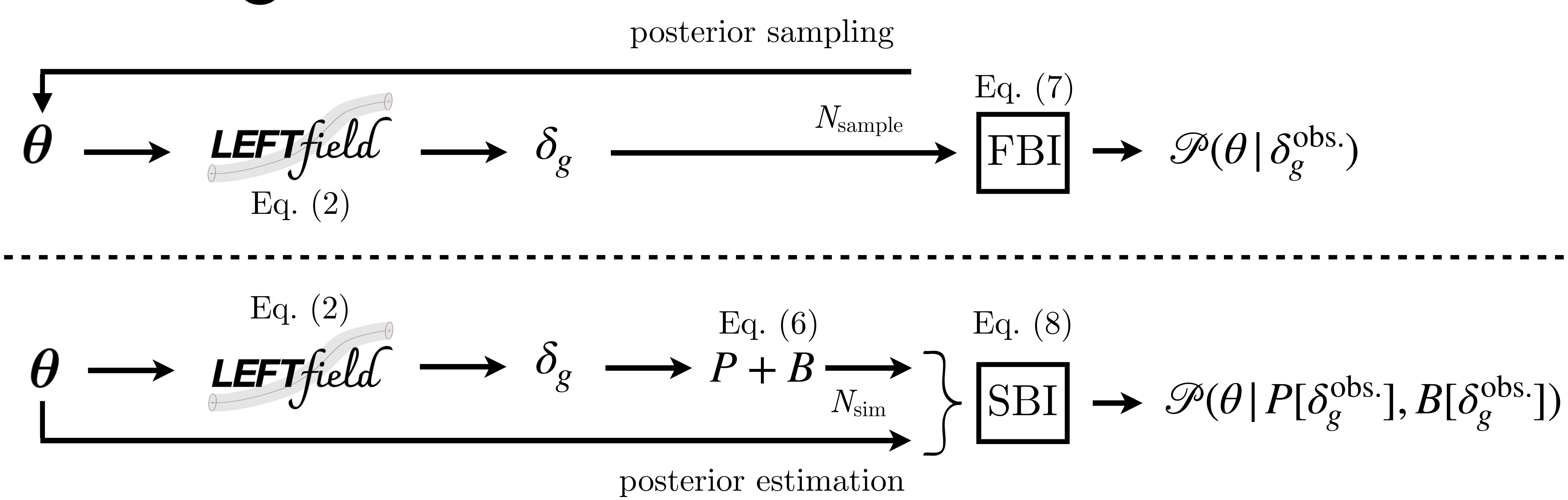
- Scheme:
  - Discretize field on grid/lattice (Nyquist frequency = cutoff Λ)
  - Draw initial conditions from prior
  - Forward-evolve using gravity
  - Evaluate likelihood on data and repeat
- Challenge: even with fairly coarse resolution, have to sample million(s) of parameters
  - Key: Hamiltonian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

Field-level inference: Inferring σ_8 from rest-frame tracers

- First results on field-level σ_8 inference from dark matter halos in real space
  - Marginalizing over bias and stochastic terms
- 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

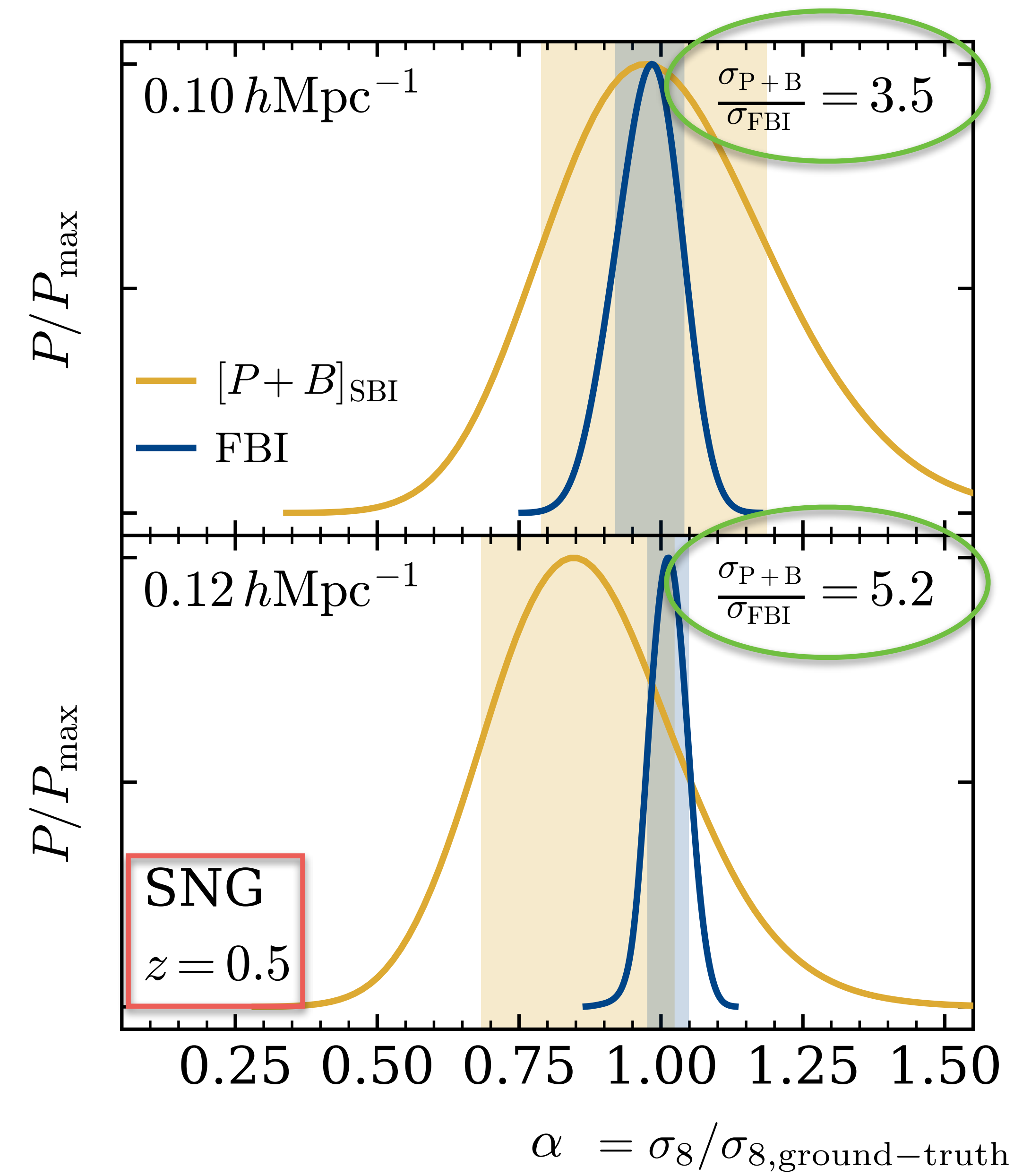


- Idea: compare field-level result with power spectrum + bispectrum using the same forward model and modes of the data
  - Via simulation-based inference (SBI) using the same forward model as in the field-level analysis

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

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

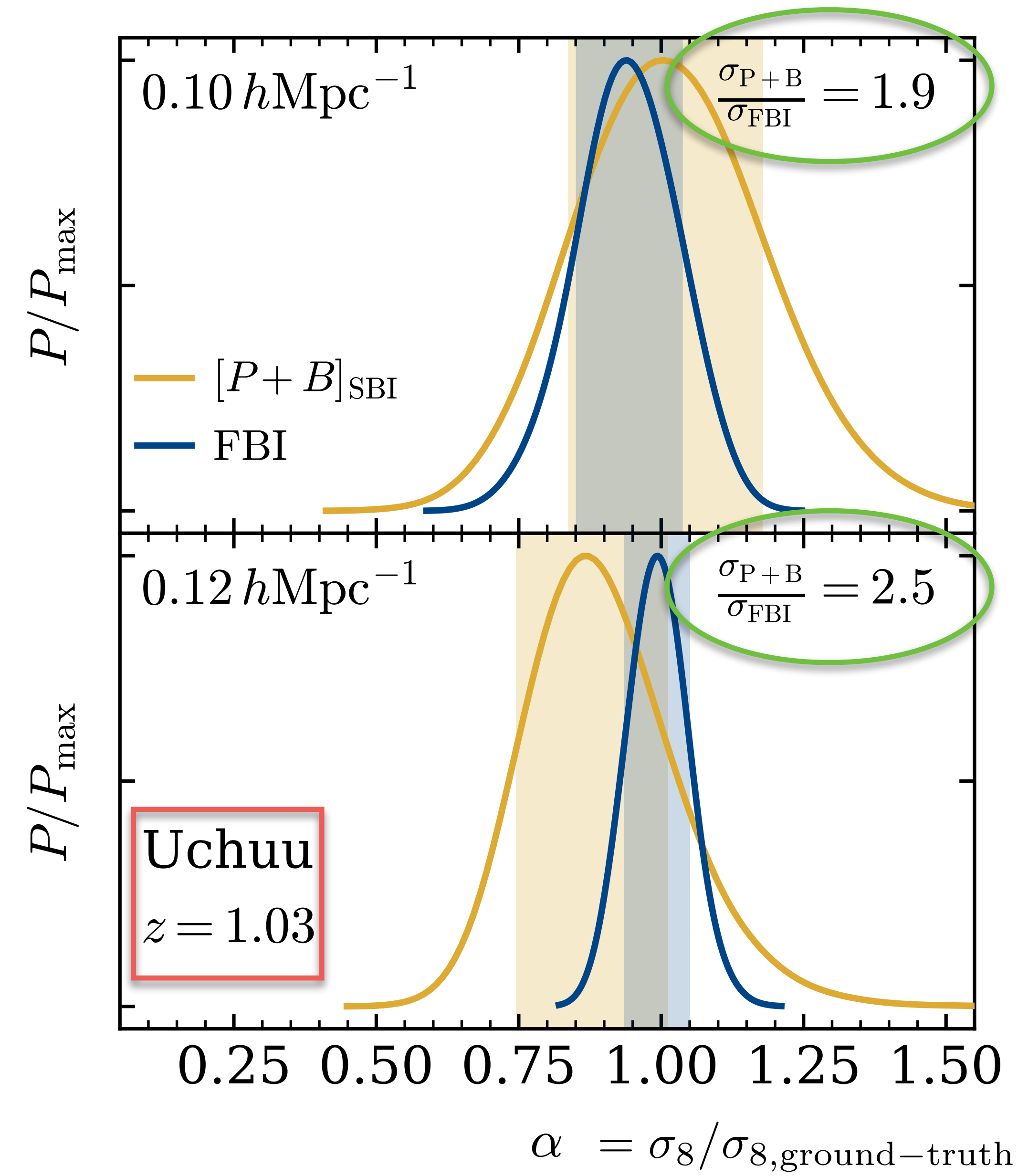
- First results on field-level σ8 inference from dark matter halos in real space
  - Marginalizing over bias and stochastic terms
- Field-level inference vs power spectrum + bispectrum using the same forward model and modes of the data



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

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

- First results on field-level σ_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

Where does the field-level information come from?

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

- Let's consider the zero-noise limit of the field-level posterior, such that likelihood becomes Dirac delta
- We can then formally perform integration over initial conditions $\delta_{\rm in}$ analytically to obtain marginalized posterior:

$$\mathcal{P}(\theta,\{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta^{-1}_{\rm fwd}[\delta_g,\{b_O\}]\middle|\theta\right)\mathcal{J}[\delta_g,\{b_O\}]$$

← Jacobian $|D\delta_{\rm fwd}/D\delta_{\rm in}|^{-1}$

Cabass, Simonović, Zaldarriaga (2023); FS (2025)

Where does the field-level information come from?

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

- Let's consider the zero-noise limit of the field-level posterior, such that likelihood becomes Dirac delta
- We can then formally perform integration over initial conditions $\delta_{\rm in}$ analytically to obtain marginalized posterior:

$$\mathcal{P}(\theta, \{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}]\middle|\theta\right) \mathcal{J}[\delta_g, \{b_O\}]$$

$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}} \frac{|\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\right] \mathcal{J}[\delta_g, \{b_O\}]$$

Jacobian $|D\delta_{\rm fwd}/D\delta_{\rm in}|^{-1}$

Cabass, Simonović, Zaldarriaga (2023); FS (2025)

Where does the field-level information come from?

$$\mathcal{P}(\theta,\{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta_{\rm fwd}^{-1}[\delta_g,\{b_O\}]\middle|\theta\right)\mathcal{J}[\delta_g,\{b_O\}]$$

$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}}\frac{|\delta_{\rm fwd}^{-1}[\delta_g,\{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\right]\mathcal{J}[\delta_g,\{b_O\}]$$

- Involves inverse of forward model, evaluated on the data
- In case of linear forward model, $\delta_{\rm fwd} = b_1\delta_{\rm in}$, marginalized field-level posterior is function of the power spectrum of the data - *$P_g(k)$ is sufficient statistic*

Cabass, Simonović, Zaldarriaga (2023); FS (2025)

Where does the field-level information come from?

$$\mathcal{P}(\theta, \{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}]\middle|\theta\right)\mathcal{J}[\delta_g, \{b_O\}]$$
$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}} \frac{|\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\right]\mathcal{J}[\delta_g, \{b_O\}]$$

- If forward model is nonlinear, $\delta_{\rm fwd}^{-1}$ is a *nonlinear functional of the data* δ_g: effectively, we add higher *n*-point functions to the posterior

FS (2025), arXiv:2504.15351

Where does the field-level information come from?

$$\mathcal{P}(\theta, \{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}]\middle|\theta\right) \mathcal{J}[\delta_g, \{b_O\}]$$

$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}} \frac{|\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\right] \mathcal{J}[\delta_g, \{b_O\}]$$

- If forward model is nonlinear, $\delta_{\rm fwd}^{-1}$ is a *nonlinear functional of the data* δ_g: effectively, we add higher *n*-point functions to the posterior
- Each term in the forward model adds a new, specific statistic to the posterior
  - Complete forward model at 2nd order: power spectrum + bispectrum
  - Complete forward model at 3d order: power spectrum + bispectrum + trispectrum …

FS (2025), arXiv:2504.15351

Where does the field-level information come from?

$$\mathcal{P}(\theta, \{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta^{-1}_{\rm fwd}[\delta_g, \{b_O\}]\Big|\theta\right) \mathcal{J}[\delta_g, \{b_O\}]$$

$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}} \frac{|\delta^{-1}_{\rm fwd}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\right] \mathcal{J}[\delta_g, \{b_O\}]$$

- Each term in the forward model adds a new, specific statistic to the posterior
- Lagrangian, LPT-based forward model as in LEFTfield: *correctly describes displacement terms at all orders, precisely those terms responsible for the degeneracy breaking*
- Impact of missing operators in forward model is proportional to scalar product of missing $O_{\rm missing}[\delta]$ with $O[\delta]$ of interest

FS (2025), arXiv:2504.15351

Cosmology results II: Field-level inference of BAO scale

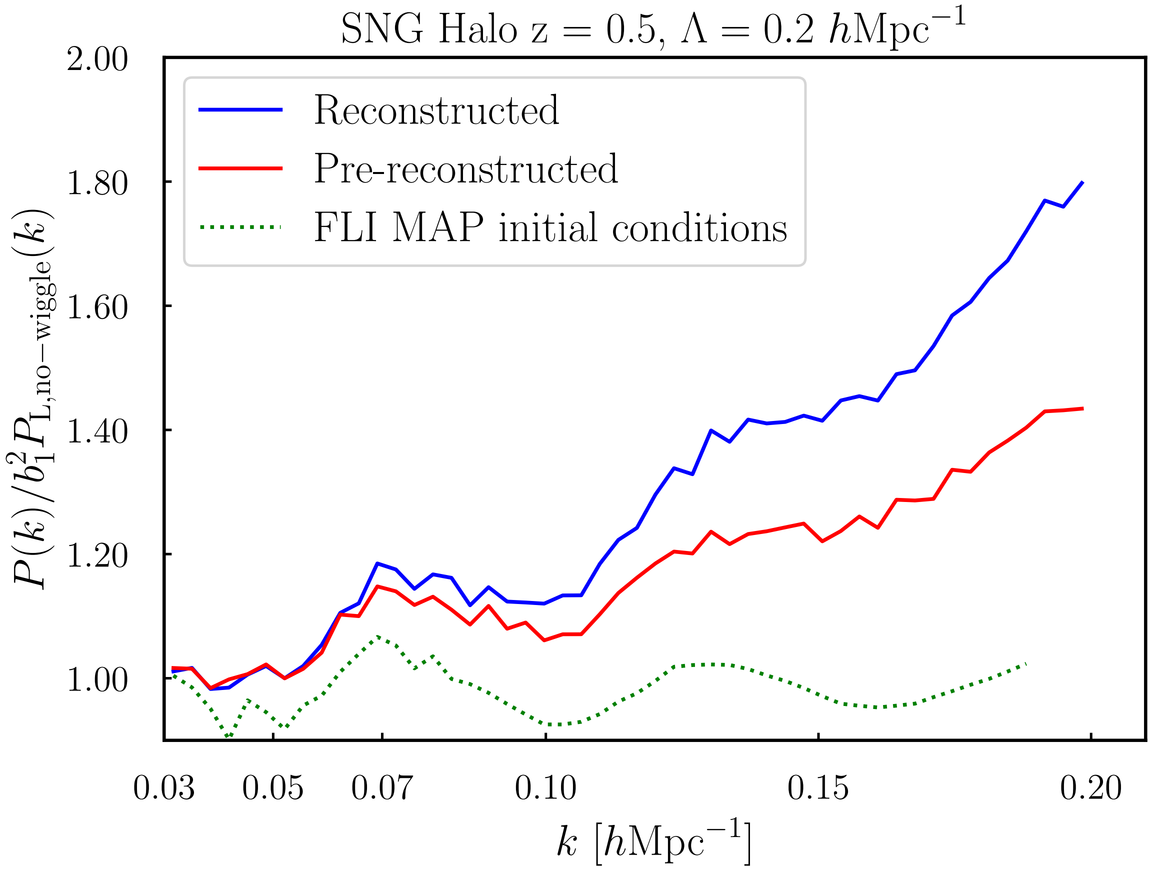
- Constraints on expansion history (dark energy) from galaxy clustering are based on the BAO standard ruler (cf. DESI results)
- These are commonly inferred by performing reconstruction procedure on galaxies, and then using the *post-reconstruction galaxy power spectrum*

Cosmology results II: 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

Cosmology results II: Field-level inference of BAO scale

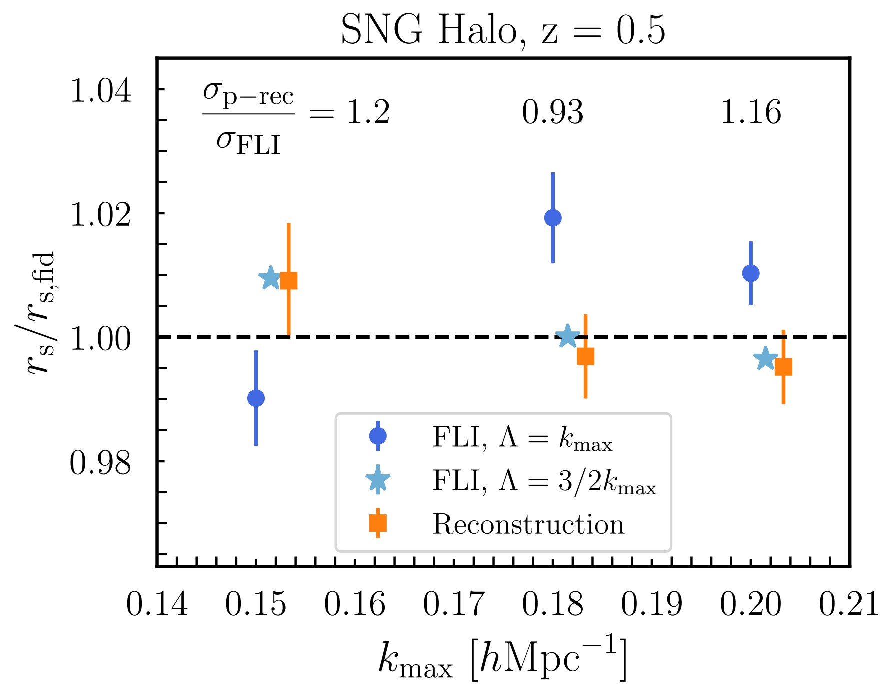
- Field-level inference of BAO scale using a trick: moving BAO feature in linear (initial) density field:

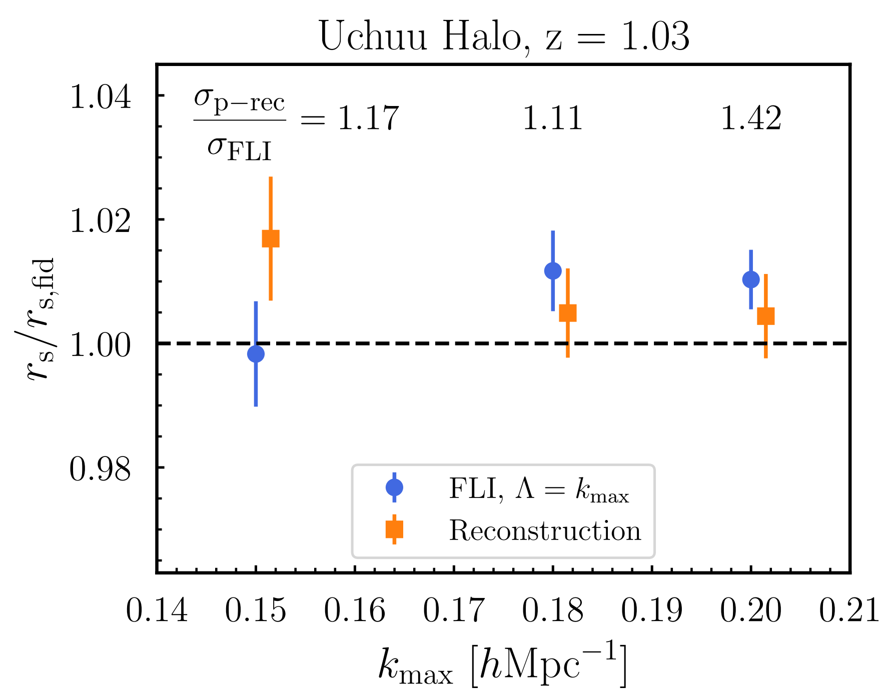
$$f(k, r_s) = \frac{T^2_{\rm BAO}(k|r_s)}{T^2_{\rm BAO}(k|r_{s,\rm fid})}\,,$$

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

- 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

Cosmology results II: Field-level inference of BAO scale

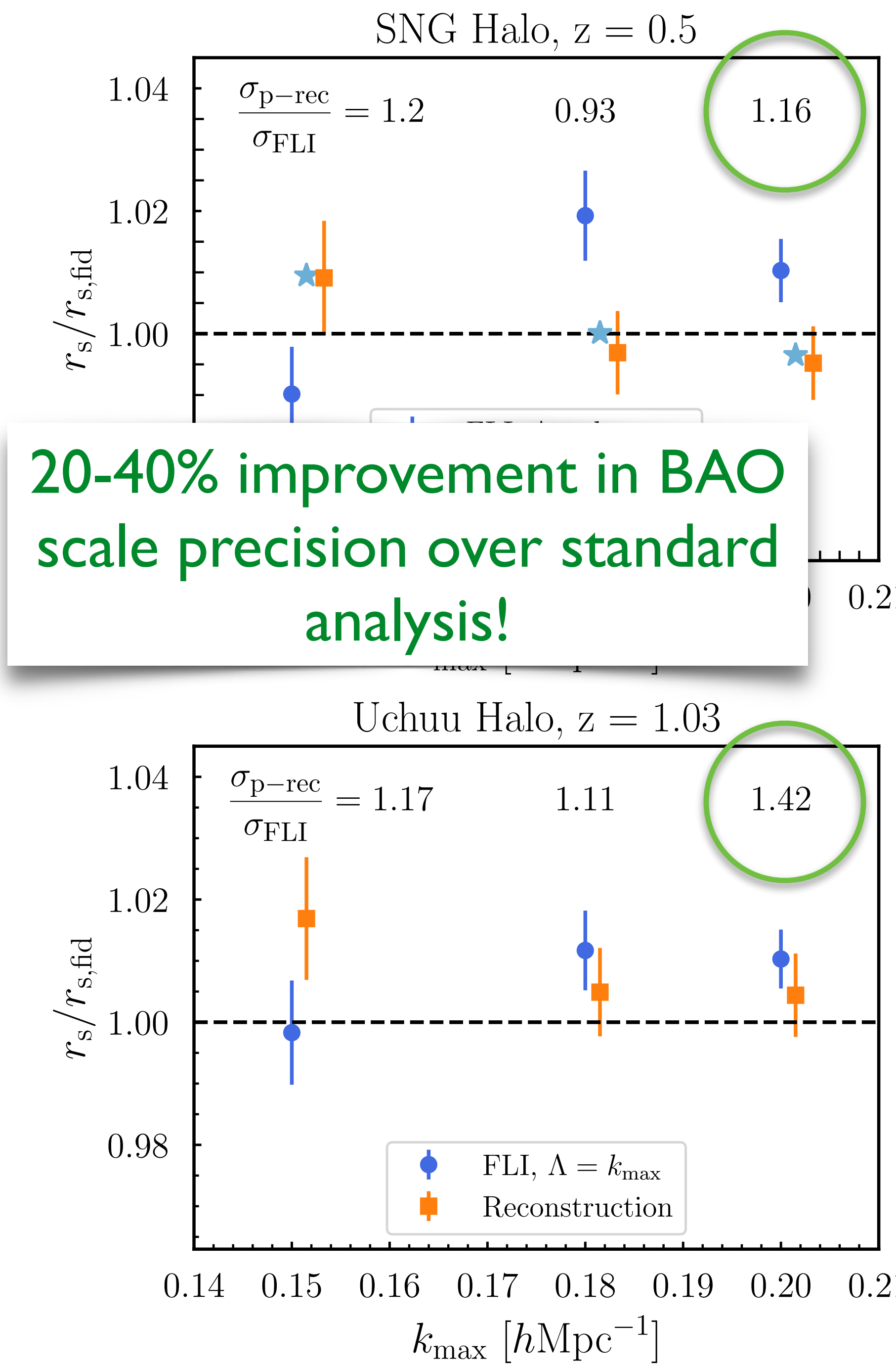
- Field-level inference of BAO scale using a trick: moving BAO feature in linear (initial) density field:

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

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

- 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

Where does the field-level *BAO* information come from?

$$\mathcal{P}(\theta, \{b_O\}|\delta_g) \propto \mathcal{P}_{\text{prior}}\left(\delta_{\text{fwd}}^{-1}[\delta_g, \{b_O\}]\middle|\theta\right)\mathcal{J}[\delta_g, \{b_O\}]$$

$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}} \frac{|\delta_{\text{fwd}}^{-1}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\text{L}}(k|\theta)}\right]\mathcal{J}[\delta_g, \{b_O\}]$$

- In case of *perfect* forward model, δ_{fwd}^{-1} is a sample from prior (Gaussian linear density field) - in fact, information obtained is precisely that contained in linear density field: *optimal inference*
    - Field-level inference "undoes" nonlinear evolution as well as nonlinear bias

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

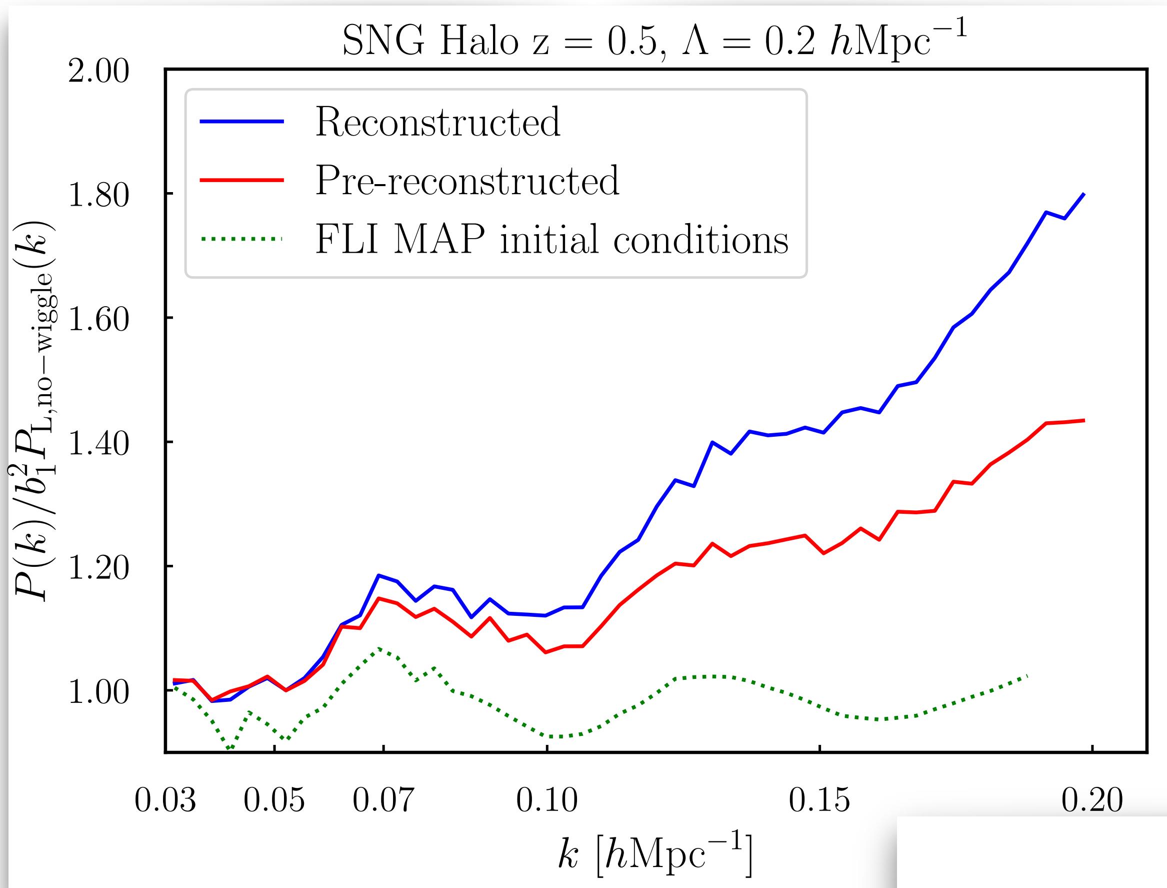
Where does the field-level *BAO* information come from?

$$\mathcal{P}(\theta, \{b_O\}|\delta_g) \propto \mathcal{P}_{\rm prior}\left(\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}]\Big|\theta\right)\mathcal{J}[\delta_g, \{b_O\}]$$
$$\propto \exp\left[-\frac{1}{2}\int_{\boldsymbol{k}} \frac{|\delta_{\rm fwd}^{-1}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\right]\mathcal{J}[\delta_g, \{b_O\}]$$

- In case of *perfect* forward model, $\delta_{\rm fwd}^{-1}$ is a sample from prior (Gaussian linear density field) - in fact, information obtained is precisely that contained in linear density field: *optimal inference*
  - Field-level inference "undoes" nonlinear evolution as well as nonlinear bias
- On the other hand, standard BAO reconstruction leaves substantial broad-band contribution to $\delta_g^{\rm post\text{-}rec}$; this explains information gain found at field level
- Cannot easily be recuperated using higher-order n-pt functions

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

Where does the field-level *BAO* information come from?



$$\ldots{}^{-1}_{\rm vd}[\delta_g, \{b_O\}]\Big|\theta\Big)\, \mathcal{J}[\delta_g, \{b_O\}]$$

$$\ldots\int_{\boldsymbol{k}} \frac{|\delta^{-1}_{\rm fwd}[\delta_g, \{b_O\}](\boldsymbol{k})|^2}{P_{\rm L}(k|\theta)}\Big]\, \mathcal{J}[\delta_g, \{b_O\}]$$

$\delta_{\rm fwd}^{-1}$ is a sample from prior (Gaussian

…mation obtained is precisely that contained

…*rence*

… nonlinear evolution as well as nonlinear

- On the other hand, stan… band contribution to δ_g… level
- Cannot easily be recupe…

$$F^{\rm FLI}_{r_s r_s} = -\left\langle \frac{\partial^2}{\partial r_s^2} \ln \mathcal{P}_{\rm FLI}[\{b_O\}, r_s|\delta_g] \right\rangle = \frac{1}{2}\sum_{\boldsymbol{k}}^{\Lambda} \frac{1}{[P_{\rm L}(k|r_{s,\rm fid})]^2} \left(\frac{\partial P_{\rm L}(k|r_{s,\rm fid})}{\partial r_{s,\rm fid}}\right)^2$$

$$F^{\rm rec-P(k)}_{r_s r_s} = -\left\langle \frac{\partial^2}{\partial r_s^2} \ln \mathcal{P}_{\rm rec-P(k)}[r_s|\delta_g] \right\rangle = \sum_{\boldsymbol{k}}^{\Lambda} \frac{1}{{\rm Var}[P_{\rm p\text{-}rec}(k|r_{s,\rm fid})]} \left(\frac{\partial P_{\rm p\text{-}rec}(k|r_s)}{\partial r_s}\right)^2$$
$$= \frac{1}{2}\sum_{\boldsymbol{k}}^{\Lambda} \frac{1}{[P_{\rm p\text{-}rec}(k|r_{s,\rm fid})]^2} \left(\frac{\partial P_{\rm p\text{-}rec}(k|r_s)}{\partial r_s}\right)^2 .$$

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

Why field-level?

- Evaluating the full posterior guarantees optimality in the context of the given forward model
- One can certainly hope to approach this optimum closely with suitably engineered summary statistics (i.e. *data compression*)
    - Calling that "field-level" does not seem to make sense however…
- Advantages of (actual) field-level inference apart from optimality:
    - Maximally interpretable: have access to all physically relevant variables
    - Allows for broad range of systematics checks (e.g. cross-correlating predicted mean field with systematics maps)
    - Many possibilities for ancillary science: cross-correlation with other tracers, shear, CMB lensing

Conclusions

- By now have a robust framework to predict galaxy clustering on large scales within GR
- Even after many years we are continuing to find new signals to search for in LSS
  - Light thermal relics
  - Spinning particles
  - Primordial parity violation
  - …
- Inference/analysis methods have made tremendous progress — now need to tie the two together