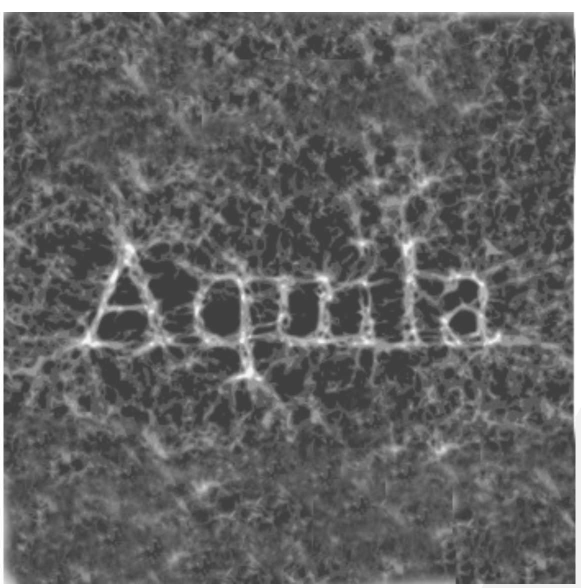
MAX-PLANCK-INSTITUT
FÜR ASTROPHYSIK

$\mathbf{x} = \mathbf{x}_{\rm fl}(\tau)$

Field-level Inference & PNG from galaxy clustering



aquila-consortium.org

Fabian Schmidt
MPA

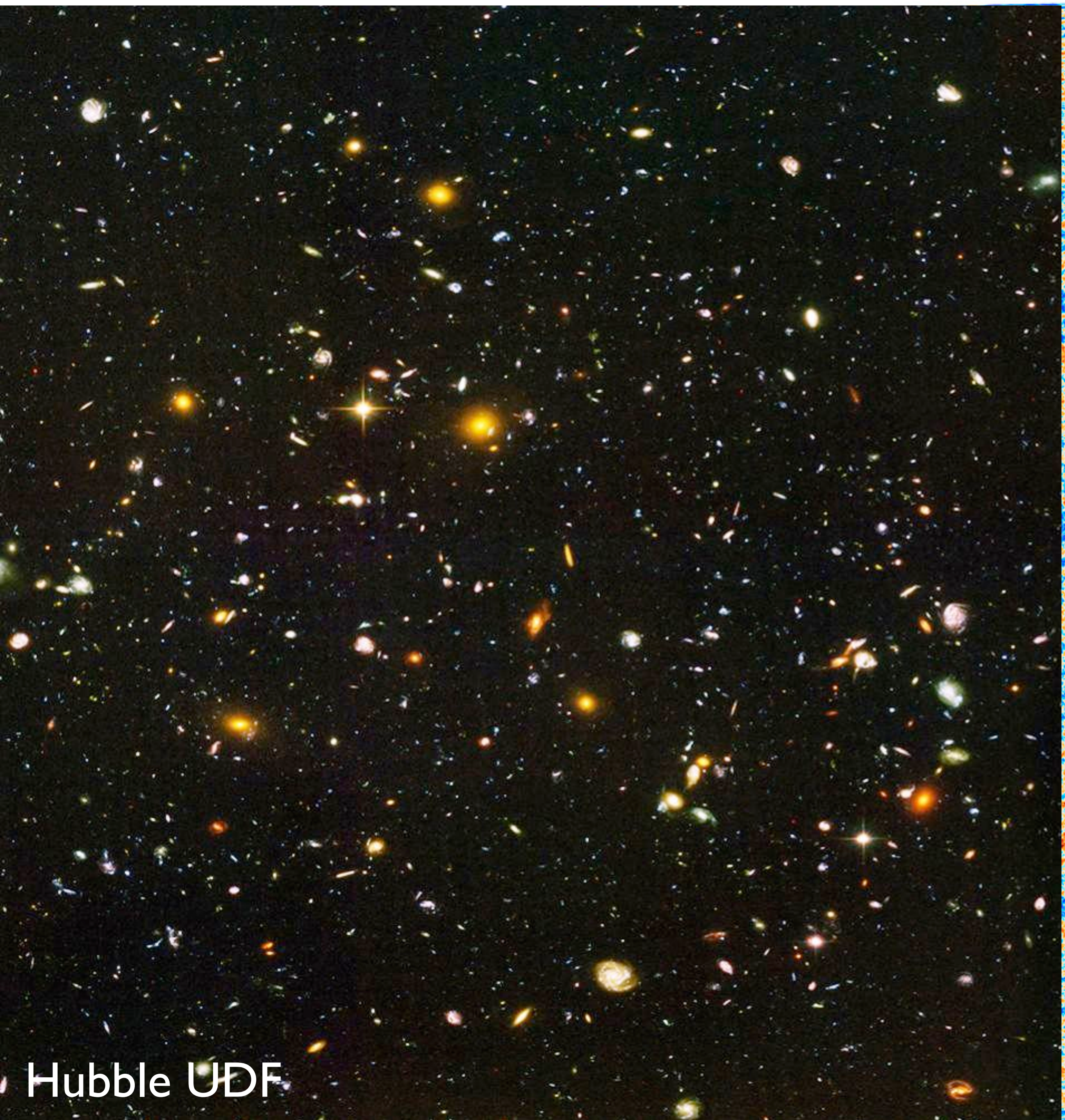
European Research Council
Established by the European Commission

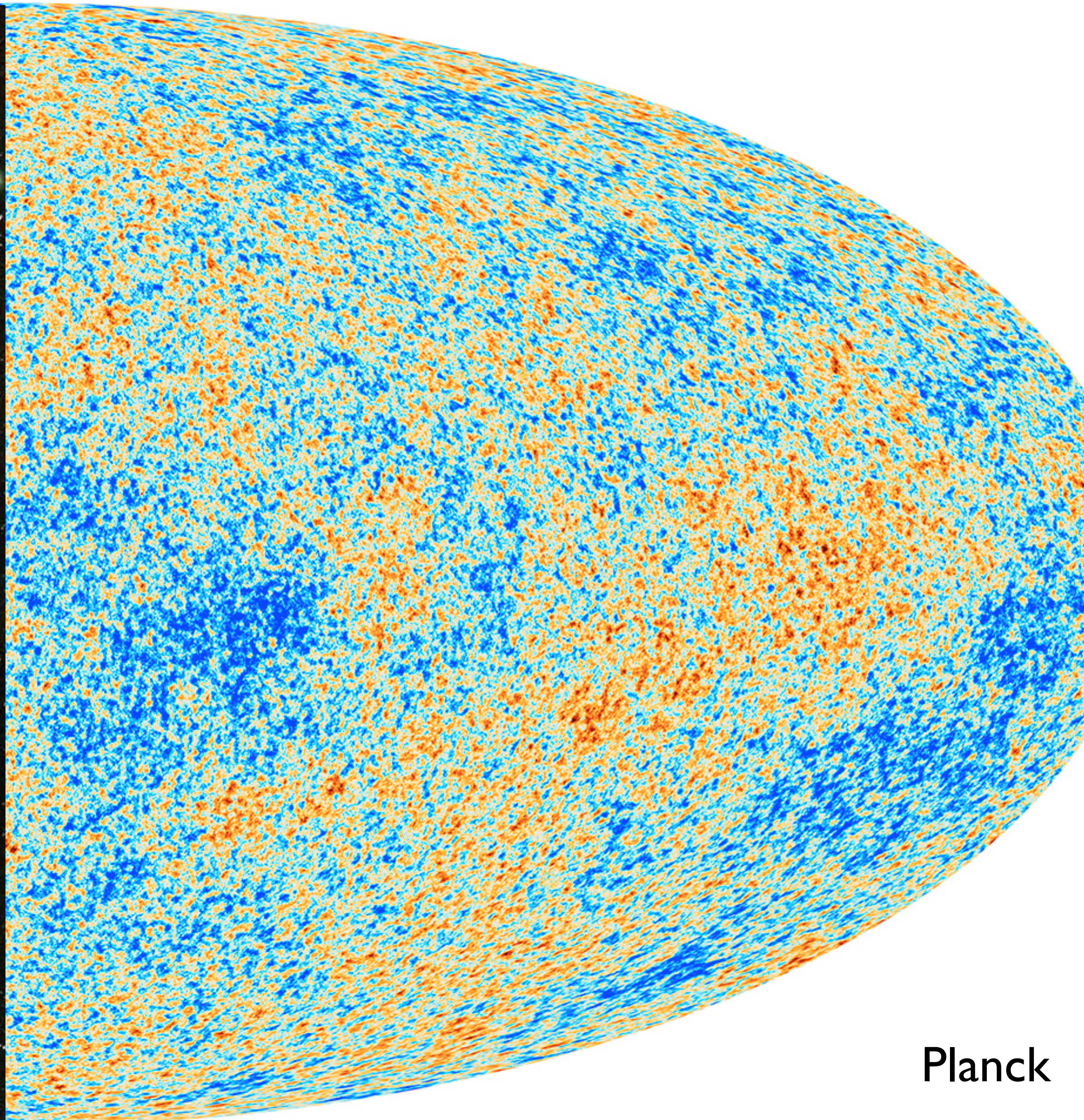
with
Giovanni Cabass, Franz Elsner, Jens Jasche,
Andrija Kostić, Guilhem Lavaux, Minh Nguyen,
Titouan Lazeyras, Alex Barreira

$\mathbf{q} = \mathbf{x}_{\rm fl}(0)$

PNG Conference, IFT Madrid, Sep 21, 2022

Main Challenge to unlocking primordial physics using LSS: unlike the CMB, every data point is nonlinear!





Outline

- Two goals of the talk:
  - Show that we can deal with complexities of galaxies* *rigorously* on large scales: **EFT of LSS**
  - Argue that there is *much more (trustable) information in galaxy clustering* than what we are using so far: **field-level Bayesian inference**

* Everything in following will apply to any tracer of LSS: clusters, Ly-α forest, intensity mapping, …

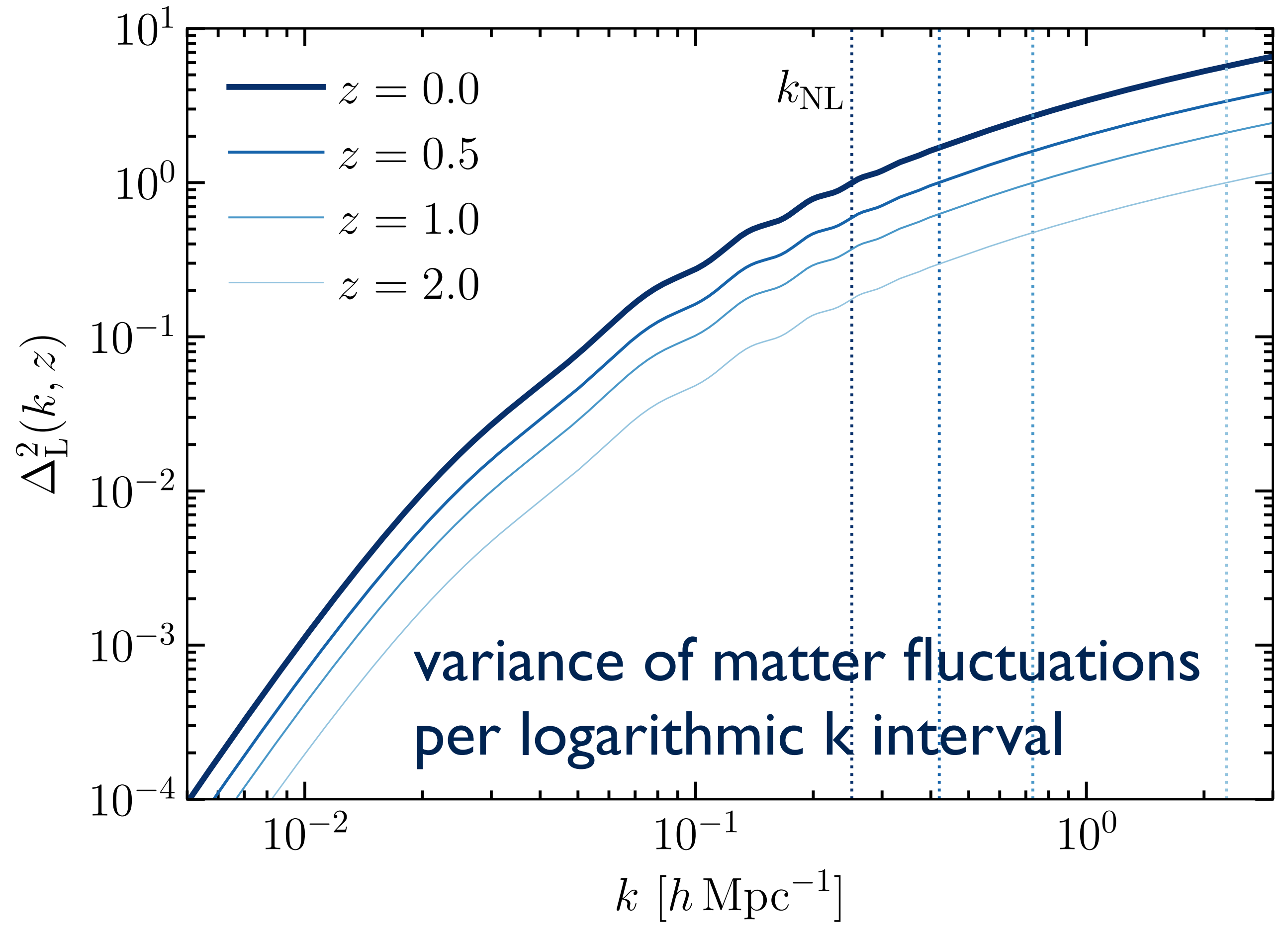
Outline

- Two goals of the talk:
  - Show that we can deal with complexities of galaxies* *rigorously* on large scales: **EFT of LSS**
  - Argue that there is *much more (trustable) information in galaxy clustering* than what we are using so far: **field-level Bayesian inference**

* Everything in following will apply to any tracer of LSS: clusters, Ly-α forest, intensity mapping, …

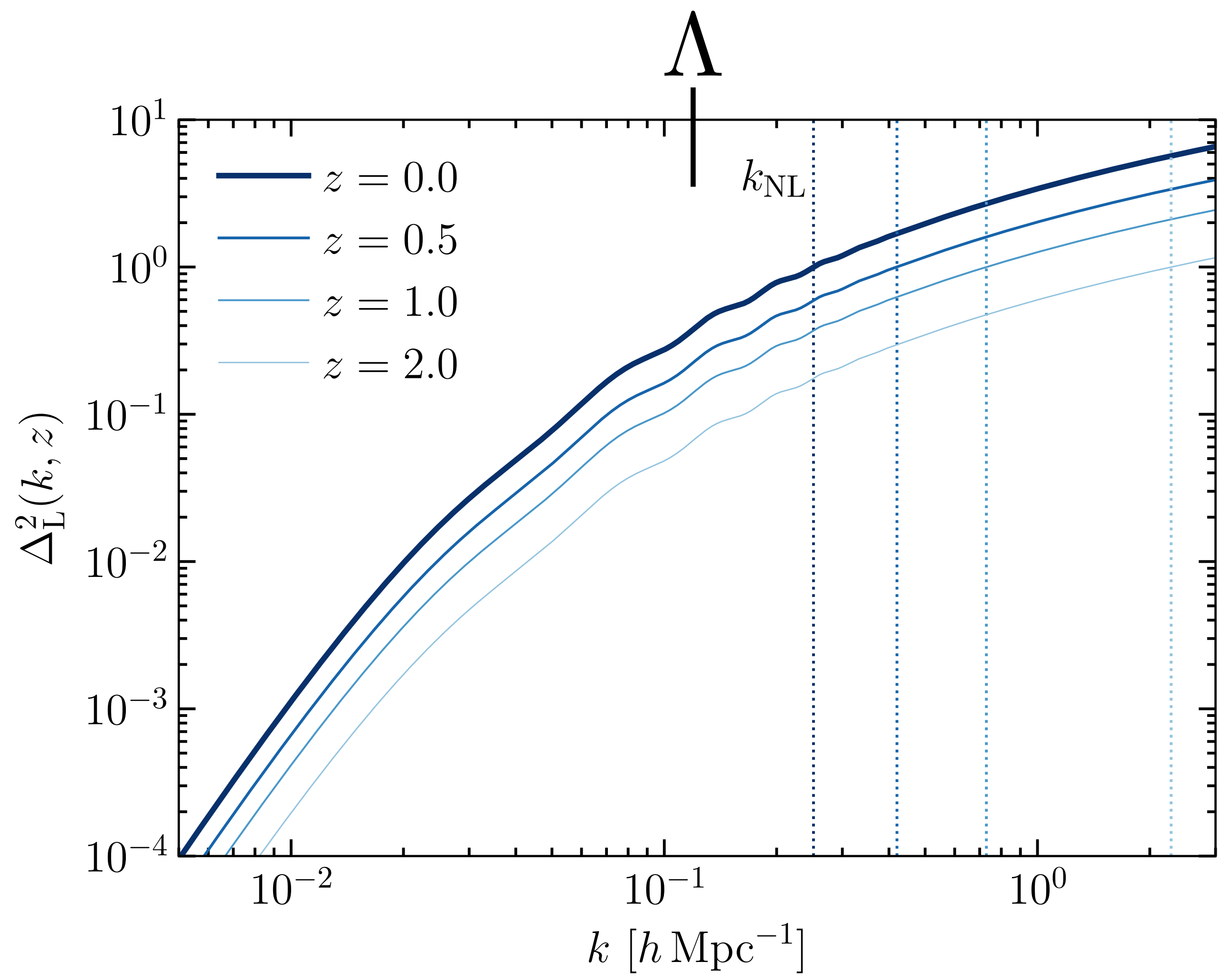
Theory of galaxy clustering

- Perturbations in our universe are small on large scales
  - Perturbation theory works on quasilinear scales $k < k_{NL}$
- Goal: describe galaxy clustering up to a given scale and accuracy using a finite number of free (A) bias parameters and (B) stochastic amplitudes



EFT approach

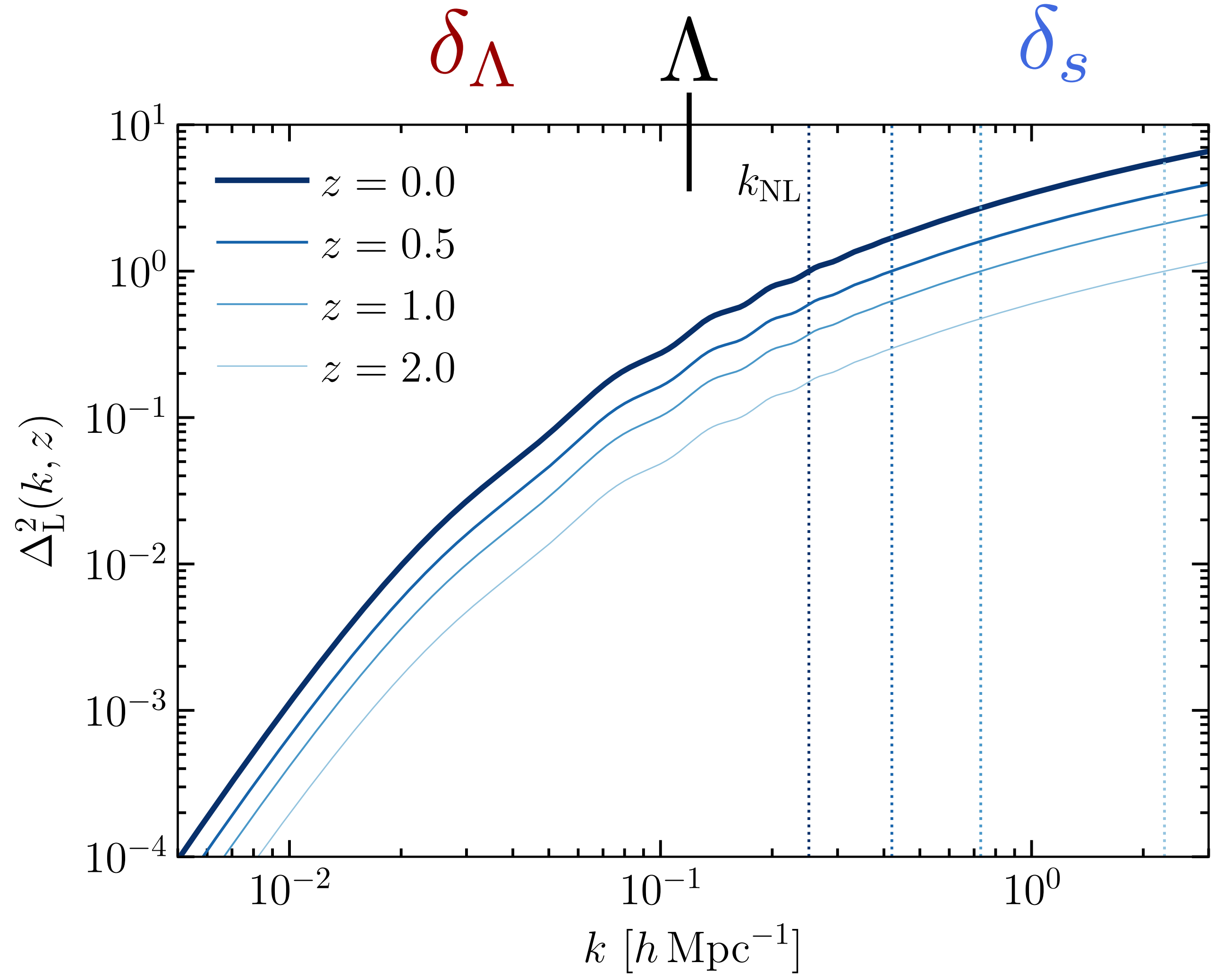
- Idea: trust our theory for $k < \Lambda$



EFT approach

- Idea: trust our theory for *k* < *Λ*
- Split *initial* perturbations into large scale (< Λ) and small scale (>= Λ):

$$\delta(\boldsymbol{x},\tau) \equiv \frac{\rho_m(\boldsymbol{x},\tau)}{\bar{\rho}_,(\tau)} - 1 = \delta_\Lambda + \delta_s$$

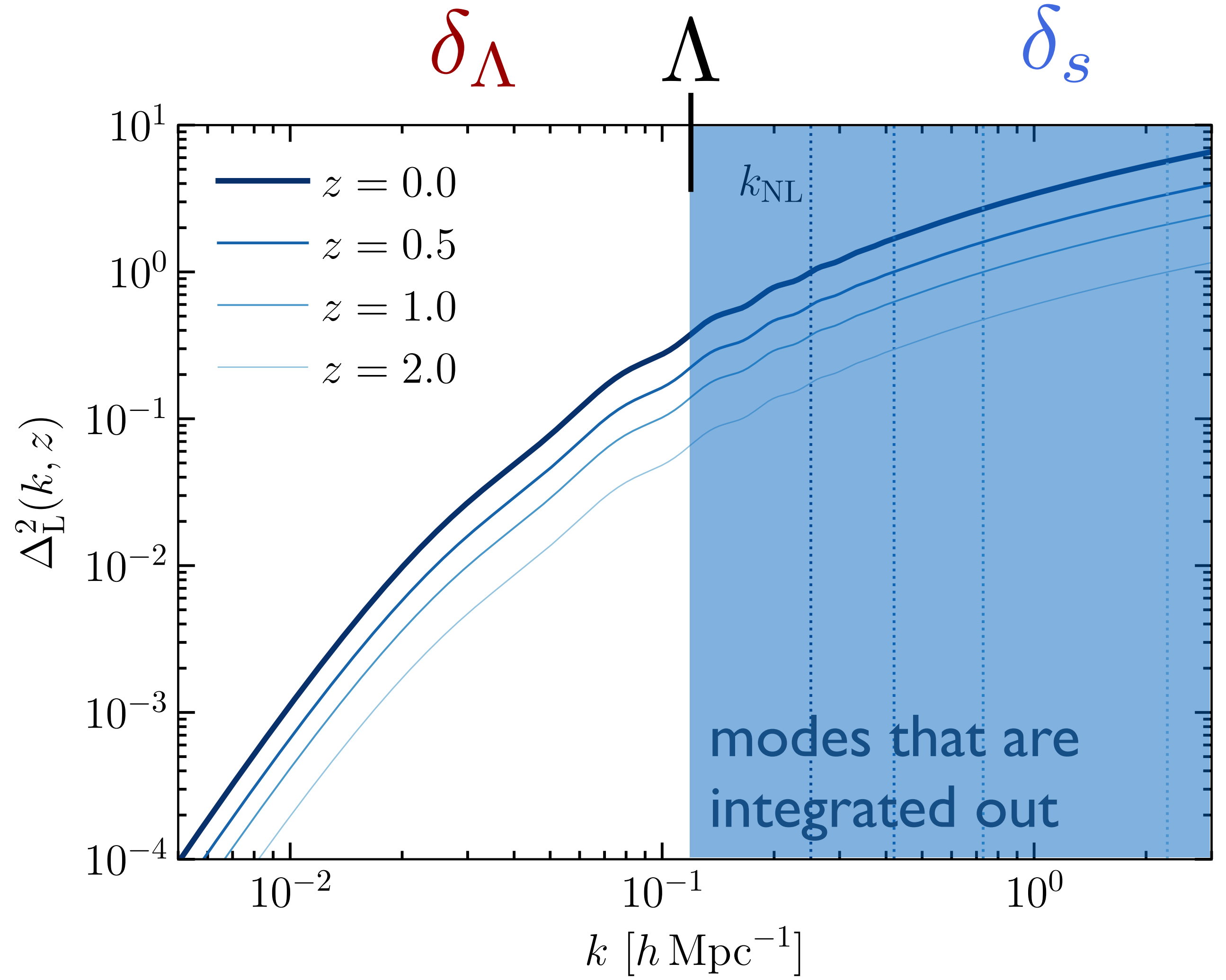


EFT approach

- Idea: trust our theory for $k < \Lambda$
- Split *initial* perturbations into large scale (< Λ) and small scale (>= Λ):

$$\delta(\boldsymbol{x},\tau) \equiv \frac{\rho_m(\boldsymbol{x},\tau)}{\bar{\rho}_,(\tau)} - 1 = \delta_\Lambda + \delta_s$$

- Then, we integrate out (marginalize over) perturbations with k > Λ

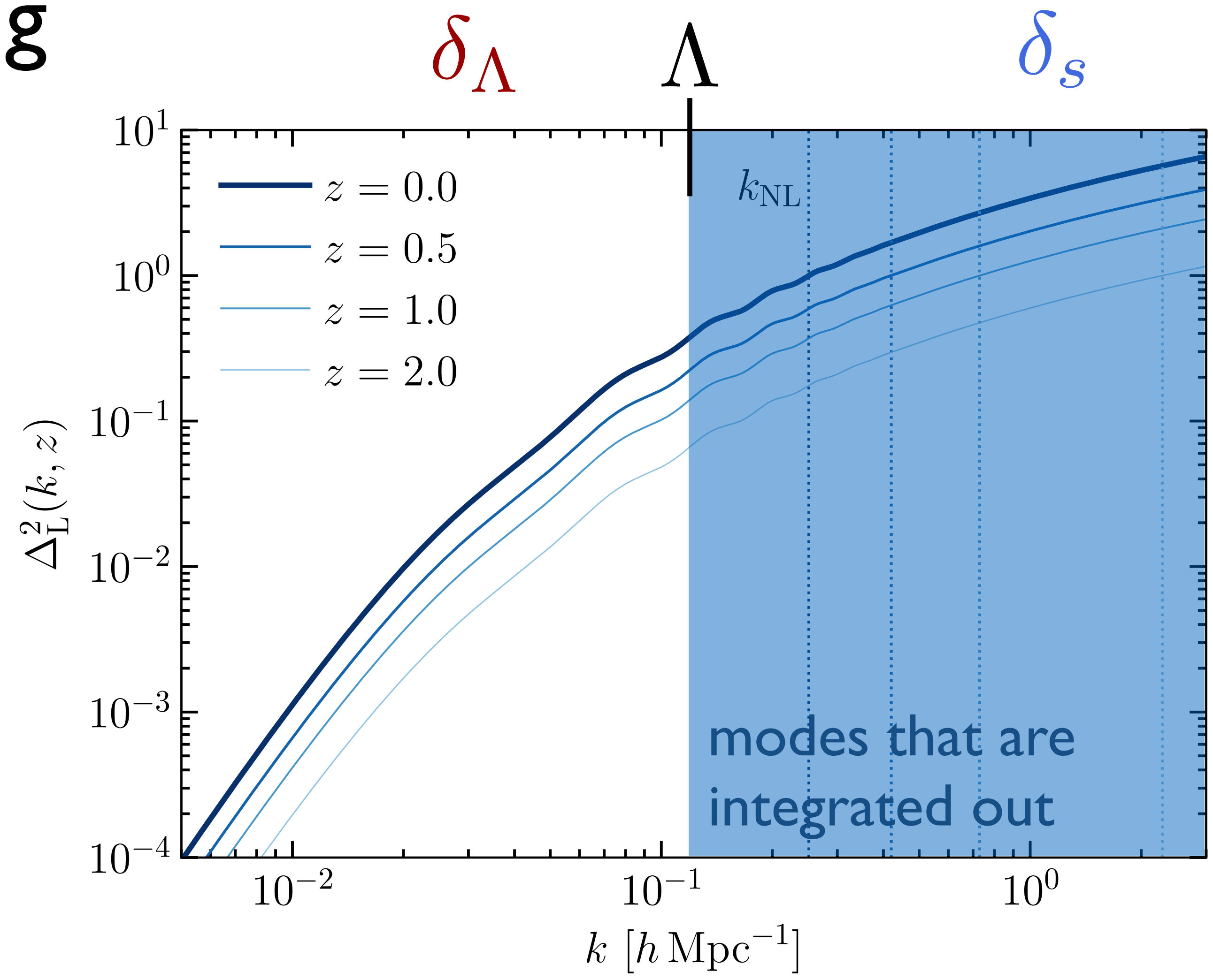


Bias and Stochasticity

- Incorporate effect of large-scale perturbations explicitly using bias expansion, with free coefficients b_O

$$\delta_g(\boldsymbol{x}) = \sum_O b_O O(\boldsymbol{x})$$

  - Fields O are constructed from δ_Λ



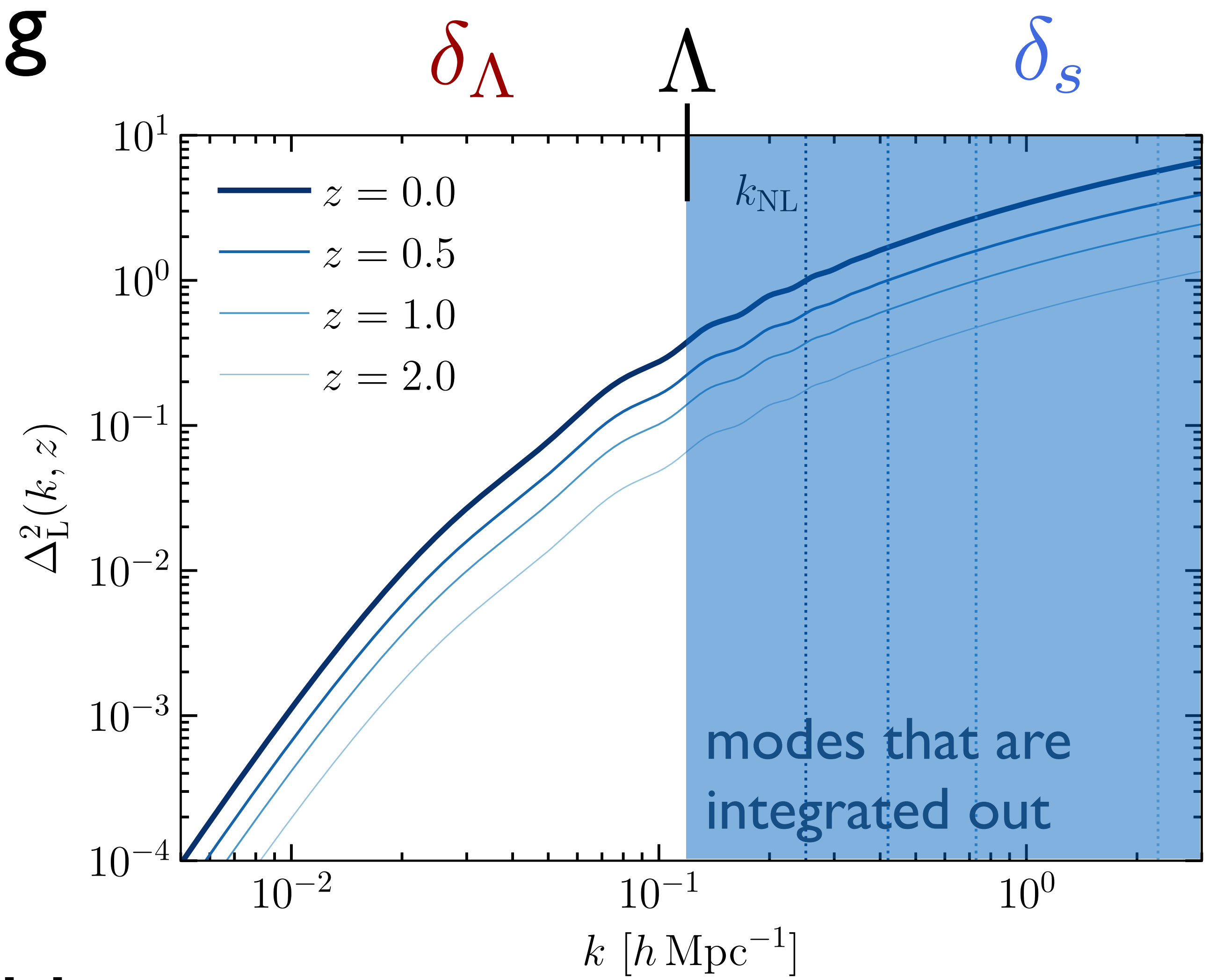
Bias and Stochasticity

- Incorporate effect of large-scale perturbations explicitly using bias expansion, with free coefficients b_O

$$\delta_g(\boldsymbol{x}) = \sum_O b_O O(\boldsymbol{x}) + \varepsilon(\boldsymbol{x})$$

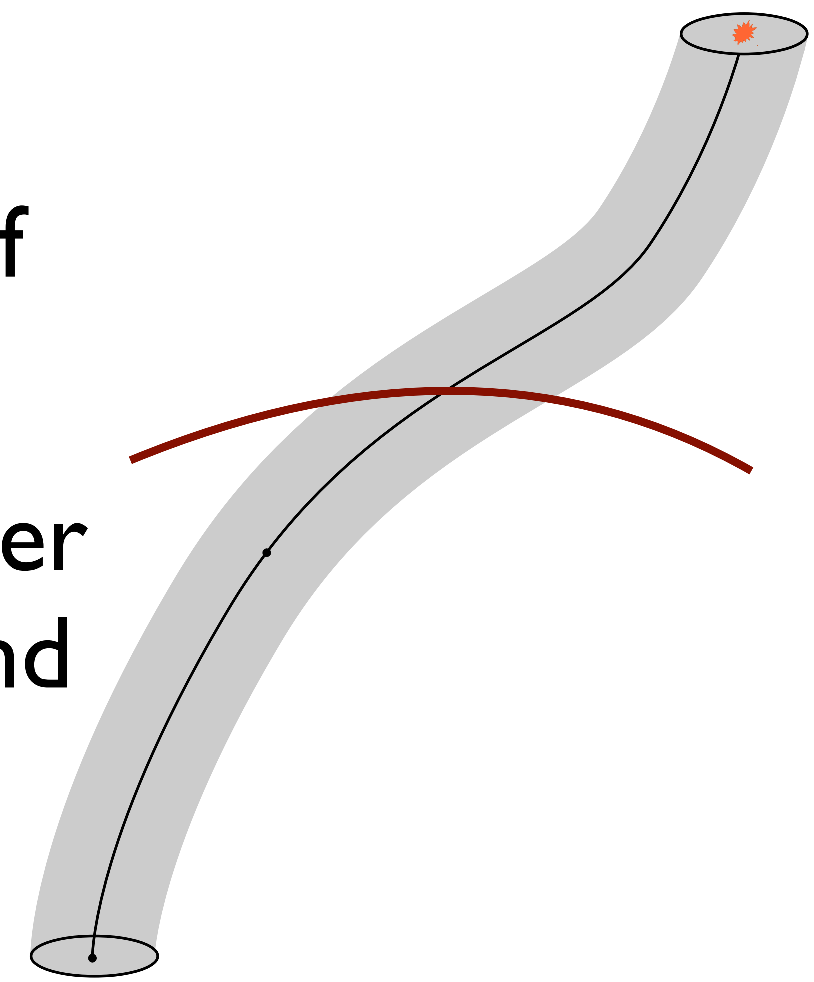
  - Fields O are constructed from δ_Λ

- Small-scale perturbations add noise ε



(A) Bias

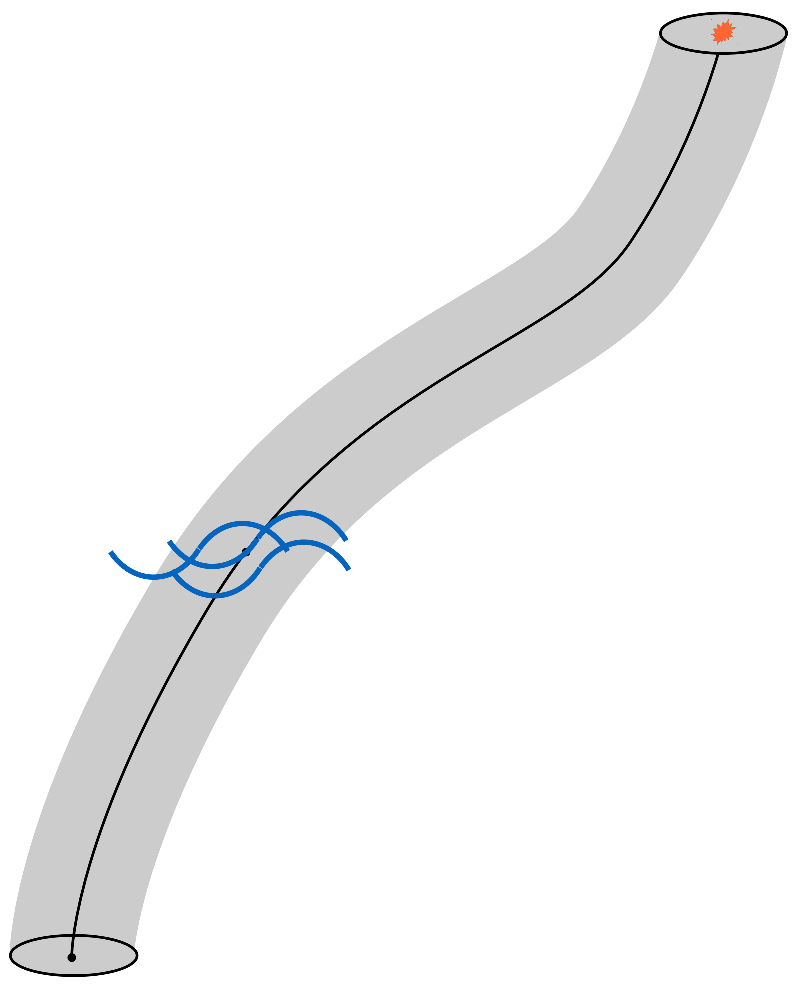
- Which bias terms $O(x)$ we need to include:
  - Well understood by now
  - Include dependence on full history of structure formation
  - Includes "local bias" (powers of matter density) as well as tidal fields, time and space derivatives thereof
- Displacement terms protected by equivalence principle have *fixed* coefficients!



Desjacques, Jeong, FS; Phys. Rept. (2018)

(B) Stochasticity

- ε arises from local (in real space) superposition of many small-scale perturbations
- Central limit theorem: ε(k) is approximately Gaussian distributed (the lower k, the more Gaussian it is)
- Local in real space: power spectrum is white noise at low k, with corrections* ~k^2:



* Also density dependence.

$$\langle \varepsilon(\boldsymbol{k})\varepsilon^*(\boldsymbol{k}')\rangle = (2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}')\left[P_\varepsilon + k^2 P_\varepsilon^{\{2\}} + \cdots\right]$$

Desjacques, Jeong, FS; Phys. Rept. (2018)

Outline

- Two goals of the talk:
  - Show that we can deal with complexities of galaxies* *rigorously* on large scales: **EFT of LSS**
  - Argue that there is *much more (trustable) information in galaxy clustering* than what we are using so far: **field-level Bayesian inference**

* Everything in following will apply to any tracer of LSS: clusters, Ly-α forest, intensity mapping, …

A broad view of cosmology inference

- Given cosmological parameters θ, we can hope to predict

1. Statistics of initial conditions
2. How a given $\delta_{\rm in}(x)$ evolves into the final density field

A broad view of cosmology inference

- Given cosmological parameters θ, we can hope to predict

1. Statistics of initial conditions

   Prior $P_{\rm prior}\left(\vec{\delta}_{\rm in}, \theta\right)$
   PNG enters here.

2. How a given $\delta_{\rm in}(x)$ evolves into the final density field

A broad view of cosmology inference

- Given cosmological parameters θ, we can hope to predict

1. Statistics of initial conditions

   Prior $P_{\rm prior}\left(\vec{\delta}_{\rm in}, \theta\right)$

   PNG enters here.

2. How a given $\delta_{\rm in}(x)$ evolves into the final density field

   deterministic evolution

   $\vec{\delta}_{\rm fwd}[\vec{\delta}_{\rm in}, \theta]$

Bayesian cosmology inference

- The *full posterior of cosmological parameters given the data* is then given by

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

Bayesian cosmology inference

- The *full posterior of cosmological parameters given the data* is then given by

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

Multivariate Gaussian, diagonal covariance in Fourier space, *plus nonlinear operation generating PNG*

Bayesian cosmology inference

- The *full posterior of cosmological parameters given the data* is then given by

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

$$= \int d\{b_O\} P\left(\vec{\delta}_g \middle| \vec{\delta}, \theta_i, \{b_O\}\right)$$

conditional probability of galaxy density given matter density
- contains all physics of galaxy formation

Bayesian cosmology inference

- The *full posterior of cosmological parameters given the data* is then given by

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

Functional integral

Bayesian cosmology inference

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

- Standard approach proceeds via *data compression*: replace galaxy density field with much smaller data vector (e.g., power spectrum in bins of k)

Bayesian cosmology inference

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

- Standard approach proceeds via *data compression*: replace galaxy density field with much smaller data vector (e.g., power spectrum in bins of k)
- Then, the functional integral over initial conditions (a.k.a taking ensemble average) is done either
  - semi-analytically (loop integrals in PT / EFT approach - formally, sending Λ to infinity)
  - numerically (emulators based on ensemble of simulations)

Bayesian cosmology inference

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

- Can we make progress *without* this (lossy) data compression?

Inference beyond the power spectrum

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

- Yes - basically by doing a Markov Chain Monte Carlo:

Inference beyond the power spectrum

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

- Yes - basically by doing a Markov Chain Monte Carlo:
  - Discretize field on grid/lattice (Nyquist frequency = cutoff Λ)
  - Draw initial conditions from prior
  - Forward-evolve using gravity
  - Compare with data and repeat

Inference beyond the power spectrum

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

- Yes - basically by doing a Markov Chain Monte Carlo:
  - Discretize field on grid/lattice (Nyquist frequency = cutoff Λ)
  - Draw initial conditions from prior
  - Forward-evolve using gravity
  - Compare with data and repeat
- Challenge: even with fairly coarse resolution, have to sample millions of parameters in $\vec{\delta}_{\rm in}$
  - Key: Hamiltonian Monte Carlo

Inference beyond the power spectrum

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

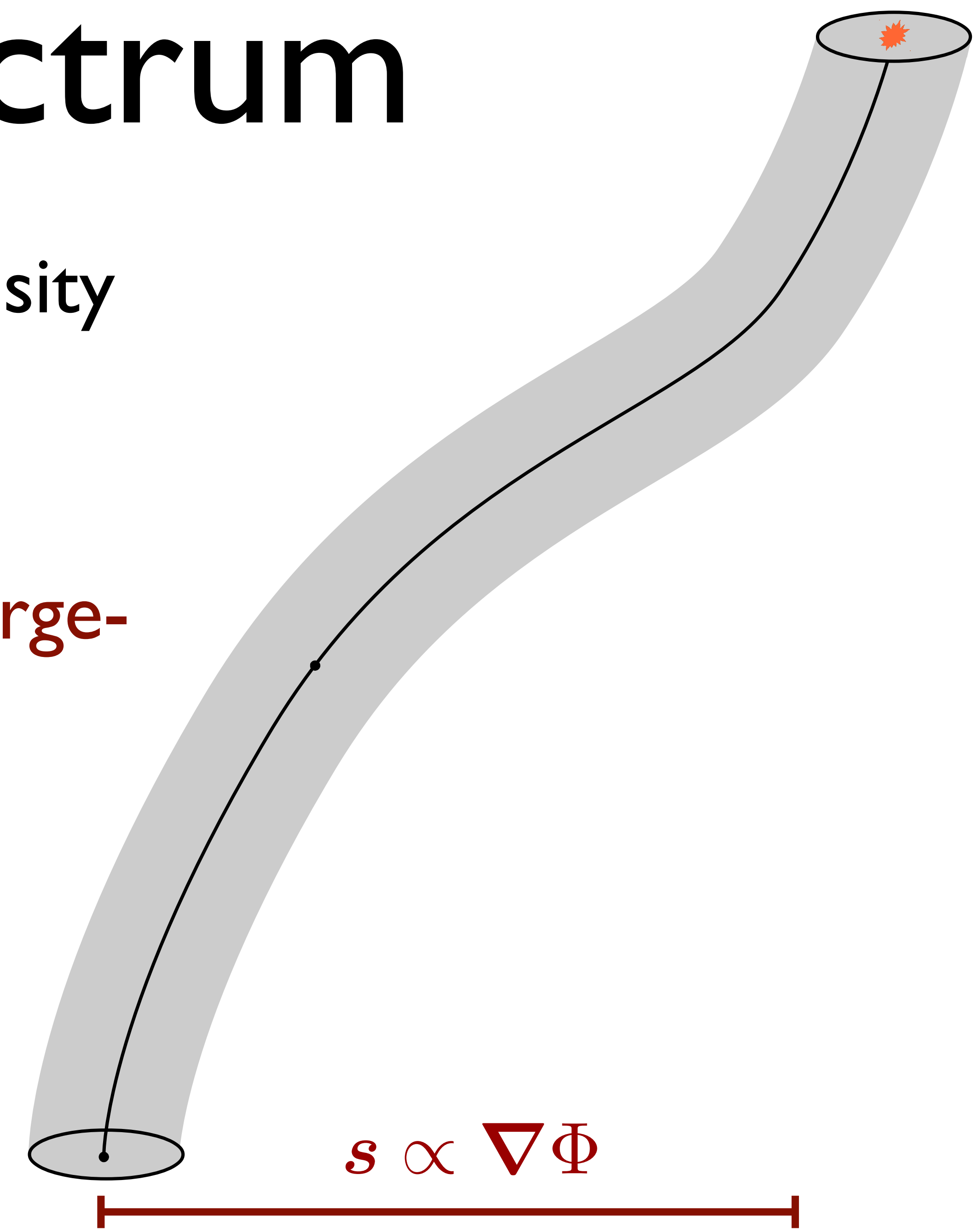
- Yes - basically by doing a Markov Chain Monte Carlo:
  - Discretize field on grid/lattice (Nyquist frequency = cutoff Λ)
  - Draw initial conditions from prior
  - Forward-evolve using gravity
  - Compare with data and repeat

- Lots of interest in this approach recently -> Adam Andrews' talk

Kitaura & Ensslin, Jasche & Wandelt, Wang, Mo et al, Seljak et al, Jasche & Lavaux (2017), …

Why we *should* go beyond the power spectrum

- At second and higher order, galaxy density contains displacement terms which are special:
  - Equivalence principle ensures that large-scale displacement is the same for galaxies and matter
  - Displacement term allows for disentangling bias and amplitude of fluctuations ($\mathcal{A}_s$ or σ_8)
- In the power spectrum, these are mixed in with other nonlinear bias contributions and impossible to disentangle



The galaxy likelihood

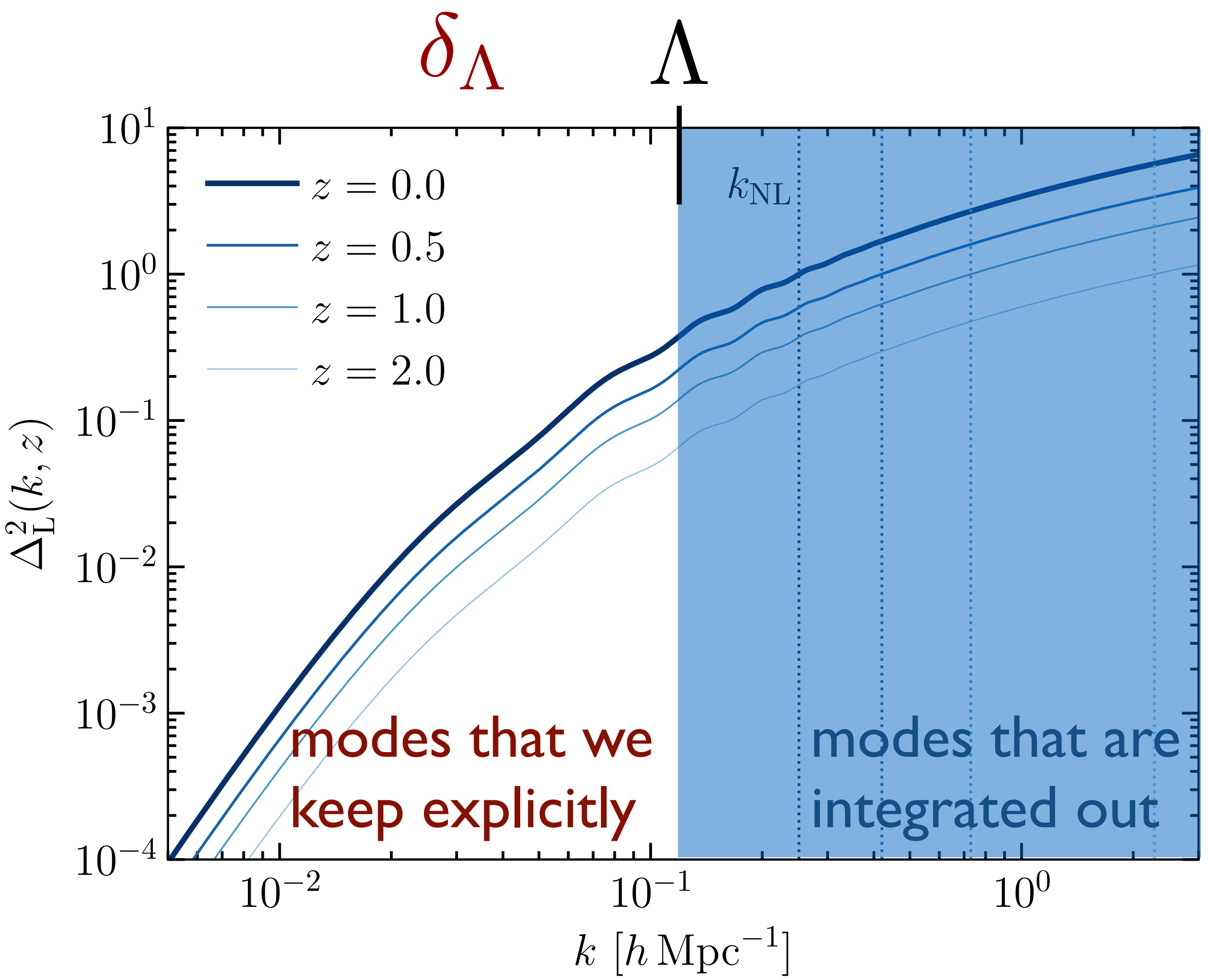
- Putting numerical challenges aside, we need an expression for the *field-level galaxy likelihood:*
  - conditional probability of galaxy density given matter density

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

$$= \int d\{b_O\} P\left(\vec{\delta}_g \middle| \vec{\delta}, \theta_i, \{b_O\}\right)$$

An EFT approach to the likelihood

- Goal is to derive $P\left(\vec{\delta}_g \middle| \vec{\delta}\right)$ in EFT approach
- Recall: split perturbations into large scale (δ_Λ) and small scale, using sharp-k filter*

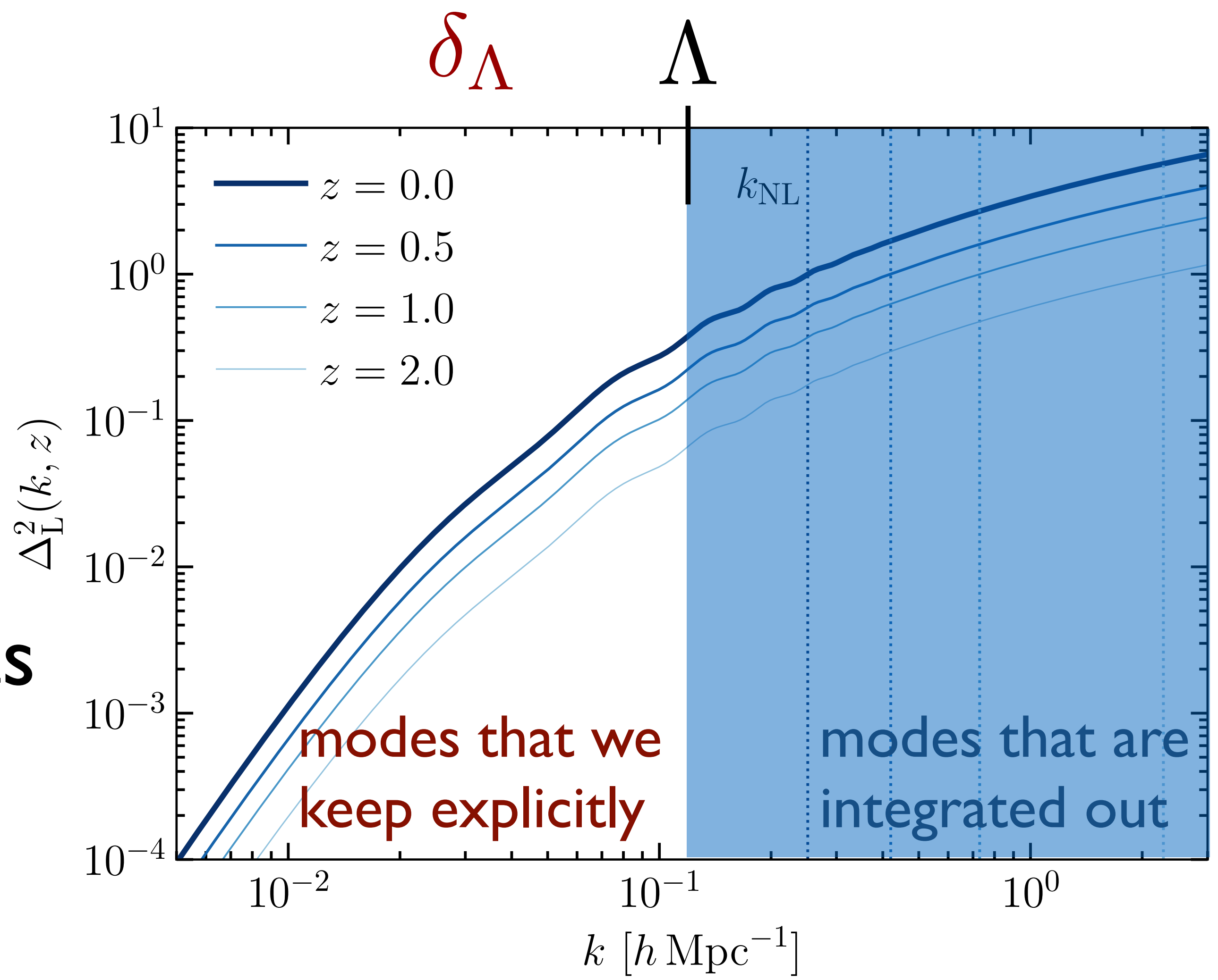


Cabass, FS, arXiv:1909.04022

* In the end, we vary Λ to check convergence.

An EFT approach to the likelihood

- Goal is to derive $P\left(\vec{\delta}_g \middle| \vec{\delta}\right)$ in EFT approach
- Recall: split perturbations into large scale (δ_Λ) and small scale, using sharp-k filter*
- Incorporate effect of large-scale perturbations explicitly using bias expansion, with free coefficients

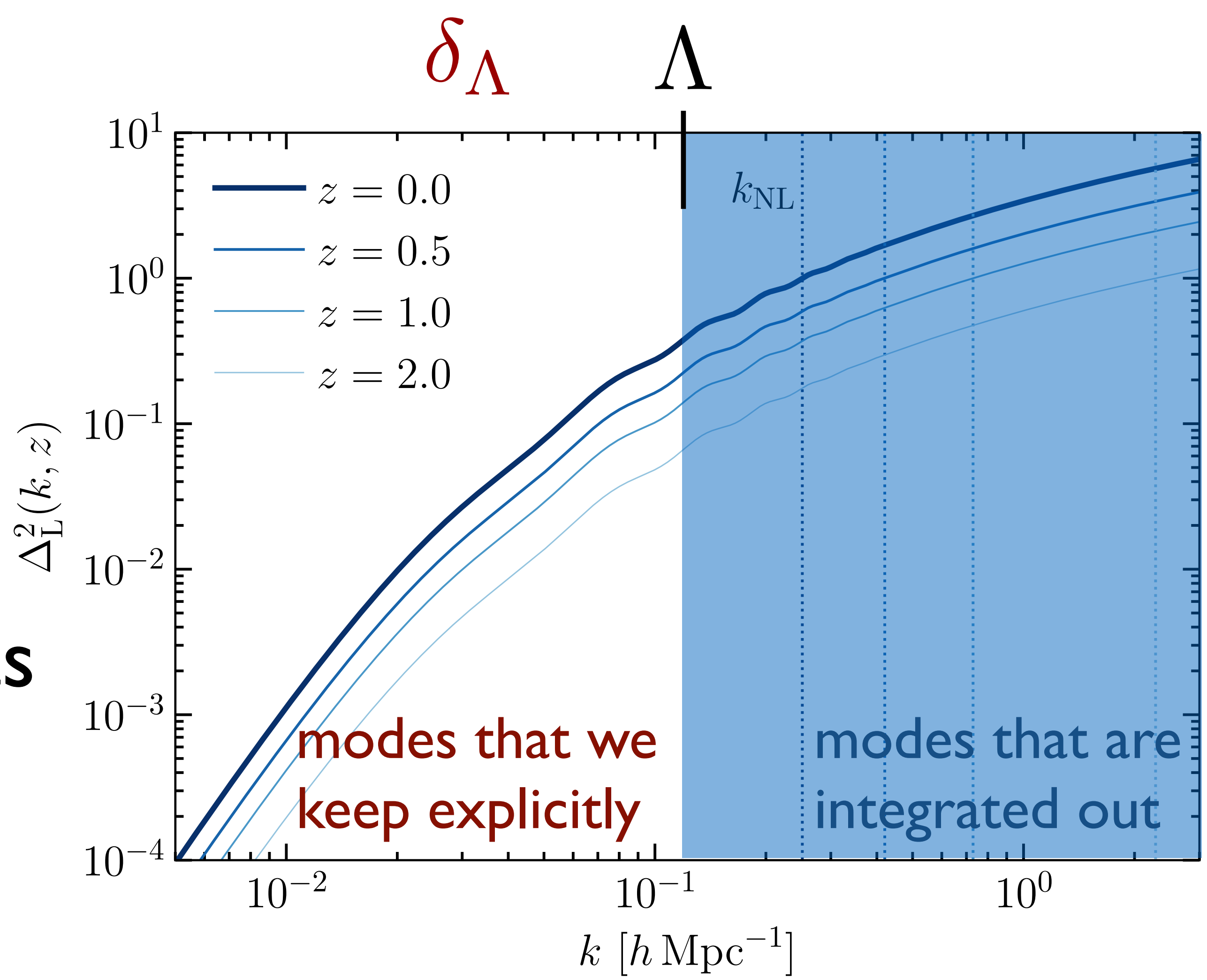


Cabass, FS, arXiv:1909.04022

* In the end, we vary Λ to check convergence.

An EFT approach to the likelihood

- Goal is to derive $P\left(\vec{\delta}_g \middle| \vec{\delta}\right)$ in EFT approach
- Recall: split perturbations into large scale (δ_Λ) and small scale, using sharp-k filter*
- Incorporate effect of large-scale perturbations explicitly using bias expansion, with free coefficients
- Then, use knowledge of PDF of noise ε(x): Gaussian with diagonal covariance in Fourier space



Cabass, FS, arXiv:1909.04022

* In the end, we vary Λ to check convergence.

EFT likelihood

- With these results, we can write:

$$\delta_g(\boldsymbol{k}) = \delta_{g,\mathrm{det}}(\boldsymbol{k}) + \varepsilon(\boldsymbol{k})$$
$$\delta_{g,\mathrm{det}}(\boldsymbol{k}) = \sum_O b_O O(\boldsymbol{k})$$

  - All fields cut at cutoff Λ
  - In addition, employ sharp-k filter $\delta_{\mathrm{in}} \to \delta_{\mathrm{in}}^{\Lambda}$ on initial conditions: crucial to regularize loop integrals involving the observed halo/galaxy field

- and insert $\varepsilon = \delta_g - \delta_{g,\mathrm{det}}$ into the Gaussian noise PDF:

$$P[\varepsilon] \propto \exp\left[-\frac{1}{2}\int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\frac{|\varepsilon(\boldsymbol{k})|^2}{P_\varepsilon(k)}\right]$$

Cabass, FS; 1909.04022
FS, Cabass, Jasche, Lavaux; arXiv:2004.06707

Dark matter halos as a test case

- Let's begin with a thought experiment:
  - We are given a halo catalog and the *normalized amplitudes of the initial conditions* for the matter density in the same volume $\delta_{\rm in}(\boldsymbol{x})$
  - Can we infer the cosmological parameters from this halo catalog?

Dark matter halos as a test case

- Let's begin with a thought experiment:
  - We are given a halo catalog and the *normalized amplitudes of the initial conditions* for the matter density in the same volume $\delta_{\rm in}(\boldsymbol{x})$
  - Can we infer the cosmological parameters from this halo catalog?
- Near optimal case: no cosmic variance
- Of course, not a real-world example, but applicable to halos (or galaxies) in simulations

Dark matter halos as a test case

- Specifically, can we recover unbiased $\mathcal{A}_s$ (σ_8) from a halo catalog (treating bias parameters as unknown) ?
- Perfect degeneracy between b_1 and σ_8 at linear order; nonlinear information (due to protected displacement) essential

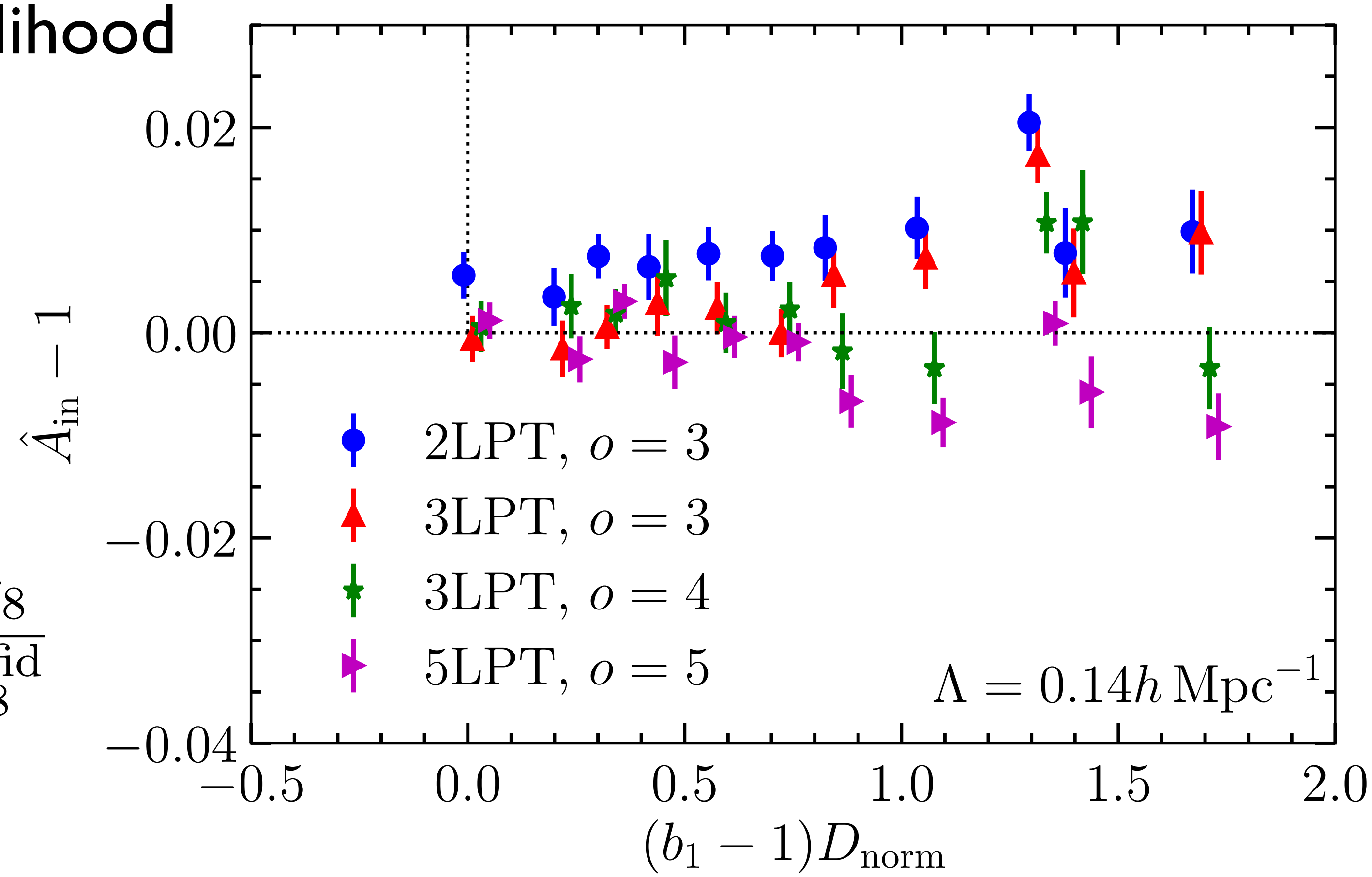
Dark matter halos as a test case

Relative deviation of maximum-likelihood value of σ_8 from ground truth, for different perturbative orders

L_{box} = 2000 Mpc/h

$$A_{\rm in} \equiv \frac{\sigma_8}{\sigma_8^{\rm fid}}$$

Results for all mass bins and redshifts for Λ = 0.14h Mpc^{-1}



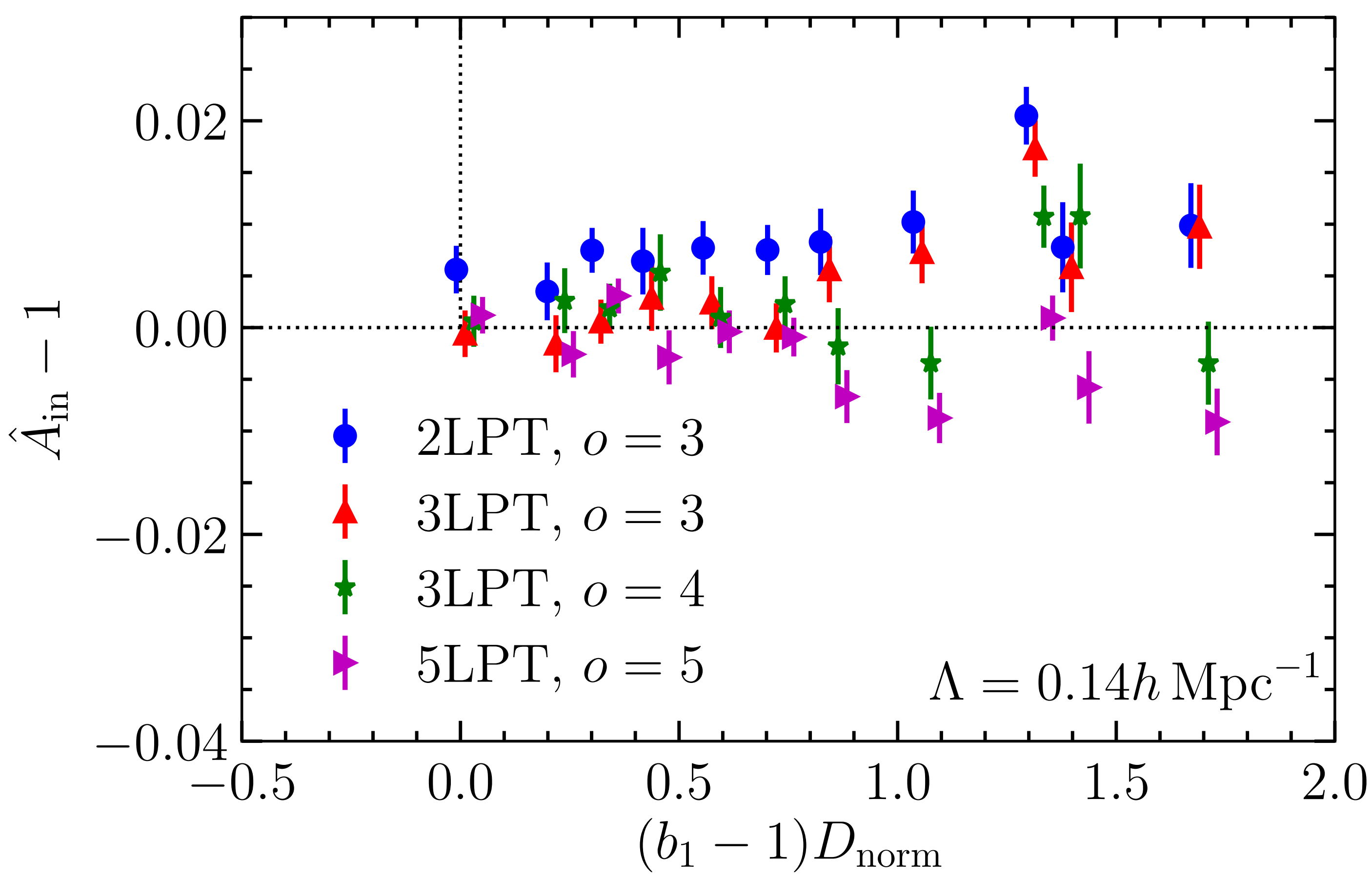
Proxy for higher-order bias terms

FS, Cabass, Jasche, Lavaux; arXiv:2004.06707; FS, arXiv:2009.14176

Dark matter halos as a test case

- Residual error in σ8 at k < 0.14h/Mpc is <~ 1-2% depending on halos mass and redshift
- Most likely due to higher-order bias, and numerical errors of simulations (transients)

Results for all mass bins and redshifts for Λ = 0.14h Mpc-1



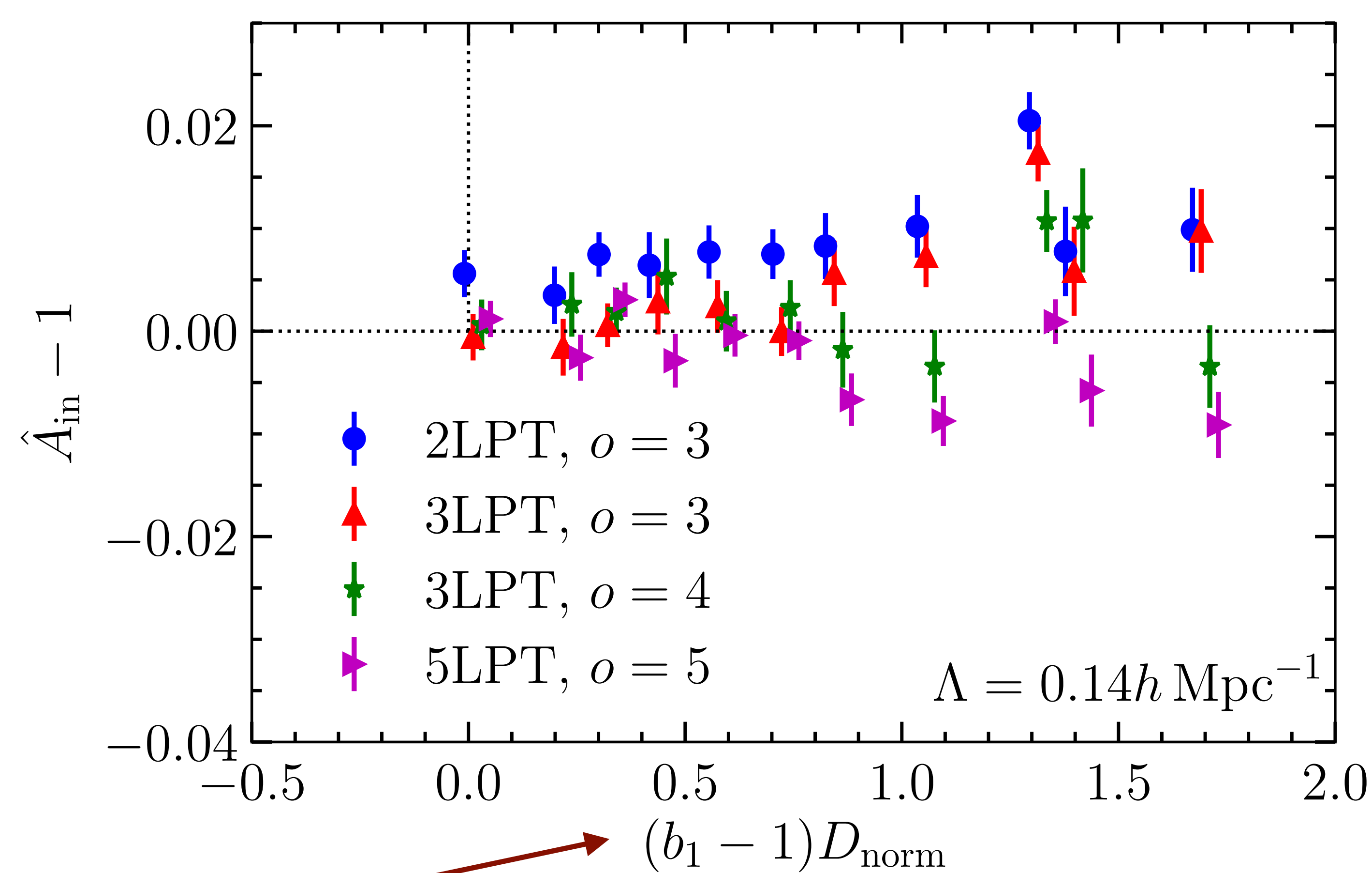
Proxy for higher-order bias terms

FS, Cabass, Jasche, Lavaux; arXiv:2004.06707; FS, arXiv:2009.14176

Dark matter halos as a test case

- Residual error in σ_8 at k < 0.14h/Mpc is <~ 1-2% depending on halos mass and redshift
- Most likely due to higher-order bias, and numerical errors of simulations (transients)

Results for all mass bins and redshifts for Λ = 0.14h Mpc^{-1}



Proxy for higher-order bias terms

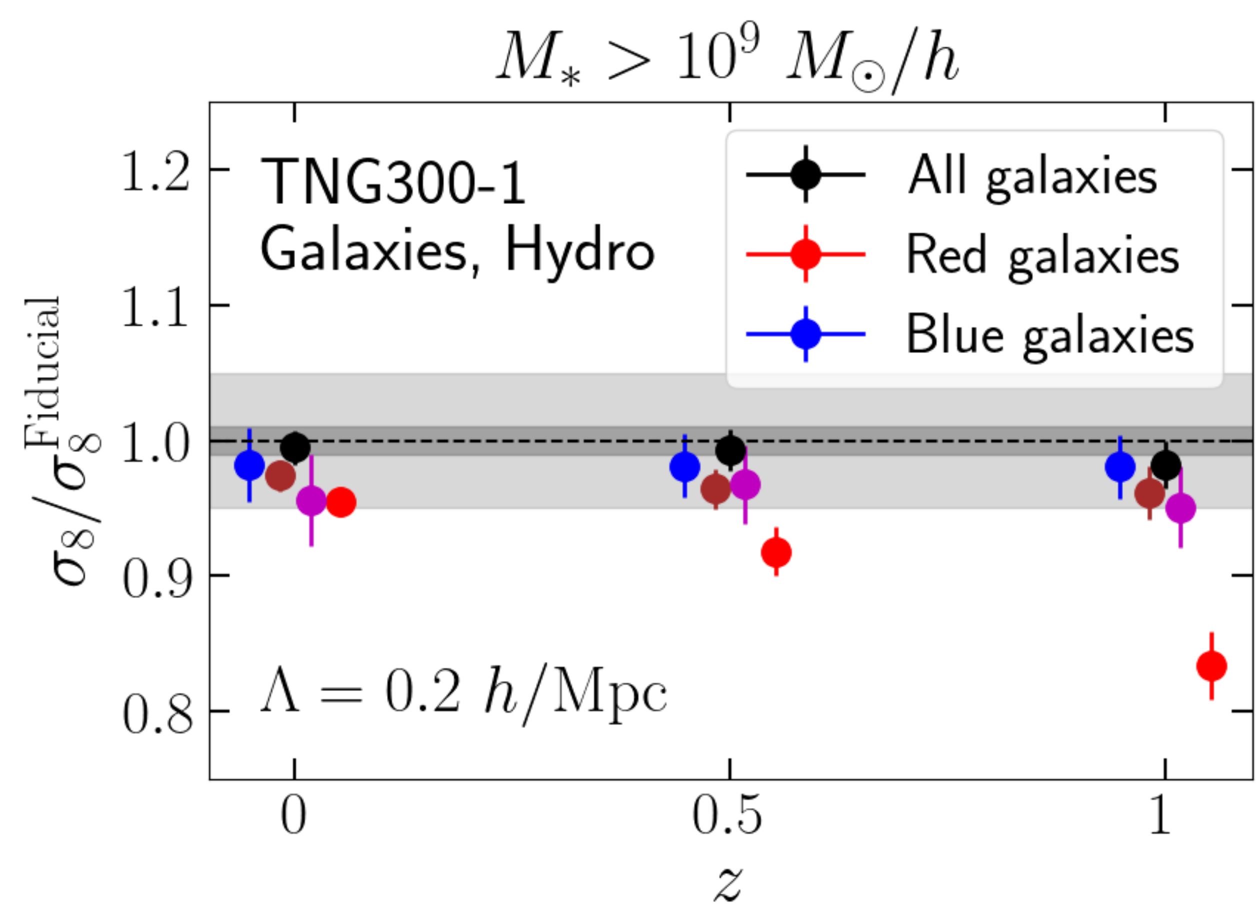
Note: this combination typically grows toward higher z; *bias loops will limit useable range of scales* of upcoming galaxy surveys, *not* matter nonlinearities!

Also works for (simulated) galaxies

- Apply the same analysis to stellar-mass-selected galaxies in IllustrisTNG

L_{box} = 300 Mpc/h

No chance to do this using power spectrum+bispectrum due to cosmic variance…



Barreira, Lazeyras, FS; arXiv:2105.02876

PNG (review)

- PNG has two effects on the clustering of galaxies (see previous talks…)

PNG (review)

- PNG has two effects on the clustering of galaxies (see previous talks…)

1. Adds bispectrum in the initial conditions for gravitational evolution

$$B_m^{(1)}(k_1,k_2,k_3) = T(k_1)T(k_2)T(k_3)B_\phi(k_1,k_2,k_3)$$
$$= T(k_1)T(k_2)T(k_3)2f_{\rm NL}^{\rm loc}\left[P_\phi(k_1)P_\phi(k_2) + {\rm perm.}\right]$$ (local shape)

- Coupled to higher n-point functions by nonlinear evolution

PNG (review)

- PNG has two effects on the clustering of galaxies (see previous talks…)

1. Adds bispectrum in the initial conditions for gravitational evolution

$$B_m^{(1)}(k_1, k_2, k_3) = T(k_1)T(k_2)T(k_3)B_\phi(k_1, k_2, k_3)$$
$$= T(k_1)T(k_2)T(k_3)2f_{\rm NL}^{\rm loc}\left[P_\phi(k_1)P_\phi(k_2) + \text{perm.}\right]$$ (local shape)

   - Coupled to higher n-point functions by nonlinear evolution

2. Mode coupling effect *on small-scale modes that are integrated out* leads to new bias term

$$\delta_g(\boldsymbol{x}, \tau) \supset b_\phi(\tau)\phi(\boldsymbol{q}[\boldsymbol{x}], \tau)$$

Primordial gravitational potential at *Lagrangian* position

Field-level gains for PNG

- For *local* f_{NL}, field-level inference will not improve upon the scale-dependent bias constraint in the power spectrum. However, still expect improvement:

Field-level gains for PNG

- For *local* f_{NL}, field-level inference will not improve upon the scale-dependent bias constraint in the power spectrum. However, still expect improvement:
- Displacement term $s \cdot \nabla\phi$, constrained in similar way as σ_8 shown above
- We reconstruct $\delta^{(1)}$ and hence ϕ: improved constraint on direct contribution of primordial bispectrum (cf. Adam Andrews talk)
  - Independent of b_ϕ (crucial; cf. Alex Barreira's talk)
  - Leading constraint for *nonlocal* f_{NL} (e.g. equilateral, orthogonal)

Conclusions

- Two main messages:
  - *We can deal with complexities of galaxies rigorously on large scales -> EFT*
  - *There is much more (trustable) information in galaxy clustering than what we are using so far -> full inference*

Conclusions

1. *We can deal with complexities of galaxies rigorously on large scales:*

- The EFT provides a complete framework for galaxy biasing
  - Many free parameters, however there are important terms that are *protected by equivalence principle*
  - Gaussian stochasticity on large scales

Conclusions

2. *There is much more (trustable) information in galaxy clustering than what we are using so far:*

- There is a lot of additional information in the phases over summary statistics like Pk+Bk
- The EFT likelihood, coupled with full Bayesian inference, allows us to extract this information with the same rigor as that in the power spectrum
  - Only at the beginning of this program, but first results on f_{NL} look promising (-> Adam Andrews' talk)
  - Reconstruction of initial density (potential) should yield *even more interesting improvement for non-local PNG*!

EFT likelihood

- We obtain the desired conditional probability for δ_g in *Fourier space*:

$$P\left(\vec{\delta}_g \middle| \vec{\delta}\right) \propto \left(\prod_{\boldsymbol{k}\neq 0}^{\Lambda} \sigma^2(k)\right)^{-1/2} \exp\left[-\frac{1}{2}\sum_{\boldsymbol{k}\neq 0}^{\Lambda} \frac{1}{\sigma^2(k)} \left|\delta_g(\boldsymbol{k}) - \delta_{g,\mathrm{det}}(\boldsymbol{k})\right|^2\right]$$

with

$$\delta_{g,\mathrm{det}}(\boldsymbol{k}) = \sum_O b_O O(\boldsymbol{k})$$

$$\sigma^2(k) = \sigma_0^2 + \sigma_2^2 k^2$$

↑ Finite volume in actual data
-> discrete Fourier representation

FS, Elsner, et al; 1808:02002
Cabass, FS; 1909.04022