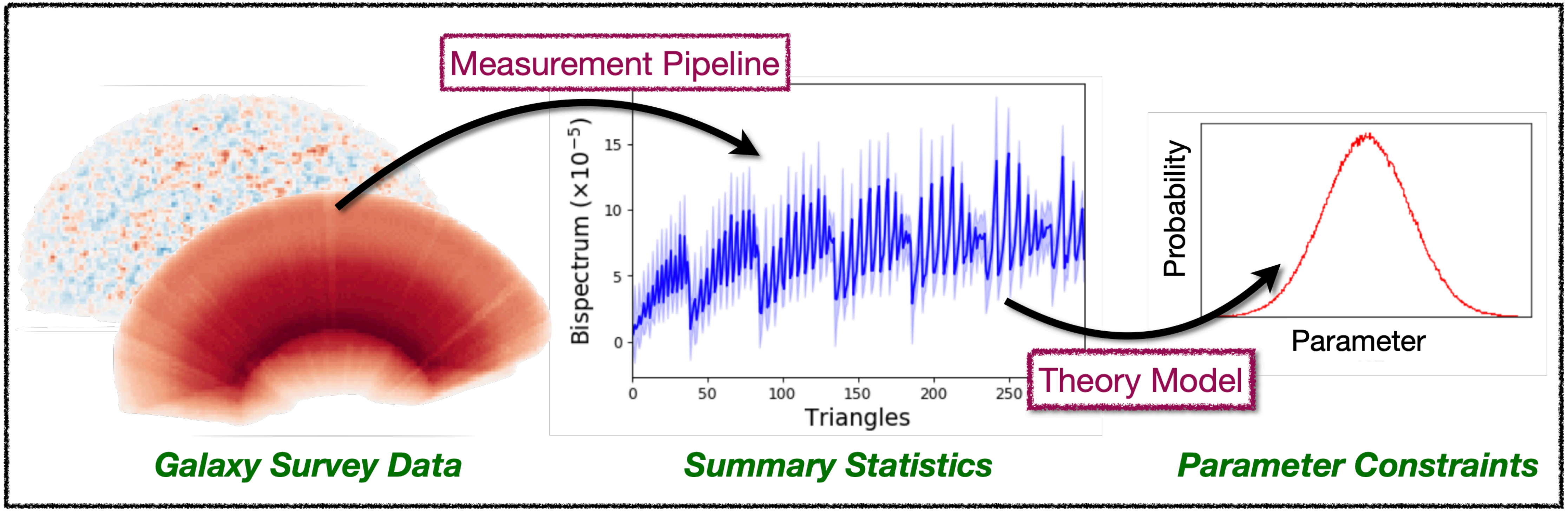
PRINCETON UNIVERSITY

Probing Fundamental Cosmology With Galaxy Surveys

IAS INSTITUTE FOR ADVANCED STUDY



Galaxy Survey Data

Summary Statistics

Parameter Constraints

OLIVER PHILCOX

Final Public Oral, *August 31st 2022*

Advisors: David Spergel & Matias Zaldarriaga

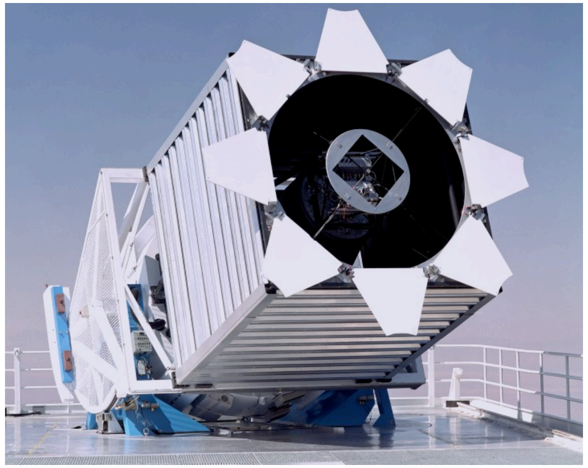
Committee: David Spergel, Matias Zaldarriaga & Jo Dunkley

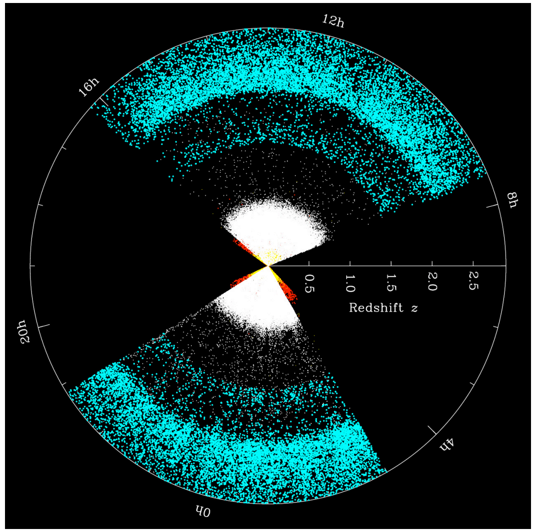
LARGE-SCALE STRUCTURE COSMOLOGY

- DESI, Euclid, SPHEREx will measure $\sim 10^8$ galaxy positions in the next decade

- New data will be **far** better than anything before

- But the proposed analysis techniques are the **same**…



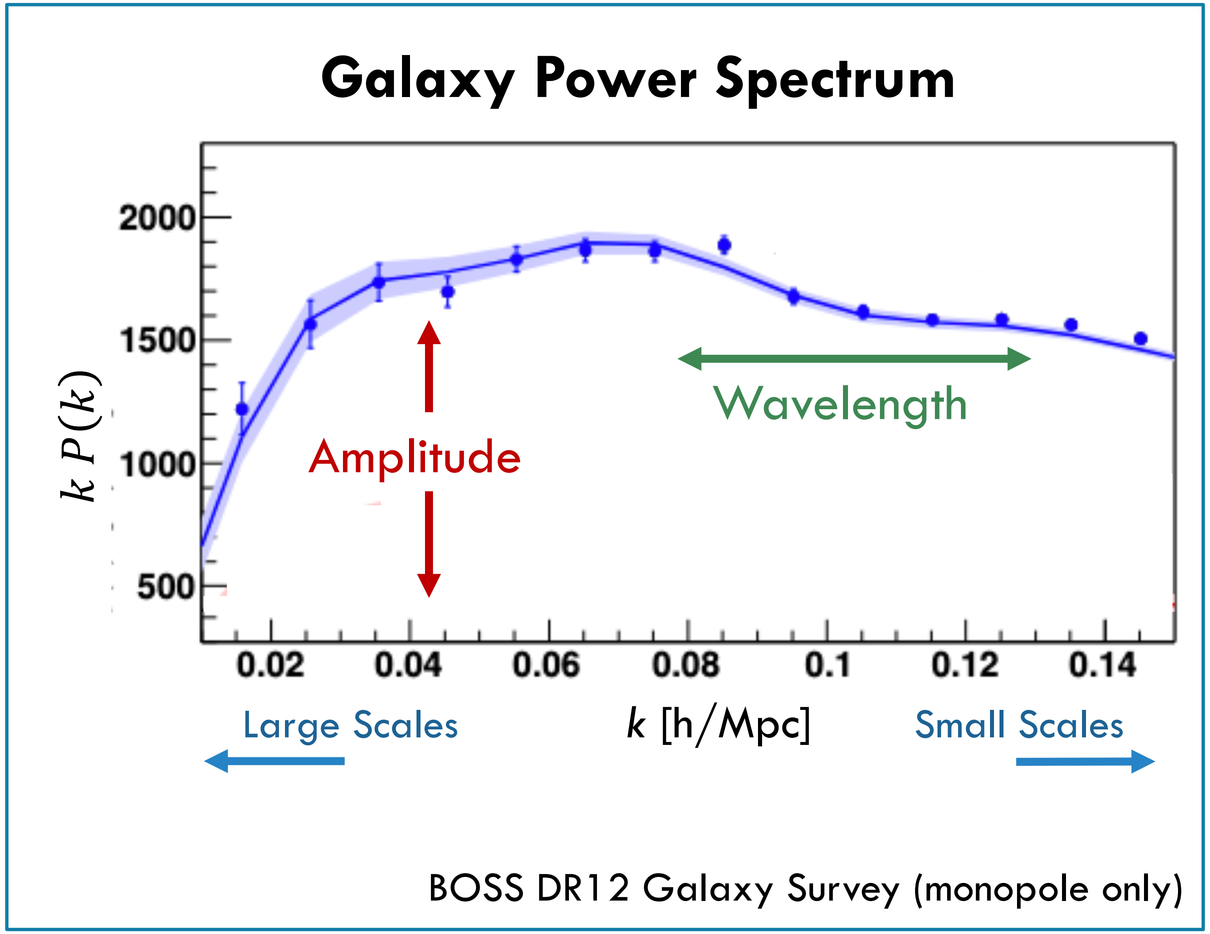


SDSS-III

CONVENTIONAL LSS ANALYSES

- Compress the 10^6 galaxy positions to a **power spectrum,** $\langle\delta_g(\boldsymbol{k})\delta_g^*(\boldsymbol{k})\rangle$

- Use a **scaling analysis** to measure:
  - Overall **amplitude** (= primordial amplitude)
  - **Wiggle** positions (= BAO feature)



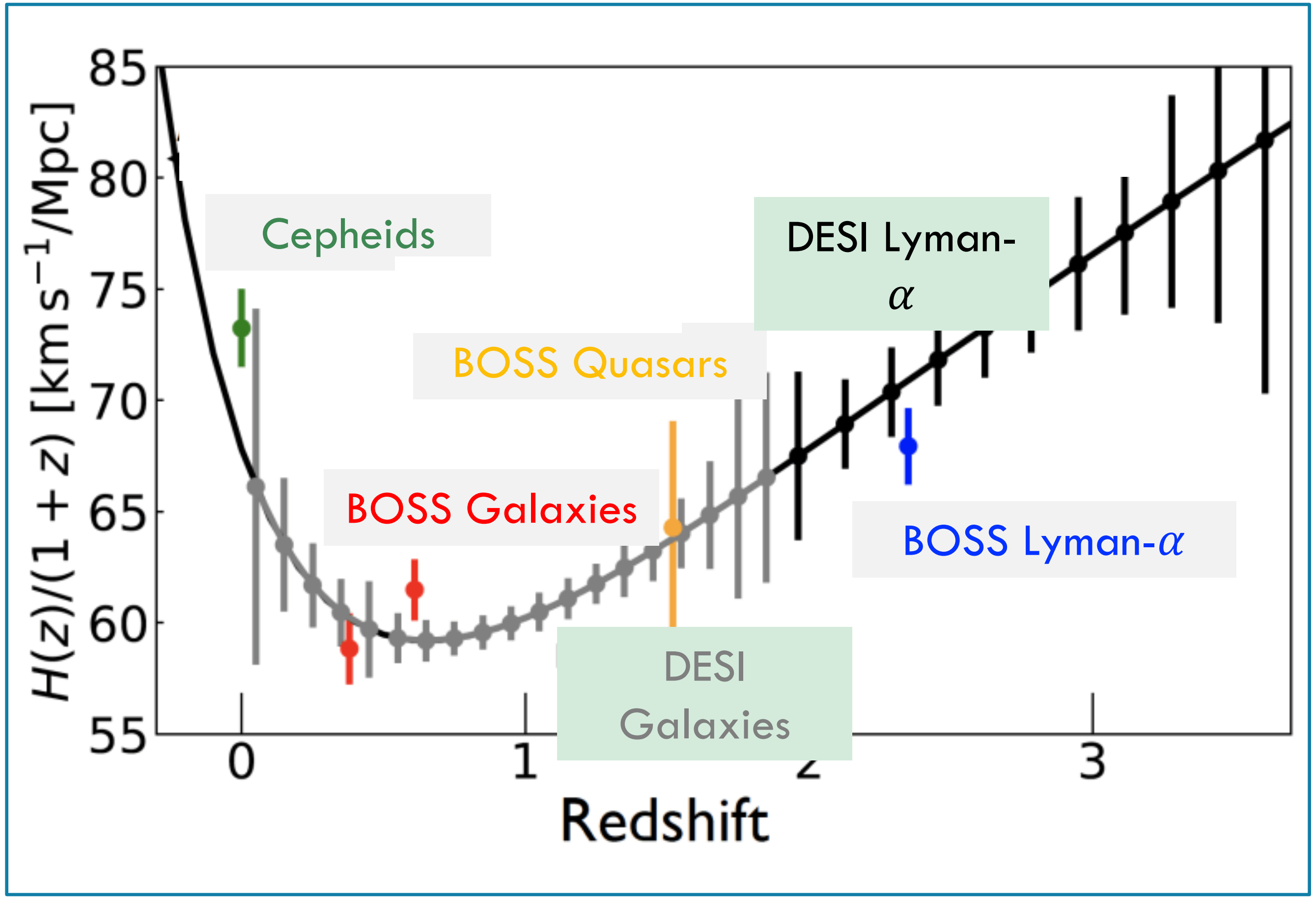
e.g. Beutler+17, Gil-Marin+15,17, Alam+20

CONVENTIONAL LSS ANALYSES

- Compress the 10^6 galaxy positions to a **power spectrum,** $\langle \delta_g(\boldsymbol{k})\delta_g^*(\boldsymbol{k})\rangle$

- Use a **scaling analysis** to measure:
  - Overall **amplitude** (= primordial amplitude)
  - **Wiggle** positions (= BAO feature)

- Robust way to constrain **structure growth** $f\sigma_8(z)$, and **expansion history** $H(z), D_A(z)$

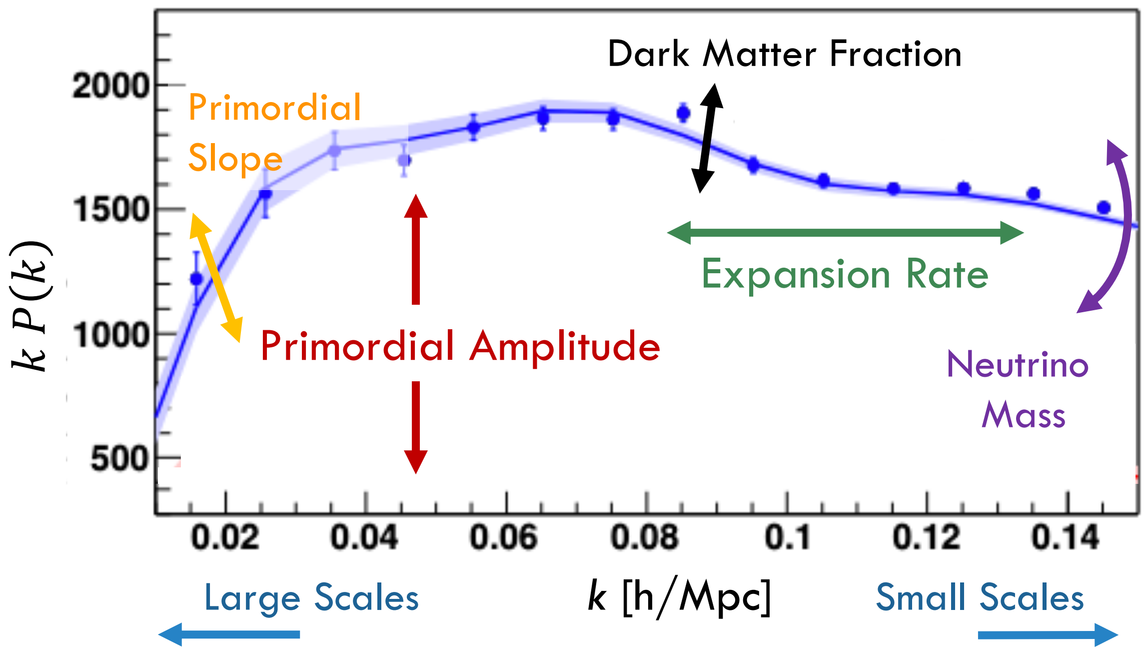


BEYOND CONVENTIONAL ANALYSES

Three opportunities for improvement:

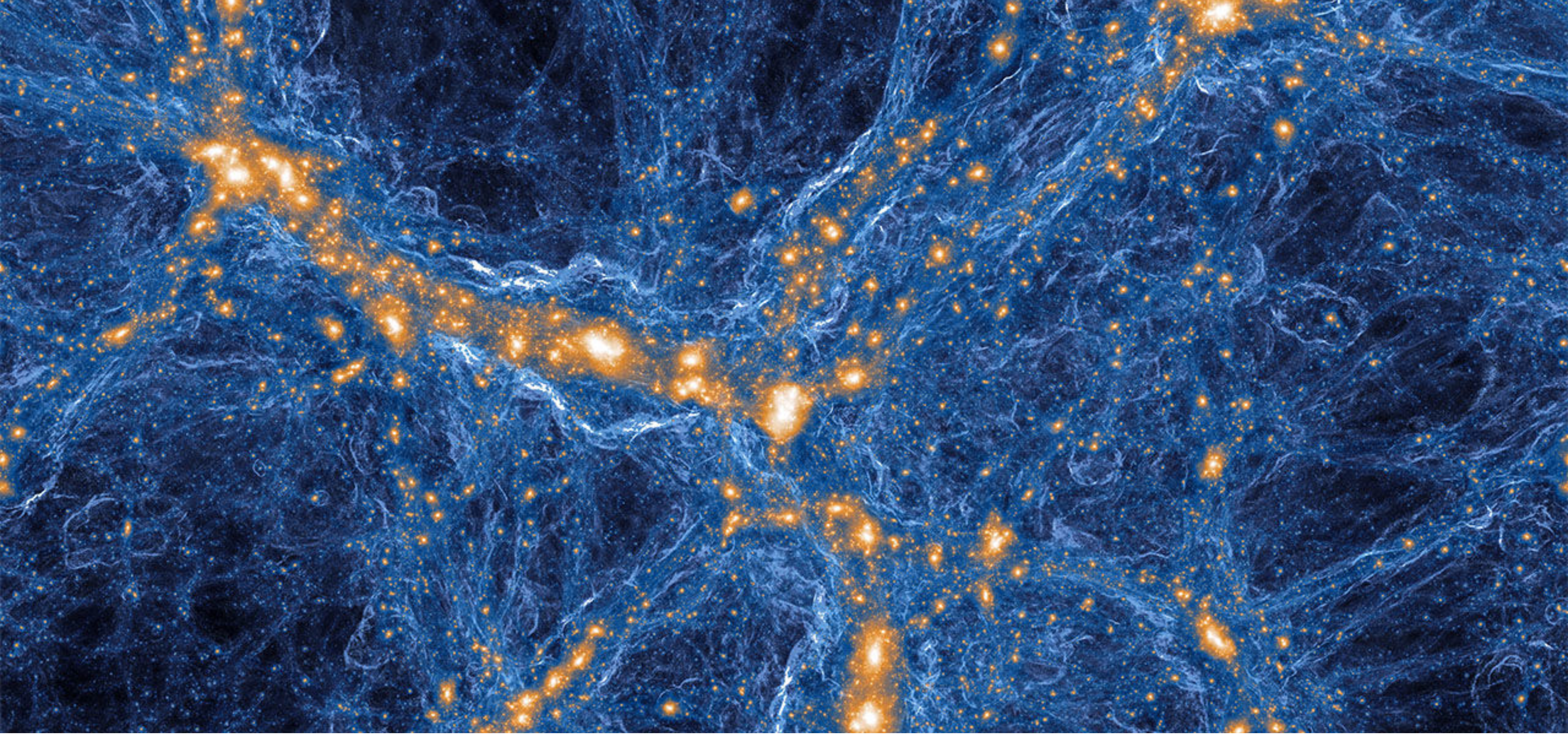
1. Measure ΛCDM parameters **directly**
2. Include statistics **beyond** the 2-point function
3. Constrain **extensions** to ΛCDM

Galaxy Power Spectrum



BOSS DR12 Galaxy Survey, Beutler+17

e.g. Ivanov+19,20, d'Amico+19,20, **Philcox**+20ab, Chen+21, Kobayashi+21



PART I: How to Measure Summary Statistics

WINDOWED STATISTICS

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

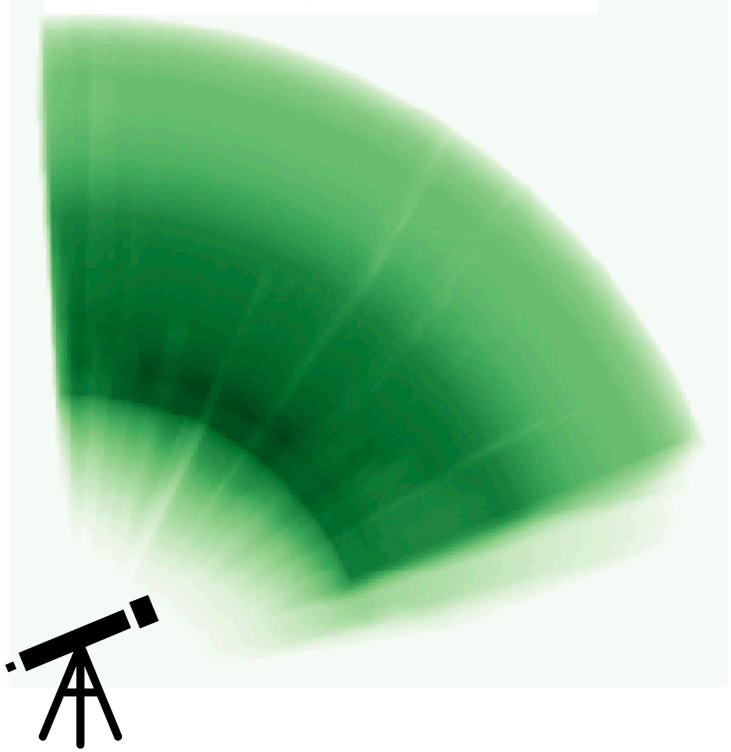
$$\delta_g(\mathbf{r}) \to W(\mathbf{r})\delta_g(\mathbf{r}) \qquad \delta_g(\mathbf{k}) \to \int \frac{d\mathbf{p}}{(2\pi)^3} W(\mathbf{k}-\mathbf{p})\delta_g(\mathbf{p})$$

Window Function

The measured statistics are **convolutions** (cf. pseudo-C_ℓ)

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

Survey Mask



Solution: Forward-model, *i.e.* convolve the **theory model**

This is hard beyond 2pt functions!

Gil-Marin 16, **Philcox** 21 (see also Pardede+22)

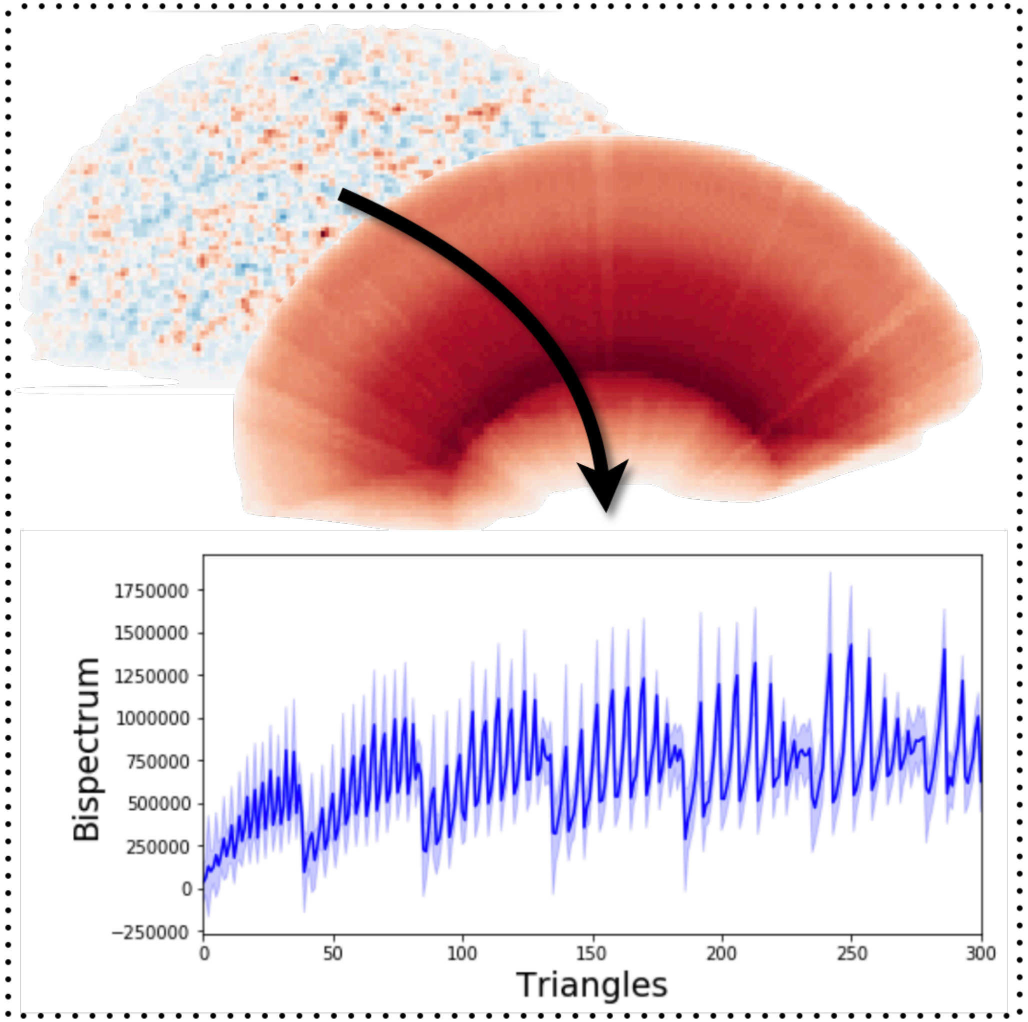
UNWINDOWED STATISTICS

Alternative: estimate **unwindowed** statistics

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

- ▷ Derive **maximum-likelihood** estimators for the **true** power spectrum and bispectrum
- ▷ Effectively **deconvolves** window → easier **modeling**

$$\boxed{\nabla_{B_g} L[\text{data}|B_g]=0 \quad \Rightarrow \quad \widehat{B}_g=\cdots}$$



See GitHub.com/oliverphilcox/Spectra-Without-Windows

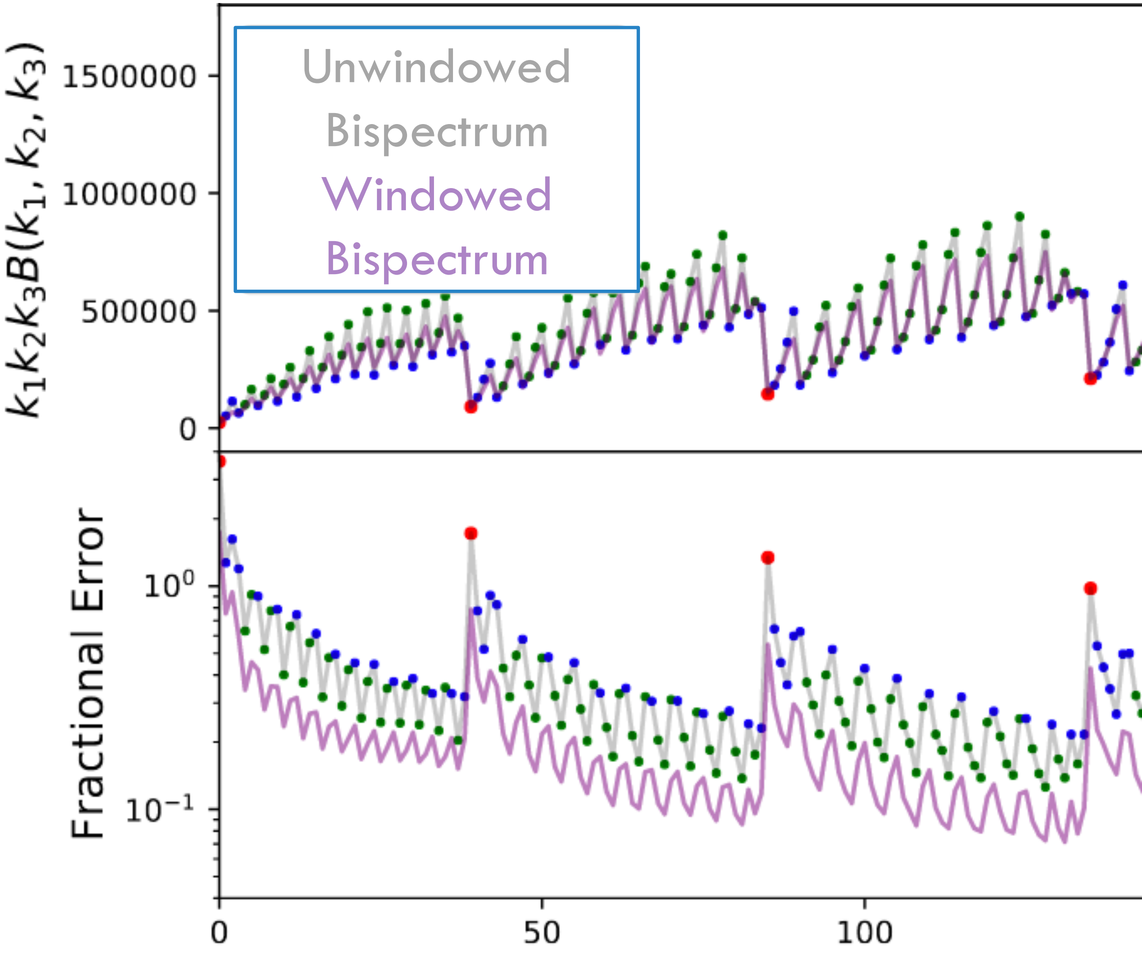
Tegmark 96, **Philcox** 20, 21

UNWINDOWED STATISTICS

Properties of the **new estimators:**

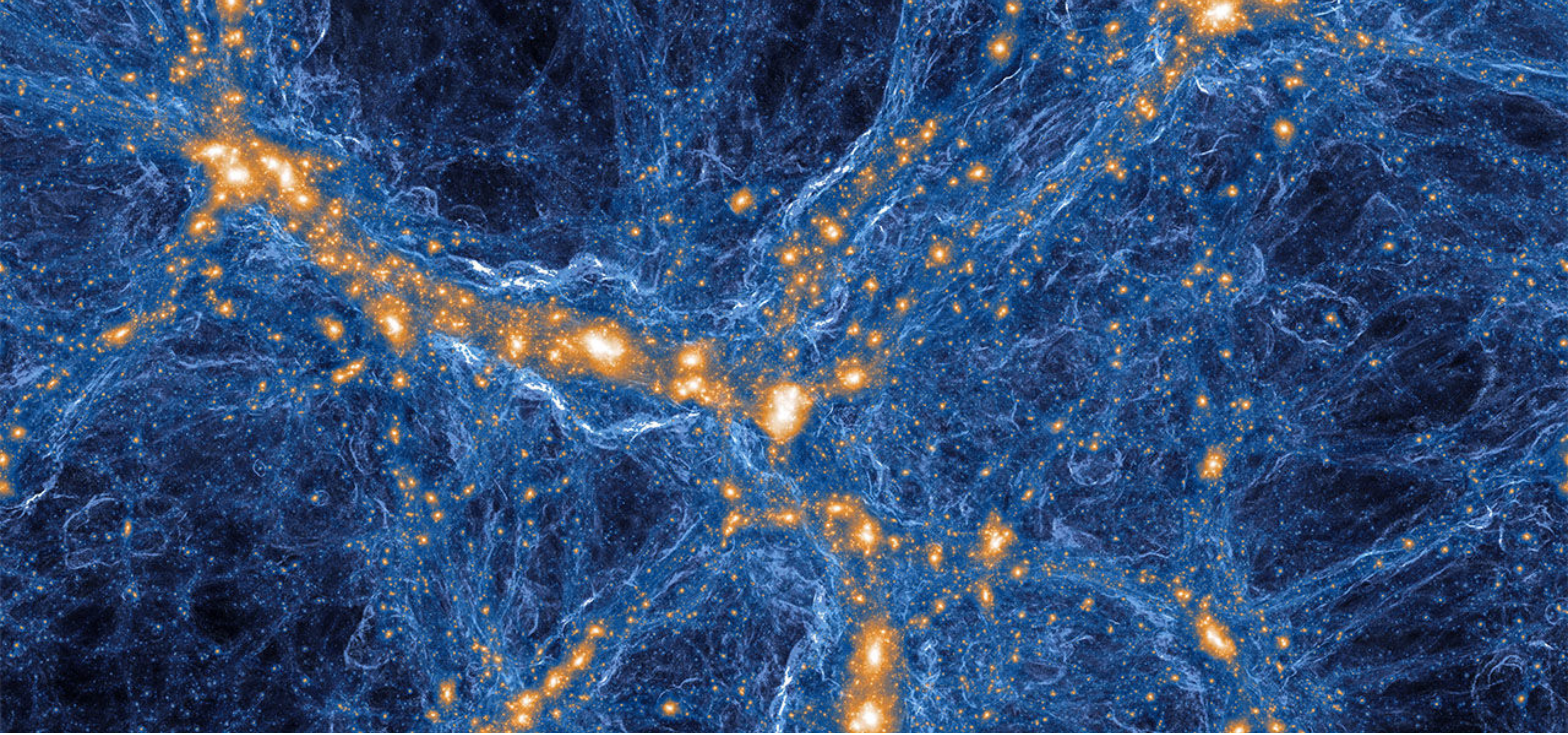
1. Unbiased
2. Minimum variance [as $B(k_1, k_2, k_3) \to 0$]
3. Window-free [effectively a deconvolution]

- ▷ Now being used within *Euclid*
- ▷ Could be extended to **trispectra**?



See GitHub.com/oliverphilcox/BOSS-Without-Windows

Philcox 20, 21



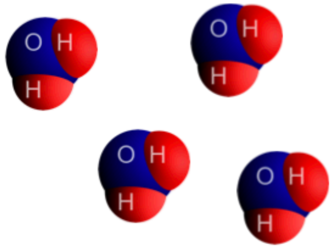
PART II: How to Model Summary Statistics

THE EFFECTIVE FIELD THEORY OF LARGE SCALE STRUCTURE

- **Analytic** theory for $\delta(\mathbf{x})$, based on the non-ideal **fluid equations**

$$\boxed{\dot{v}^i + H v^i + v^j \partial_j v^i = \frac{1}{\rho} \partial_j \tau^{ij}}$$

- A <u>controlled</u> **Taylor series** in k/k_{NL}



- **Major Ingredient:** *self-consistent back-reaction* of small-scale physics on large-scale modes

large scales

e.g. Baumann, Carrasco, Assassi, Senatore, Zaldarriaga, etc.

HOW TO MODEL BISPECTRA AT O(1)

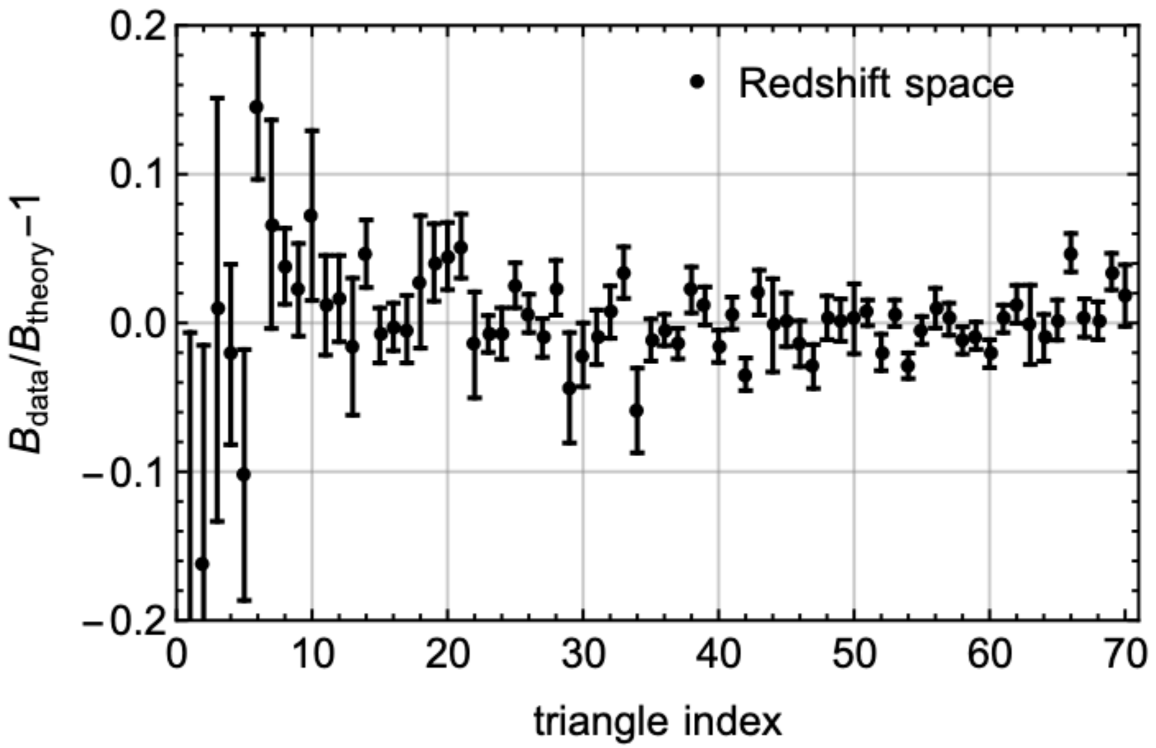
Tree-Level Galaxy Bispectrum

$$B_{\rm ggg}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \; 2Z_2(\mathbf{k}_1,\mathbf{k}_2)Z_1(\mathbf{k}_1)Z_1(\mathbf{k}_2)P_{\rm lin}(k_1)P_{\rm lin}(k_2)$$
$$+ P_\epsilon(k_2)2d_1\left(d_2b_1 + d_1 f\mu_1^2\right)Z_1(\mathbf{k}_1)P_{\rm lin}(k_1) + \text{cycl.} + d_1^3 B_\epsilon(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$$

- Second-order galaxy bias
- Large-scale displacements
- Coordinate transformations
- Fingers-of-God
- Stochasticity

12 physical parameters

Accurate up to $k_{\max} = \mathbf{0.08}\ h/\mathrm{Mpc}$



See GitHub.com/michalychforever/CLASS-PT

Ivanov, **Philcox**+21, see also d'Amico++

HOW TO MODEL BISPECTRA AT O(2)

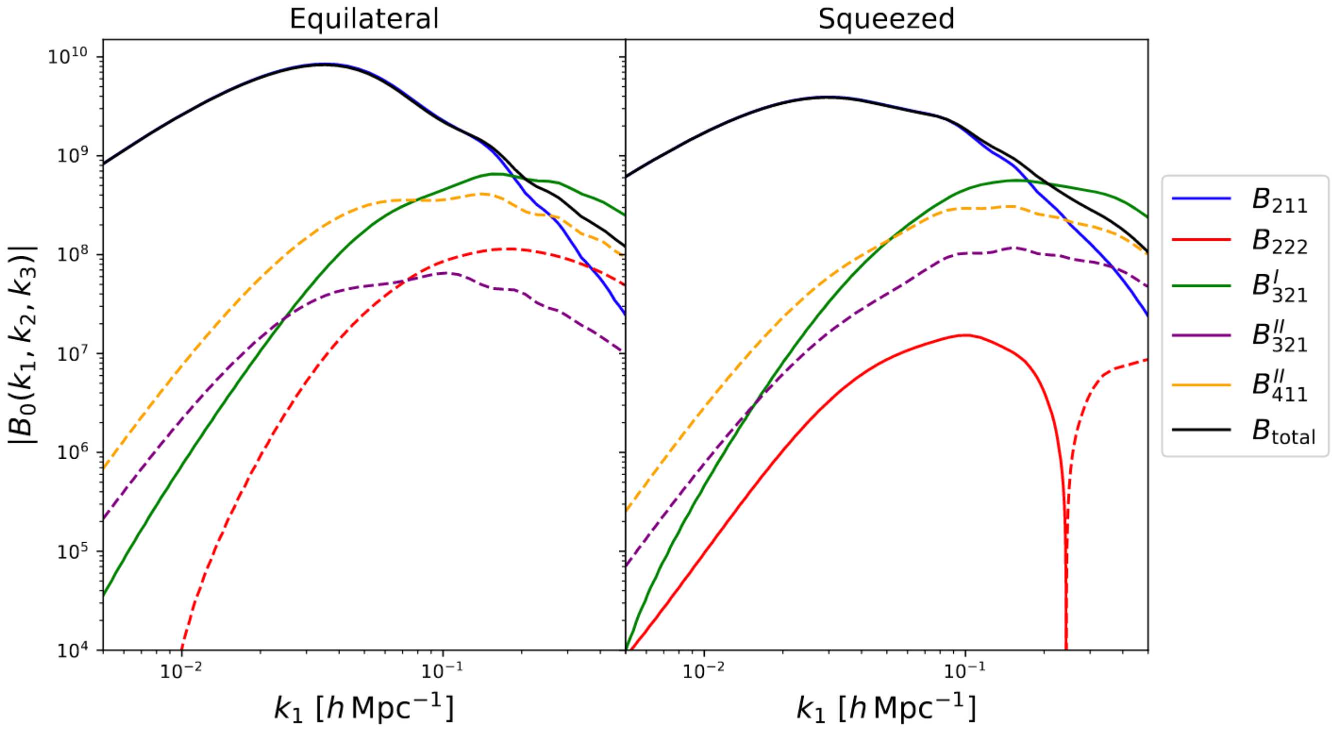
One-Loop Galaxy Bispectrum

$$B_{1-\mathrm{loop}}(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = B_{211} + \left[B_{222} + B_{321}^{I} + B_{321}^{II} + B_{411}\right] + B_{\mathrm{ct}} + B_{\mathrm{stoch}},$$

- Fourth-order galaxy bias
- Counterterms
- Large-scale displacements
- Coordinate transformations
- Fingers-of-God
- Stochasticity

44 physical parameters
(not independent)

Accurate up to $k_{\max} = \mathbf{0.15}\ h/\mathrm{Mpc}$



See GitHub.com/oliverphilcox/OneLoopBispectrum

Philcox+22, see also d'Amico++

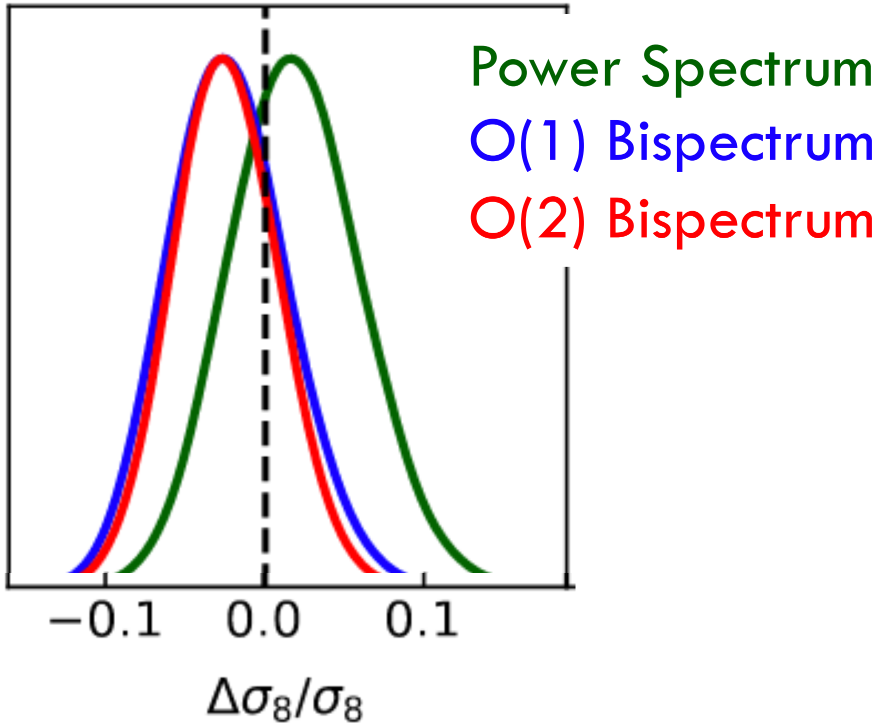
HOW TO MODEL BISPECTRA AT O(2)

- More loops → **many** more parameters
- More loops → **little** increase in cosmological parameter constraints

Is this a problem?

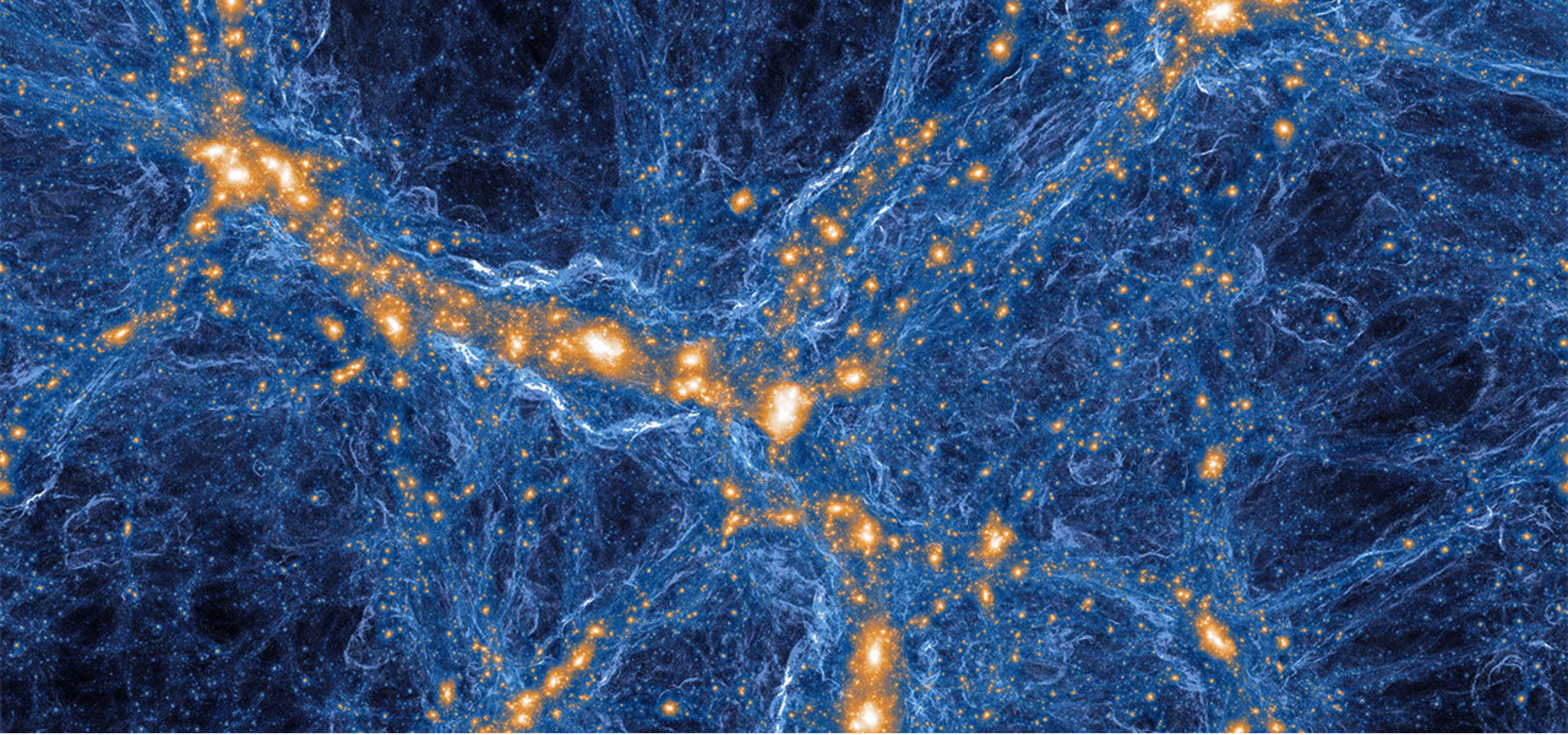
To make better use of loop corrections we need:

- Better **priors** on higher-order parameters
- Better **statistics**, e.g., bispectrum multipoles



See GitHub.com/oliverphilcox/OneLoopBispectrum

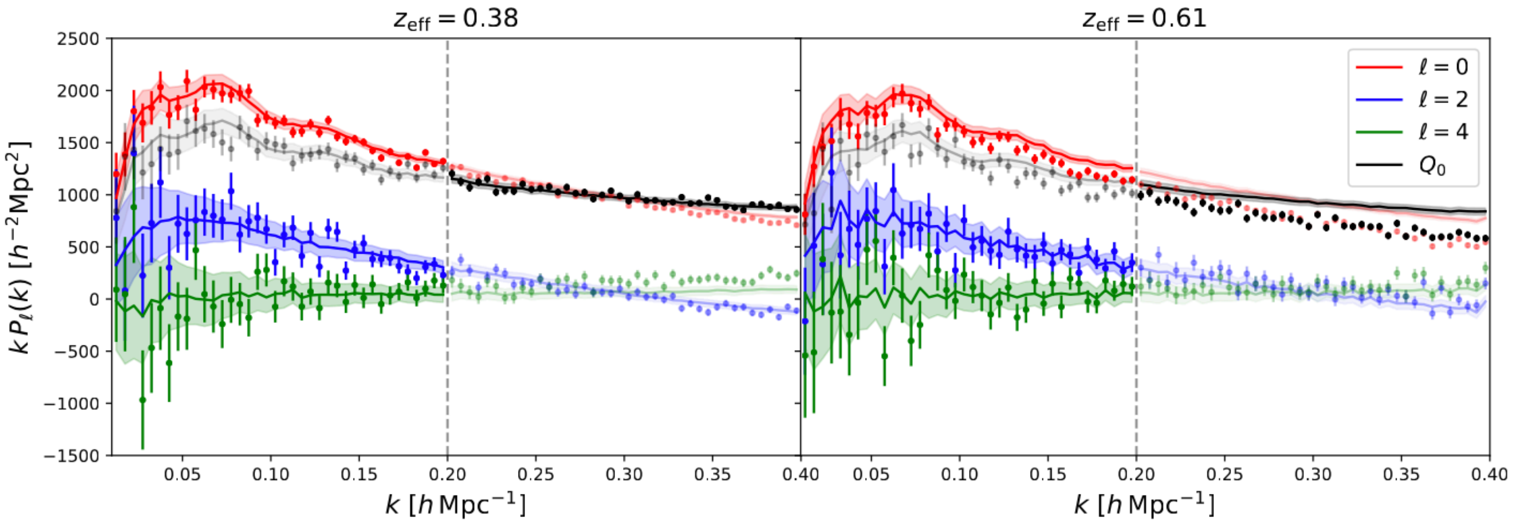
Philcox+22, see also d'Amico++



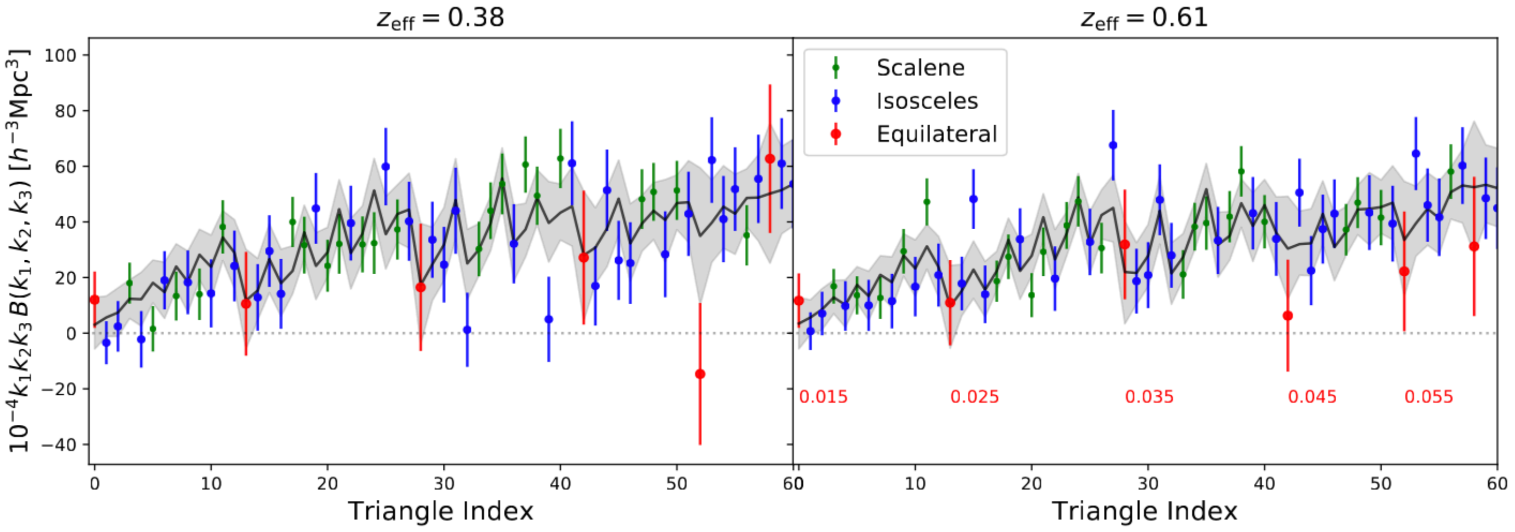
PART III: How to Interpret Summary Statistics

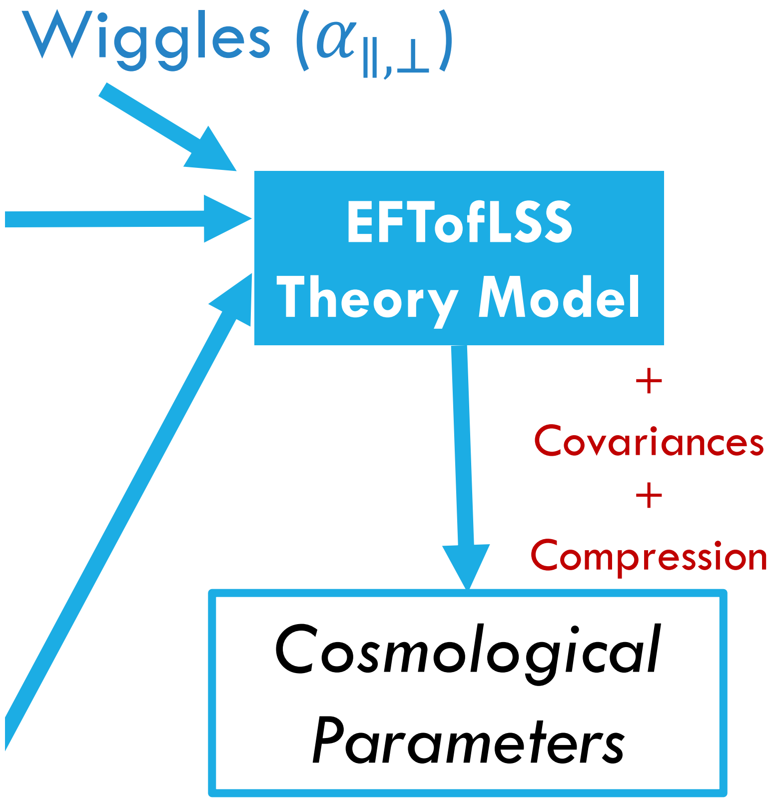
THE *UNOFFICIAL* BOSS DR12 ANALYSIS

$P_\ell + Q_0$



Bispectrum





Wiggles ($\alpha_{\parallel,\perp}$)

EFTofLSS Theory Model

\+

Covariances

\+

Compression

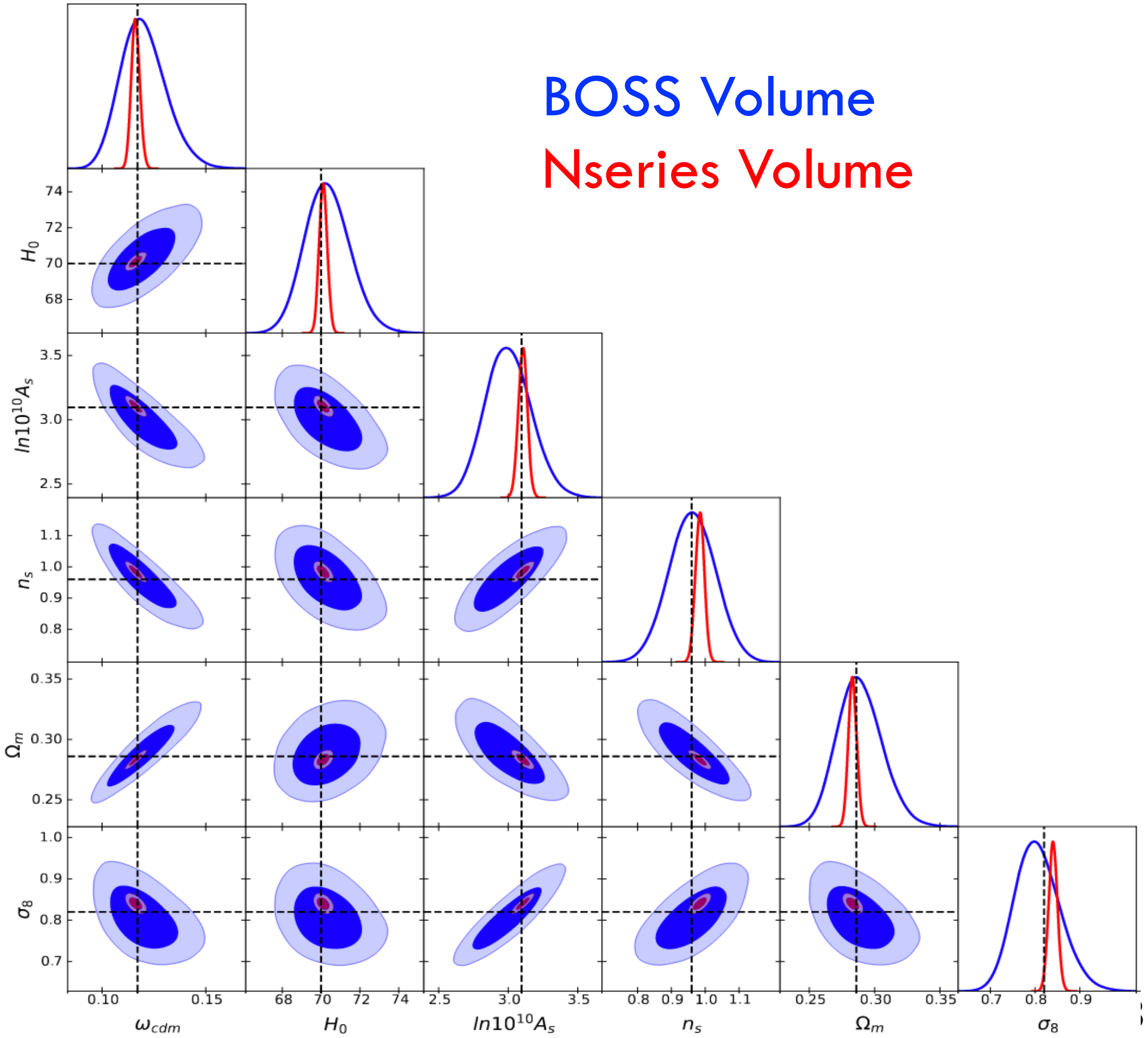
Cosmological Parameters

Philcox+21

Ivanov, **Philcox**+21

Cabass, **Philcox**+21,22

MODEL VALIDATION



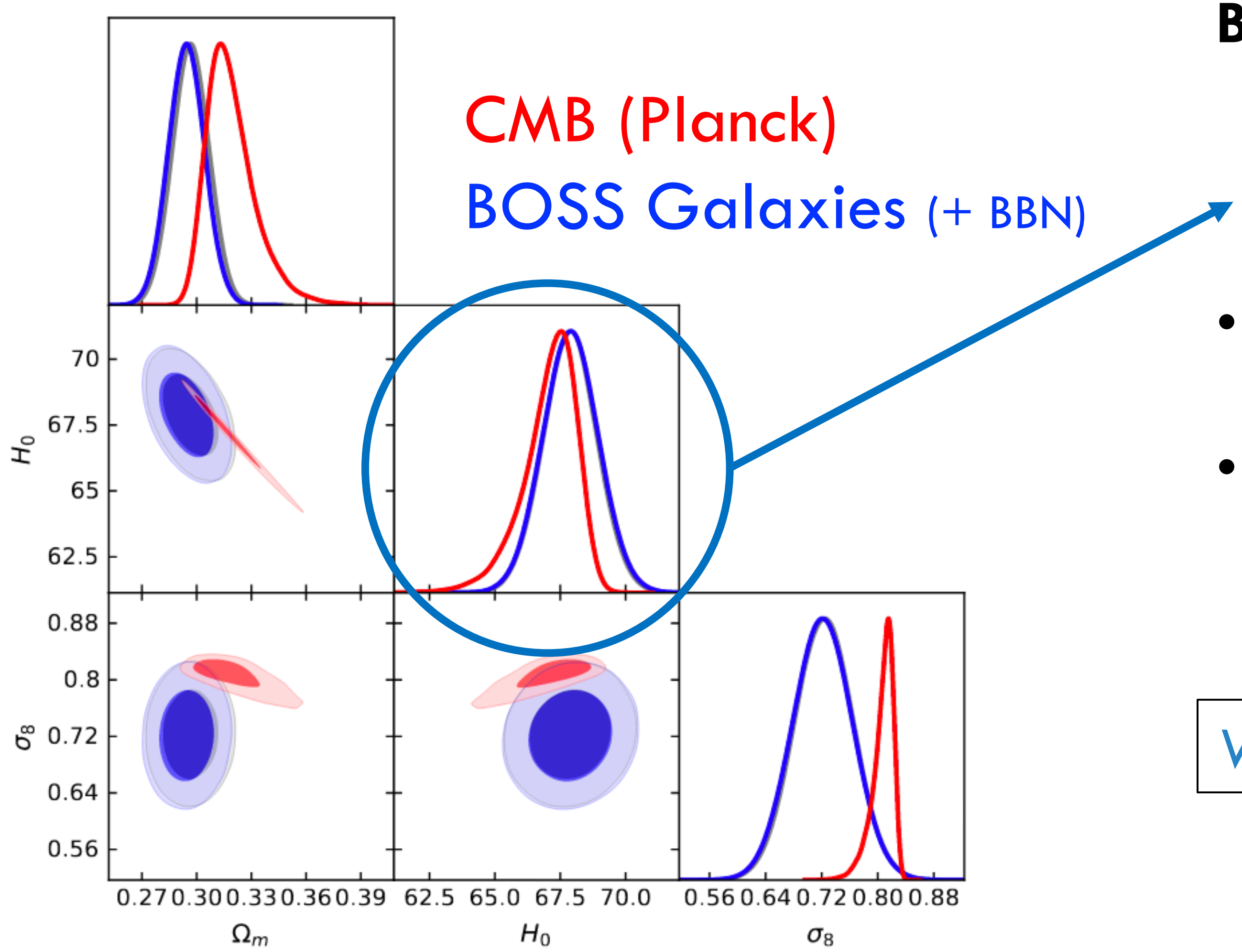
Validate with high-resolution **Nseries** mocks

- All parameters recovered at $\ll 1\sigma$
- Theory model works!
- Window function works!
- Fiber collisions work!

See *GitHub.com/oliverphilcox/full_shape_likelihoods*

Philcox+21

CONSTRAINING ΛCDM: H_0



BOSS Power Spectrum + Bispectrum:

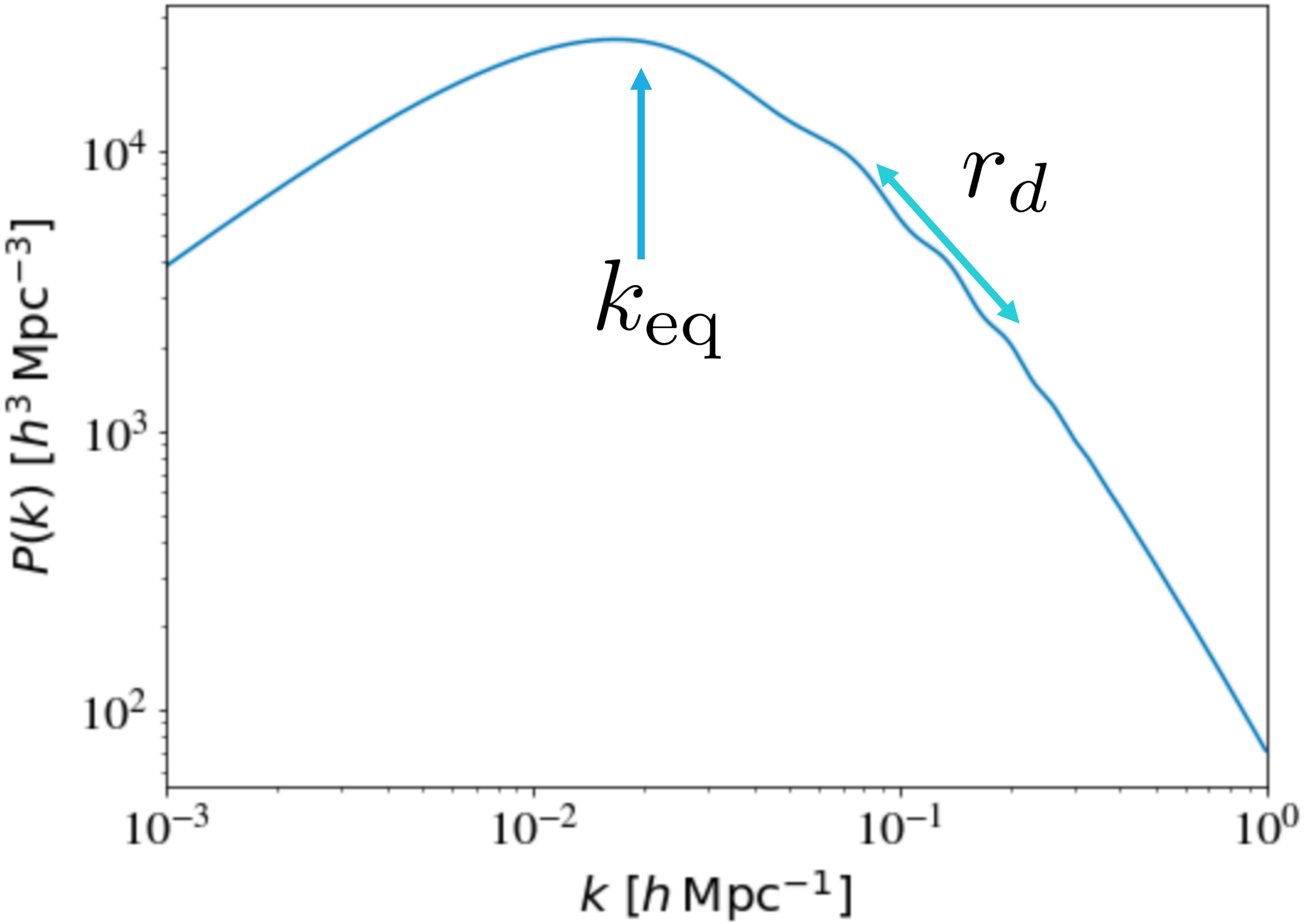
$$H_0 = 68.3 \pm 0.8\ \mathrm{km\,s^{-1}Mpc^{-1}}$$

- H_0 agrees with *Planck*
- 3.7σ *discrepant with SH0ES!*

Where does this information come from?

Philcox+21,22 (also Chen+21, d'Amico+21)

TWO STANDARD RULERS FOR H_0

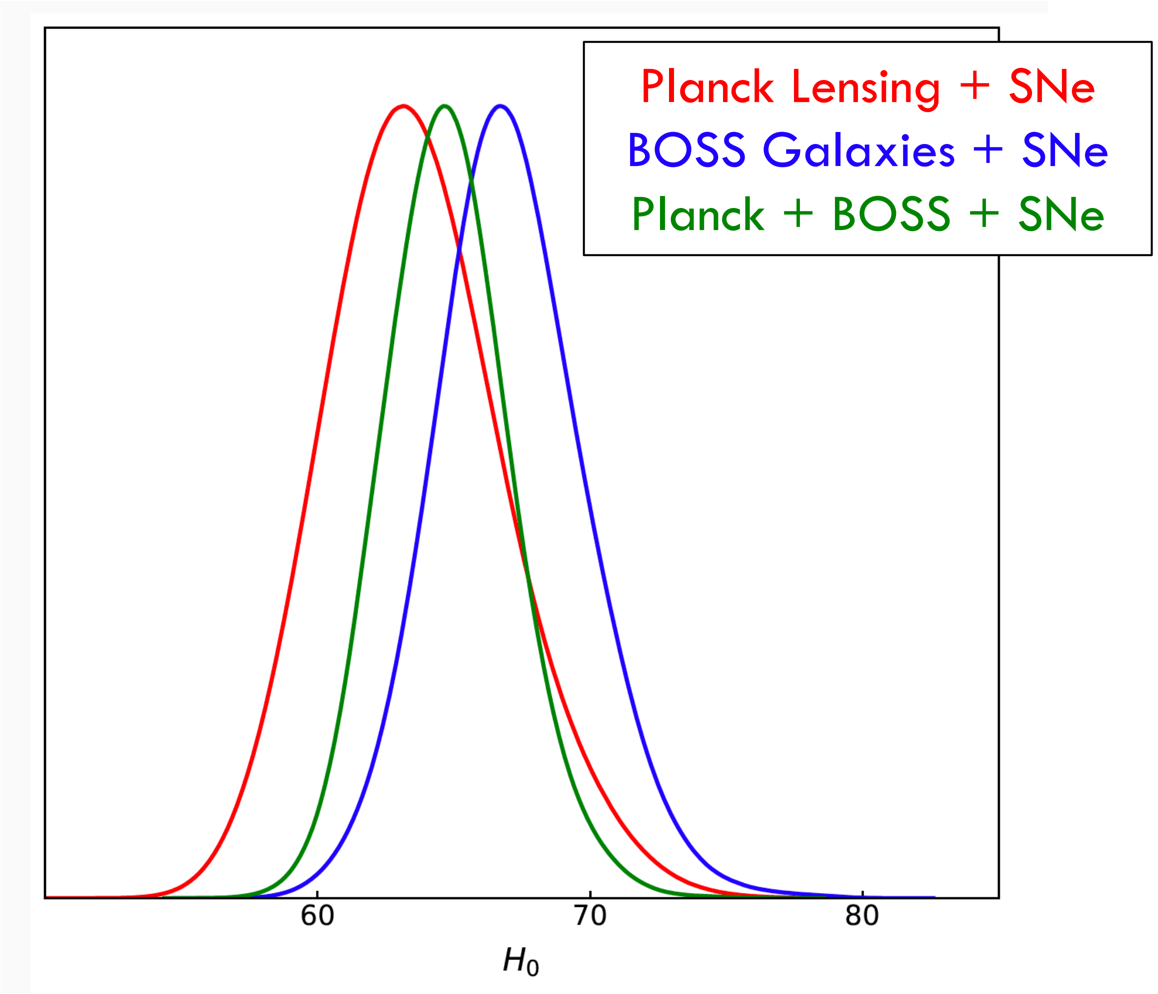


1. The Sound Horizon: r_d
   - ▷ The **sound horizon** at baryon drag ($z \sim 1100$)
2. The Equality Scale: $k_{\rm eq}^{-1}$
   - ▷ The **horizon** at radiation-matter equality ($z \sim 3600$)

***Both** can be used to extract H_0*

Ivanov+21
Philcox+21

CONSTRAINTS ON H_0



Sound-Horizon Independent Constraints

BOSS Full Power Spectrum + Bispectrum:

$$(z \approx 1100\,) \quad H_0 \;= 68.3 \pm 0.8 \text{ km s}^{-1}\text{Mpc}^{-1}$$

BOSS-without-the-sound-horizon:

(using new r_d-marginalized pipeline)

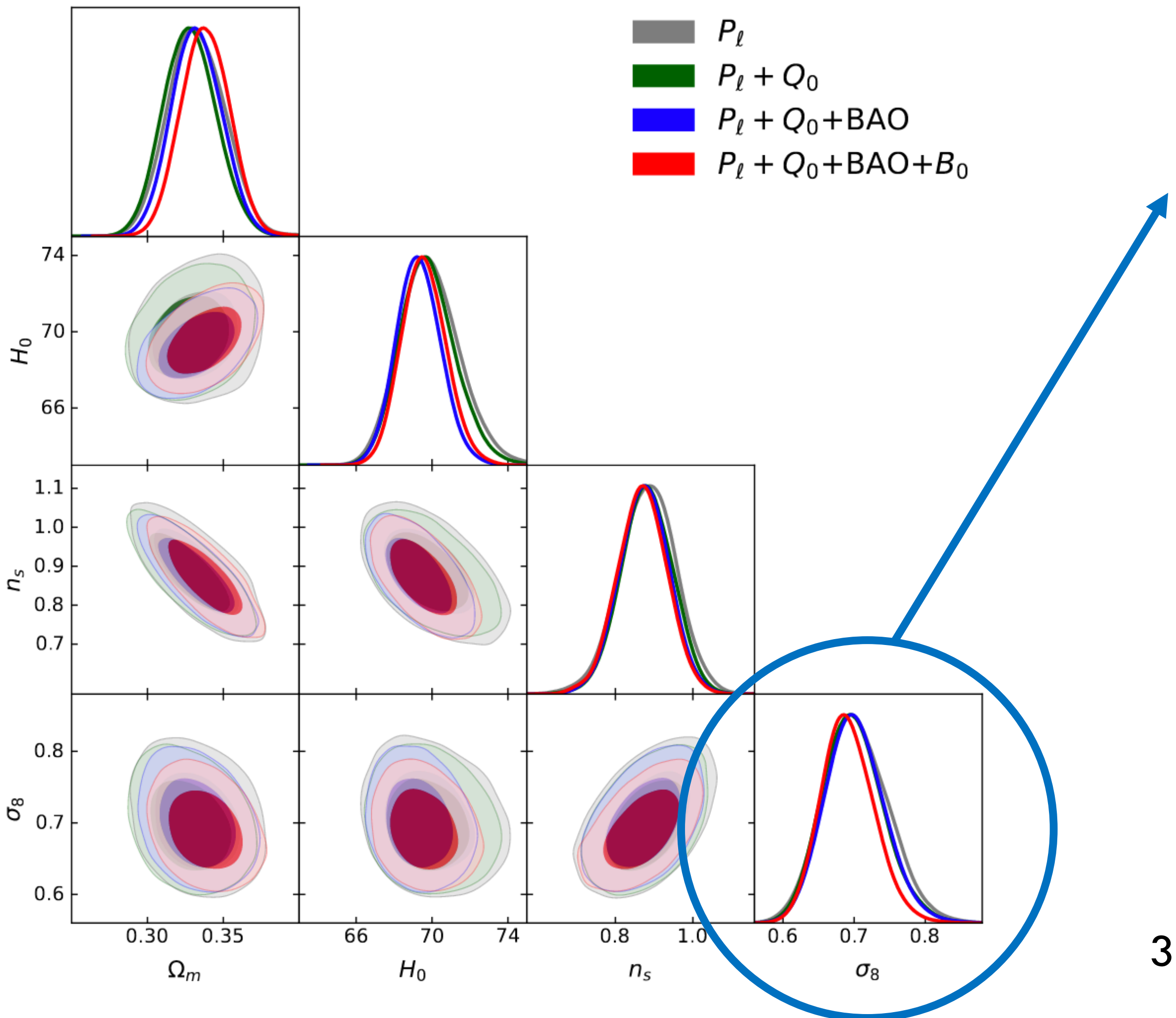
$$(z \approx 3500) \quad H_0 \;= 67.1 \pm 2.7 \text{ km s}^{-1}\text{Mpc}^{-1}$$

3.0σ *tension with SH0ES!*

No evidence for new physics from BOSS!

Philcox+21,22, Farren+21

CONSTRAINING ΛCDM: σ_8


BOSS (+ BBN) Constraints

BOSS Power Spectrum + Bispectrum:

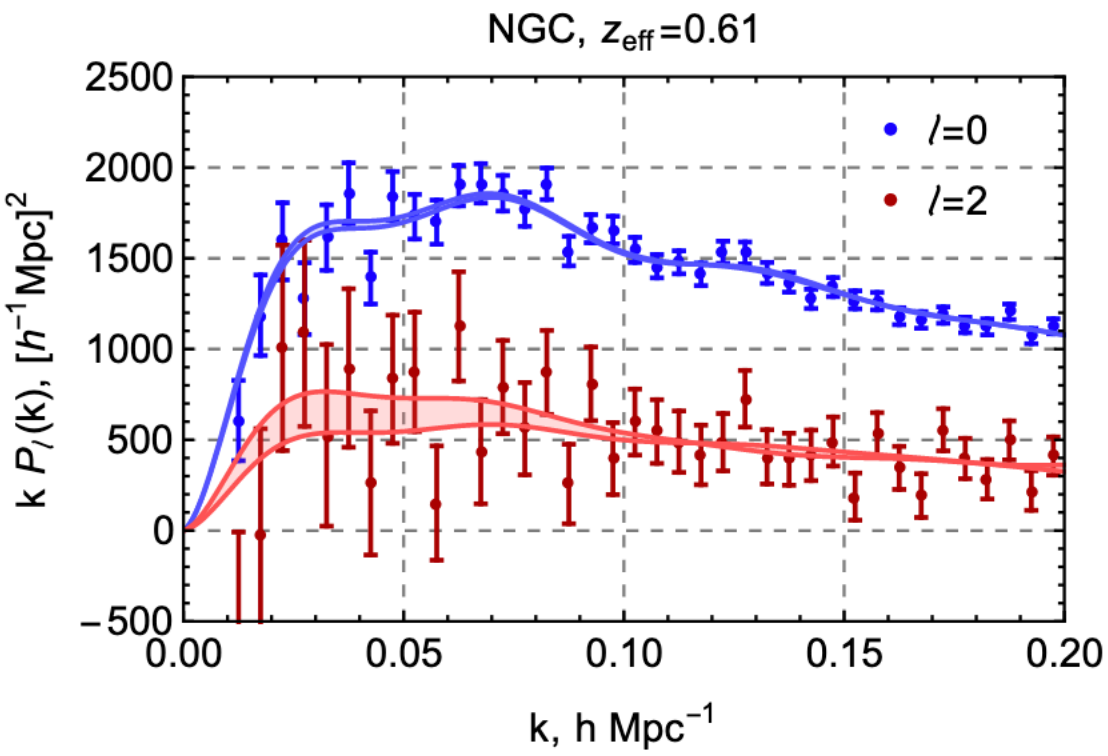
$S_8 = \mathbf{0.73 \pm 0.04}$ (BOSS, with *Planck* n_s)

This is consistent with weak lensing, but somewhat lower than *Planck*:

$S_8 = \mathbf{0.83 \pm 0.01}$ (*Planck*)

Philcox+21 (see also Chen+21, d'Amico+21)

WHERE DOES THE σ_8 INFORMATION COME FROM?



$\boldsymbol{\sigma_8}$ is set by the **large-scale** ($k < 0.1h/\mathrm{Mpc}$) quadrupole

This is hard to change!

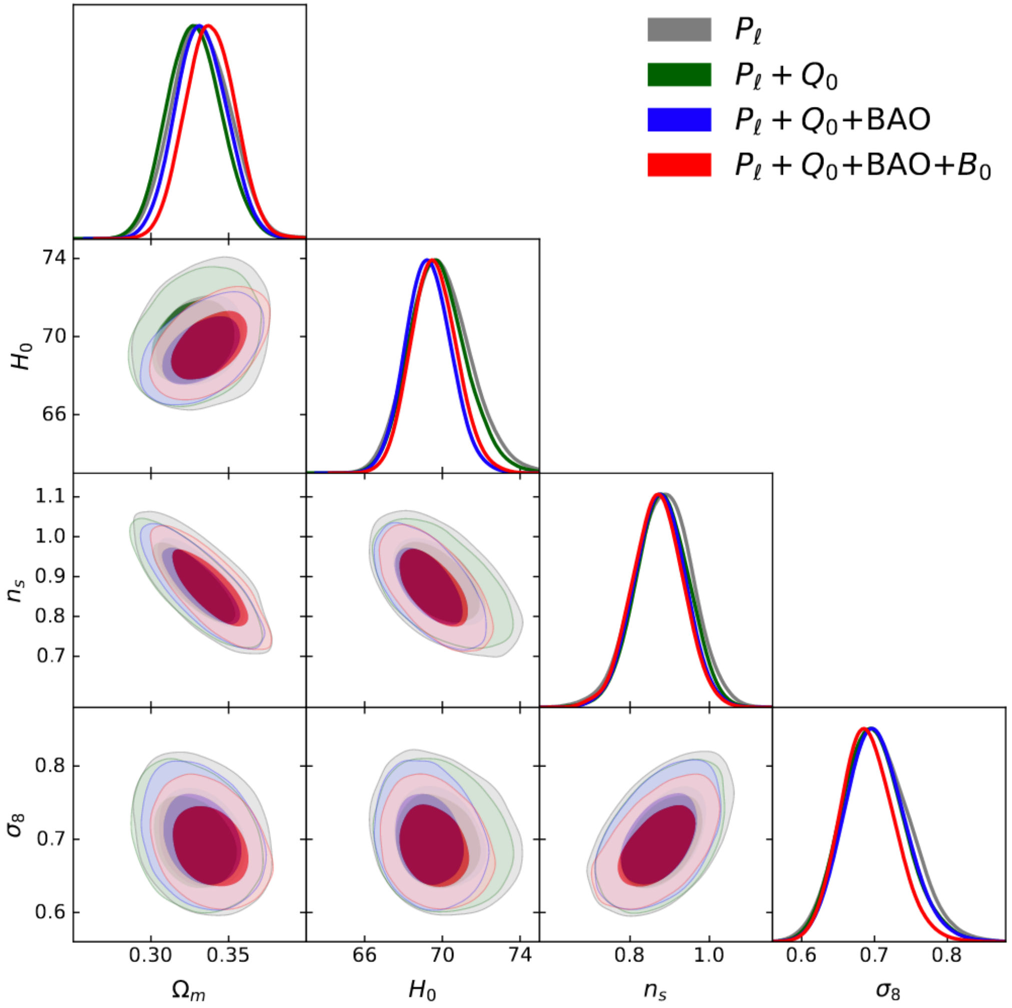
- ▷ Mostly linear scales
- ▷ Bias well understood
- ▷ Fingers-of-God suppressed

<u>But</u> priors are 1σ effect! [Simon+22]

Philcox+21 (see also Chen+21, d'Amico+21)

CONSTRAINING ΛCDM: OTHER PARAMETERS

BOSS (+ BBN) Constraints



Matter Density:

$$\Omega_m = 0.34 \pm 0.02$$

Consistent with Pantheon+ supernovae!

Spectral Slope:

$$n_s = 0.87 \pm 0.07$$

Consistent with *Planck*

Neutrino Mass:

$$\sum m_\nu < 0.14 \text{ eV } (95\% \text{ CL})$$

Philcox+20,21 (see also Chen+21, d'Amico+21)

CONSTRAINING INFLATION

In Single-Field Slow-Roll Inflation:

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

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

Non-standard inflation can beat this:

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

Maldacena 03, Creminelli+04

CONSTRAINING INFLATION

Need to **model** PNG in **power spectra** and **bispectra:**

- Primordial bispectrum:

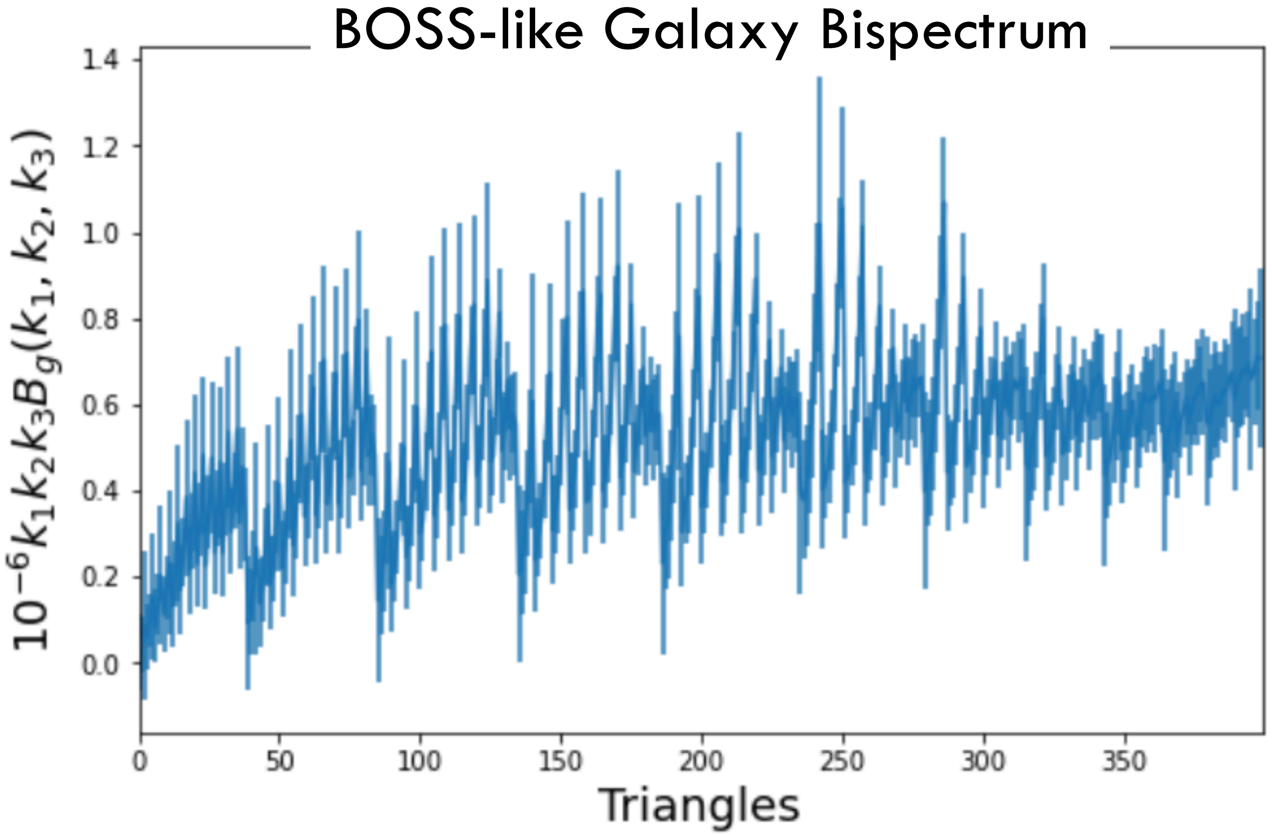
$$\left\langle \delta^{(1)}\delta^{(1)}\delta^{(1)} \right\rangle \sim f_{\rm NL} P^2(k)$$

- Scale dependent bias:

$$b_1(f_{\rm NL}) \to b_1 + (b_\phi f_{\rm NL})/k^2$$

- Loop corrections:

$$P_{gg}(\boldsymbol{k}) \to P_{gg}(\boldsymbol{k}) + f_{\rm NL}\int d\boldsymbol{q}\, \alpha\, P(\boldsymbol{q})P(\boldsymbol{k}-\boldsymbol{q})$$



$$\boxed{B_g = B_g(f_{\rm NL}^{\rm eq}, f_{\rm NL}^{\rm orth}, f_{\rm NL}^{\rm loc})}$$

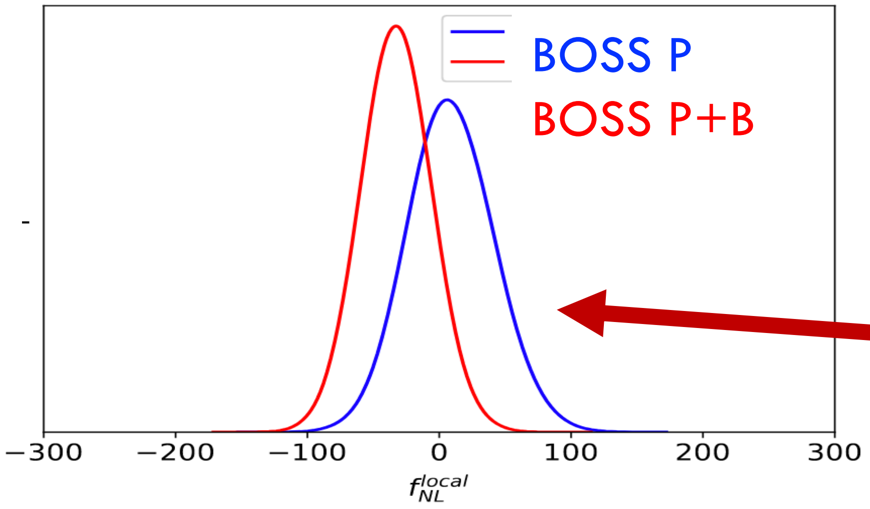
See GitHub.com/michalychforever/CLASS-PT

Cabass, **Philcox**+21,22 (see also d'Amico+22)

CONSTRAINING INFLATION

BOSS Power Spectrum + Bispectrum + $\boldsymbol{O(f_{\mathrm{NL}})}$ Theory Model

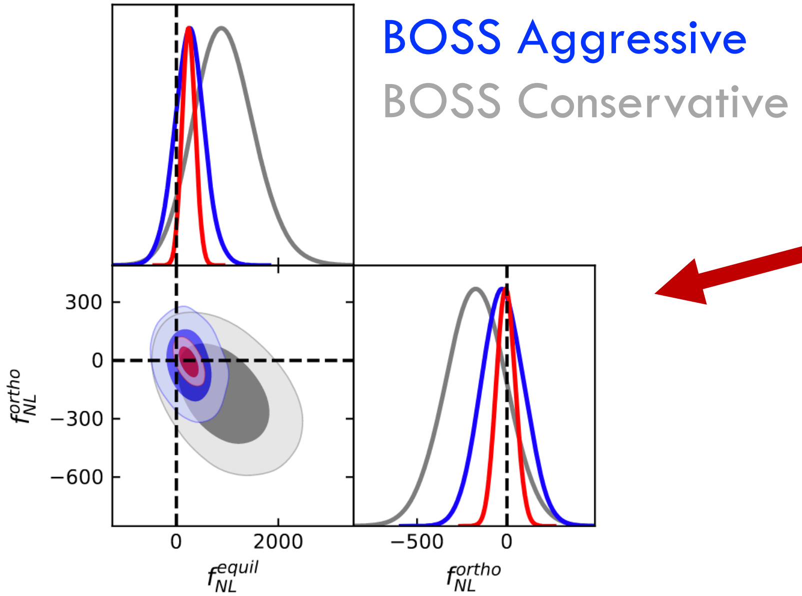
Multi-Field Inflation



$$f_{\mathrm{NL}}^{\mathrm{local}} = -33 \pm 28$$

(Really measuring $b_\phi f_{\mathrm{NL}}$ - see Barreira+22)

Single-Field Inflation

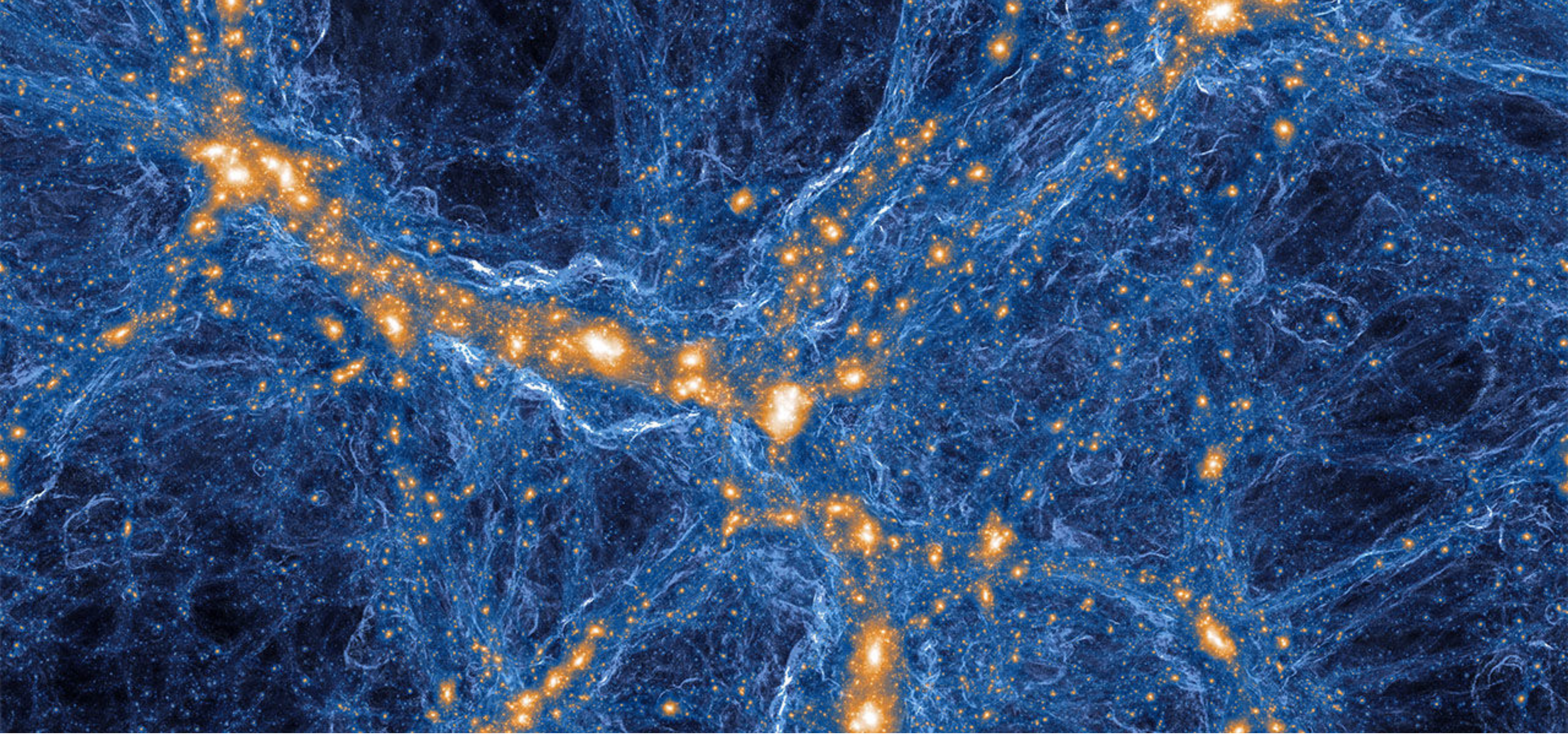


$$f_{\mathrm{NL}}^{\mathrm{equil}} = 260 \pm 300$$

$$f_{\mathrm{NL}}^{\mathrm{orth}} = -23 \pm 120$$

- *First measurement without CMB*
- *Needs bispectrum*

Cabass, **Philcox**+21,22 (see also d'Amico+22)

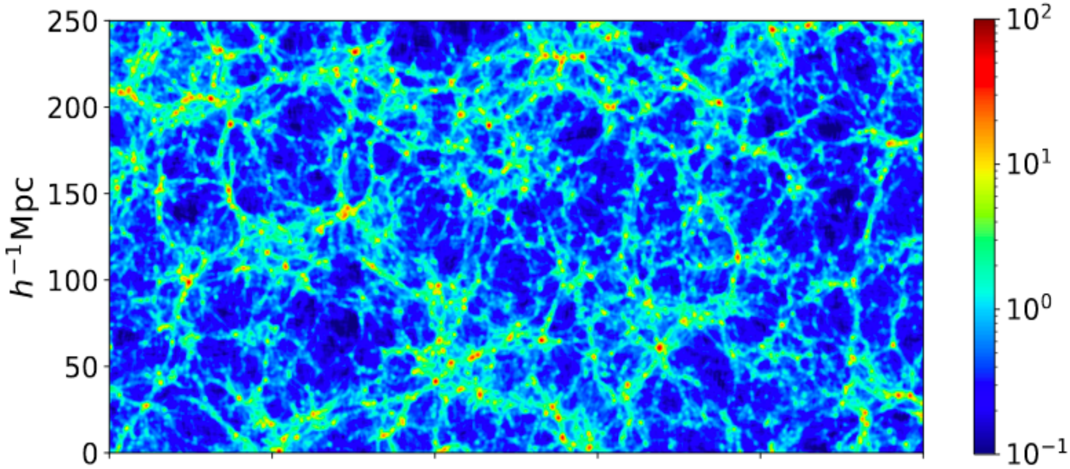


PART IV: Beyond Polyspectra

BEYOND THE DENSITY FIELD

- **Non-Gaussian** Universes need **higher-order** statistics

- Various **transformed** fields have been proposed:
  - Reconstructed Density Fields [e.g. Eisenstein+07
  - Lognormal Transforms [Neyrinck+09, Wang+11
  - Gaussianized Density Fields [Weinberg 92, Neyrinck+17]



Overdensity Field

Massara+20

BEYOND THE DENSITY FIELD

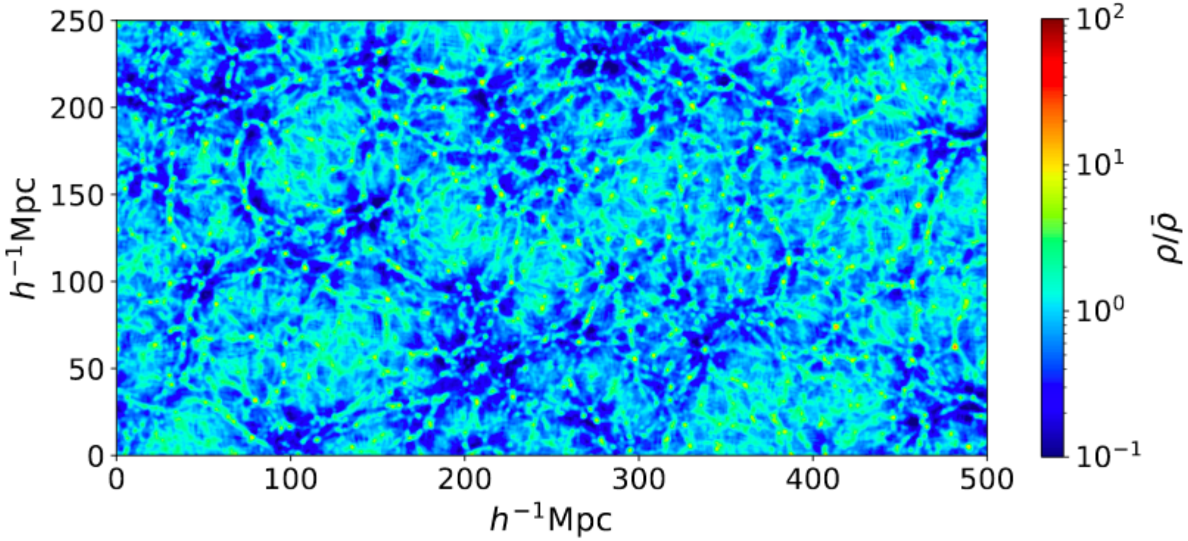
- The **marked** density field **upweights** low density regions

$$\delta(\mathbf{x}) \to \delta(\mathbf{x}) \left(\frac{1}{1 + \alpha\, \delta_R(\mathbf{x})}\right)^p$$

Smoothed density field

- Expected to improve constraints on parameters e.g., **neutrino mass** by O(10x)

- Can we **understand** what's going on?



Marked Density Field

Massara+20

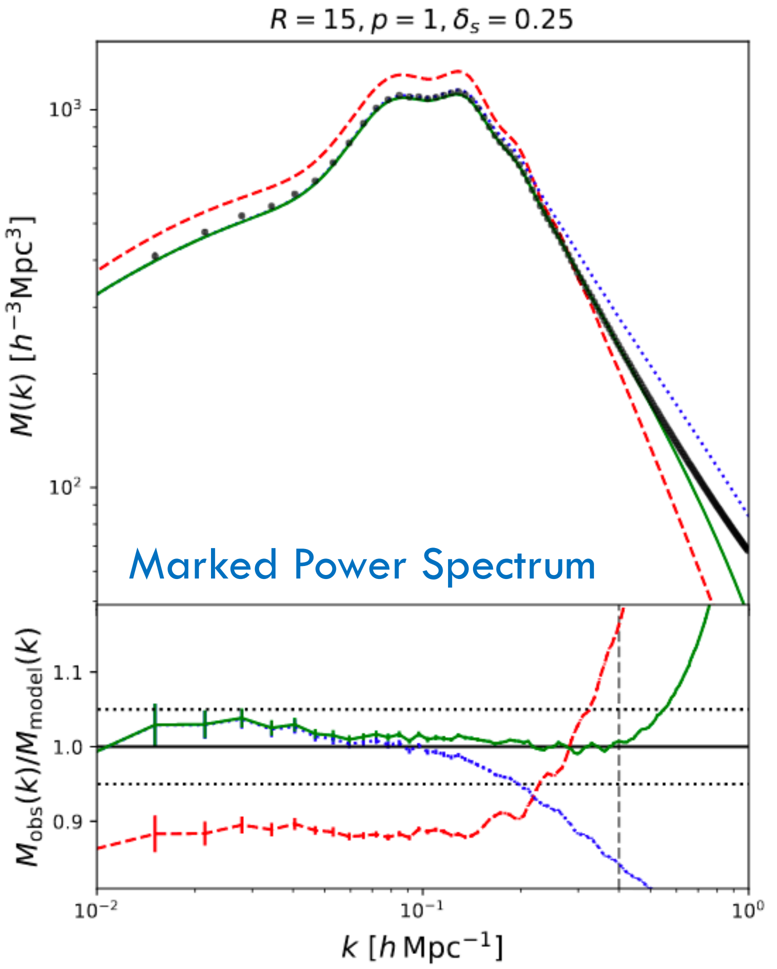
THE MARKED DENSITY FIELD

- Model marked field using 1-loop EFT
- The mark couples **small-scale** non-Gaussianities to **large-scale** modes
- Neutrino information leaks into low-k!

However:

- Modelling is **difficult** at low-z – no scale separation!
- Is it still useful for galaxies – absorbed by bias freedoms?

[cf. Massara+22]



Matter at $z = 1$

Massara+20, **Philcox**+20ce

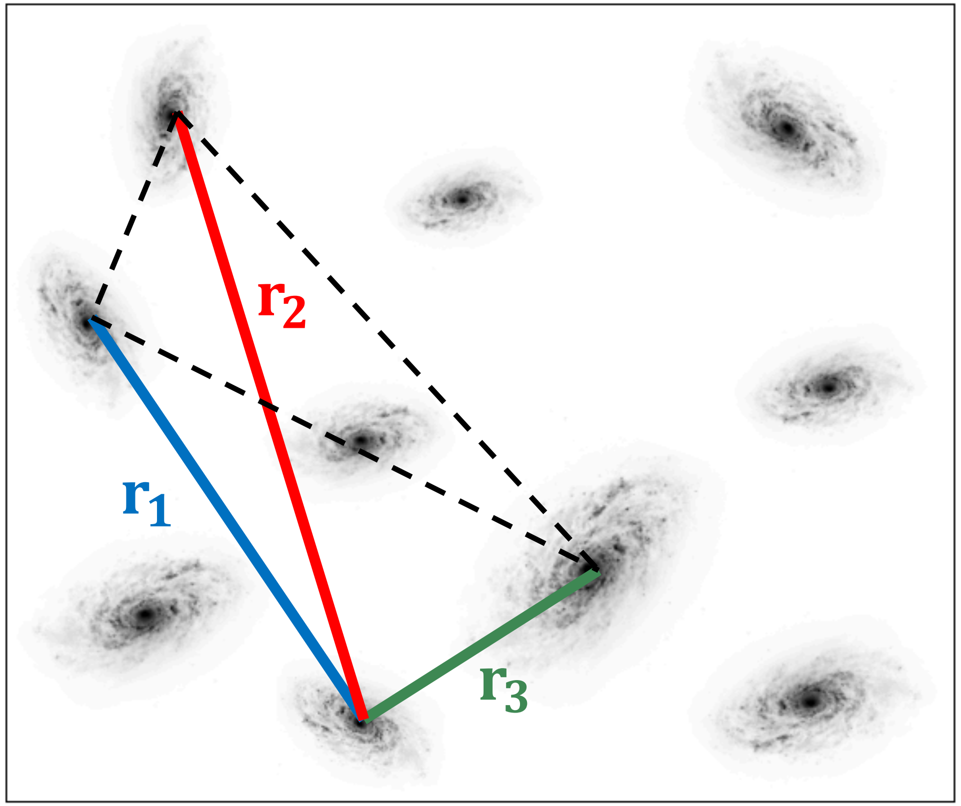
COSMOLOGY WITH CORRELATION FUNCTIONS

- N-point correlation functions (NPCFs) are **equivalent** to polyspectra

- In **real-space**, windows are much easier to deal with!

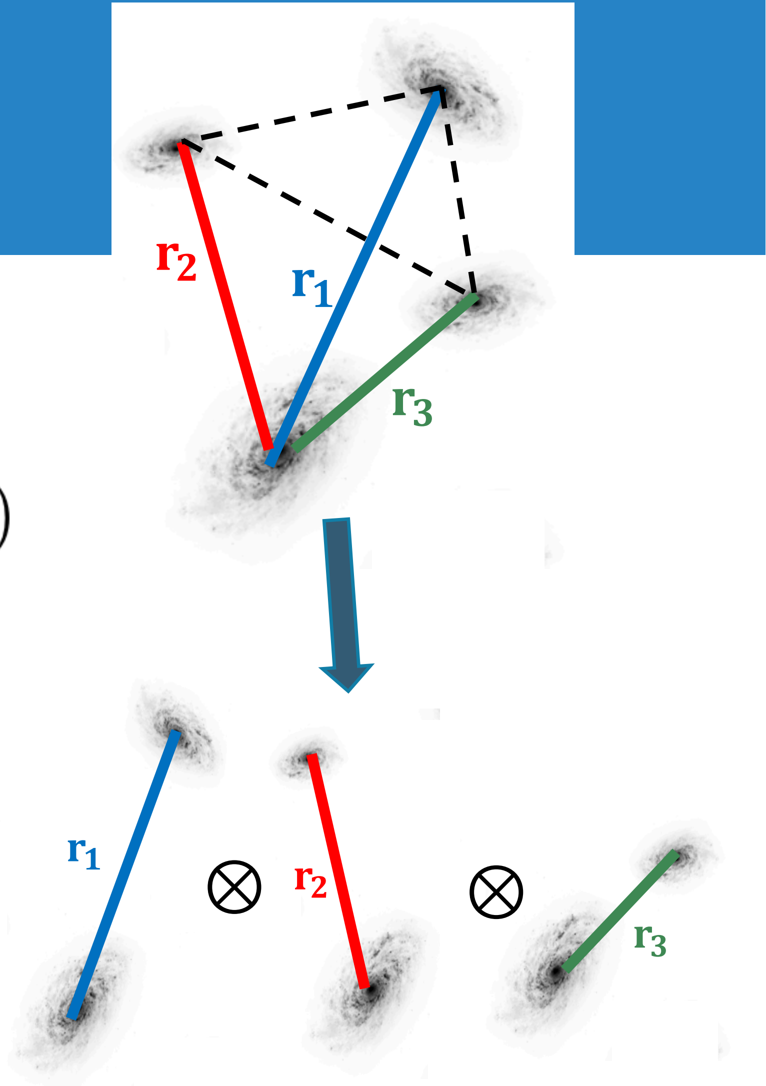
- Correlators usually estimated using particle counts, e.g., counting **quadruplets** for the 4PCF

Total number of quadruplets: $\mathcal{O}\left(N_{\text{gal}}^4\right)$!



Philcox+21

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

Separable basis formed from **angular momentum** addition

$$\mathcal{P}_{\ell_1\ell_2\ell_3}(\hat{\mathbf{r}}_1,\hat{\mathbf{r}}_2,\hat{\mathbf{r}}_3) = \sum_{m_1m_2m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} Y^*_{\ell_1m_1}(\hat{\mathbf{r}}_1)Y^*_{\ell_2m_2}(\hat{\mathbf{r}}_2)Y^*_{\ell_3m_3}(\hat{\mathbf{r}}_3)$$

We can count **pairs** of galaxies to compute the 4PCF!

See GitHub.com/oliverphilcox/encore, GitHub.com/oliverphilcox/NPCFs.jl

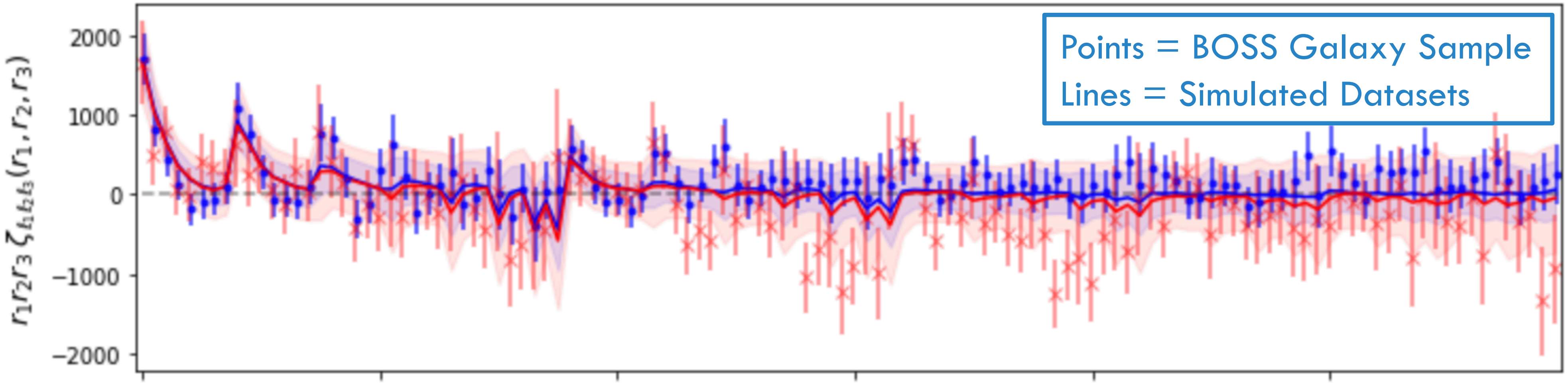
Cahn+20, **Philcox**+21

MEASURING THE 4-POINT FUNCTION

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

Do we detect a signal?

Even 4PCF (one of 42 components)



Points = BOSS Galaxy Sample
Lines = Simulated Datasets

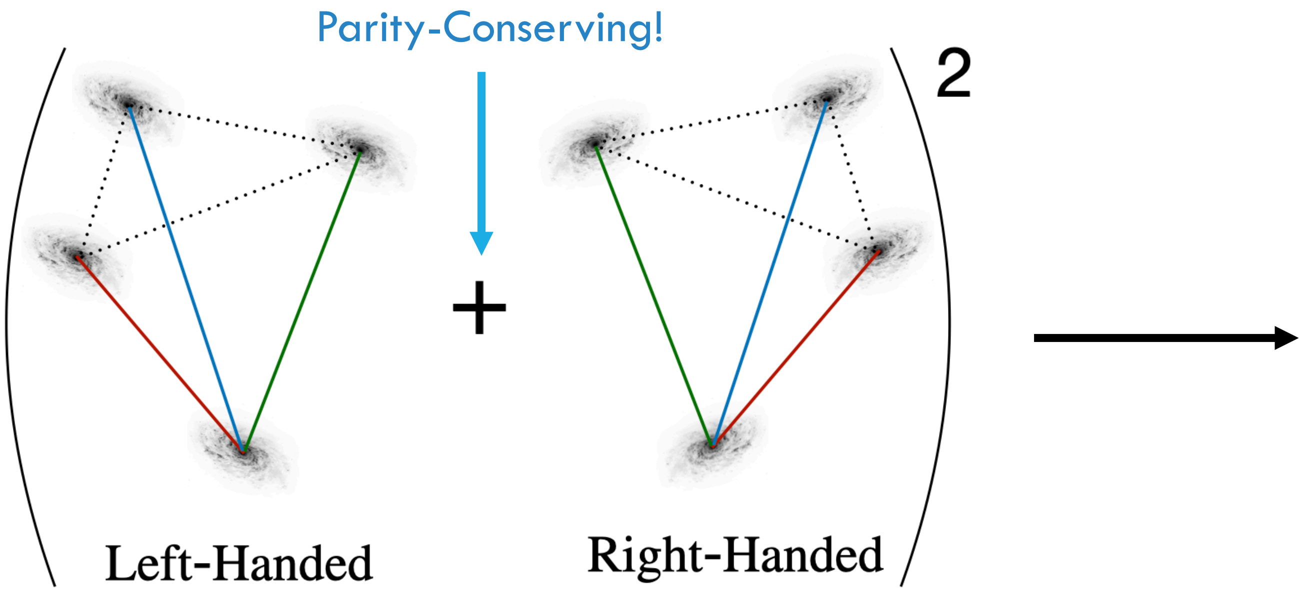
$\sim$20 Mpc/h

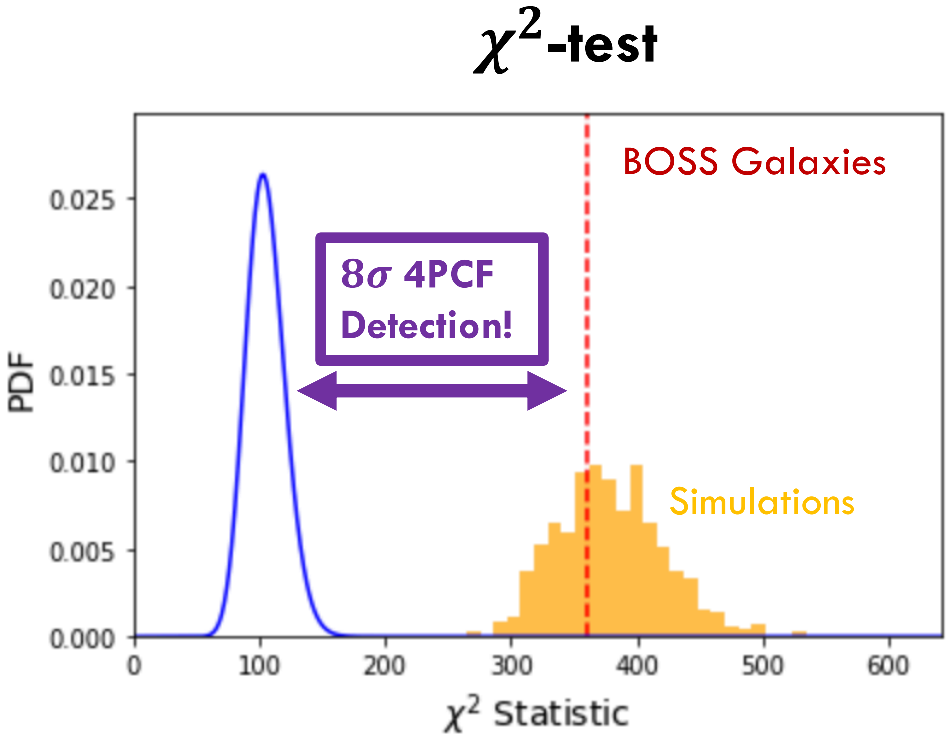
Increasing Radius

$\sim$ 150 Mpc/h

Philcox+21

PARITY-EVEN 4-POINT FUNCTIONS

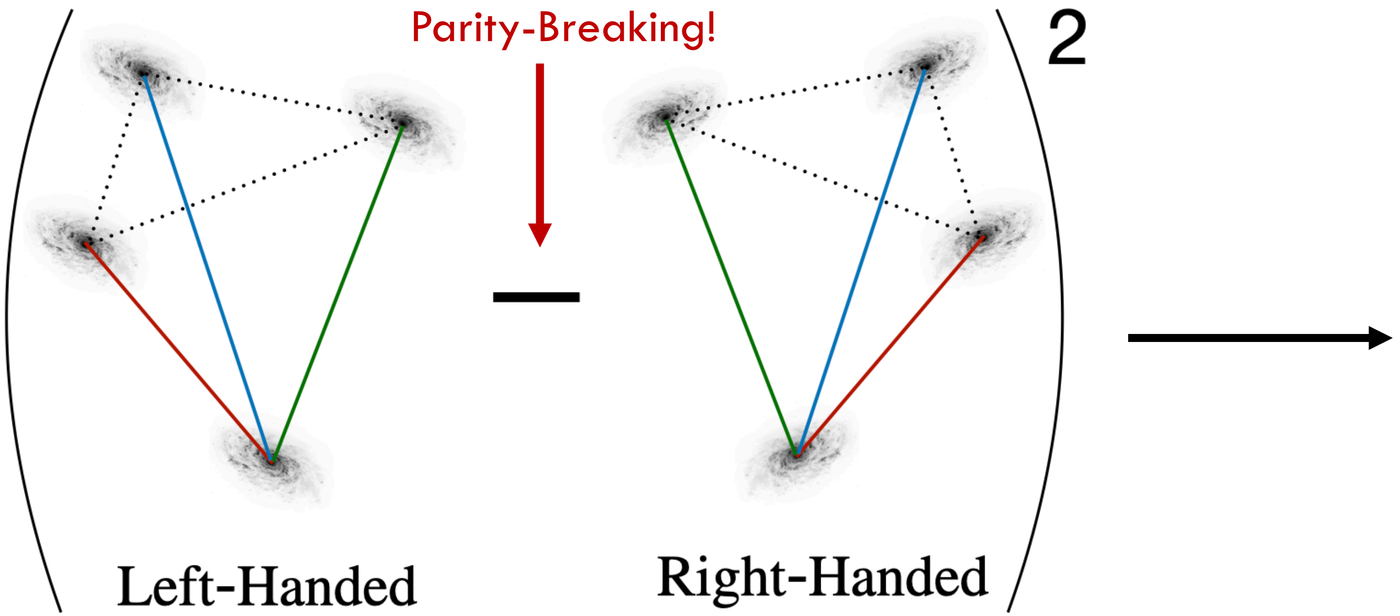


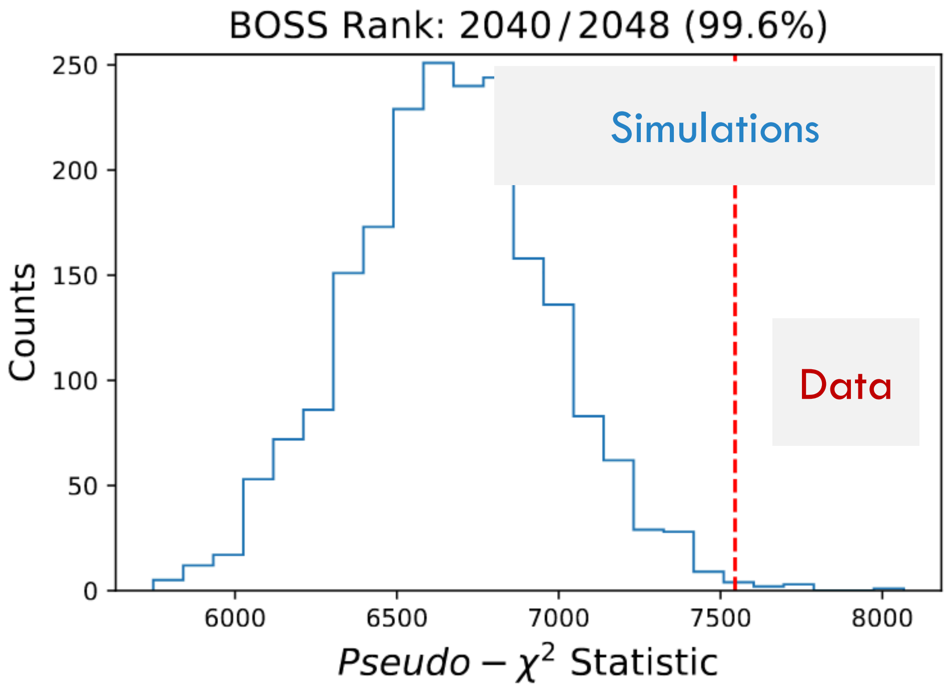


- Strong detection of **gravitational** non-Gaussianity
- ***But***, it's hard to model and interpret!

Philcox+22

PARITY-ODD 4-POINT FUNCTIONS



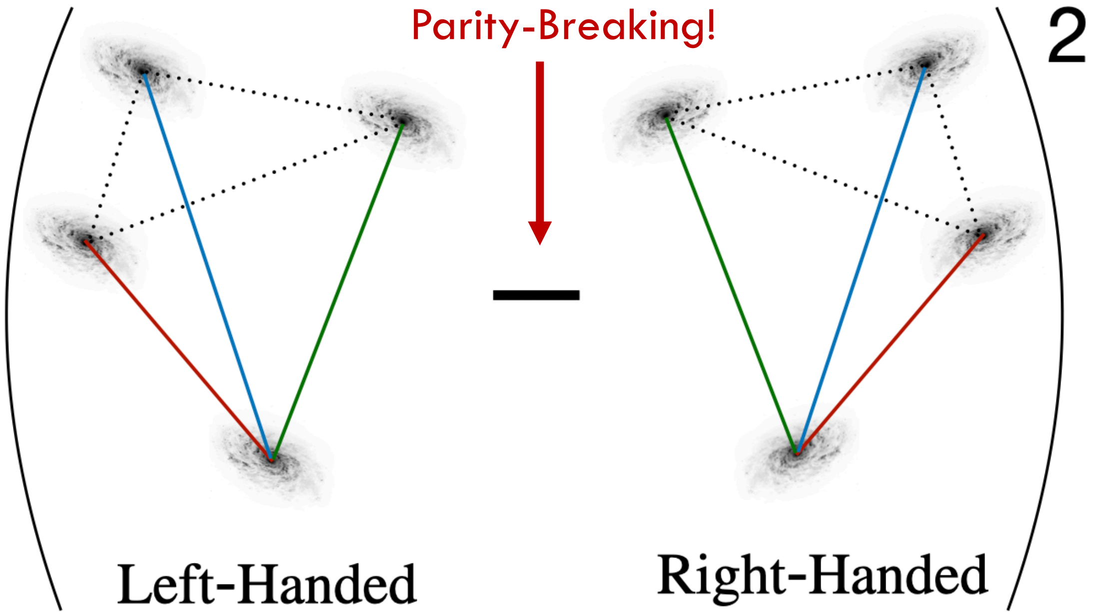


Weak detection of parity-violation signal

- Simulations do not capture noise properties of the data
- <u>Or</u> **systematics** are creeping in!
- <u>Or</u> we have detected ***parity-violating physics*** at 3σ???

Philcox 22 (see also Hou+22)

PARITY-ODD 4-POINT FUNCTIONS



NEWSLETTERS
Sign up to read our regular email newsletters

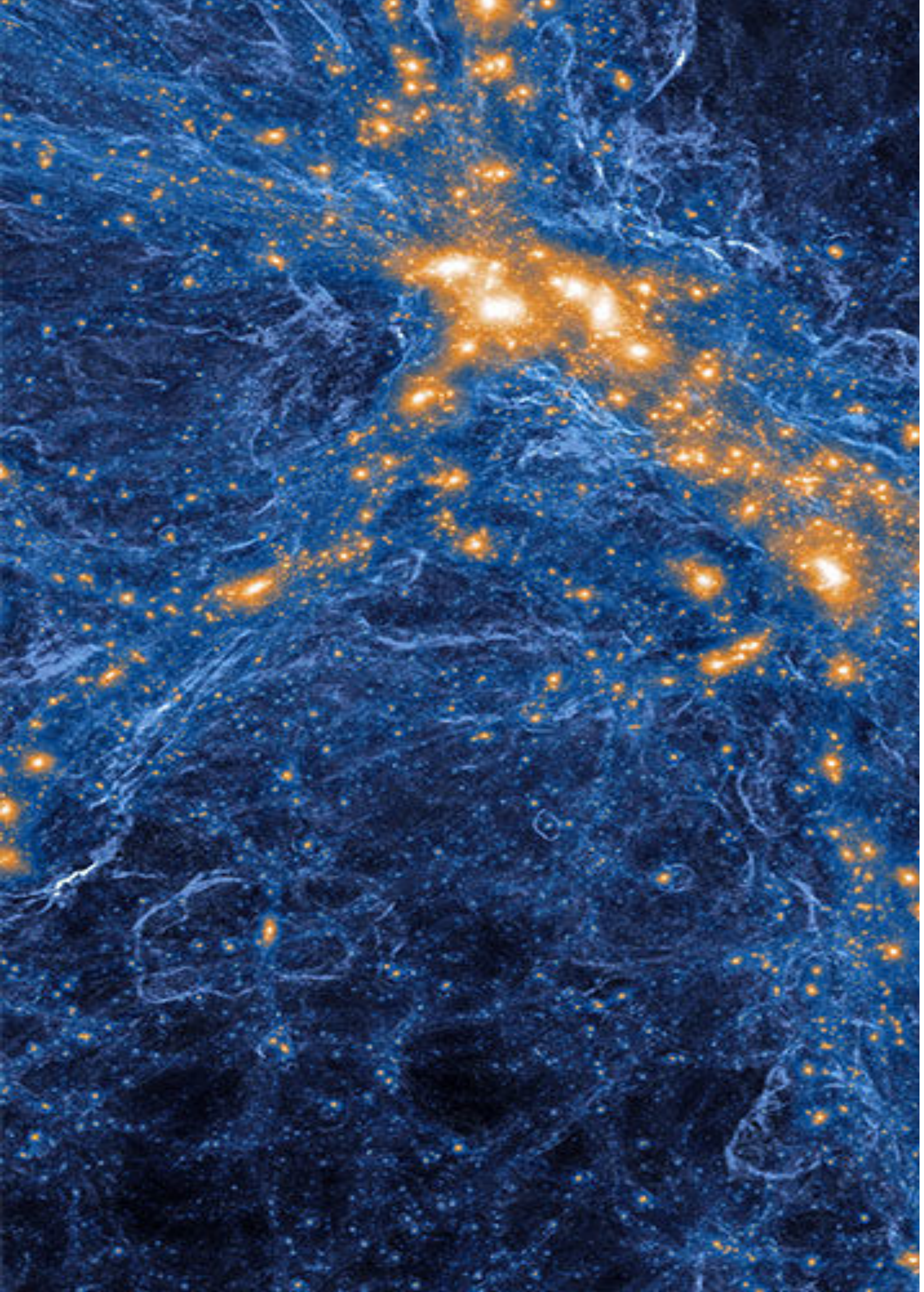
The universe is surprisingly lopsided and we don't know why

Two analyses of a million galaxies show that their distribution may not be symmetrical, which may mean that our understandings of gravity and the early universe are incorrect

Weak detection of parity-violation signal

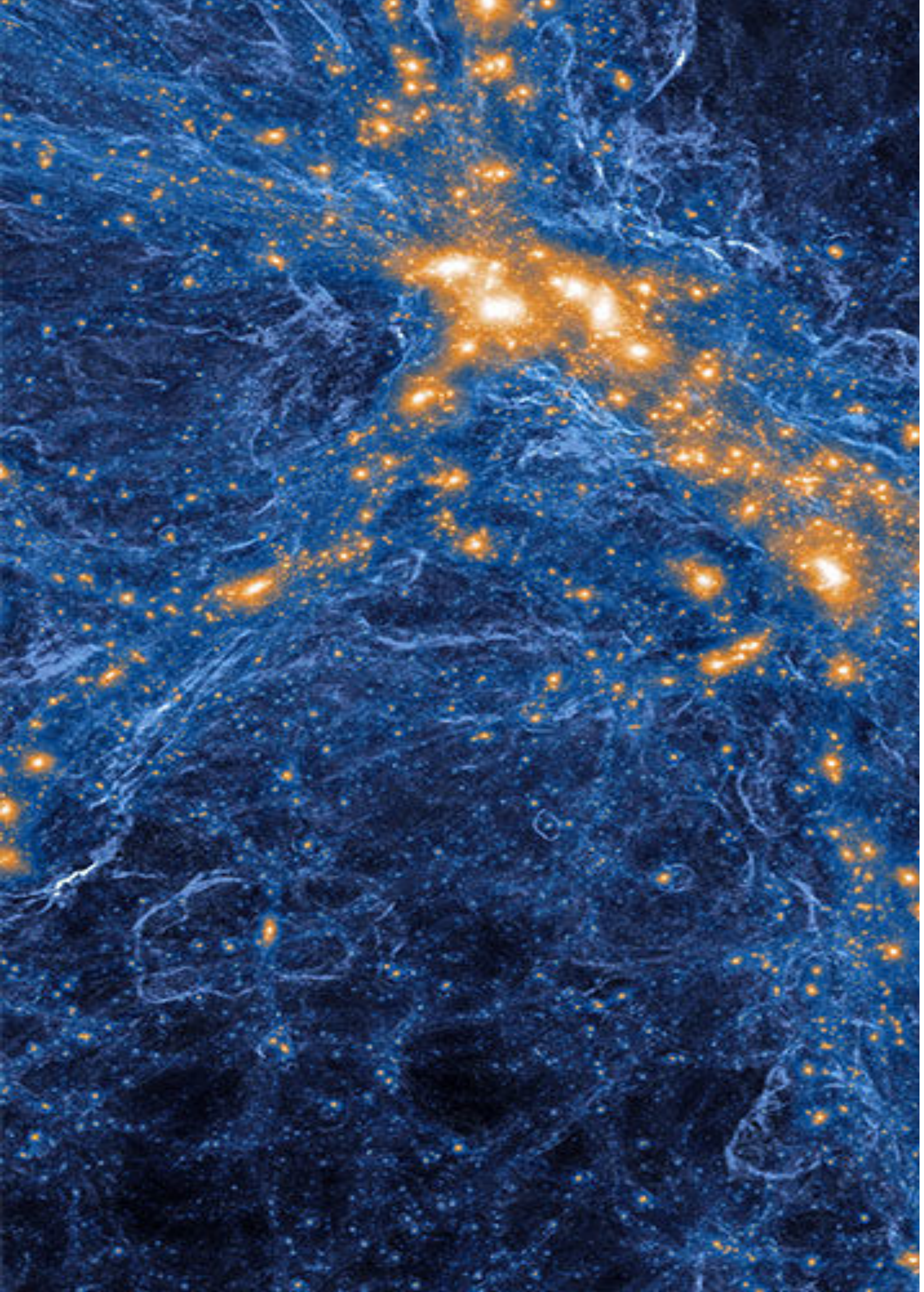
- Simulations do not capture noise properties of the data
- Or **systematics** are creeping in!
- Or we have detected ***parity-violating physics*** at 3σ???

Philcox 22 (see also Hou+22)



SUMMARY

- **New techniques** are needed to make the most of upcoming LSS surveys
- We can now **measure, model**, and **interpret** the power spectrum, bispectrum, and various higher-order statistics
- Direct **parameter inference** is now possible for ΛCDM and extensions



WHAT'S NEXT?

○ **More statistics**: bispectrum multipoles, trispectrum, field-level inference, etc.

○ **More loops:** higher $k_{\rm max}$, **but** more parameters!

○ **More models:** non-standard inflation, non-standard dark matter, etc.

○ **More data:** Euclid / DESI / SPHEREx [all codes are public] + better treatment of **systematics**

Collaborators

THANK YOU!!!

Advisors