

The DESI survey: overview and missing-observation countermeasures

Davide Bianchi

ICC - University of Barcelona

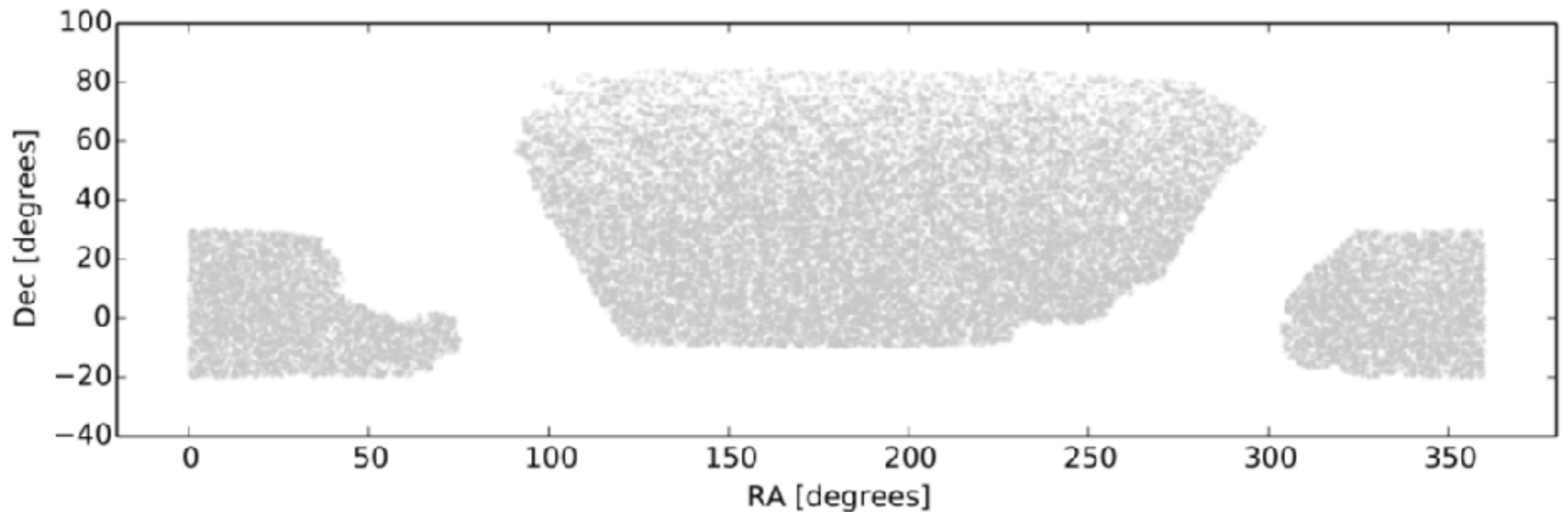
Barolo 10/09/2021

The Dark Energy Spectroscopic Instrument (DESI) galaxy survey



Mayall telescope (4m) in Kitt Peak, Arizona

DESI footprint



total area $\sim 14,000 \text{ deg}^2$

total number of targets $\sim 30 \times 10^6$

DESI targets

Galaxy type	Redshift range	Bands used	Targets per deg ²	Exposures per deg ²	Good z 's per deg ²	Baseline sample
LRG	0.4–1.0	$r, z, W1$	350	580	285	4.0 M
ELG	0.6–1.6	g, r, z	2400	1870	1220	17.1 M
QSO (tracers)	< 2.1	$g, r, z, W1, W2$	170	170	120	1.7 M
QSO (Ly- α)	> 2.1	$g, r, z, W1, W2$	90	250	50	0.7 M
Total in dark time			3010	2870	1675	23.6 M
BGS	0.05–0.4	r	700	700	700	9.8 M
Total in bright time			700	700	700	9.8 M

DESI collaboration arXiv:1611.00036

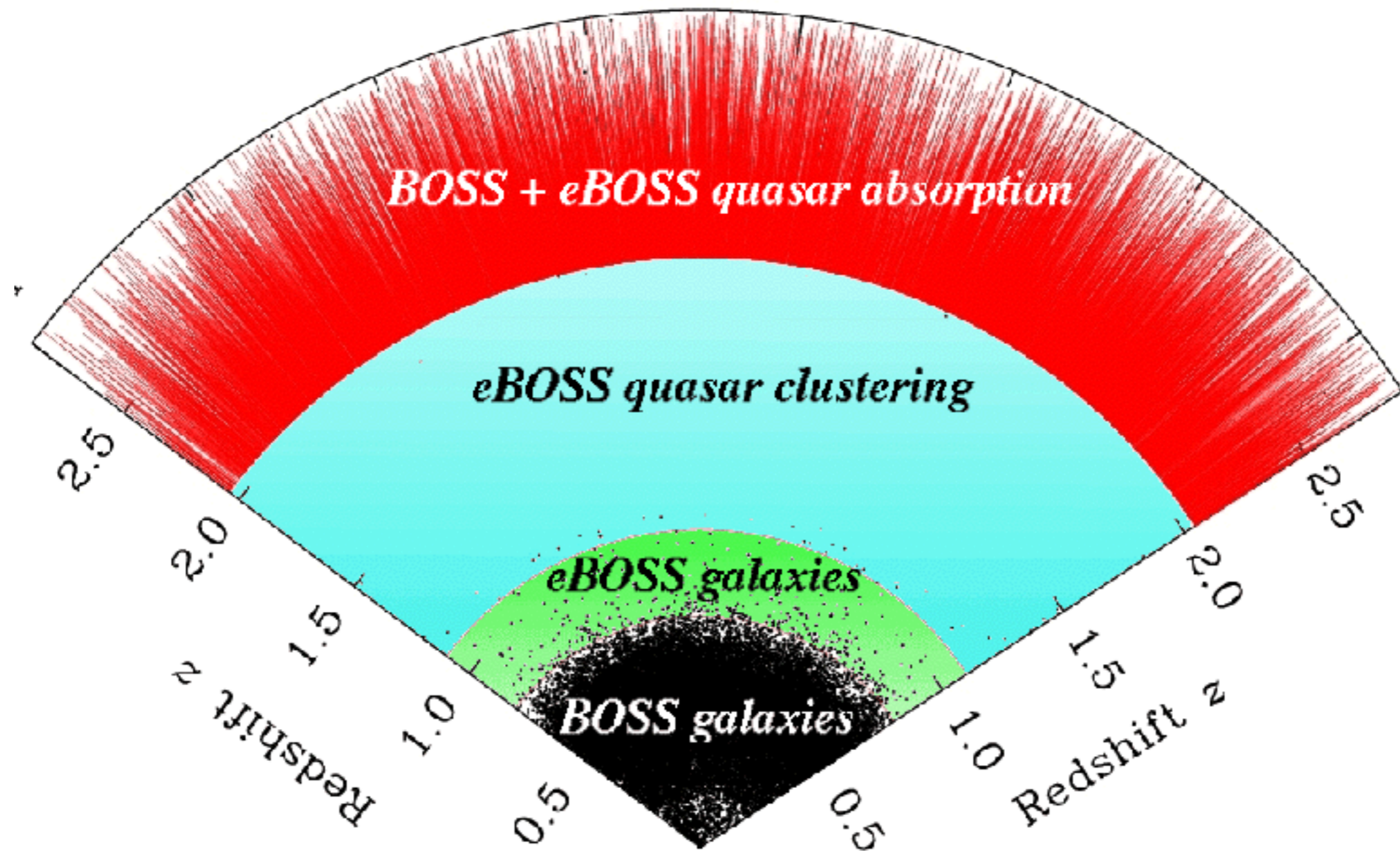
ELG = emission line galaxies

LRG = luminous red galaxies

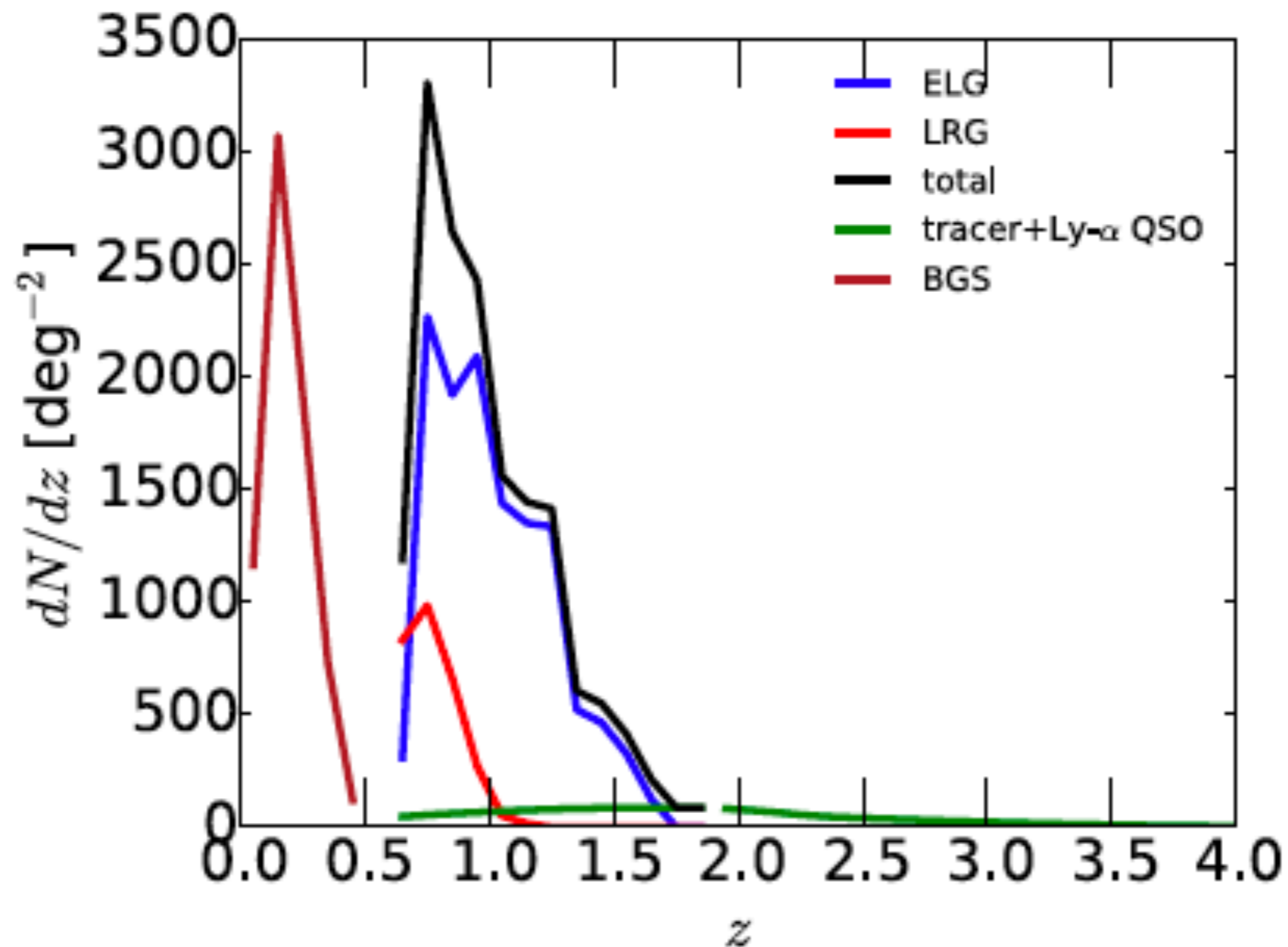
QSO = quasars

BGS = bright galaxy survey

Visual impression of the z -coverage with different tracers (borrowed from eBOSS)



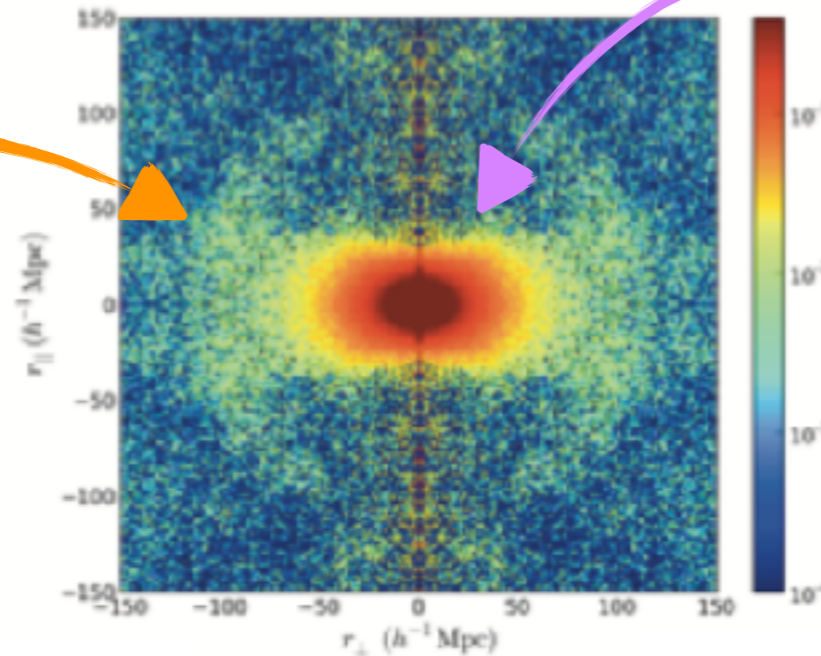
DESI redshift distributions for the different tracers (illustrative)



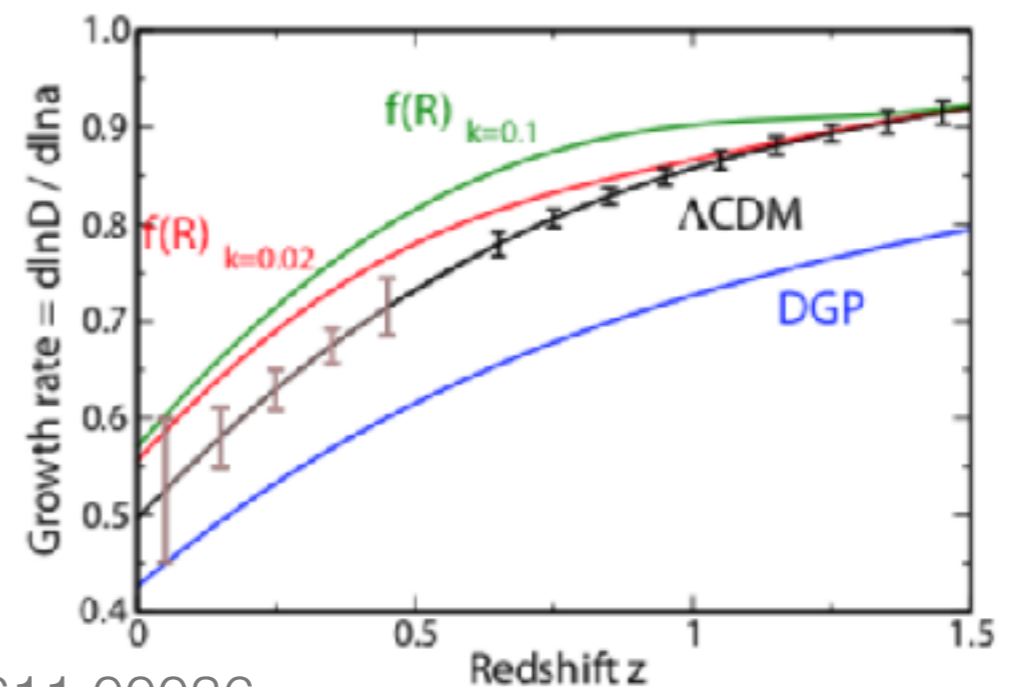
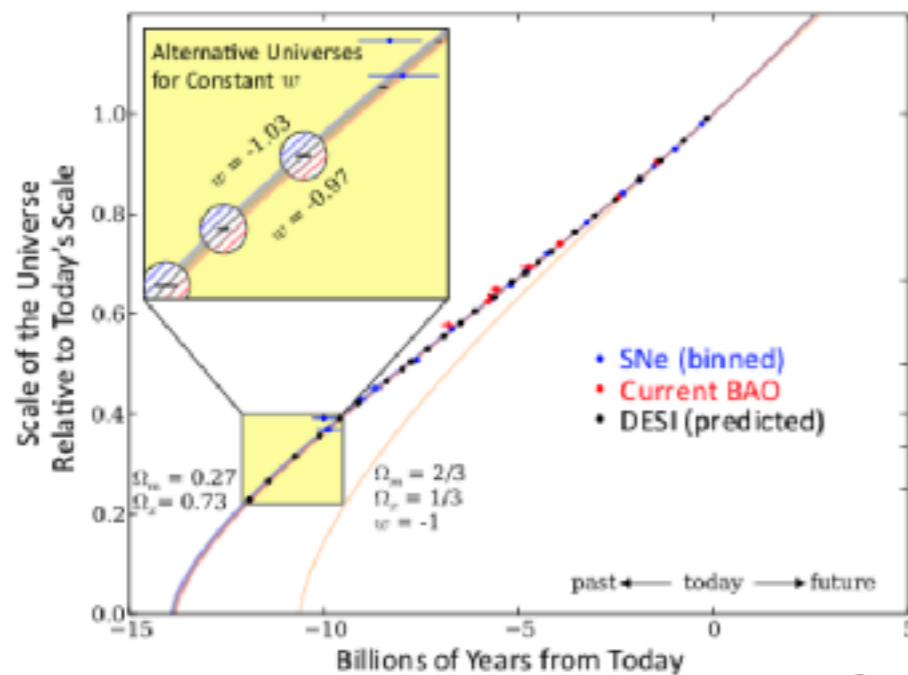
Why do we want to measure the 3D position of 30M galaxies?

2-point correlation function

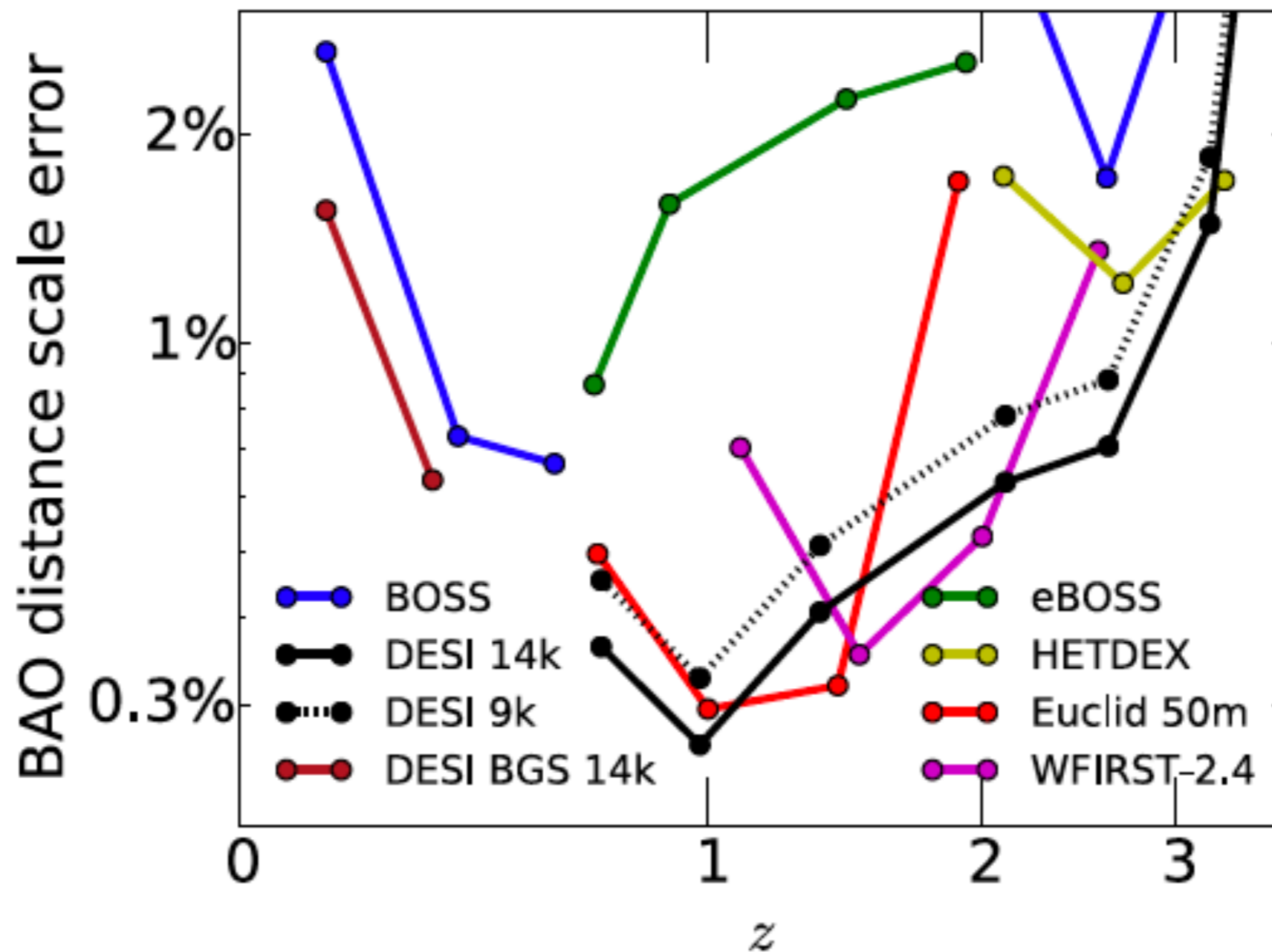
baryon acoustic oscillation (BAO) peak



redshift space distortion (RSD) squashing

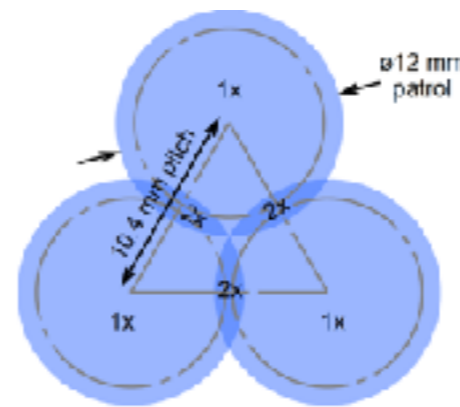
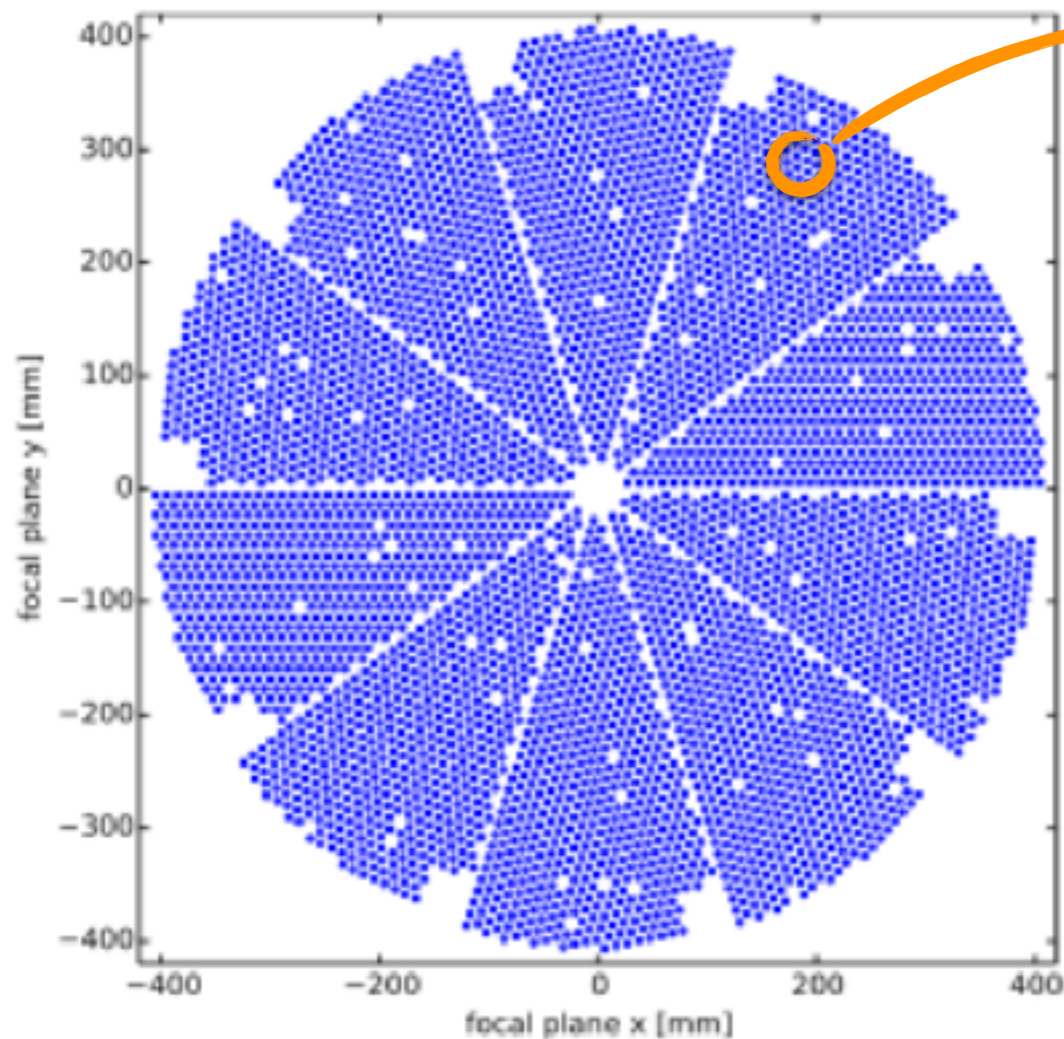


DESI precision compared to other surveys

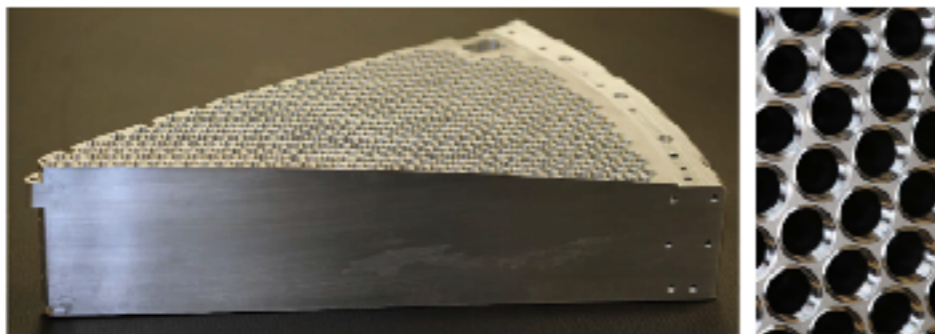
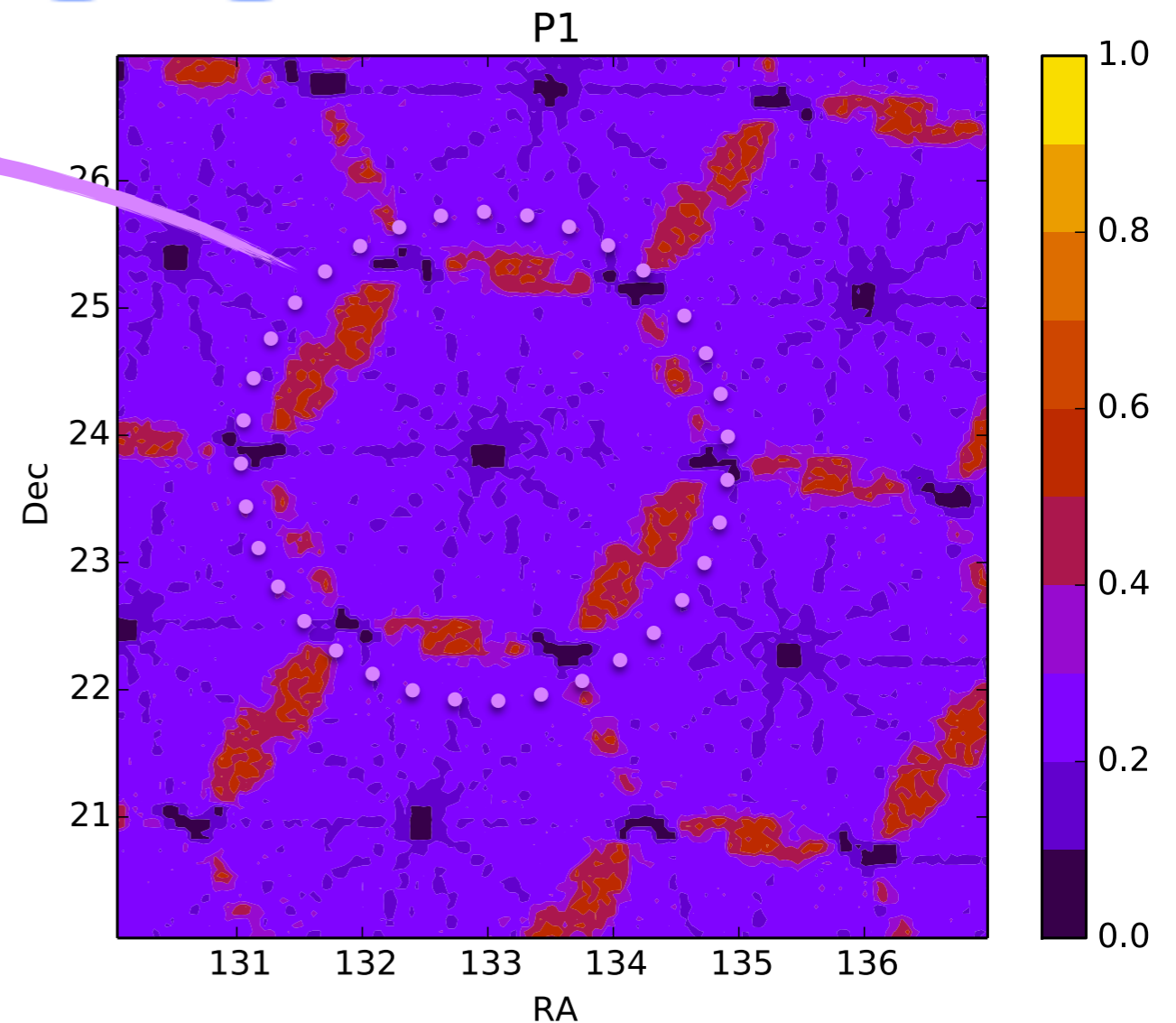


Fibre assignment and missing observations

5,000 fibers focal plane area = 7.5 deg²



completeness
after 1 pass of the
DESI instrument



2 point correlation function

galaxy-galaxy pairs

$$\xi(\vec{s}) = \frac{DD(\vec{s})}{RR(\vec{s})} - 2\frac{DR(\vec{s})}{RR(\vec{s})} + 1$$

galaxy-random pairs

Landy & Szalay 1993

random-random pairs

Pairwise-Inverse-Probability (PIP) weights: correlation-function estimator

$$\xi(\vec{s}) = \frac{DD(\vec{s})}{RR(\vec{s})} - 2 \frac{DR(\vec{s})}{RR(\vec{s})} + 1 \quad \text{Landy \& Szalay 1993}$$

$$DD(\vec{s}) = \sum_{\vec{x}_m - \vec{x}_n \approx \vec{s}} w_{mn}$$

$$w_{mn} = \frac{1}{p_{mn}} = \frac{1}{p_m p_n (1 + c_{mn})}$$

Pairwise-Inverse-Probability (PIP) weights: correlation-function estimator

$$\xi(\vec{s}) = \frac{DD(\vec{s})}{RR(\vec{s})} - 2 \frac{DR(\vec{s})}{RR(\vec{s})} + 1 \quad \text{Landy \& Szalay 1993}$$

$$DD(\vec{s}) = \sum_{\vec{x}_m - \vec{x}_n \approx \vec{s}} w_{mn}$$

$$w_{mn} = \frac{1}{p_{mn}} = \frac{1}{p_m p_n (1 + c_{mn})}$$

$$DR(\vec{s}) = \sum_{\vec{x}_m - \vec{y}_n \approx \vec{s}} w_m$$

$$w_m = \frac{1}{p_m}$$

Pairwise-Inverse-Probability (PIP) weights: correlation-function estimator

$$\xi(\vec{s}) = \frac{DD(\vec{s})}{RR(\vec{s})} - 2 \frac{DR(\vec{s})}{RR(\vec{s})} + 1 \quad \text{Landy \& Szalay 1993}$$

$$DD(\vec{s}) = \sum_{\vec{x}_m - \vec{x}_n \approx \vec{s}} w_{mn} \frac{DD_a^{(p)}(\theta)}{DD_a(\theta)}$$

$$DD_a(\theta) = \sum_{\vec{u}_m \cdot \vec{u}_n \approx \cos(\theta)} w_{mn}$$

$$w_{mn} = \frac{1}{p_{mn}} = \frac{1}{p_m p_n (1 + c_{mn})}$$

$$DR(\vec{s}) = \sum_{\vec{x}_m - \vec{y}_n \approx \vec{s}} w_m \frac{DR_a^{(p)}(\theta)}{DR_a(\theta)}$$

$$DR_a(\theta) = \sum_{\vec{u}_m \cdot \vec{v}_n \approx \cos(\theta)} w_m$$

$$w_m = \frac{1}{p_m}$$

Inverse-probability formalism: basic concept

We introduce the binary random variable b (selected/discarded)

$$\mathcal{P}(b) = p^b (1 - p)^{1-b}$$



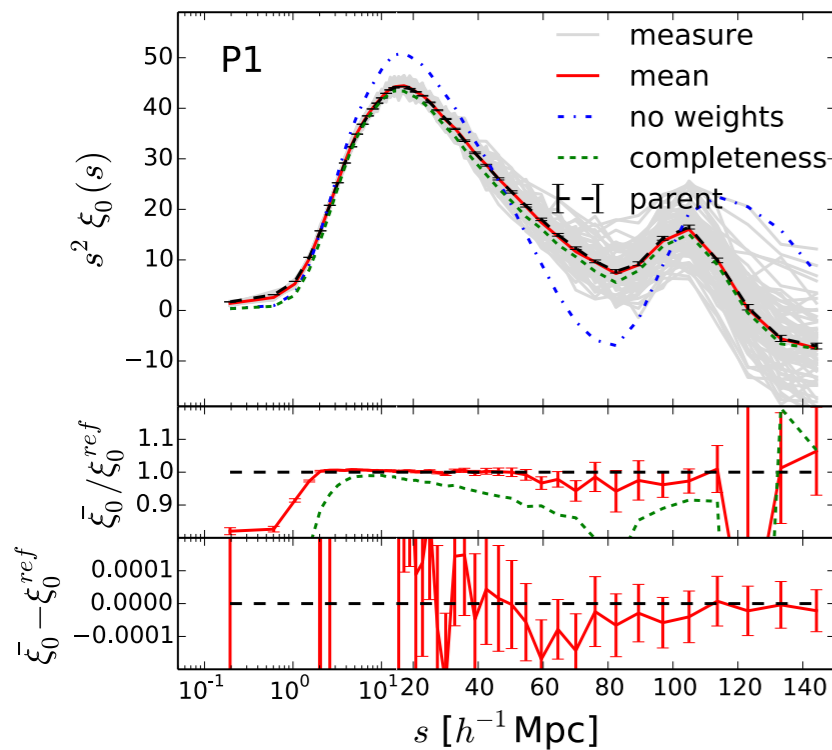
Selection probability

$$\langle b \rangle = p \quad \Rightarrow \quad \langle b/p \rangle = 1$$

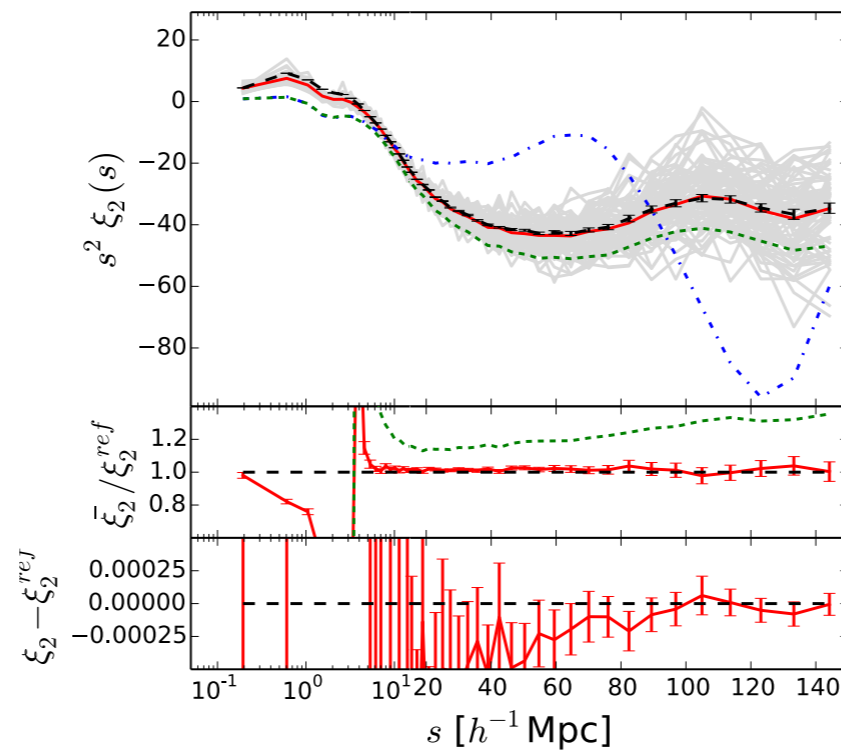
Comparison to DESI mocks: pass 1

completeness = 25%

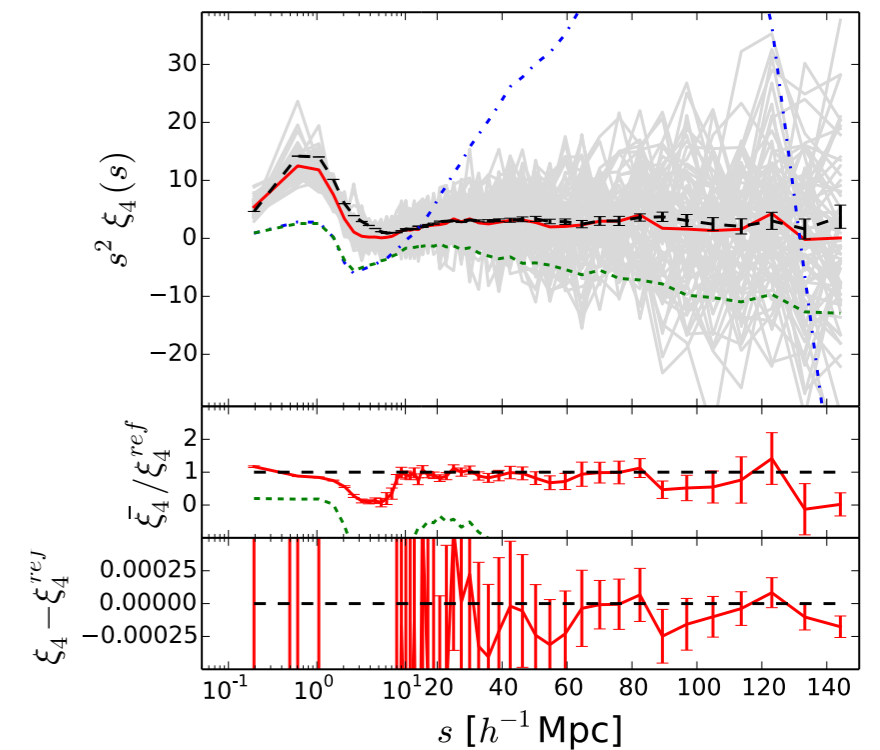
Monopole



Quadrupole



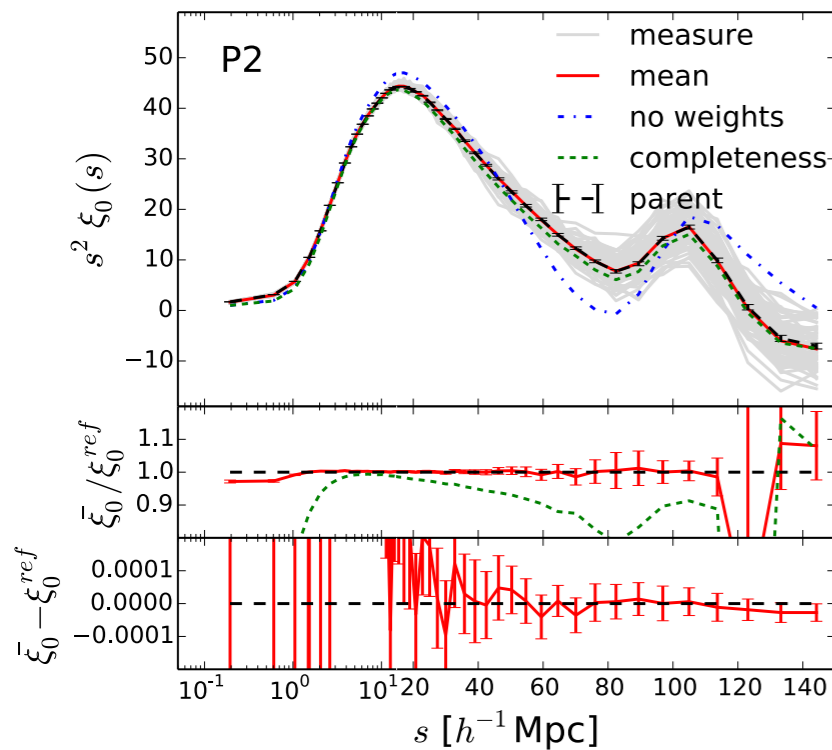
Hexadecapole



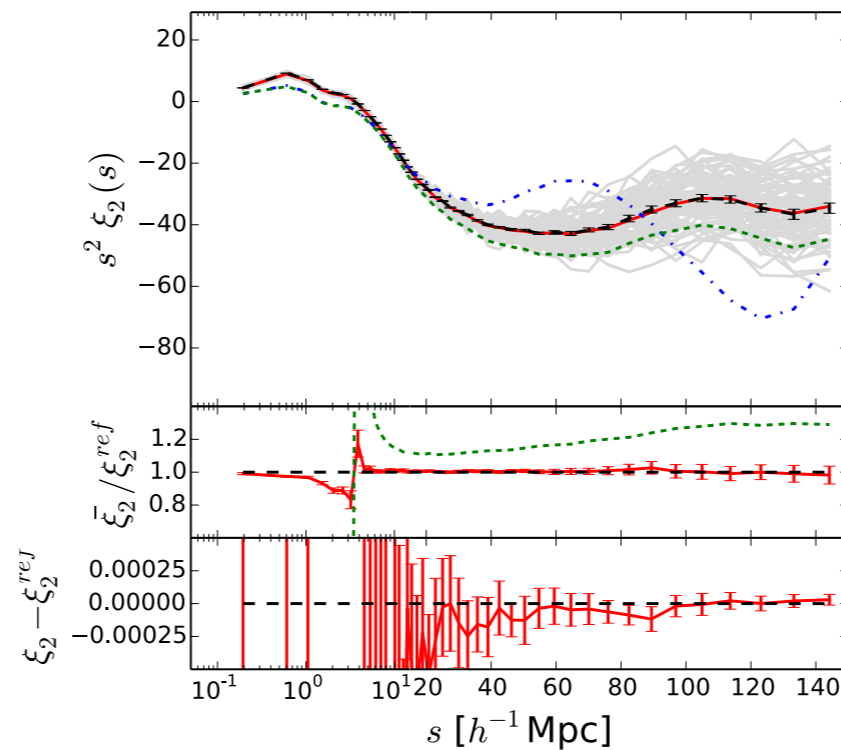
Comparison to DESI mocks: pass 2

completeness = 48%

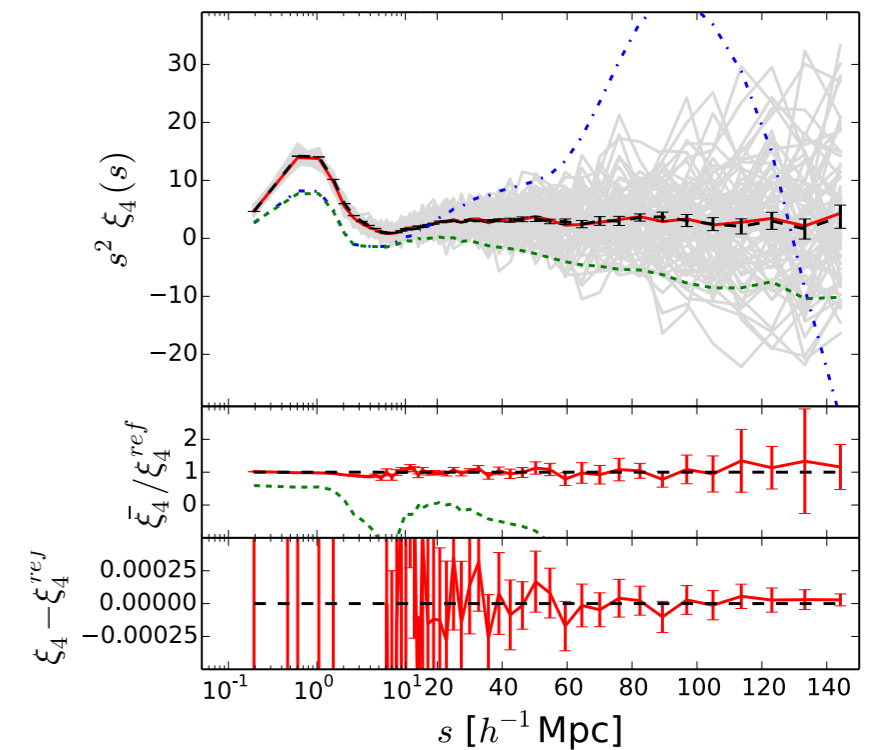
Monopole



Quadrupole



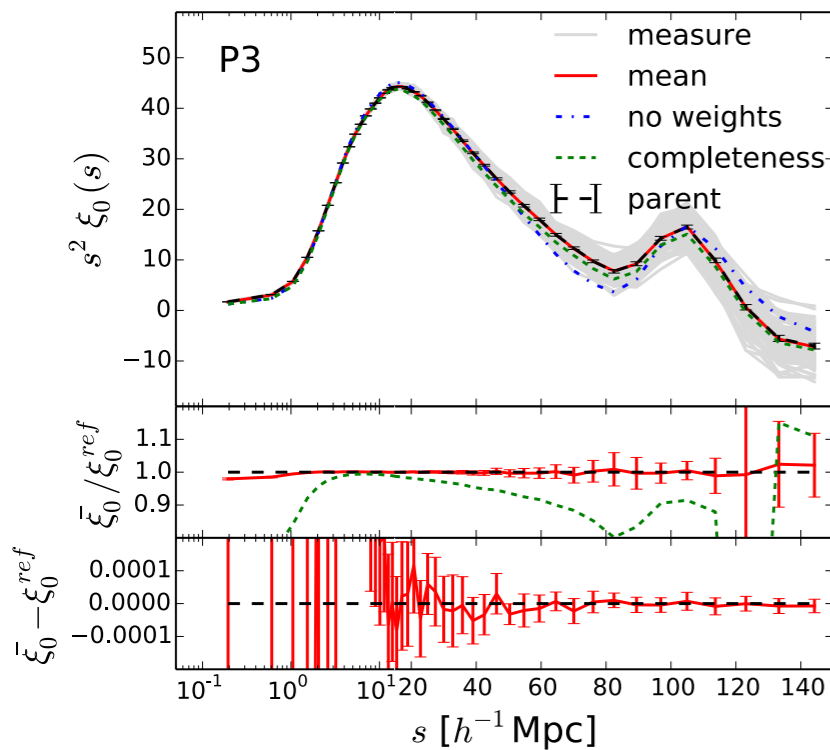
Hexadecapole



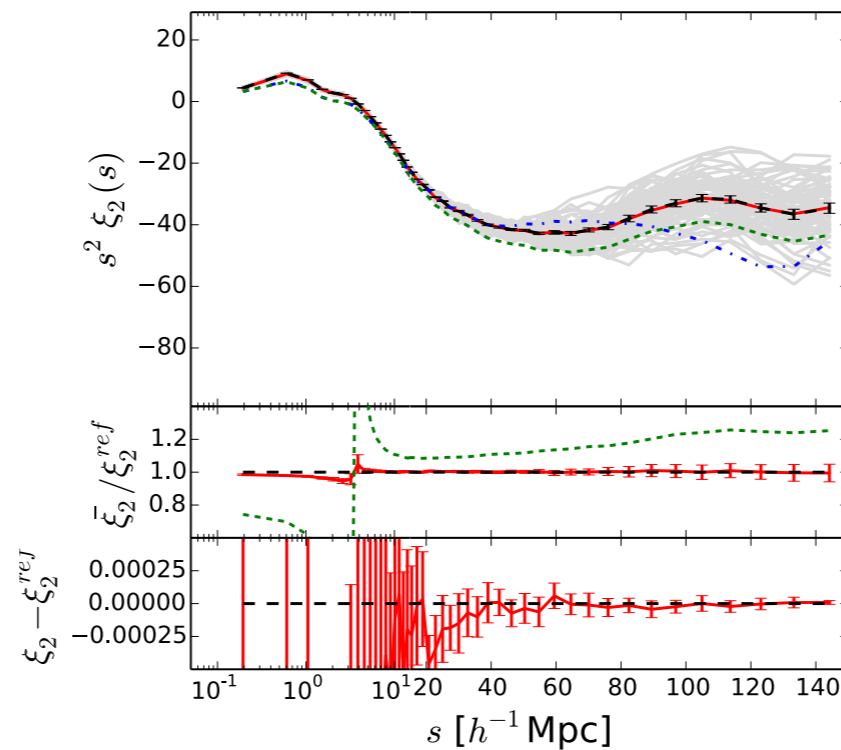
Comparison to DESI mocks: pass 3

completeness = 67%

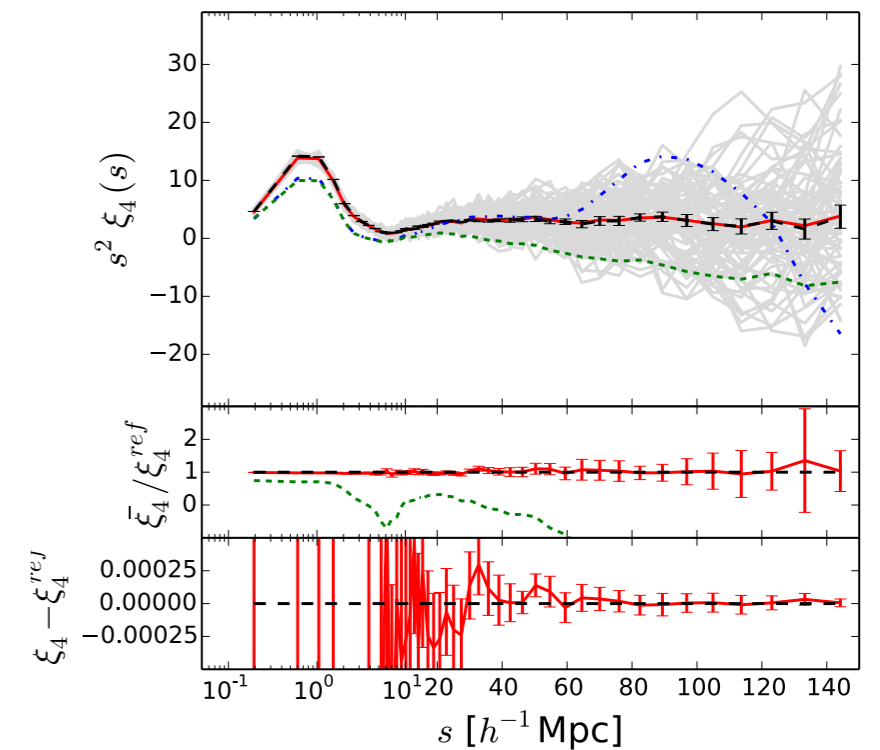
Monopole



Quadrupole



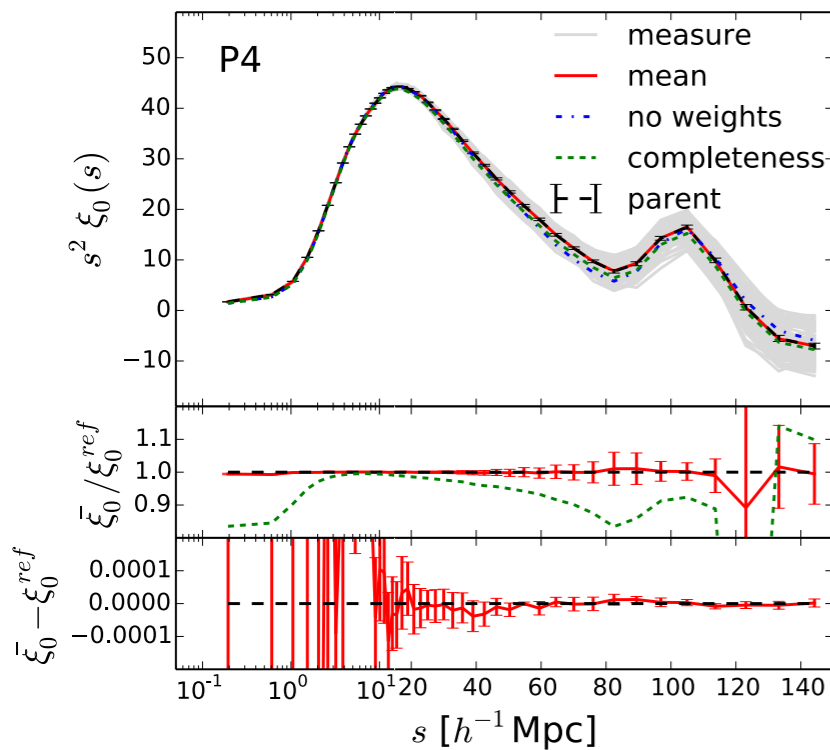
Hexadecapole



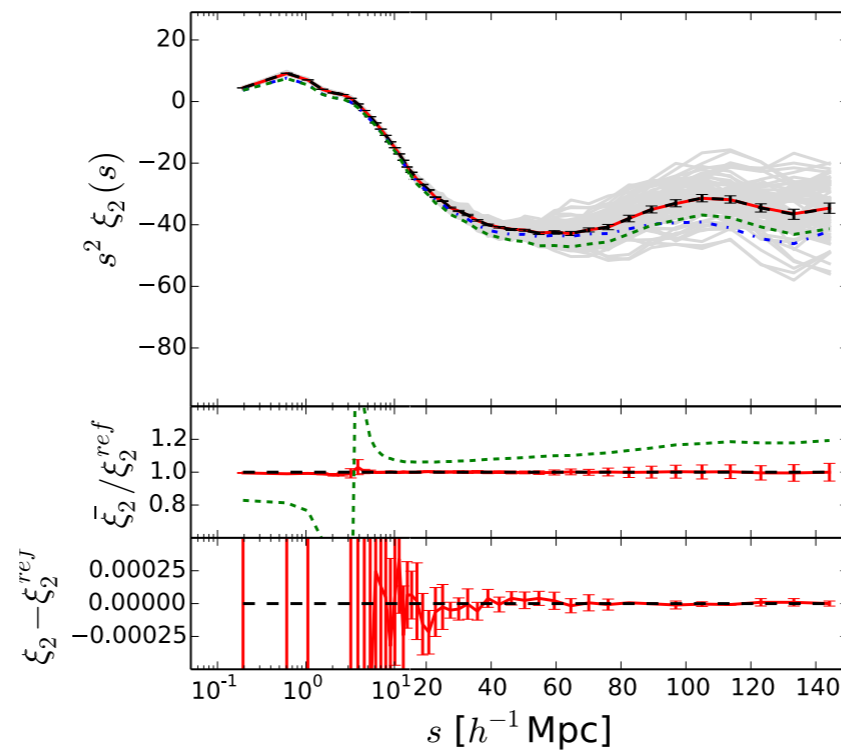
Comparison to DESI mocks: pass 4

completeness = 81%

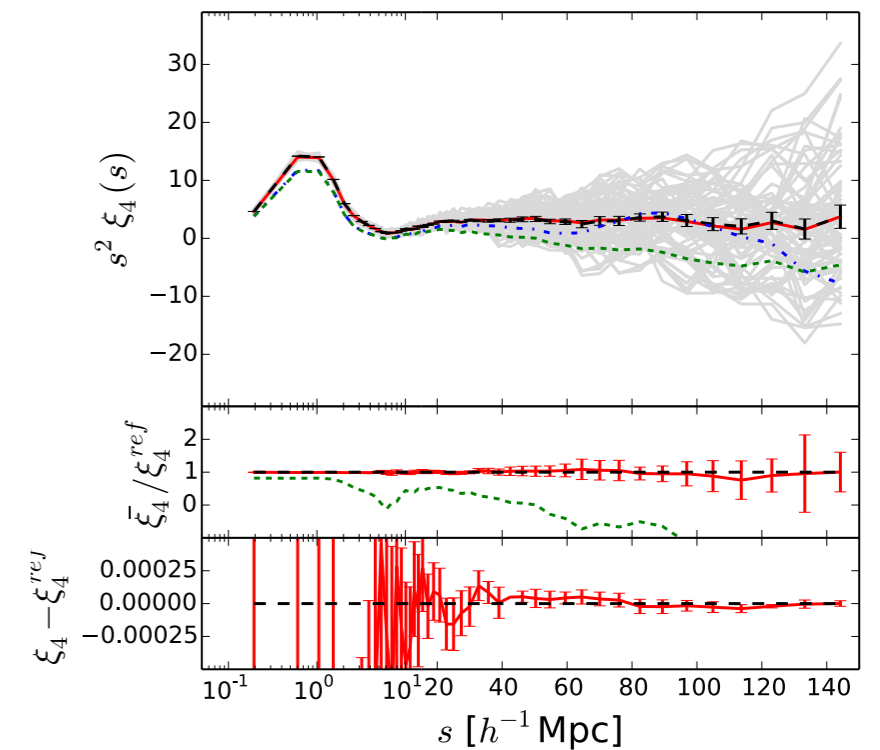
Monopole



Quadrupole



Hexadecapole

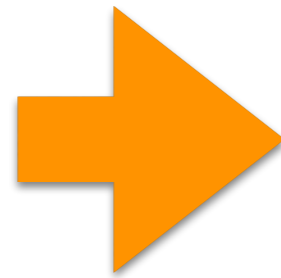


Practical implementation: bitwise weights

The pair weights are computed (as usual) while doing pair counts via a simple function of individual weights

Standard approach

$$w_{mn} = w_m w_n$$



Bitwise weights

$$w_{mn} = \frac{N_{bits}}{\text{popcnt} \left[w_m^{(b)} \text{ and } w_n^{(b)} \right]}$$

Practical implementation: summary

The pair weights are computed (as usual) while doing pair counts via a simple function of individual weights

*Standard
approach*

$$w_{mn} = w_m w_n$$

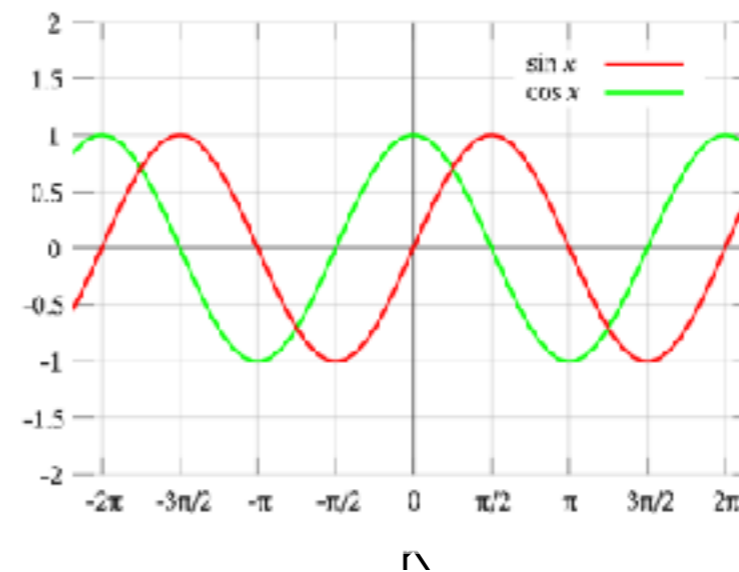
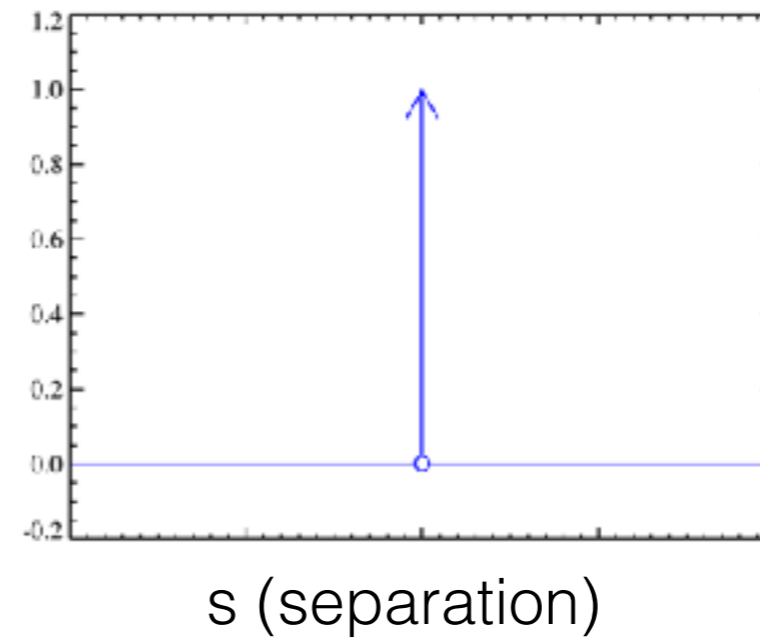
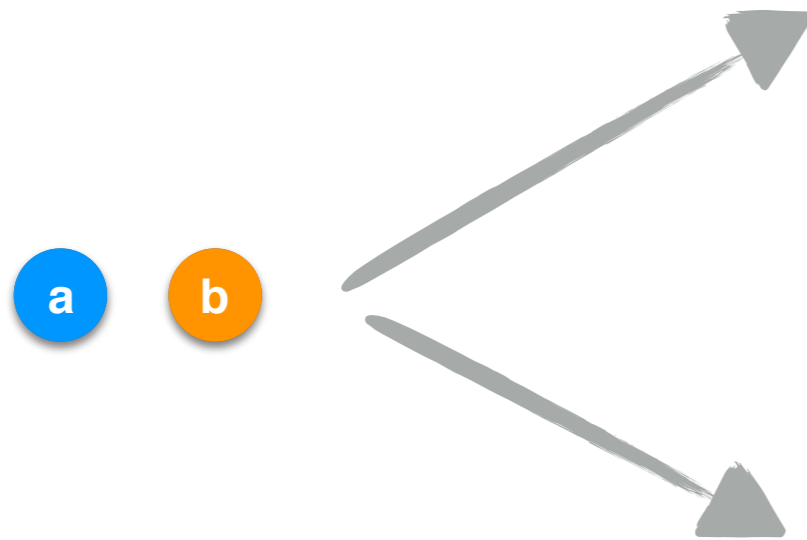


Bitwise weights

$$w_{mn} = \frac{N_{bits}}{\text{popcnt} [w_m^{(b)} \text{ and } w_n^{(b)}]}$$

$$\langle w_m^{(b)} | w_n^{(b)} \rangle$$

From pairs to k-space modes



Here s
is a wave
number

Power spectrum

(local plane-parallel approx)

$$P_\ell(k) = \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \left[\int d^3r_1 \int d^3r_2 F(\mathbf{r}_1) F(\mathbf{r}_2) e^{i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\eta}}) - S(\mathbf{k}) \right]$$

Line of sight
↓
 $\hat{\boldsymbol{\eta}}$

Legendre polynomials
↗

Shot noise
↖

$$F(\mathbf{r}) = n(\mathbf{r}) - \alpha n_s(\mathbf{r})$$

Yamamoto 2005

$$P_\ell(k) = \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \left\{ \left[\int d^3r_1 F(\mathbf{r}_1) e^{i\mathbf{k}\cdot\mathbf{r}_1} \right] \left[\int d^3r_2 F(\mathbf{r}_2) e^{-i\mathbf{k}\cdot\mathbf{r}_2} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{r}}_2) \right] - S_\ell(\mathbf{k}) \right\}$$

with further manipulation
(Bianchi et al. 2015) this term
can be evaluated via FFTs
(see also Scoccimarro 2015, Hand
et al. 2017)

Power spectrum

(local plane-parallel approx)

$$P_\ell(k) = \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \left[\int d^3r_1 \int d^3r_2 F(\mathbf{r}_1) F(\mathbf{r}_2) e^{i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\eta}}) - S(\mathbf{k}) \right]$$

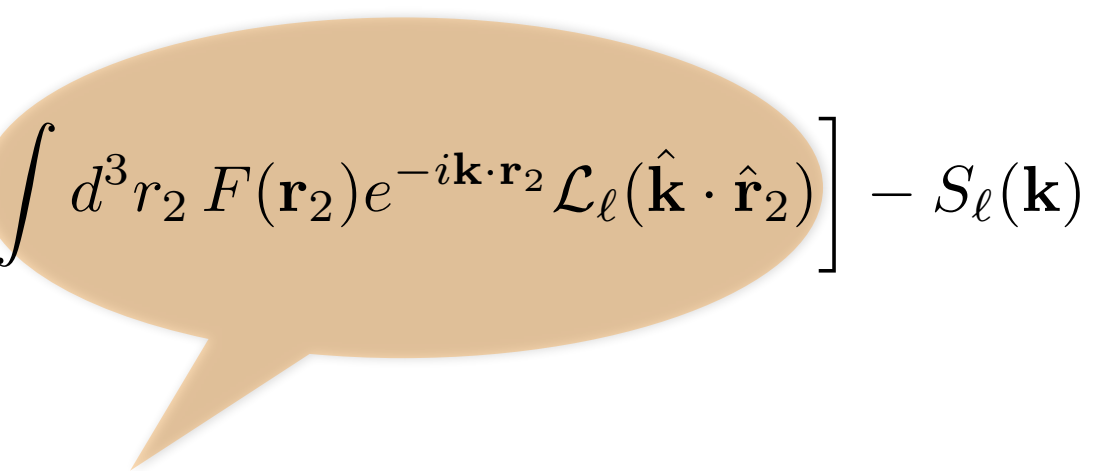
Line of sight
↓
 $\hat{\boldsymbol{\eta}}$

Legendre polynomials
↗

Shot noise
↖

$$F(\mathbf{r}) = n(\mathbf{r}) - \alpha n_s(\mathbf{r})$$

Yamamoto 2005

$$P_\ell(k) = \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \left\{ \left[\int d^3r_1 F(\mathbf{r}_1) e^{i\mathbf{k} \cdot \mathbf{r}_1} \right] \left[\int d^3r_2 F(\mathbf{r}_2) e^{-i\mathbf{k} \cdot \mathbf{r}_2} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{r}}_2) \right] - S_\ell(\mathbf{k}) \right\}$$


with further manipulation
(Bianchi et al. 2015) this term
can be evaluated via FFTs
(see also Scoccimarro 2015, Hand
et al. 2017)

Pairwise inverse probability (PIP) correction

$$P_{\ell}^{\text{PIP}}(k) = P_{\ell}^{\text{IIP}}(k) + \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \sum_{ij} A_{ij} e^{i\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \mathcal{L}_{\ell}(\hat{\mathbf{k}} \cdot \hat{\eta}_{ij})$$

$A_{ij} = w_{ij}^{\text{PIP}} - w_i^{\text{IIP}} w_j^{\text{IIP}}$

If $A_{ij} = 0$ above some angular separation \ll size of the survey, this term can be evaluated efficiently

Pairwise inverse probability (PIP) correction

$$P_{\ell}^{\text{PIP}}(k) = P_{\ell}^{\text{IIP}}(k) + \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \sum_{ij} A_{ij} e^{i\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \mathcal{L}_{\ell}(\hat{\mathbf{k}} \cdot \hat{\eta}_{ij})$$

$A_{ij} = w_{ij}^{\text{PIP}} - w_i^{\text{IIP}} w_j^{\text{IIP}}$

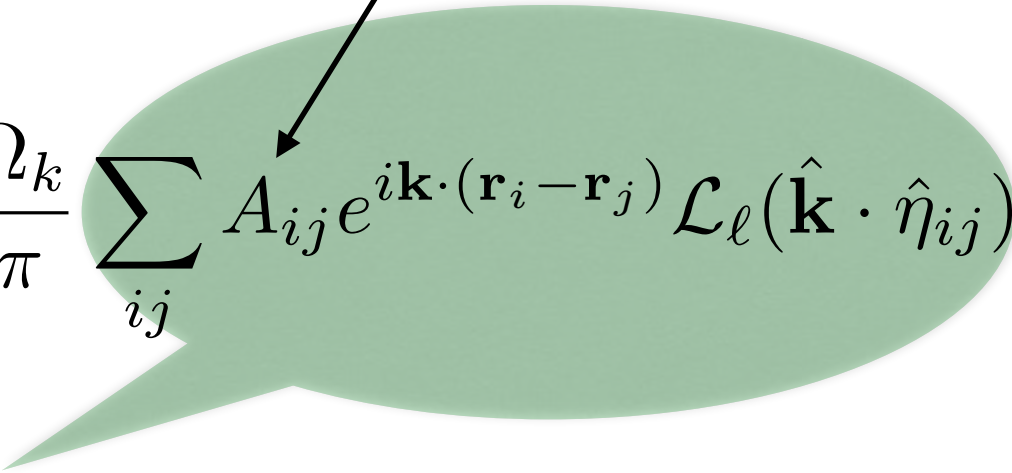
If $A_{ij} = 0$ above some angular separation \ll size of the survey, this term can be evaluated efficiently

In practice, for the PIP term, convenient to carry out the Ω_k integration first, i.e. we sum over Bessel rather than Fourier k -modes (see e.g. Wilson et al. 2015)

$$P_{\ell}^{\text{PIP}}(k) = P_{\ell}^{\text{IIP}}(k) + (-i)^{\ell} \frac{(2\ell + 1)}{I} \sum_{ij} A_{ij} j_{\ell}(ks_{ij}) \mathcal{L}_{\ell}(\hat{\mathbf{s}}_{ij} \cdot \hat{\eta}_{ij})$$

Pairwise inverse probability (PIP) correction

$$P_{\ell}^{\text{PIP}}(k) = P_{\ell}^{\text{IIP}}(k) + \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \sum_{ij} A_{ij} e^{i\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \mathcal{L}_{\ell}(\hat{\mathbf{k}} \cdot \hat{\eta}_{ij})$$

$$A_{ij} = w_{ij}^{\text{PIP}} - w_i^{\text{IIP}} w_j^{\text{IIP}}$$


If $A_{ij} = 0$ above some angular separation \ll size of the survey, this term can be evaluated efficiently

In practice, for the PIP term, convenient to carry out the Ω_k integration first, i.e. we sum over Bessel rather than Fourier k -modes (see e.g. Wilson et al. 2015)

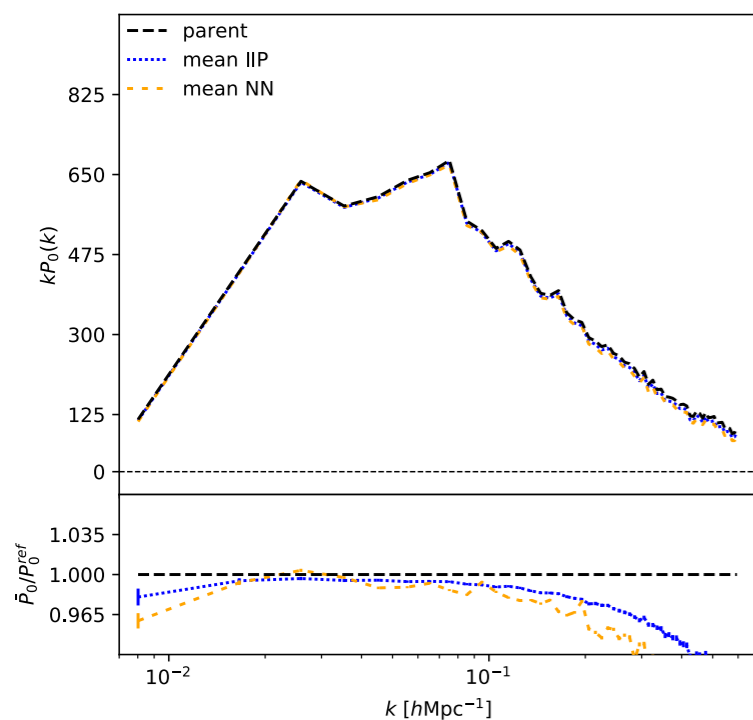
$$P_{\ell}^{\text{PIP}}(k) = P_{\ell}^{\text{IIP}}(k) + (-i)^{\ell} \frac{(2\ell + 1)}{I} \sum_{ij} A_{ij} j_{\ell}(ks_{ij}) \mathcal{L}_{\ell}(\hat{\mathbf{s}}_{ij} \cdot \hat{\eta}_{ij})$$

Also see Hahn et al. 2016 for interesting similarities and differences in the overall modelling

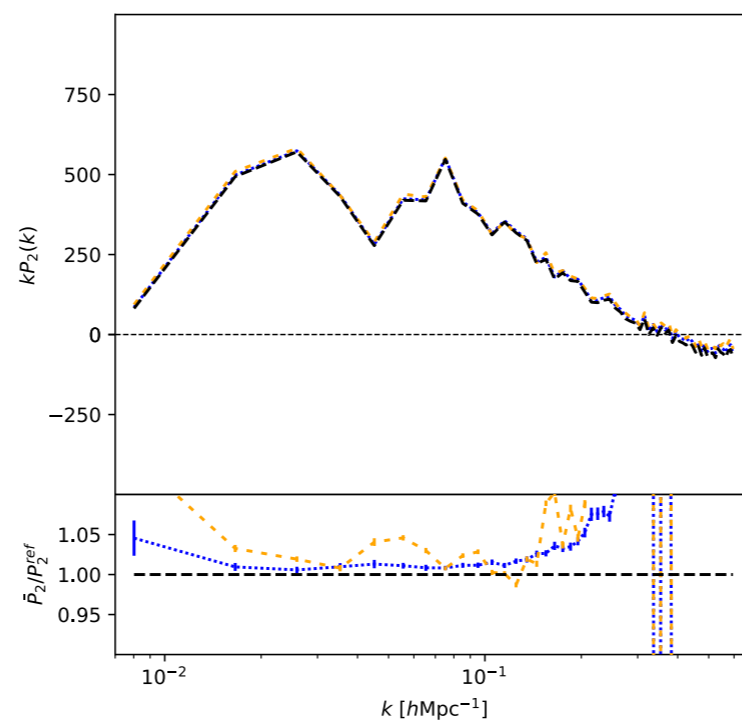
Using individual inverse probabilities (IIP)

MDR1 simulation, 1 Gpc periodic box, 2 passes of pure-fibre-collision targeting algorithm, 84% completeness

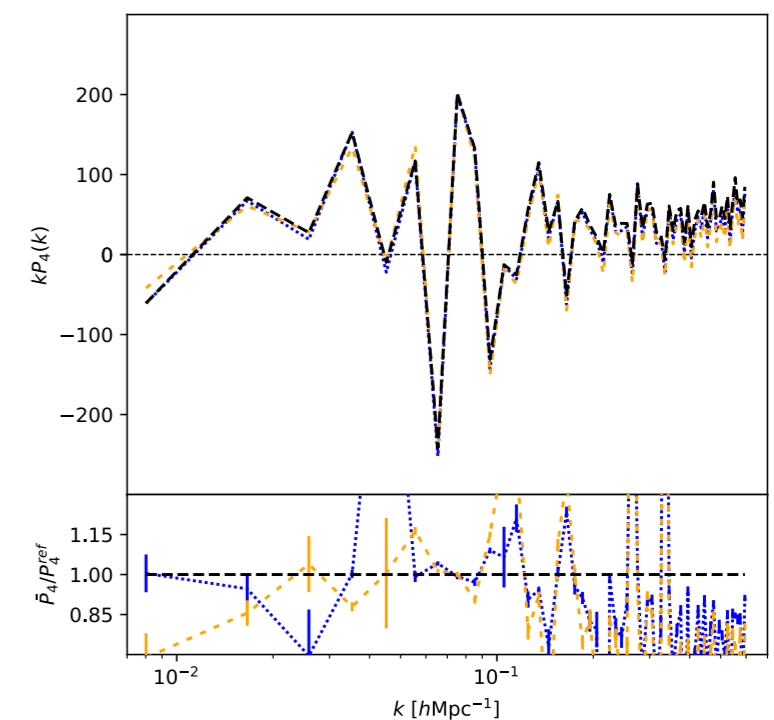
Monopole



Quadrupole



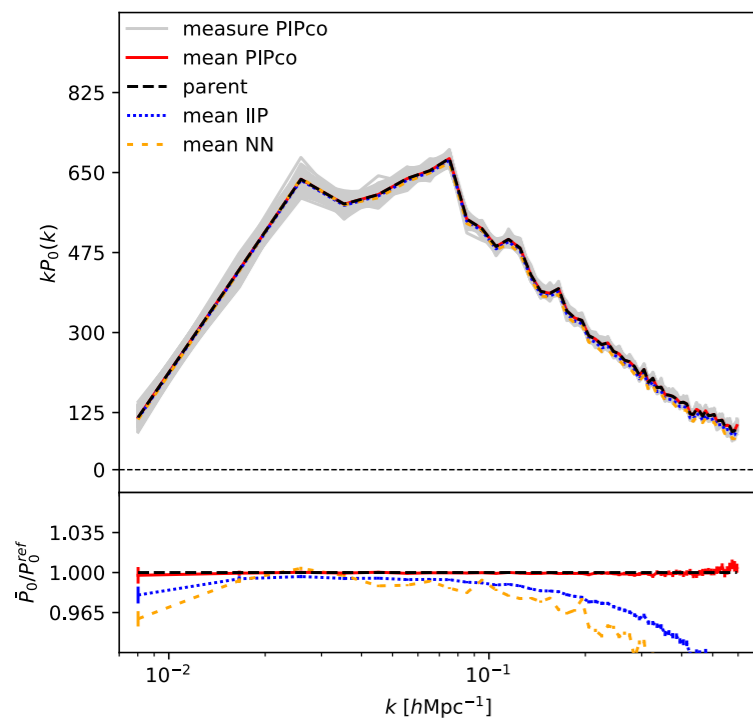
Hexadecapole



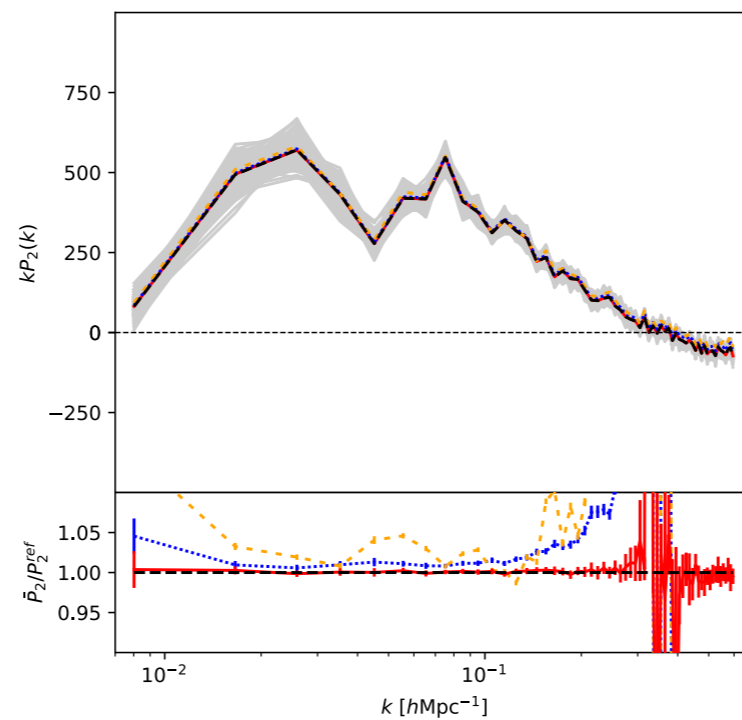
Comparison to simulations

MDR1 simulation, 1 Gpc periodic box, 2 passes of pure-fibre-collision targeting algorithm, 84% completeness

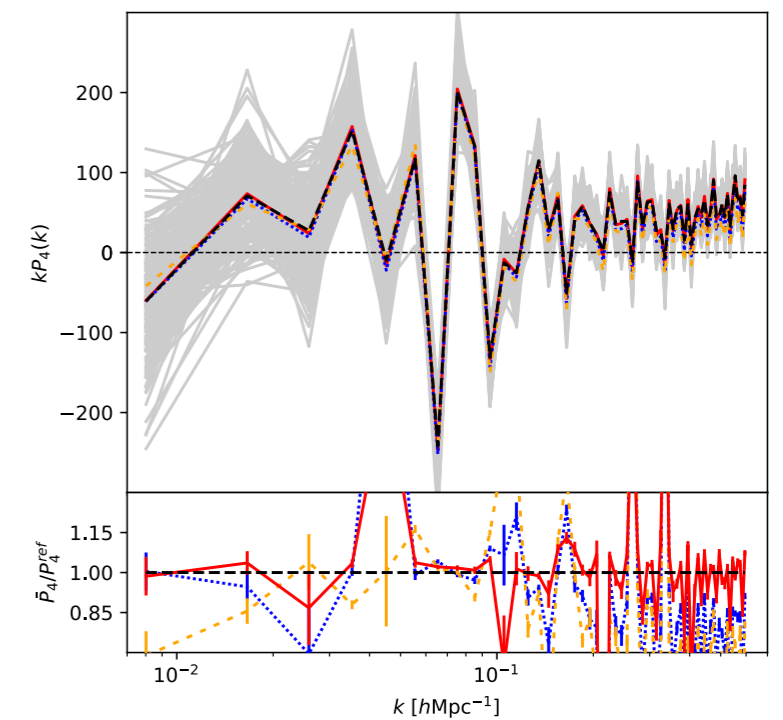
Monopole



Quadrupole



Hexadecapole



Summary

- We are at the dawn of a the era of percent precision. DESI it's already collecting redshifts, with unprecedented efficiency (around 2M galaxies in a few months).
- Such a statistical power requires exquisite control on systematics. We have developed an unbiased-by-construction approach (PIP) to deal with the missing observation issue
- More systematics will show up for DESI and the other next-generation surveys, which will require dedicated effort (statistical modelling, machine learning, cross correlations between different probes, etc.)

DESI imaging parent sample

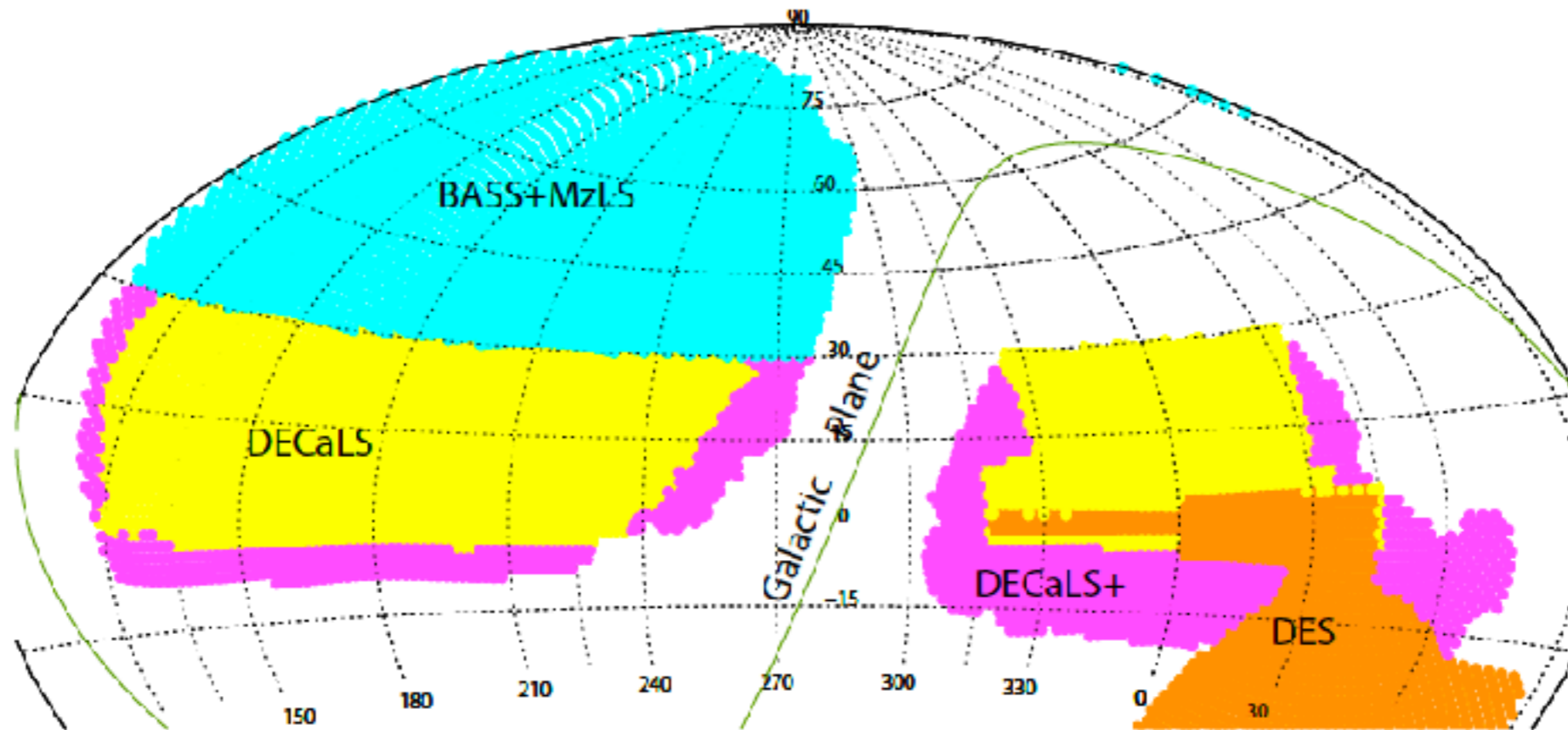


Figure 3.19: The primary imaging surveys that will result in targeting data for the DESI project. The footprint at $\text{DEC} < +34^\circ$ will be covered using the Dark Energy Camera (DECam) on the Blanco 4m telescope at Cerro Tololo Inter American Observatory. The Dark Energy Camera Legacy Survey (DECaLS, in yellow), the Dark Energy Survey (DES, in orange), and the extended DECaLS in the North Galactic Cap (DECaLS₁, in purple on left) are underway. A proposal for the remaining extended DECaLS in the South Galactic Cap (DECaLS₊, in purple on right) will be submitted. Imaging of the North Galactic Cap region at $\text{DEC} \geq +34^\circ$ (cyan) will be covered with the 90Prime camera at the Bok 2.3-m telescope in g - and r -bands (BASS: the Beijing-Arizona Sky Survey) and with the upgraded MOSAIC-3 camera on the Mayall 4m telescope in z -band (MzLS: the MOSAIC z -band Legacy Survey). Both the Bok and Mayall telescopes are located on Kitt Peak National Observatory.

ELG photometric selection (illustrative example)

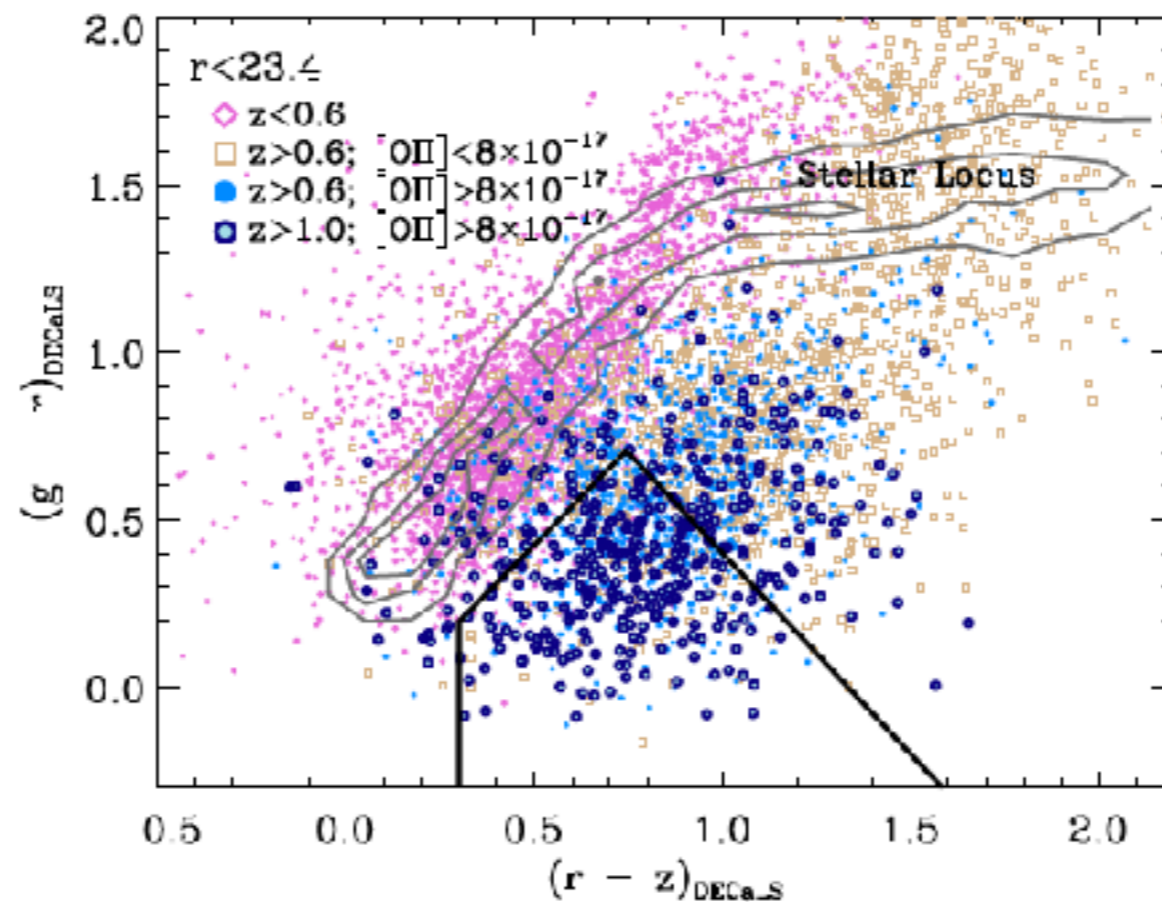


Figure 3.10: Optical $g-r$ vs. $r-z$ color-color diagram based on spectroscopy from the DEEP2 Galaxy Redshift Survey, illustrating our preliminary selection for ELGs at $z > 0.6$ with significant [O II] emission-line flux. Although the galaxy photometry is based on deep CFHTLS imaging [208], the colors have been transformed and degraded to the expected depth of the DECaLS imaging. This plot shows that strong [O II]-emitting galaxies at $z > 0.6$ (blue points) are in general well-separated from both the population of lower-redshift galaxies (pink diamonds) and from the locus of stars in this color space (grey contours). The selection box (thick black polygon) selects those galaxies with strong [O II]-emission while minimizing contamination from stars and lower-redshift interlopers.