SDSS3-BOSS
Baryon Oscillations Spectroscopic Survey
What do we learn from it?

Shirley Ho (Carnegie Mellon University)

Work with Xiaoying Xu, Mariana Vargas, Ross O'Connel, Nishant Agarwal, Hy Trac, Martin White, Daniel Eisenstein, Antonio Cuesta, Nikhil Padmanabhan, Anze Slosar, Angela Burden, Jeff Schneider (CS), Barnabas Poczos (CS), Junier Oliver (CS), Andrea Klein (CS), SDSS3-BOSS Collaboration


BRUCE AND ASTRID MCWILLIA
Center for Cosmolog
Carnegie Mellon

Outline

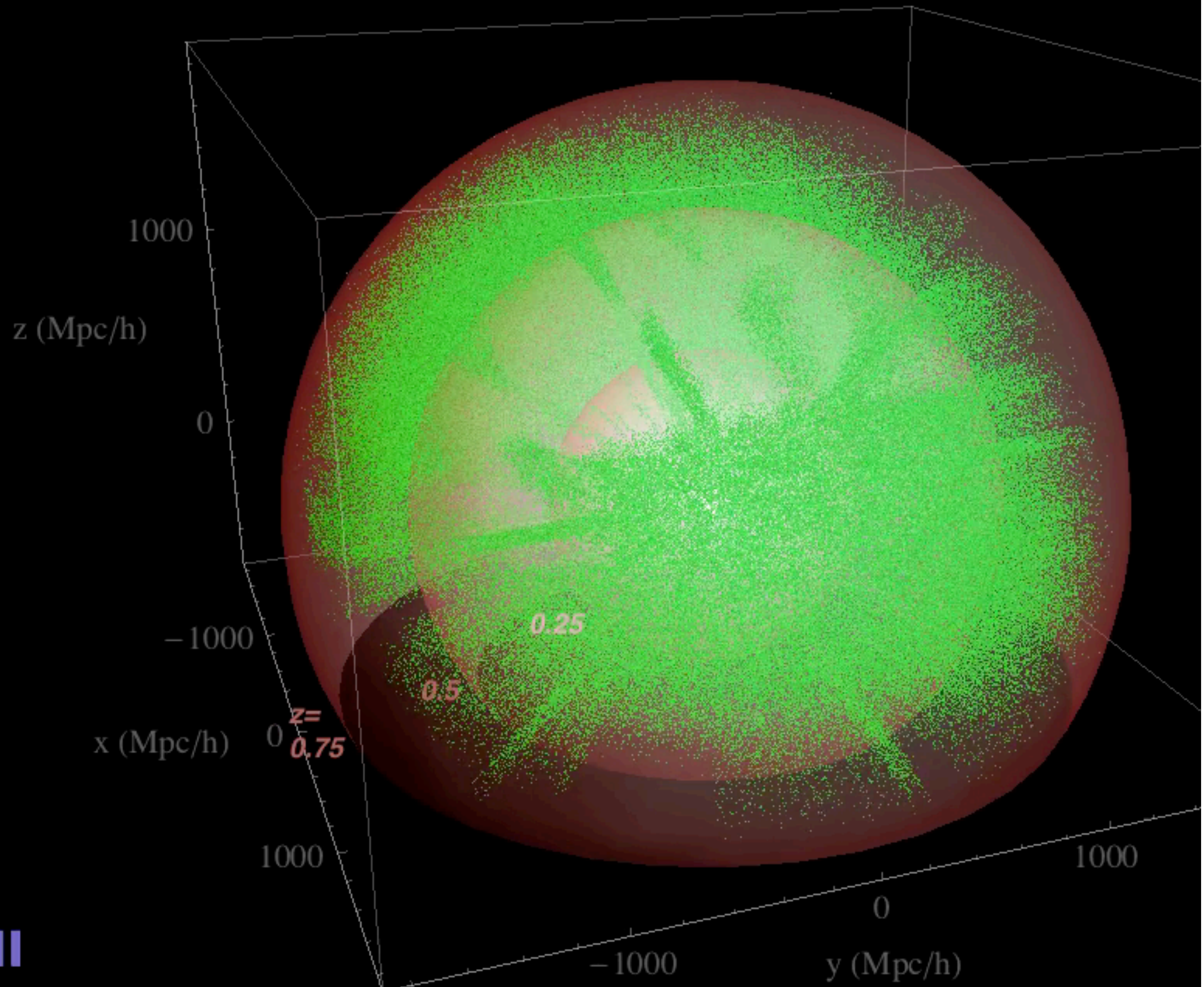
- SDSS III- BOSS :
  - Mini-summary and looking forward
- BAO Reconstruction
- Ultra-fast simulations of the Universe?
- BAO in cross-correlations?

Outline

- SDSS III- BOSS :
  - Mini-summary and looking forward
- BAO Reconstruction
- Ultra-fast simulations of the Universe?
- BAO in cross-correlations

SDSS3-BOSS
Data Release 8 (DR8)

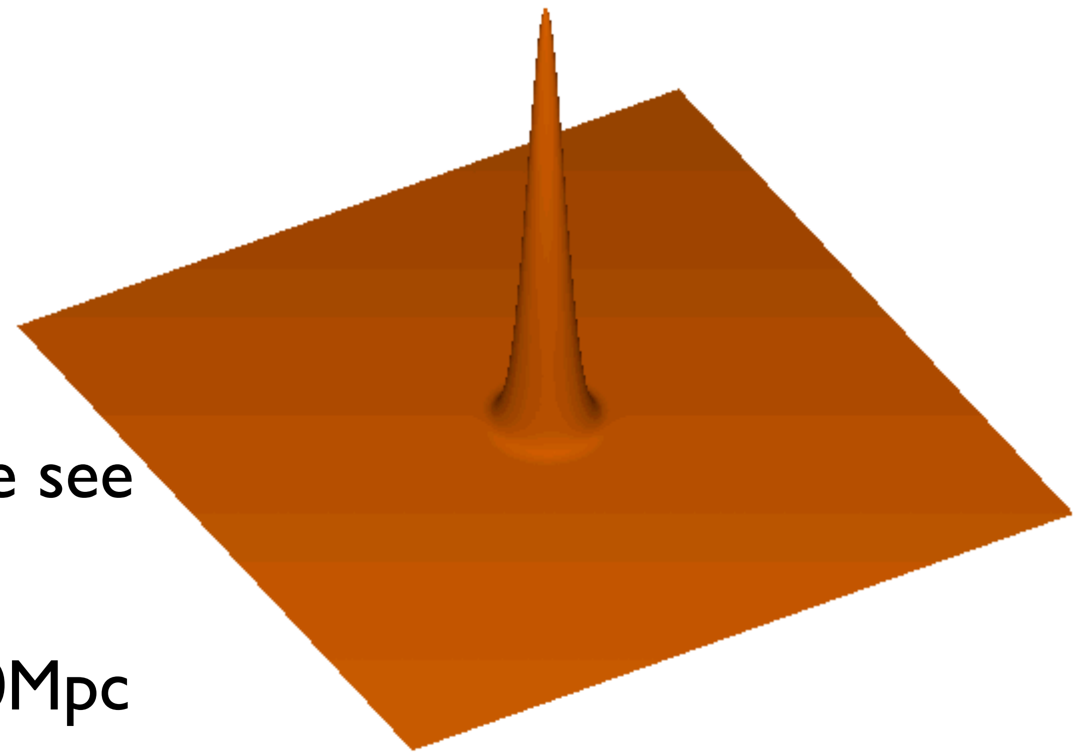
video credit: David Kirkby



SDSS III

Luminous Galaxies

- Each initial overdensity is an over-pressure that launches a spherical sound wave.
- This sound wave travel at 57% speed of light.
- Pressure providing photons decouple at recombination. We see these photons as the CMB
- Wave stalls at a radius of ~150Mpc
- Overdensity in shell and center both seed formation of galaxies. Preferred separation of 150Mpc

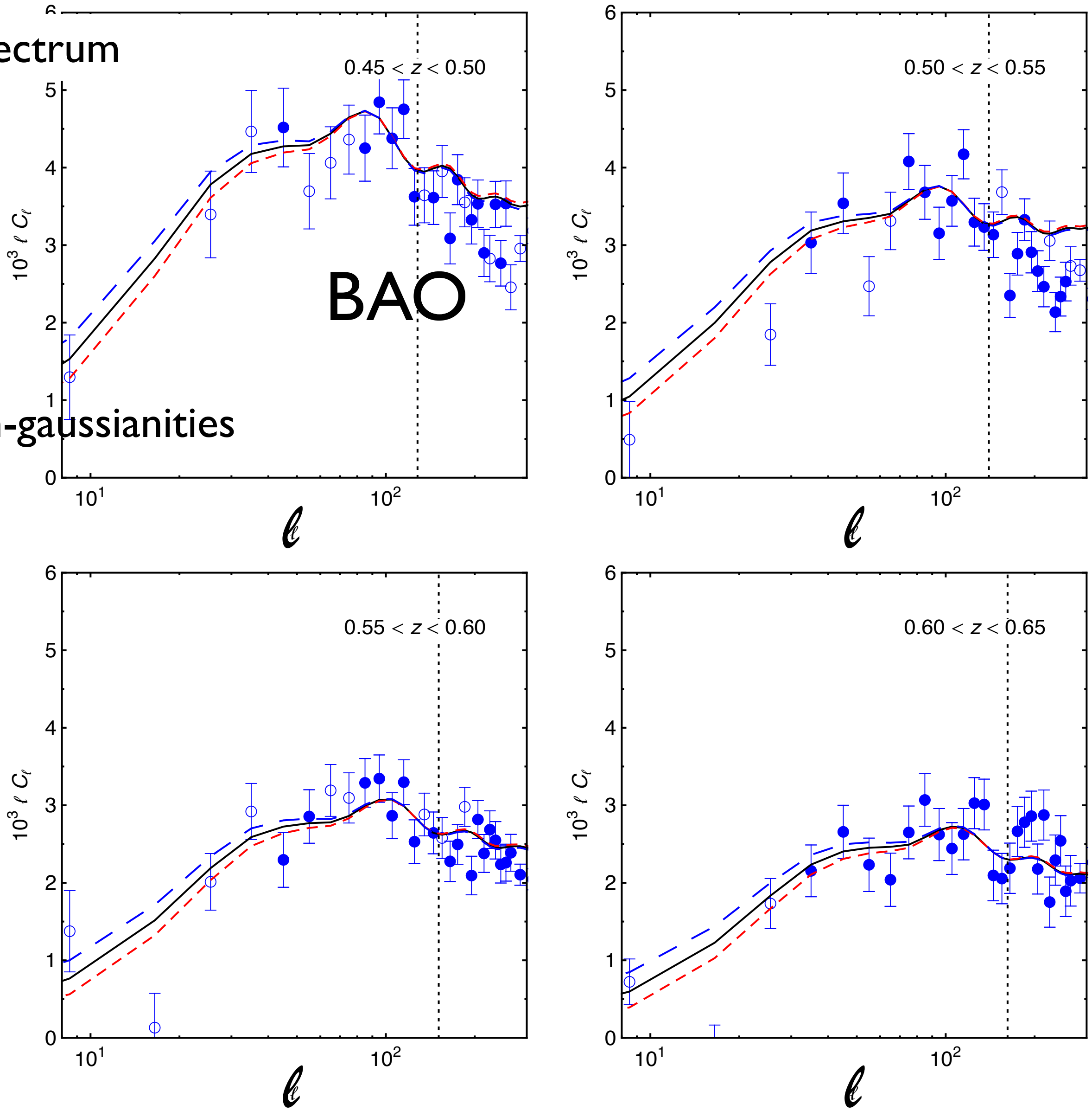


In the beginning, a sound wave …

What can you do with imaging data?

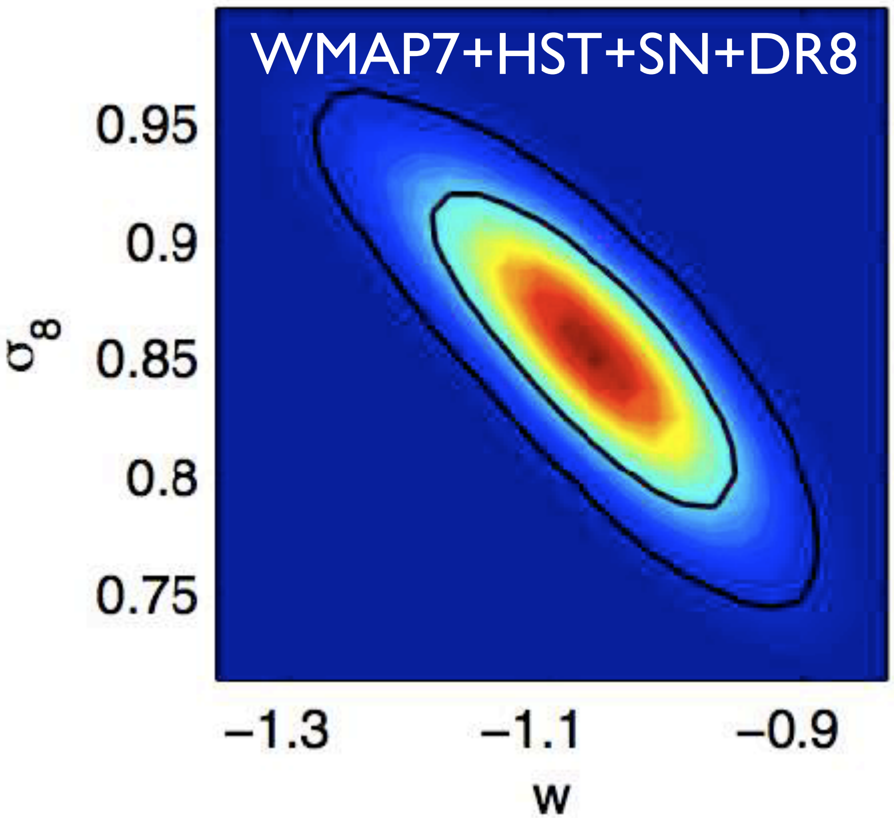
Angular power-spectrum

primordial non-gaussianities



SH, Agarwal, Myers et al. 2013

Cosmological Constraints from the overall shape

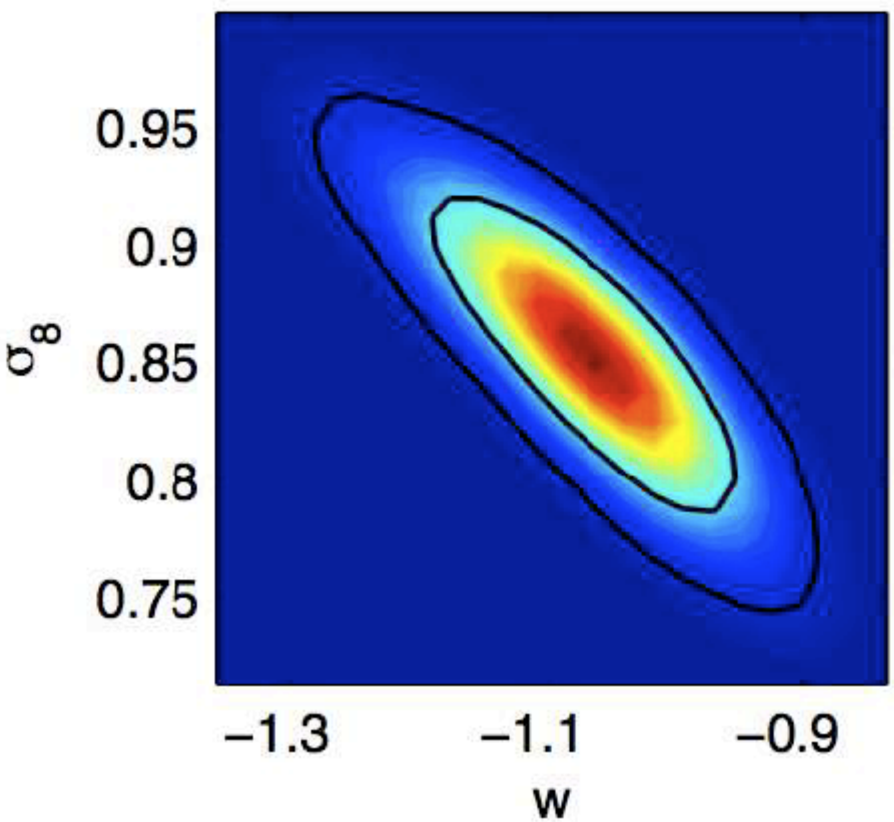


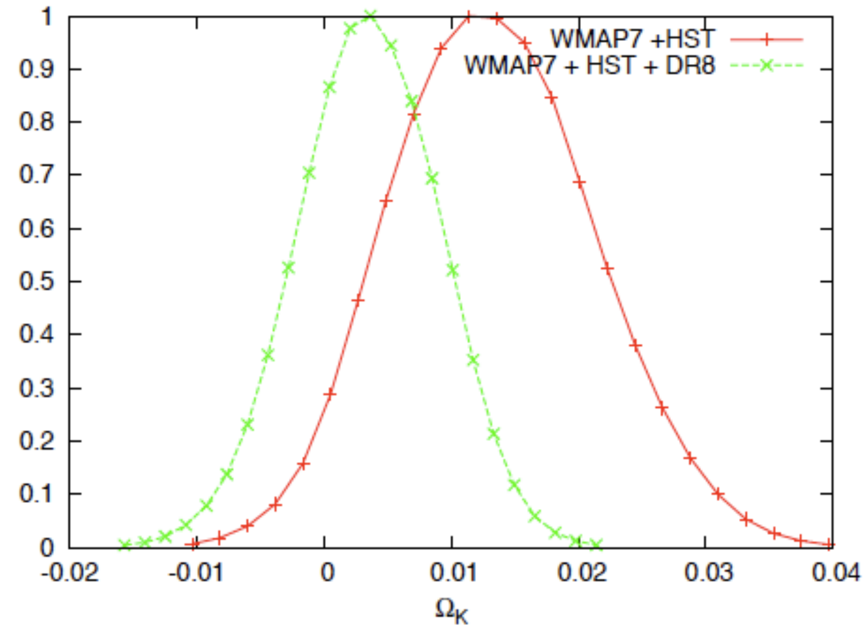
Combining with WMAP7+SN+HST, Dark Energy equation of state is constrained to 7% (1-sigma)

SDSS III

SH, Cuesta, Seo, Ross, DePutter et al. (2012)

Cosmological Constraints from the overall shape



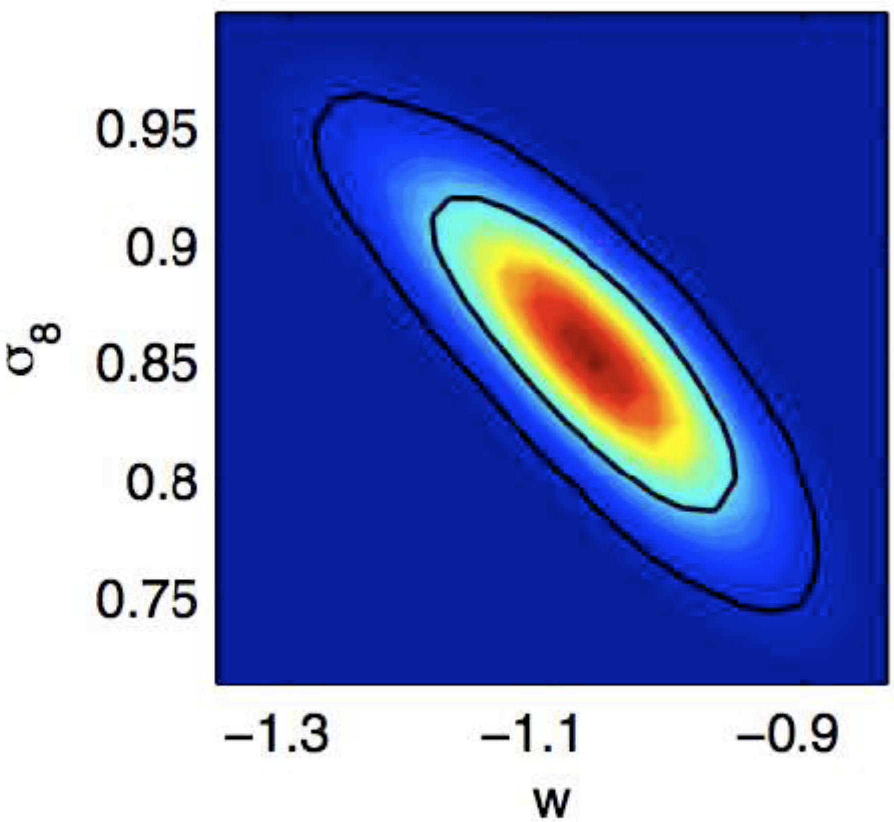


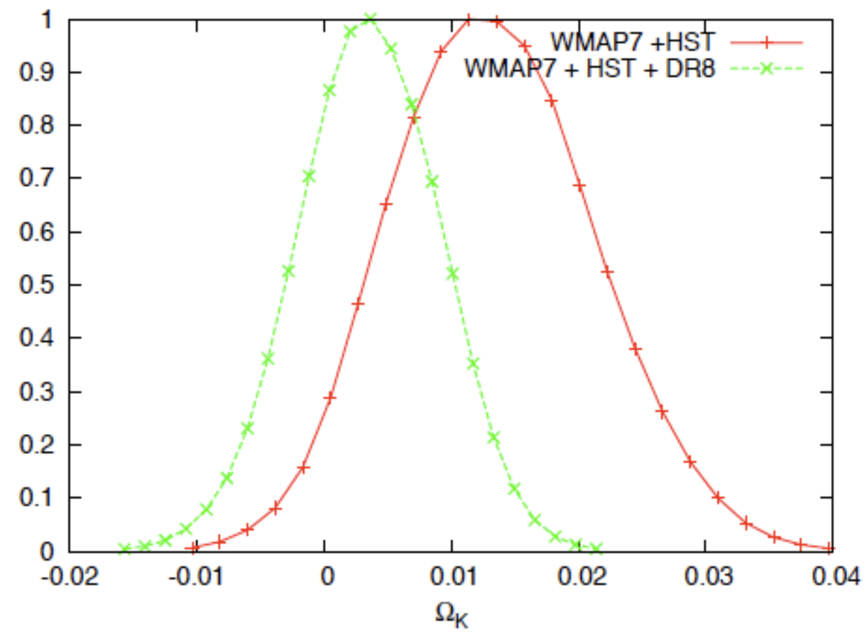
By including DR8 angular clustering (+WMAP+HST), we improve the constraint on flatness of the Universe by 40% over WMAP7+HST

SDSS III

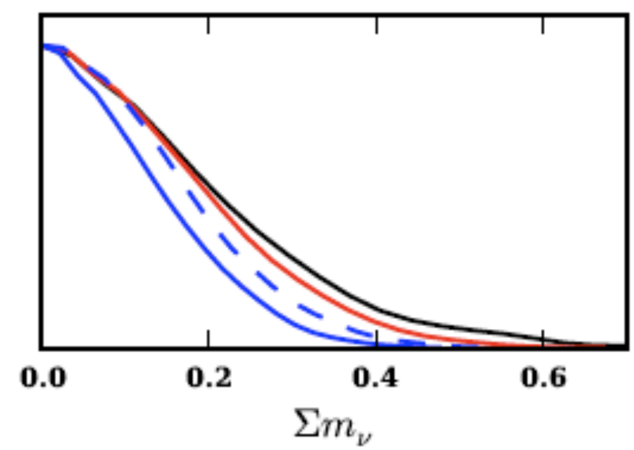
SH, Cuesta, Seo, Ross, DePutter et al. (2012)

Cosmological Constraints from the overall shape





The sum of neutrino masses is constrained to less than 0.26eV



DePutter, Mena, Guisarma, SH, Seo et al. (2012)
SH, Cuesta, Seo, Ross, DePutter et al. (2012)

What else can you do with imaging data?

What else can you do with imaging data?

Testing initial conditions of the Universe

- The power spectrum of dark matter halos is given by

$$P_{\text{halo}}(M,k,z,f_{\text{NL}}) = [b_1(M,z) + \Delta b(M,k,z,f_{\text{NL}})]^2 P_{\text{matter}}(k,z)$$

with

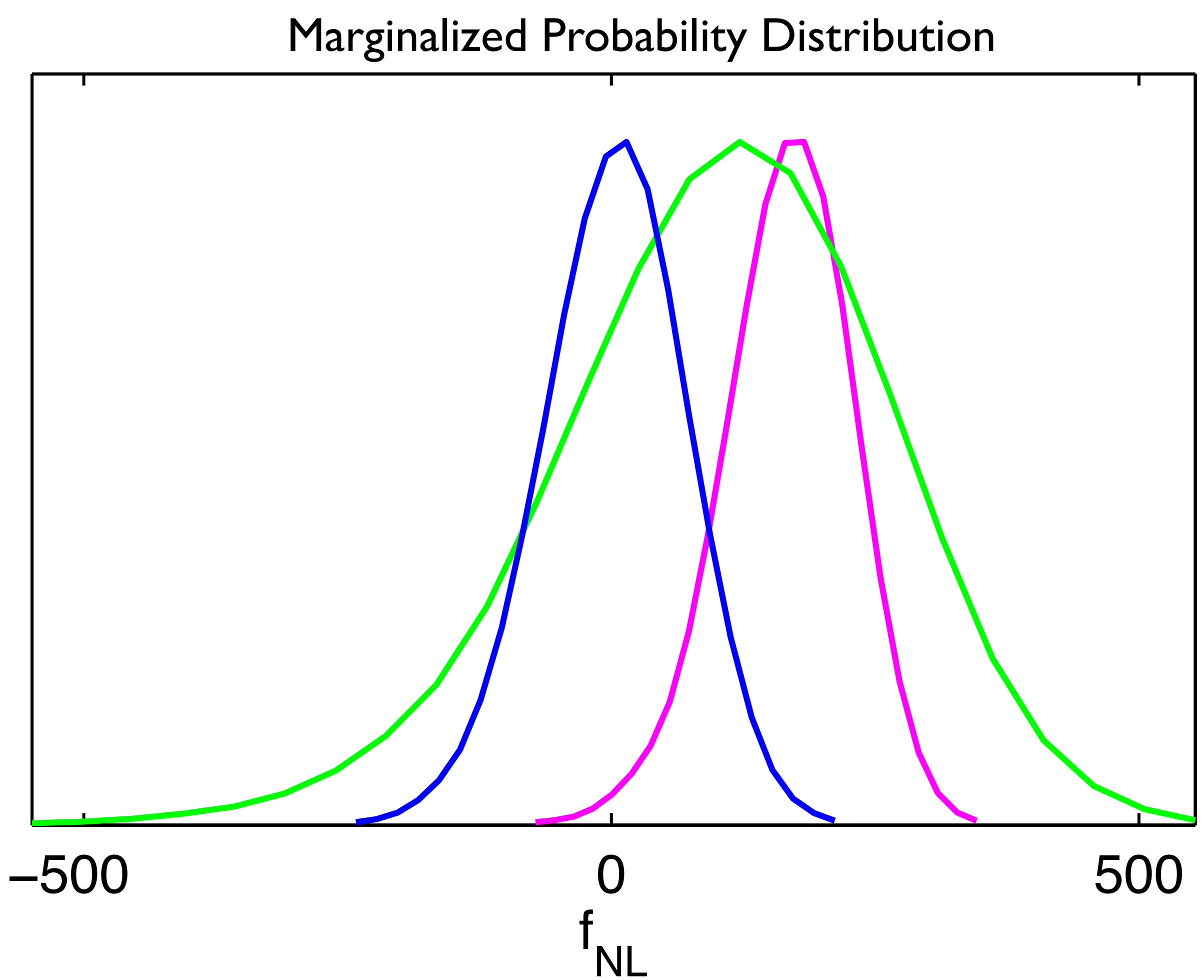
$$\Delta b_{\text{non-Gaussian}} \propto f_{\text{NL}} \frac{(b_1(M,z) - p)}{k^2}$$

S. Matarrese, F. Lucchin, and S. A. Bonometto, 1986; N. Dalal, O. Dore, D. Huterer, and A. Shirokov, 2008; S. Matarrese and L. Verde, 2008; A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, 2008

What else can you do with imaging data?

Testing initial conditions of the Universe

Dalal et al., 2008, Slosar et al. 2008...

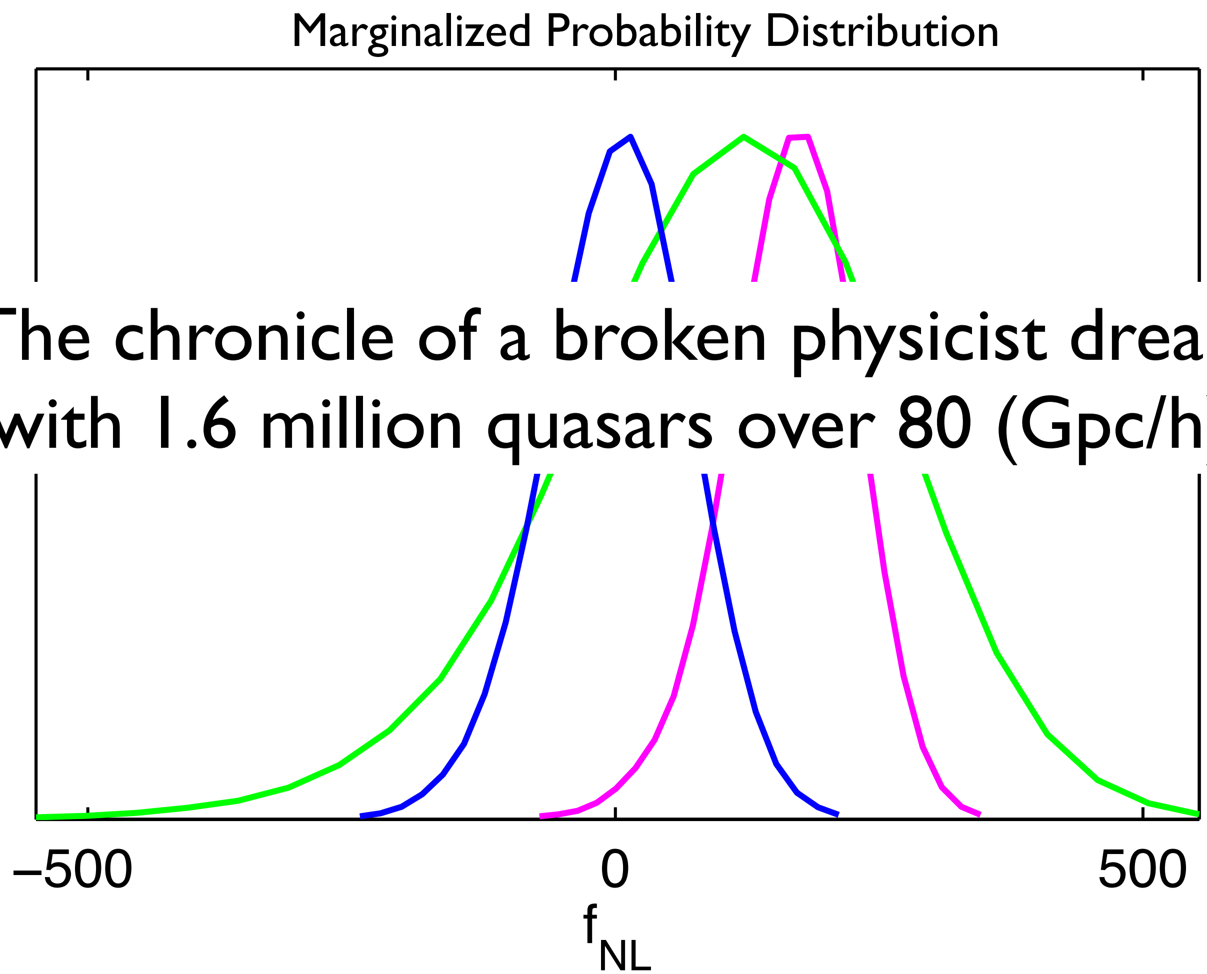


SH, Agarwal, Myers et al. 2013 (under SDSS3 review)

What else can you do with imaging data?

Testing initial conditions of the Universe

Dalal et al., 2008, Slosar et al. 2008...



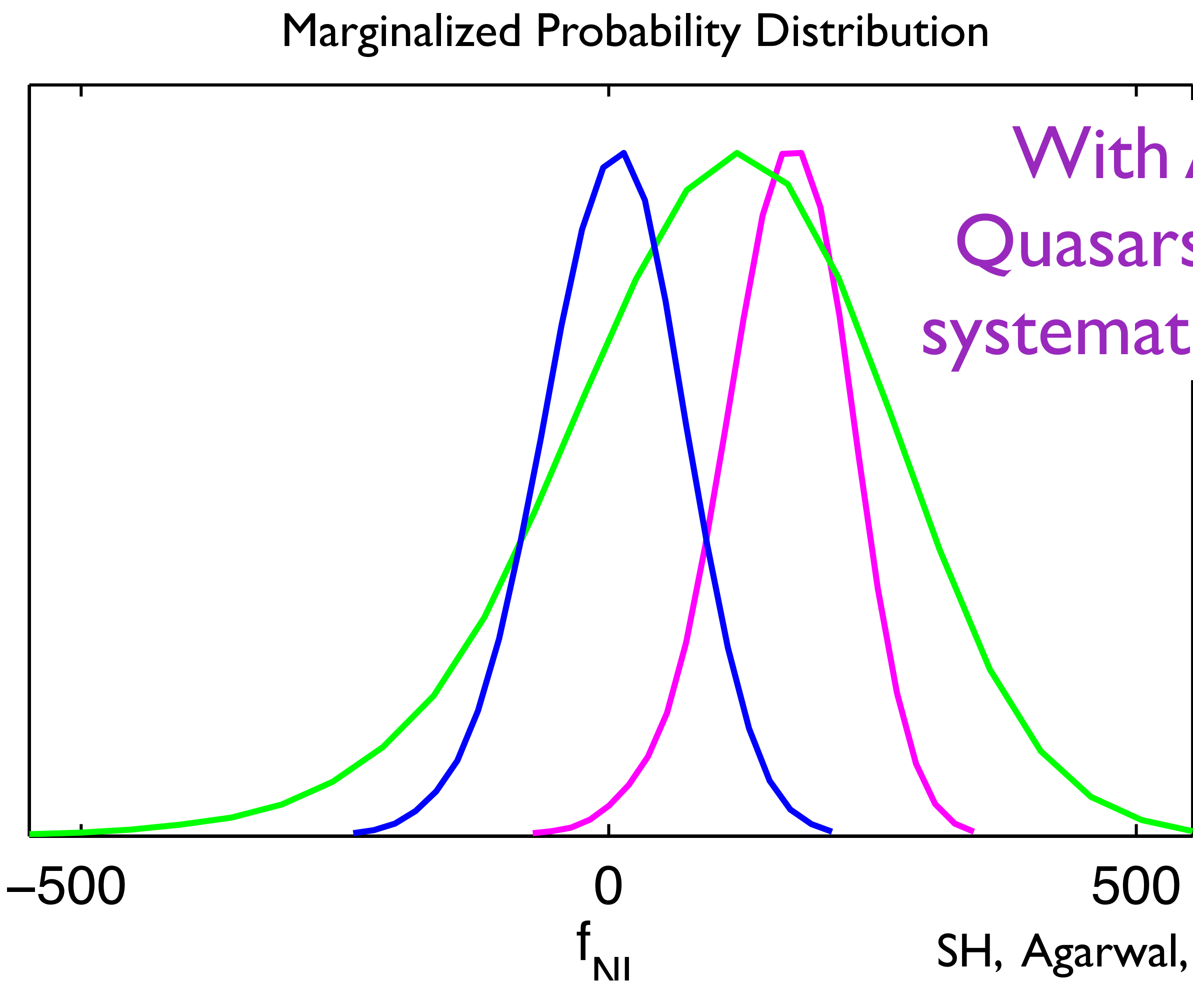
The chronicle of a broken physicist dream....
with 1.6 million quasars over 80 (Gpc/h)^3

SH, Agarwal, Myers et al. 2013 (under SDSS3 review)

What else can you do with imaging data?

Testing initial conditions of the Universe

Dalal et al., 2008, Slosar et al. 2008...



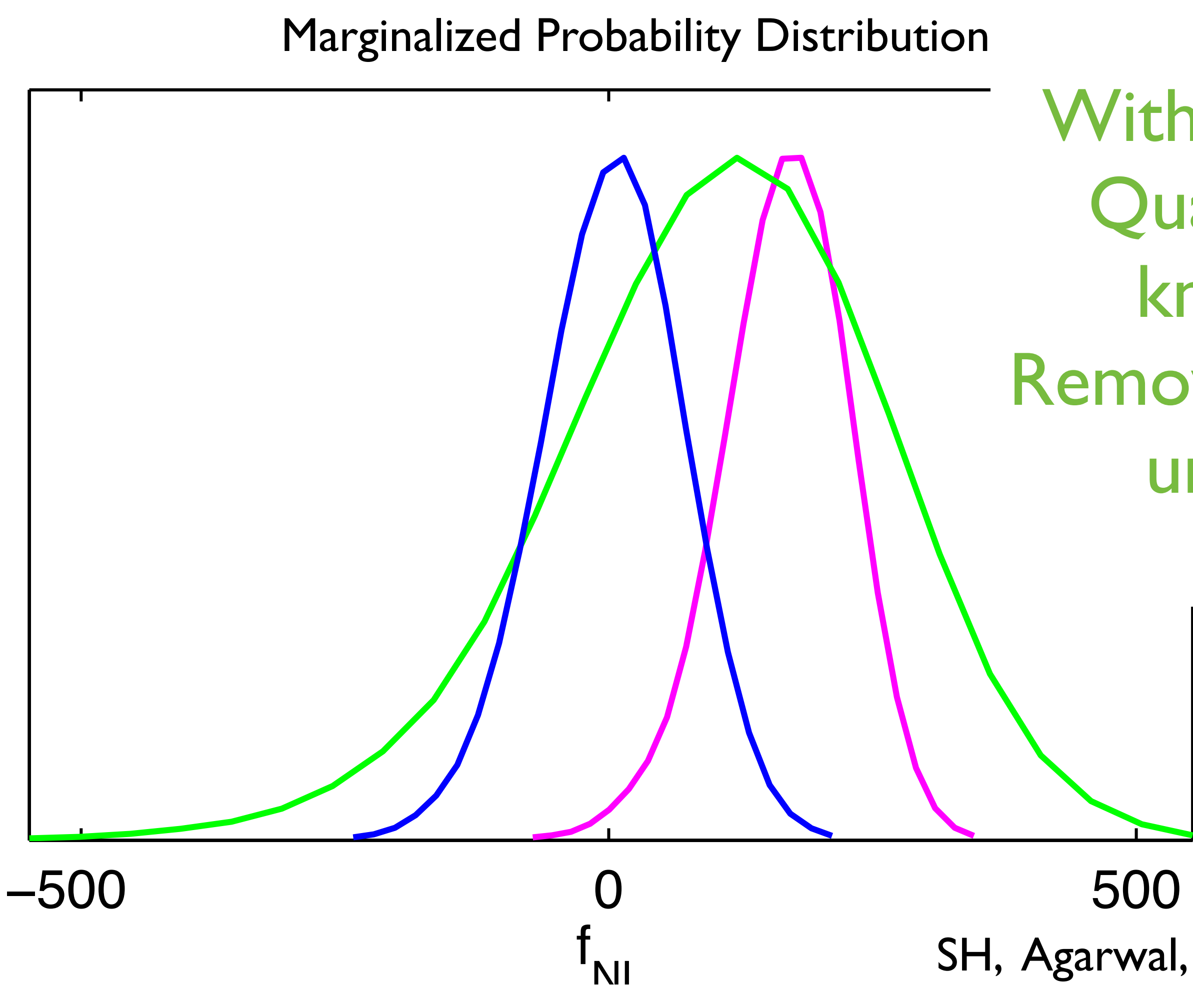
With Angular Clustering of Quasars: Project out all known systematics + starting from l >30

SH, Agarwal, Myers et al. 2013 (under SDSS3 review)

What else can you do with imaging data?

Testing initial conditions of the Universe

Dalal et al., 2008, Slosar et al. 2008...



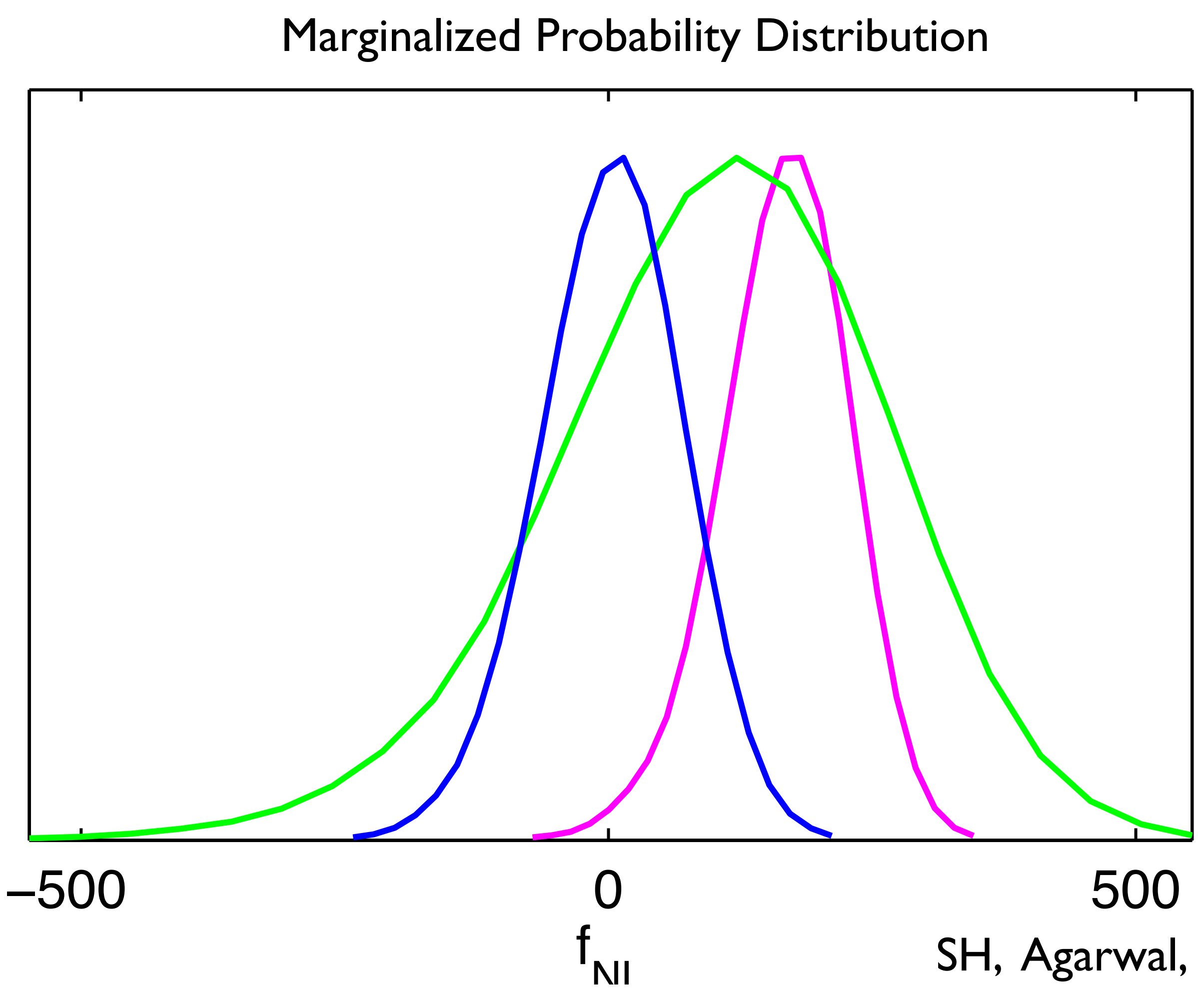
With Angular Clustering of Quasars: Project out all known systematics + Removing angular scales with unknown systematic contaminations

SH, Agarwal, Myers et al. 2013 (under SDSS3 review)

What else can you do with imaging data?

Testing initial conditions of the Universe

Dalal et al., 2008, Slosar et al. 2008...



Adding both cleaned QSOs + LRGs

SH, Agarwal, Myers et al. 2013 (under SDSS3 review)

What else can you do with imaging data?

Testing initial state of inflation !

- For non-trivial initial states, the squeezed limit bispectrum is proportional to

$$1/\left(k_S^2 k_L^4\right) = 1/\left(k_S^3 k_L^3\right) \times \left(k_S/k_L\right)$$

- This changes the scale-dependence of the non-Gaussian halo bias: $\Delta b_{\mathrm{non-Gaussian}} \propto 1/k^3$

J. Ganc and E. Komatsu, 2012; I. Agullo and S. Shandera, 2012

What else can you do with imaging data?

Testing initial state of inflation !

$$\Delta b\left(M,k,z,\mathcal{A}_{\rm NL},\alpha\right) = 3\mathcal{A}_{\rm NL}(b_1(M,z))[b_1(M,z)-p]\frac{\Omega_m H_0^2}{k^2(k/k_p)^{\alpha-2}T(k)D(z)},$$

- Therefore, we can parameterize the bias correction as

$$\Delta b_{\rm non-Gaussian} \quad \propto \quad \mathcal{A}_{\rm NL}\left(b_1(M,z)\right)\frac{(b_1(M,z)-p)}{k^{\alpha}}$$

- Exact local ansatz: $\alpha = 2$

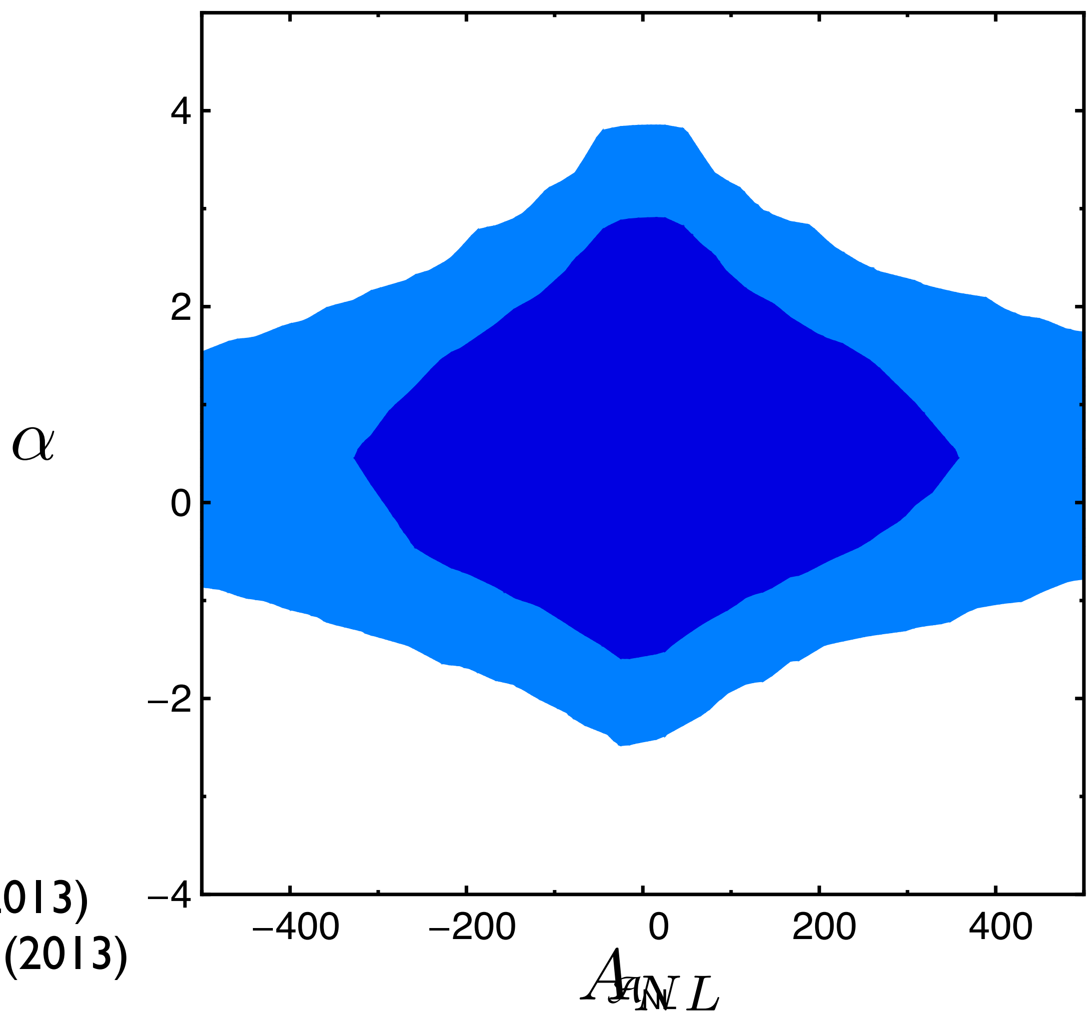
  General initial states: $\alpha \sim 3$

  Multiple fields: $0 \leq \alpha \lesssim 2 + \mathcal{O}(\epsilon)$ (S. Shandera, N. Dalal, and D. Huterer, 2011; E. Sefusatti, J. R. Fergusson, X. Chen, and E. Shellard, 2012; M. Dias, R. Ribeiro, and D. Seery, 2013)

What else can you do with imaging data?

Testing initial state of inflation !

$$\Delta b\,(M,k,z,\mathcal{A}_{\mathrm{NL}},\alpha) = 3\mathcal{A}_{\mathrm{NL}}(b_1(M,z))[b_1(M,z)-p]\frac{\Omega_m H_0^2}{k^2(k/k_p)^{\alpha-2}T(k)D(z)},$$



Giannantonio et al. (2013)
Agarwal, SH & Shadera (2013)

Testing initial state of inflation ! Why I shouldn't despair yet...

This is what would have happened if we have no systematics!

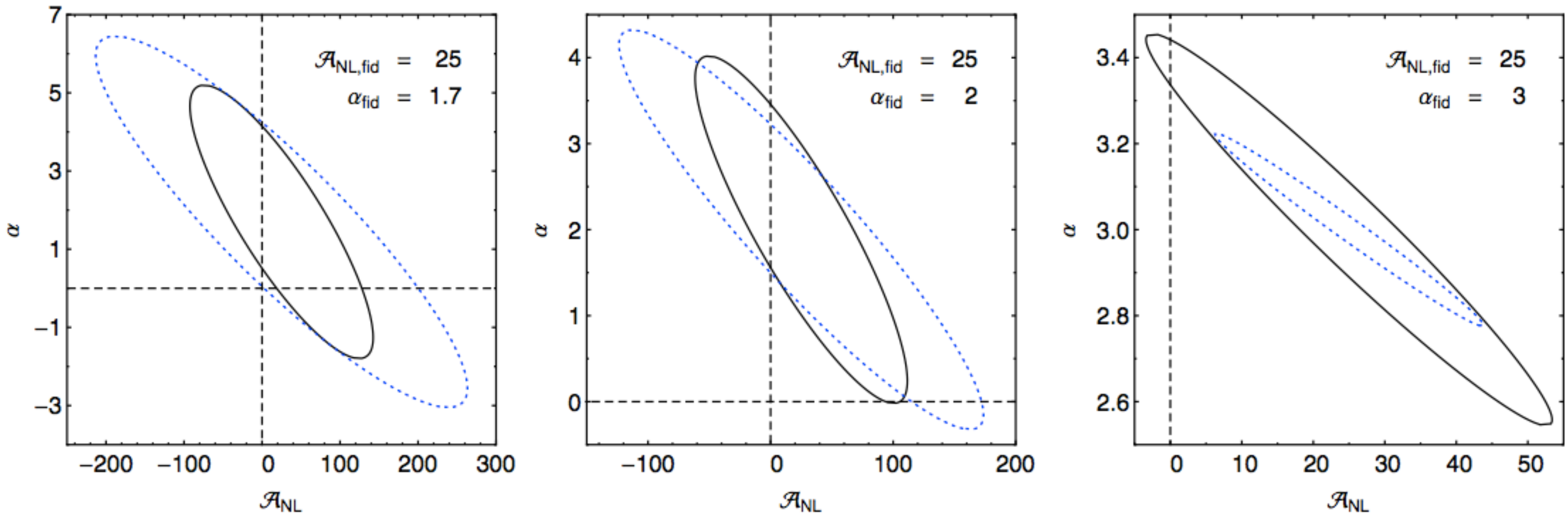


Figure 5. The one-sigma ellipses for LRGs (solid black) and quasars (dotted blue), with fiducial $\mathcal{A}_{\rm NL}(k_p = 0.1\ {\rm Mpc}^{-1}) = 25$ and $\alpha = 1.7,\ 2,\ 3$. See also table 5, which lists the results at the pivot points where $\mathcal{A}_{\rm NL}$ and α are uncorrelated for each choice of $\alpha_{\rm fid}$ and each data set (LRGs or quasars).

Agarwal, SH & Shadera (2013)

What happens when we have BOSS Spectroscopy ?

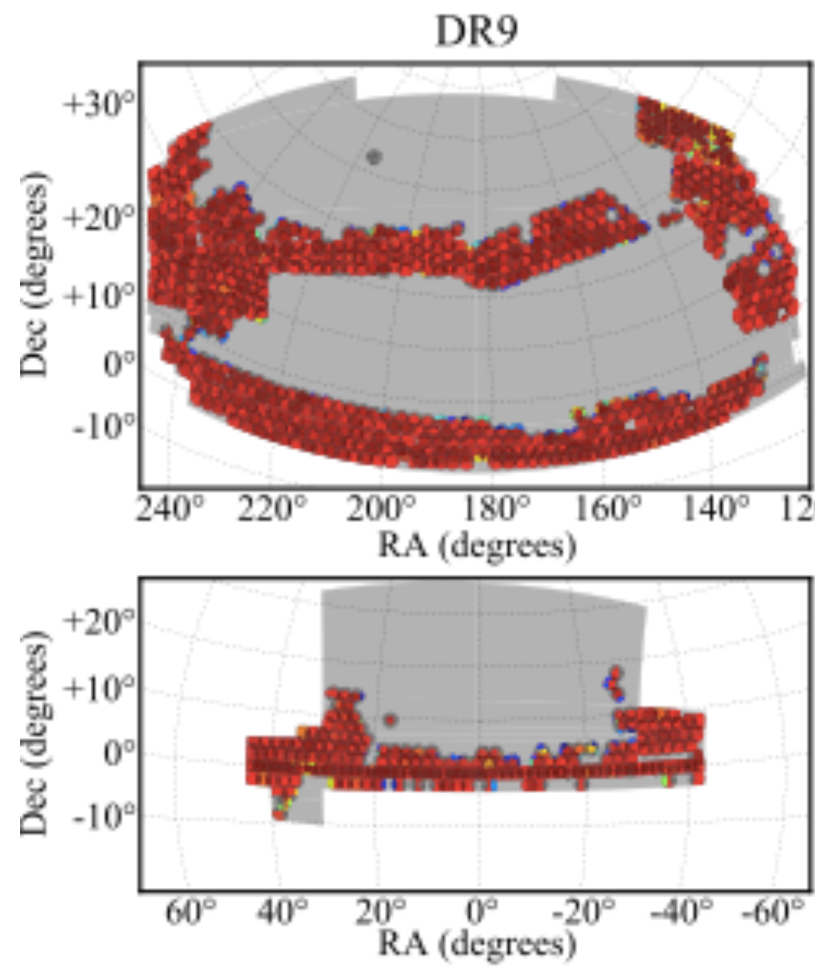
What happens when we have BOSS Spectroscopy ?

Claudia plugging a BOSS plate and getting it wrong!

SDSS III

Data Release 9-12 Spectroscopy

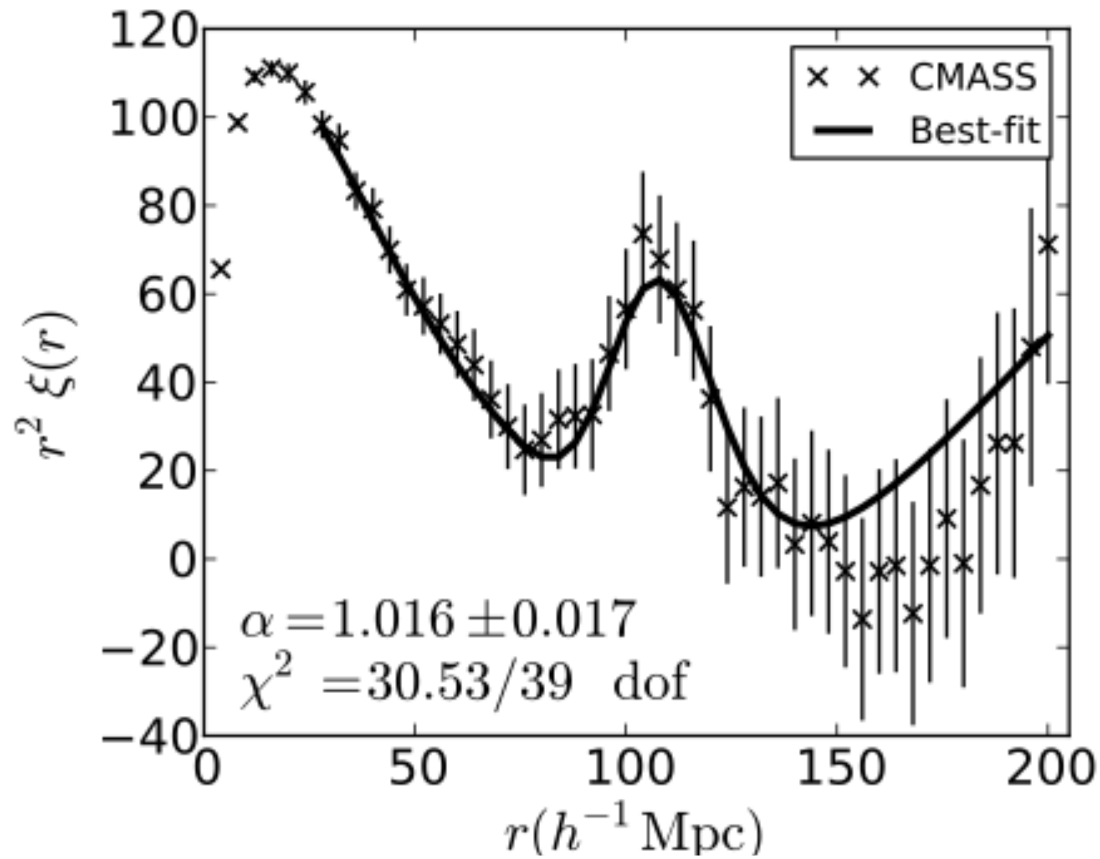
- Public data release:
  - July 2012: 700,000 spectra, 1/3 footprint; + 1.5 million spectra from SDSS 1+2 and 14,000 square degrees of imaging
- Our first BAO analysis uses only the higher redshift portion of the sample (called CMASS)
- 264k galaxies over 3275 sqdeg at median z=0.57



SDSS III

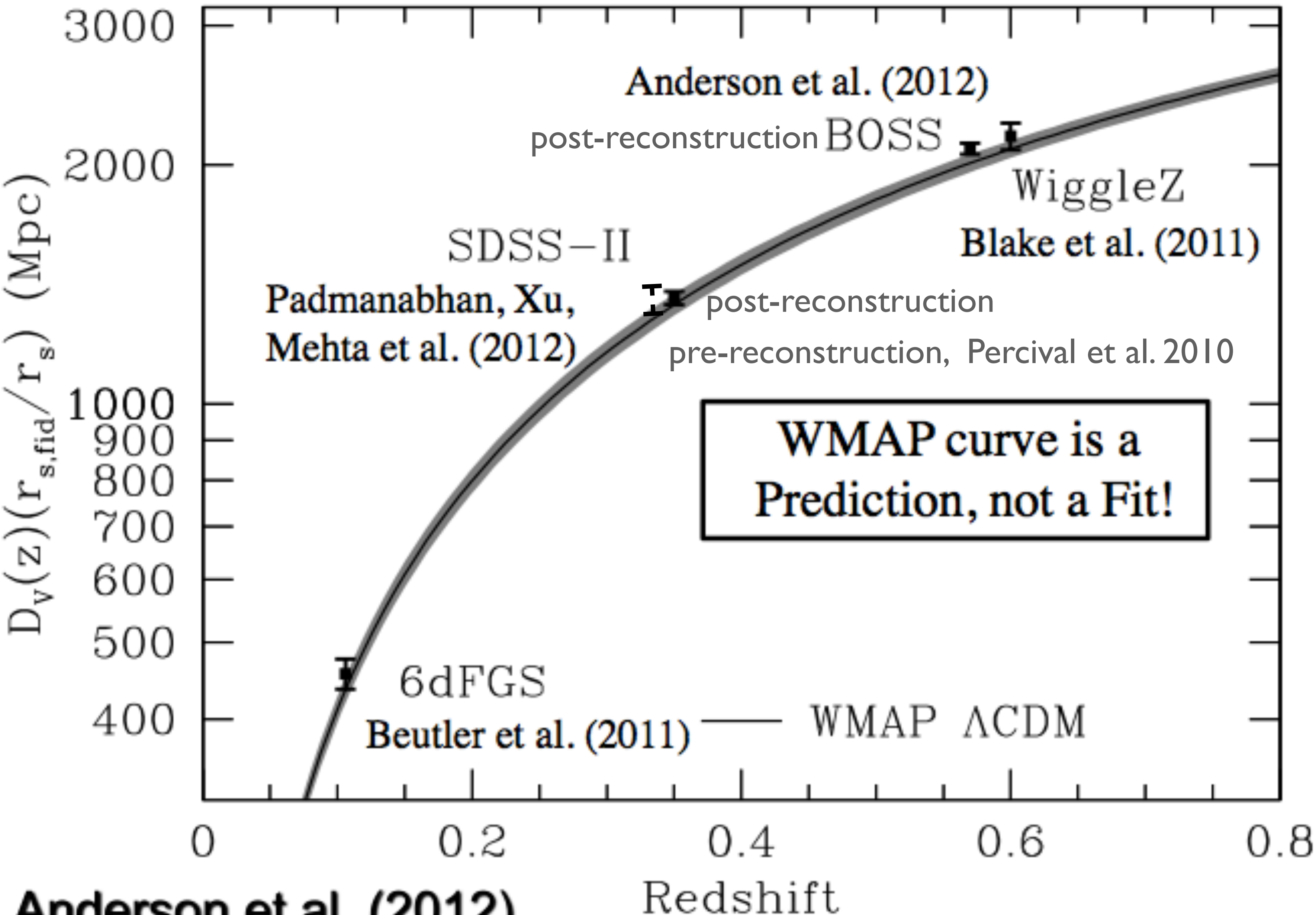
DR9- All Spectroscopy

Extremely clear BAO detection at z=0.57;
measure the distance to z=0.57 with a 1.7% precision



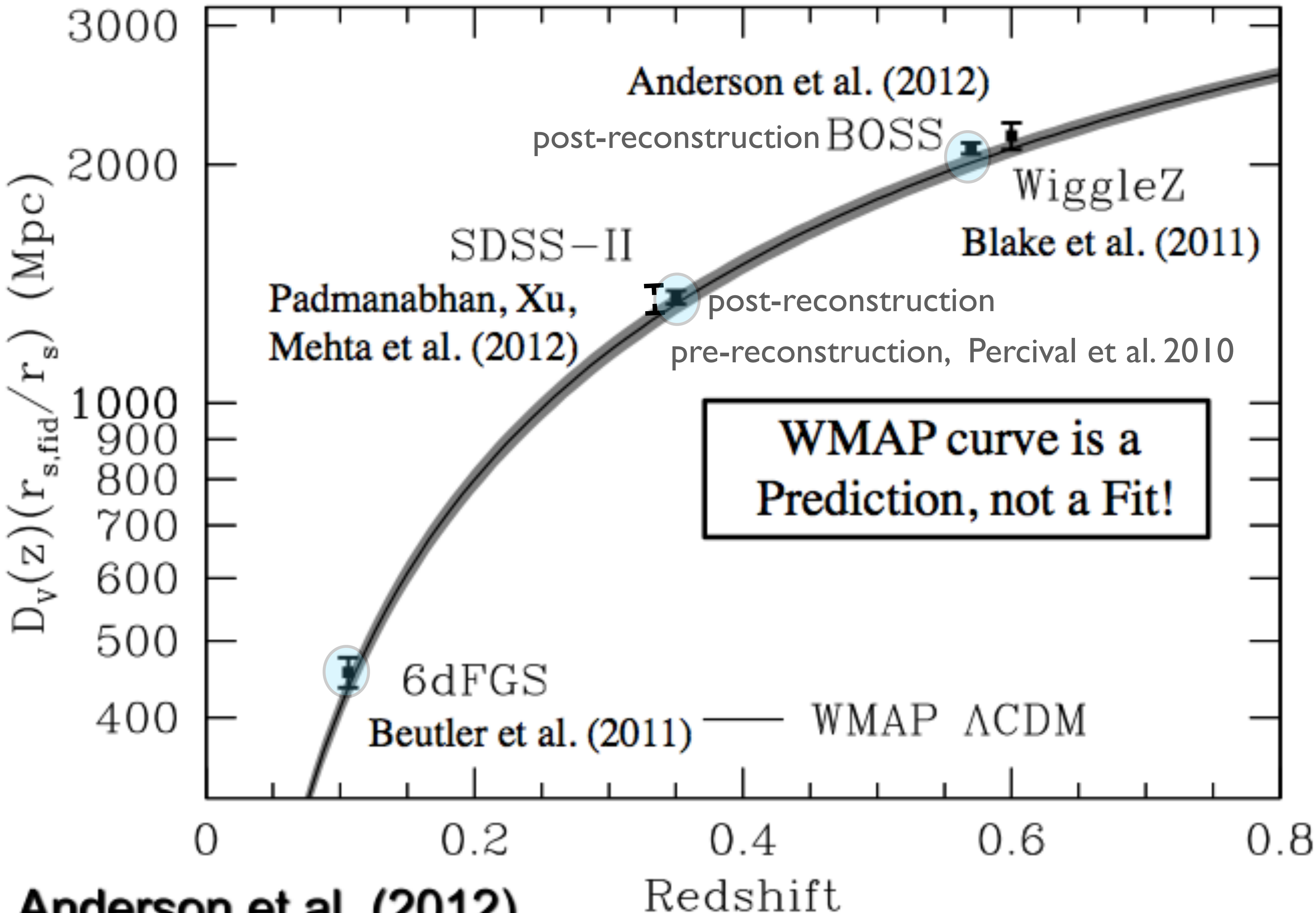
Anderson et al. 2012

BAO Hubble Diagram



Anderson et al. (2012)

BAO Hubble Diagram



Anderson et al. (2012)

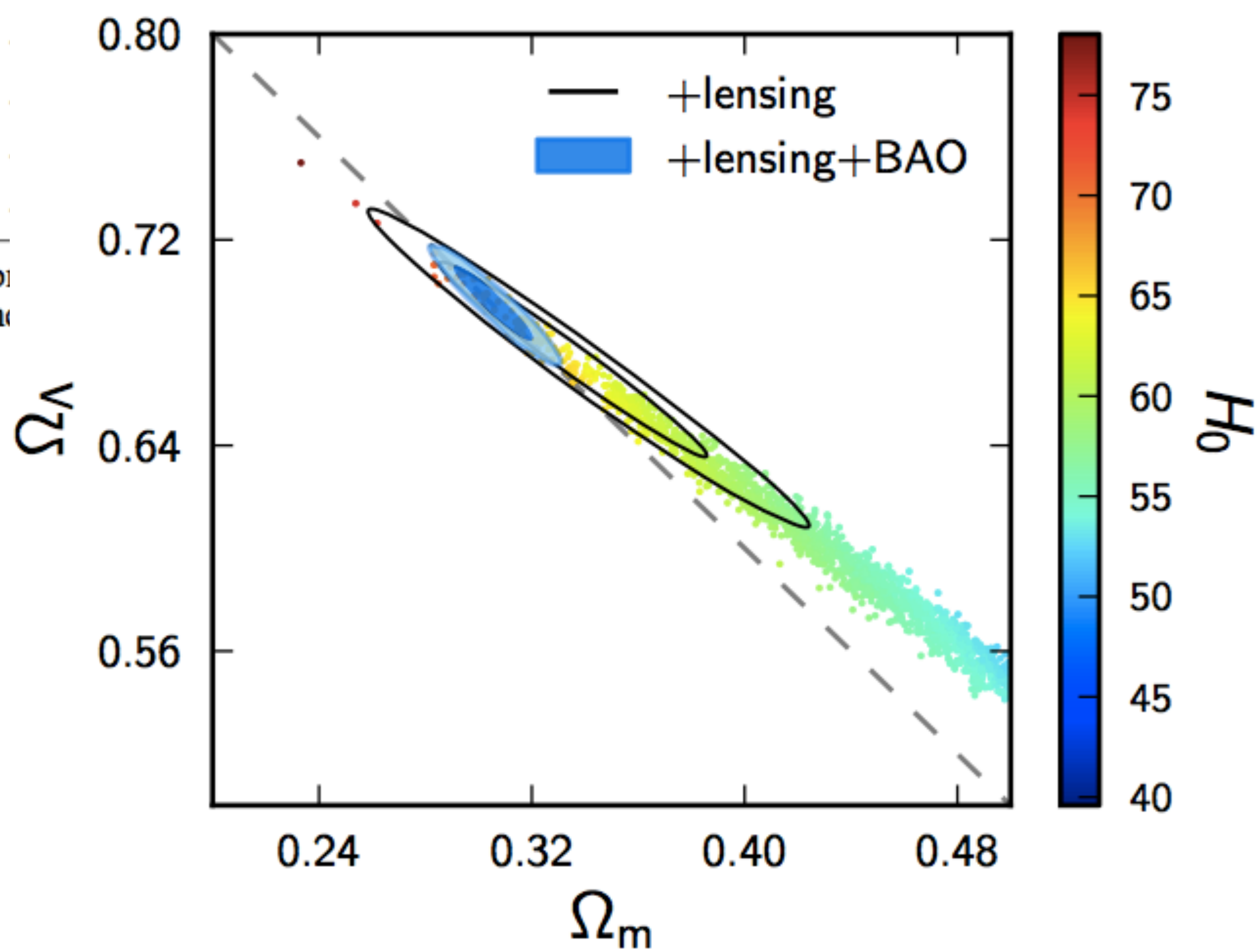
Parameter	*Planck*+WP Best fit	*Planck*+WP 95% limits	*Planck*+WP+BAO Best fit	*Planck*+WP+BAO 95% limits	*Planck*+WP+highL Best fit	*Planck*+WP+highL 95% limits	*Planck*+WP+highL+BAO Best fit	*Planck*+WP+highL+BAO 95% limits
Ω_K	-0.0105	$-0.037^{+0.043}_{-0.049}$	0.0000	$0.0000^{+0.0066}_{-0.0067}$	-0.0111	$-0.042^{+0.043}_{-0.048}$	0.0009	$-0.0005^{+0.0065}_{-0.0066}$
Σm_ν [eV]	0.022	< 0.933	0.002	< 0.247	0.023	< 0.663	0.000	< 0.230
$N_{\rm eff}$	3.08	$3.51^{+0.80}_{-0.74}$	3.08	$3.40^{+0.59}_{-0.57}$	3.23	$3.36^{+0.68}_{-0.64}$	3.22	$3.30^{+0.54}_{-0.51}$
$Y_{\rm P}$	0.2583	$0.283^{+0.045}_{-0.048}$	0.2736	$0.283^{+0.043}_{-0.045}$	0.2612	$0.266^{+0.040}_{-0.042}$	0.2615	$0.267^{+0.038}_{-0.040}$
$dn_s/d\ln k$	-0.0090	$-0.013^{+0.018}_{-0.018}$	-0.0102	$-0.013^{+0.018}_{-0.018}$	-0.0106	$-0.015^{+0.017}_{-0.017}$	-0.0103	$-0.014^{+0.016}_{-0.017}$
$r_{0.002}$	0.000	< 0.120	0.000	< 0.122	0.000	< 0.108	0.000	< 0.111
w	-1.20	$-1.49^{+0.65}_{-0.57}$	-1.076	$-1.13^{+0.24}_{-0.25}$	-1.20	$-1.51^{+0.62}_{-0.53}$	-1.109	$-1.13^{+0.23}_{-0.25}$

Table 10. Constraints on one-parameter extensions to the base ΛCDM model. Data combinations all include *Planck* combined with *WMAP* polarization, and results are shown for combinations with high-ℓ CMB data and BAO. Note that we quote 95% limits here.

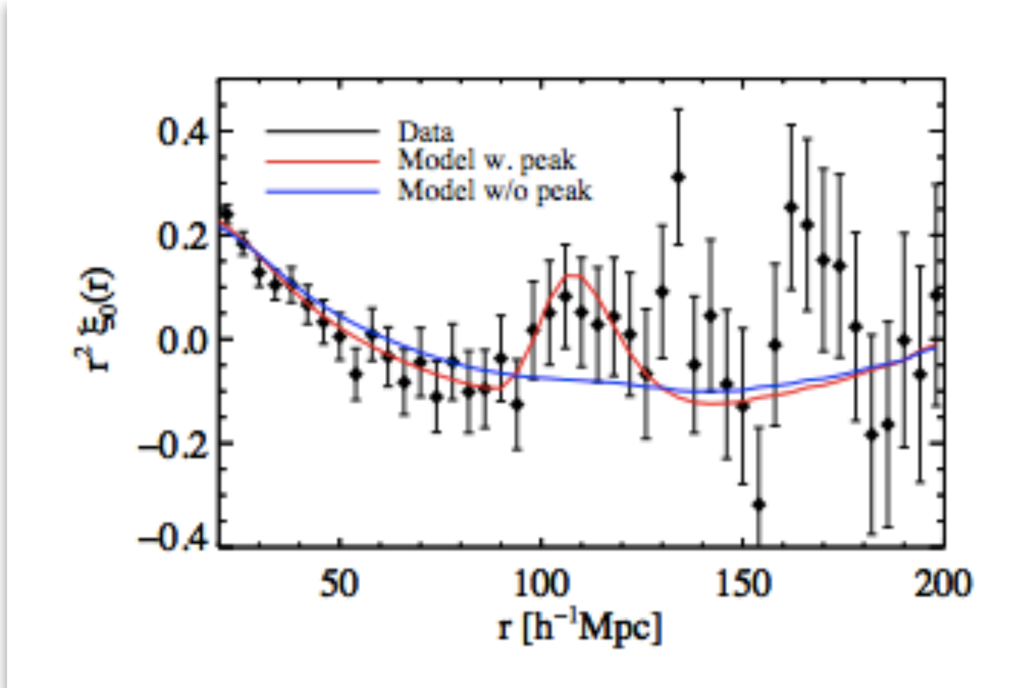
highL = ACT (148,218)+ SPT (95,150,220)
WP= WMAP polarization
BAO= 6dF+SDSS(R)+BOSS

	Planck+WP		*Planck*+WP+BAO		*Planck*+WP+highL		*Planck*+WP+highL+BAO	
Parameter	Best fit	95% limits	Best fit	95% limits	Best fit	95% limits	Best fit	95% limits
Ω_K	−0.0105	$-0.037^{+0.043}_{-0.049}$	0.0000	$0.0000^{+0.0066}_{-0.0067}$	−0.0111	$-0.042^{+0.043}_{-0.048}$	0.0009	$-0.0005^{+0.0065}_{-0.0066}$
Σm_ν [eV]	0.022	< 0.933	0.002	< 0.247	0.023	< 0.663	0.000	< 0.230
$N_{\rm eff}$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$3.30^{+0.54}_{-0.51}$
$Y_{\rm P}$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$[illegible].267^{+0.038}_{-0.040}$
$dn_{\rm s}/d\ln k$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$[illegible]0.014^{+0.016}_{-0.017}$
$r_{0.002}$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	< 0.111
w	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$-1.13^{+0.23}_{-0.25}$

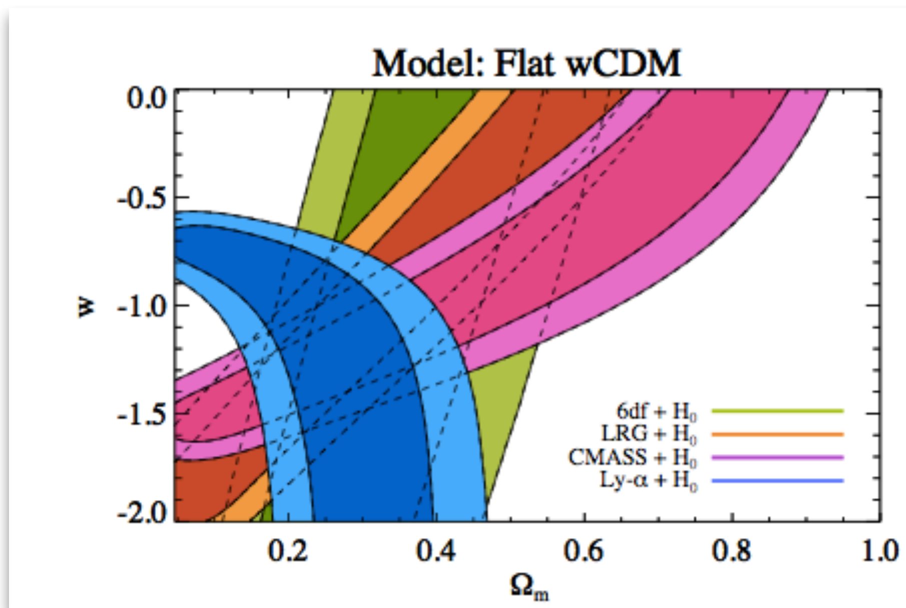
Table 10. Constraints o[illegible] *[illegible]anck* combined with *WMAP* polarization, an[illegible] [illegible]ote 95% limits here.



- See Anze Slosar's talk!
- BAO in Lyman-alpha forest
- First BAO analysis at z>2



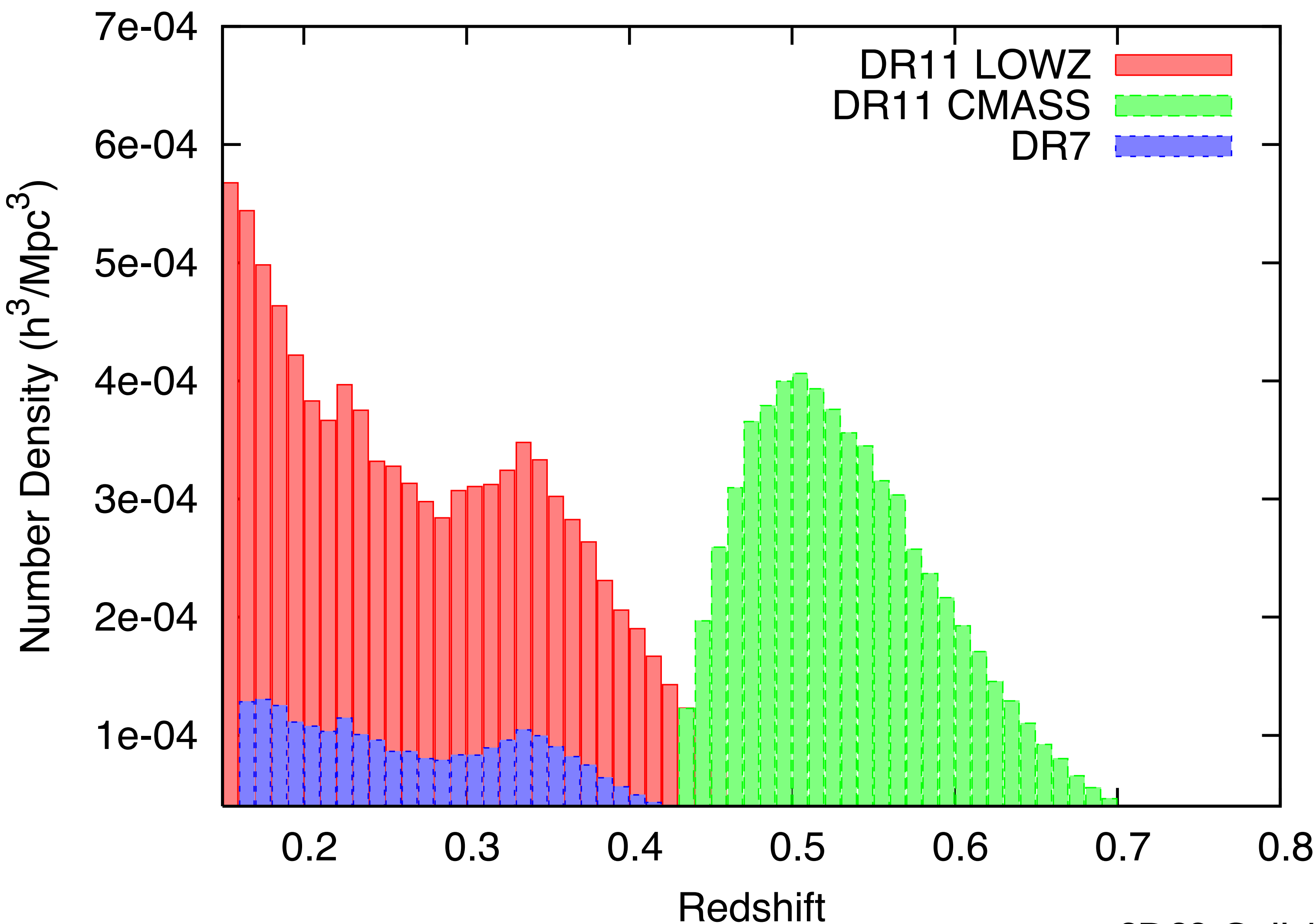
Detected BAO using Lyman alpha forest at z=2.3



with NO CMB at all!!

SDSS III

What about after DR9 ?



SDSS Collaboration 2013

SDSS III
Data Release 9-12 Spectroscopy

~30K high-z galaxy spectra ~60K 1million

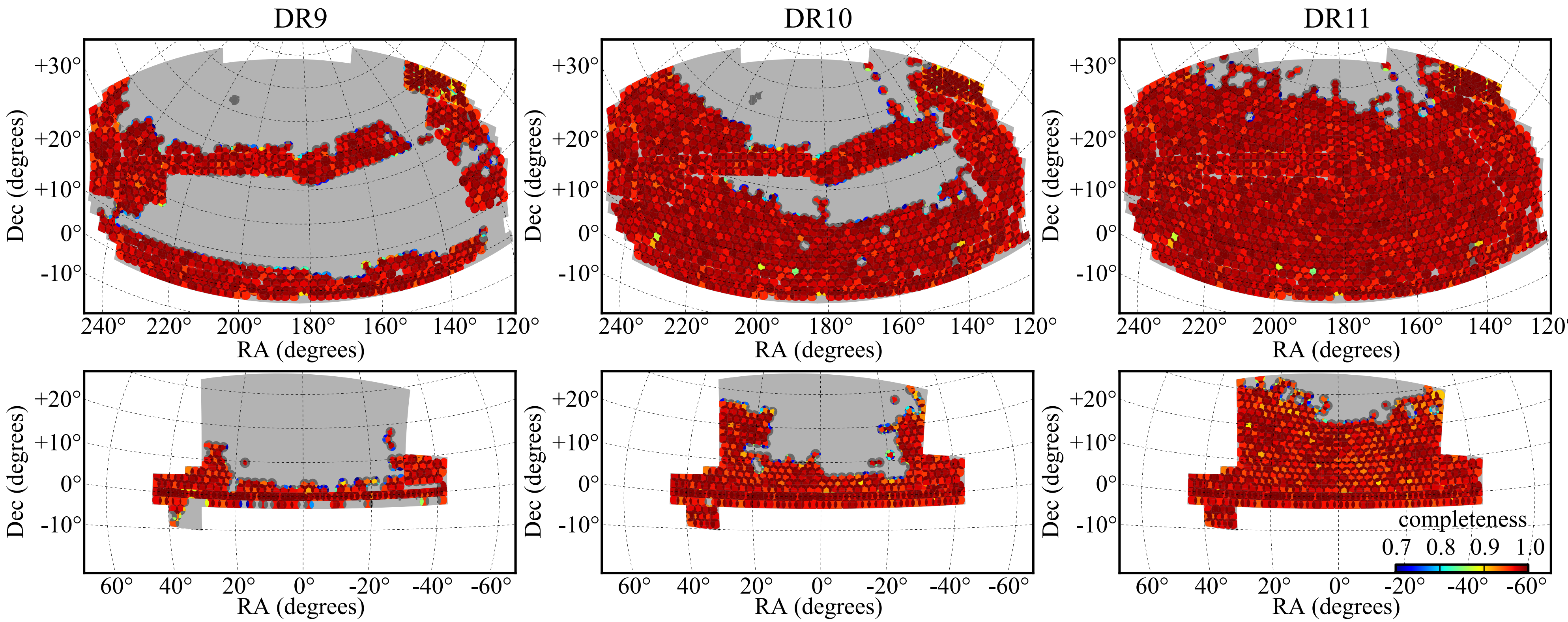
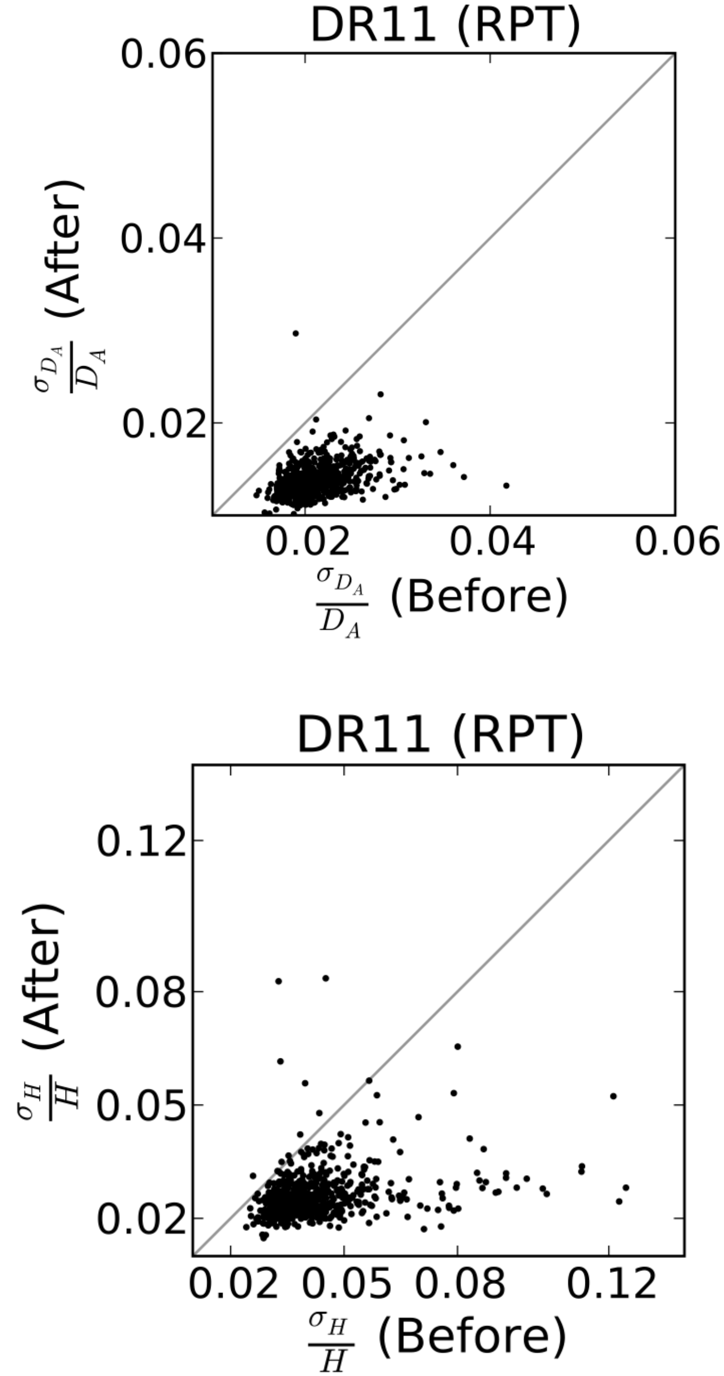


Figure 1. Evolution of the BOSS sky coverage from DR9 to DR11.

SDSS Collaboration (in prep)

- In DR9, reconstruction didn't improve our signal-to-noise of BAO. This is probably due to poor survey window (too much area near a boundary).
- DR11 has a filled geometry. Reconstruction improves mocks dramatically.
- Average Constraints (mocks) on D_A goes from 2.1% to 1.5%.
- Average Constraints (mocks) on H(z) goes from 4.4% to 2.7%
- Expects BAO >8 sigma
- Stay Tuned!



Magana-Vargas, Ho, Xu et al. (in prep)
SDSS Collaboration (in prep)

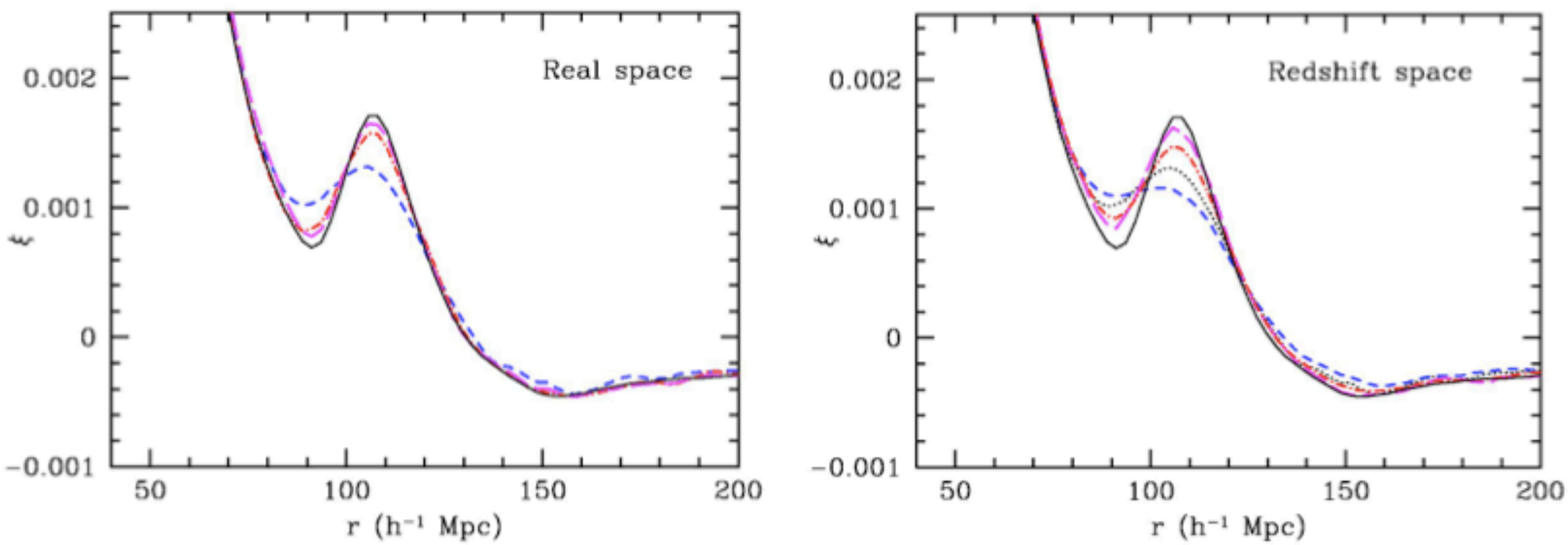
Outline

- SDSS III- BOSS :
  - Mini-summary and looking forward
- BAO Reconstruction

You can reconstruct to significantly different large scale structure, but still get the same/similar BAO post-reconstruction!

Bringing in Reconstruction ... (Going back in time)

- Most of the non-linear degradation is due to large scale flows. These are produced by the same large scale structure that we are measuring for the BAO signature.
- Map of galaxies tells us where the mass is that sources the gravitational forces that created the bulk flows
- Can run this backwards and undo most non-linearity
- Restore the statistical precision available per unit volume



Eisenstein, Seo, Sirko, Spergel, 2007

Reconstruction: general idea

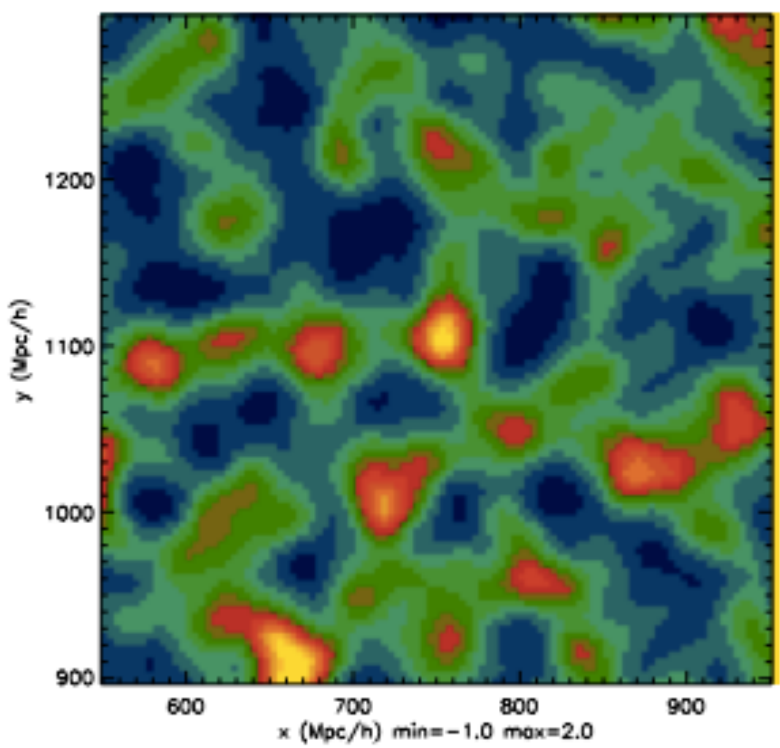
- Smooth the density field to filter out high k modes, which are difficult to model.
- Compute the negative Zel'dovich displacement, s, from the smoothed density field: $\mathrm{s}(\mathbf{k}) = -i(\mathbf{k}/k^2)\delta(\mathbf{k})\mathcal{S}(k)$, where $\mathcal{S}$ is the smoothing kernel (see below).
- Shift the original particles by s and compute the "displaced" density field, δ_d.
- Shift an initially spatially uniform distribution of particles by s to form the "shifted" density field, δ_s.
- The reconstructed density field is defined as $\delta_r \equiv \delta_d - \delta_s$ with power spectrum $P_r(k) \propto \langle |\delta_r^2| \rangle$.

Seo, Eisenstein, Sirki & Spergel 2008

Noh, White & Padmanabhan 2009

Reconstruction: general idea

- Smooth the density field to filter out high k modes, which are difficult to model.



Seo, Eisenstein, Sirki & Spergel 2008

Noh, White & Padmanabhan 2009

Reconstruction: general idea

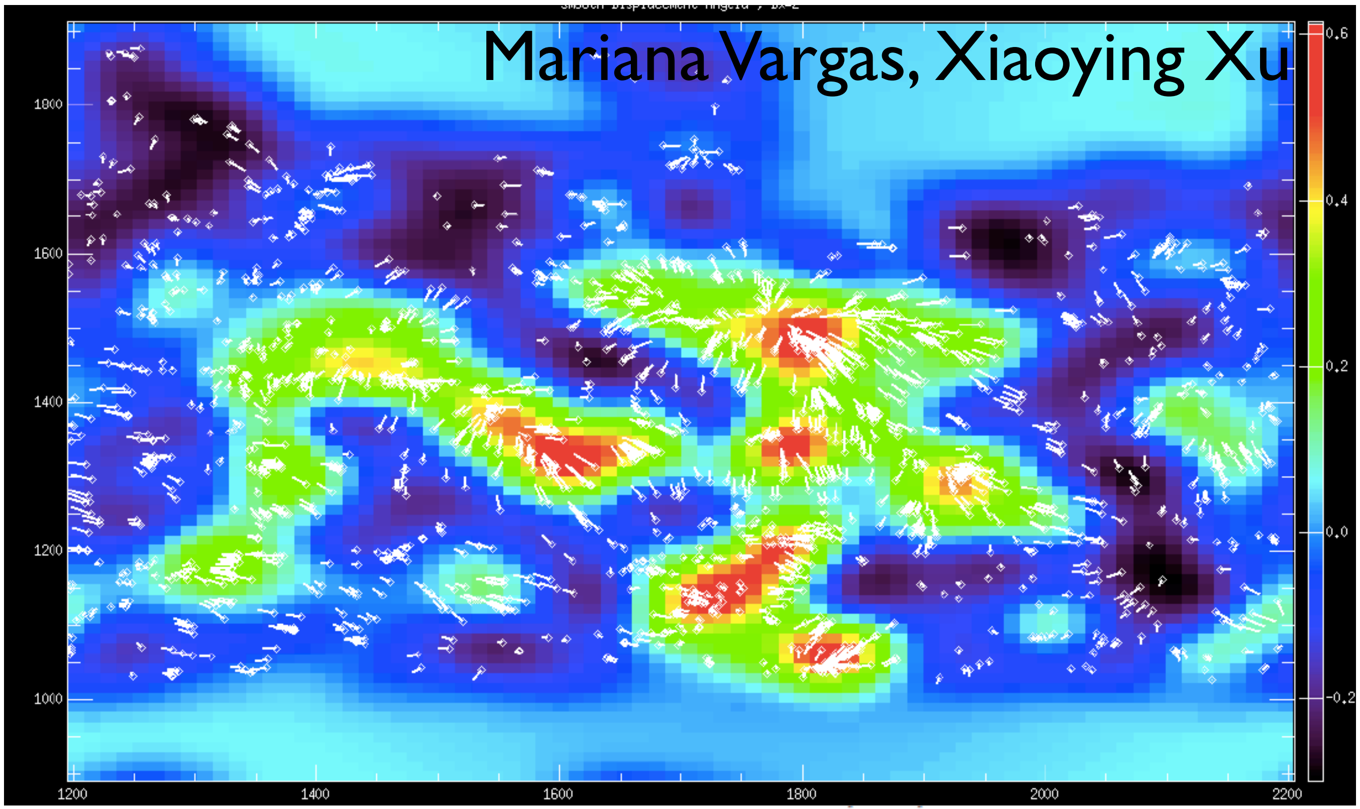
- Smooth the density field to filter out high k modes, which are difficult to model.
- Compute the negative Zel'dovich displacement, $\mathbf{s}$, from the smoothed density field: $\mathbf{s}(\mathbf{k}) = -i(\mathbf{k}/k^2)\delta(\mathbf{k})\mathcal{S}(k)$, where $\mathcal{S}$ is the smoothing kernel (see below). $\mathcal{S}(k) = e^{-(kR)^2/2}$.
- Shift the original particles by s and compute the "displaced" density field, δ_d.
- Shift an initially spatially uniform distribution of particles by s to form the "shifted" density field, δ_s.
- The reconstructed density field is defined as $\delta_r \equiv \delta_d - \delta_s$ with power spectrum $P_r(k) \propto \langle|\delta_r^2|\rangle$.

Seo, Eisenstein, Sirki & Spergel 2008

Noh, White & Padmanabhan 2009

Reconstruction: general idea

- Smooth the density field to filter out high k modes, which are difficult to model.



contours: density field

white lines: displacement field

Seo, Eisenstein, Sirki & Spergel 2008

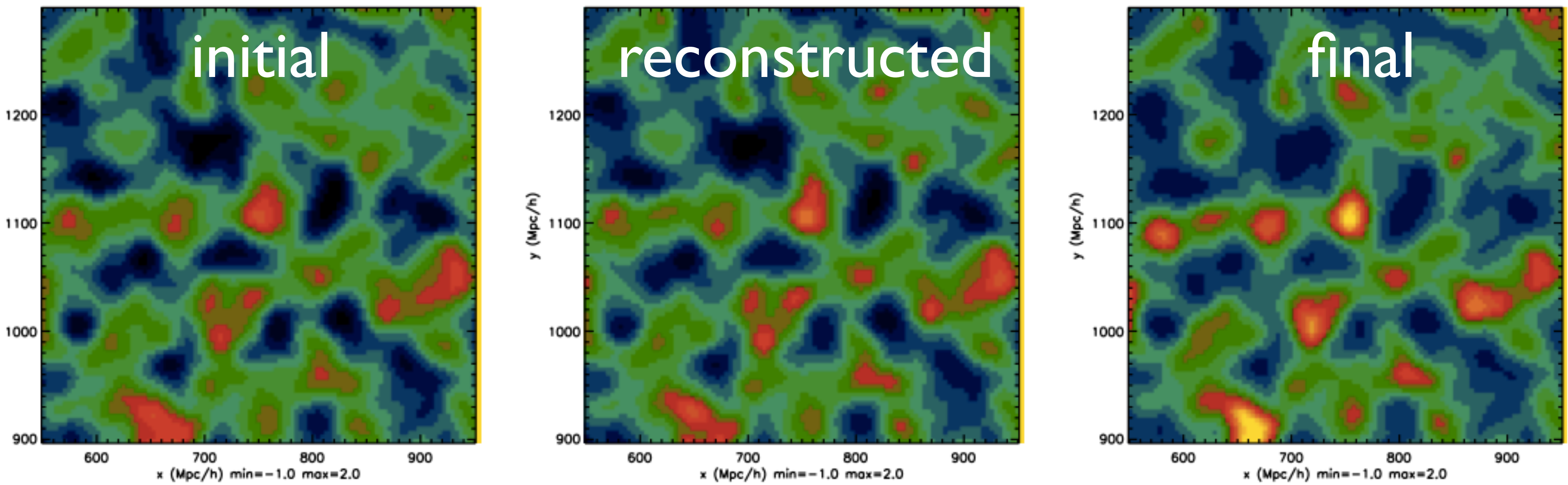
Noh, White & Padmanabhan 2009

Reconstruction: general idea

- Smooth the density field to filter out high k modes, which are difficult to model.
- Compute the negative Zel'dovich displacement, $\mathbf{s}$, from the smoothed density field: $\mathbf{s}(\mathbf{k}) = -i(\mathbf{k}/k^2)\delta(\mathbf{k})\mathcal{S}(k)$, where $\mathcal{S}$ is the smoothing kernel (see below).
- Shift the original particles by $\mathbf{s}$ and compute the "displaced" density field, δ_d.
- Shift an initially spatially uniform distribution of particles by $\mathbf{s}$ to form the "shifted" density field, δ_s.
- The reconstructed density field is defined as $\delta_r \equiv \delta_d - \delta_s$ with power spectrum $P_r(k) \propto \langle|\delta_r^2|\rangle$.

Seo, Eisenstein, Sirki & Spergel 2008

Noh, White & Padmanabhan 2009



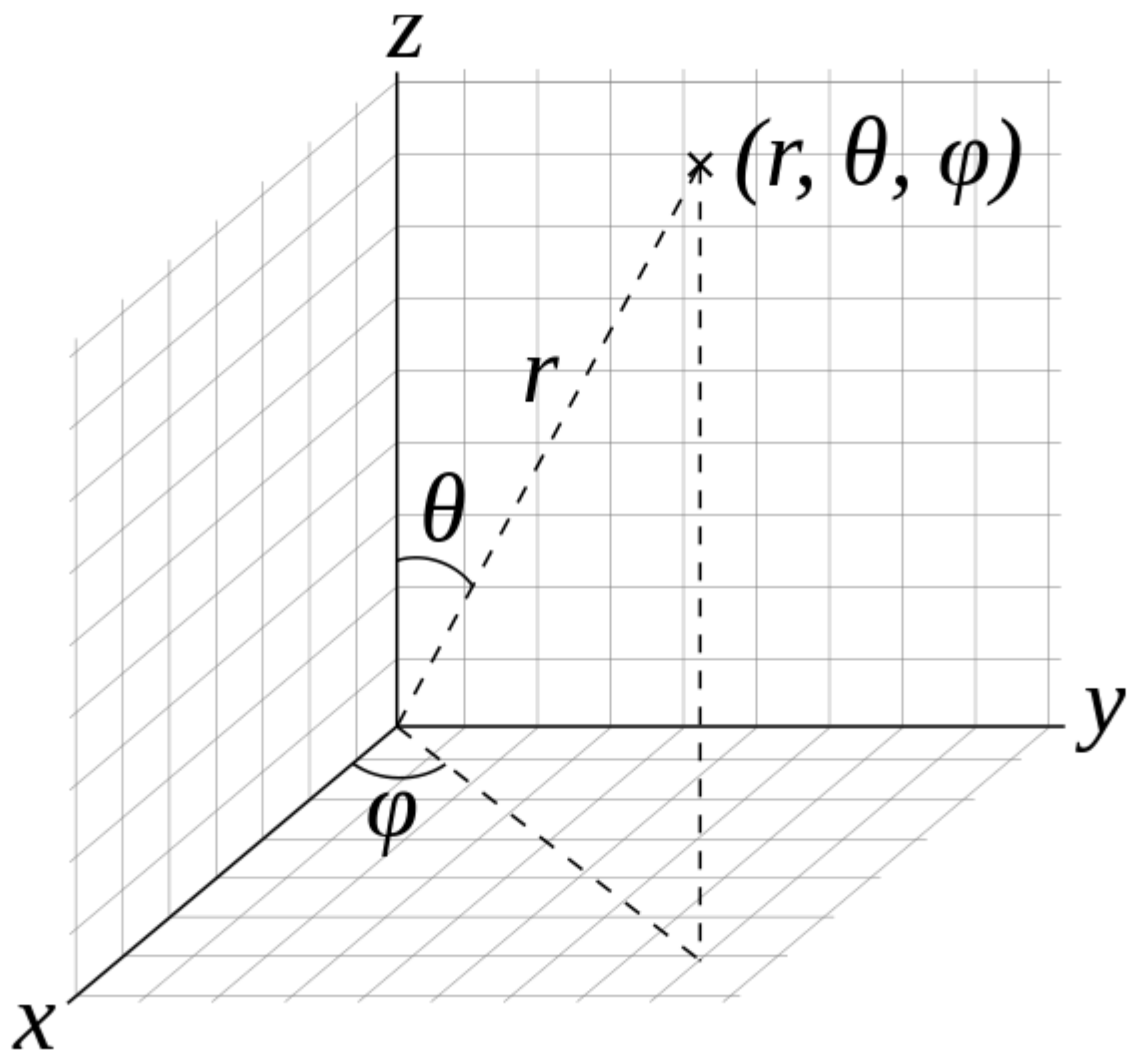
- Shift the original particles by s and compute the "displaced" density field, δ_d.
- Shift an initially spatially uniform distribution of particles by s to form the "shifted" density field, δ_s.
- The reconstructed density field is defined as $\delta_r \equiv \delta_d - \delta_s$ with power spectrum $P_r(k) \propto \langle|\delta_r^2|\rangle$.

Seo, Eisenstein, Sirki & Spergel 2008

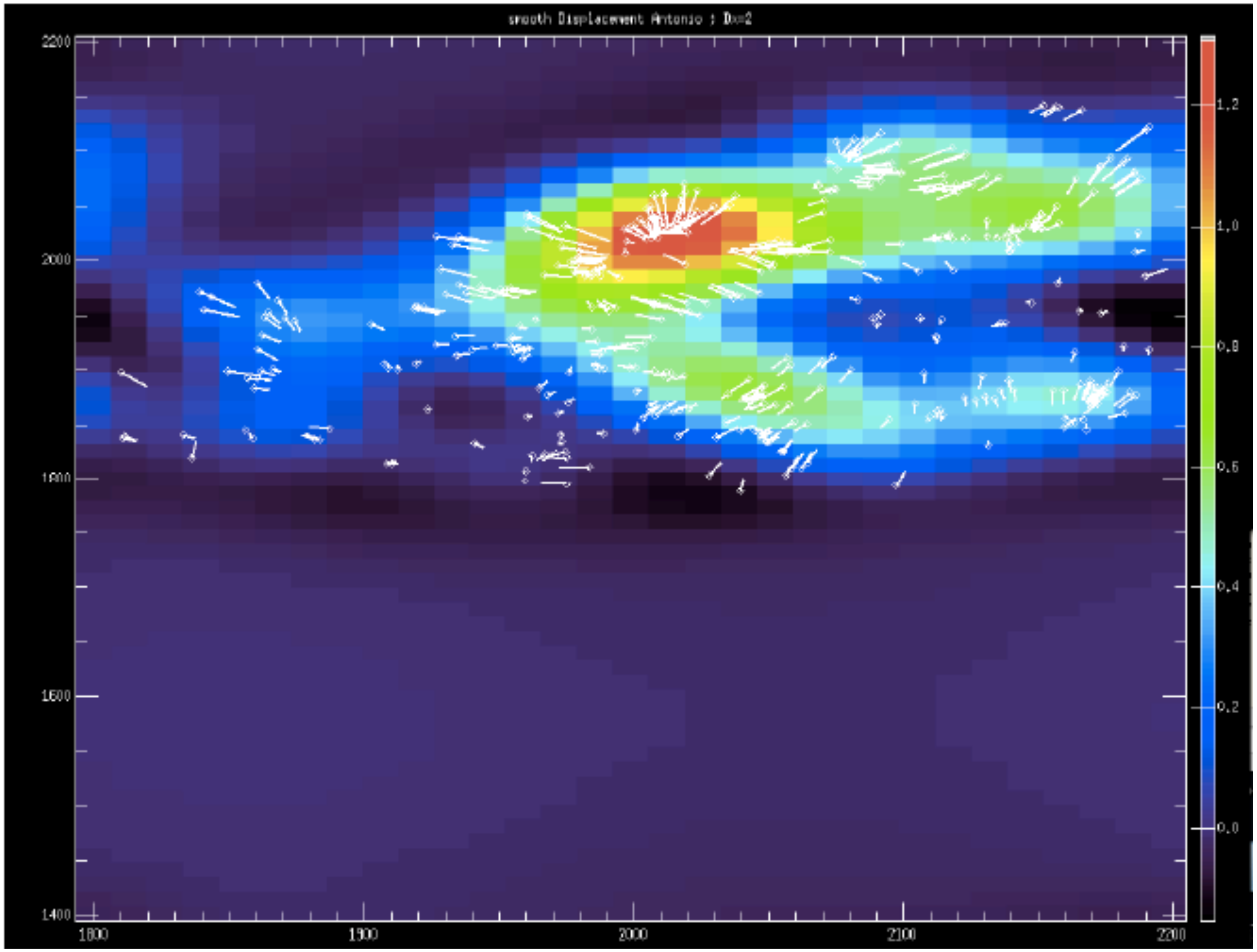
Noh, White & Padmanabhan 2009

As simple as the algorithm sounds like...
three different codes following similar methodology generate different displacement field given the same input.

Just reminding us the coordinate system



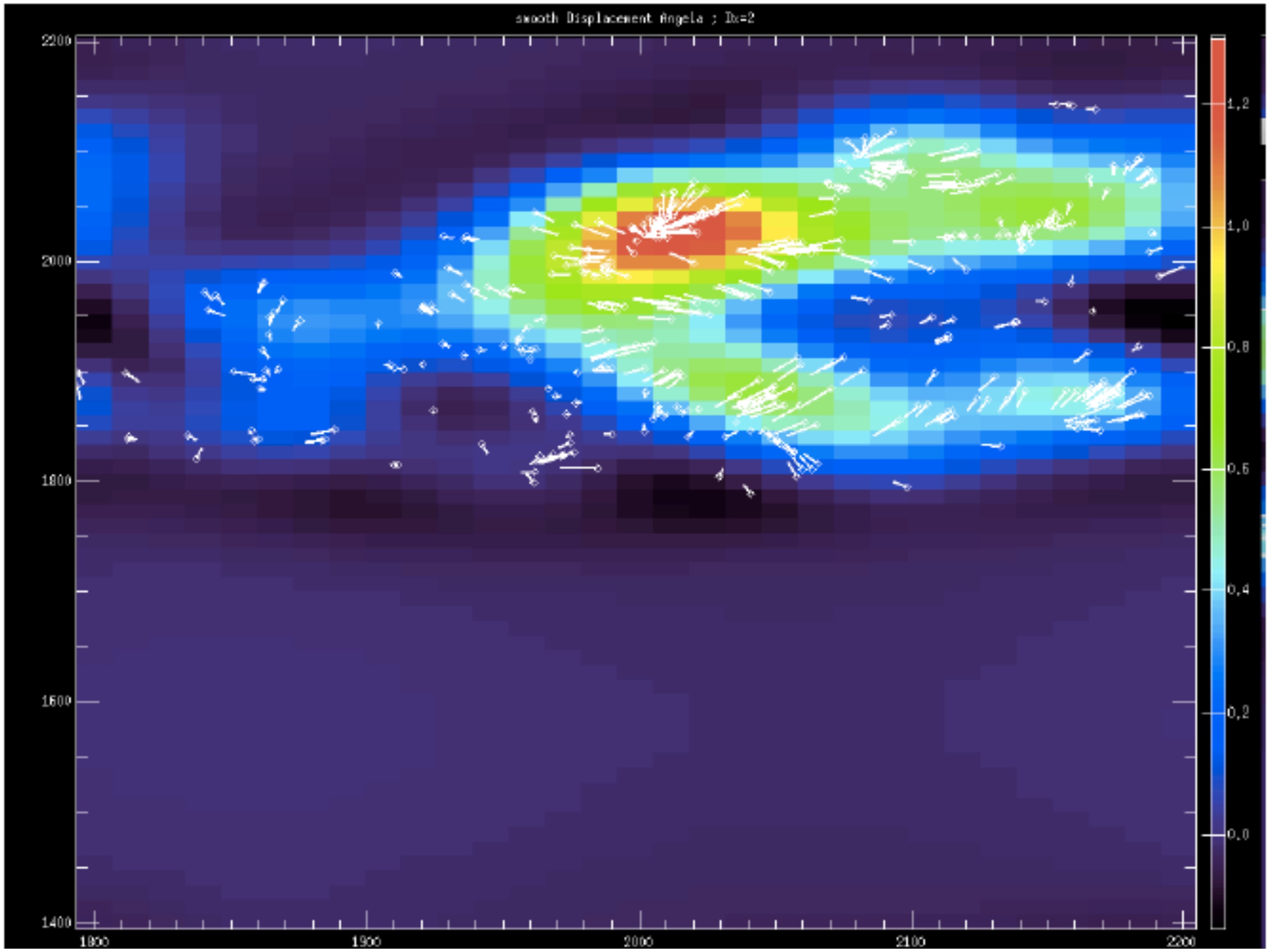
Displacement in r



Method 1

Magana-Vargas, SH , Xu + BOSS Clustering WG (in prep)

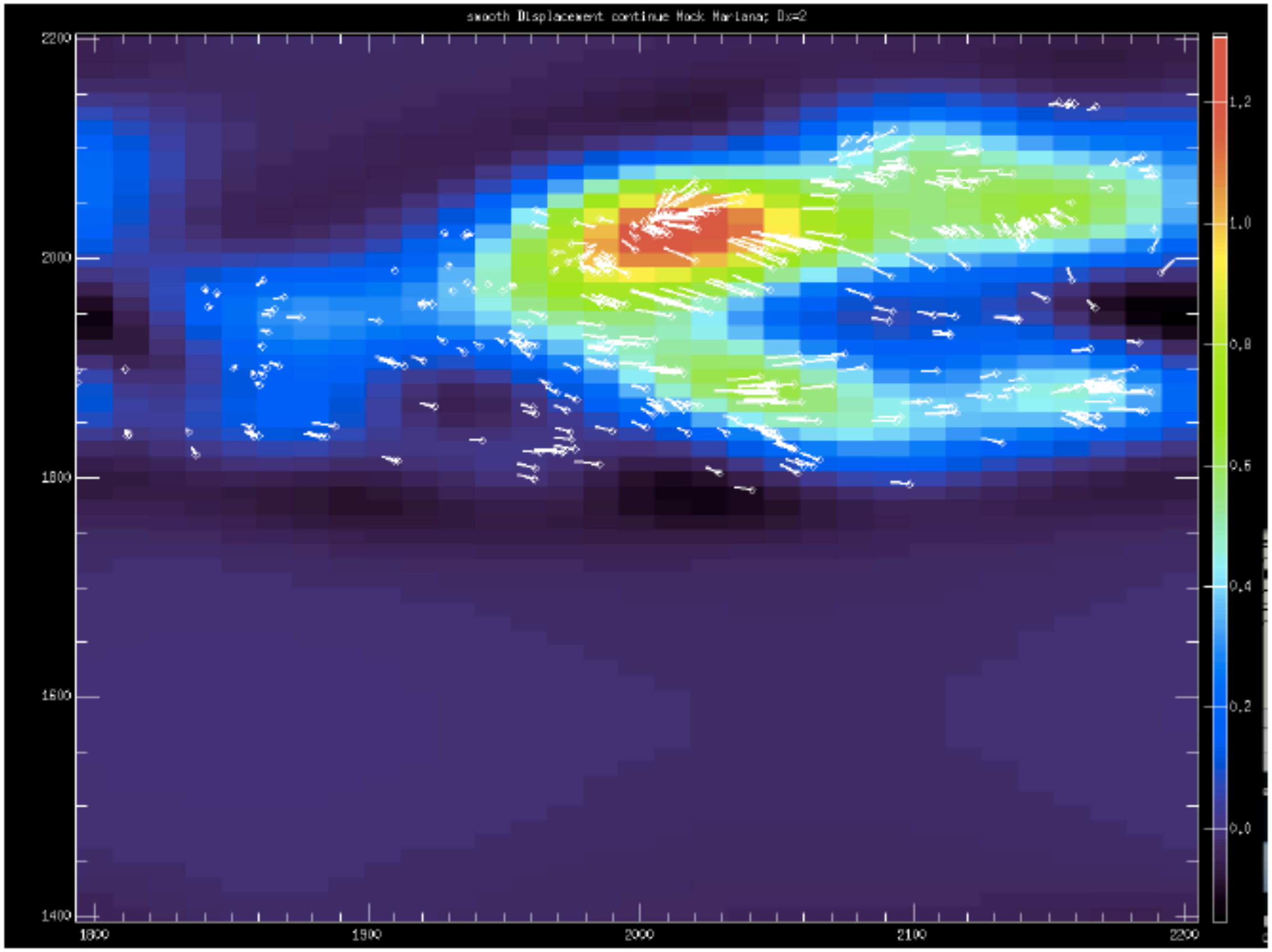
Displacement in r



Method 2

Magana-Vargas, SH , Xu + BOSS Clustering WG (in prep)

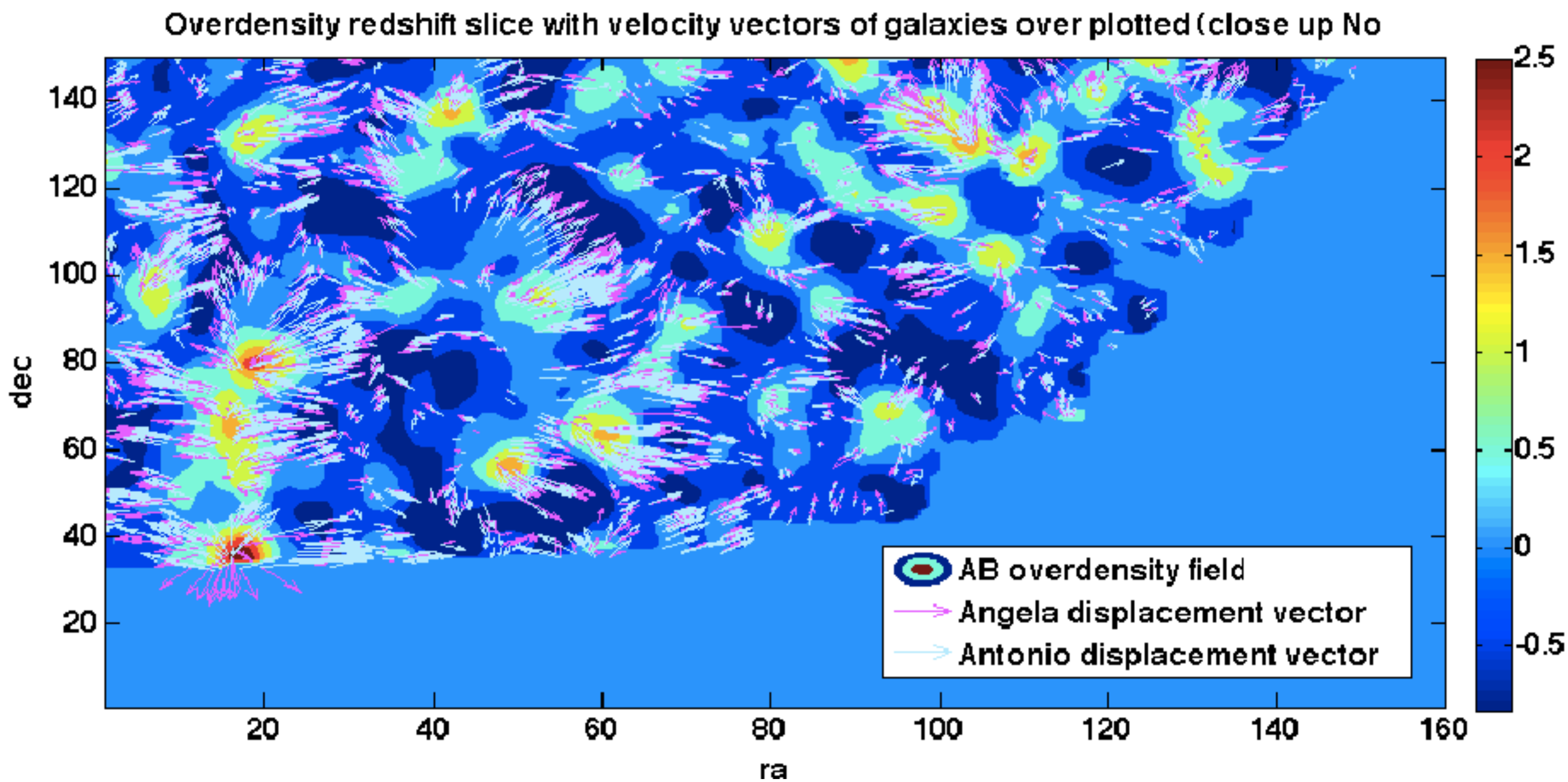
Displacement in r



Method 3

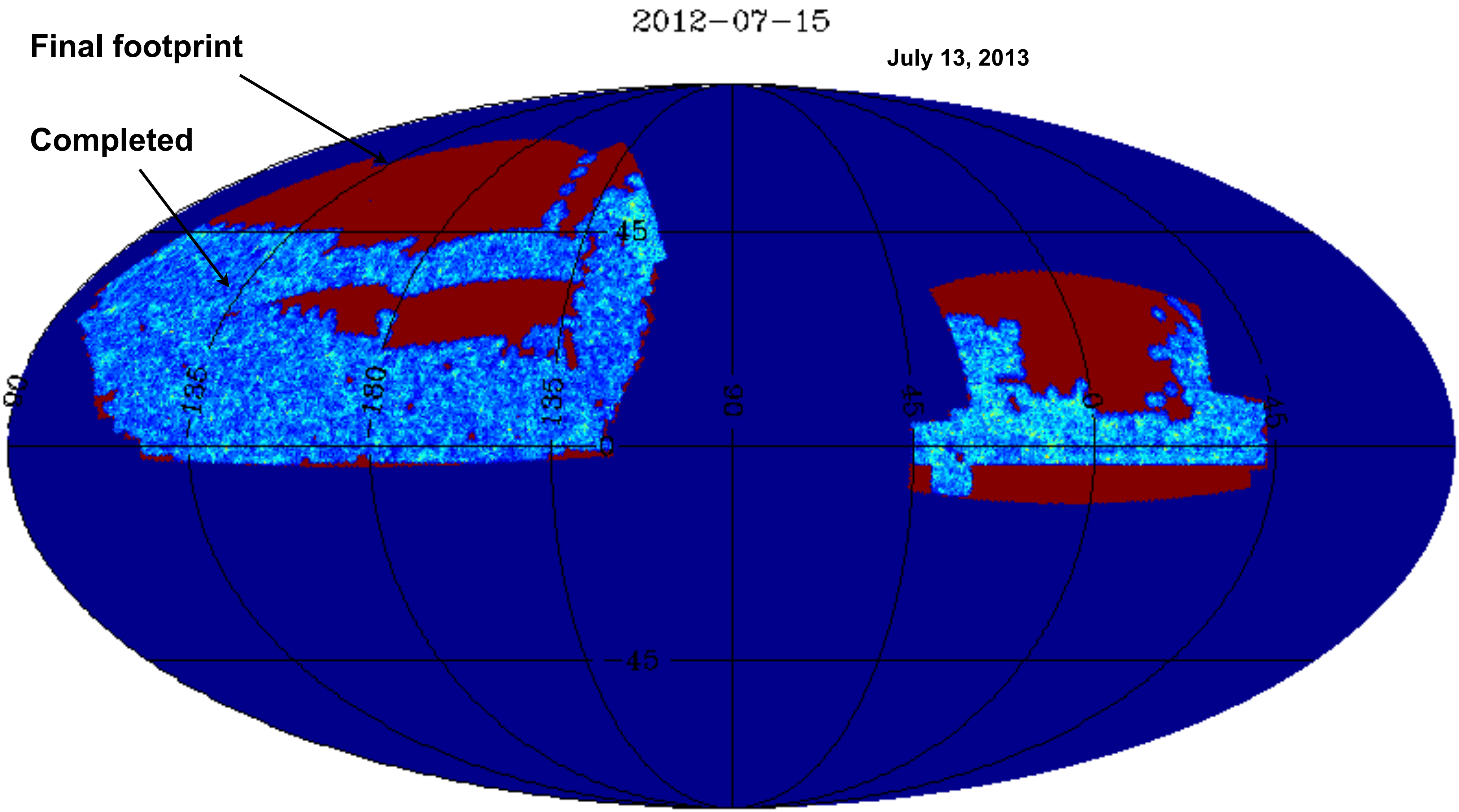
Magana-Vargas, SH , Xu + BOSS Clustering WG (in prep)

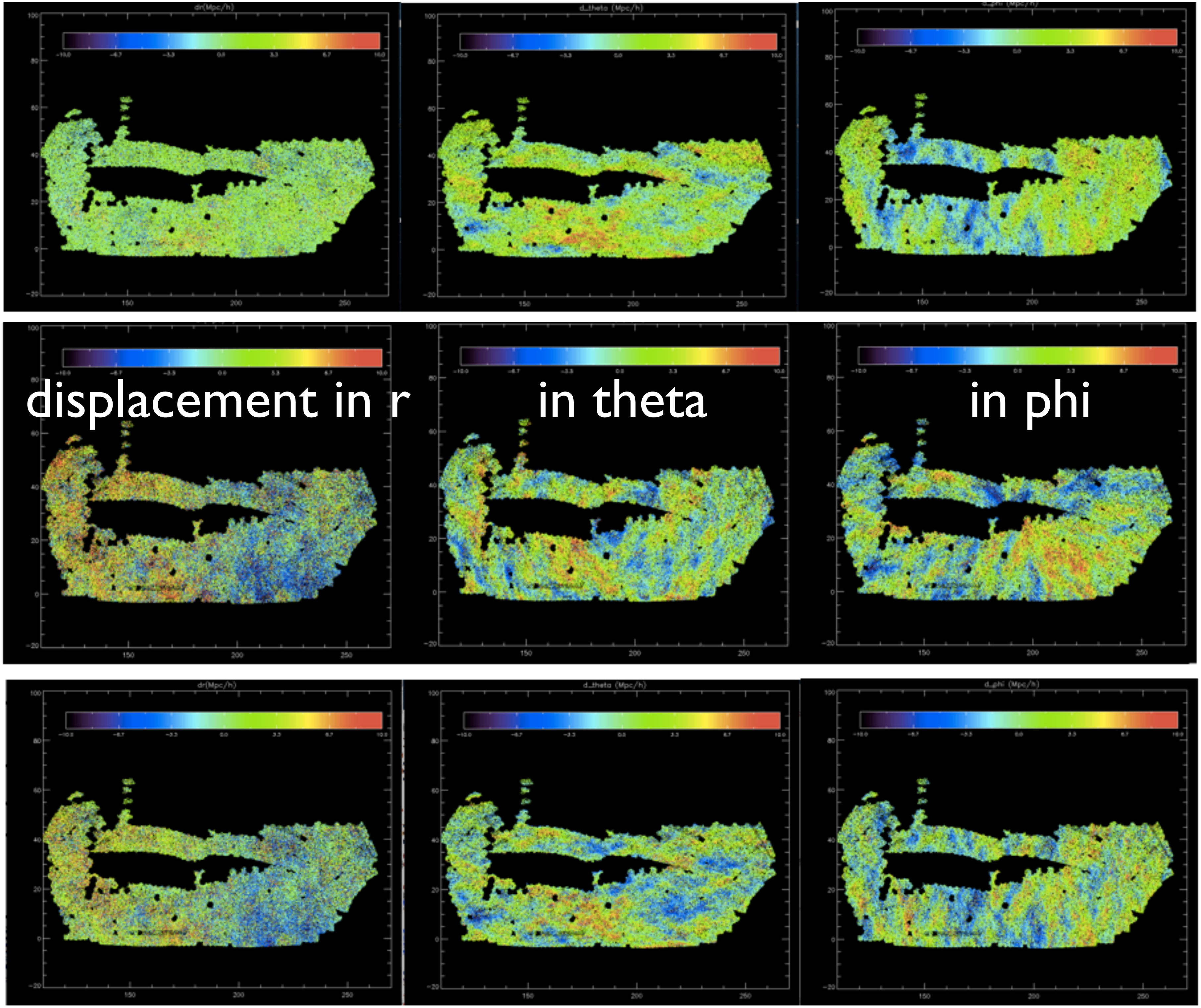
Effects of different reconstruction method choices



Courtesy plots from Angela Burden

How about applying 3 methods to the same BOSS dataset?



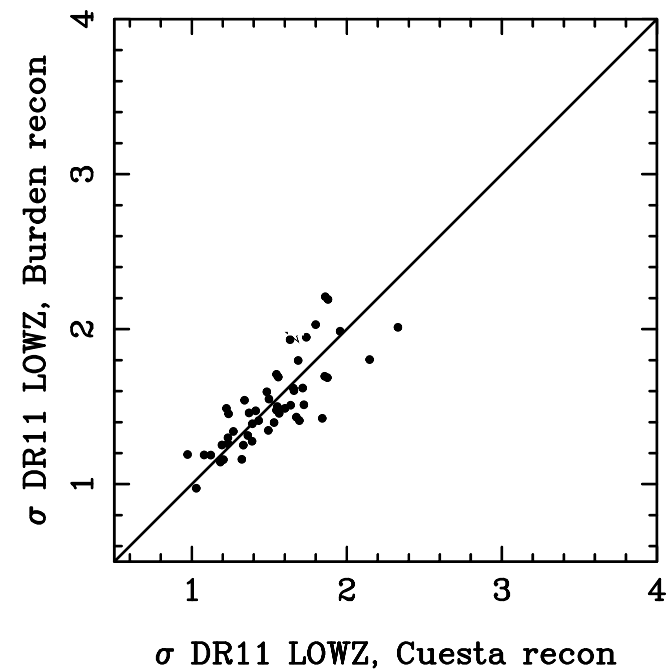

displacement in r
in theta
in phi

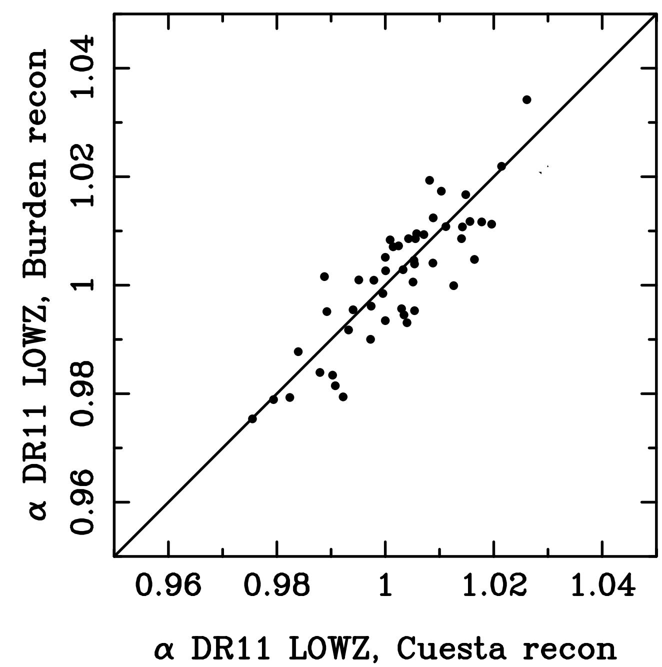
Test results from two methods on 50 mocks

$$\bar{\alpha} = 1.00289 \pm 0.00161$$

$$\bar{\alpha} = 1.00169 \pm 0.00172$$

Courtesy plots from Will Percival





- Conclusions?
- All methods are relatively unbiased.
- The scatter plots show that they correlate with each other very well.
- But there is a scatter between the methods

Outline

- SDSS III- BOSS :
  - Mini-summary and looking forward
- BAO isotropic and anisotropic fitting
- BAO Reconstruction

You can reconstruct to significantly different large scale structure, but still get the same/similar BAO post-reconstruction!
Now we need to understand and include the systematic errors we are introducing when using reconstruction.

Outline

- SDSS III- BOSS :
  - Mini-summary and looking forward
- BAO isotropic and anisotropic fitting
- BAO Reconstruction
- Ultra-fast simulations of the Universe?
- BAO in cross-correlations?

Aim: Efficiently make lots of mocks relatively accurately. BOSS galaxy clustering working group used ~600 mocks using PTHalos method. Can we do better?

Aim: Efficiently make lots of mocks relatively accurately. BOSS galaxy clustering working group used ~600 mocks using PTHalos method. Can we do better?

New Direction: Use a non-parametric Machine learning algorithm that does distribution-to-distribution regression to learn the non-linear evolution of the Universe.

What the heck is Machine Learning?

- Machine Learning algorithms learns trends from the data itself, it does not impose pre-assumed models on the data.
- The advantage of ML is that it is fully non-parametric:
  - The only assumption necessary is that some relationship **does** exist between halo properties (features) and the number of galaxies that will reside in it and that this relationship is continuous.

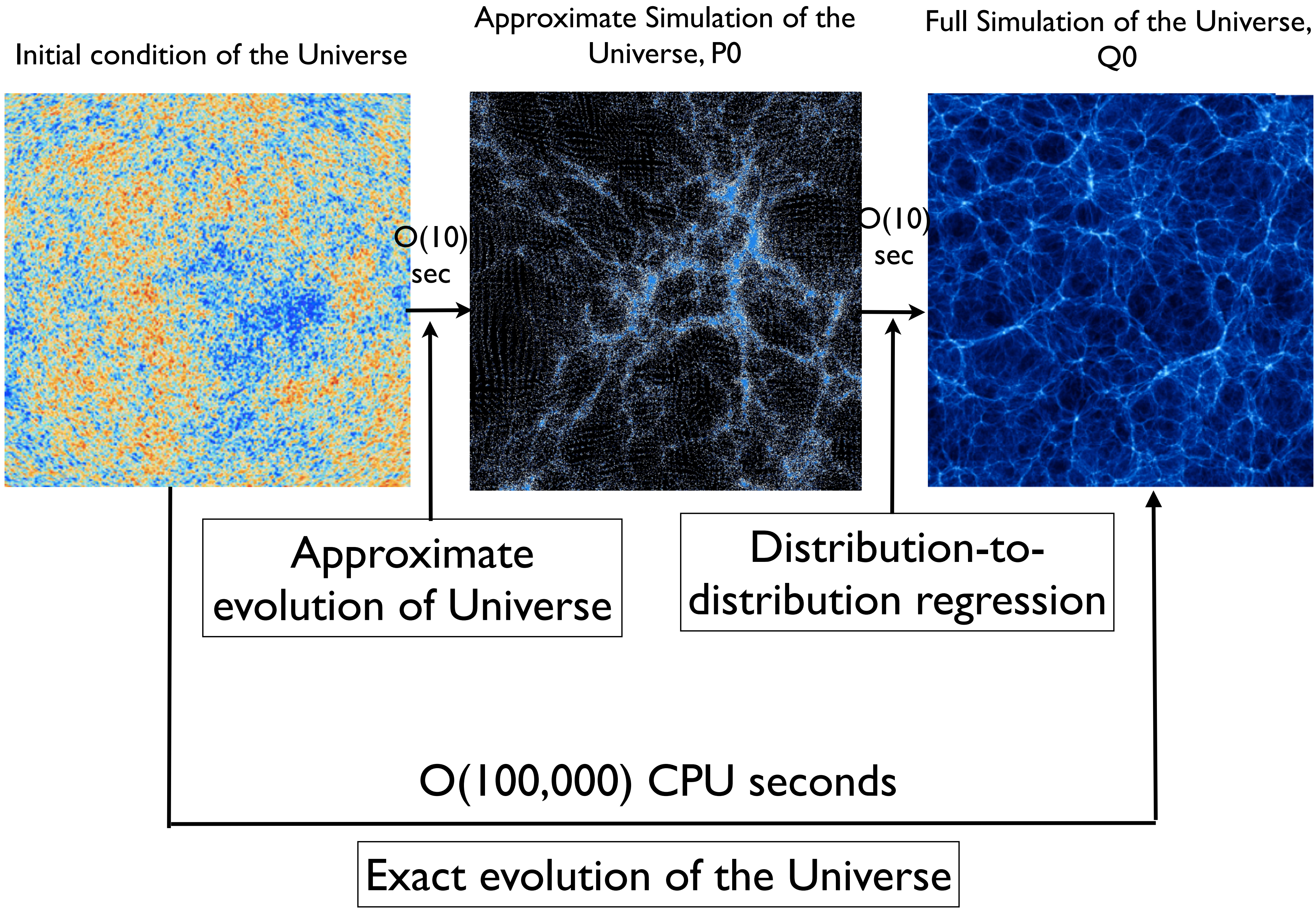
Cool Examples of ML applications (Courtesy Slide from Kayvon Fatahalian)

[Shrivastava 2011]
"Find images that are similar to a query image (even if not similar in individual pixel values)."

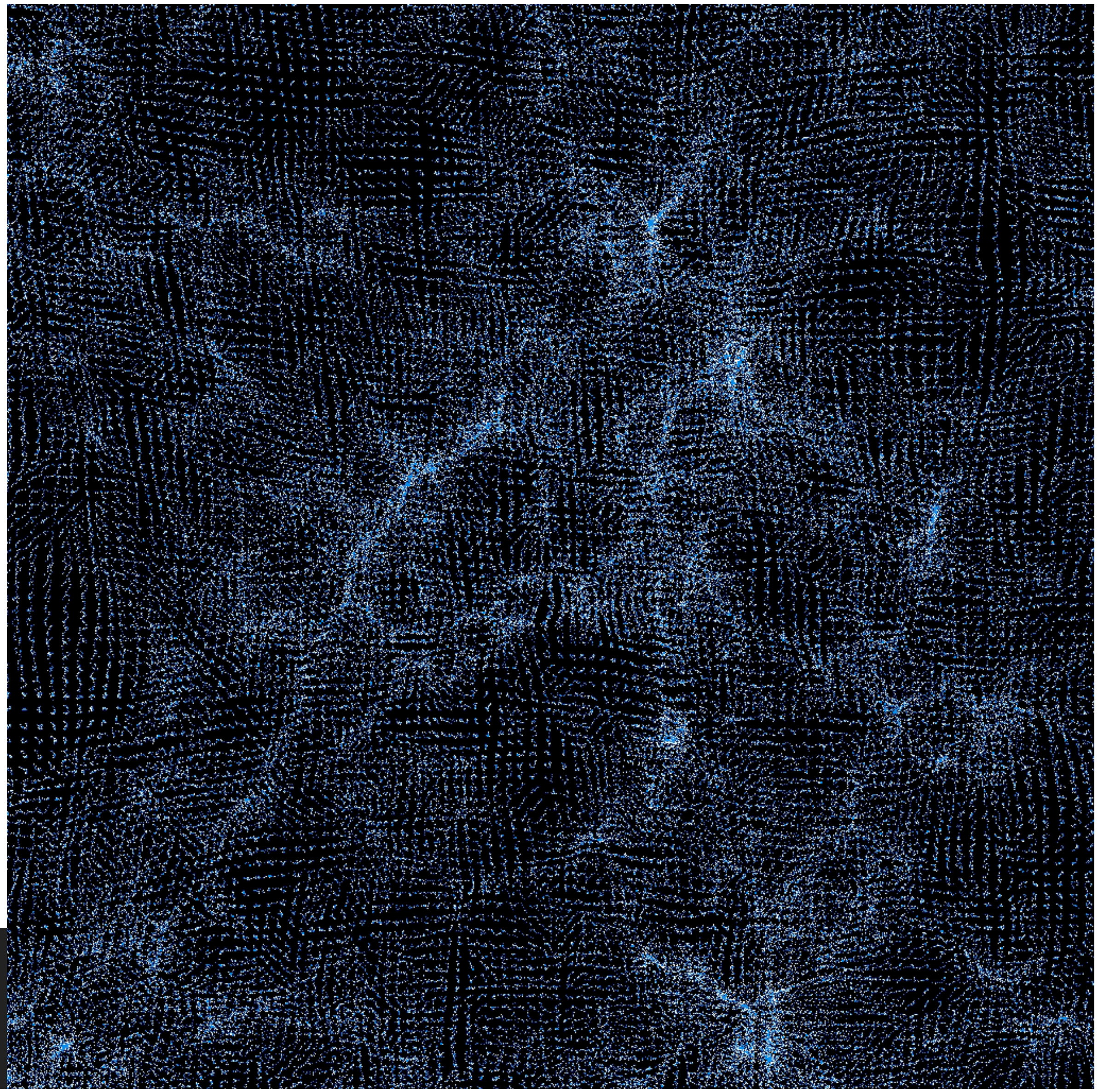
Query image (snowy day)

Matches

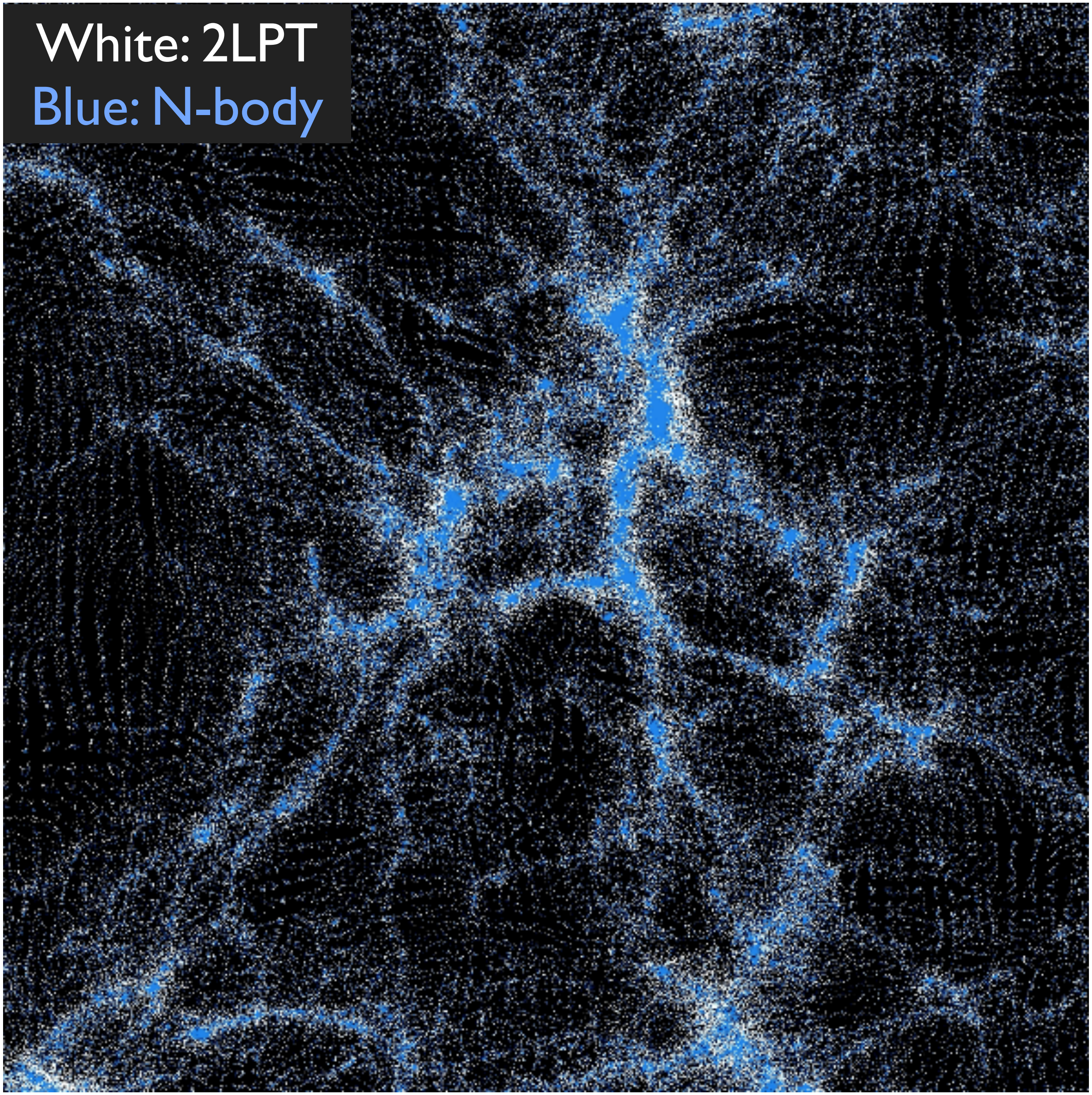
[Doersch 2012]
"Find meaningful visual elements that are unique to Paris"


Initial condition of the Universe
Approximate Simulation of the Universe, P0
Full Simulation of the Universe, Q0
O(10) sec
O(10) sec
Approximate evolution of Universe
Distribution-to-distribution regression
O(100,000) CPU seconds
Exact evolution of the Universe

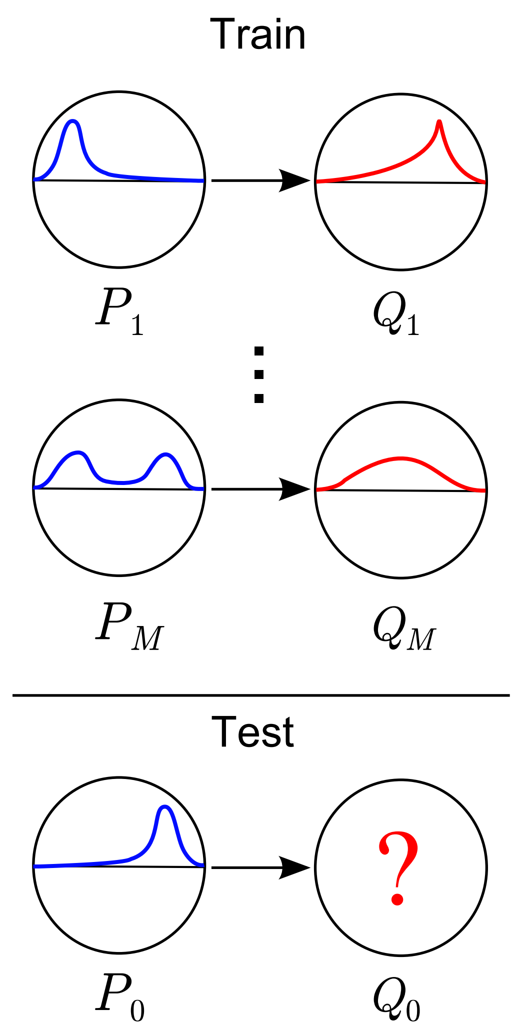
How does approximate (2LPT) field compares to N-body



White: 2LPT
Blue: N-body

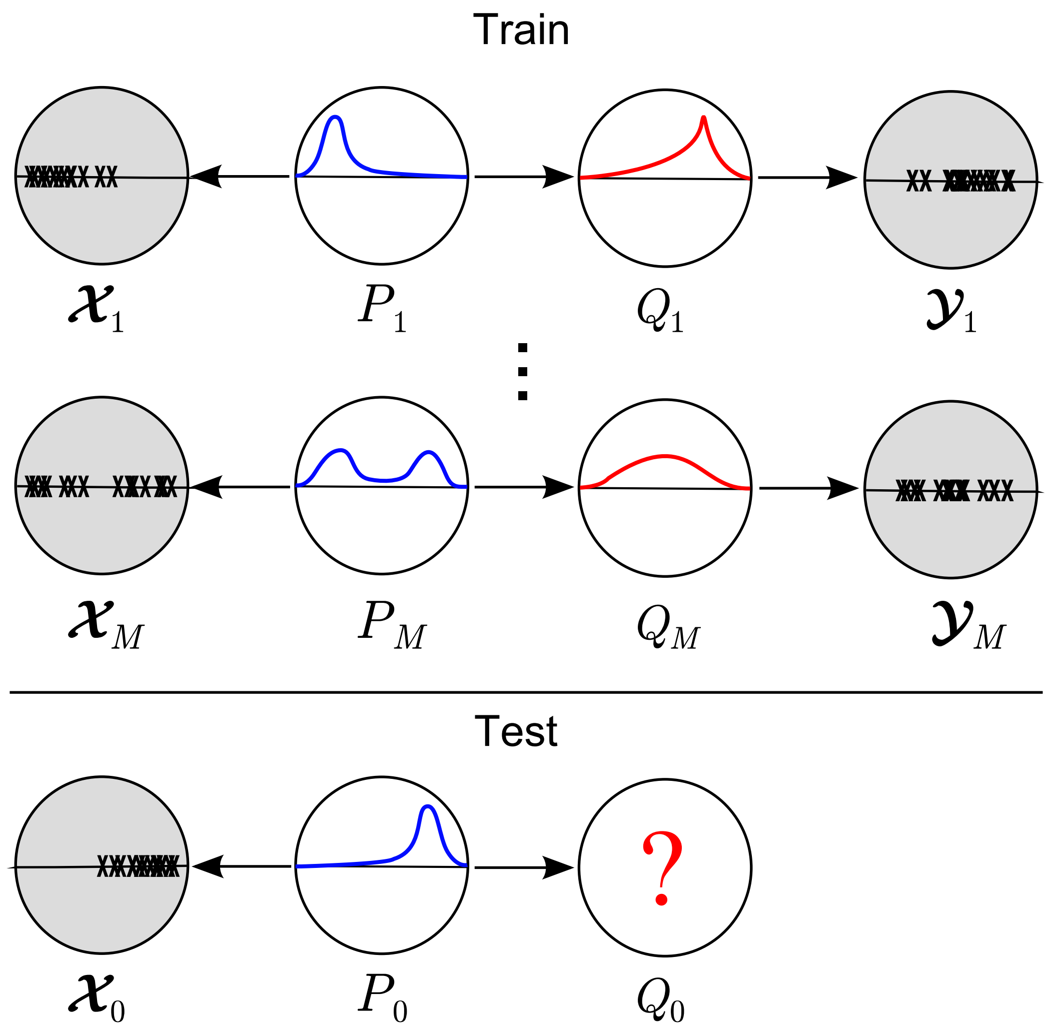

White: 2LPT
Blue: N-body

(2LPT) Distribution to (N-body) Distribution Machine Learning algorithms



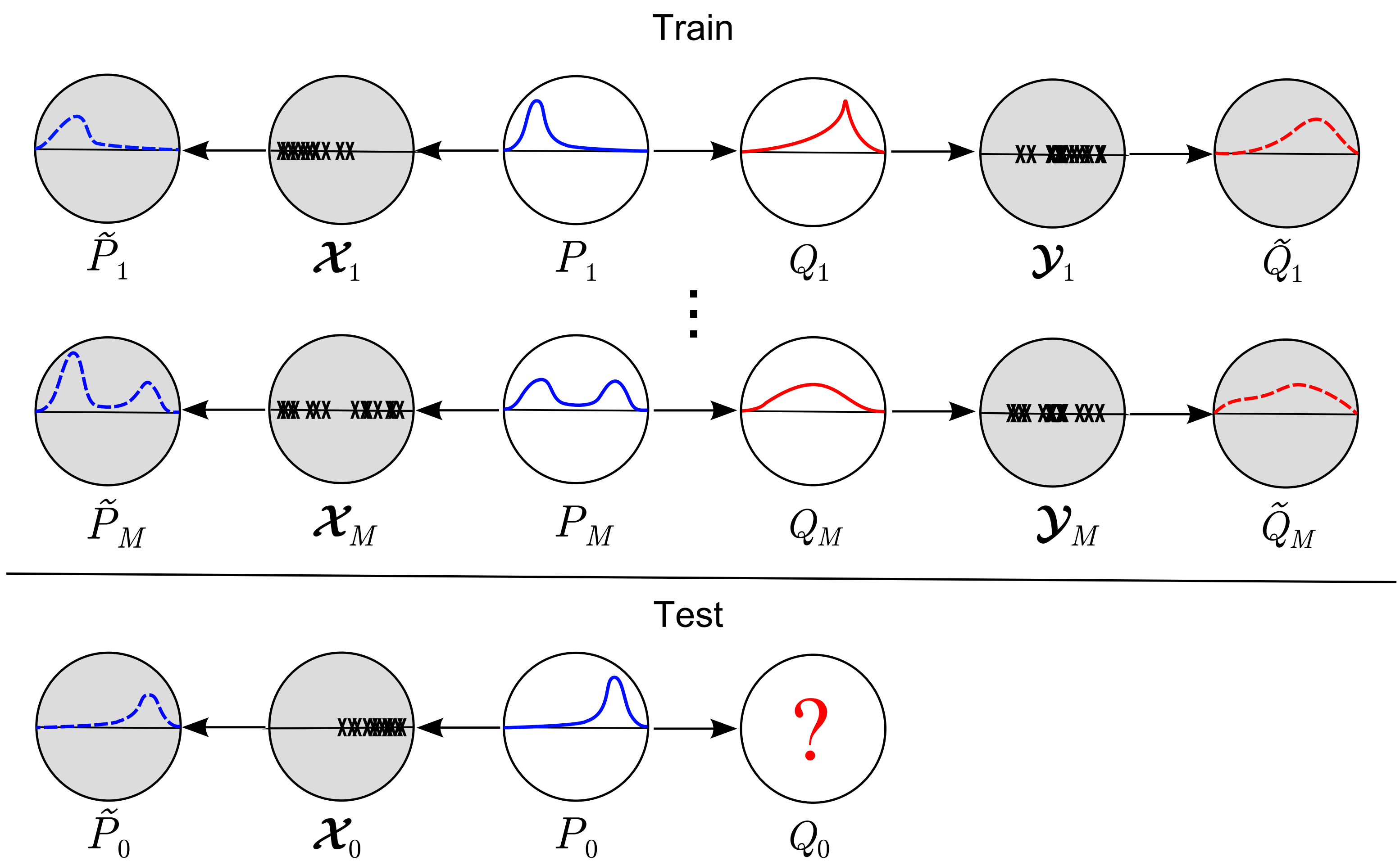
Oliver, Poczos, Schneider, 2013, ICML

(2LPT) Distribution to (N-body) Distribution Machine Learning algorithms

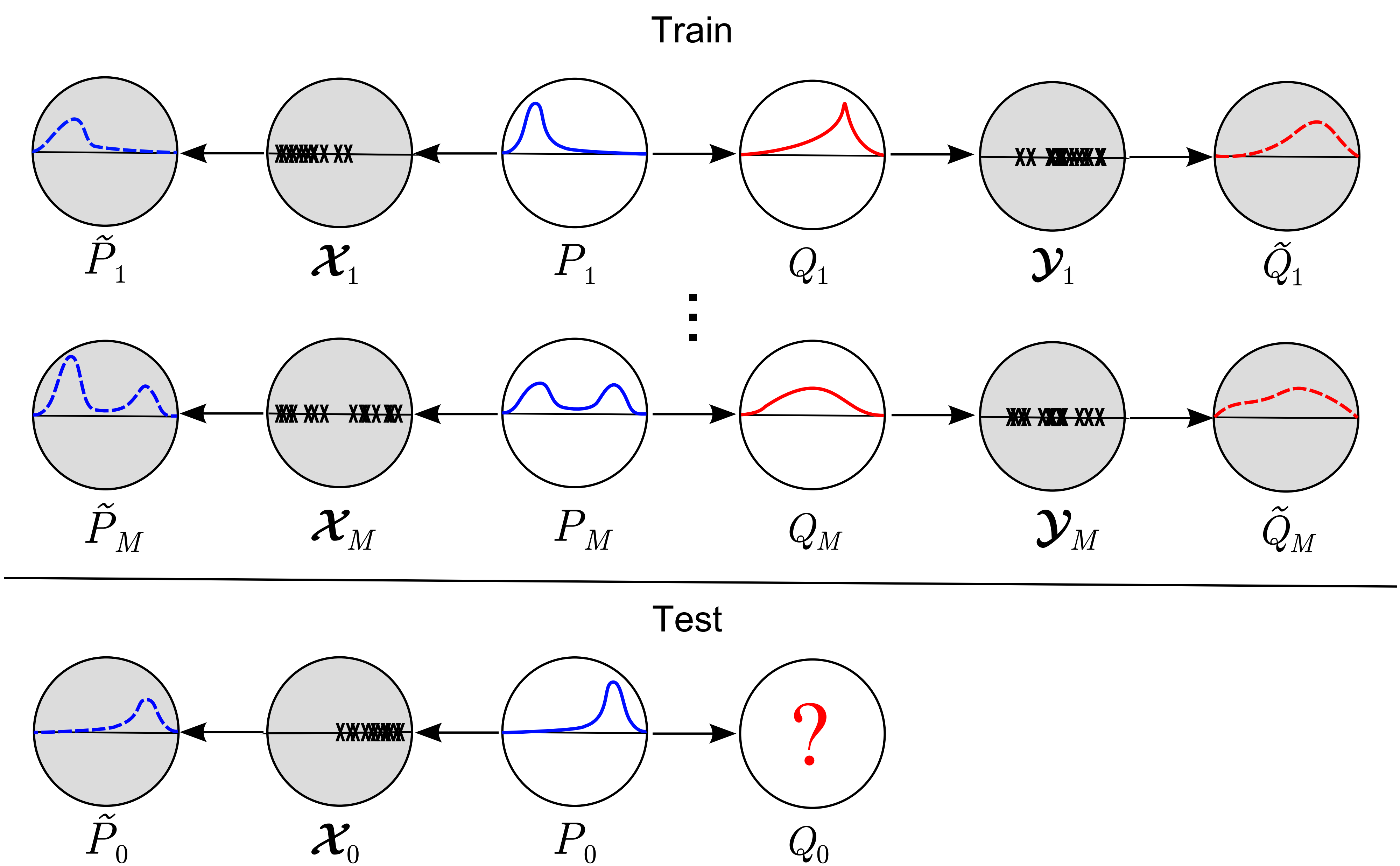


Oliver, Poczos, Schneider, 2013, ICML

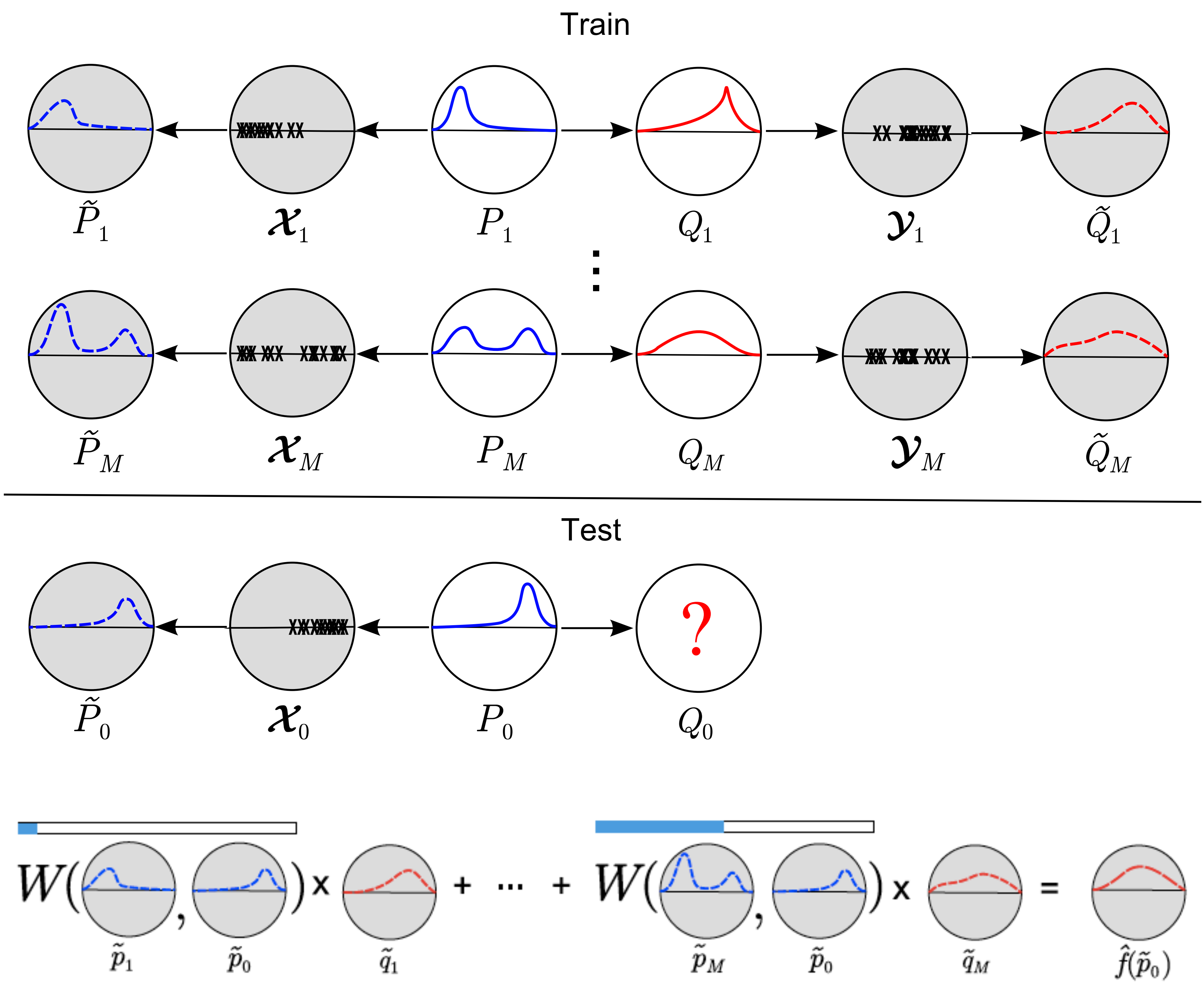
(2LPT) Distribution to (N-body) Distribution Machine Learning algorithms



Oliver, Poczos, Schneider, 2013, ICML



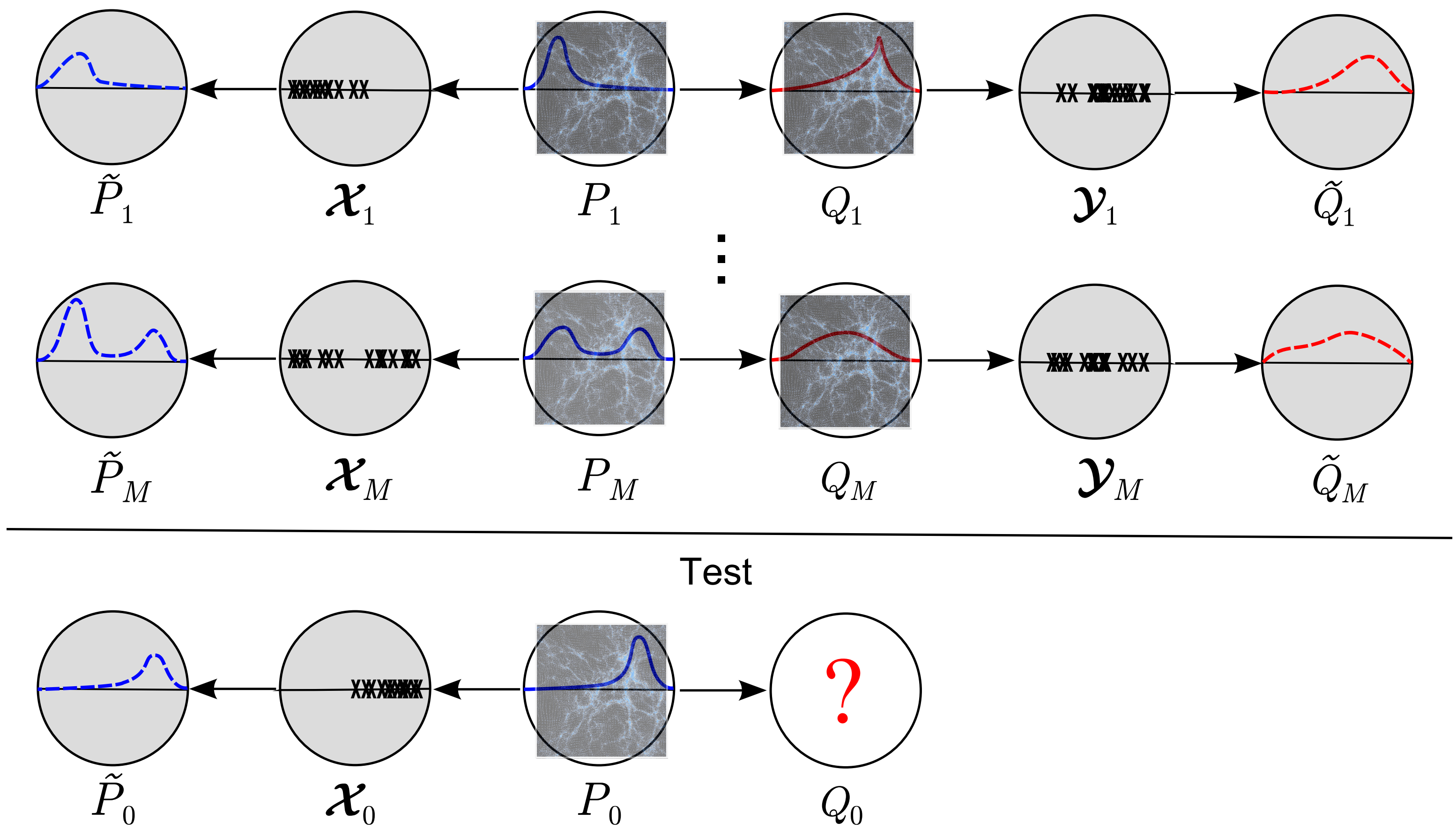
Oliver, Poczos, Schneider, 2013, ICML



Oliver, Poczos, Schneider, 2013, ICML

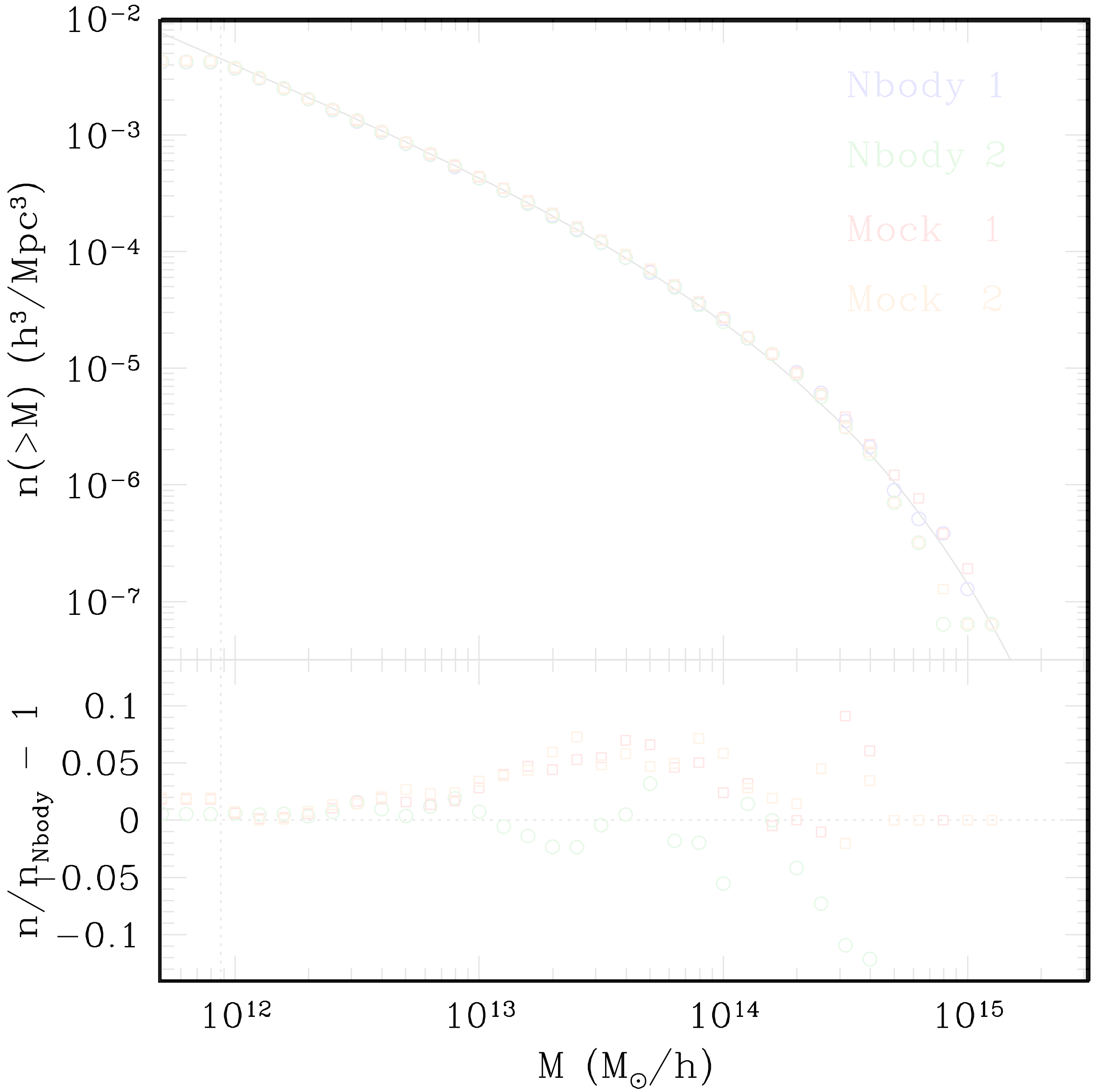
(2LPT) Distribution to (N-body) Distribution
Machine Learning algorithms

2LPT (P) -> N-body(Q)



Very much in prep and speculation

Very preliminary:
let's compare the halos statistics



Outline

- SDSS III- BOSS :
  - Mini-summary and looking forward
- BAO Reconstruction
- Ultra-fast simulations of the Universe?
- BAO in cross-correlations?

You can pull out a BAO signal
even when one of the tracers is really sparse

BAO in QSO x Lyα Cross-Correlations

BOSS at High Redshift

- ~ 100,000 z = 2.15 - 3.5 quasar spectra
- ~ 45 million Lyα pixels
- First anisotropic BAO measurement at $z > 2$ done this year using Lyα autocorrelations.
- Quasar autocorrelation not yet good enough for BAO, but we can cross-correlate with Lyα!

BAO in QSO x Lyα Cross-Correlations

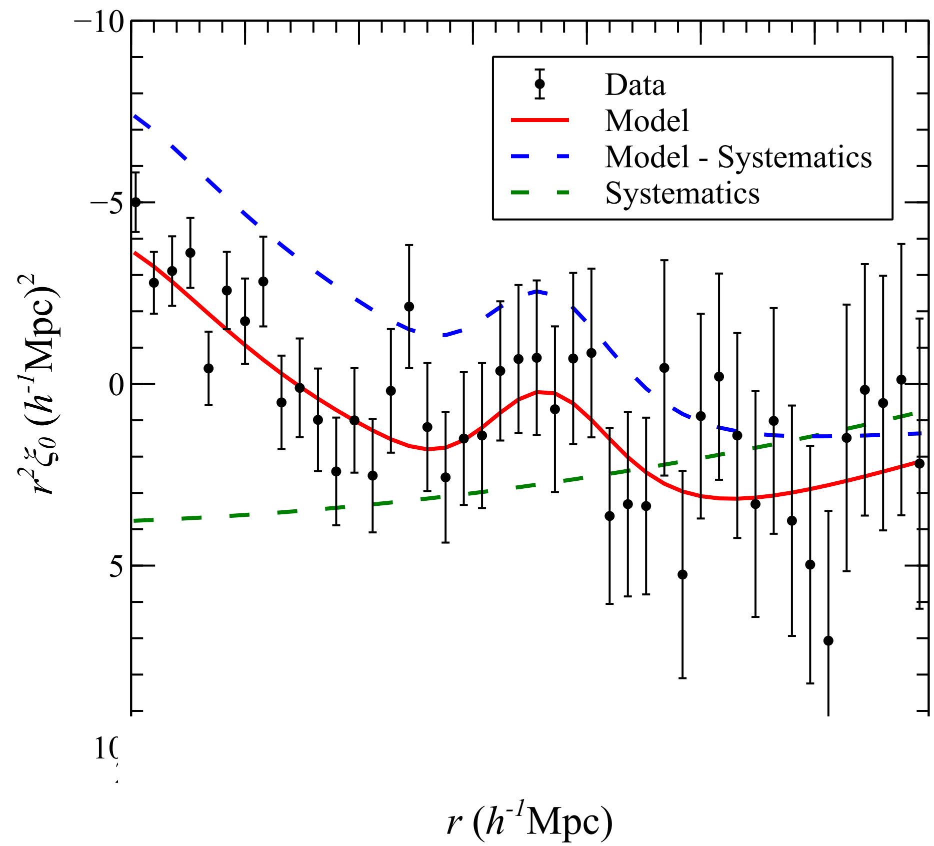
BOSS at High Redshift

- ~ 100,000 z = 2.15 - 3.5 quasar spectra
- ~ 45 million Lyα pixels
- First anisotropic BAO measurement at z > 2 done this year using Lyα autocorrelations.
- Quasar autocorrelation not yet good enough for BAO, but we can cross-correlate with Lyα!

O'Connel, SH , et al. (in prep)
Font-Ribera, Kirkby et al. (in prep)

Our Measurement

- First BAO detection in quasars (via Cross-correlations).Developed new estimator, method for determining covariance matrix, systematic corrections for Lyα forest.



BOSS: What did we learn?

- SDSS III- BOSS : Lots of cosmological constraints and other cool things I didn't get to talk about...
- BAO Reconstruction : Different methods give rise to similar (but not the same) BAO. Possible improvement?
- Ultra-fast simulations of the Universe! A way forward to make many synthetic Universes quickly?
- BAO in cross-correlations! We can look for BAO in the most unlikely places, such as very sparsely sampled quasars. [Other applications of correlating neutral hydrogen and halos?]

The not-so-new anymore assistant faculty
Shirley Ho