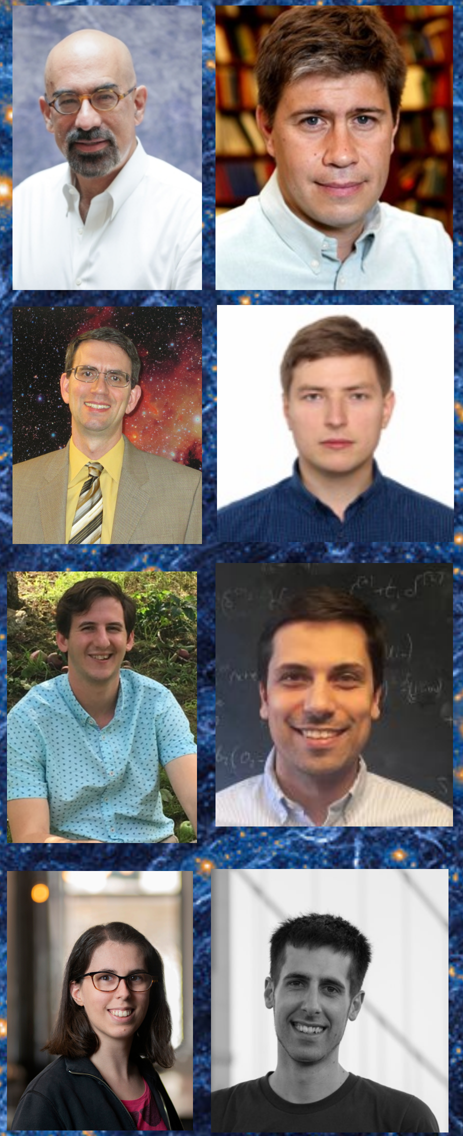
PRINCETON UNIVERSITY

IAS INSTITUTE FOR ADVANCED STUDY



Have We Exhausted the Galaxy Two-Point Function?

Oliver Philcox (Princeton / IAS)

Cosmology & Particle Physics Seminar, University of Geneva

Based on: 1912.01010, 2005.01739 2009.03311, 2006.10055, 2012.09389

Extracting Information from Galaxy Surveys

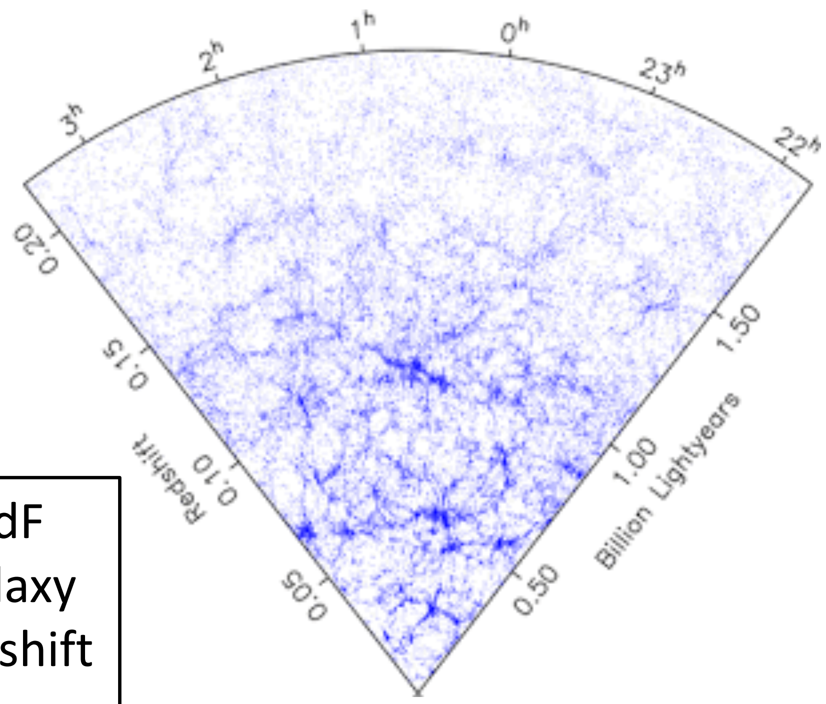
- Fundamental observable: the galaxy **overdensity** field

$$\delta(\mathbf{r}) = \frac{1}{\bar{n}} \left[n_g(\mathbf{r}) - n_r(\mathbf{r}) \right]$$

Random particle positions

Mean density

Galaxy positions



2dF Galaxy Redshift Survey

Extracting Information from Galaxy Surveys

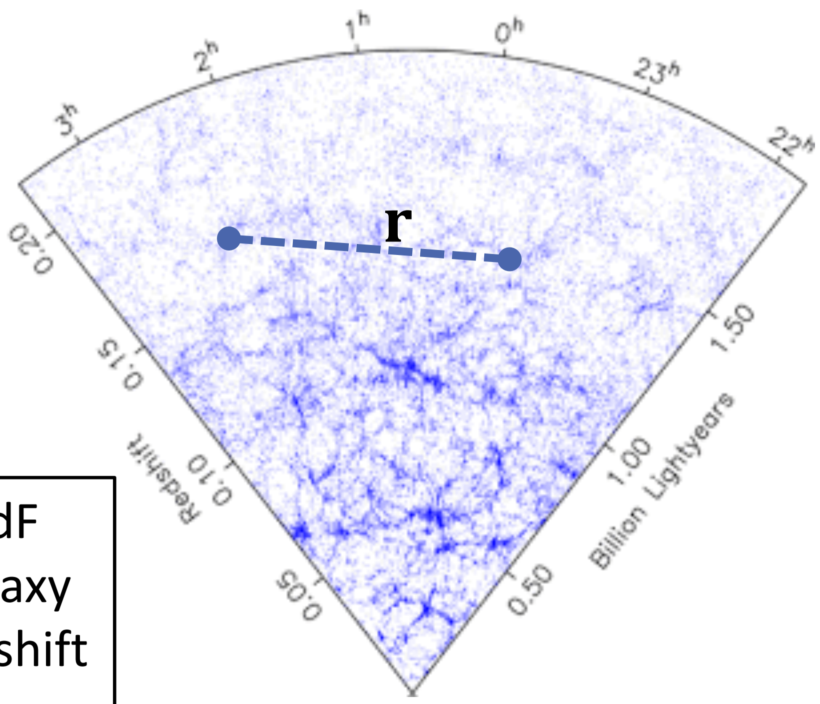
- Fundamental observable: the galaxy **overdensity** field

$$\delta(\mathbf{r}) = \frac{1}{\bar{n}}\left[n_g(\mathbf{r}) - n_r(\mathbf{r})\right]$$

Random particle positions

Mean density

Galaxy positions



2dF Galaxy Redshift Survey

- Analyze with **summary statistics:**
  - Two-point correlation function (2PCF), $\xi(\mathbf{r})$

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

  - Power spectrum, $P(\mathbf{k})$

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

Extracting Information from Galaxy Surveys

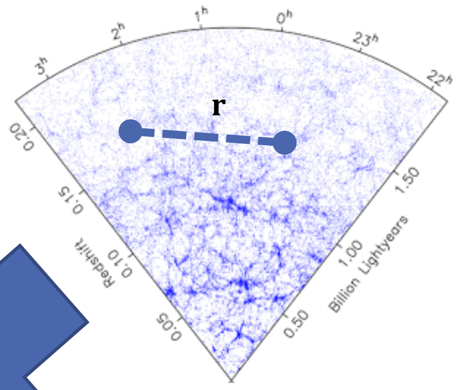
- Fundamental observable: the galaxy **overdensity** field

$$\delta(\mathbf{r}) = \frac{1}{\bar{n}}\left[n_g(\mathbf{r}) - n_r(\mathbf{r})\right]$$

Random particle positions

Mean density

Galaxy positions

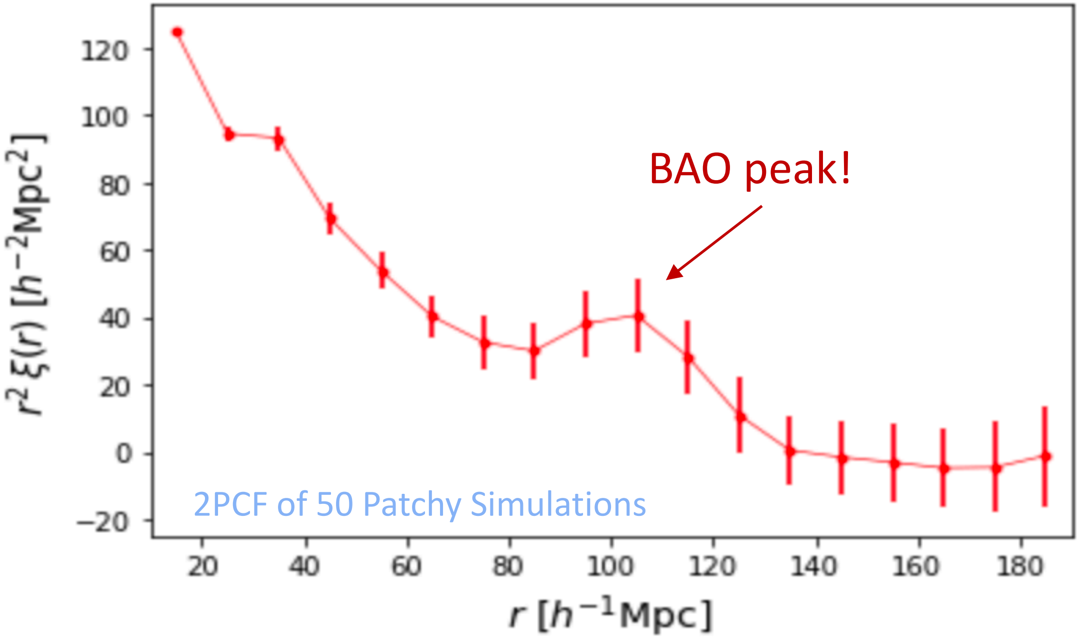


- Analyze with **summary statistics:**
  - Two-point correlation function (2PCF), $\xi(\mathbf{r})$

$$\boxed{\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle}$$

  - Power spectrum, $P(\mathbf{k})$

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



Extracting Information from Galaxy Surveys

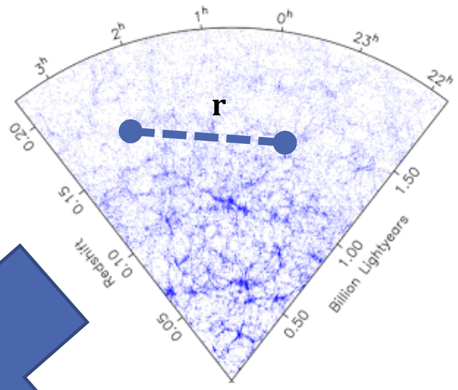
- Fundamental observable: the galaxy **overdensity** field

$$\delta(\mathbf{r}) = \frac{1}{\bar{n}}\left[n_g(\mathbf{r}) - n_r(\mathbf{r})\right]$$

Random particle positions

Mean density

Galaxy positions

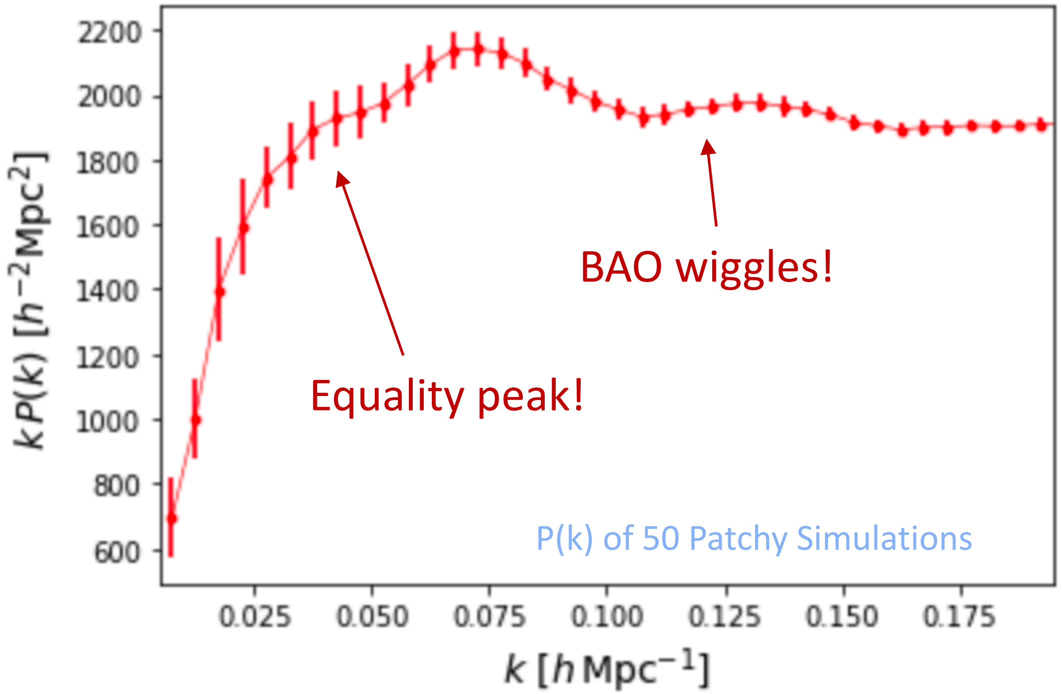


- Analyze with **summary statistics:**
  - Two-point correlation function (2PCF), $\xi(\mathbf{r})$

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

  - Power spectrum, $P(\mathbf{k})$

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



Understanding Anisotropy

- Due to **redshift-space distortions** the two-point correlators are **anisotropic**

  - Parametrize by **galaxy separation** and angle to **line-of-sight**, $\hat{\mathbf{n}}$

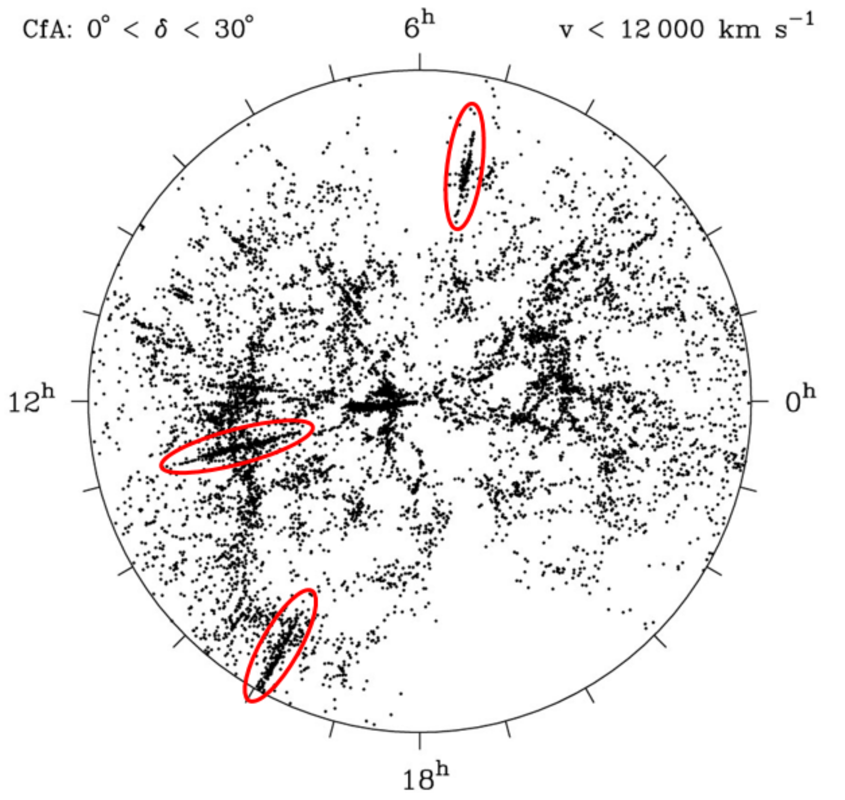
$$\xi(\mathbf{r}) = \sum_{\ell} \xi_\ell(r) L_\ell(\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}) \qquad P(\mathbf{k}) = \sum_{\ell} P_\ell(k) L_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}})$$

  - The **multipoles** are estimated from the density field:

$$\hat{\xi}_\ell(r) = (2\ell+1) \int \frac{d\Omega_r}{4\pi} \int d\mathbf{x}\, \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) L_\ell(\hat{\mathbf{r}} \cdot \hat{\mathbf{n}})$$

Legendre Polynomials

$$\hat{P}_\ell(k) = \frac{(2\ell+1)}{V} \int \frac{d\Omega_k}{4\pi} \underbrace{\int d\mathbf{r}_1\, d\mathbf{r}_2\, e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{=\,\delta(\mathbf{k})\delta^*(\mathbf{k})} L_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}})$$



CfA Redshift Survey

Understanding Anisotropy

- Due to **redshift-space distortions** the two-point correlators are **anisotropic**
  - Parametrize by **galaxy separation** and angle to **line-of-sight**, $\hat{\mathbf{n}}$

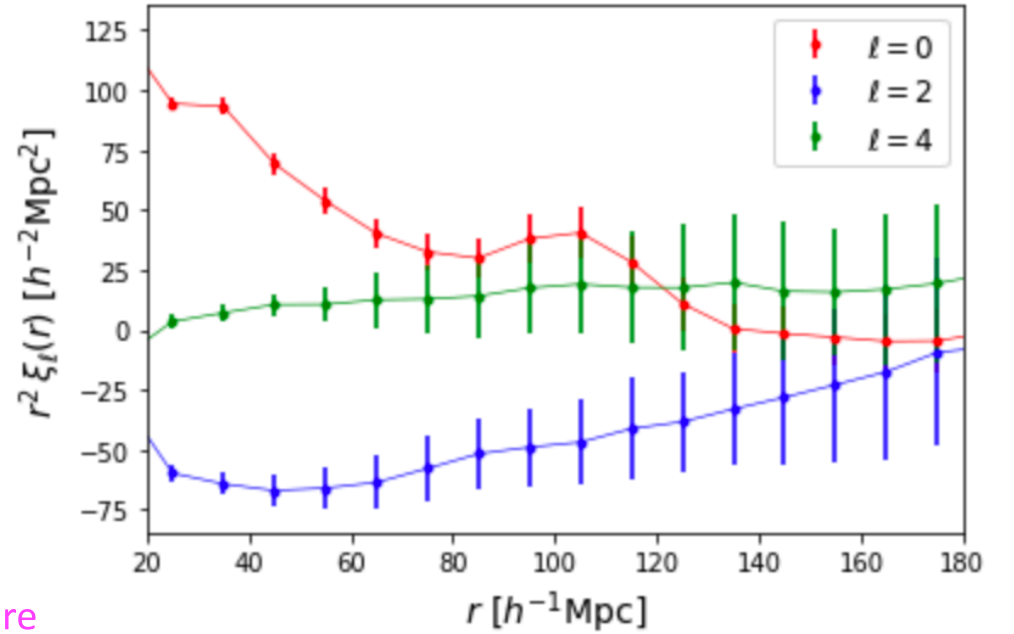
$$\xi(\mathbf{r}) = \sum_\ell \xi_\ell(r) L_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{n}}) \qquad P(\mathbf{k}) = \sum_\ell P_\ell(k) L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})$$

  - The **multipoles** are estimated from the density field:

$$\boxed{\hat{\xi}_\ell(r) = (2\ell+1)\int\frac{d\Omega_r}{4\pi}\int d\mathbf{x}\,\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})L_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{n}})}$$

Legendre Polynomials

$$\hat{P}_\ell(k) = \frac{(2\ell+1)}{V}\int\frac{d\Omega_k}{4\pi}\int d\mathbf{r}_1\,d\mathbf{r}_2\,e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)}\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})$$



MultiDark-Patchy Mocks (Kitaura+16)

Understanding Anisotropy

- Due to **redshift-space distortions** the two-point correlators are **anisotropic**

  - Parametrize by **galaxy separation** and angle to **line-of-sight**, $\hat{\mathbf{n}}$

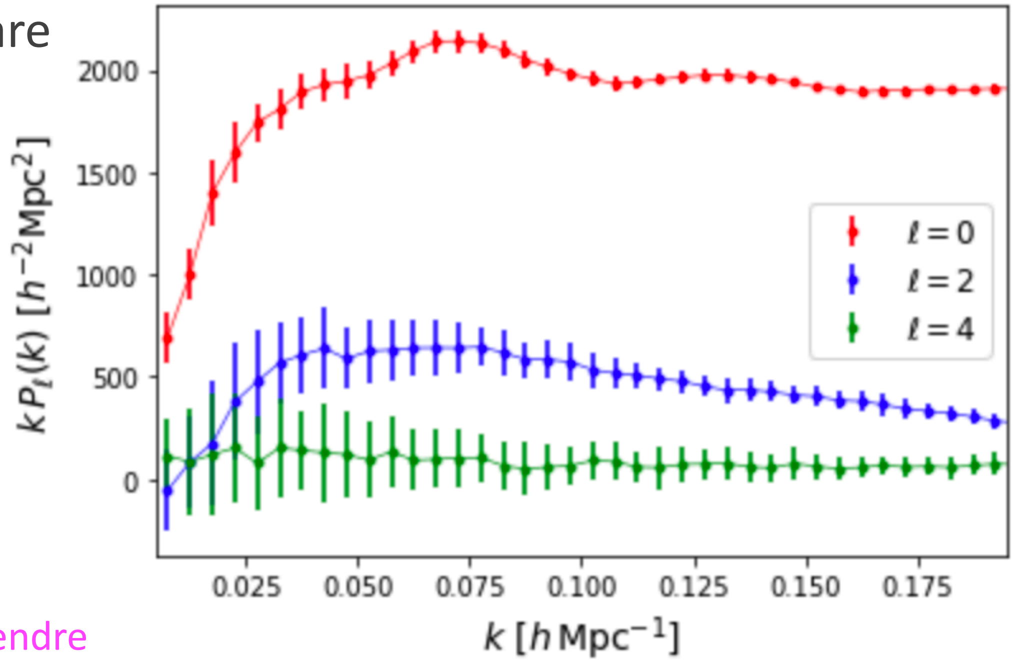
$$\xi(\mathbf{r}) = \sum_\ell \xi_\ell(r) L_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{n}}) \qquad P(\mathbf{k}) = \sum_\ell P_\ell(k) L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})$$

  - The **multipoles** are estimated from the density field:

$$\hat{\xi}_\ell(r) = (2\ell+1)\int \frac{d\Omega_r}{4\pi}\int d\mathbf{x}\,\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})L_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{n}})$$

Legendre Polynomials

$$\boxed{\hat{P}_\ell(k) = \frac{(2\ell+1)}{V}\int \frac{d\Omega_k}{4\pi}\int d\mathbf{r}_1\,d\mathbf{r}_2\,e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)}\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})}$$

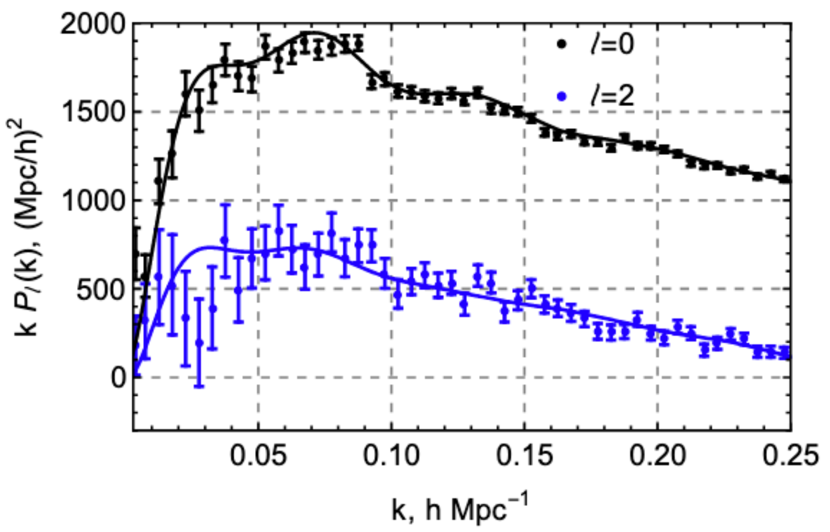


MultiDark-Patchy Mocks
(Kitaura+16)

Parameter Inference

CMB-Strength Parameter Constraints, including 1.6% on H_0!

BOSS DR12 Spectra



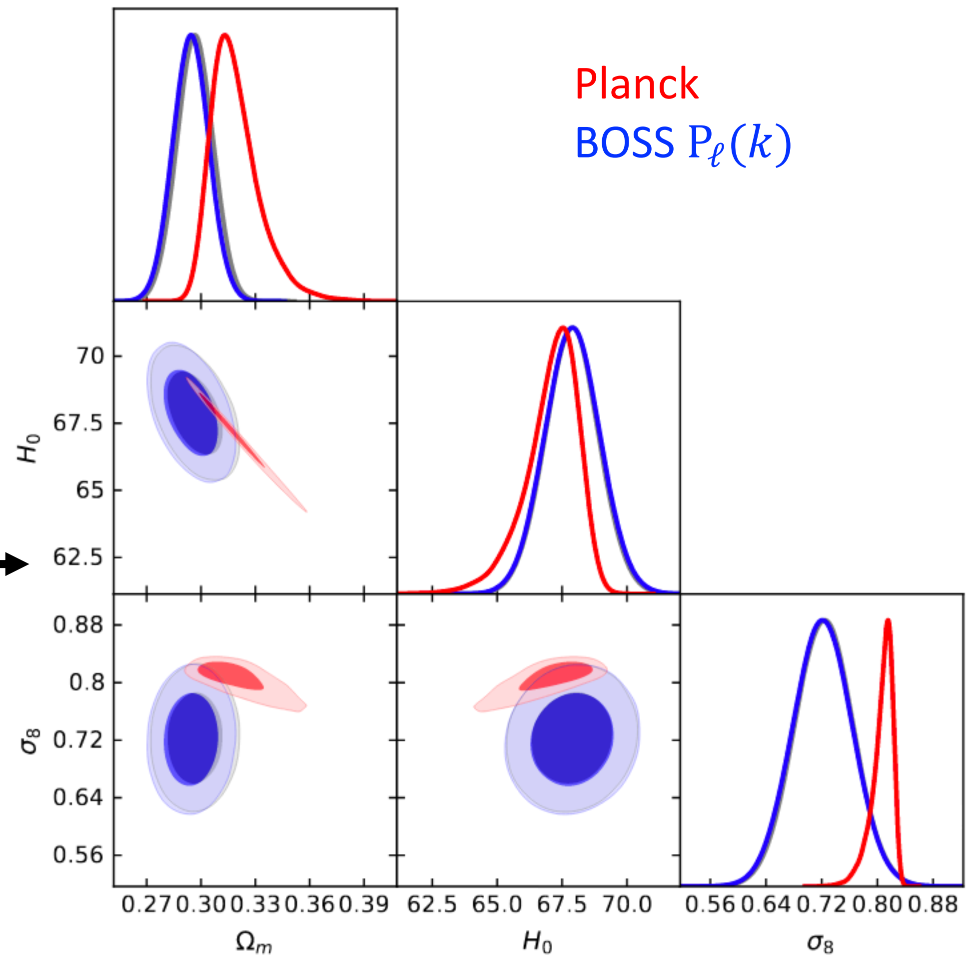
$$P_{g,\ell}(k) = P^{\rm tree}_{g,\ell}(k) + P^{\rm 1-loop}_{g,\ell}(k) + P^{\rm noise}_{g,\ell}(k) + P^{\rm ctr}_{g,\ell}(k)$$

Linear Theory | 1-loop PT | Shot-noise | Counterterms

Theory Model
(Effective Field Theory)

MCMC

Mock Datasets
(at least N_{bin} mocks)



Ivanov+19, Philcox+20a

Beyond 2-Point Statistics

- In a **non-Gaussian** universe, there is more information in **higher-point** functions, e.g.

  - Bispectrum / 3PCF [Gil-Marín+16, Slepian+15, d'Amico+19]

$$(2\pi)^3\delta_D\left(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3\right)B(\mathbf{k}_1,\mathbf{k}_2)=\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle$$

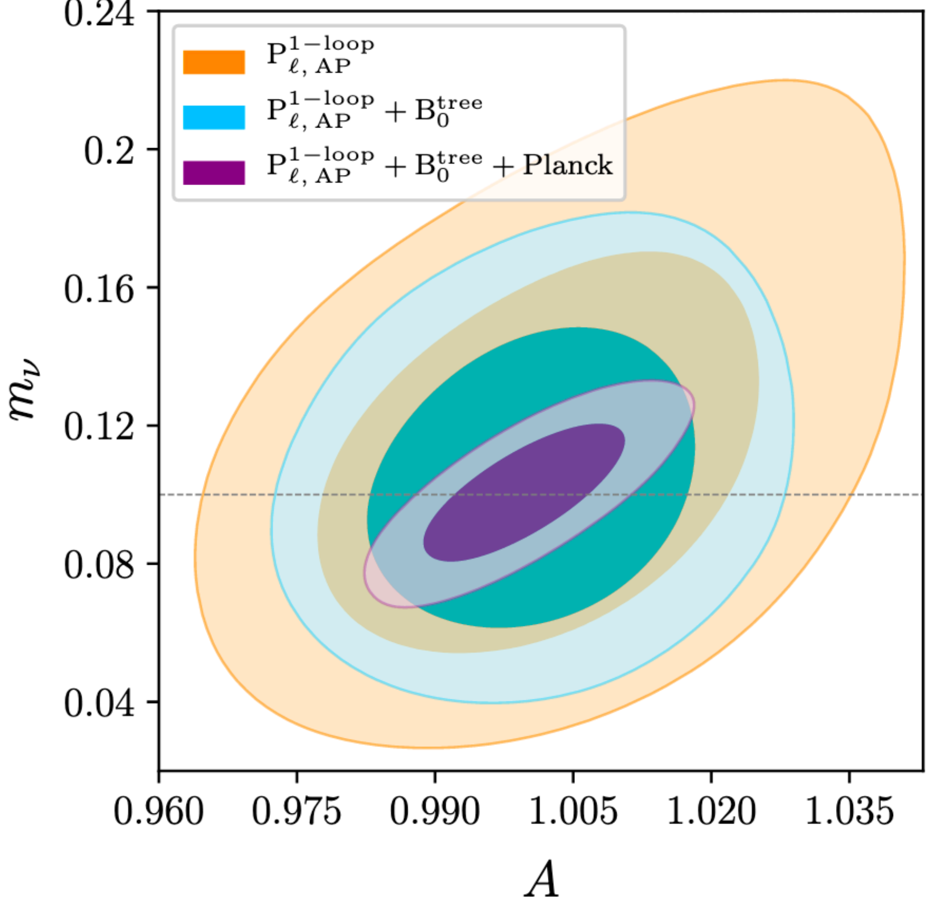
  - Trispectrum / 4PCF [Gualdi+20]

$$(2\pi)^3\delta_D\left(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4\right)T(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)=\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4)\rangle$$

- These get steadily larger and harder to measure.

  - Not used in many cosmological analyses yet!

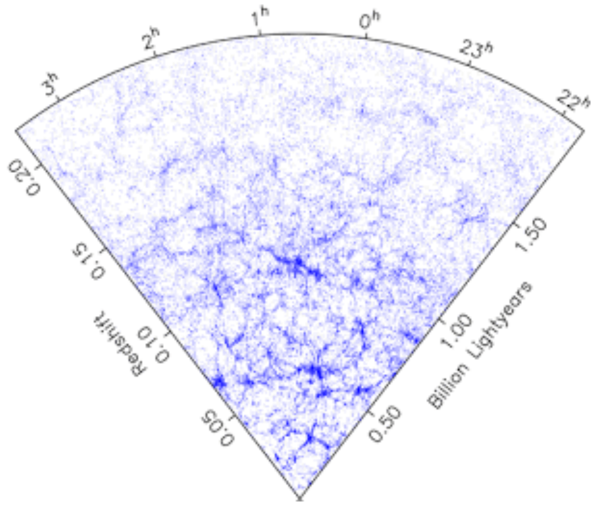
3-point Forecasts for Euclid

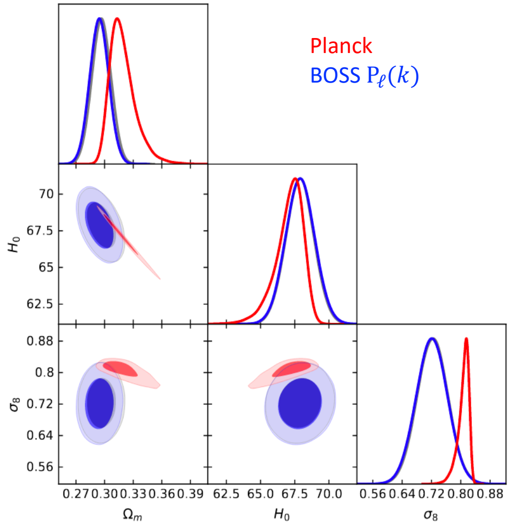


Chudaykin & Ivanov 19

Cosmology from $P_\ell(k)$: A Summary

- Fundamental observable: the galaxy **overdensity field**
- $P_\ell(k)$ parametrized by pair separation and **line-of-sight angle**
- Power spectrum estimators measure $|\delta(\mathbf{k})|^2 L_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}})$
- Computed using **Fast Fourier Transforms** (FFTs)
- Derive constraints by comparing data and theory with MCMC, using **mock galaxy catalogs** to define covariances





Cosmology from $P_\ell(k)$: A Summary

- Fundamental observable: the galaxy **overdensity field** — *Is this the best field to use?*
- $P_\ell(k)$ parametrized by pair separation and **line-of-sight angle** — *How do we define this angle?*
- Power spectrum estimators measure $|\delta(\mathbf{k})|^2 L_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}})$ — *Can we estimate it more optimally?*
- Computed using **Fast Fourier Transforms** (FFTs) — *Are FFTs always the most efficient?*
- Derive constraints by comparing data and theory with MCMC, using **mock galaxy catalogs** to define covariances — *Can data-compression help?*

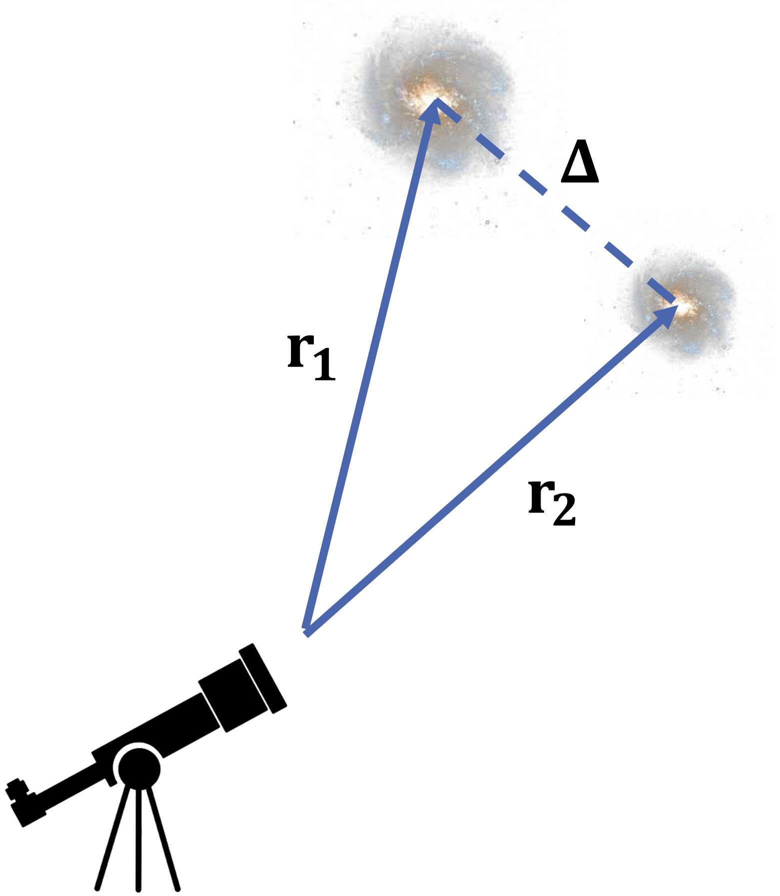
1. Parametrizing Anisotropy

Philcox & Slepian (in prep.)

Choosing the Line-of-Sight

- Galaxy correlation function depends on the angle between the separation vector $\boldsymbol{\Delta}$ and the line-of-sight $\hat{\mathbf{n}}$:

$$\hat{\xi}_\ell(r) = \frac{2\ell+1}{V}\int d\mathbf{r}_1\, d\mathbf{r}_2\, \underbrace{\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{\text{Density Fields}} \overbrace{L_\ell(\hat{\boldsymbol{\Delta}}\cdot\hat{\mathbf{n}})}^{\text{Angular Dependence}} \underbrace{\left[\frac{\delta_D(r-\Delta)}{4\pi r^2}\right]}_{\text{Binning}}$$



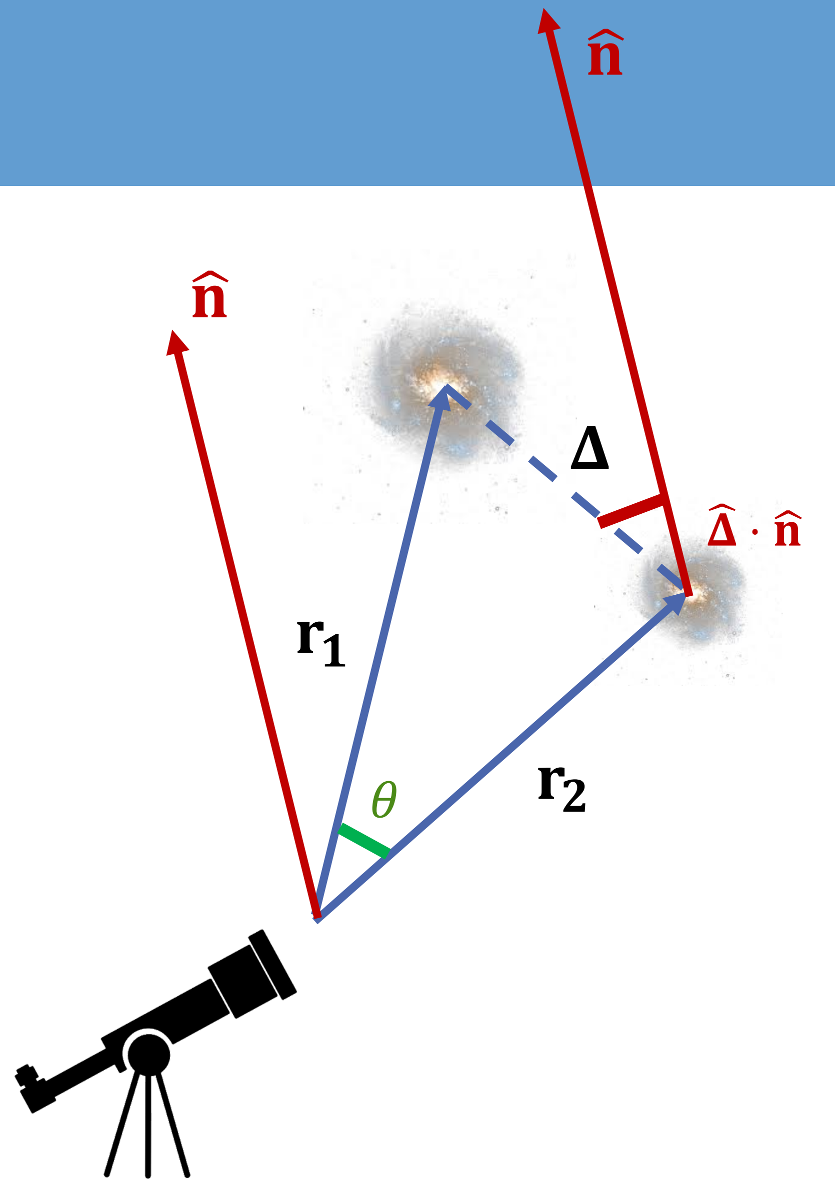
Slepian & Eisenstein 15,
Philcox & Slepian (in prep.)

Choosing the Line-of-Sight

- Galaxy correlation function depends on the angle between the separation vector $\boldsymbol{\Delta}$ and the line-of-sight $\hat{\mathbf{n}}$:

$$\hat{\xi}_\ell(r) = \frac{2\ell+1}{V}\int d\mathbf{r}_1\, d\mathbf{r}_2\, \underbrace{\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{\text{Density Fields}}\overbrace{L_\ell(\hat{\boldsymbol{\Delta}}\cdot\hat{\mathbf{n}})}^{\text{Angular Dependence}}\underbrace{\left[\frac{\delta_D(r-\Delta)}{4\pi r^2}\right]}_{\text{Binning}}$$

- Options:
  - **Fixed $\hat{\mathbf{n}}$** : $\mathcal{O}(\theta^0)$ error, for survey size θ



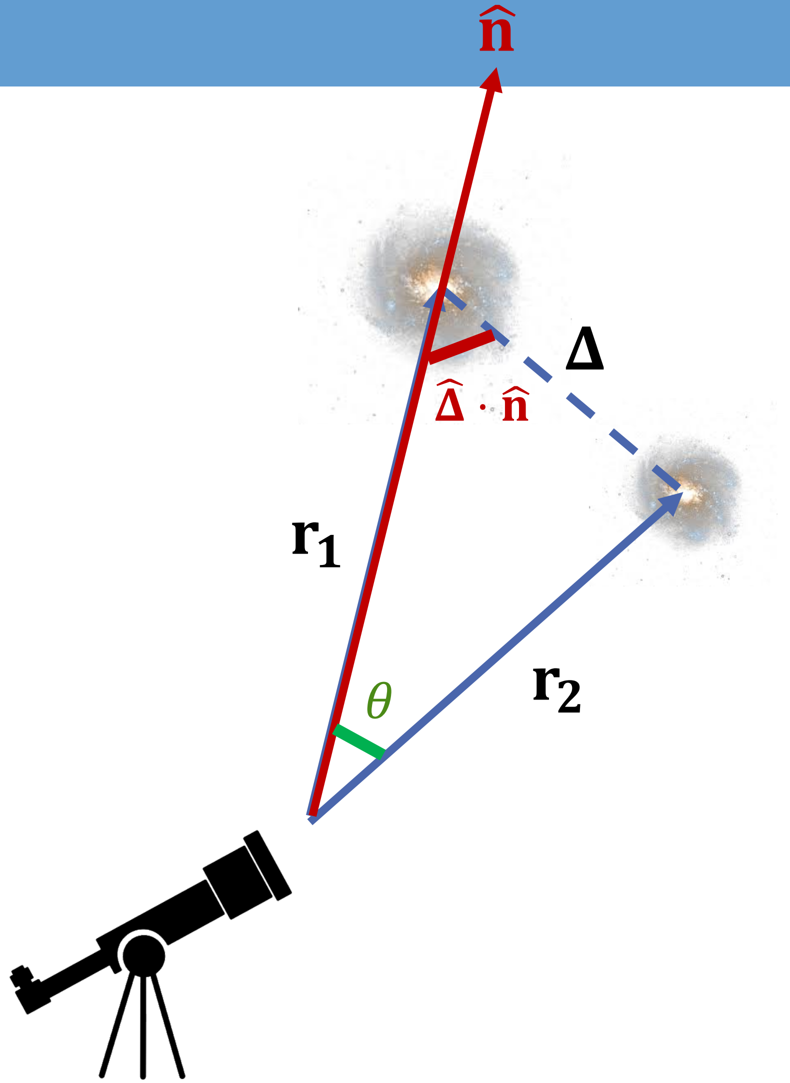
Slepian & Eisenstein 15,
Philcox & Slepian (in prep.)

Choosing the Line-of-Sight

- Galaxy correlation function depends on the angle between the separation vector $\mathbf{\Delta}$ and the line-of-sight $\hat{\mathbf{n}}$:

$$\hat{\xi}_\ell(r) = \frac{2\ell+1}{V}\int d\mathbf{r}_1\, d\mathbf{r}_2\, \underbrace{\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{\text{Density Fields}}\overbrace{L_\ell(\hat{\mathbf{\Delta}}\cdot\hat{\mathbf{n}})}^{\text{Angular Dependence}}\underbrace{\left[\frac{\delta_D(r-\Delta)}{4\pi r^2}\right]}_{\text{Binning}}$$

- Options:
  - **Fixed** $\hat{\mathbf{n}}$: $\mathcal{O}(\theta^0)$ error, for survey size θ
  - **Yamamoto** approximation: $\hat{\mathbf{n}} = \hat{\mathbf{r}}_1$, $\mathcal{O}(\theta^2)$ error



Slepian & Eisenstein 15,
Philcox & Slepian (in prep.)

Choosing the Line-of-Sight

- Galaxy correlation function depends on the angle between the separation vector $\boldsymbol{\Delta}$ and the line-of-sight $\hat{\mathbf{n}}$:

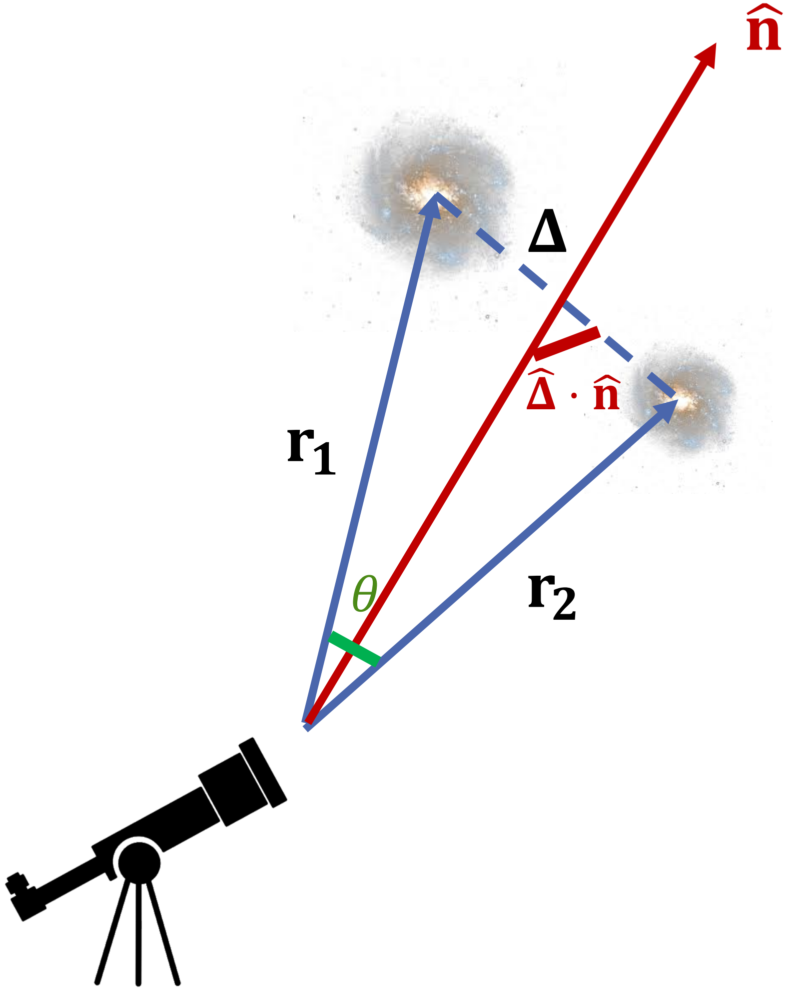
Angular Dependence

$$\hat{\xi}_\ell(r) = \frac{2\ell+1}{V}\int d\mathbf{r}_1\, d\mathbf{r}_2\, \underbrace{\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{\text{Density Fields}} \overbrace{L_\ell(\hat{\boldsymbol{\Delta}}\cdot\hat{\mathbf{n}})}^{\text{Angular Dependence}} \underbrace{\left[\frac{\delta_D(r-\Delta)}{4\pi r^2}\right]}_{\text{Binning}}$$

Density Fields

Binning

- Options:
  - **Fixed** $\hat{\mathbf{n}}$: $\mathcal{O}(\theta^0)$ error, for survey size θ
  - **Yamamoto** approximation: $\hat{\mathbf{n}} = \hat{\mathbf{r}}_1$, $\mathcal{O}(\theta^2)$ error
  - **Midpoint** method: $\hat{\mathbf{n}} = \widehat{\mathbf{r}_1 + \mathbf{r}_2}$, $\mathcal{O}(\theta^{4+})$ error



Slepian & Eisenstein 15,
Philcox & Slepian (in prep.)

Lines-of-Sight in the Power Spectrum

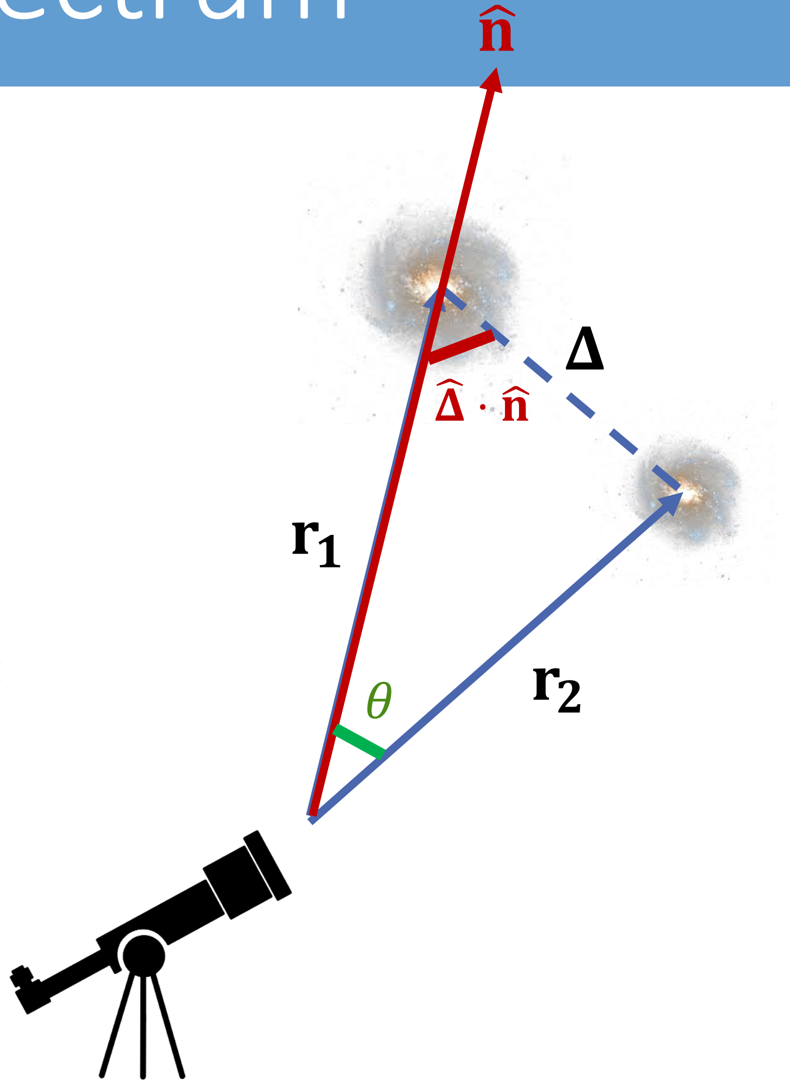
- We can do the same for the **power spectrum**

$$\hat{P}_\ell(k) = \frac{2\ell+1}{V}\int_{\Omega_k}\int d\mathbf{r}_1\, d\mathbf{r}_2\, \overbrace{e^{-i\mathbf{k}\cdot(\mathbf{r}_2-\mathbf{r}_1)}}^{\text{Fourier Transform}}\, \underbrace{\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{\text{Density Fields}}\, \overbrace{L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})}^{\text{Angular Dependence}}$$

- This is **easy to implement** for the Yamamoto approximation, $\hat{\mathbf{n}} = \hat{\mathbf{r}}_1$:

$$\hat{P}_\ell^{\text{Yama}}(k) = \frac{4\pi}{V}\int_{\Omega_k}\left[\sum_{m=-\ell}^{\ell} Y_\ell^{m*}(\hat{\mathbf{k}})\mathcal{F}\left[Y_\ell^m\delta\right](\mathbf{k})\right]\delta^*(\mathbf{k})$$

Spherical Harmonics



Hand+17,
Philcox & Slepian (in prep.)

Lines-of-Sight in the Power Spectrum

- We can do the same for the **power spectrum**

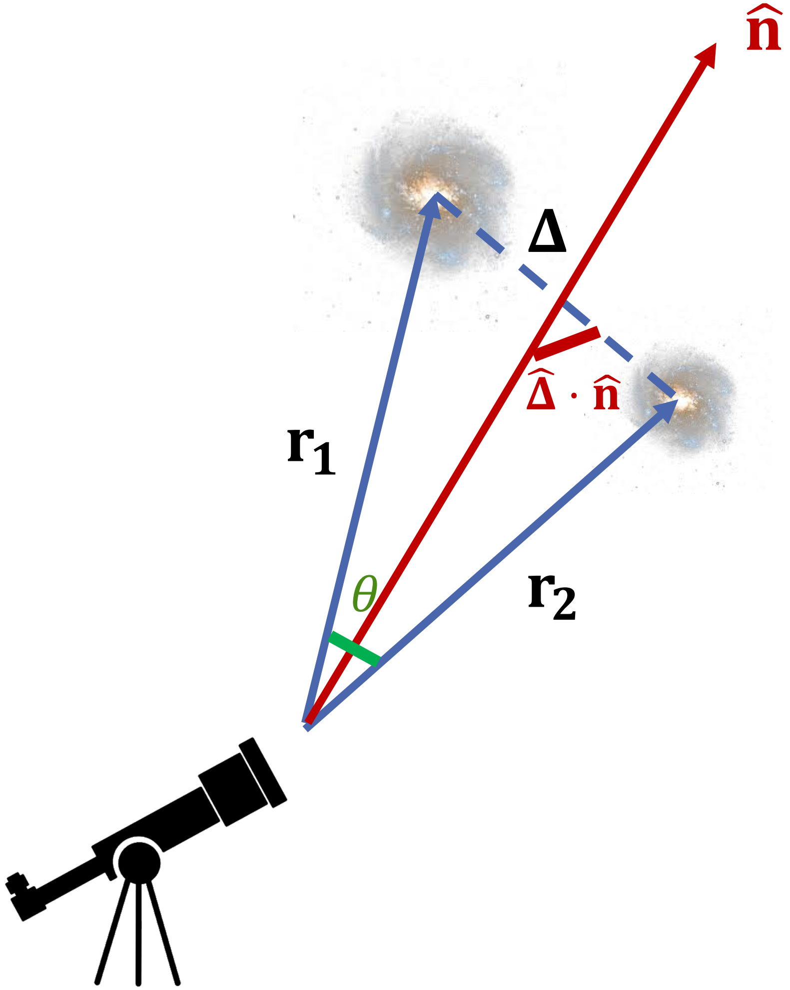
$$\hat{P}_\ell(k) = \frac{2\ell+1}{V}\int_{\Omega_k}\int d\mathbf{r}_1\, d\mathbf{r}_2\, \overbrace{e^{-i\mathbf{k}\cdot(\mathbf{r}_2-\mathbf{r}_1)}}^{\text{Fourier Transform}}\underbrace{\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)}_{\text{Density Fields}}\overbrace{L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})}^{\text{Angular Dependence}}$$

- This is **easy to implement** for the Yamamoto approximation, $\hat{\mathbf{n}} = \hat{\mathbf{r}}_1$:

$$\hat{P}_\ell^{\text{Yama}}(k) = \frac{4\pi}{V}\int_{\Omega_k}\left[\sum_{m=-\ell}^{\ell} Y_\ell^{m*}(\hat{\mathbf{k}})\mathcal{F}\left[Y_\ell^m\delta\right](\mathbf{k})\right]\delta^*(\mathbf{k})$$

Spherical Harmonics

- But **not separable** for the midpoint method!



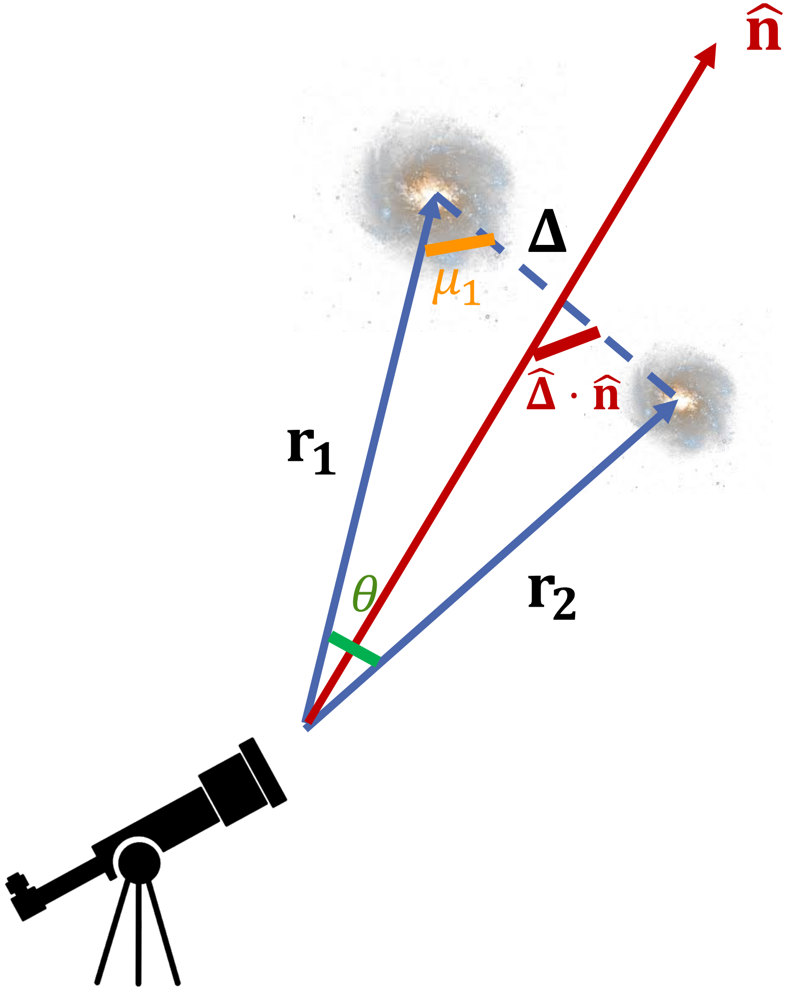
Hand+17,
Philcox & Slepian (in prep.)

Implementing the Midpoint Method

- Use a trick to make the integrals separable:
  - Expand in powers of $\theta \sim \Delta / r_1$:

$$L_\ell(\hat{\mathbf{\Delta}} \cdot \widehat{\mathbf{r}_1 + \mathbf{r}_2}) = \sum_{\alpha=0}^{\infty} \sum_{J=0}^{\ell+\alpha} \underbrace{f_J^{\alpha,\ell}}_{\text{Coefficients}} \overbrace{\left(\frac{\Delta}{2r_1}\right)^\alpha}^{\text{Survey Angle, } \ll 1} \underbrace{L_J(\hat{\mathbf{\Delta}} \cdot \mathbf{r}_1)}_{\text{Yamamoto Piece}}$$

$$\begin{aligned} L_2(\hat{\mathbf{\Delta}} \cdot \widehat{\mathbf{r}_1 + \mathbf{r}_2}) &= L_2(\mu_1) + \frac{6}{5}\left(\frac{\Delta}{2r_1}\right)[L_1(\mu_1) - L_3(\mu_1)] \\ &+ \frac{1}{35}\left(\frac{\Delta}{2r_1}\right)^2 [7L_0(\mu_1) - 55L_2(\mu_1) + 48L_4(\mu_1)] \\ &- \frac{4}{105}\left(\frac{\Delta}{2r_1}\right)^3 [9L_1(\mu_1) - 49L_3(\mu_1) + 40L_5(\mu_1)] \\ &+ \frac{1}{385}\left(\frac{\Delta}{2r_1}\right)^4 [11L_0(\mu_1) + 165L_2(\mu_1) - 816L_4(\mu_1) + 640L_6(\mu_1)] \\ &+ \ldots \end{aligned}$$



Slepian & Eisenstein 15,
Philcox & Slepian (in prep.)

Implementing the Midpoint Method

- Use a trick to make the integrals separable:

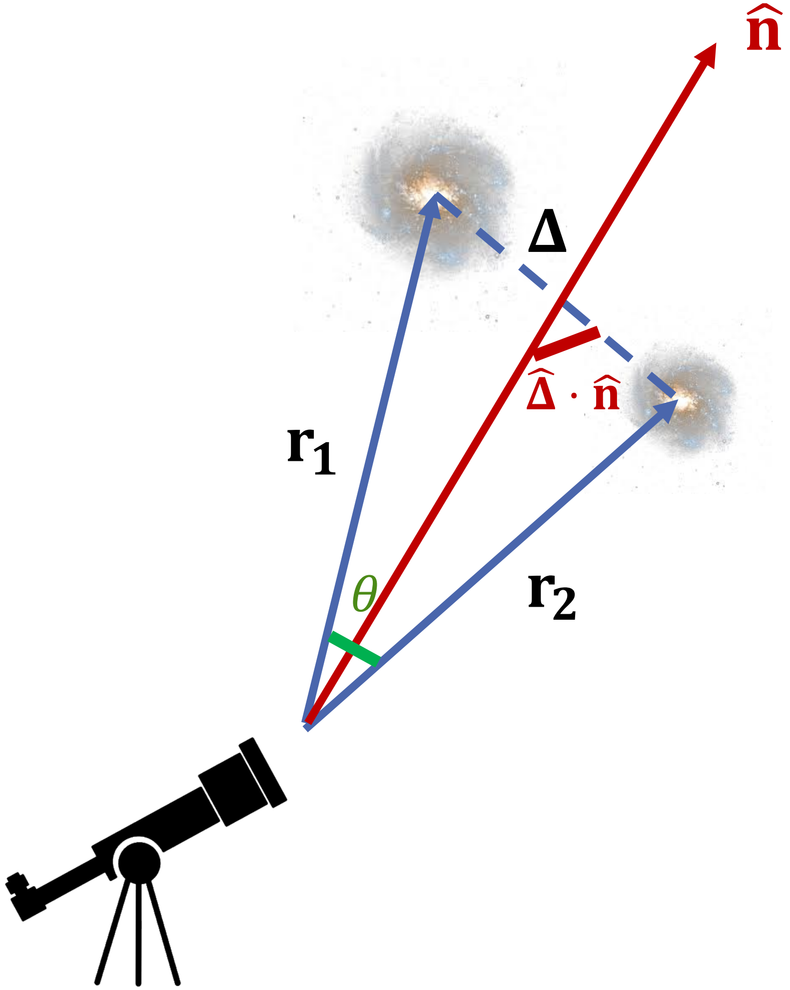
  - Expand in powers of $\theta \sim \Delta / r_1$:

$$L_\ell(\hat{\mathbf{\Delta}} \cdot \widehat{\mathbf{r}_1 + \mathbf{r}_2}) = \sum_{\alpha=0}^{\infty} \sum_{J=0}^{\ell+\alpha} \underbrace{f_J^{\alpha,\ell}}_{\text{Coefficients}} \overbrace{\left(\frac{\Delta}{2r_1}\right)^{\alpha}}^{\text{Survey Angle, } \ll 1} \underbrace{L_J(\hat{\mathbf{\Delta}} \cdot \mathbf{r}_1)}_{\text{Yamamoto Piece}}$$

  - Can now compute the 2PCF using Fourier transforms!

  - This applies also to the **power spectrum**, with a more complex expansion

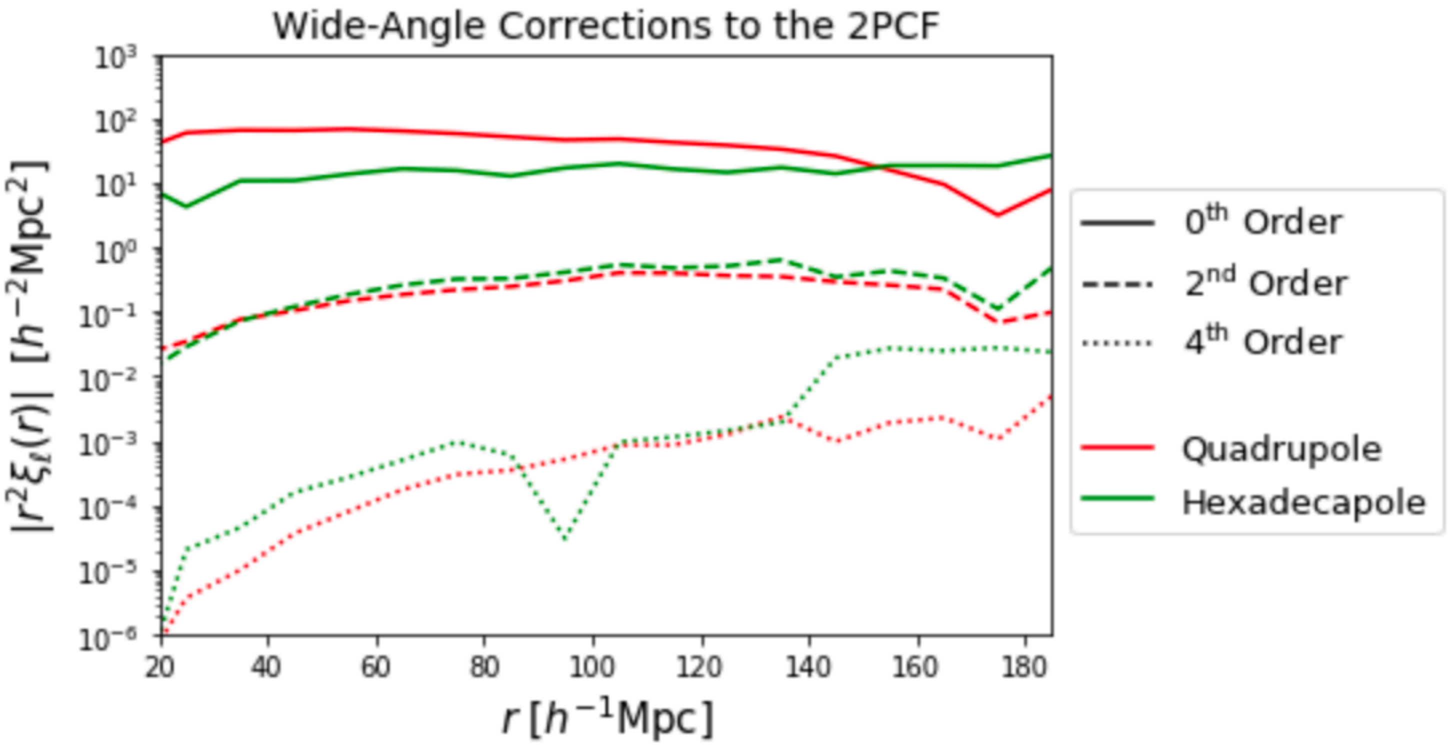
  - Same computational complexity as Yamamoto approximation



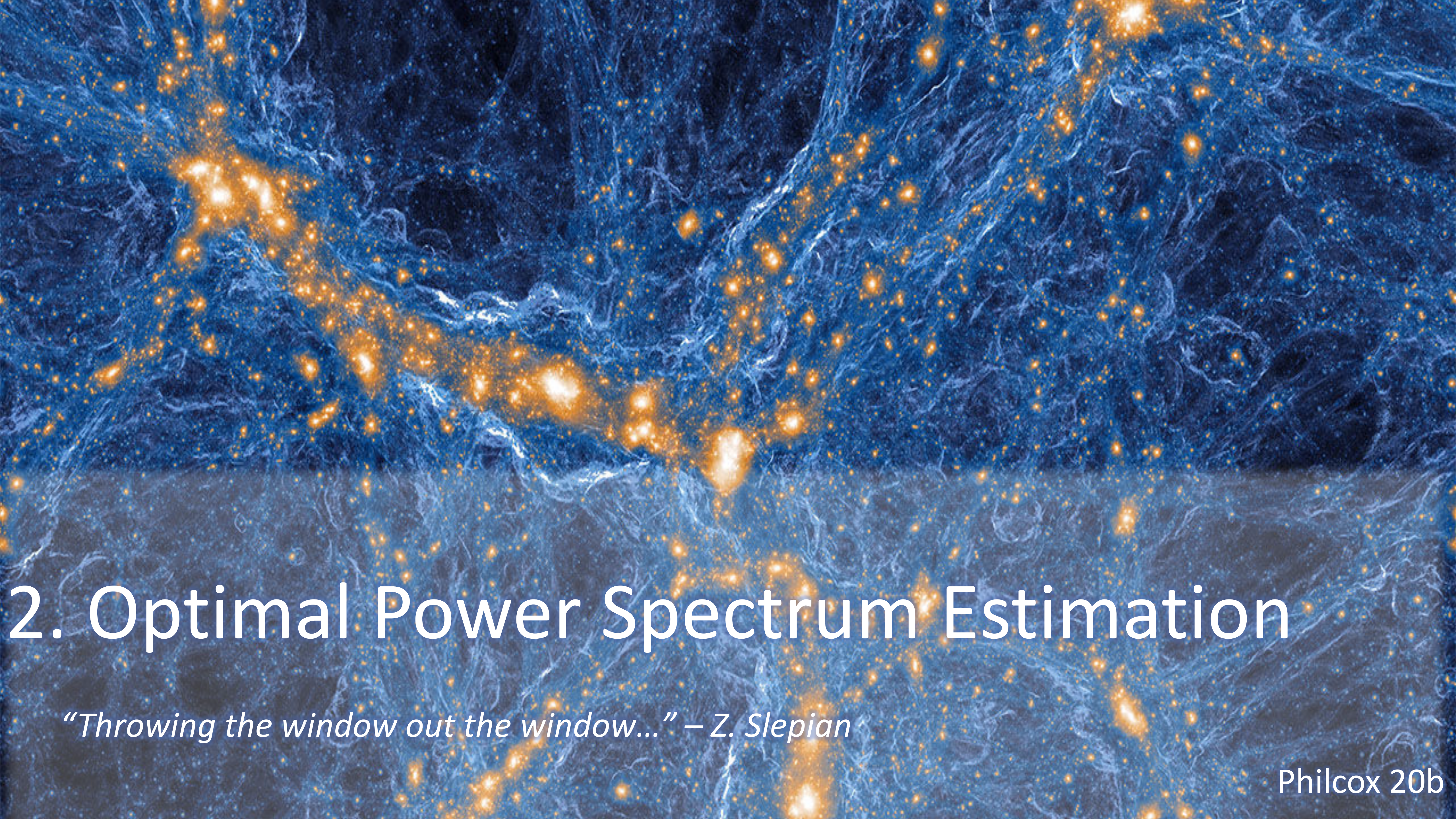
Philcox & Slepian (in prep.)

The Midpoint Method in Practice

- For BOSS correlation function:
  - $\theta \sim 0.1 - 0.2$ at the BAO scale
  - Corrections are **marginally** important
- For BOSS $P(k)$:
  - Spectrum depends on **all** scales:
  - $\theta \sim 1$ for the largest-modes
  - Corrections **are** important
- Most important for **wide surveys** at **low redshifts**
- *Full results coming soon…*



Philcox & Slepian (in prep.)



2. Optimal Power Spectrum Estimation

“Throwing the window out the window…” – Z. Slepian

Philcox 20b

The FKP Estimator

The power spectrum **isn't** simply the density field squared.

- Neglects **inhomogeneous noise** and **survey window functions**

1. Define $\delta(\mathbf{r})$ as the difference between **galaxy** and **random** densities
2. Add an **FKP weight** to incorporate **Poisson noise** densities (and systematics)

This is the **optimal solution** on small-scales with Poisson noise

But:

- Not optimal on large scales
- Measures the **window-convolved** power spectrum

$$\hat{P}(k) = \int \frac{d\Omega_k}{4\pi} \int d\mathbf{r}_1\, d\mathbf{r}_2\, e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \delta(\mathbf{r}_1)\delta(\mathbf{r}_2)$$

Galaxies *Randoms*

$$\delta(\mathbf{r}) \to \frac{w(\mathbf{r})[n_g(\mathbf{r}) - \alpha_r n_r(\mathbf{r})]}{I^{1/2}}, \qquad I \equiv \int d\mathbf{r}\, w^2(\mathbf{r})\bar{n}^2(\mathbf{r})$$

Systematics

$$w(\mathbf{r}) = \frac{w_{\mathrm{sys}}(\mathbf{r})}{1 + P_{\mathrm{FKP}}\, n(\mathbf{r})}$$

Poisson Noise Correction, $P_{\mathrm{FKP}} \sim 10^4$

Feldman+94, Tegmark+98

Optimal Estimators

- **Optimal approach:**
  - Maximize the **likelihood** of the data, **d**, given band-powers **p** and data-covariance C(**p**)

$$-2\log L[\mathbf{d}](\mathbf{p}) = \mathbf{d}^T \mathsf{C}^{-1}(\mathbf{p})\mathbf{d} + \mathrm{Tr}\log \mathsf{C}(\mathbf{p}) + \text{const.} \longleftarrow \boxed{\textit{Gaussian likelihood}}$$

  - Gives a **maximum-likelihood** estimator for the **unwindowed** power spectrum**:**

$$\underbrace{\hat{p}_\alpha^{\mathrm{ML}}}_{\text{Band-powers}} = \overbrace{p_\alpha^{\mathrm{fid}}}^{\text{Fiducial Model}} + \frac{1}{2}\sum_\beta \underbrace{F^{-1}_{\alpha\beta}}_{\text{Fisher Matrix}} \left(\underbrace{\mathbf{d}^T \mathsf{C}^{-1}}_{} \overbrace{\mathsf{C}_{,\beta}}^{\text{Covariance Derivative}} \underbrace{\mathsf{C}^{-1}\mathbf{d}}_{\text{Weighted Data}} - \overbrace{\mathrm{Tr}\left[\mathsf{C}^{-1}\mathsf{C}_{,\beta}\right]}^{\text{Bias Term}}\right)$$

  - The estimator is a **quadratic** function of the data, $\hat{q}_\beta$, plus a normalization $F_{\alpha\beta}$, and a fiducial model, p_α^{fid}

$$\hat{p}_\alpha^{\mathrm{ML}} = \underset{\text{Known!}}{p_\alpha^{\mathrm{fid}}} + \sum_\beta \underset{\text{From sims}}{F^{-1}_{\alpha\beta}} \left(\underset{\text{From data}}{\hat{q}_\beta} - \bar{q}_\beta\right) \longleftarrow \text{From sims}$$

Tegmark 97,
Philcox 20b

Implementing the ML Estimator

$$\hat{p}_\alpha^{\mathrm{ML}} = p_\alpha^{\mathrm{fid}} + \sum_\beta F_{\alpha\beta}^{-1}\left(\hat{q}_\beta - \bar{q}_\beta\right)$$

- This depends on the **quadratic estimator** $\hat{q}_\beta$:

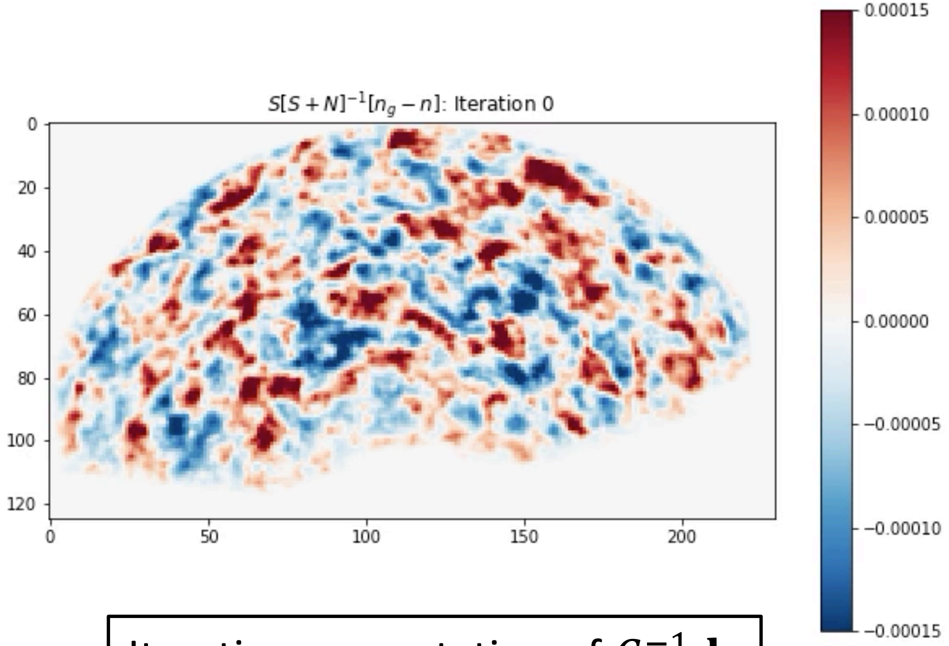
$$\hat{q}(k) = \int \frac{d\Omega_k}{4\pi} \int d\mathbf{r}\, d\mathbf{r}'\, e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \left[\mathsf{C}^{-1}\mathbf{d}\right](\mathbf{r}) \left[\mathsf{C}^{-1}\mathbf{d}\right](\mathbf{r}')$$

  - Just a power spectrum of the **inverse-covariance** weighted data

- We need the the covariance for each pair of pixels:

$$\mathsf{C}(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}') \int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \sum_\ell P_\ell(k) L_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}') + (1+\alpha)n(\mathbf{r})\delta_D(\mathbf{r}-\mathbf{r}')$$

- This covariance is **gigantic** ($N_{\mathrm{pix}} \times N_{\mathrm{pix}}$)
  - Never store directly
  - Invert using **conjugate gradient descent** methods



Iterative computation of $C^{-1}\mathbf{d}$ from an initial guess

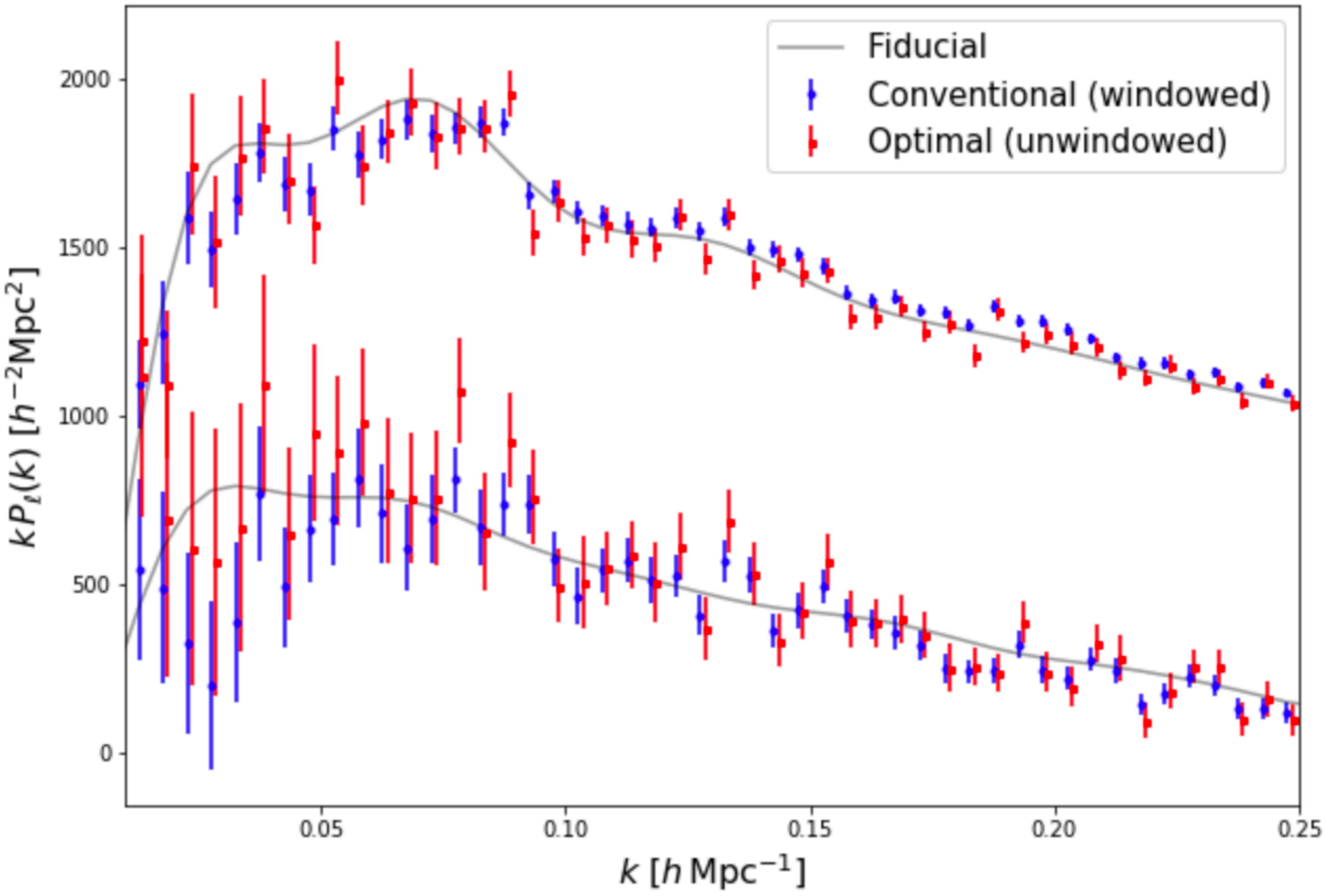
Philcox 20b

Implementing the ML Estimator

$$\hat{p}_\alpha^{\mathrm{ML}} = p_\alpha^{\mathrm{fid}} + \sum_\beta F_{\alpha\beta}^{-1}\left(\hat{q}_\beta - \bar{q}_\beta\right)$$

<u>Basic Pipeline:</u>

1. Choose a **fiducial** set of cosmological parameters
2. Define the **pixel covariance** $C(\mathbf{p})$
3. Compute the **quadratic estimator** on the data, $\hat{q}_\beta$
4. Repeat on a suite of **simulations**, **m**, to get the bias, $\bar{q}_\beta$ and Fisher matrix, $F_{\alpha\beta}$
5. Combine to get the **quadratic estimator**
6. ***Optional:*** Repeat with new fiducial parameters



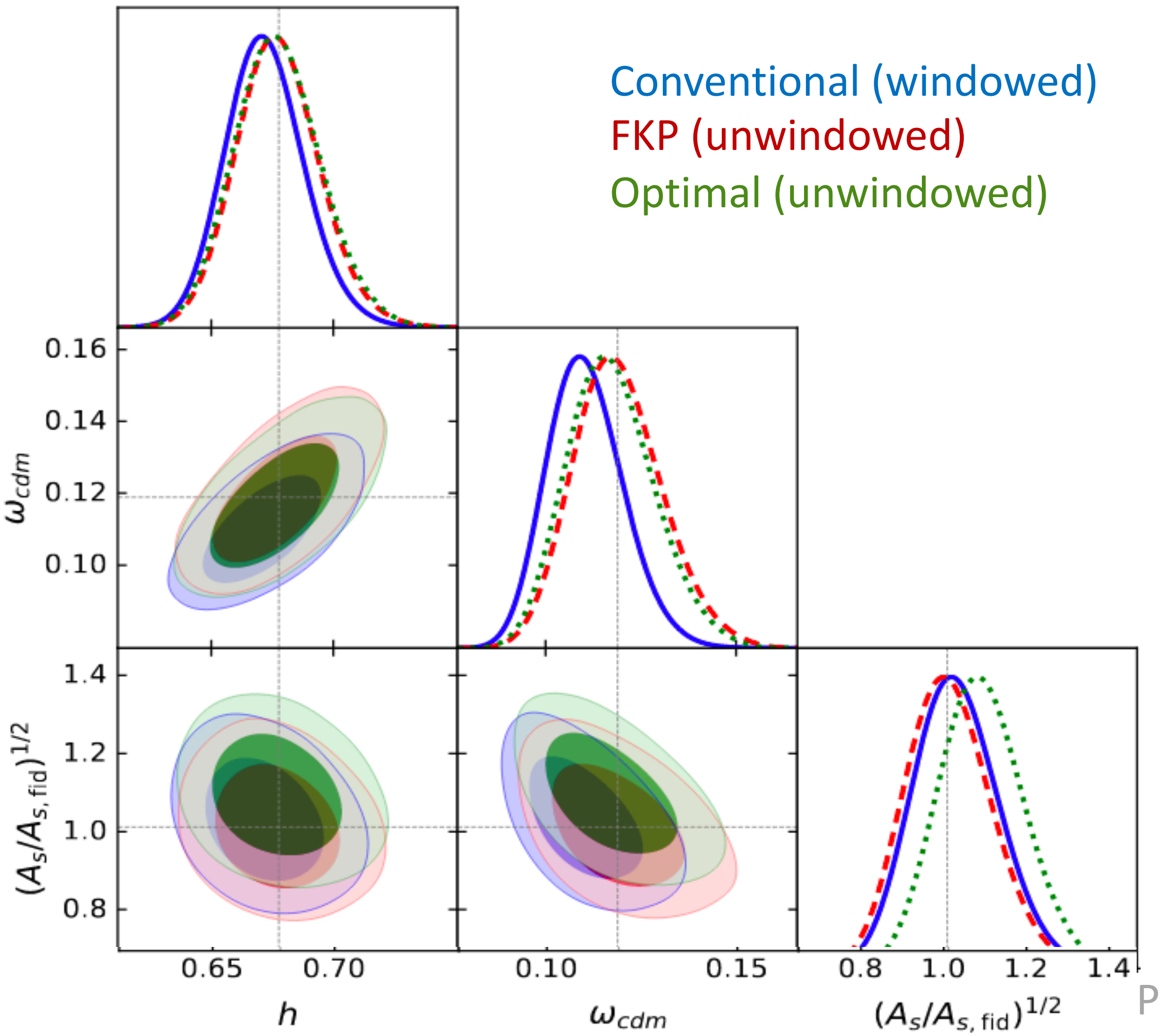
Philcox 20b

Implementing the ML Estimator

$$\hat{p}_\alpha^{\mathrm{ML}} = p_\alpha^{\mathrm{fid}} + \sum_\beta F_{\alpha\beta}^{-1}\left(\hat{q}_\beta - \bar{q}_\beta\right)$$

<u>Basic Pipeline:</u>

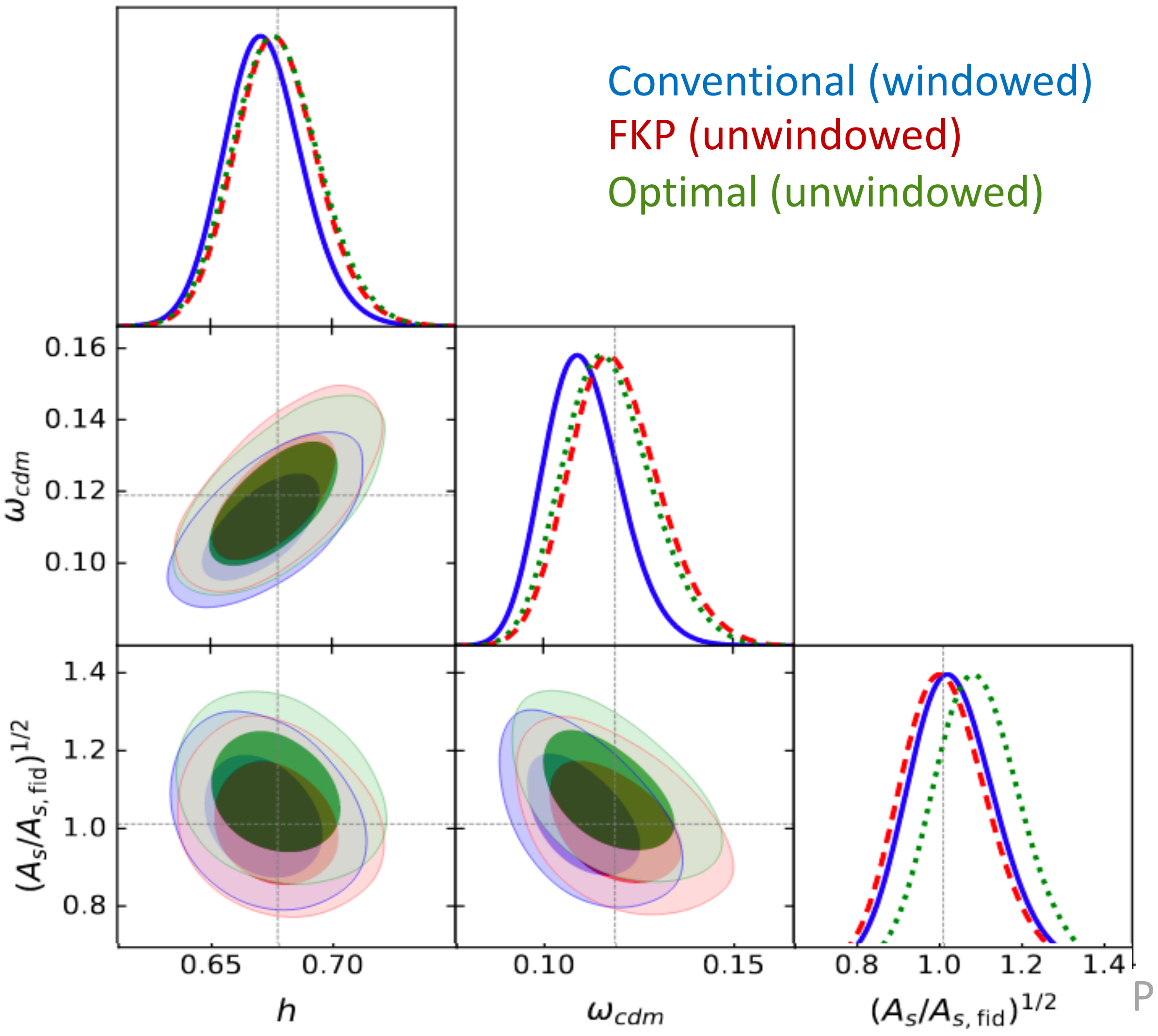
1. Choose a **fiducial** set of cosmological parameters
2. Define the **pixel covariance** $C(\mathbf{p})$
3. Compute the **quadratic estimator** on the data, $\hat{q}_\beta$
4. Repeat on a suite of **simulations**, **m**, to get the bias, $\bar{q}_\beta$ and Fisher matrix, $F_{\alpha\beta}$
5. Combine to get the **quadratic estimator**
6. ***Optional:*** Repeat with new fiducial parameters



Philcox 20b

Is this useful?

- **Benefits**:
  - No window-convolution required!
  - No leading-order **gridding** effects
  - No leading-order non-Poissonian **shot noise** effects
  - **Optimal** error-bars if Gaussian
    - Most useful for **small**, **dense**, **anisotropic** surveys, and **large-scale** modes
- **Extensions:**
  - Optimal cubic $P_\ell(k)$ estimators for weak **non-Gaussianity**
  - Window-free **bispectrum** estimators



Philcox 20b

3. Power Spectra without FFTs

Philcox & Eisenstein 19, Philcox 20a

Configuration-Space $P(k)$ Estimators

- Conventionally, $P(k)$ is estimated using Fast Fourier Transforms

$$\hat{P}(k) = \int \frac{d\Omega_k}{4\pi} \int d\mathbf{r}_1\, d\mathbf{r}_2\, e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \delta(\mathbf{r}_1)\delta(\mathbf{r}_2) = \int \frac{d\Omega_k}{4\pi} \left|\mathrm{FFT}\left[\delta\right](\mathbf{k})\right|^2$$

- This has complexity $\mathcal{O}(N_g \log N_g)$ for N_g grid points, with **Nyquist frequency** $k_{\mathrm{Nyq}} \propto N_g/L_{\mathrm{box}}$
  - On **small scales**, we need large N_g => **slow computation** and **high memory** usage!

$$\boxed{Time \propto k_{\max} \log k_{\max}}$$

- The 2PCF can be computed by **counting pairs of particles** with $\mathcal{O}(N^2)$ complexity. In bin a:

Weights

$$\xi^a = \int d\mathbf{r}_1\, d\mathbf{r}_2\, \delta(\mathbf{r}_1)\delta(\mathbf{r}_2)\Theta^a(|\mathbf{r}_1 - \mathbf{r}_2|) = \sum_{i\neq j} w_i w_j \Theta^a(|\mathbf{r}_i - \mathbf{r}_j|)$$

Binning function

Sum over galaxies

- This is **fast** on small scales!

$$\boxed{Time \propto N n R_{\max}^3}$$

Philcox & Eisenstein 19, Philcox 20

Configuration-Space $P(k)$ Estimators

- We can do the same for $P(k)$:

$$P(k) \propto \int \frac{d\Omega_k}{4\pi} \sum_{i \neq j} w_i w_j e^{-i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} = \sum_{i\neq j} w_i w_j j_0(k|\mathbf{r}_i - \mathbf{r}_j|)$$

Weights

0^{th} order Bessel function

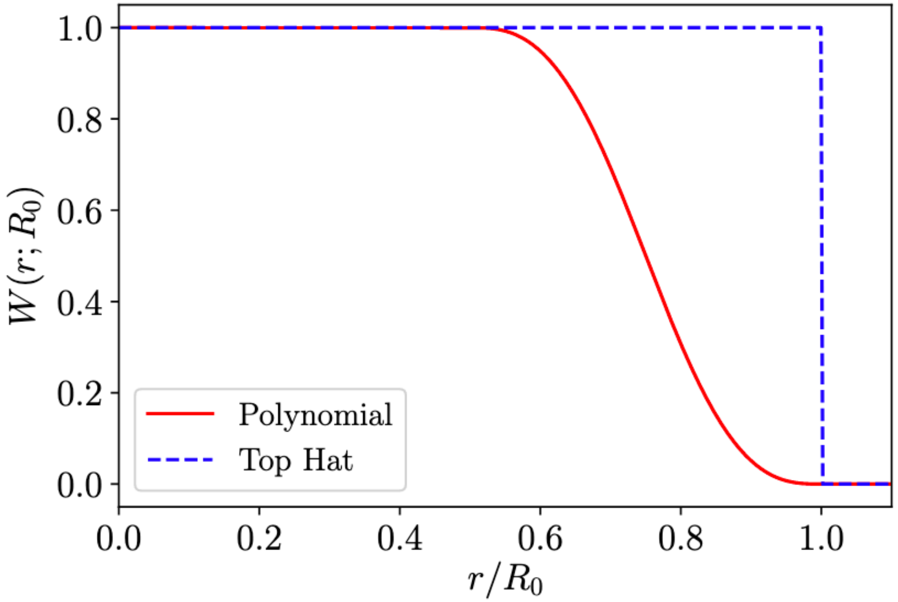
Sum over galaxies

- But we need to sum over all N^2 pairs of galaxies in the survey!

  - Only sum up to some maximum radius R_0, via a smooth function $W(r; R_0)$

$$P(k; R_0) \propto \sum_{i\neq j} w_i w_j j_0(k|\mathbf{r}_i - \mathbf{r}_j|) W(|\mathbf{r}_i - \mathbf{r}_j|; R_0)$$

$$\boxed{Time \propto N n R_0^3 \propto k_{\min}^{-3}}$$



Philcox & Eisenstein 19, Philcox 20

Configuration-Space $P_\ell(k)$ Estimators

This extends to the multipoles too:

Weights

l^{th} order Bessel function

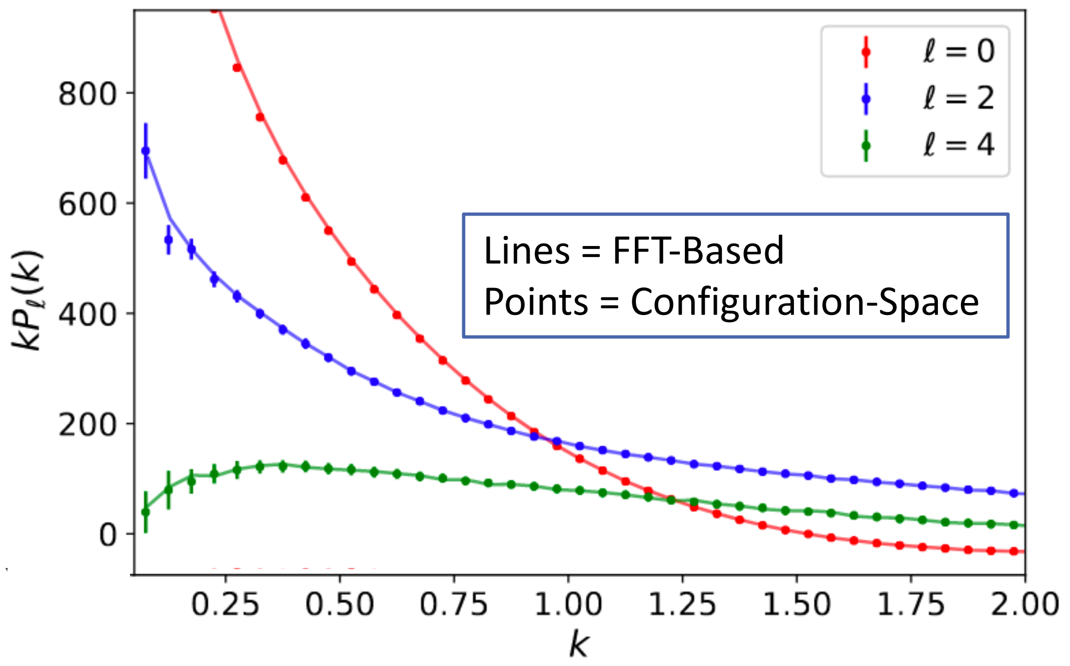
$$P_\ell(k;R_0) \propto (2\ell+1)(-i)^\ell \sum_{i\neq j} w_i w_j j_\ell(k|\mathbf{r}_i-\mathbf{r}_j|)W(|\mathbf{r}_i-\mathbf{r}_j|;R_0)$$

Window function

Sum over galaxies

Benefits

- **Speed**
  - Time scales as k_{min}^{-3}
- **Memory**
  - No storage of large FFT grids
- **Aliasing**
  - No gridding!
- **Shot-noise**
  - Removes self-counts -> Poissonian shot-noise!
- **Window function**
  - Can remove survey window, just as for 2PCF

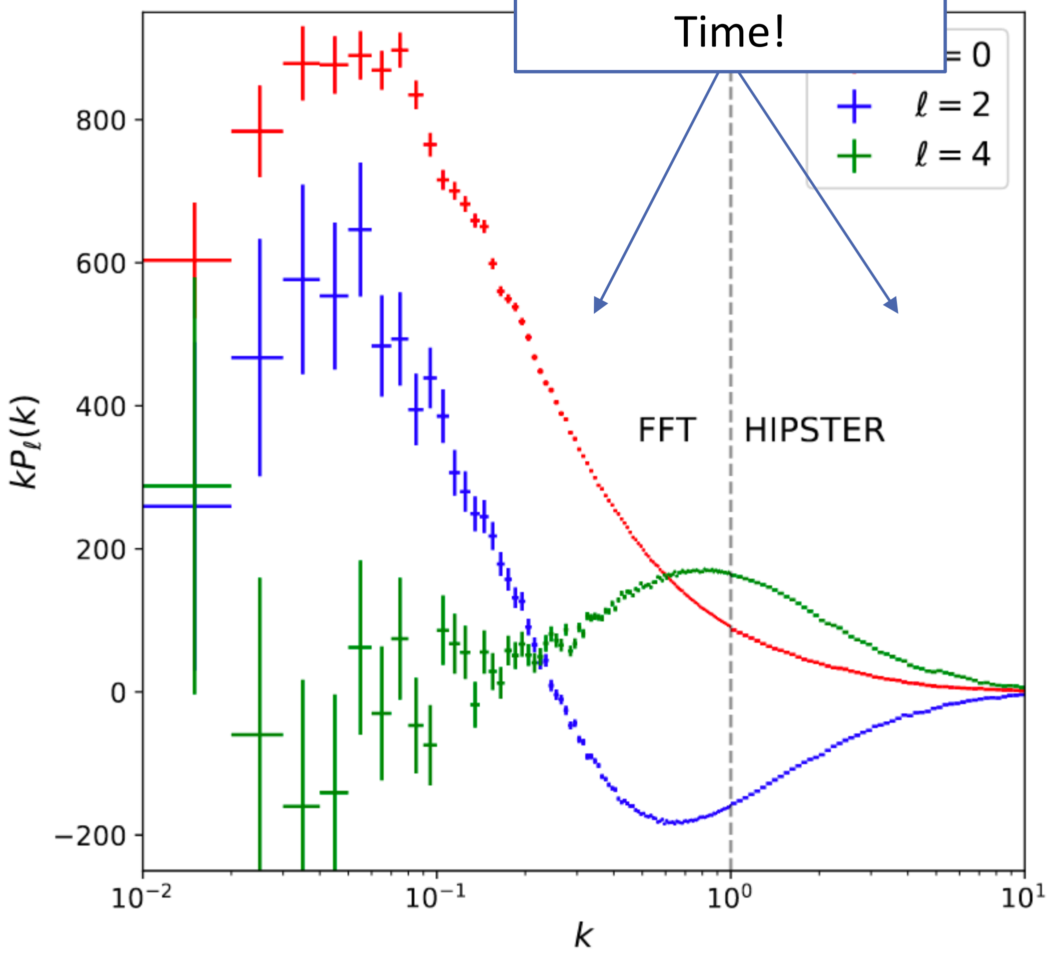


Philcox & Eisenstein 19, Philcox 20

Configuration-Space $P(k)$ Estimators

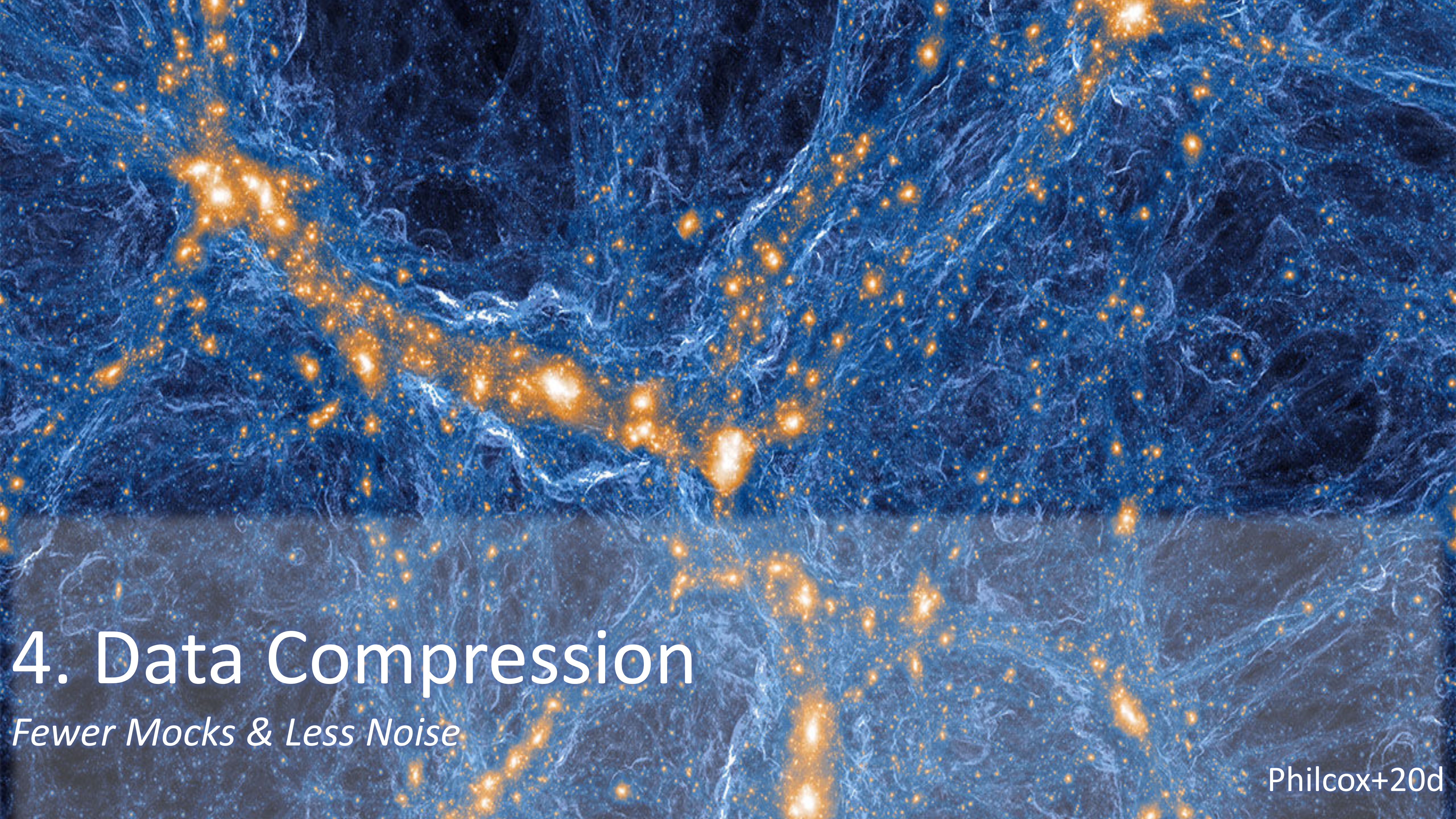
- Implemented in the **HIPSTER** code
- For fastest implementation, **combine** with FFT-based treatments
  - FFTs are fastest on **large** scales (time $\sim k_{\max} \log \mathrm{k}_{\max}$)
  - HIPSTER is fastest on **small** scales (time $\sim k_{\min}^{-3}$)
- Can be similarly applied to **bispectra**
  - Here, time $\propto Nn^2R_0^6 \propto k_{\min}^{-6}$
  - Same scaling with number density as for $P(k)$!

Same Computation Time!



Philcox & Eisenstein 19, Philcox 20

HIPSTER.rtfd.io



4. Data Compression

Fewer Mocks & Less Noise

Philcox+20d

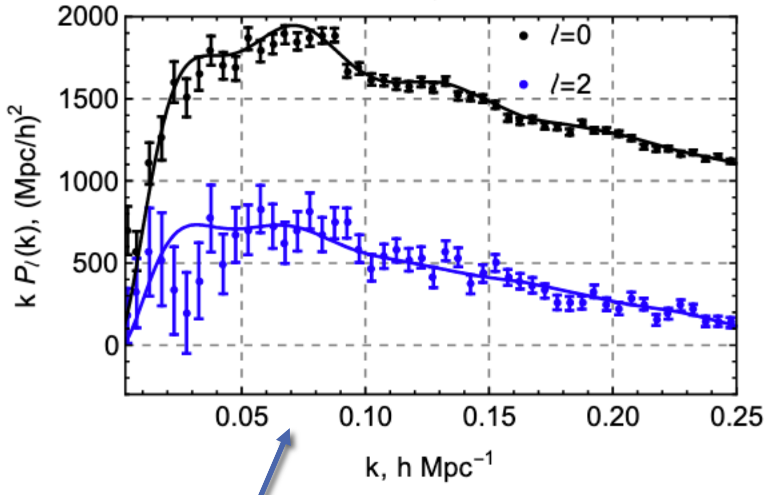
The Curse of Dimensionality

- The power spectrum is **high-dimensional**, e.g.;
  - BOSS has ~ 100 power spectrum bins
  - Higher order statistics will have many more!
  - Only use these to measure ~ 10 parameters

- For a Gaussian likelihood, we need the **inverse covariance**

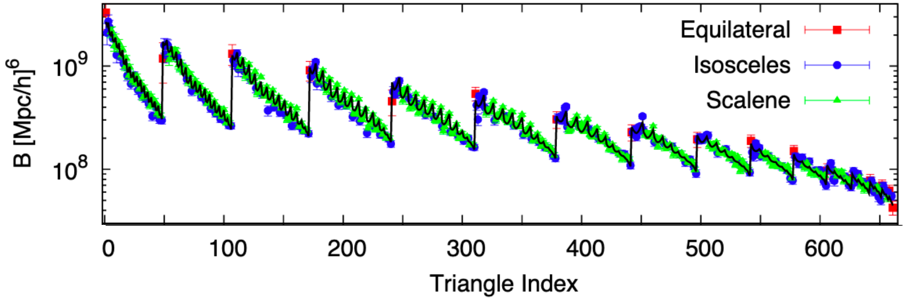
- For a **sample** covariance matrix:
  - Need $N_{mocks} > N_{bins}$ for invertibility
  - Covariance matrix noise gives parameter **shifts**, unless $N_{mocks} \gg N_{bins}$
  - Common solution: **inflate** the error bars and lose constraining power

- What's the best way to compress our data-set?



Power Spectrum

Bispectrum



Gil-Marin+16, Ivanov+19

Data Compression via PCA

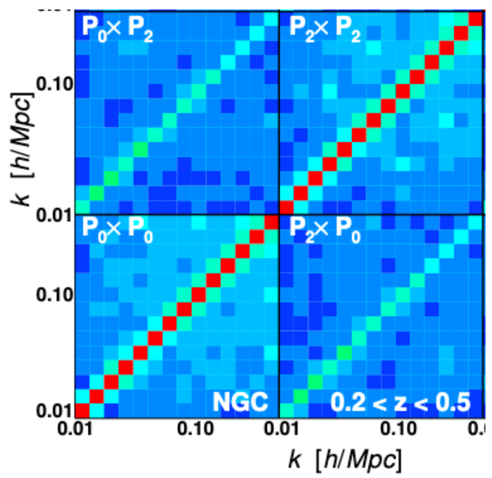
Beutler+16

- A canonical approach: [e.g. Scoccimarro 2000]
  - Form the theoretical covariance matrix of the observable
  - Perform a **P**rincipal **C**omponent **A**nalysis on this
  - Project the data onto the first few components

- This chooses the basis vectors which contribute most to the signal-to-noise of the statistic

- But they might not be the directions we're most interested in!

Power Spectrum Covariance



PCA

$$P(k) \approx \sum_i \mathrm{a}_i W_i(k)$$

Basis Vectors

Coefficients

See also MOPED: Heavens+00, Alsing+18 , KL: Tegmark+97

Data Compression via Subspace Projection

A new* approach, given a **theory model**:

- Draw sets of parameters from the **priors**
- Compute the **theory model** at each point
- Perform a **S**ingular **V**alue **D**ecomposition on the **noise-weighted** samples
- Use these **basis vectors** to perform the compression

Parameters used in the analysis

$$\theta = \{\omega_{\rm cdm}, A_s/A_{s,\rm fid}, h, \ldots\} \times \{b_1, b_2, b_{G_2}, b_4, c_{s,0}, c_{s,2}, P_{\rm shot}\}$$

$$X_a(\theta) \equiv \sum_{ab} \mathsf{C}_{ab}^{-1/2}\left[P_b(\theta) - \overline{P}_b\right]$$

Covariance Estimate *Theory Model*

$$X_{ia} = \sum_\alpha U_{i\alpha} D_\alpha V_{\alpha a}$$

Basis Vectors

$$X_a^{(i)} \approx \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha^{(i)} V_{\alpha a}$$

Subspace Coefficients

The basis vectors are the components that contribute most to the **log-likelihood**

*somewhat inspired by gravitational wave analyses [e.g. Roulet+19]

Philcox+20d

Data Compression via Subspace Projection

Benefits

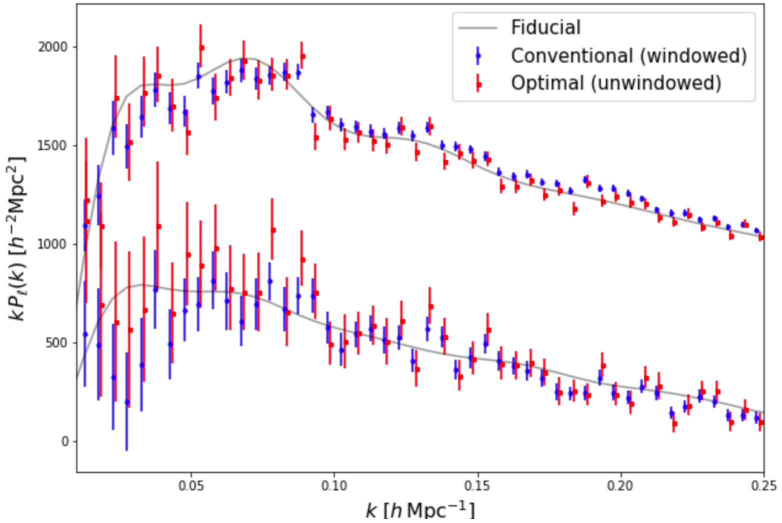
- This is the **best** linear compression for a **specific** analysis
- The subspace covariance is almost **diagonal**
- Can set the **number** of basis vectors **robustly** by fixing χ^2-error
- Can estimate coefficients **optimally** and **directly**

For BOSS 10-parameter analysis:

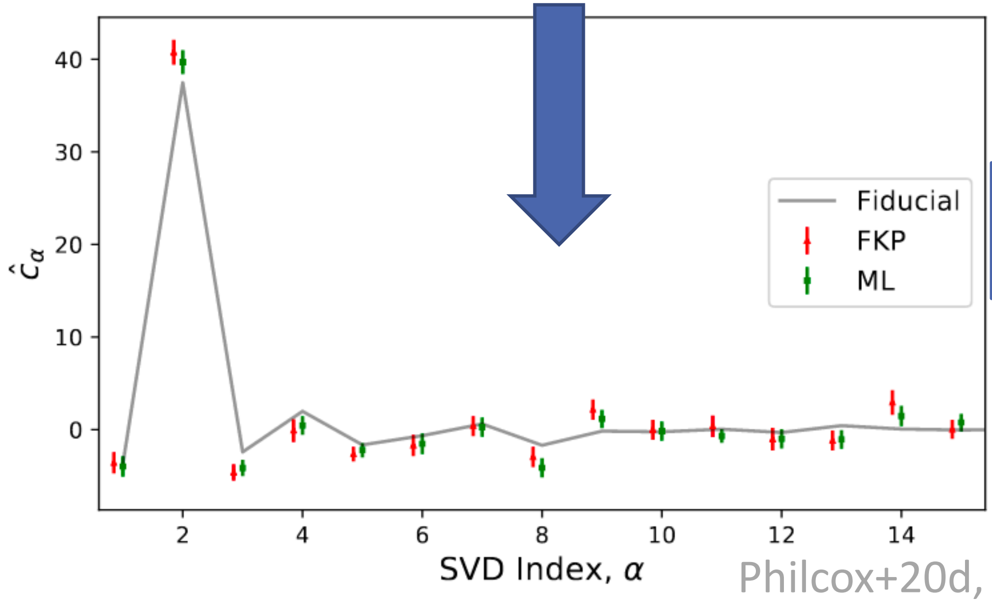
- **96**-bin P(k) -----> **12** subspace coefficients with $\Delta\chi^2 < 0.1$
- **2135**-bin B(k_1,k_2) ----> **8** subspace coefficients

This is applicable to **any** analysis given:

1. Theory Model
2. Parameter Priors
3. Rough Covariance Estimate



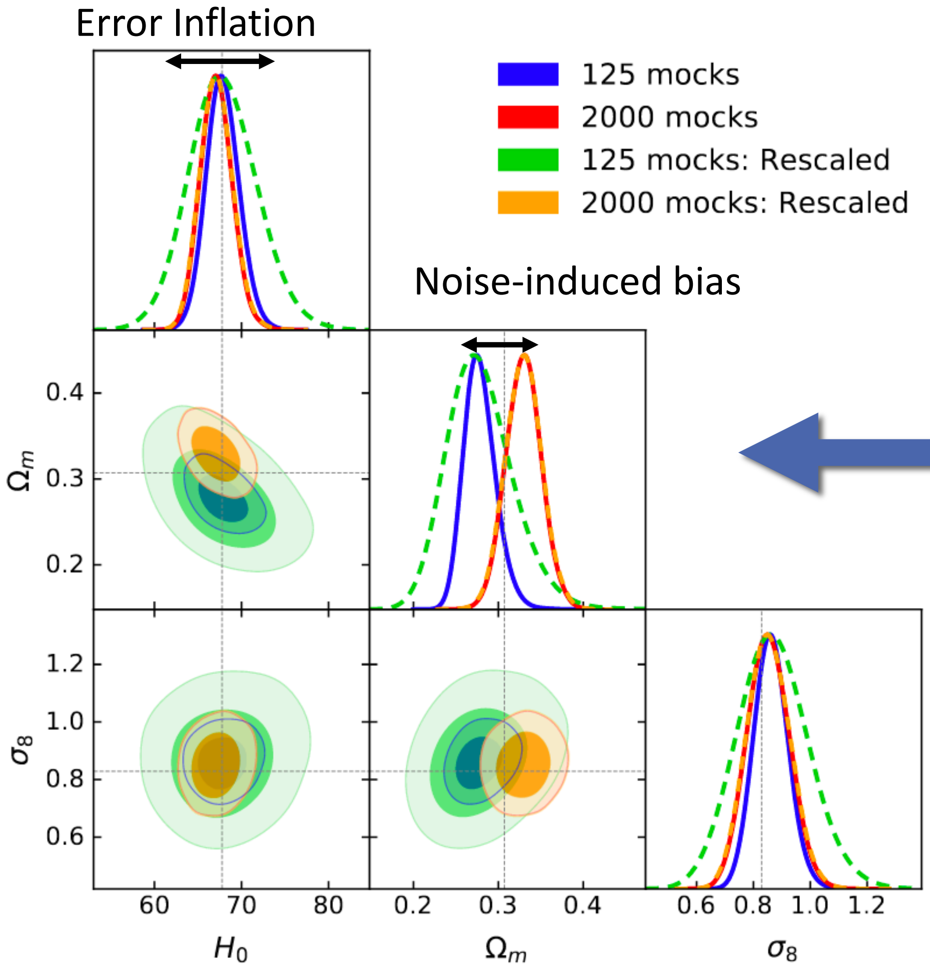
Power Spectrum



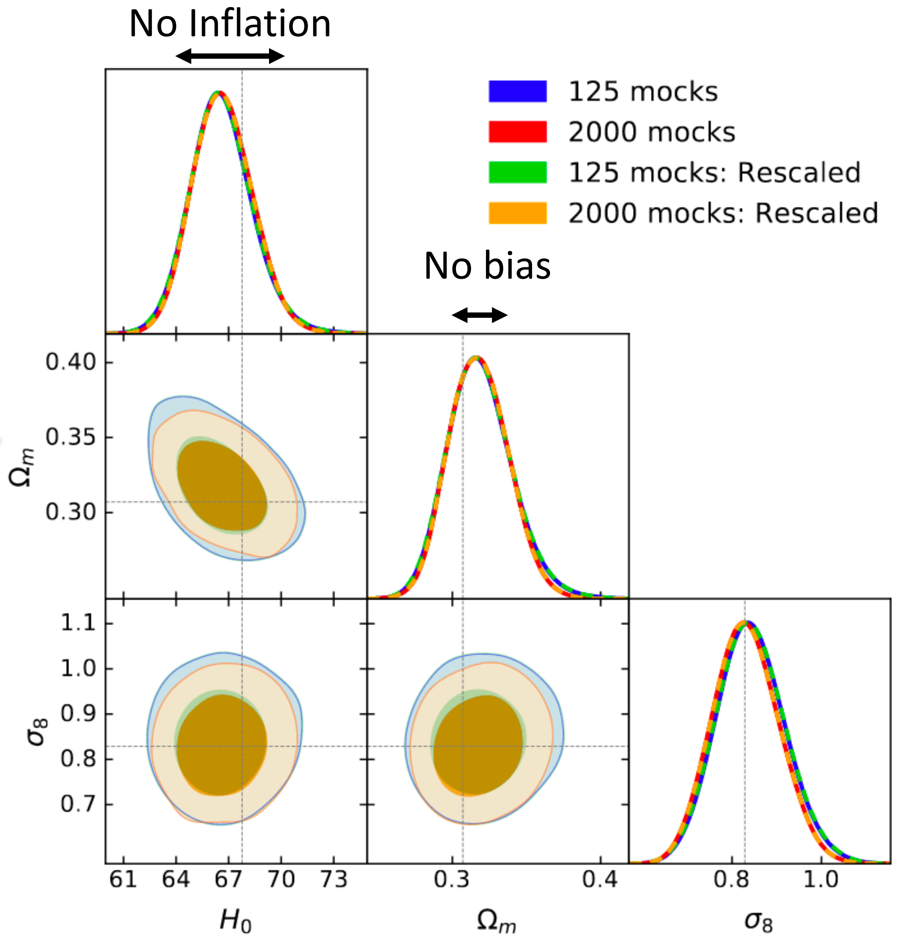
Subspace Coefficients

Philcox+20d, Philcox 20b

Too Few Mocks -> Parameter Biases

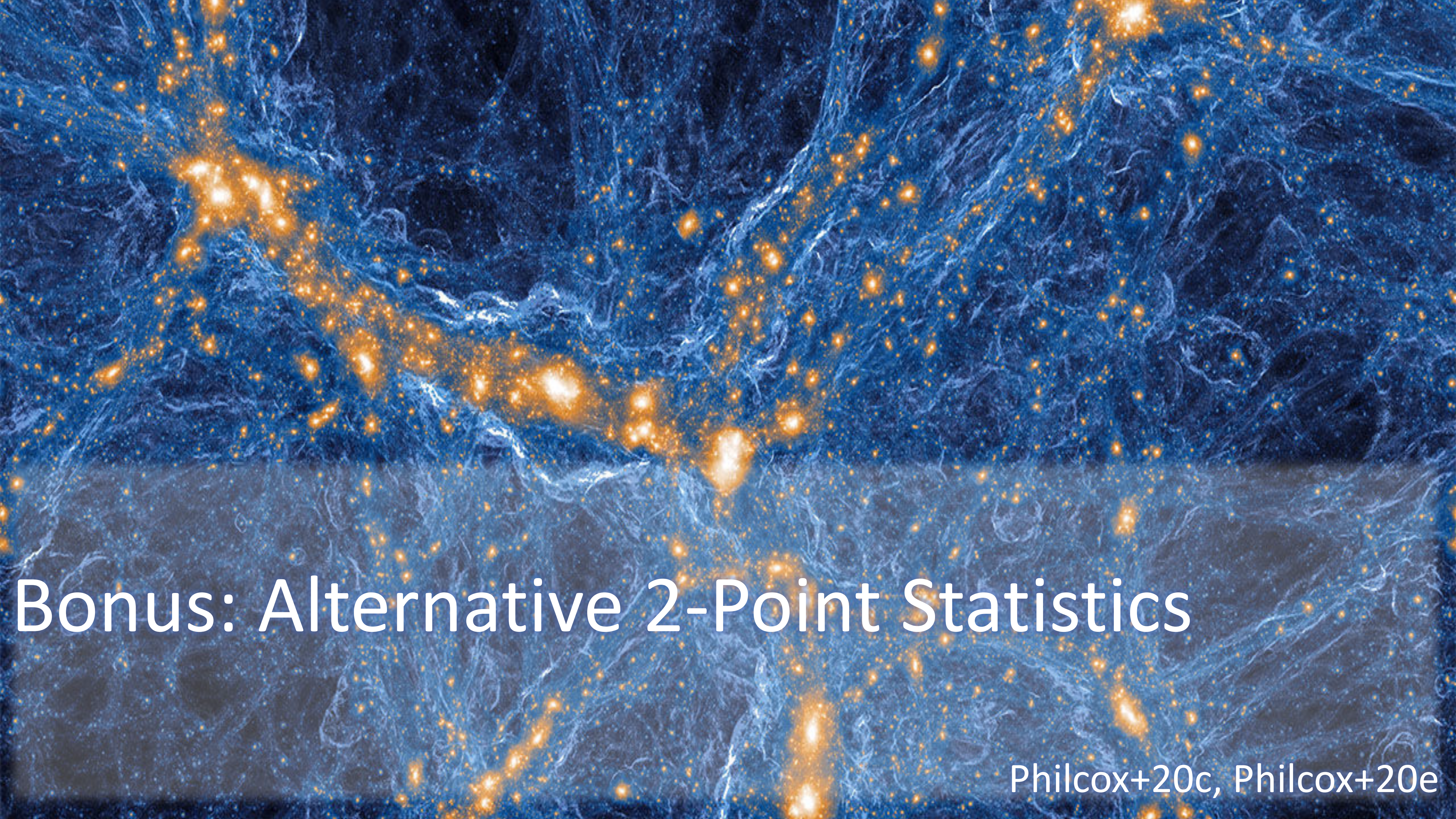


(a) 96-bin Power Spectrum



(c) 12 Subspace Coefficients

Philcox+20d

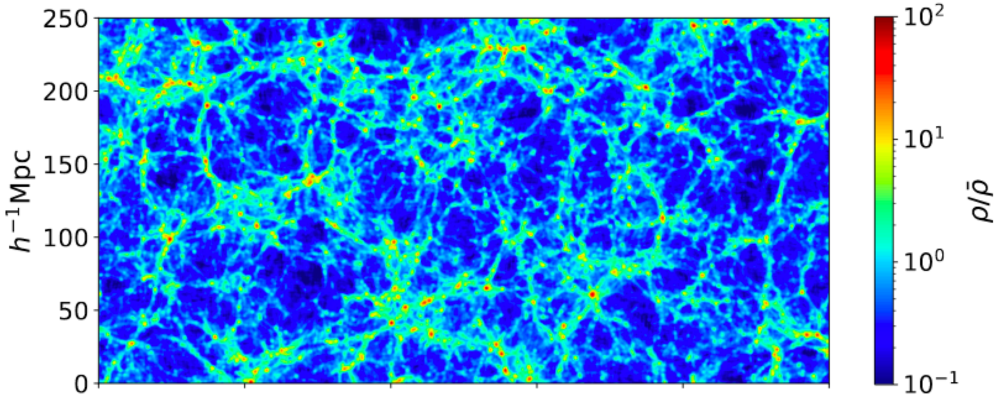

Bonus: Alternative 2-Point Statistics
Philcox+20c, Philcox+20e

Beyond the Density Field

- What's the best field to compute the two-point function of?
  - For a **Gaussian** universe, the power-spectrum of galaxy overdensity contains **all** the information

- The Universe is **not** Gaussian:
  - Information **cascades** to the higher-point functions
  - **Low-density regions** carry a lot of cosmological information, and contribute little to δ [e.g. Pisani+19]

- Various **transformed** fields have been proposed:
  - Reconstructed Density Fields [e.g. Eisenstein+07]
  - Log-normal Transforms [Neyrinck+09, Wang+11]
  - Gaussianized Density Fields [Weinberg 92, Neyrinck+17]
  - Marked Density Fields [Stoyan 84, White 16, Massara+20]



Overdensity Field

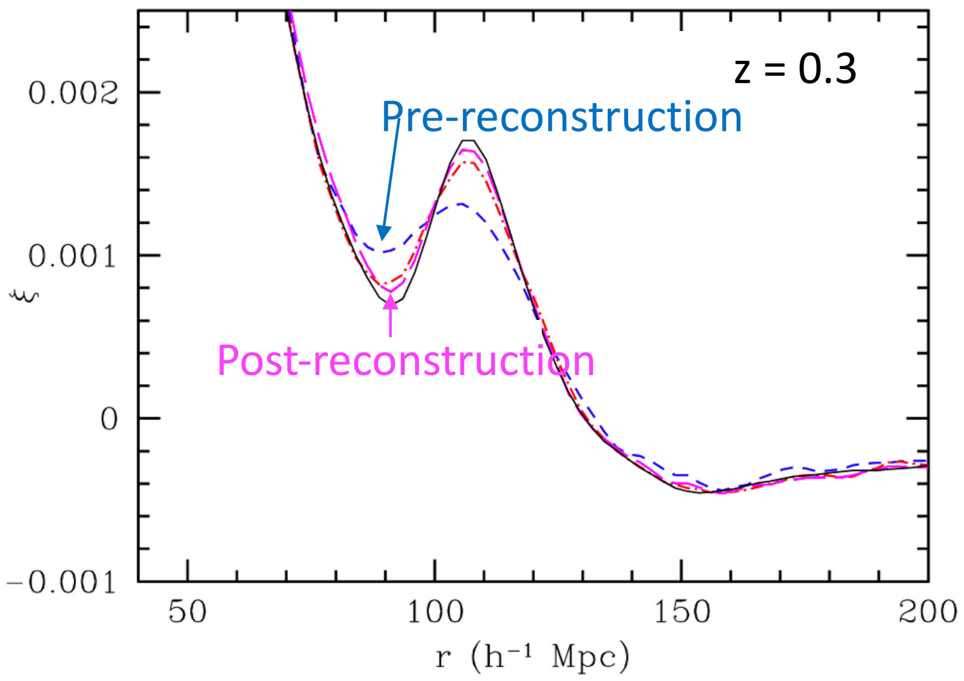
Massara+20

Beyond the Density Field

- What's the best field to compute the two-point function of?
  - For a **Gaussian** universe, the power-spectrum of galaxy overdensity contains **all** the information

- The Universe is **not** Gaussian:
  - Information **cascades** to the higher-point functions
  - **Low-density regions** carry a lot of cosmological information, and contribute little to δ [e.g. Pisani+19]

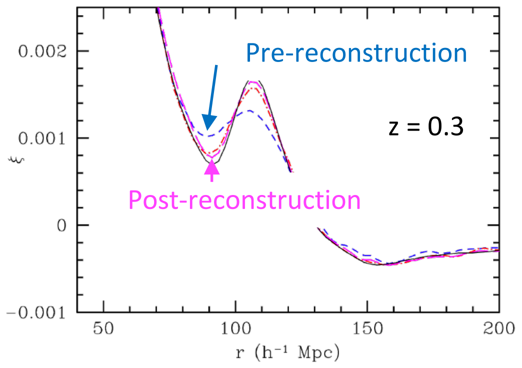
- Various **transformed** fields have been proposed:
  - Reconstructed Density Fields [e.g. Eisenstein+07]
  - Log-normal Transforms [Neyrinck+09, Wang+11]
  - Gaussianized Density Fields [Weinberg 92, Neyrinck+17]
  - Marked Density Fields [Stoyan 84, White 16, Massara+20]



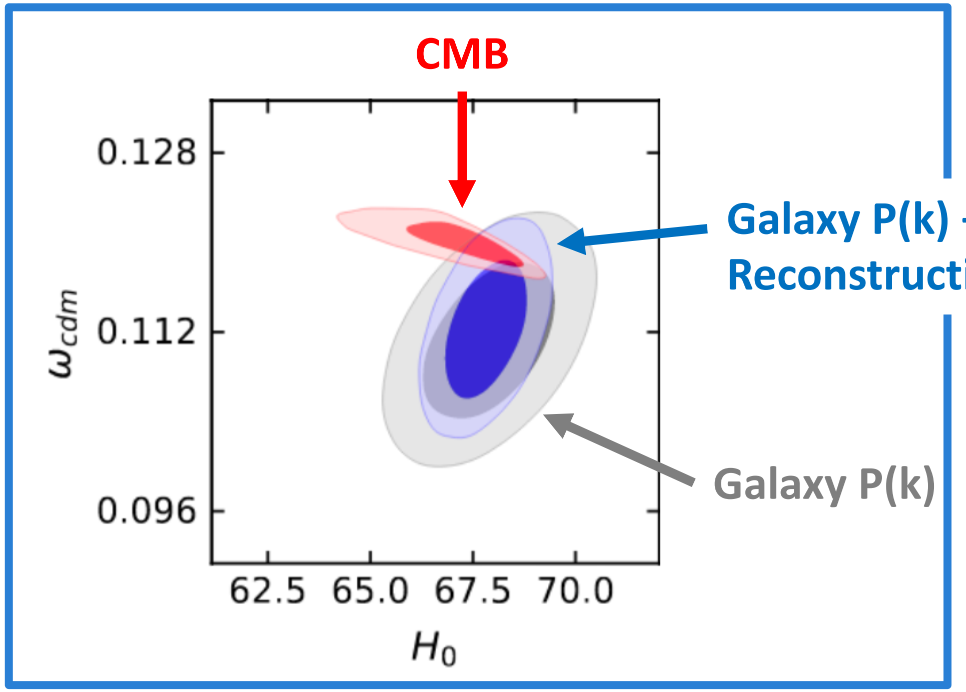
Eisenstein+06, Padamanbhan+12

Beyond the Density Field

- What's the best field to compute the two-point function of?
  - For a **Gaussian** universe, the power-spectrum of galaxy overdensity contains **all** the information



- The Universe is **not** Gaussian:
  - Information **cascades** to the higher-point functions
  - **Low-density regions** carry a lot of cosmological information, and contribute little to δ [e.g. Pisani+19]



- Various **transformed** fields have been proposed:
  - Reconstructed Density Fields [e.g. Eisenstein+07]
  - Log-normal Transforms [Neyrinck+09, Wang+11]
  - Gaussianized Density Fields [Weinberg 92, Neyrinck+17]
  - Marked Density Fields [Stoyan 84, White 16, Massara+20]

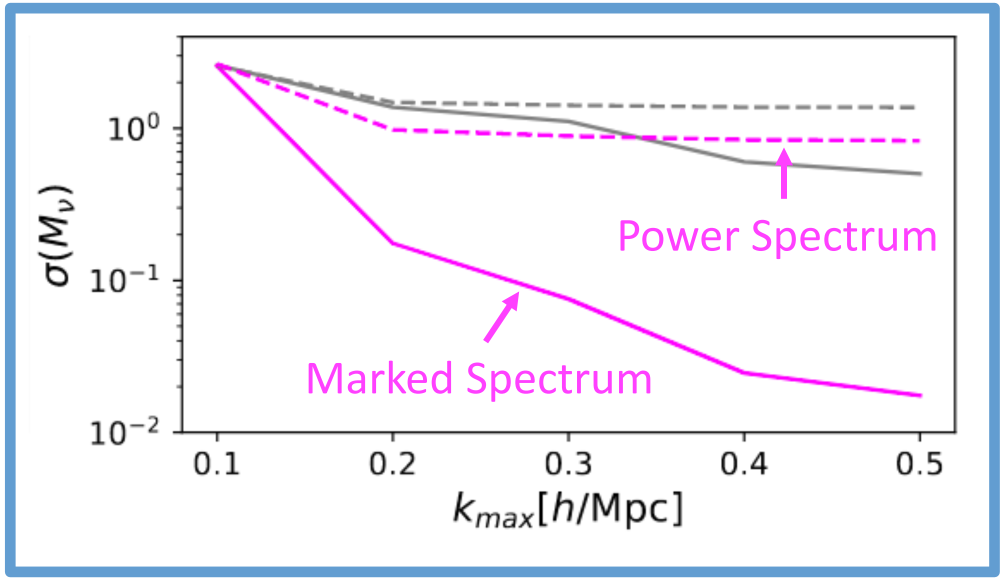
Philcox+20a

Beyond the Density Field

- What's the best field to compute the two-point function of?
  - For a **Gaussian** universe, the power-spectrum of galaxy overdensity contains **all** the information

- The Universe is **not** Gaussian:
  - Information **cascades** to the higher-point functions
  - **Low-density regions** carry a lot of cosmological information, and contribute little to δ [e.g. Pisani+19]

- Various **transformed** fields have been proposed:
  - Reconstructed Density Fields [e.g. Eisenstein+07]
  - Log-normal Transforms [Neyrinck+09, Wang+11]
  - Gaussianized Density Fields [Weinberg 92, Neyrinck+17]
  - Marked Density Fields [Stoyan 84, White 16, Massara+20]



Fisher Matrix Constraints on Neutrino Mass

Massara+20

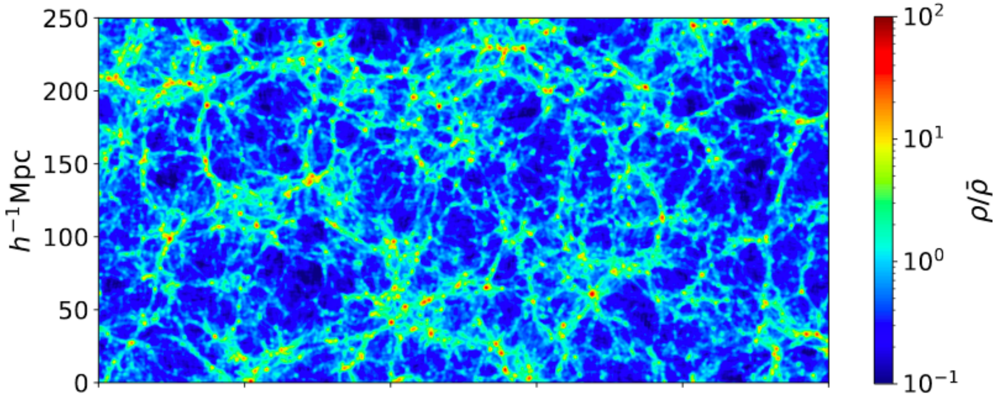
The Marked Density Field

- Define a new density field by weighting by the **mark**

$$m(\mathbf{x}) = \left(\frac{1+\delta_s}{1+\delta_s+\delta_R(\mathbf{x})}\right)^p$$

$$\rho_M(\mathbf{x}) = m(\mathbf{x})n(\mathbf{x}) = m(\mathbf{x})\bar{n}\left[1+\delta(\mathbf{x})\right]$$

depending on **smoothed** overdensity $\delta_R(\mathbf{x})$



Unweighted Density Field

Massara+20

The Marked Density Field

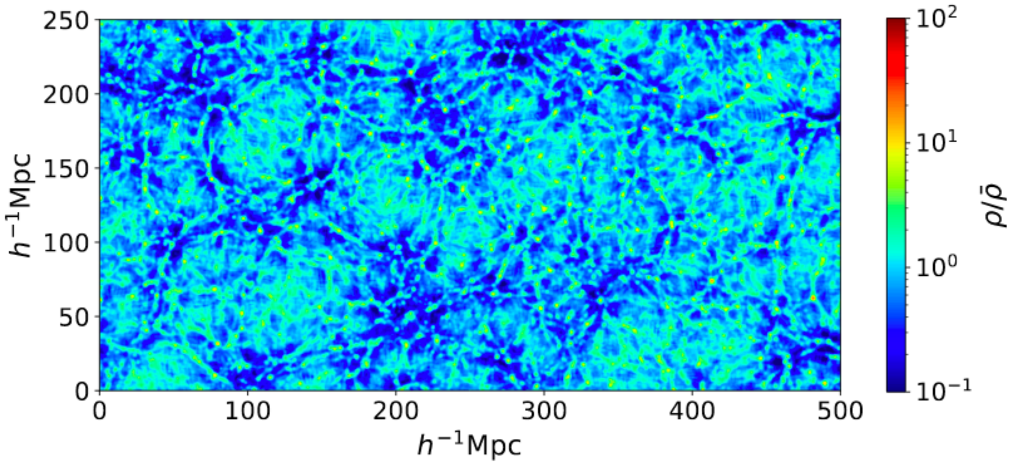
- Define a new density field by weighting by the **mark**

$$m(\mathbf{x}) = \left(\frac{1+\delta_s}{1+\delta_s+\delta_R(\mathbf{x})}\right)^p$$

$$\rho_M(\mathbf{x}) = m(\mathbf{x})n(\mathbf{x}) = m(\mathbf{x})\bar{n}\left[1+\delta(\mathbf{x})\right]$$

depending on **smoothed** overdensity $\delta_R(\mathbf{x})$

- Controlled by **mark parameters:**
  - Exponent p ($p > 0$ to upweight low-density regions)
  - Cut-off δ_s
  - Smoothing scale, R
- For **real-space matter**, significantly enhances constraints on parameters including:
  - **Neutrino masses** [Massara+20]
  - **Modified gravity** [White 16]

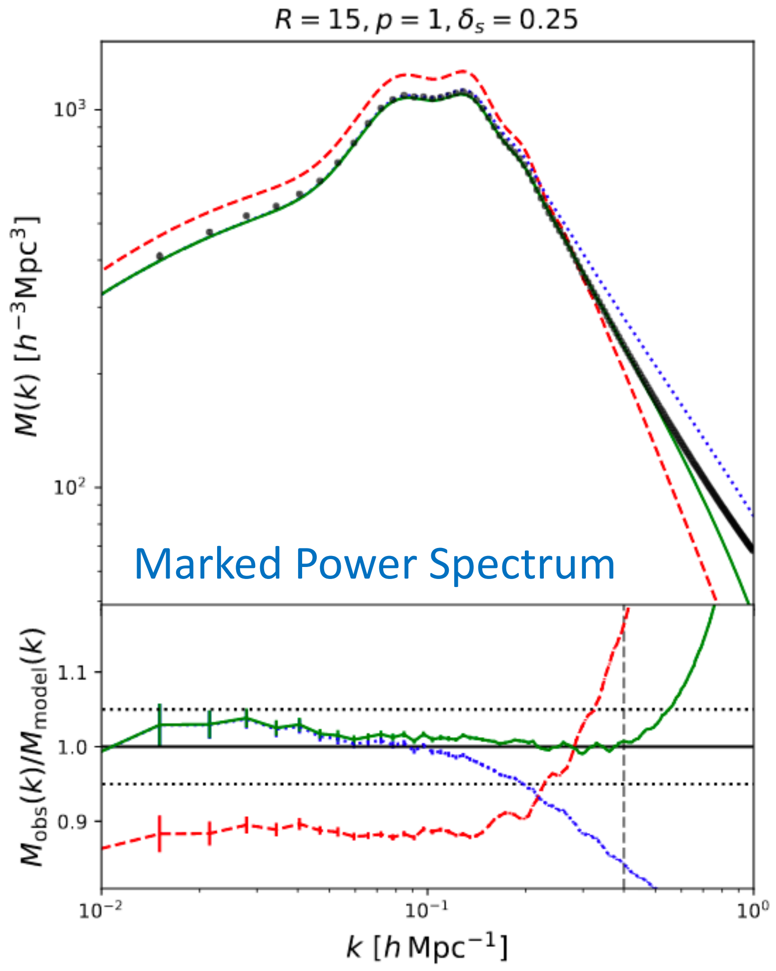


Marked Density Field

Massara+20

The Marked Density Field

- Can we **model** the marked spectrum?
  - Yes! Using **E**ffective **F**ield **T**heory
- Can we **understand** the impressive information content?
  - The mark couples **small-scale** non-Gaussianities to **large-scale** modes
  - So we find more neutrino information at low-k!
- But:
  - Modelling is **difficult** at low-z
  - Is it still useful for galaxies?

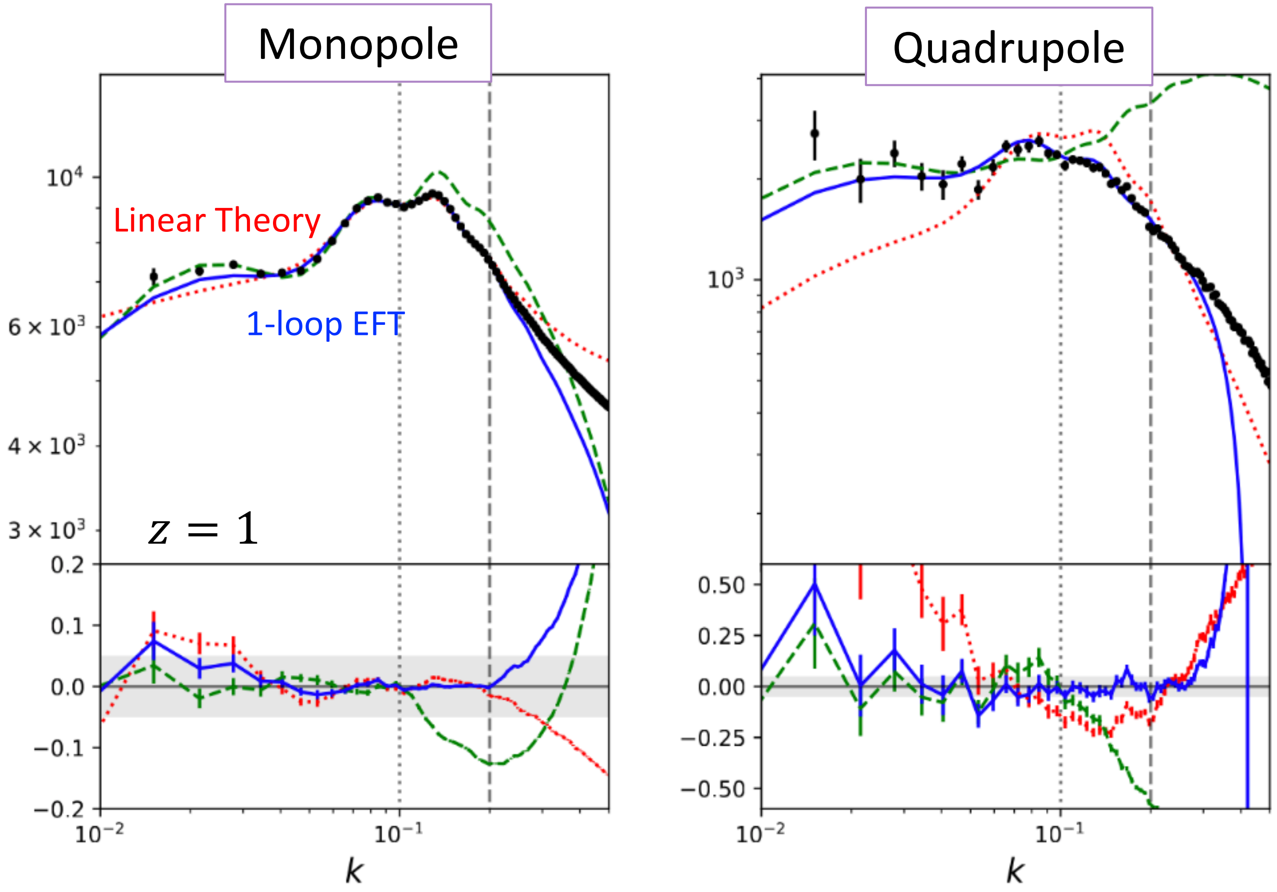


Matter at $z = 1$

Massara+20, Philcox+20ce

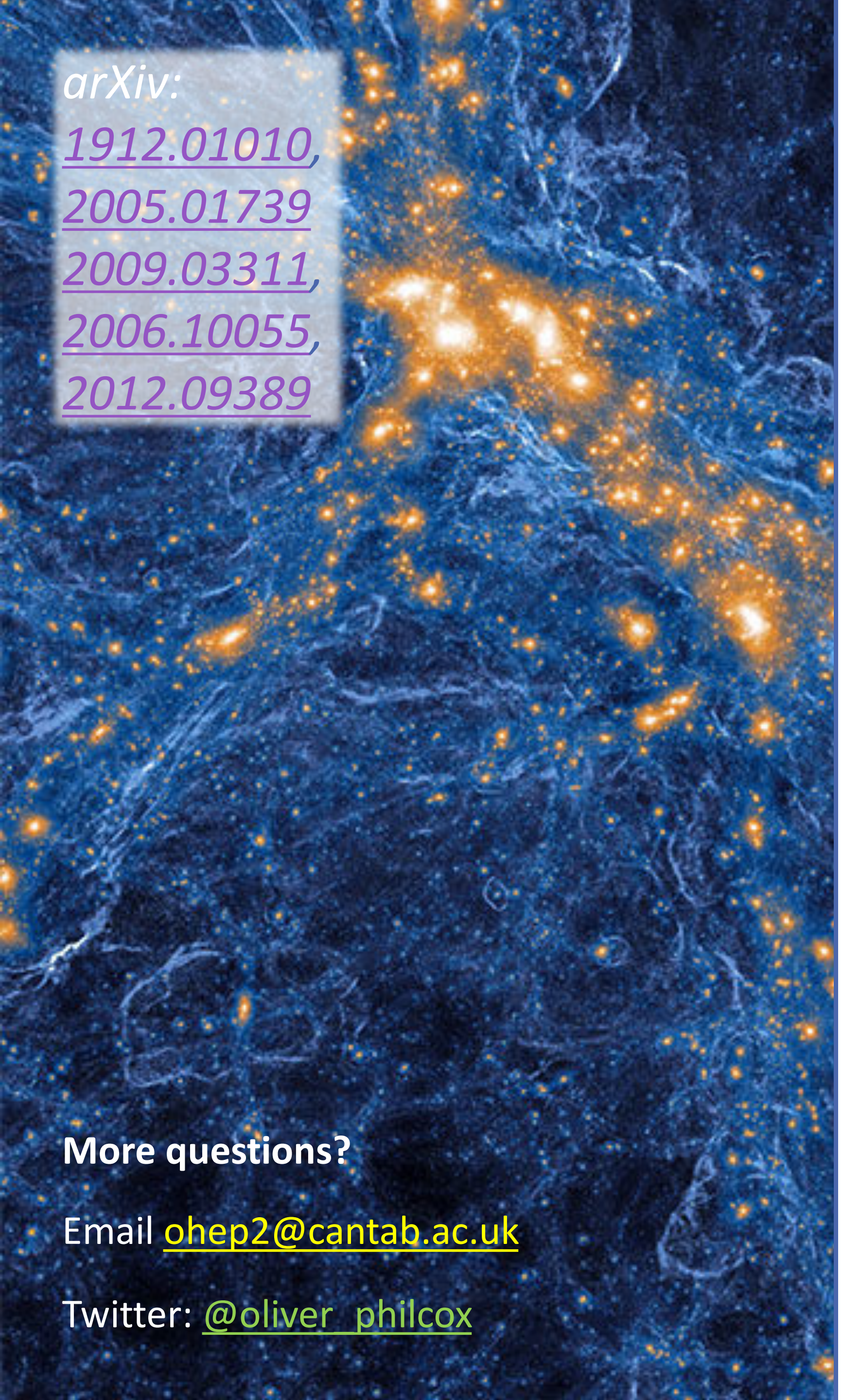
The Marked Density Field

- Can we **model** the marked spectrum?
  - Yes! Using **E**ffective **F**ield **T**heory
- Can we **understand** the impressive information content?
  - The mark couples **small-scale** non-Gaussianities to **large-scale** modes
  - So we find more neutrino information at low-k!
- But:
  - Modelling is **difficult** at low-z
  - Is it still useful for galaxies?



Galaxies at $z = 1$

Massara+20, Philcox+20ce



arXiv:
1912.01010, *2005.01739* *2009.03311*, *2006.10055*, *2012.09389*

More questions?

Email ohep2@cantab.ac.uk

Twitter: @oliver_philcox

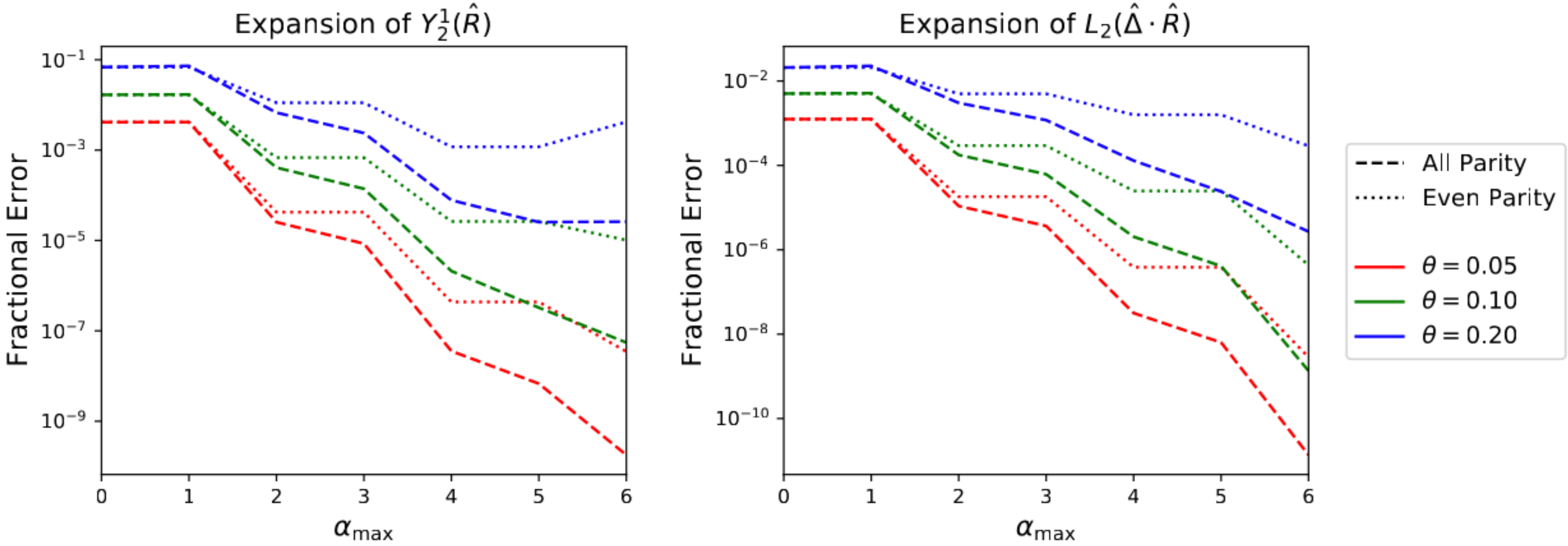
Conclusions

- We're not finished with the galaxy power spectrum yet!
- Recent updates include:
  - More accurate lines-of-sight
  - Closer to optimal large-scale $P_\ell(k)$ estimation
  - Faster small-scale computation *without* FFTs
  - Powerful analysis-specific data compression

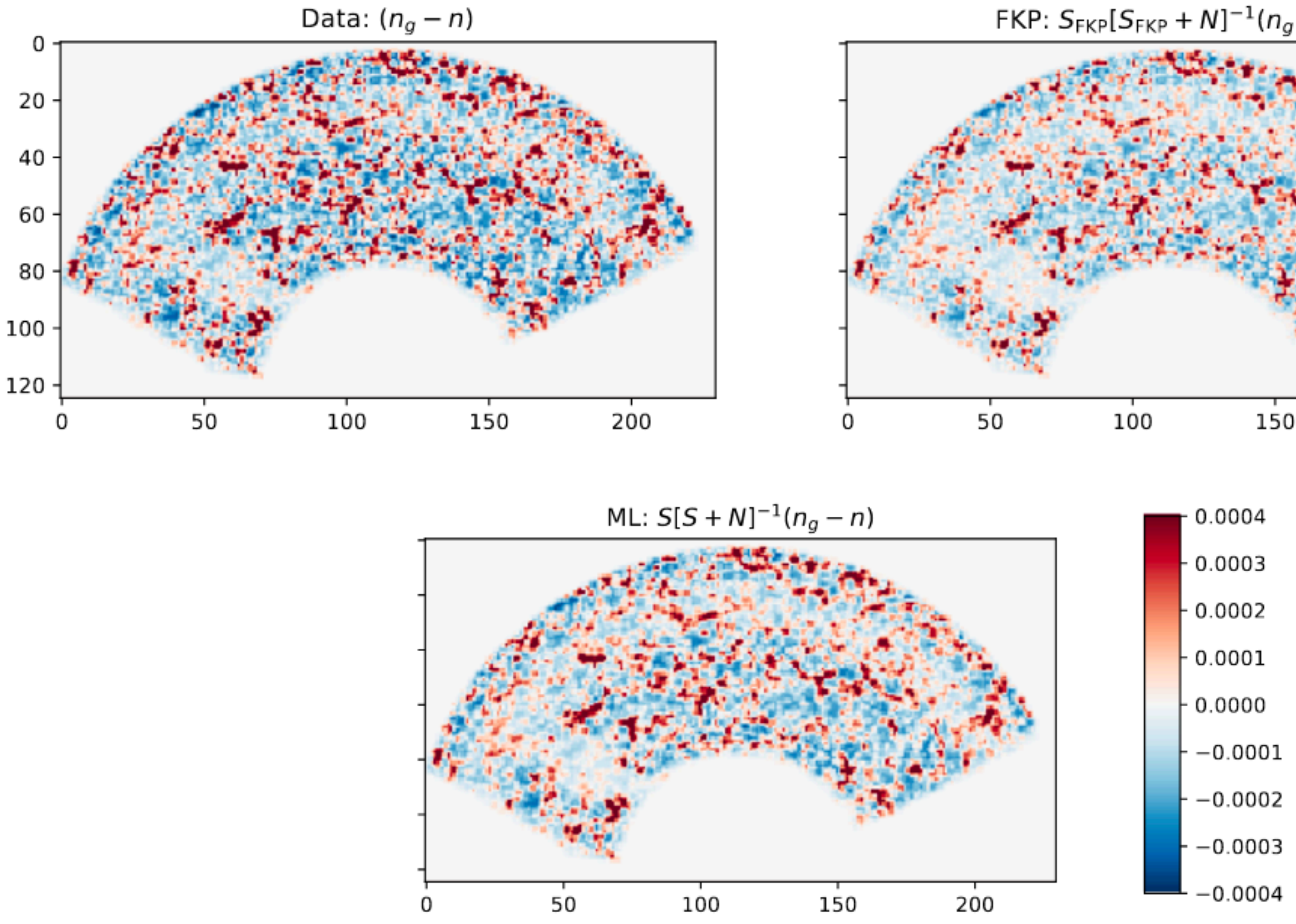
Coming soon:

- Estimating the bispectrum and beyond!

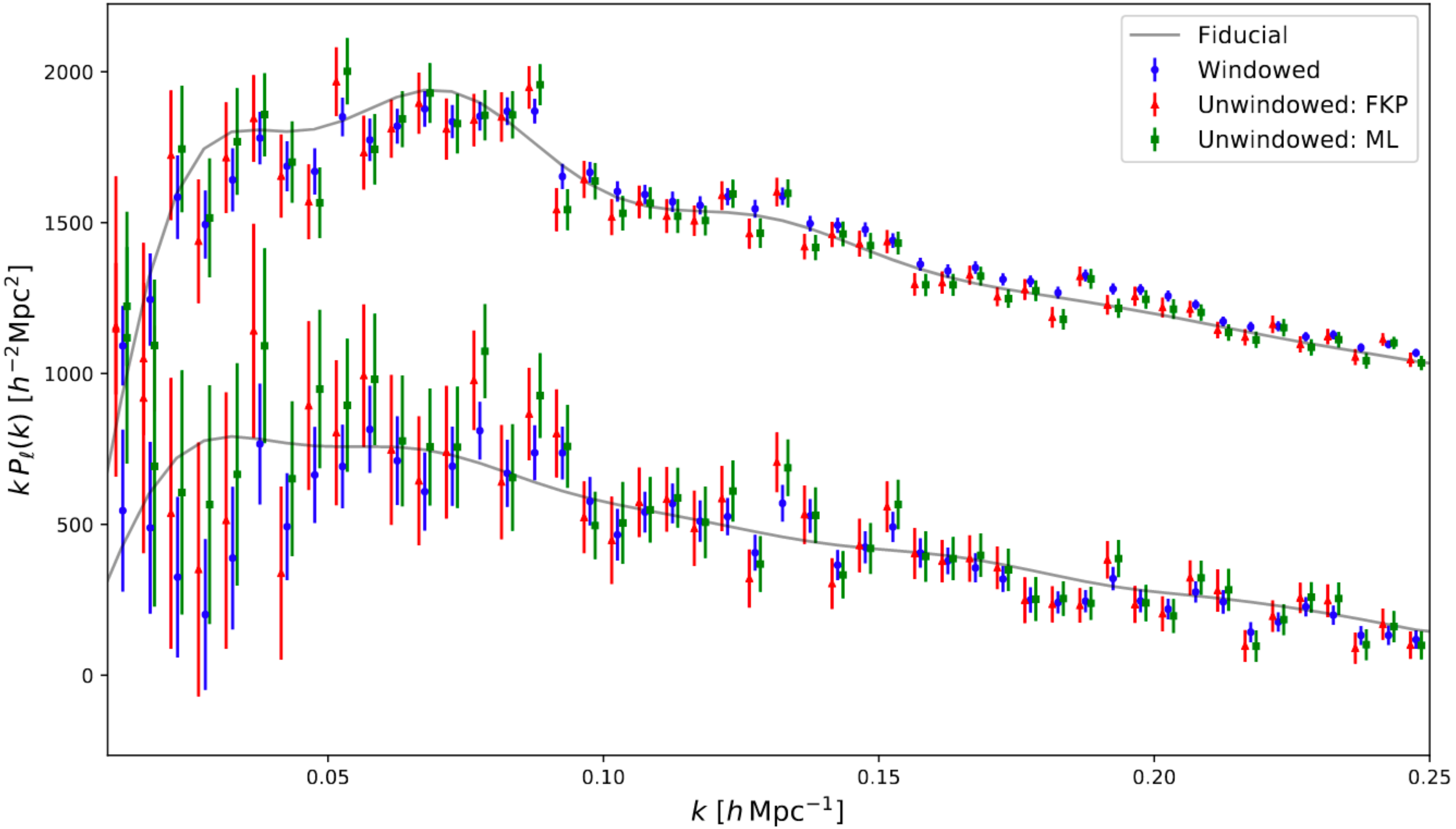
Shift Theorem Convergence



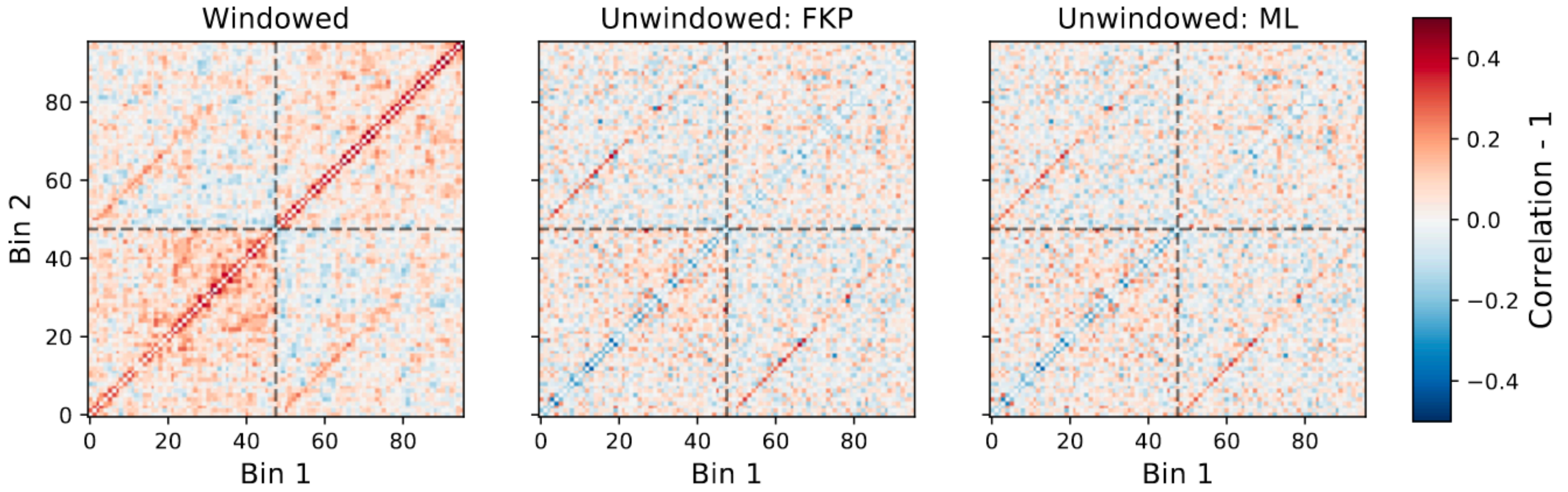
Optimal Estimators: Filtering



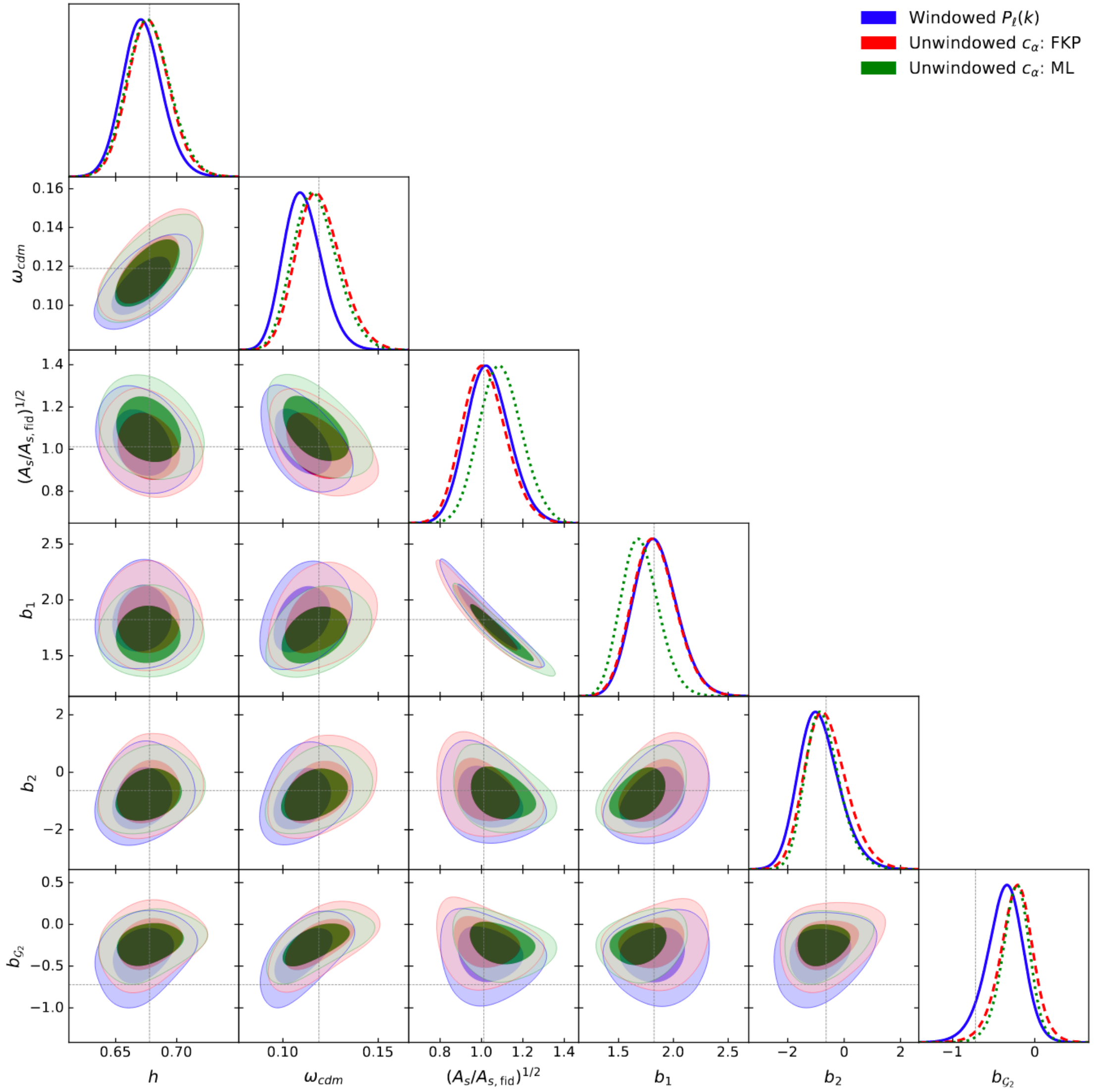
Optimal Estimators: Spectra



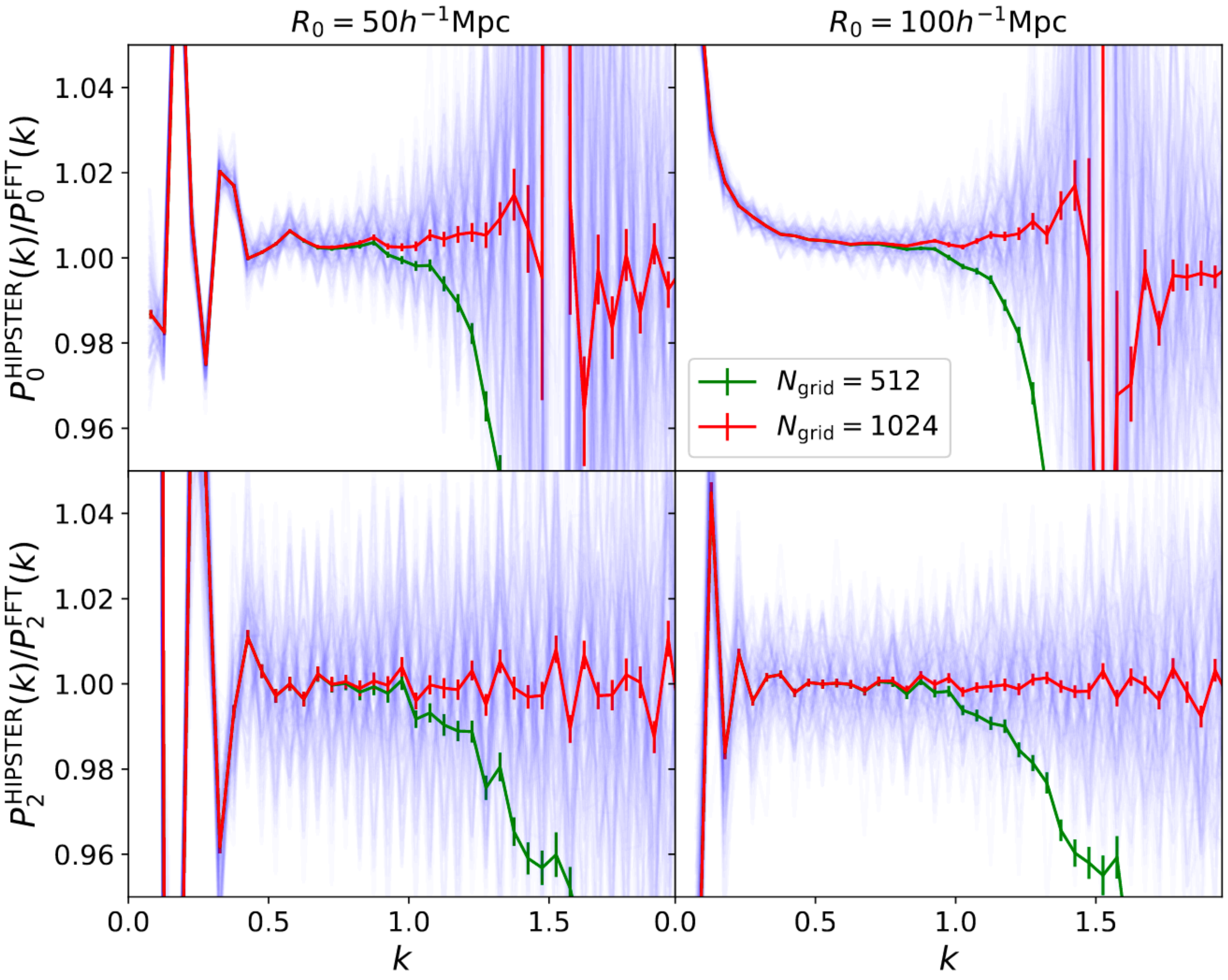
Optimal Estimators: Covariance



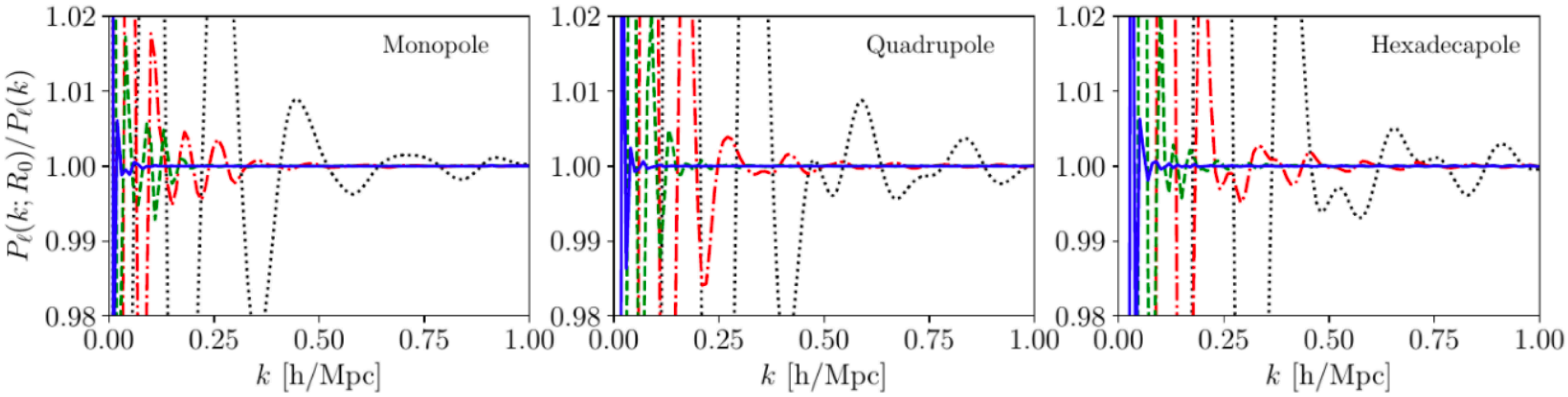
Optimal Estimators: Results



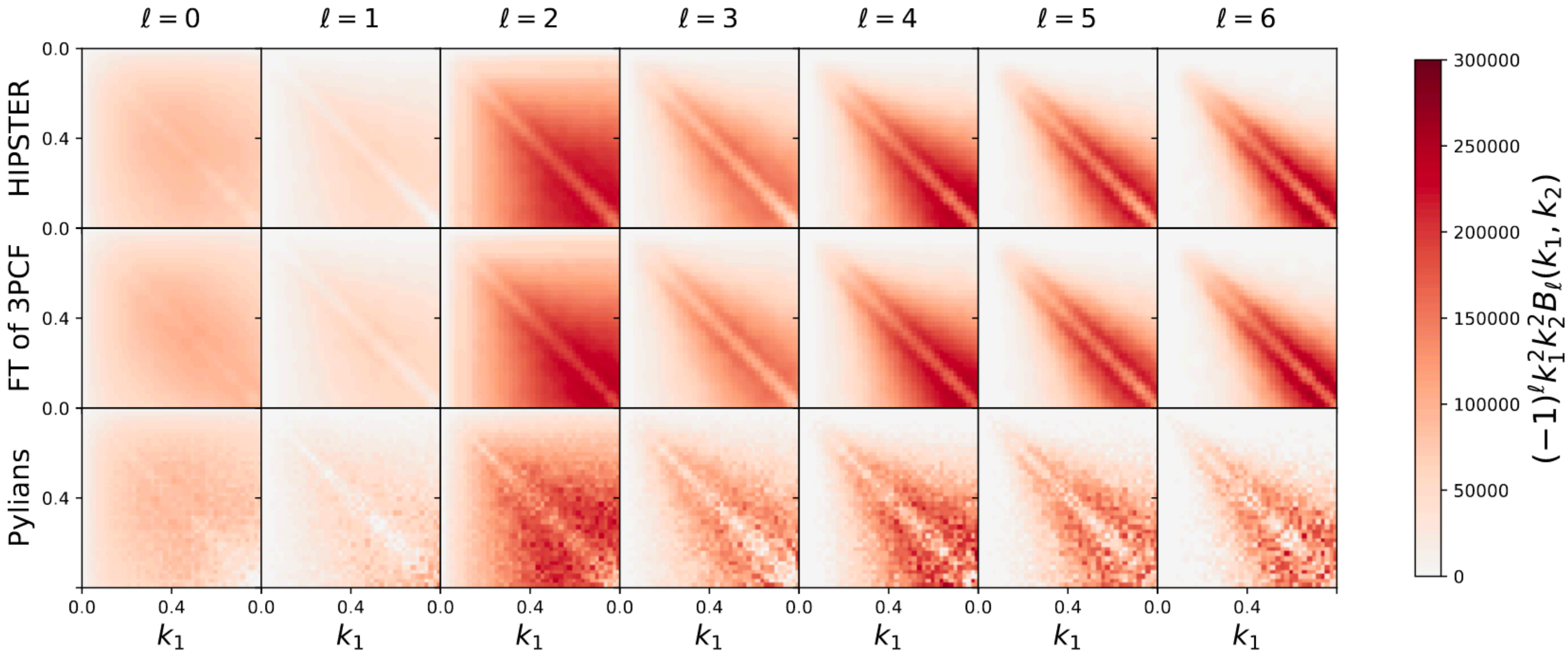
HIPSTER: Accuracy



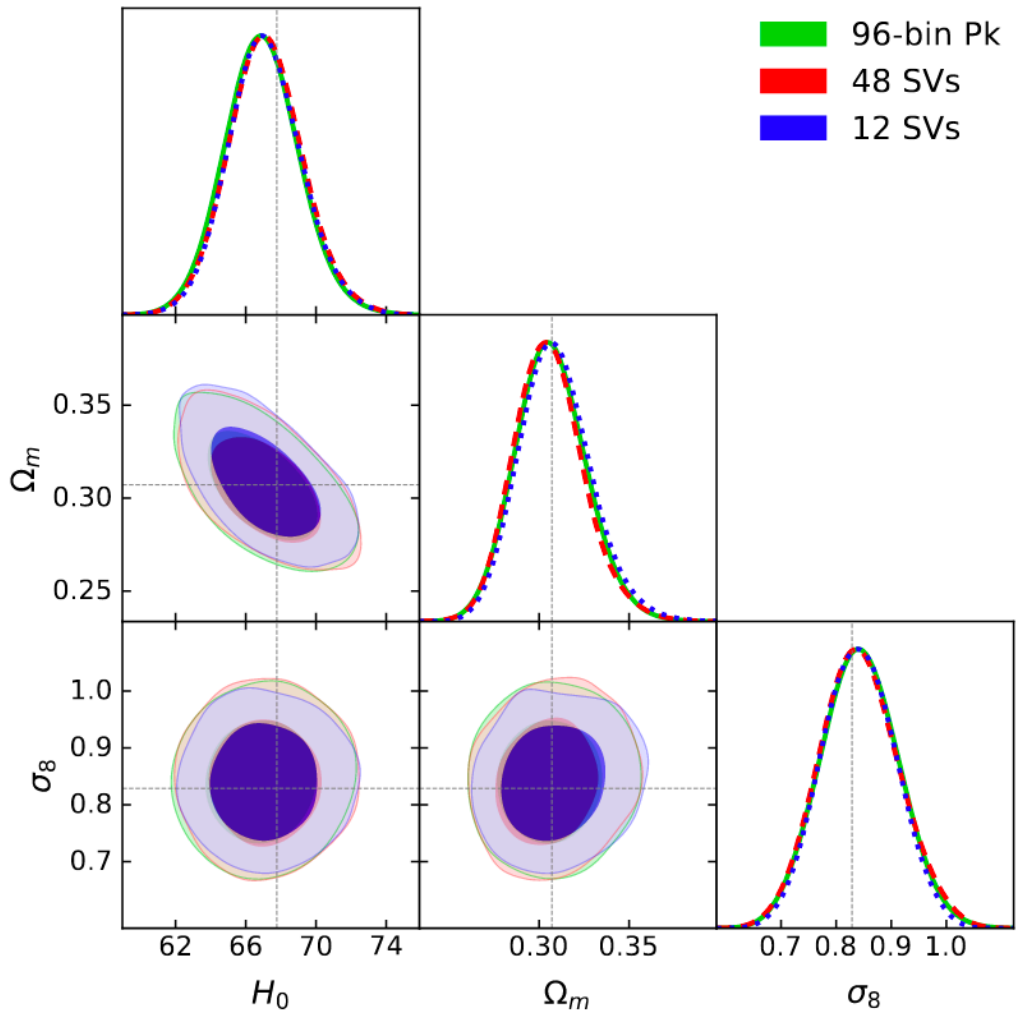
HIPSTER: Effects of Windowing

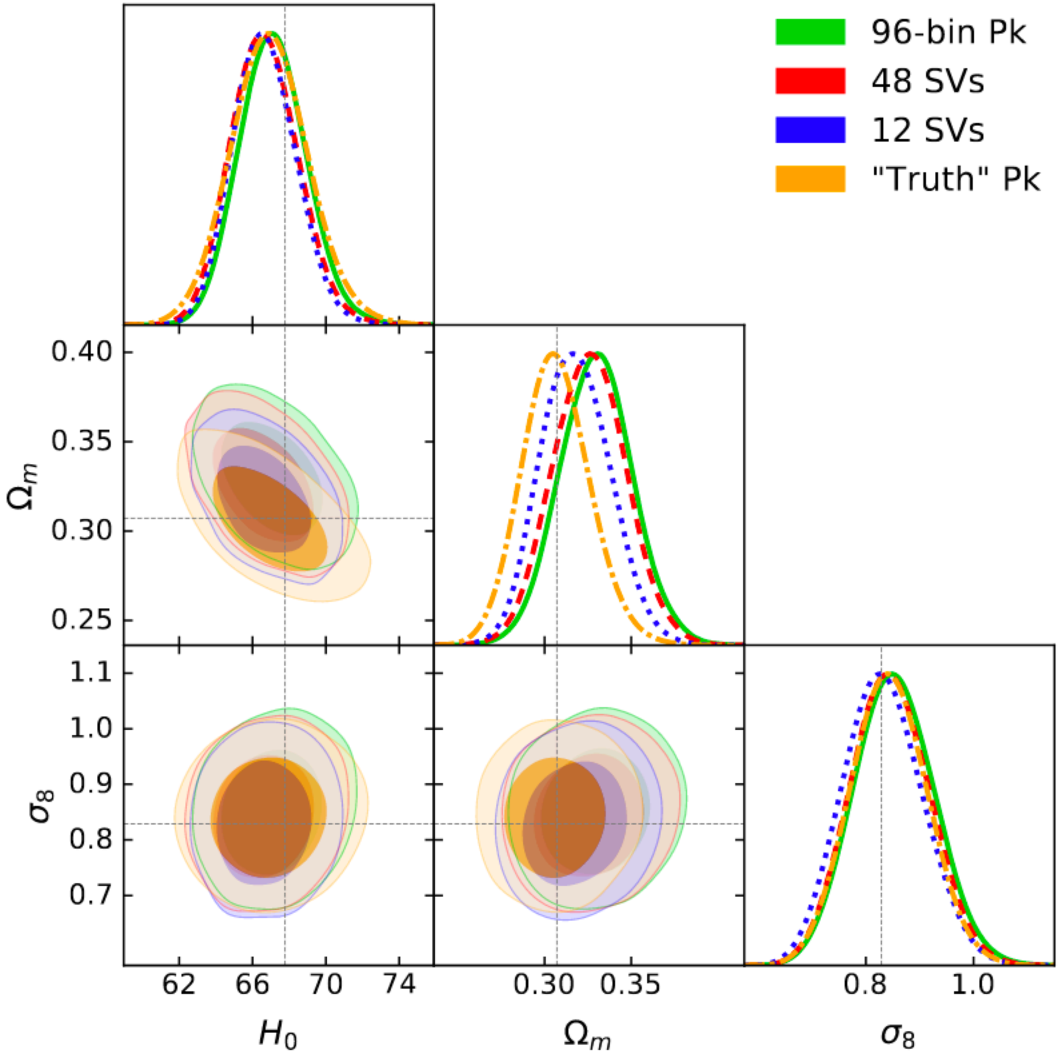


HIPSTER: Bispectra

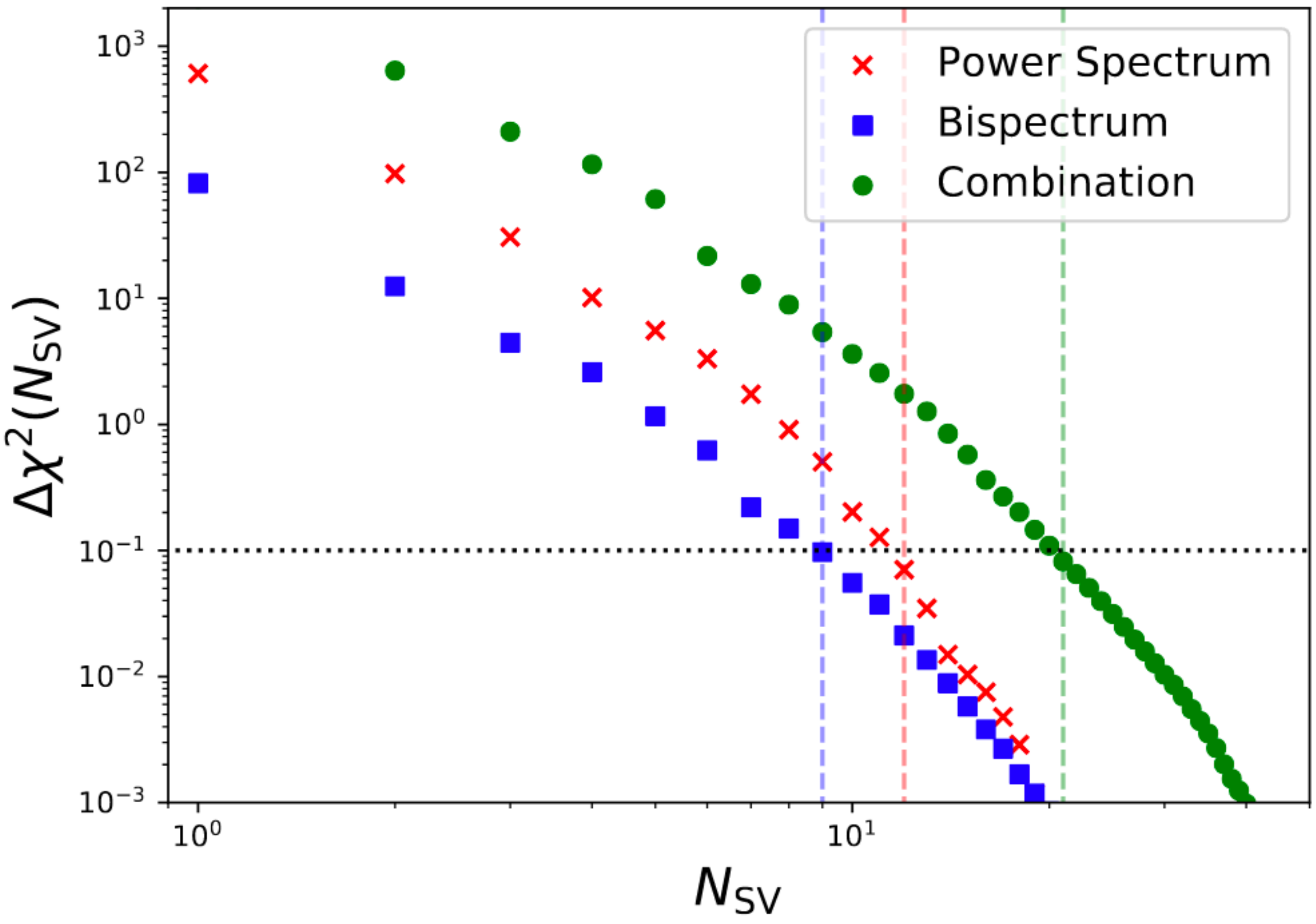


Compression: Mean of Mocks & Single Mock

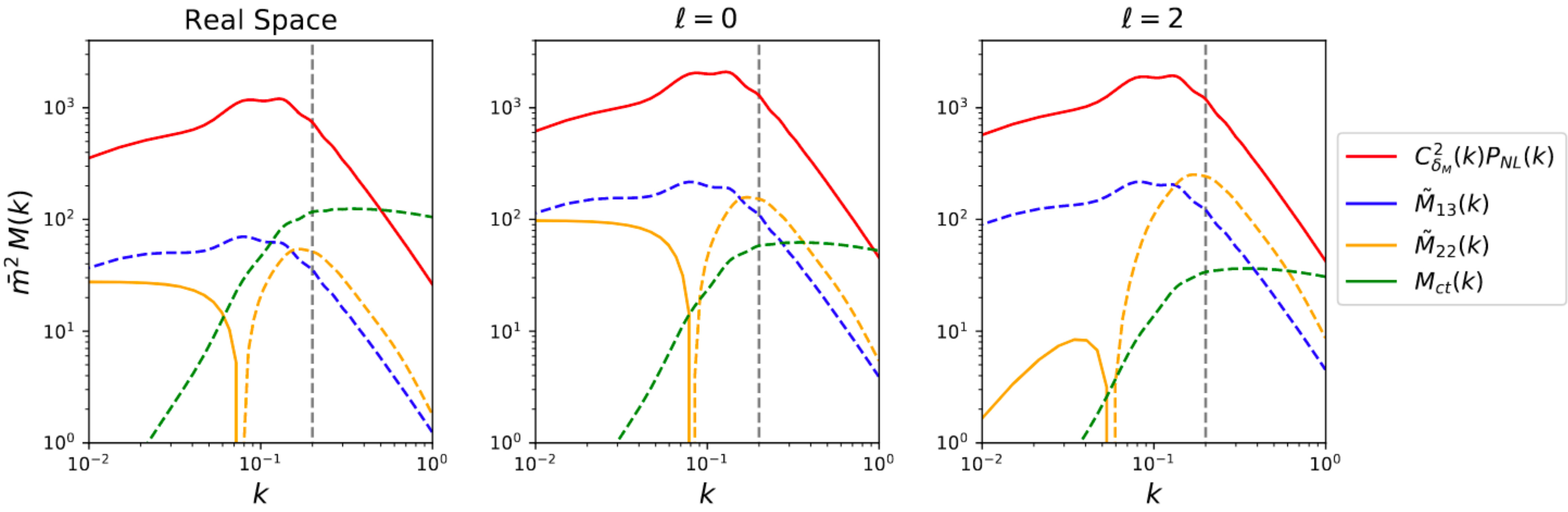




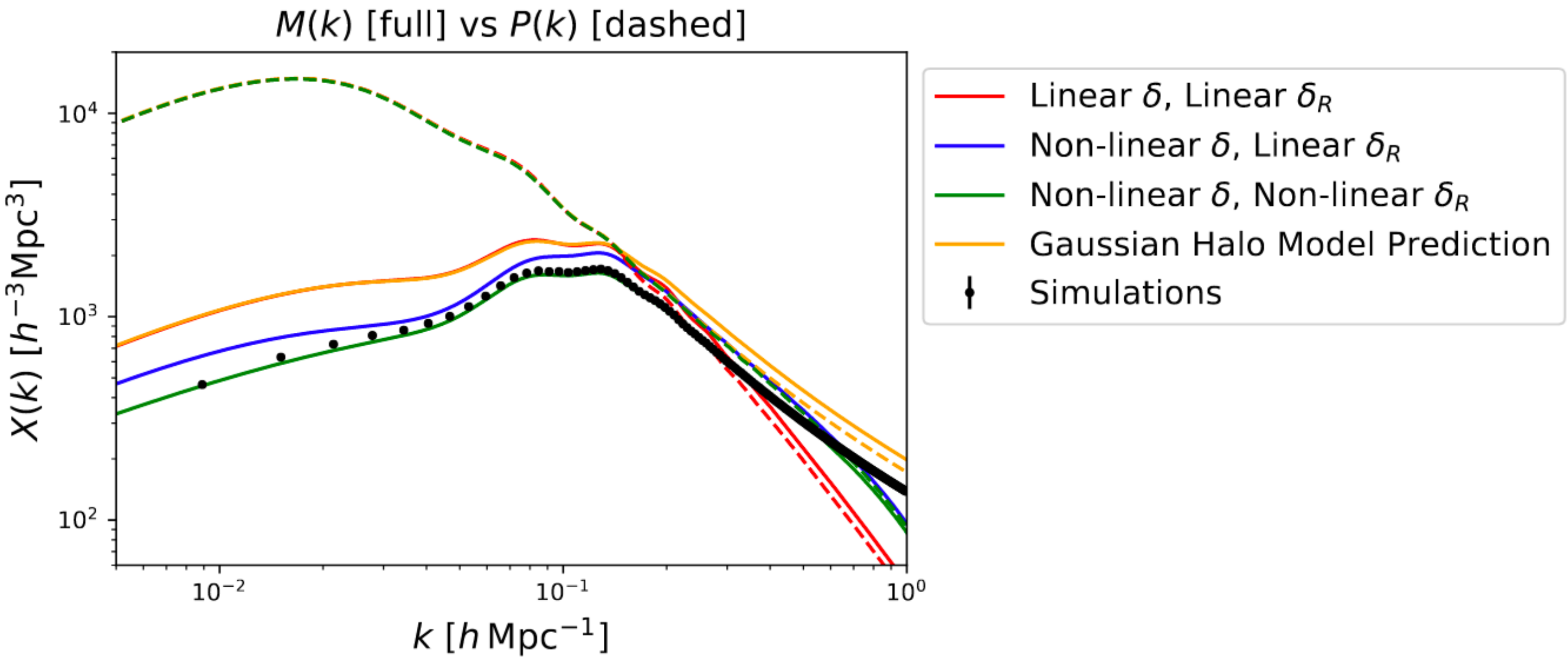
Compression: Number of Basis Vectors

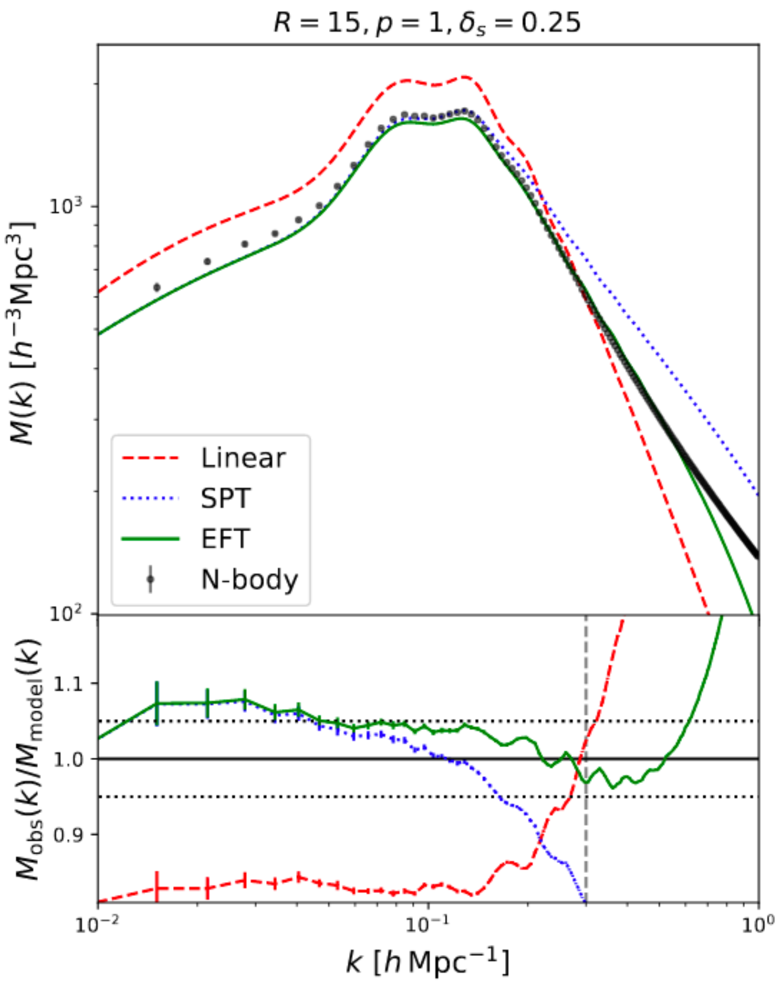


Marked Spectra: Matter Contributions



Marked Spectra: Information Content & Low-z





(a) $z = 0.5$