

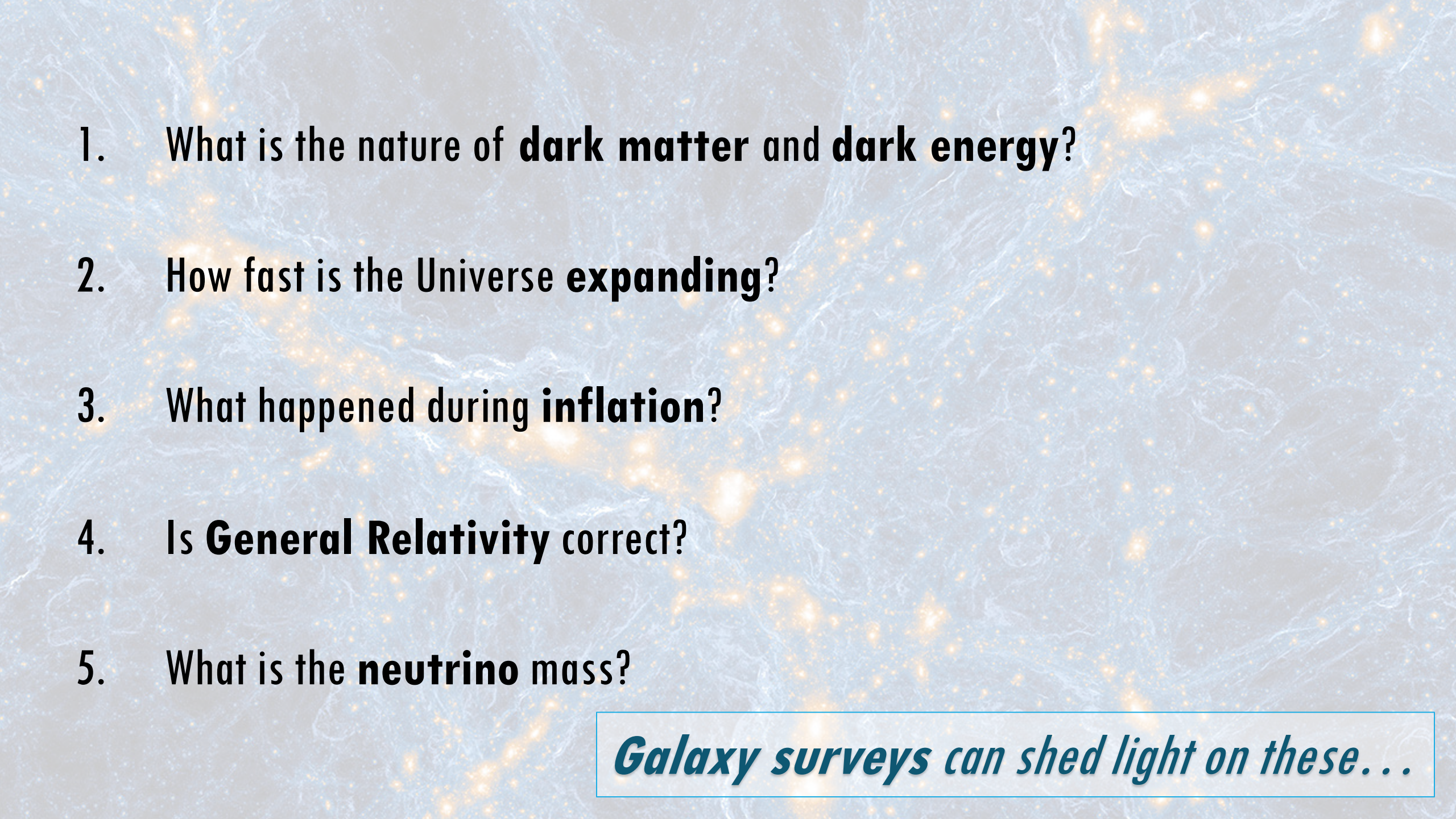


# Large Scale Structure Beyond the Two-Point Function

**Oliver Philcox (Princeton / IAS)**

Physics Division Seminar, LBNL 11/23/21



- 
- The background of the slide is a visualization of the cosmic web, showing a complex network of blue filaments and nodes with yellow and orange galaxy clusters.
1. What is the nature of **dark matter** and **dark energy**?
  2. How fast is the Universe **expanding**?
  3. What happened during **inflation**?
  4. Is **General Relativity** correct?
  5. What is the **neutrino** mass?

*Galaxy surveys can shed light on these...*



# THE QUANTUM UNIVERSE

- ▷ Inflation produces **quantum fluctuations** in the **primordial** matter distribution

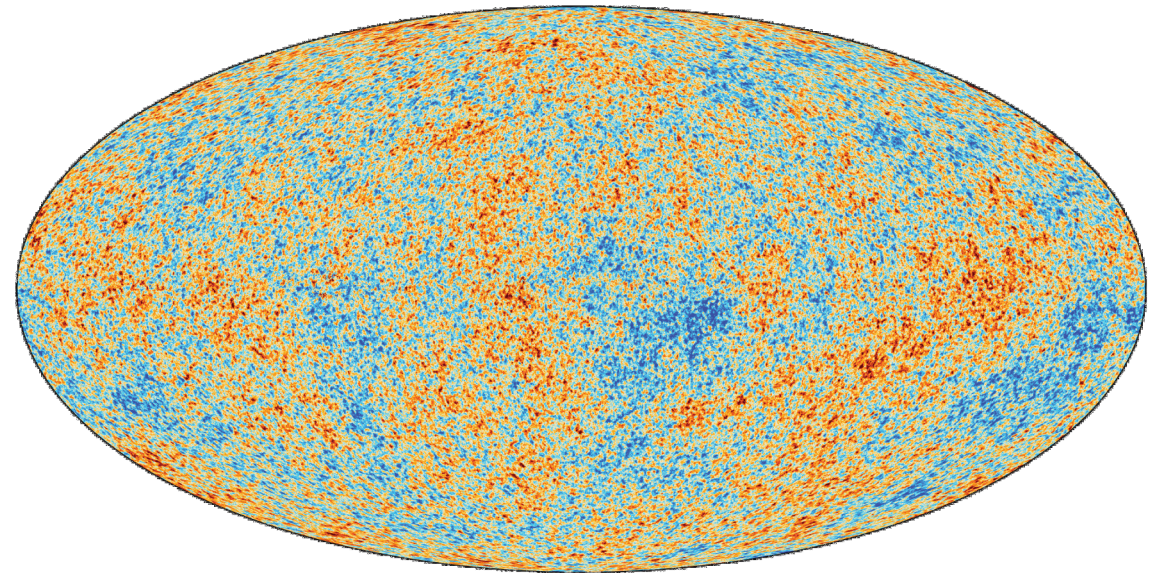
$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \langle \rho \rangle}{\langle \rho \rangle}$$

- ▷ These are (mostly) **adiabatic** and **Gaussian**

$$\delta(\mathbf{k}) \sim \mathcal{N}(0, P_L(\mathbf{k}))$$

- ▷ All information is in the **power spectrum**

*Quantum Fluctuations in the Cosmic Microwave Background*



# THE QUANTUM UNIVERSE

- ▷ Inflation produces **quantum fluctuations** in the **primordial** matter distribution

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \langle \rho \rangle}{\langle \rho \rangle}$$

- ▷ These are (mostly) **adiabatic** and **Gaussian**

$$\delta(\mathbf{k}) \sim \mathcal{N}(0, P_L(\mathbf{k}))$$

- ▷ All information is in the **power spectrum** or two-point correlation function (**2PCF**)

## Fourier Space

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}') P_L(\mathbf{k})$$

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \delta(\mathbf{k}'') \rangle = 0$$

## Configuration Space

$$\langle \delta(\mathbf{r}_1) \delta(\mathbf{r}_2) \rangle = \xi_L(\mathbf{r}_1 - \mathbf{r}_2)$$

$$\langle \delta(\mathbf{r}_1) \delta(\mathbf{r}_2) \delta(\mathbf{r}_3) \rangle = 0$$

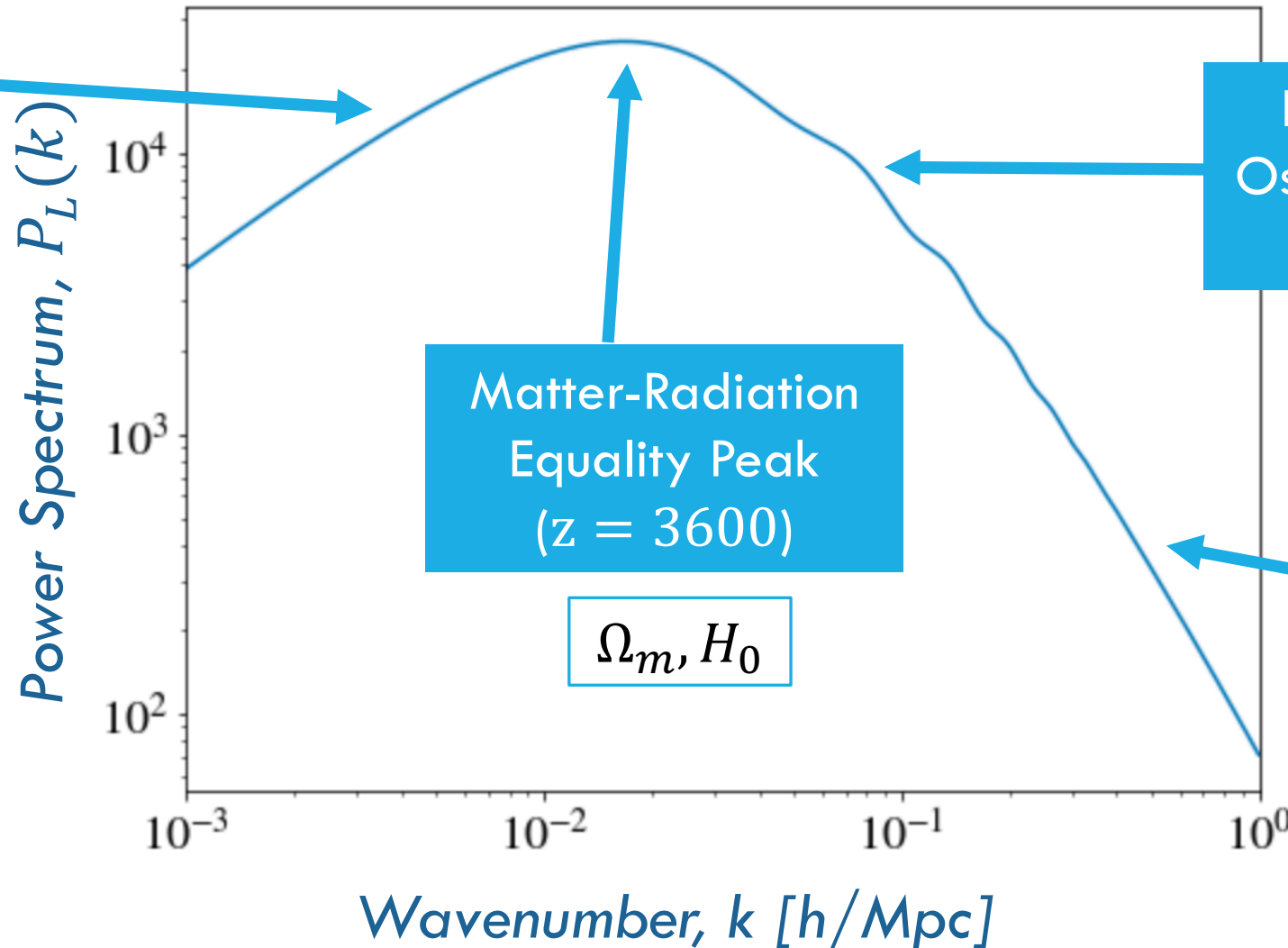


# LINEAR POWER SPECTRUM

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}') P_L(\mathbf{k})$$

Slope from Inflation  
( $z \rightarrow \infty$ )

$$n_s, A_s$$



Matter-Radiation  
Equality Peak  
( $z = 3600$ )

$$\Omega_m, H_0$$

Baryon Acoustic  
Oscillation Wiggles  
( $z = 1090$ )

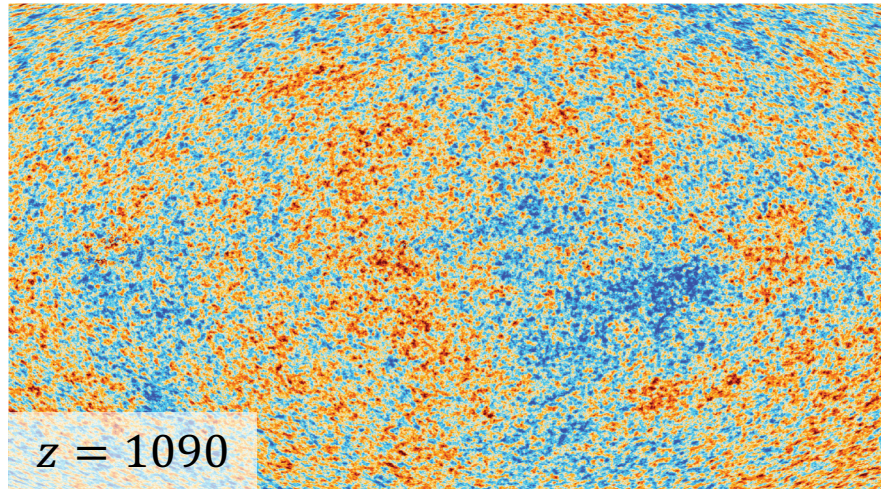
$$\Omega_b, H_0$$

Baryon and  
Neutrino  
Suppression

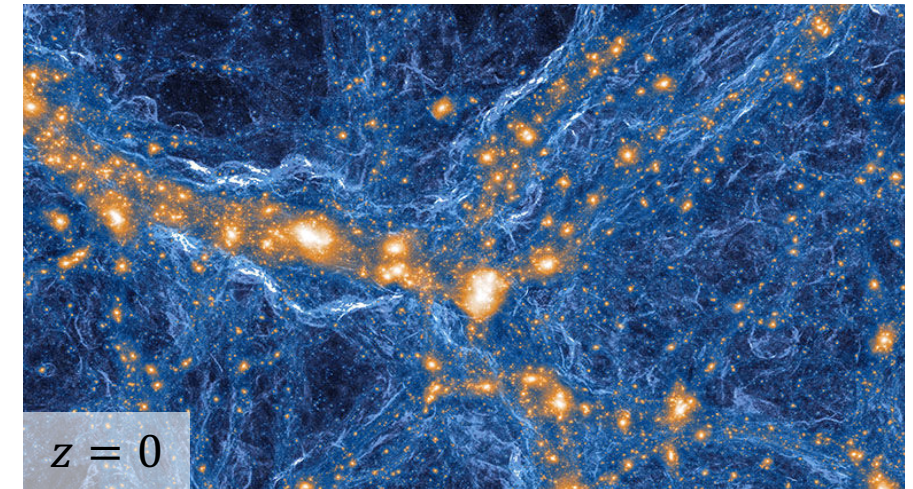
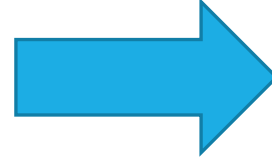
$$\Omega_b, \sum m_\nu$$



# THE LATE UNIVERSE IS NOT GAUSSIAN



Gravitational  
Collapse



$$\delta(\mathbf{k}) \sim \mathcal{N}(0, P_L(\mathbf{k}))$$

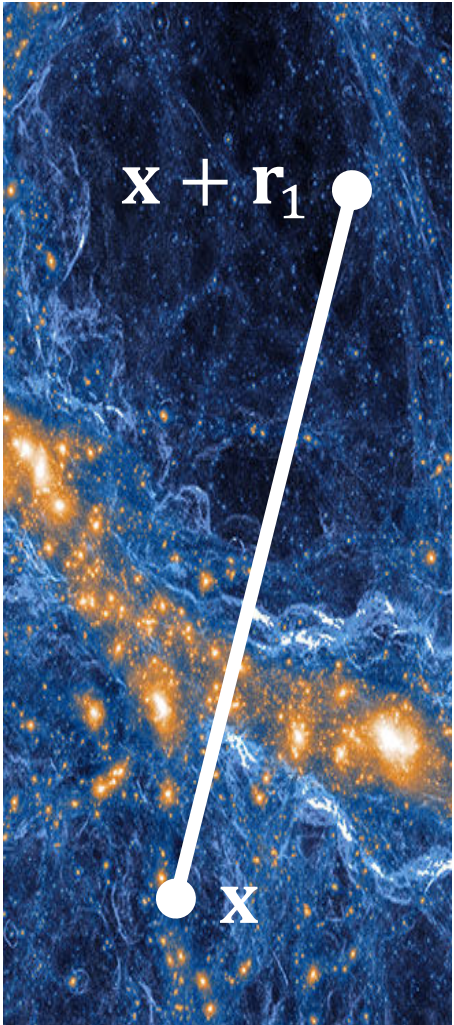
$$\delta(\mathbf{k}) \not\sim \mathcal{N}(0, P_L(\mathbf{k}))$$

- ▶ All information contained in the power spectrum  $[P_L(\mathbf{k})]$
- ▶ **No** higher order statistics needed!

- ▶ **Not** all information contained in the power spectrum
- ▶ Higher-order statistics needed!



# NON-GAUSSIAN DENSITY $\Rightarrow$ NON-GAUSSIAN STATISTICS



## Gaussian

1. Power Spectrum:

$$P(\mathbf{k}_1) = \langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \rangle'$$

2. 2-Point Correlation Function:

$$\xi(\mathbf{r}_1) \equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}_1) \rangle$$

## Non-Gaussian

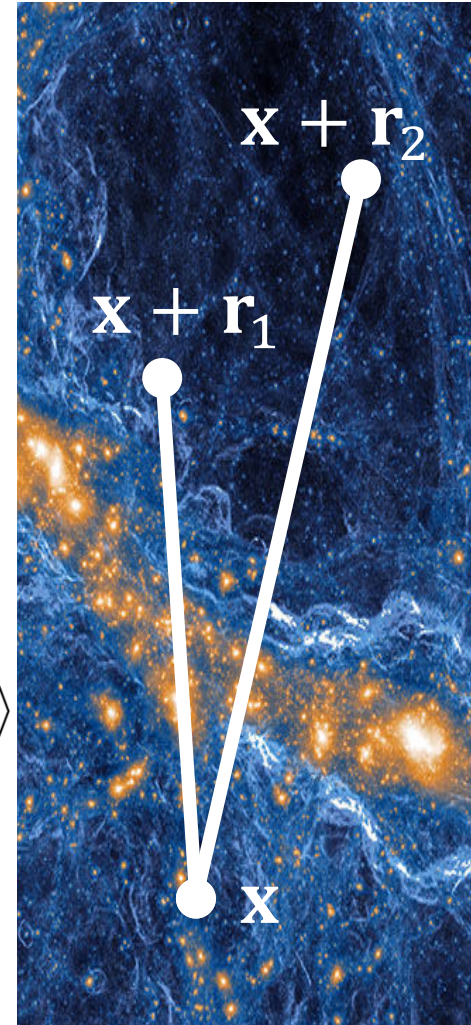
1. Bispectrum:

$$B(\mathbf{k}_1, \mathbf{k}_2) = \langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \delta(\mathbf{k}_3) \rangle'$$

2. 3-Point Correlation Function:

$$\zeta_3(\mathbf{r}_1, \mathbf{r}_2) \equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}_1) \delta(\mathbf{x} + \mathbf{r}_2) \rangle$$

*And beyond...*





# WHAT MAKES UP THE BISPECTRUM?

$$B_g(\mathbf{k}_1, \mathbf{k}_2) = \left[ 2b_1^3 F_2(\mathbf{k}_1, \mathbf{k}_2) + b_2 b_1^2 + 2b_{s^2} b_1^2 \left( \hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2 - 1/3 \right) \right] P_L(k_1) P_L(k_2) + 2 \text{ perms.}$$

The galaxy bispectrum depends on **galaxy formation physics**, **gravity**, and **early-Universe cosmology**.\*

▷ To obtain **all** the large-scale information in the initial conditions, we need:\*

- Power Spectra  $\sim P_L(k)$
- Bispectra  $\sim P_L^2(k)$
- Trispectra  $\sim P_L^3(k)$
- *etc.*



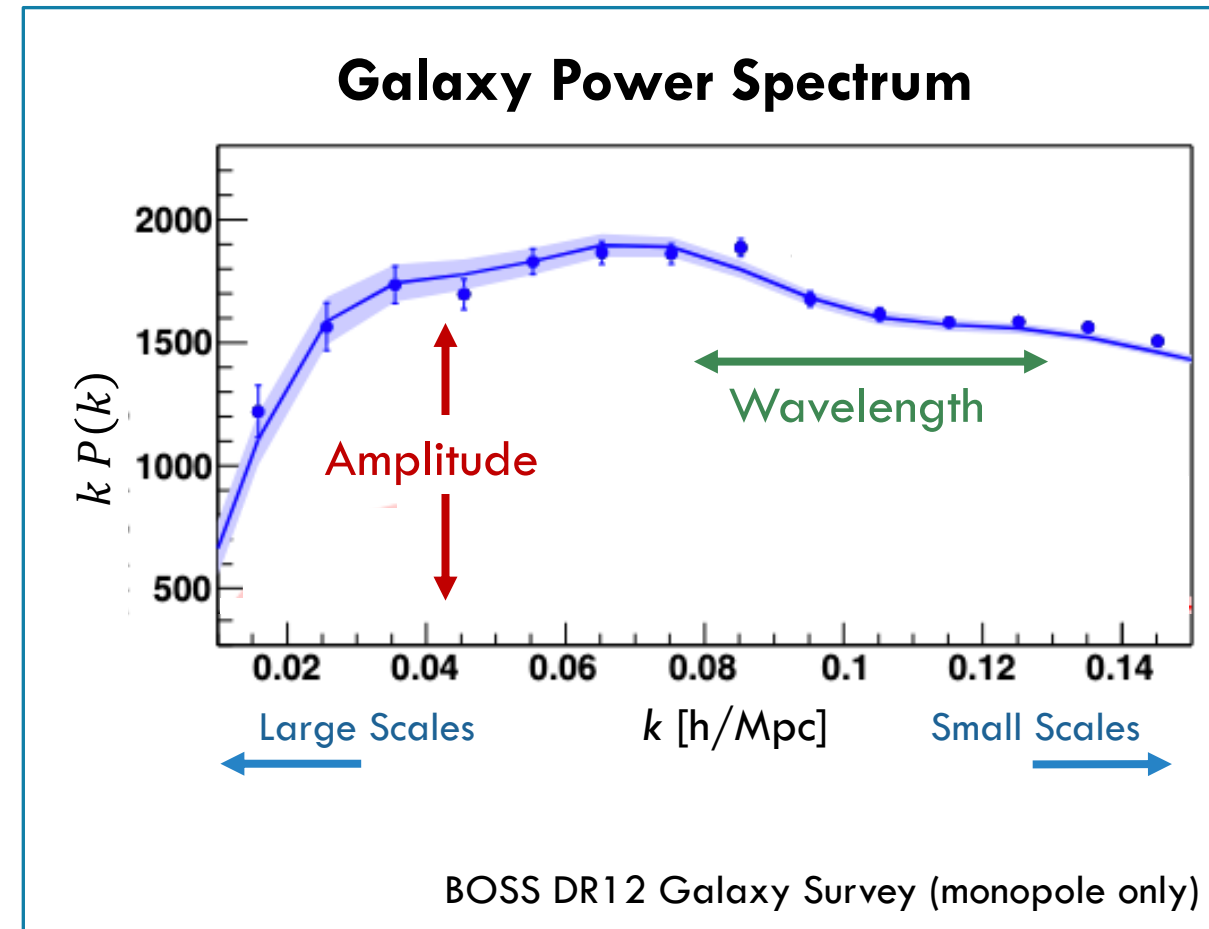
# THE CURRENT STATE OF PLAY

▷ Analyze the galaxy **power spectrum** using a **scaling analysis**

▷ This measures:

- ▷ Overall **amplitude** (= primordial amplitude)
- ▷ **Wiggle** positions (= BAO feature)

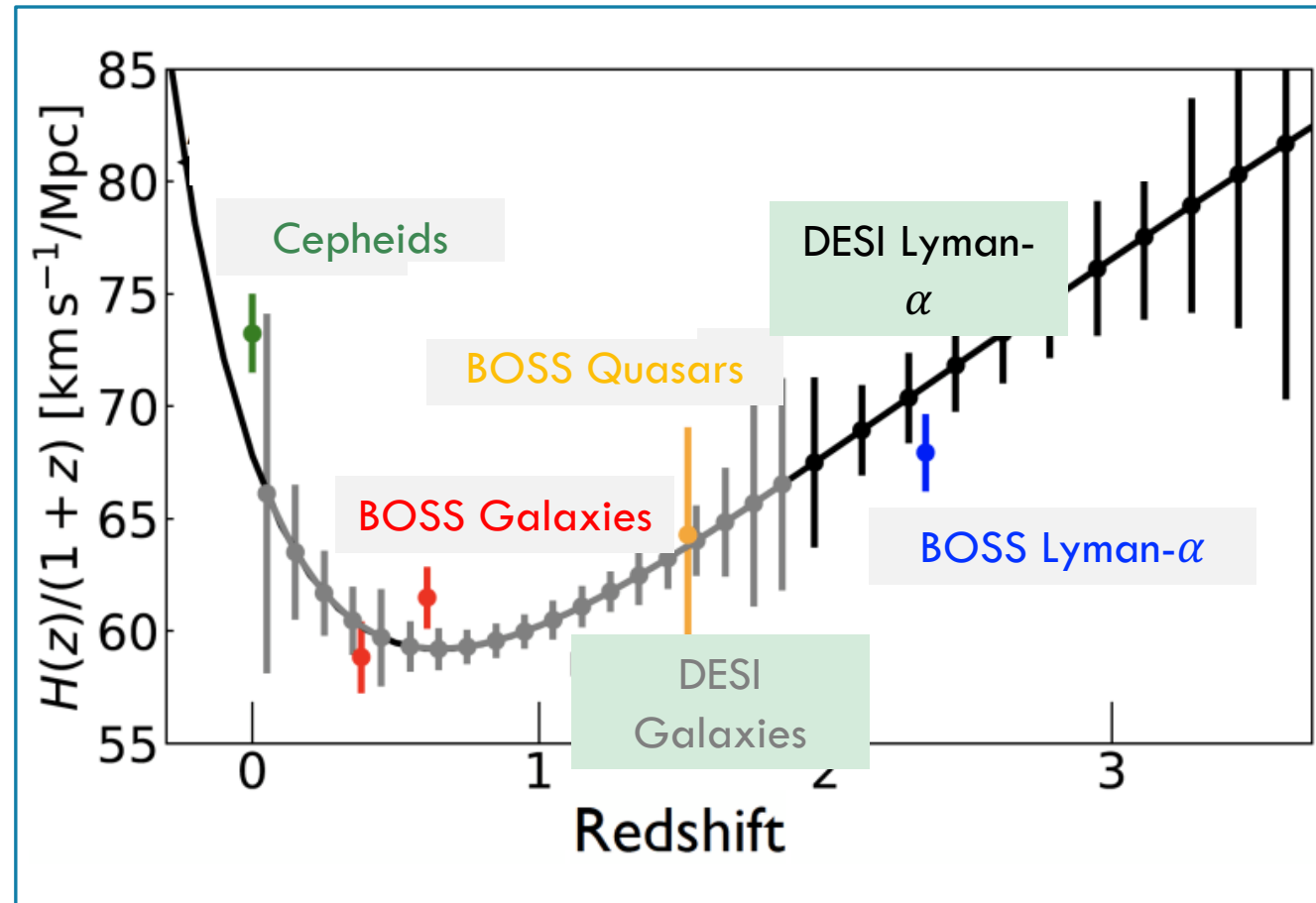
▷ Robust way to constrain **growth rate** and **expansion history**  $H(z)$





# THE CURRENT STATE OF PLAY

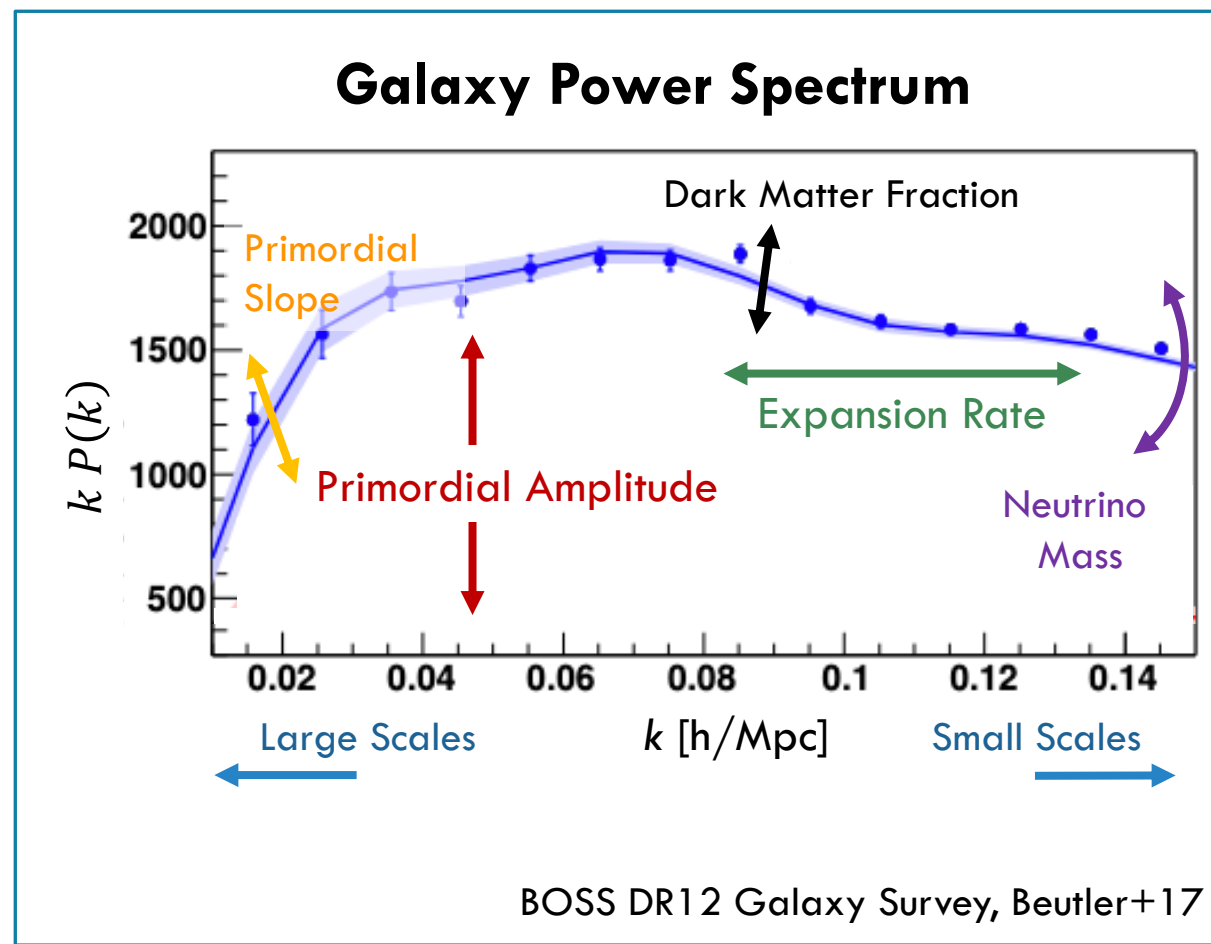
- ▶ Analyze the galaxy **power spectrum** using a **scaling analysis**
- ▶ Measure **wiggle positions** (= BAO feature) and **overall amplitude**
- ▶ Robust way to constrain **growth rate** and **expansion history**  $H(z)$



# THE CURRENT STATE OF PLAY

▷ This is *not* all the available information!

▷ Measure parameters **directly** from the **full shape** of the galaxy power spectrum

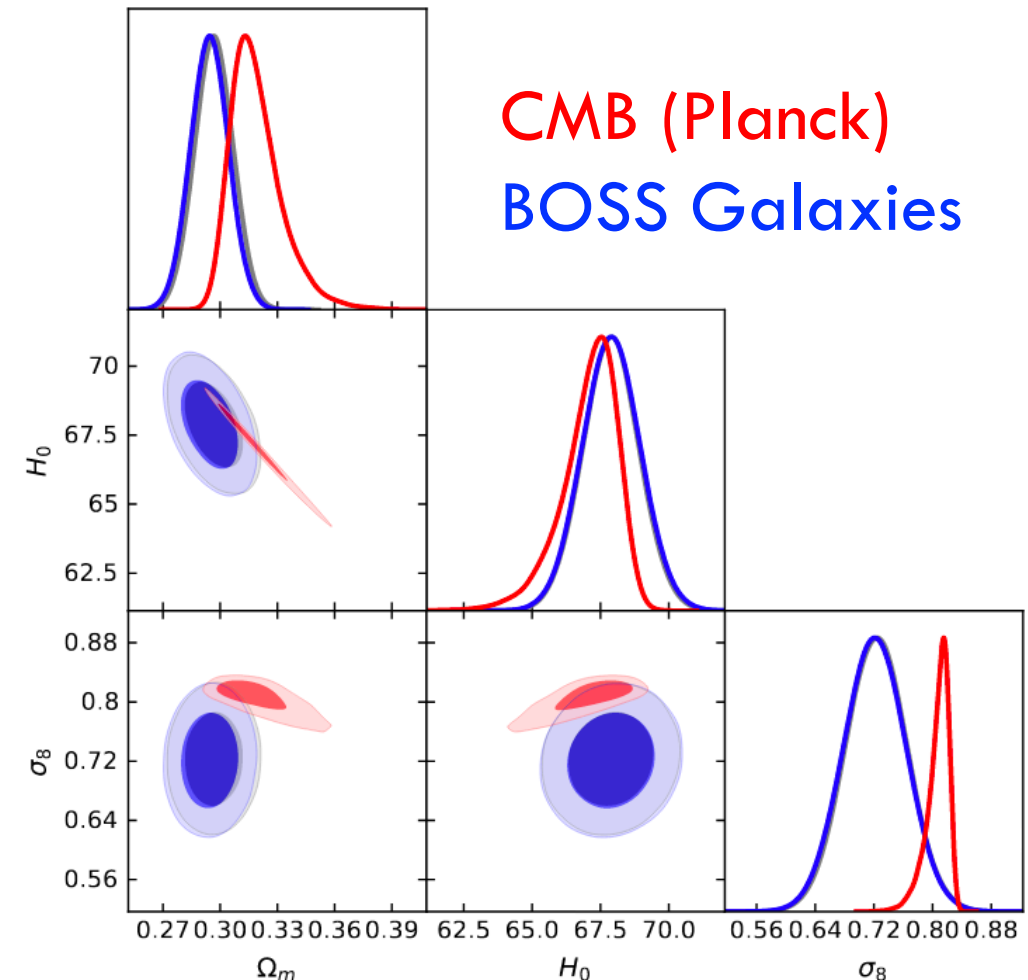




# THE CURRENT STATE OF PLAY

- ▶ This is *not* all the available information!
- ▶ Measure parameters **directly** from the **full shape** of the galaxy power spectrum
- ▶ Constrain parameters in **new** ways e.g. expansion rate from **equality** scale

Can we go *beyond* the power spectrum?



# WHY USE HIGHER-ORDER STATISTICS?

▷ **Sharpen** parameter constraints!

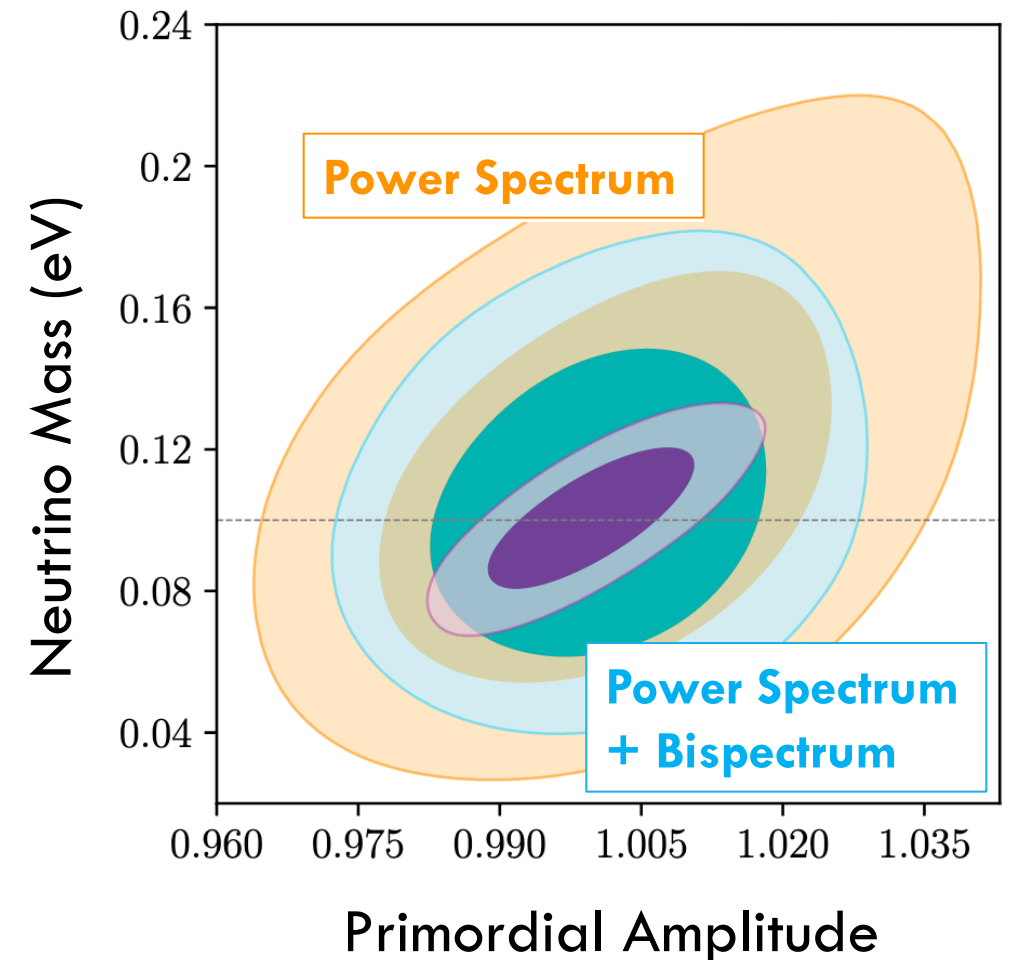
▷ **Break** parameter **degeneracies**!

$$[\text{e.g. } P_g \sim b_1^2 \sigma_8^2, B_g \sim b_1^3 \sigma_8^4]$$

## Euclid Clustering Forecast

▷ Bispectrum improves constraints by  $\approx 40\%$

▷  $1\sigma$  constraint of  $\sigma_{M_\nu} = 13 \text{ meV}$  [including *Planck*]





# NON-GAUSSIAN INFLATION

Are the *primordial perturbations* **Gaussian**?

## Standard Model of Inflation:


▷ Scalar field  $\phi$  rolling down a potential  $V(\phi)$

$$S = \int d^4x \sqrt{-g} \left[ \frac{1}{2} R - \frac{1}{2} \partial^\mu \phi \partial_\mu \phi - V(\phi) \right]$$

Gravity

Kinetic Energy

Potential


$$B_\zeta(\mathbf{k}_1, \mathbf{k}_2) \approx \frac{6}{5} f_{\text{NL}} P_\zeta(k_1) P_\zeta(k_2) + 2 \text{ perms.}$$

## Simplest Inflationary Model:

$$f_{\text{NL}} \sim (1 - n_s) \ll 1$$

**Non-standard** inflation can beat this:

- ▷ Multifield Inflation [Local Bispectrum]
- ▷ New Kinetic Terms [Equilateral Bispectrum]
- ▷ New Vacuum States [Folded Bispectrum]

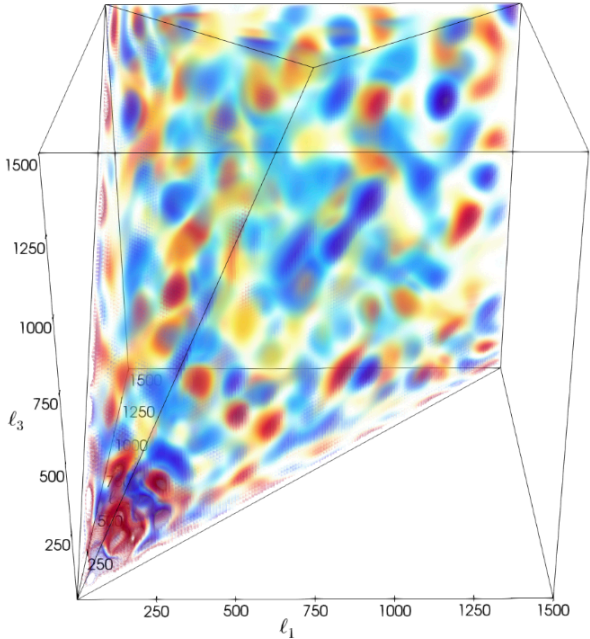
# NON-GAUSSIAN INFLATION

$$B_{\zeta}(\mathbf{k}_1, \mathbf{k}_2) \approx \frac{6}{5} f_{\text{NL}} P_{\zeta}(k_1) P_{\zeta}(k_2) + 2 \text{ perms.}$$

How do we measure this?

## 1. CMB Bispectrum

Planck TTT Bispectrum



$\approx 2\times$  better  
with CMB-S4!

$f_{\text{NL}}$  Constraints

Local . . . . .	$6.7 \pm 5.6$
Equilateral . . . . .	$6 \pm 66$
Orthogonal . . . . .	$-38 \pm 36$



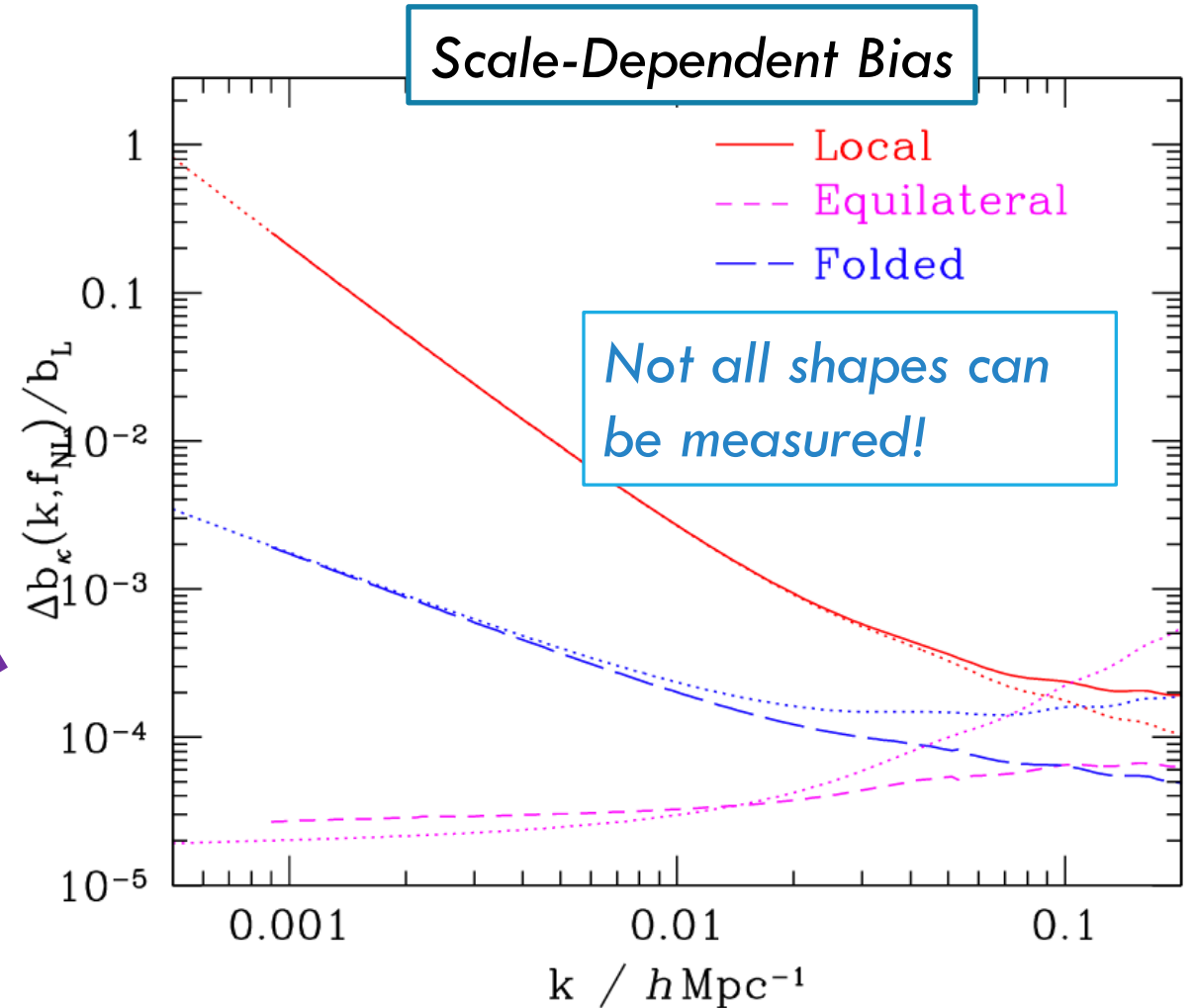
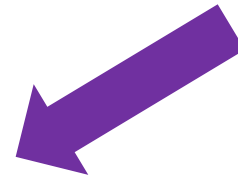
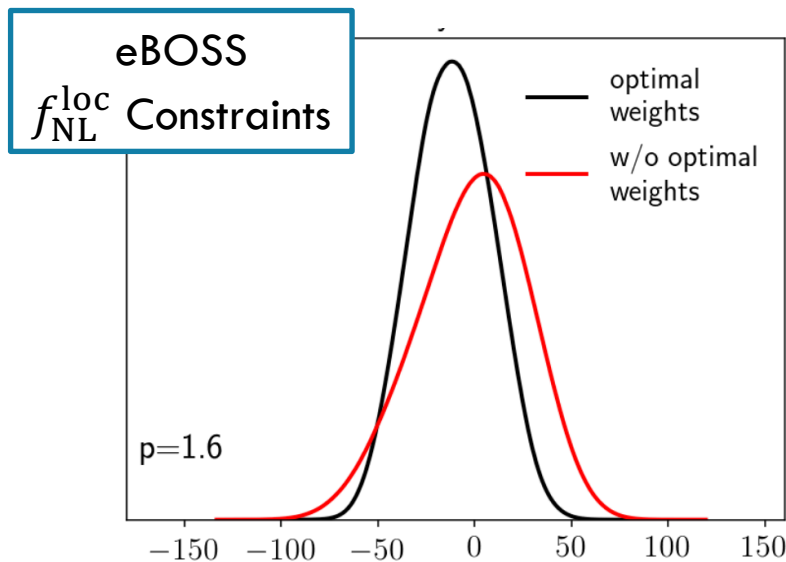
# NON-GAUSSIAN INFLATION

$$B_{\zeta}(\mathbf{k}_1, \mathbf{k}_2) \approx \frac{6}{5} f_{\text{NL}} P_{\zeta}(k_1) P_{\zeta}(k_2) + 2 \text{ perms.}$$

How do we measure this?

1. CMB Bispectrum

2. Galaxy Power Spectrum

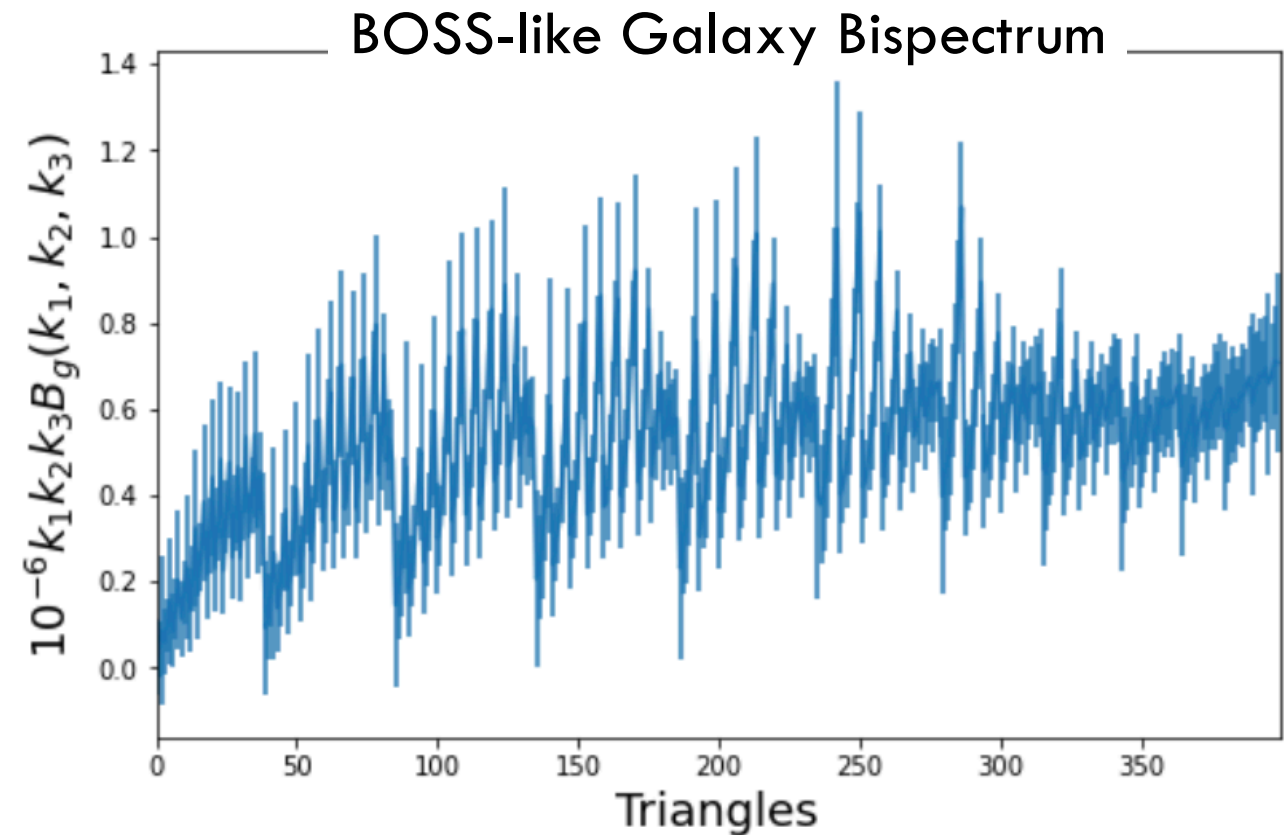


# NON-GAUSSIAN INFLATION

$$B_{\zeta}(\mathbf{k}_1, \mathbf{k}_2) \approx \frac{6}{5} f_{\text{NL}} P_{\zeta}(k_1) P_{\zeta}(k_2) + 2 \text{ perms.}$$

How do we measure this?

1. CMB Bispectrum
2. Galaxy Power Spectrum
3. **Galaxy Bispectrum**





# CHERN-SIMONS INTERACTIONS VIOLATE PARITY

$$S = \int d^4x \sqrt{-g} \left[ \frac{1}{2} R - \frac{1}{2} \partial^\mu \phi \partial_\mu \phi - V(\phi) - \frac{1}{4} f(\phi) F_{\mu\nu} F^{\mu\nu} + \frac{\gamma}{4} f(\phi) F_{\mu\nu} \tilde{F}^{\mu\nu} \right]$$

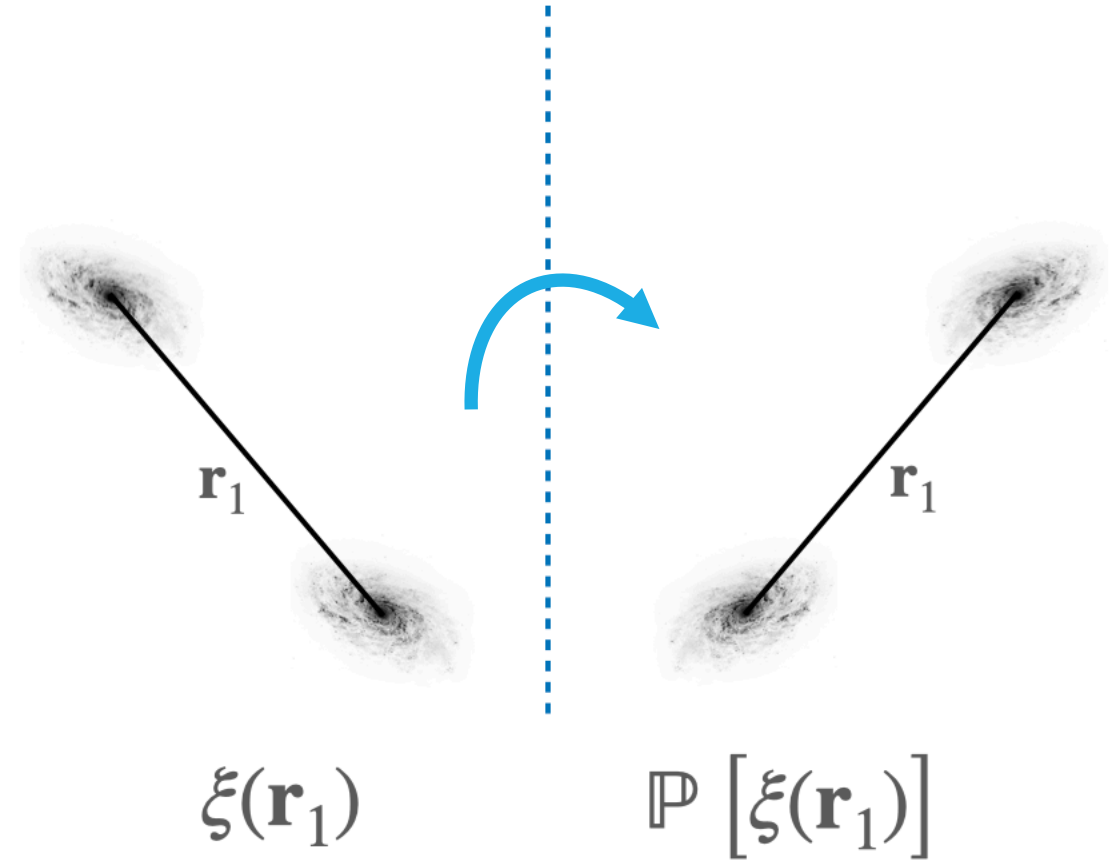
- ▶ Add a **gauge field**  $A_\mu$  to the inflationary action, via  $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$
- ▶ This can include a **Chern-Simons coupling** to the (pseudo-)scalar  $\phi$  [motivated by baryogenesis]
- ▶  $f(\phi) F_{\mu\nu} \tilde{F}^{\mu\nu}$  violates **parity symmetry**  $\Rightarrow$  parity-violating correlators!

*Where should we look for these signatures?*

# THE 2PCF + 3PCF ARE NOT SENSITIVE TO PARITY

## 2-Point Correlation Function (2PCF):

Parity Inversion = Rotation



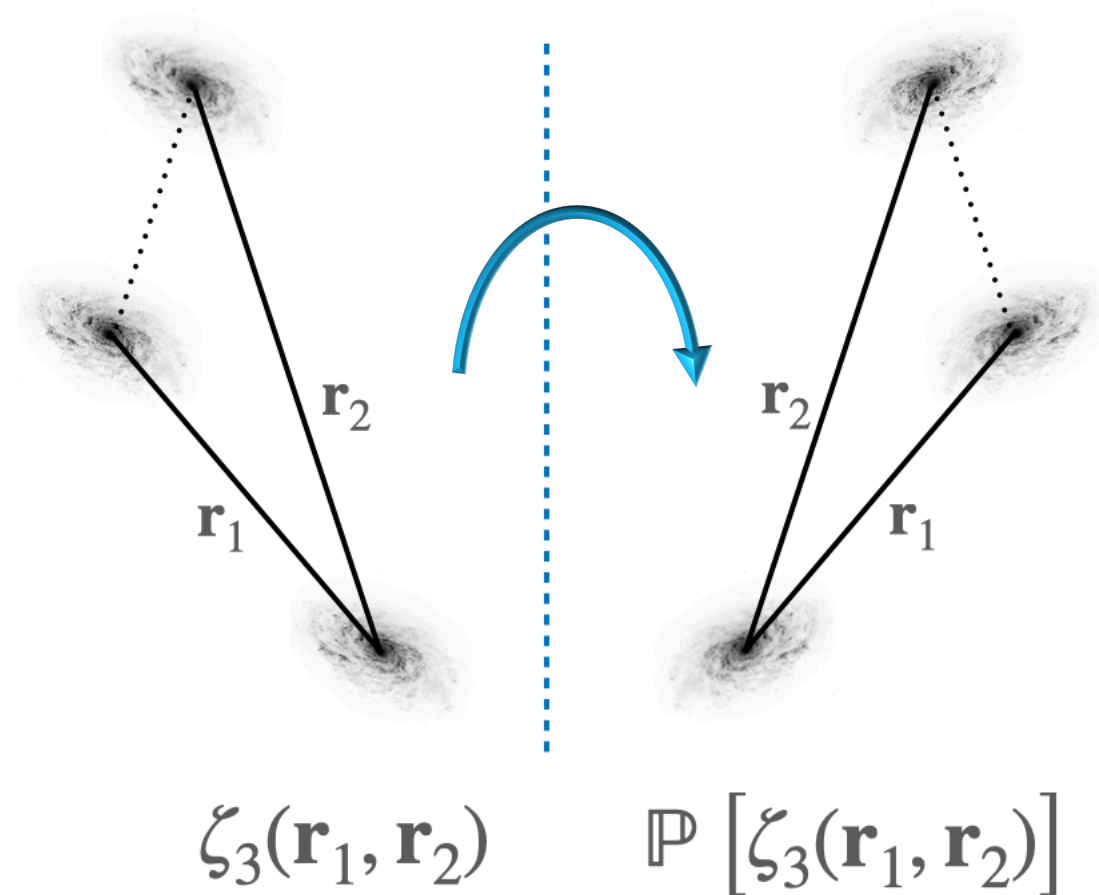
# THE 2PCF + 3PCF ARE NOT SENSITIVE TO PARITY

## 2-Point Correlation Function (2PCF):

Parity Inversion = Rotation

## 3-Point Correlation Function (3PCF):

Parity Inversion = Rotation





# THE 2PCF + 3PCF ARE NOT SENSITIVE TO PARITY

## 2-Point Correlation Function (2PCF):

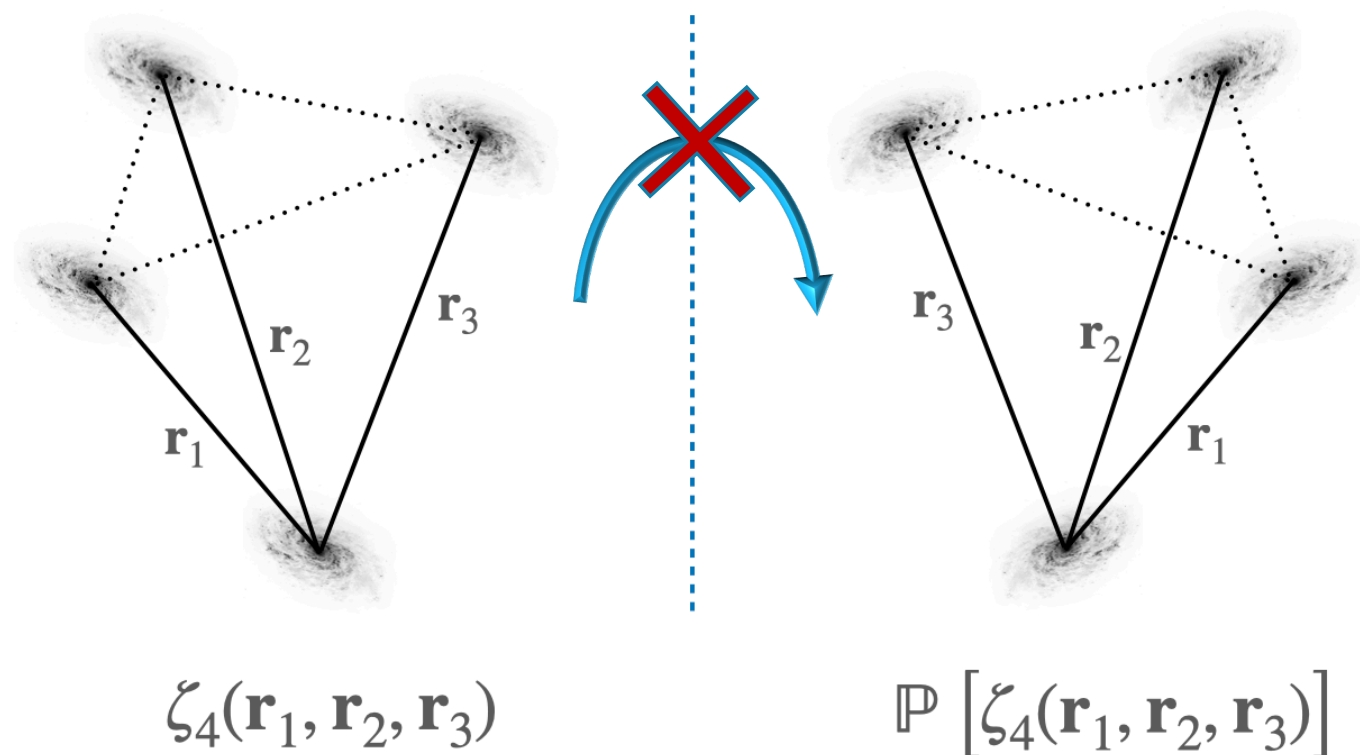
Parity Inversion = Rotation

## 3-Point Correlation Function (3PCF):

Parity Inversion = Rotation

## 4-Point Correlation Function (4PCF):

Parity Inversion  $\neq$  Rotation

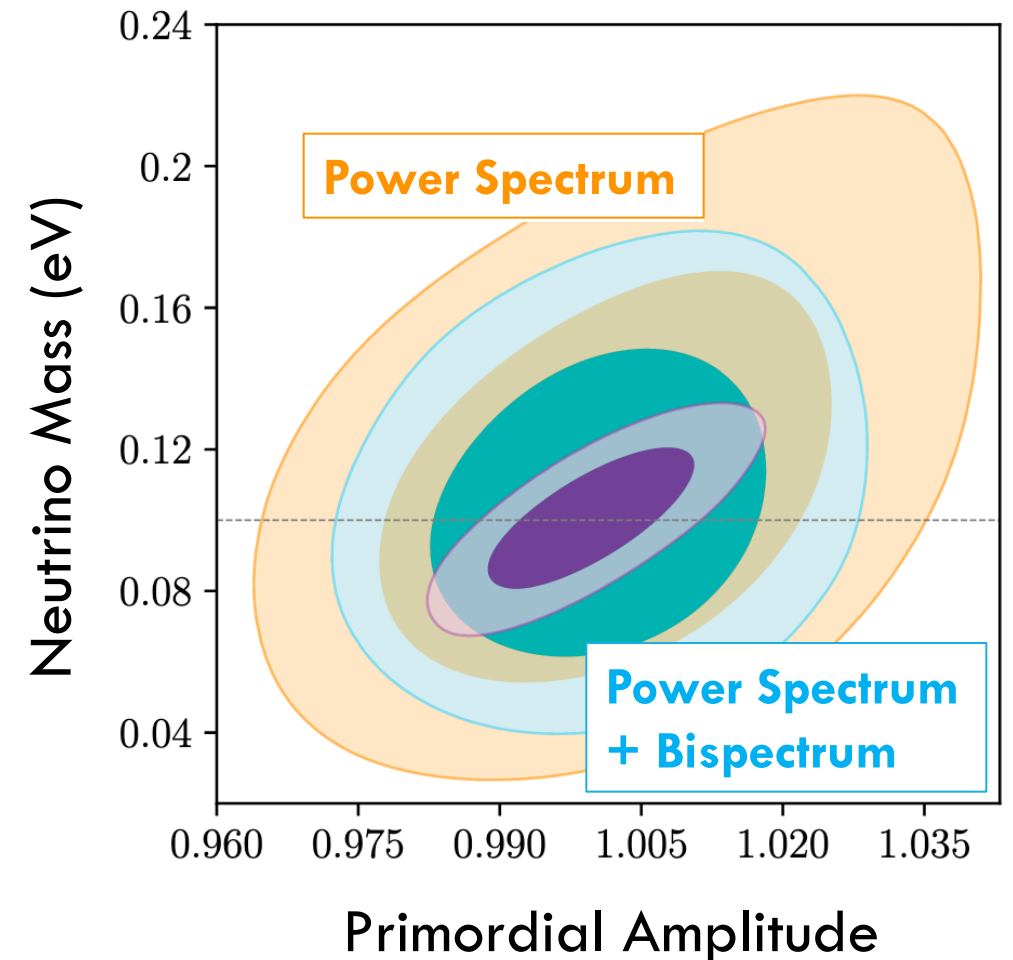


# WHY USE HIGHER-ORDER STATISTICS?

- ▷ **Sharpen** parameter constraints!
- ▷ **Break** parameter **degeneracies**!
- ▷ Test **non-standard** physics models!

## Why Use Large Scale Structure?

- Signal-to-Noise is **cubic** in number of modes unlike CMB
- New physics constraints **don't** dilute with redshift



# HOW TO MEASURE A BISPECTRUM

$$\hat{B}_g(k_1, k_2, k_3) = \int_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3 \in \text{bins}} \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \delta_g(\mathbf{k}_3) (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)$$

**Problem:** We don't measure the density field directly.

$$\delta_g(\mathbf{r}) \rightarrow W(\mathbf{r})\delta_g(\mathbf{r}) \quad \delta_g(\mathbf{k}) \rightarrow \int \frac{d\mathbf{p}}{(2\pi)^3} W(\mathbf{k} - \mathbf{p})\delta_g(\mathbf{p})$$

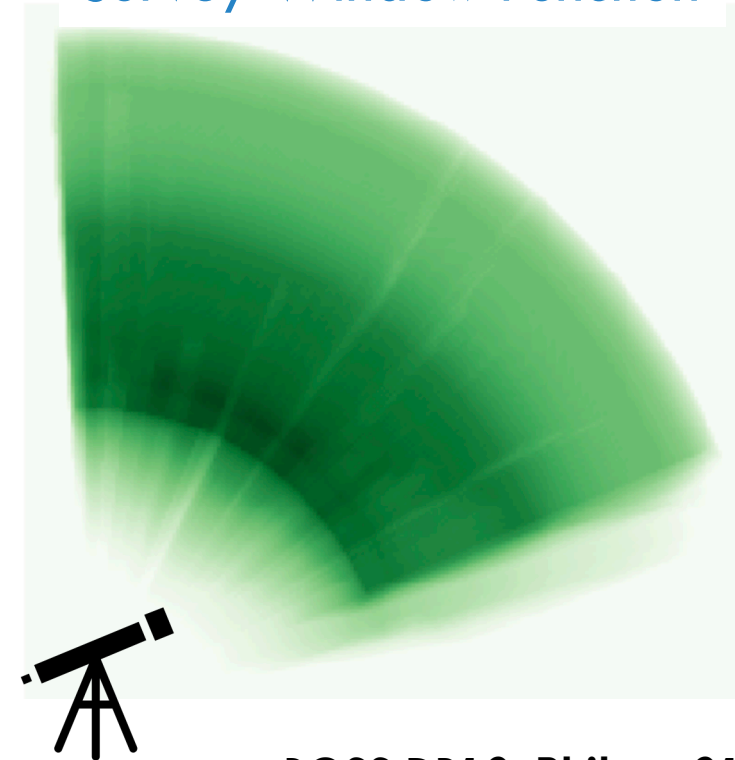
Window Function

The measured bispectrum is a triple **convolution**

$$B_g(\mathbf{k}_1, \mathbf{k}_2) \rightarrow \int_{\mathbf{p}_1 \mathbf{p}_2} W(\mathbf{k}_1 - \mathbf{p}_1) W(\mathbf{k}_2 - \mathbf{p}_2) W(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{k}_1 - \mathbf{k}_2) B_g(\mathbf{p}_1, \mathbf{p}_2)$$

**Solution:** Convolve the **theory model** too

Survey Window Function





# CONVOLUTION IS EXPENSIVE

$$B_g^{\text{win}}(\mathbf{k}_1, \mathbf{k}_2) = \int_{\mathbf{p}_1 \mathbf{p}_2} W(\mathbf{k}_1 - \mathbf{p}_1) W(\mathbf{k}_2 - \mathbf{p}_2) W(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{k}_1 - \mathbf{k}_2) B_g(\mathbf{p}_1, \mathbf{p}_2)$$

- ▶ Window convolution is too costly to do repeatedly!
- ▶ Common approximation: apply the window **only** to the power spectrum

$$B_g(\mathbf{k}_1, \mathbf{k}_2) \supset P_L(k_1) P_L(k_2)$$

**But:**

- This gives **systematic errors** on large scales
- Spectra cannot be used to search for new physics!

# BISPECTRA WITHOUT WINDOWS

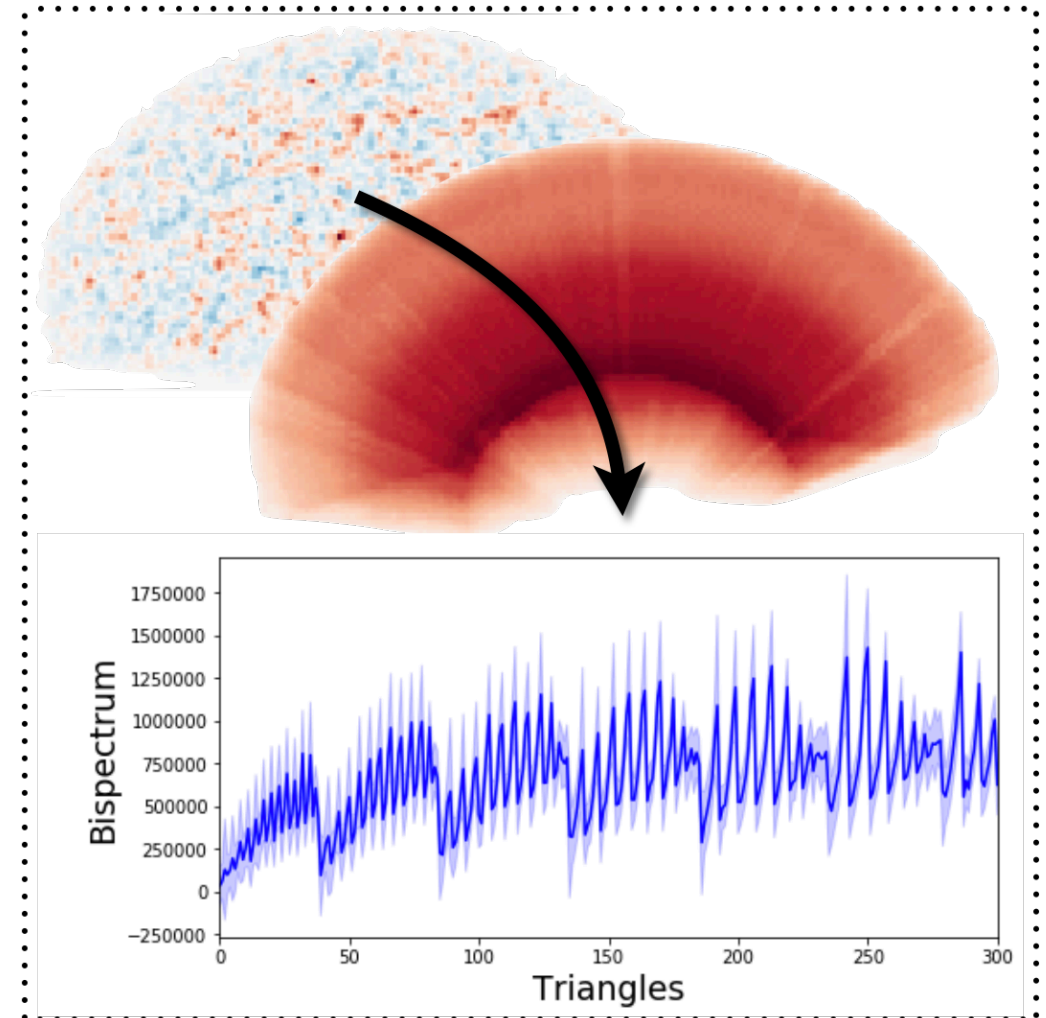
**Alternatively:** estimate the **unwindowed** bispectrum directly

$$B_g^{\text{win}}(\mathbf{k}_1, \mathbf{k}_2) = \int_{\mathbf{p}_1 \mathbf{p}_2} W(\mathbf{k}_1 - \mathbf{p}_1) W(\mathbf{k}_2 - \mathbf{p}_2) W(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{k}_1 - \mathbf{k}_2) B_g(\mathbf{p}_1, \mathbf{p}_2)$$

▷ Derive a **maximum-likelihood** estimator for the **true** bispectrum

▷ Effectively **deconvolves** the window

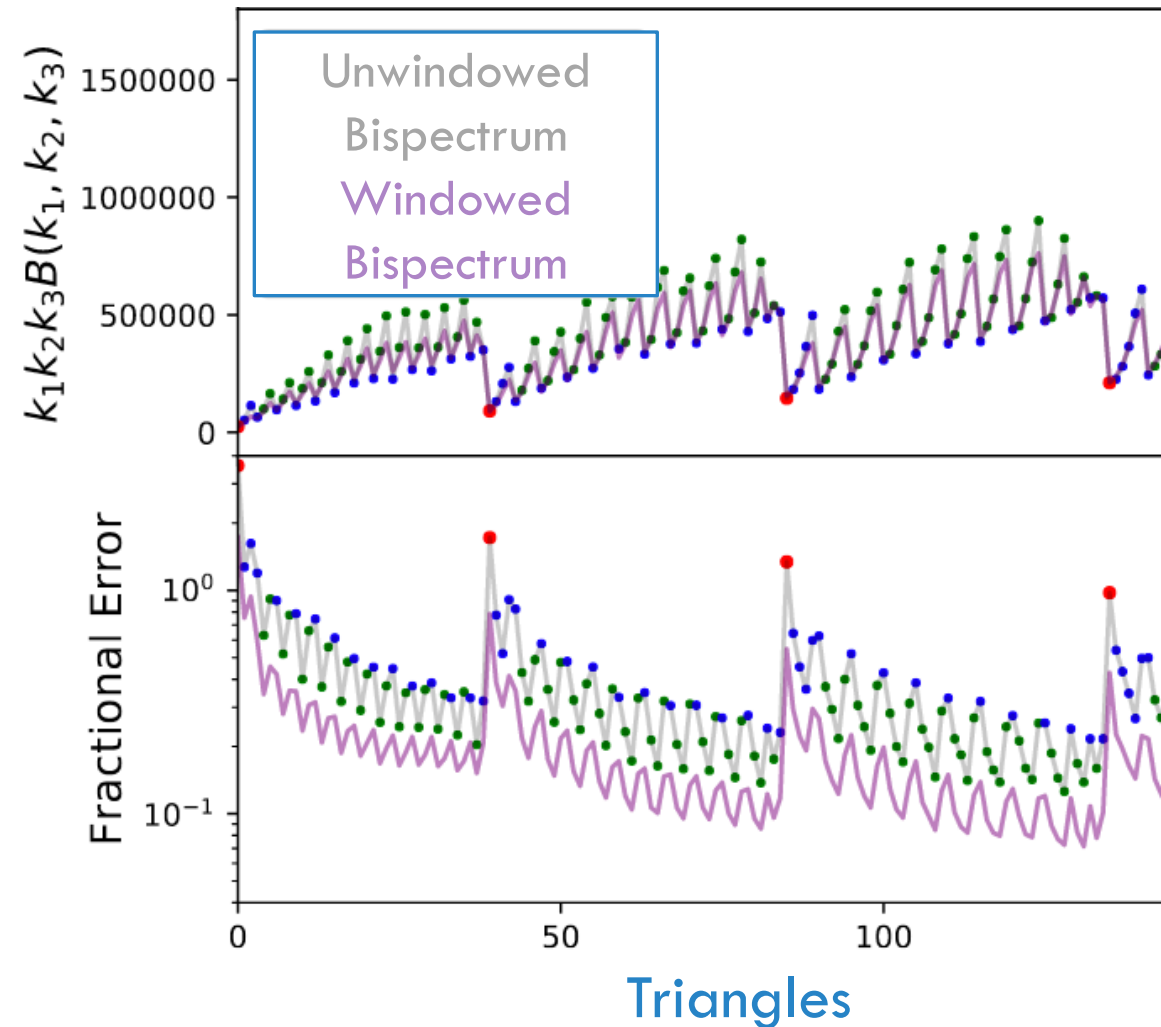
$$\nabla_{B_g} L[\text{data} | B_g] = 0 \quad \Rightarrow \quad \hat{B}_g = \dots$$



# BISPECTRA WITHOUT WINDOWS

Properties of the **cubic estimator**:

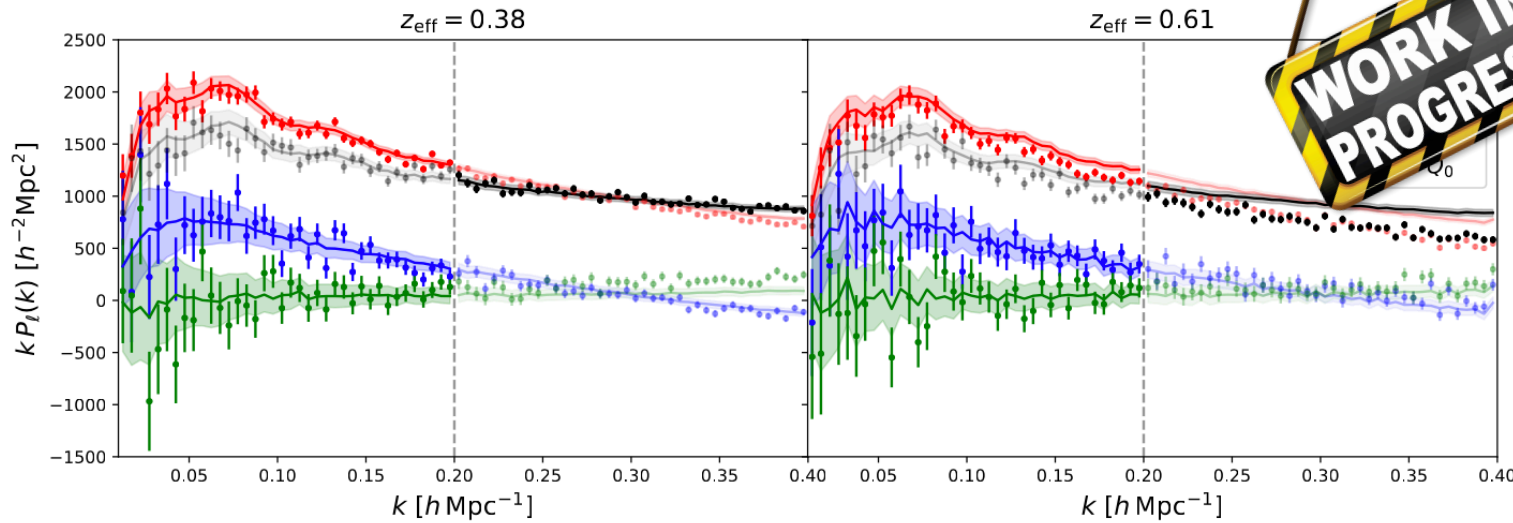
1. Unbiased
  2. Minimum variance [as  $B(k_1, k_2, k_3) \rightarrow 0$ ]
  3. Window-free [effectively a deconvolution]
- Requires various tricks for dealing with high-dimensional data [e.g. conjugate gradient descent, Monte Carlo estimation etc.]



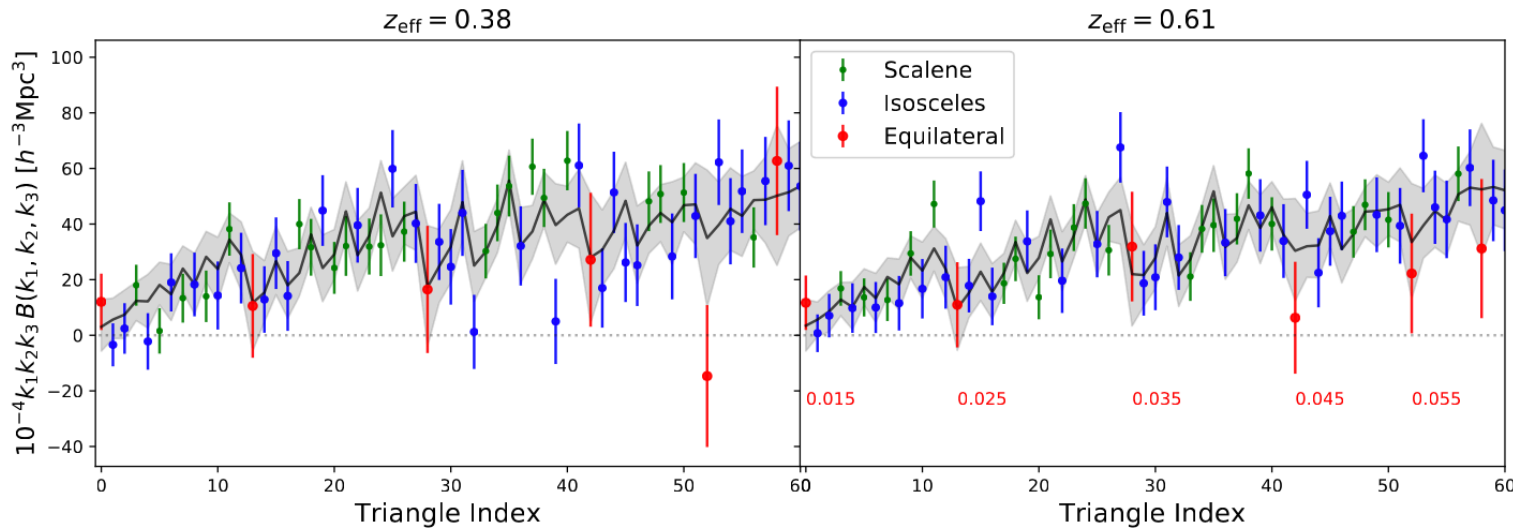


# BOSS WITHOUT WINDOWS

Power Spectra



Bispectra

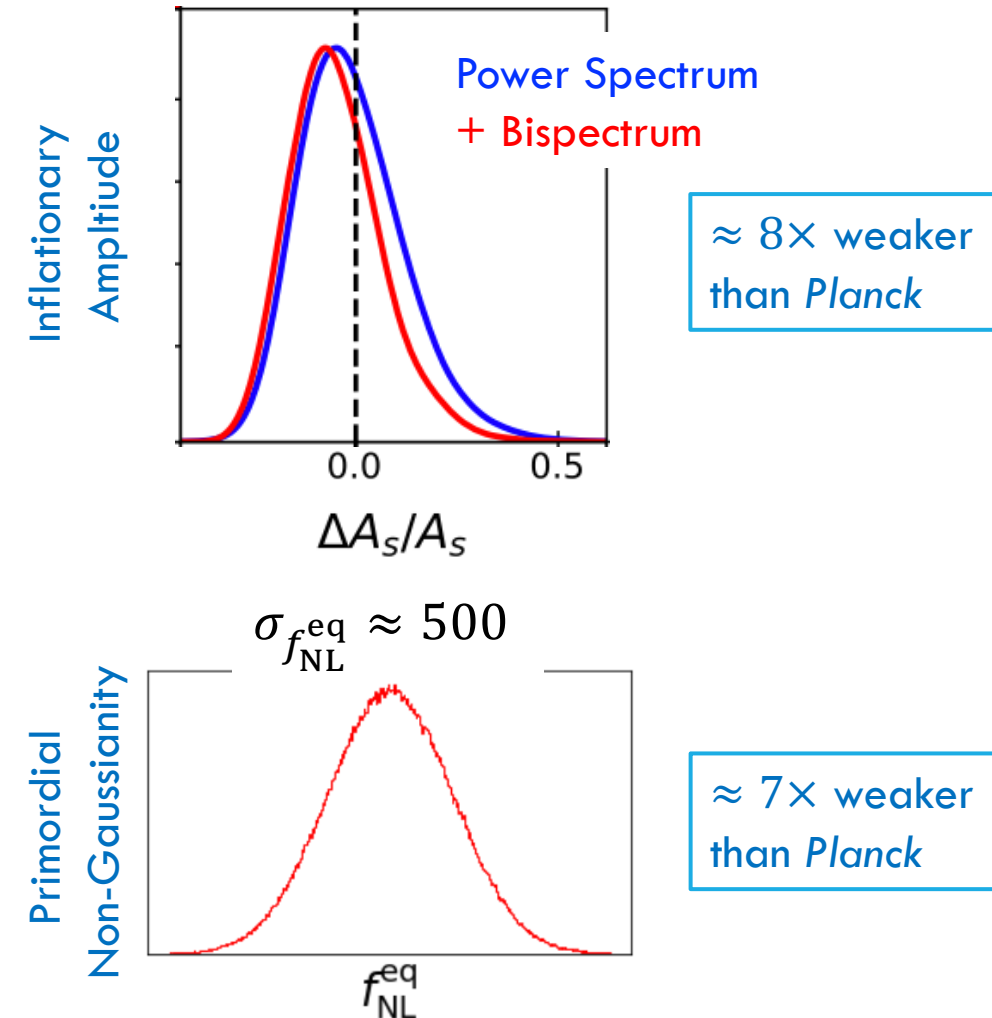


Theory Model

Cosmological  
Parameters

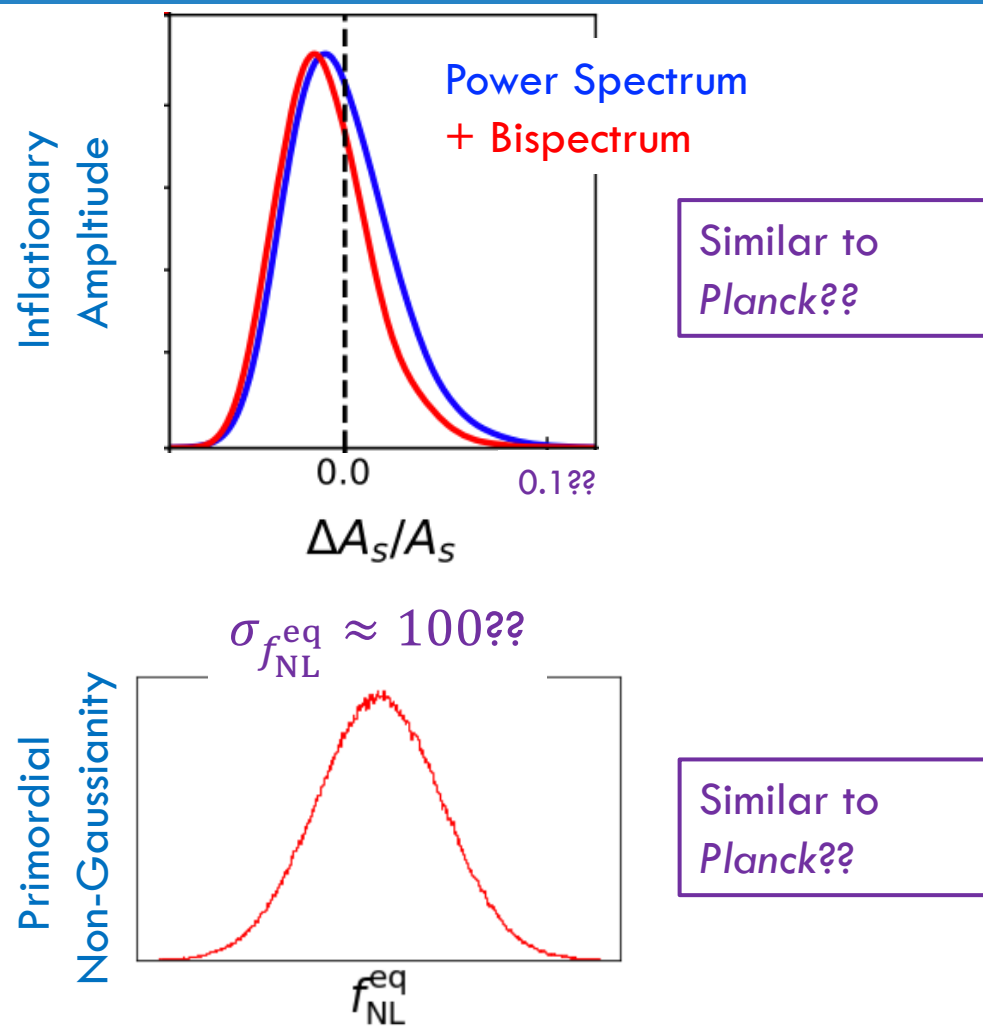
# WHAT WILL WE MEASURE?

- ▶ Tighter constraints on **cosmological** and **galaxy formation** parameters
  - ▶  $\sigma_8$  improves by 10%
  - ▶ Tidal bias improves by 50%
- ▶ Bounds on **all** flavors of **Primordial Non-Gaussianity**
  - ▶ First equilateral-type measurement from LSS



# WHAT'S NEXT FOR BISPECTRA?

- ▶ Improve bispectrum **modeling**
- ▶ Higher-order perturbation theory
- ▶ Add **redshift-space** information
- ▶ Better treatment of **fingers-of-God**
- ▶ Apply to **DESI** data
- ▶ **Pipelines** already available and **tested**
- ▶ Expect  $O(5)\times$  **stronger** constraints

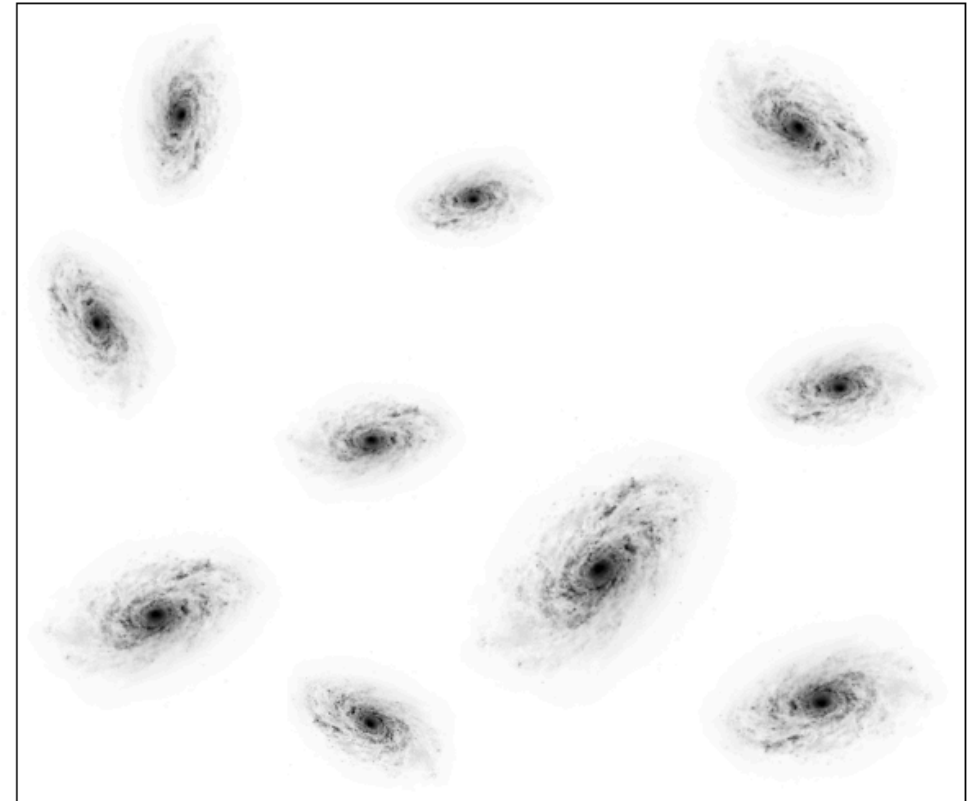




# CORRELATION FUNCTIONS = GALAXY COUNTS

$$\hat{\zeta}_{g,4}(r_1, r_2, r_3, \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_3, \hat{\mathbf{r}}_2 \cdot \hat{\mathbf{r}}_3) = \int \frac{d\mathbf{x}}{V} \int_{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \in \text{bins}} \delta_g(\mathbf{x}) \delta_g(\mathbf{x} + \mathbf{r}_1) \delta_g(\mathbf{x} + \mathbf{r}_2) \delta_g(\mathbf{x} + \mathbf{r}_3)$$

Measuring the 4PCF involves  
counting **quadruplets** of galaxies



# CORRELATION FUNCTIONS = GALAXY COUNTS

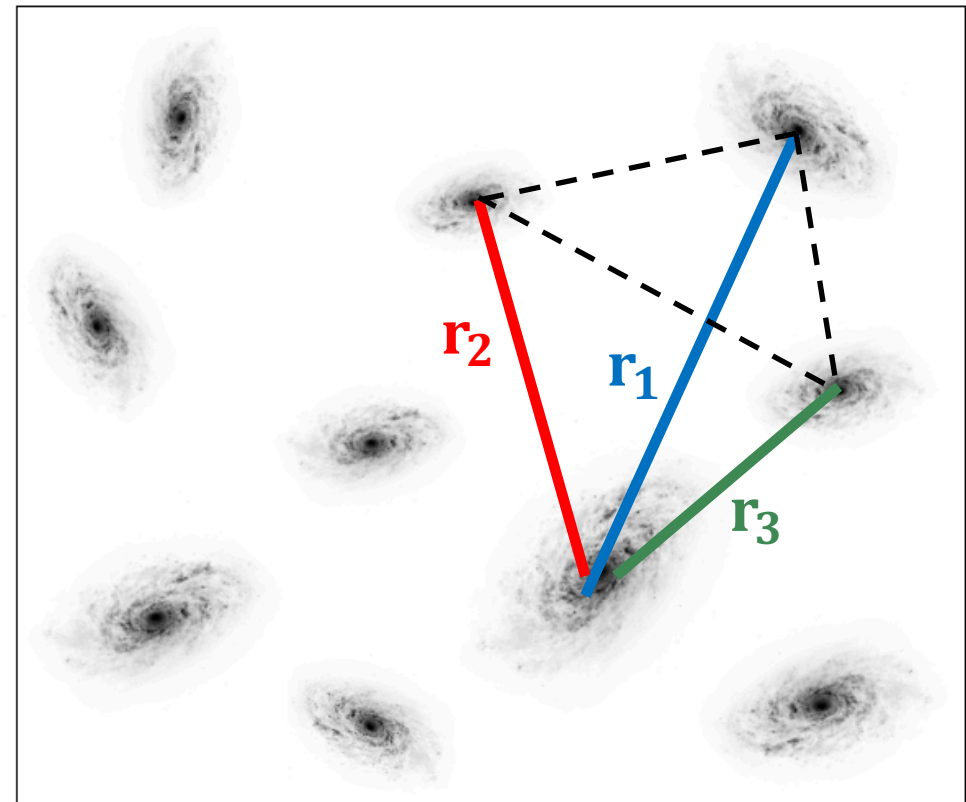
$$\hat{\zeta}_{g,4}(r_1, r_2, r_3, \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_3, \hat{\mathbf{r}}_2 \cdot \hat{\mathbf{r}}_3) = \int \frac{d\mathbf{x}}{V} \int_{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \in \text{bins}} \delta_g(\mathbf{x}) \delta_g(\mathbf{x} + \mathbf{r}_1) \delta_g(\mathbf{x} + \mathbf{r}_2) \delta_g(\mathbf{x} + \mathbf{r}_3)$$

Measuring the 4PCF involves counting **quadruplets** of galaxies

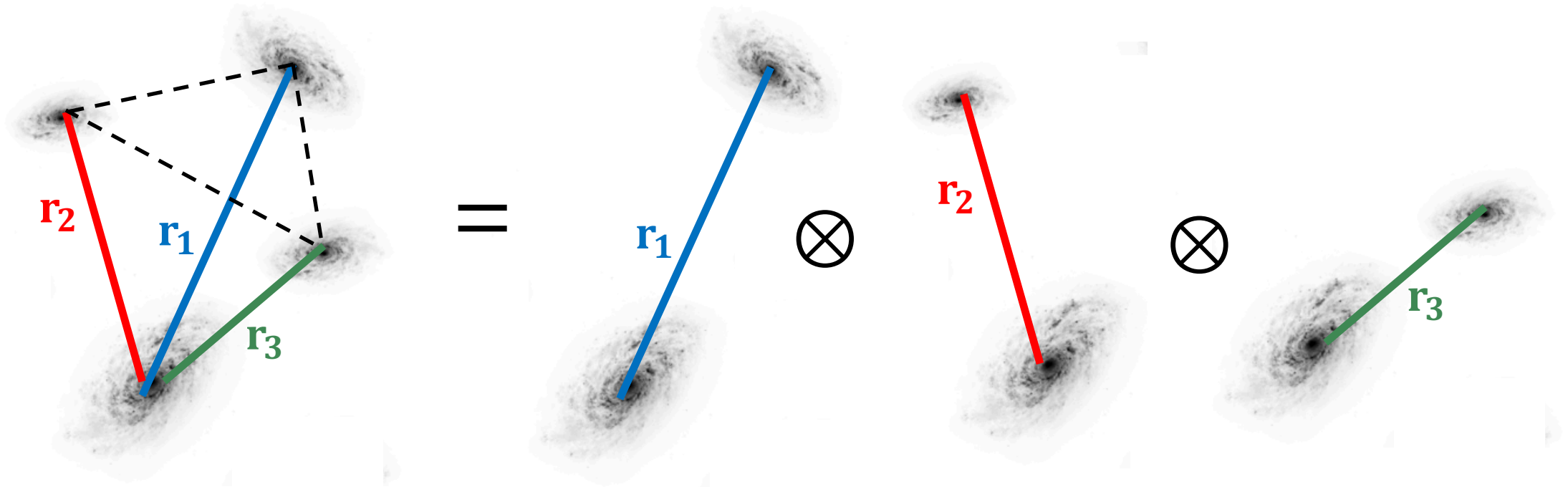
Total number of quadruplets:

$$\mathcal{O}(N_g^4)$$

**This is too many to count...**



# ONE TETRAHEDRON = THREE VECTORS



3 lengths + 3 angles

$(r_1, r_2, r_3, \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_3, \hat{\mathbf{r}}_2 \cdot \hat{\mathbf{r}}_3)$

1 length + 1 direction

$(r_1, \hat{\mathbf{r}}_1)$

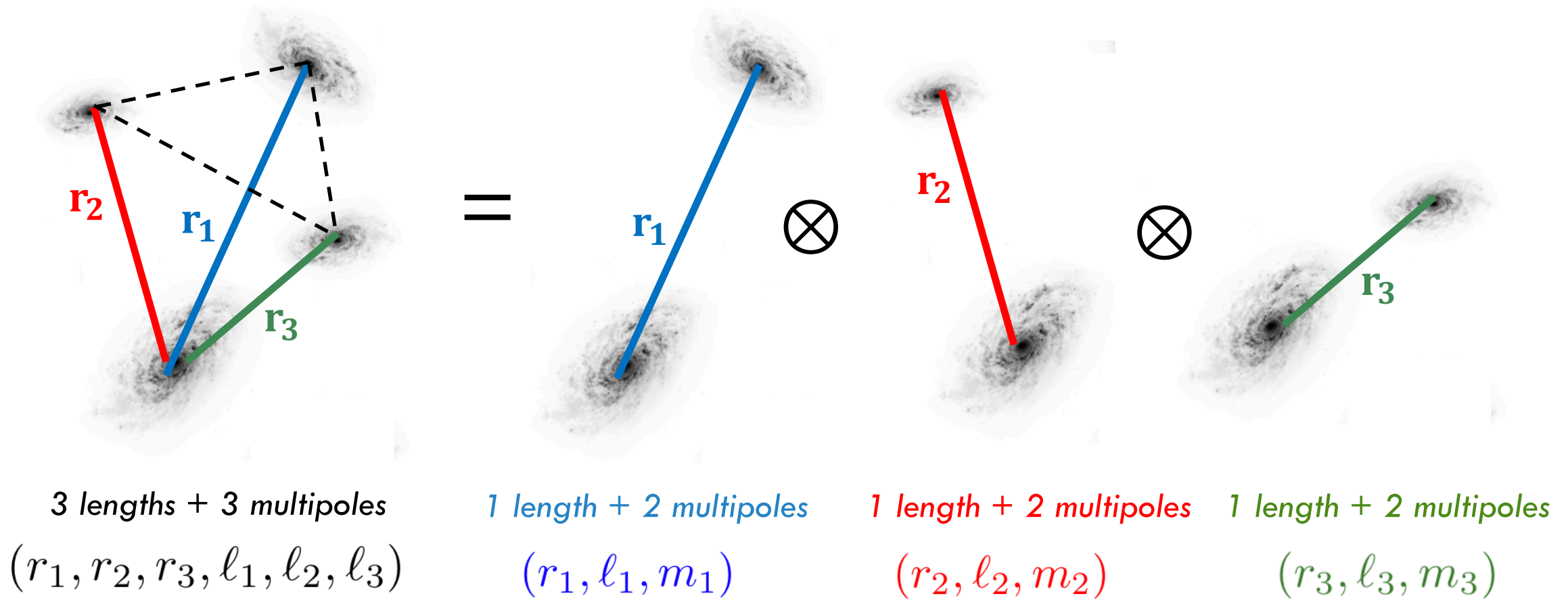
1 length + 1 direction

$(r_2, \hat{\mathbf{r}}_2)$

1 length + 1 direction

$(r_3, \hat{\mathbf{r}}_3)$

# ONE TETRAHEDRON = THREE VECTORS





# ANGULAR MOMENTUM BASIS

Expand 4PCF in basis of **isotropic functions**

$$\zeta_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \sum_{\ell_1 \ell_2 \ell_3} \zeta_{\ell_1 \ell_2 \ell_3}(r_1, r_2, r_3) \mathcal{P}_{\ell_1 \ell_2 \ell_3}(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_3)$$

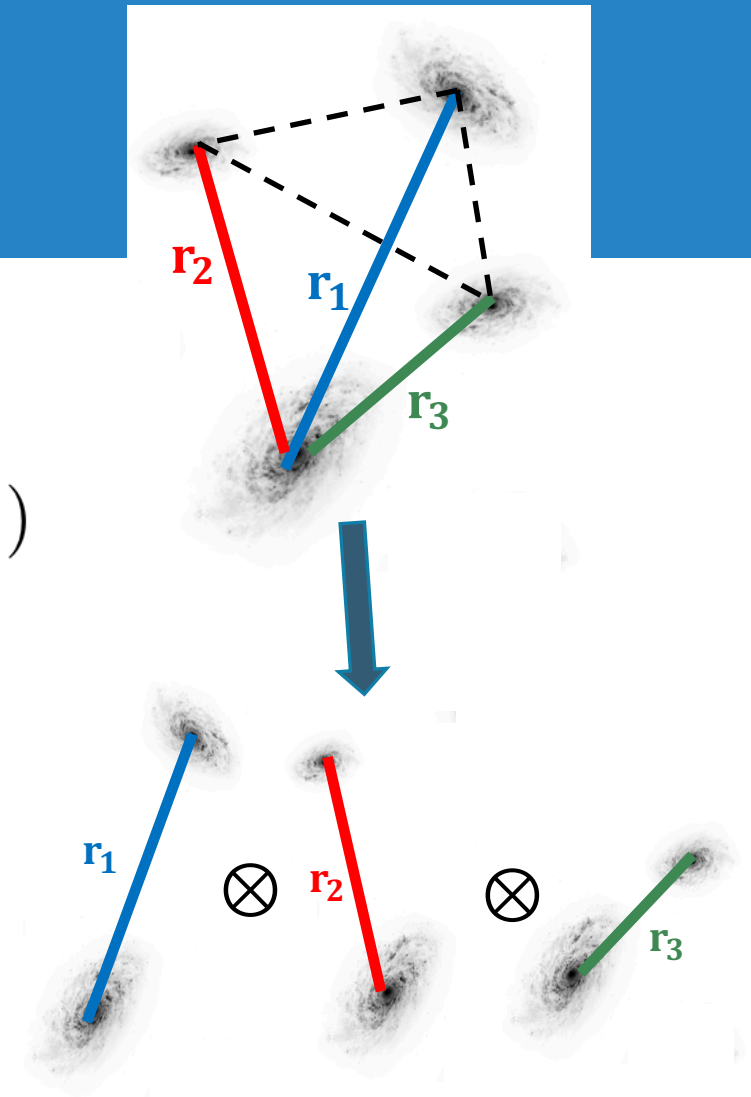
Coefficients

Basis Functions

Basis formed from **angular momentum addition** in 3D

$$\mathcal{P}_{\ell_1 \ell_2 \ell_3}(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_3) = \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} Y_{\ell_1 m_1}^*(\hat{\mathbf{r}}_1) Y_{\ell_2 m_2}^*(\hat{\mathbf{r}}_2) Y_{\ell_3 m_3}^*(\hat{\mathbf{r}}_3)$$

This is **separable** in  $\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_3$



# A SEPARABLE BASIS $\Rightarrow$ A QUADRATIC ESTIMATOR

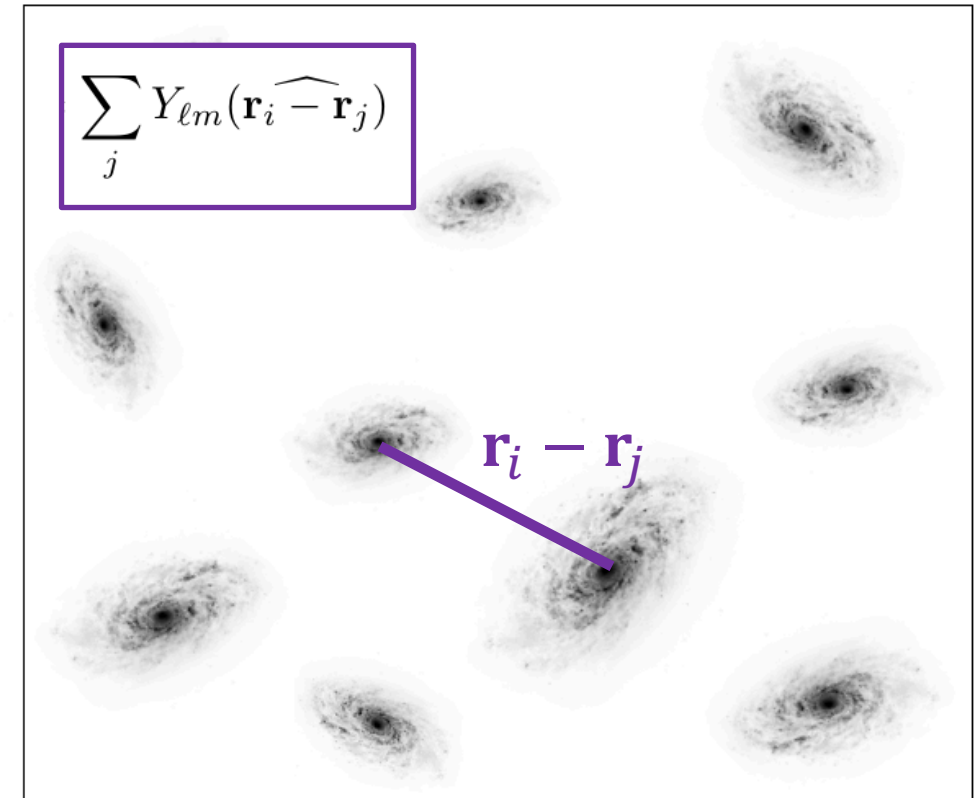
$$\hat{\zeta}_{\ell_1 \ell_2 \ell_3}(r_1, r_2, r_3) = \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \int d\mathbf{x} \delta_g(\mathbf{x}) \left[ \int_{\mathbf{r}_1} \delta_g(\mathbf{x} + \mathbf{r}_1) Y_{\ell_1 m_1}(\hat{\mathbf{r}}_1) \right] \left[ \int_{\mathbf{r}_2} \delta_g(\mathbf{x} + \mathbf{r}_2) Y_{\ell_2 m_2}(\hat{\mathbf{r}}_2) \right] \left[ \int_{\mathbf{r}_3} \delta_g(\mathbf{x} + \mathbf{r}_3) Y_{\ell_3 m_3}(\hat{\mathbf{r}}_3) \right]$$

The estimator **factorizes** into **independent** pieces

To compute the 4PCF: count *pairs* of galaxies

Total number of pairs:  $\mathcal{O}(N_g^2)$

**This can be computed!**




# ENCORE: ULTRA-FAST N-POINT FUNCTIONS


- ▶ Public C++/CUDA code
- ▶ Corrects for **survey geometry**
- ▶ Requires ~ 10 CPU-hours to compute 4PCF of current data
- ▶ Extends to curved geometries, higher-dimensions, 5PCF, 6PCF and beyond...


See [GitHub.com/oliverphilcox/encore](https://github.com/oliverphilcox/encore)


oliverphilcox/  
**encore**





encore: Efficient isotropic 2-, 3-, 4-, 5- and 6-point correlation functions in C++ and CUDA

 2  
Contributors

 0  
Issues


 4  
Stars

 1  
Fork



oliverphilcox/encore

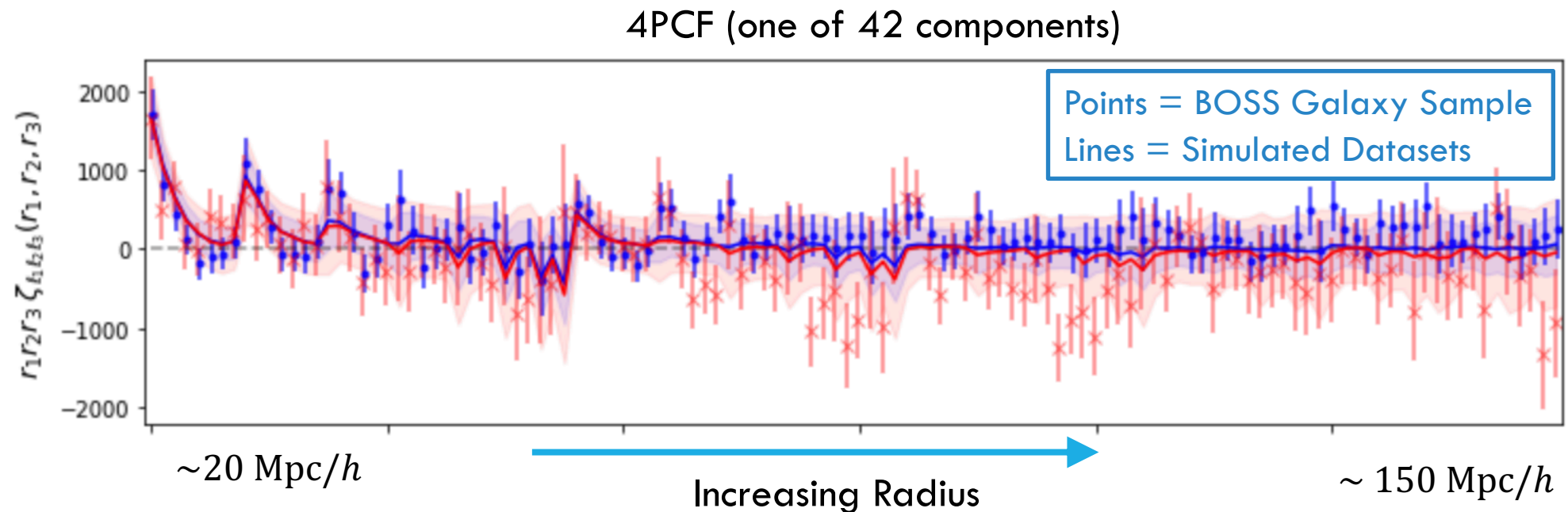
encore: Efficient isotropic 2-, 3-, 4-, 5- and 6-point correlation functions in C++ and CUDA - oliverphilcox/encore

 [github.com](https://github.com/oliverphilcox/encore)

# MEASURING THE 4-POINT FUNCTION

Compute the (connected) 4PCF from  $\sim 10^6$  **BOSS galaxies**

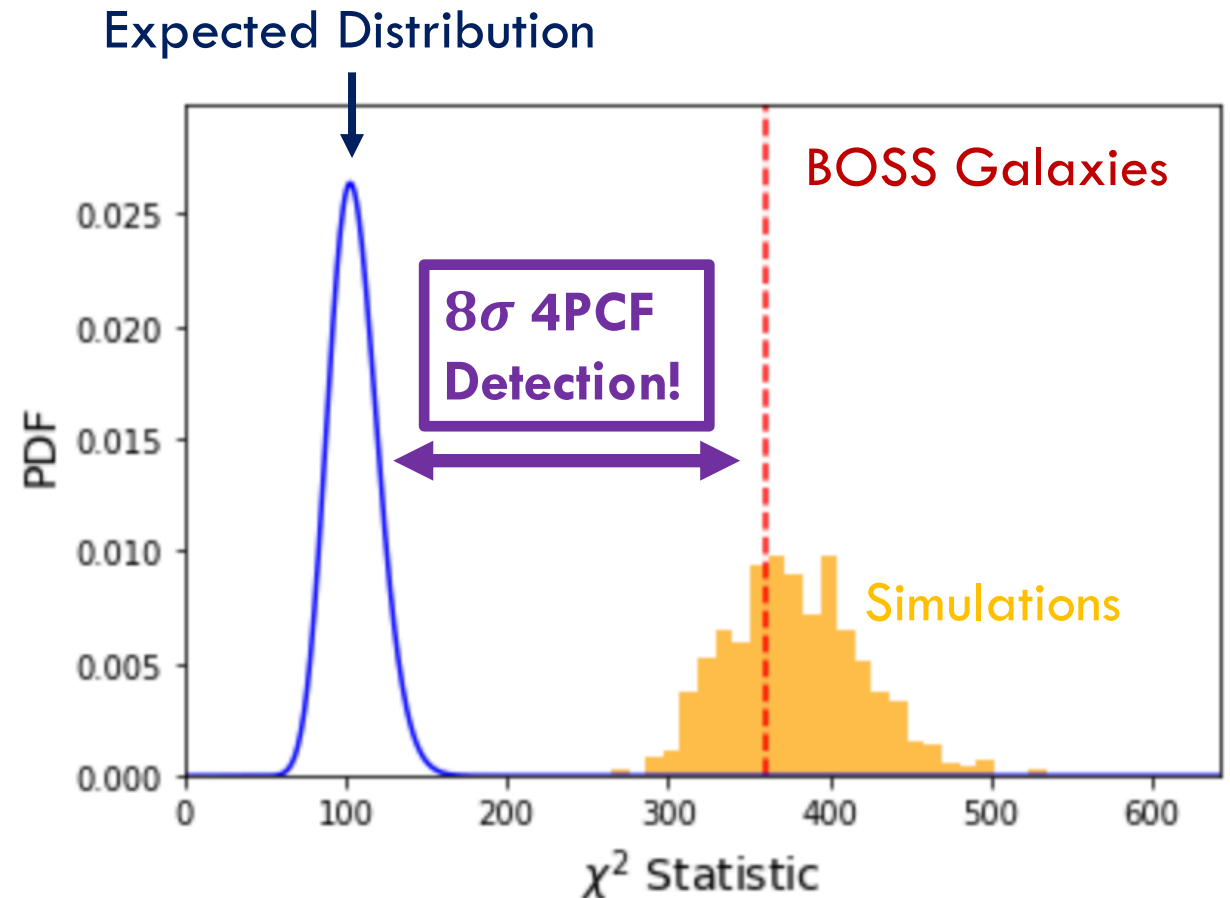
**Do we detect a signal?**





# CAN WE DETECT THE GRAVITATIONAL 4PCF?

- ▶ Perform a  $\chi^2$ -test to search for a **gravitational** 4PCF
- ▶ Null Hypothesis: **4PCF = 0**.
- ▶ **Strong** detection of non-Gaussianity!



# WHAT'S NEXT FOR THE 4-POINT FUNCTION?

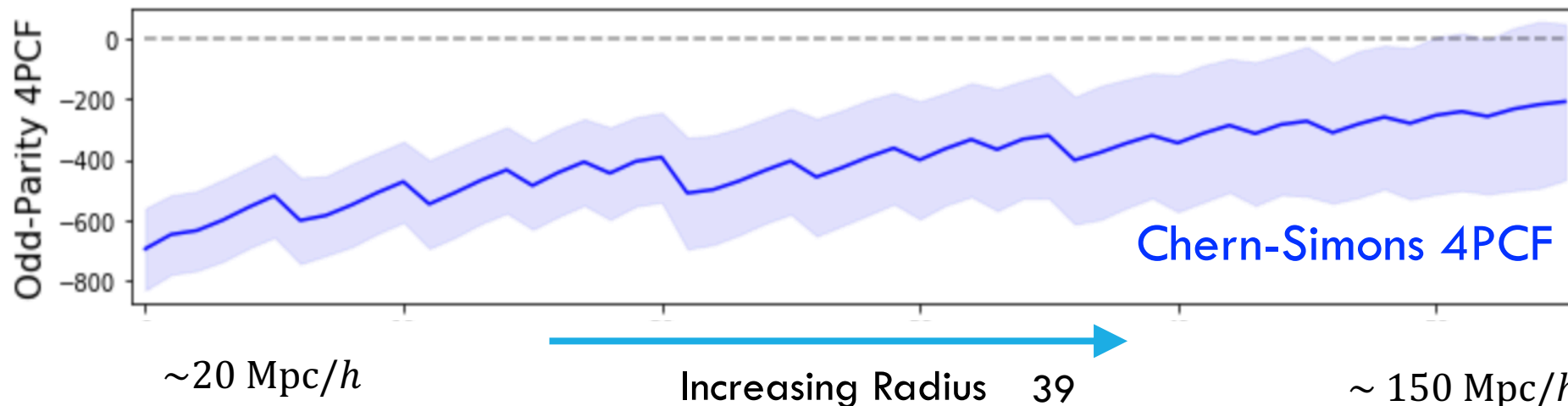
▷ Create a **theory** model and **quantify** information content:

▷ Allows  $\Lambda$ **CDM** information to be extracted

▷ Search for **parity-violating** physics in the BOSS 4PCF

$$\zeta_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) - \mathbb{P}[\zeta_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)]$$

▷ Apply to **DESI** data [2× higher precision] and combine with the **CMB**



Cahn+21, **Philcox** (in prep.)

A visualization of the cosmic web, showing a complex network of blue filaments and orange galaxy clusters against a dark background.

arXiv

[2012.09389](#)

[2105.08722](#)

[2106.10278](#)

[2107.06287](#)

[2108.01670](#)

[2110.10161](#)

## Contact

[ohp2@cantab.ac.uk](mailto:ohp2@cantab.ac.uk)

[@oliver\\_philcox](#)

# CONCLUSIONS

- Non-Gaussian statistics:
  1. **Sharpen** cosmological constraints
  2. Probe **non-standard** physics in the early Universe
- **Fast** and **accurate** estimators now available
- Extract **more** information from LSS surveys **without** additional cost