SDSS III

How to learn to Love the BOSS
Baryon Oscillations Spectroscopic Survey

Shirley Ho, Zonggie Liu, Yen-Chi Chen, Mariana Vargas, Hy Trac, Jeff Schneider, Barnabas Poczos, Junier Oliver, Chris Genovese, Peter Freeman

+BOSS collaboration

Carnegie Mellon University

Planck 2014, Ferrara, Italy

Dec 2nd, 2014

What are Baryon Acoustic Oscillations?

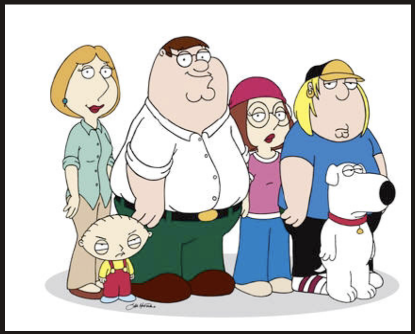
To measure BAO, we usually calculate the correlation function

What are Baryon Acoustic Oscillations?

What is the correlation function of population during the day?

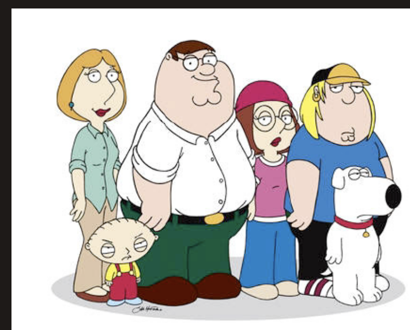
What are Baryon Acoustic Oscillations?

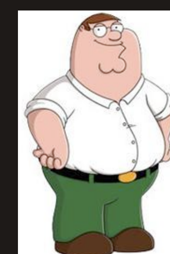
What is the correlation function of population during the day?

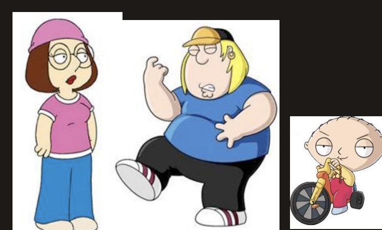


What are Baryon Acoustic Oscillations?

What is the correlation function of population during the day?

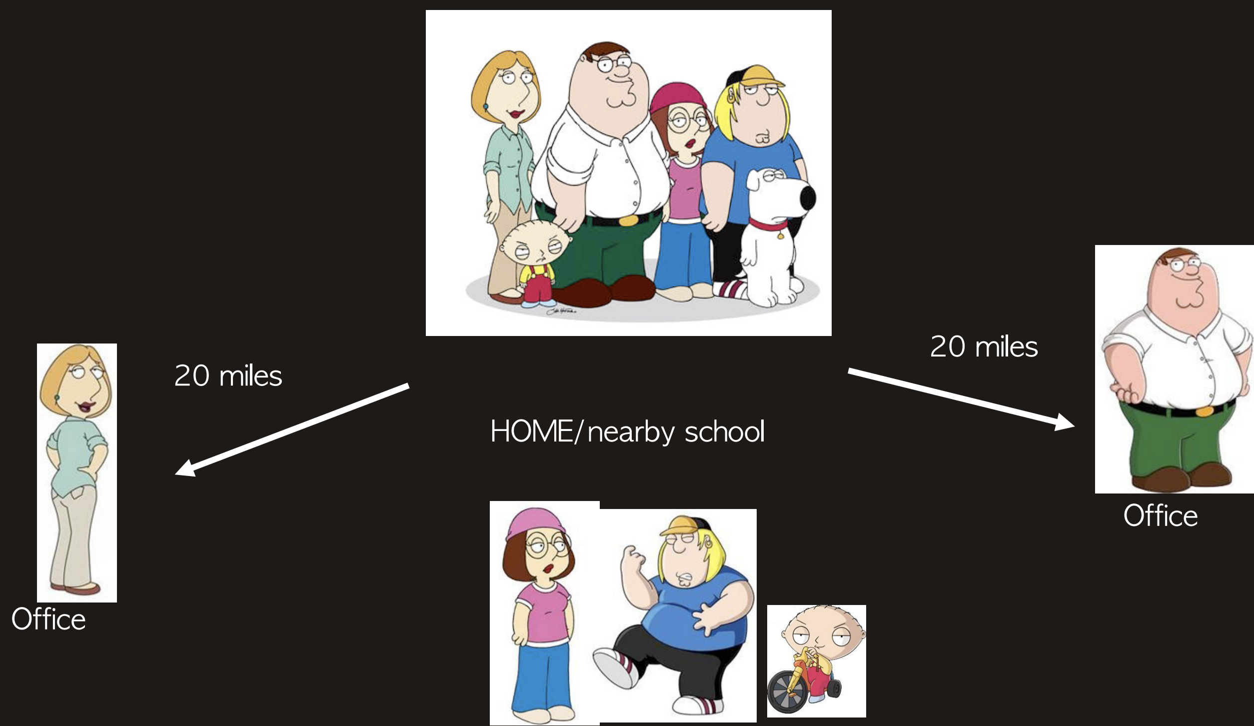






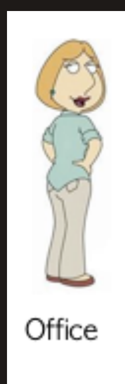
What are Baryon Acoustic Oscillations?

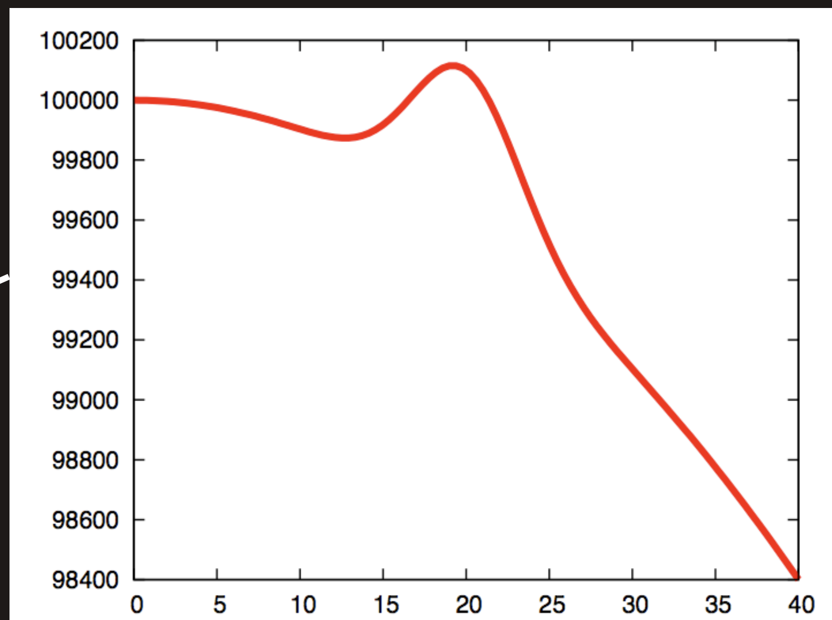
What is the correlation function of population during the day?

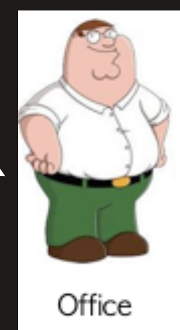


What are Baryon Acoustic Oscillations?

What is the correlation function of population during the day?



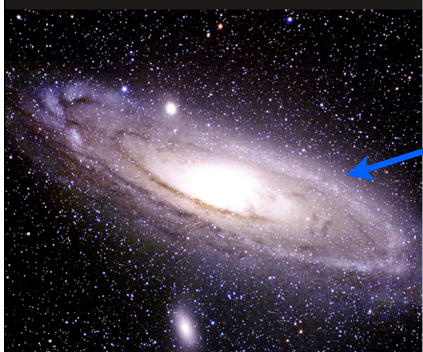


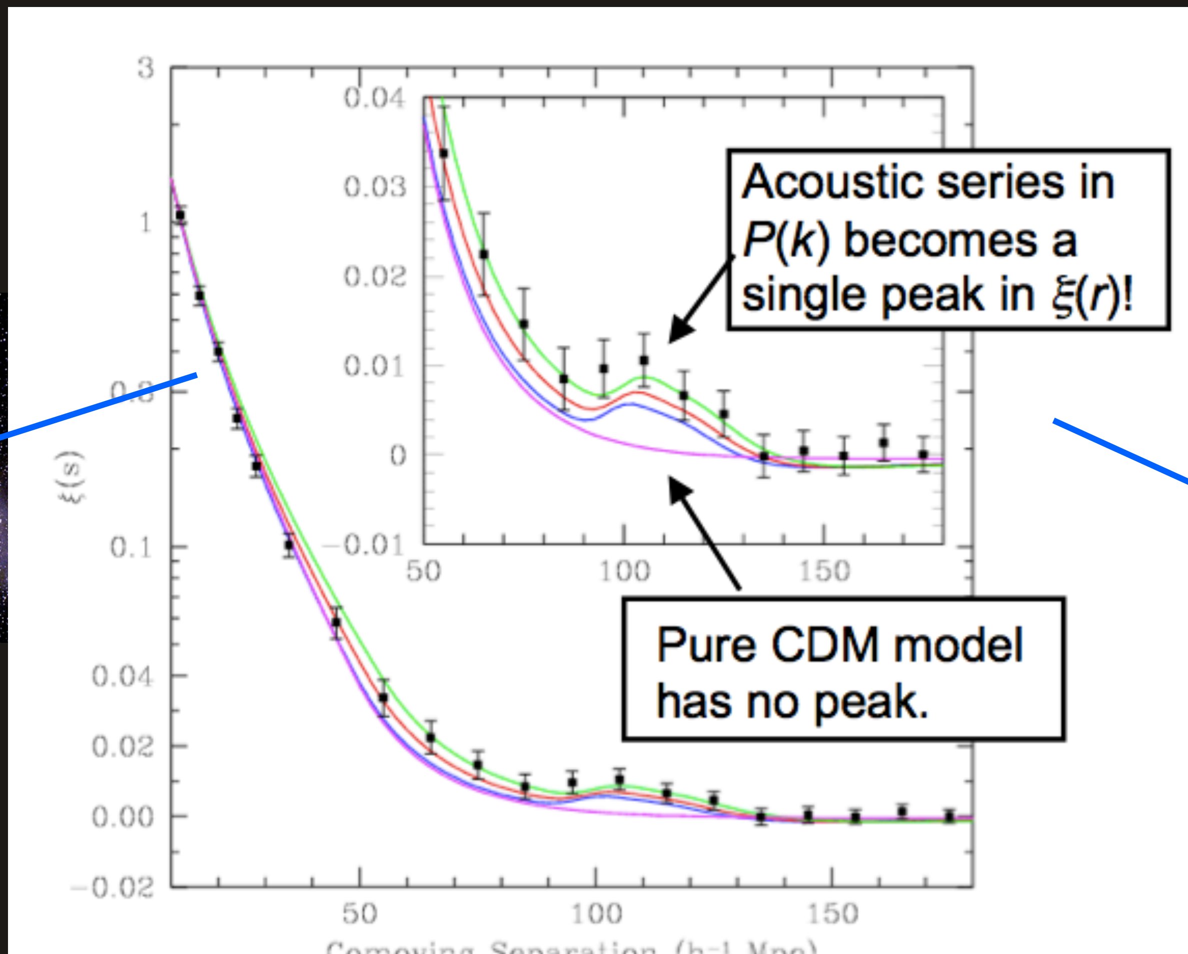


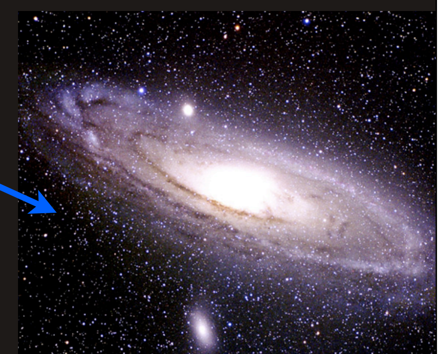
A bump in 20 miles!

What are Baryon Acoustic Oscillations?

To measure BAO, we first calculate the correlation function







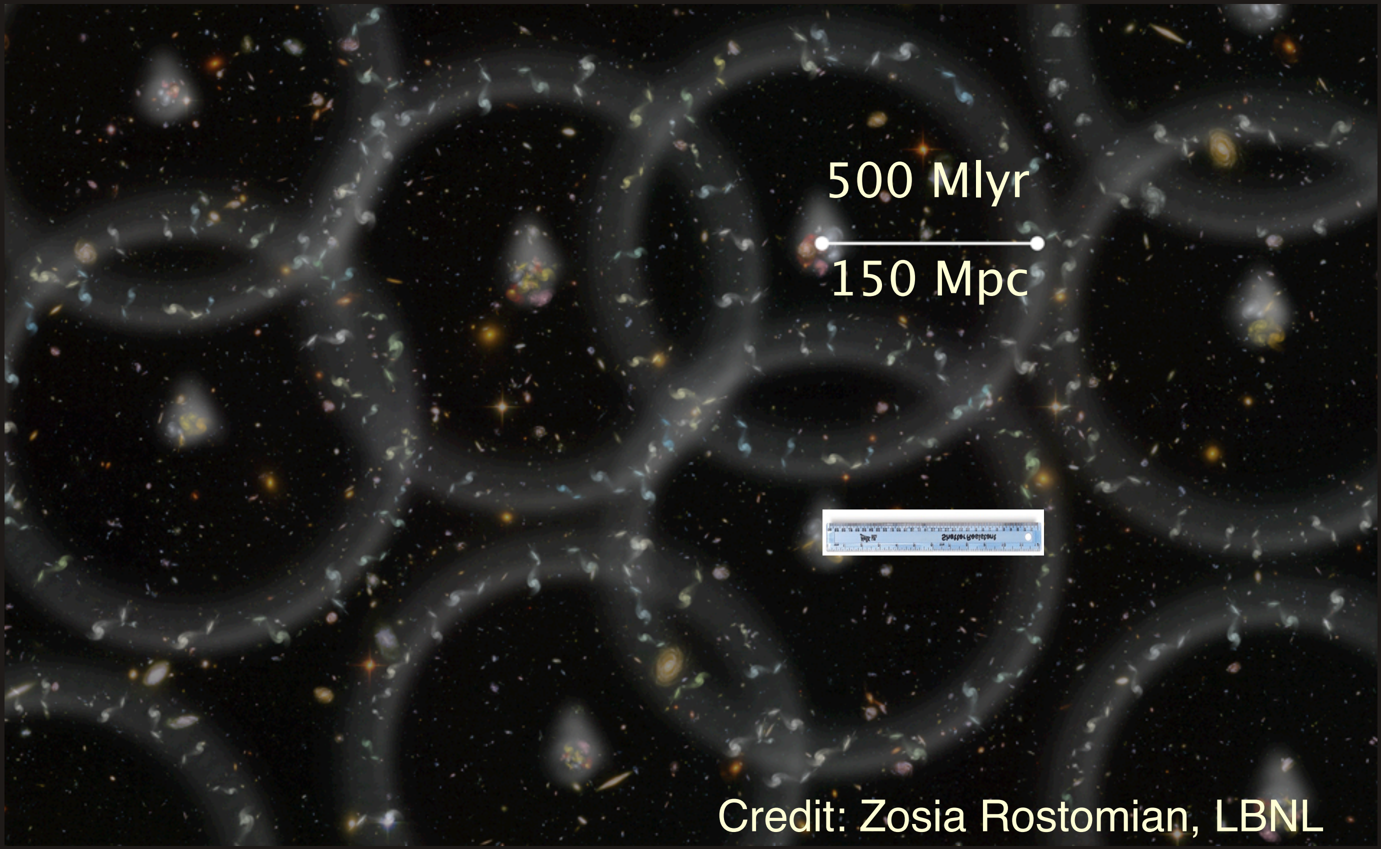
A bump at ~150Mpc

Shirley Ho, Planck 2014, Ferrara

SDSS III

BAO and Galaxies

- Pairs of galaxies are slightly more likely to be separated by 150 Mpc than 120 Mpc or 170 Mpc.



NOTE: BAO effects highly exaggerated here

SDSSIII

BAO as a Standard Ruler

- This distance of 150 Mpc is very accurately computed from the anisotropies of the CMB.
  - 0.4% calibration with current CMB.

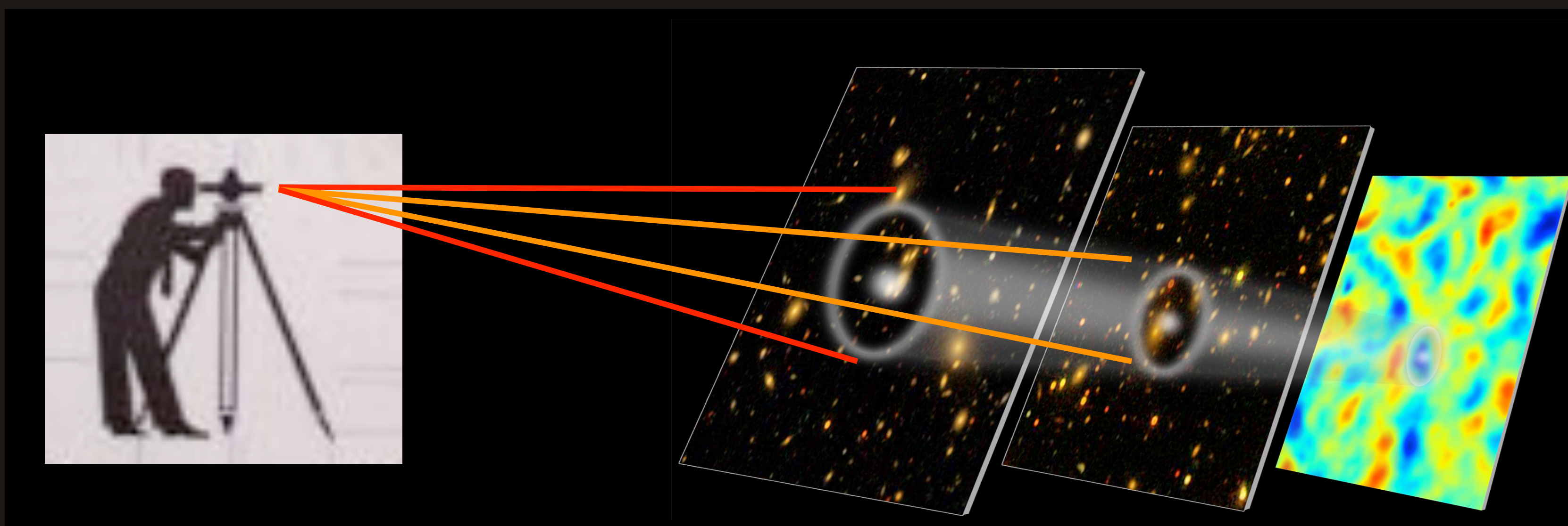
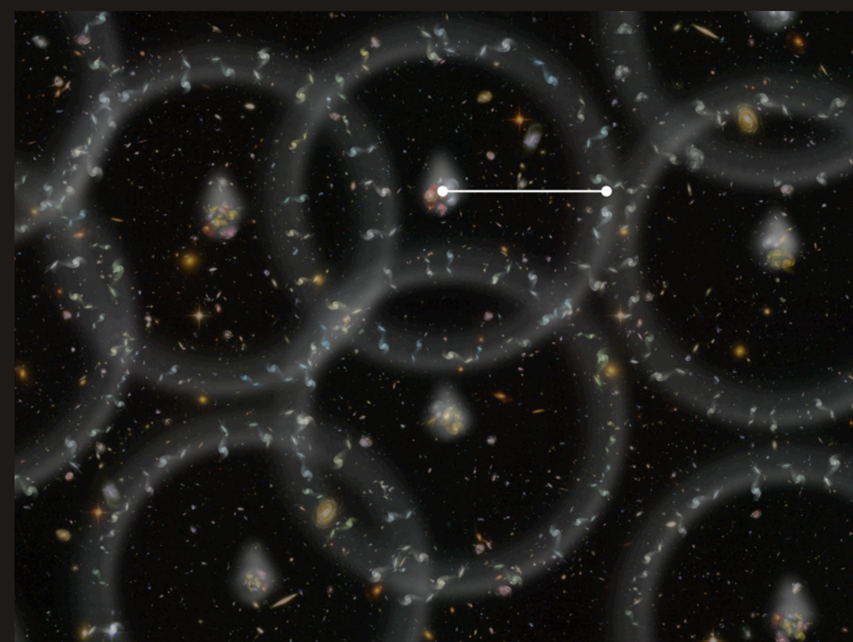


Image Credit: E.M. Huff, the SDSS-III team, and the South Pole Telescope team. Graphic by Zosia Rostomian

SDSSIII

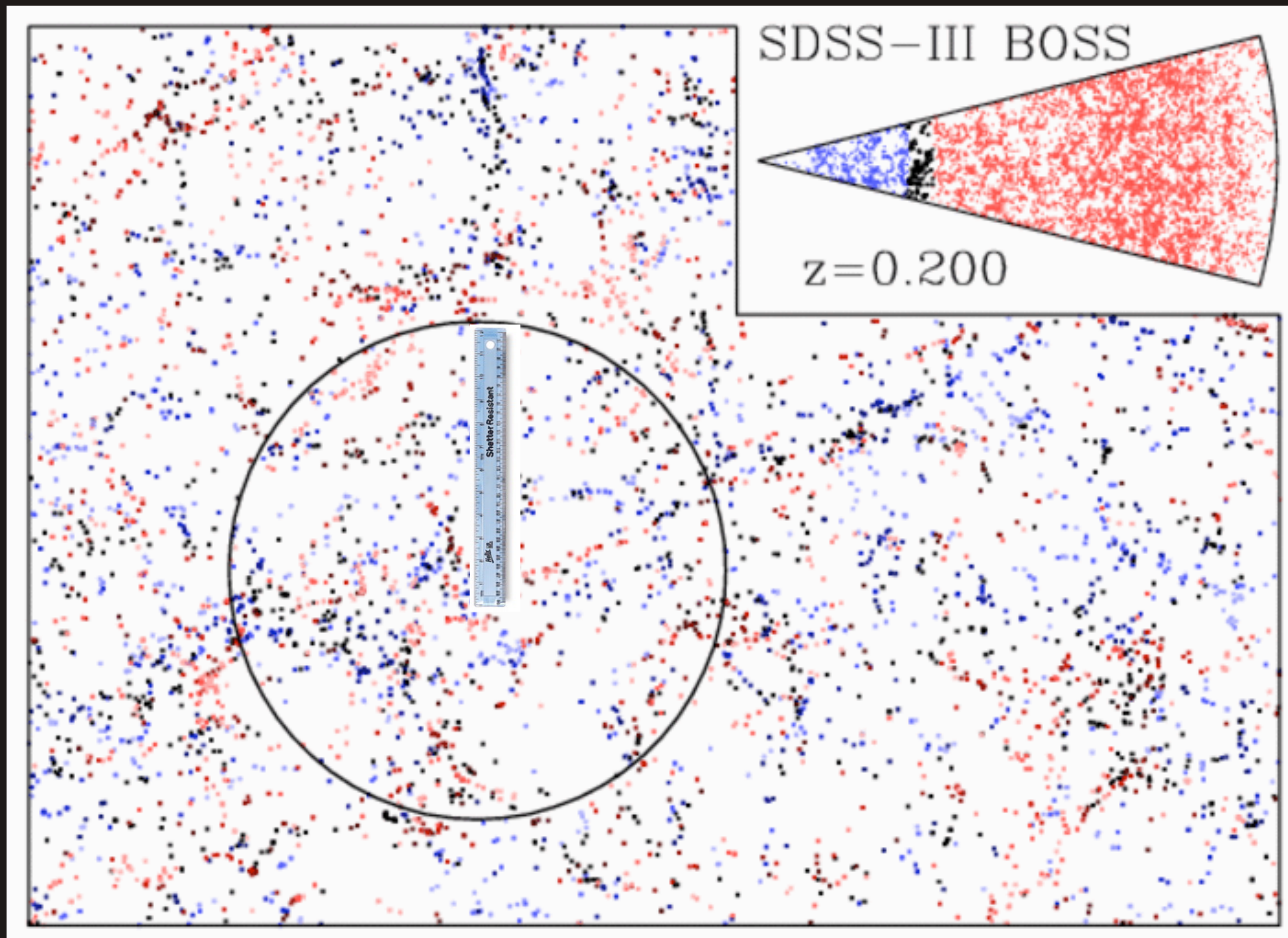
SDSS III - BOSS

- In SDSS-III, we use maps of the large-scale structure of the Universe to detect the imprint of the sound waves.
- We use 3 different tracers of the cosmic density map:
  - Galaxies at redshifts 0.2 to 0.7.
  - Quasars at redshifts 2.1 to 3.5.
  - The intergalactic medium as revealed by the Lyman α Forest, at redshifts 2.1 to 3.5.
- We look for an excess clustering of overdensity regions separated by 150 Mpc



SDSS III

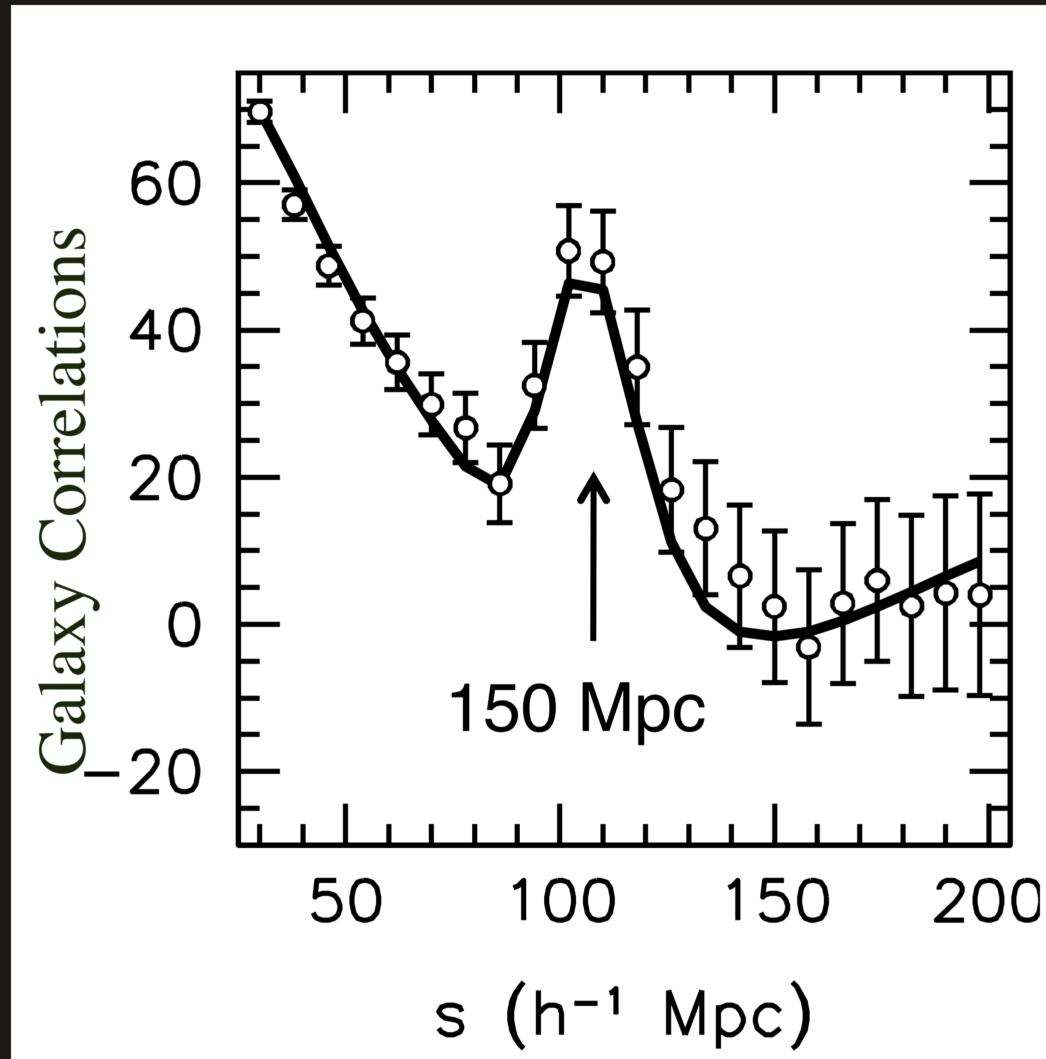
A Slice of BOSS



Credit: D. Eisenstein

BAO in BOSS Galaxies

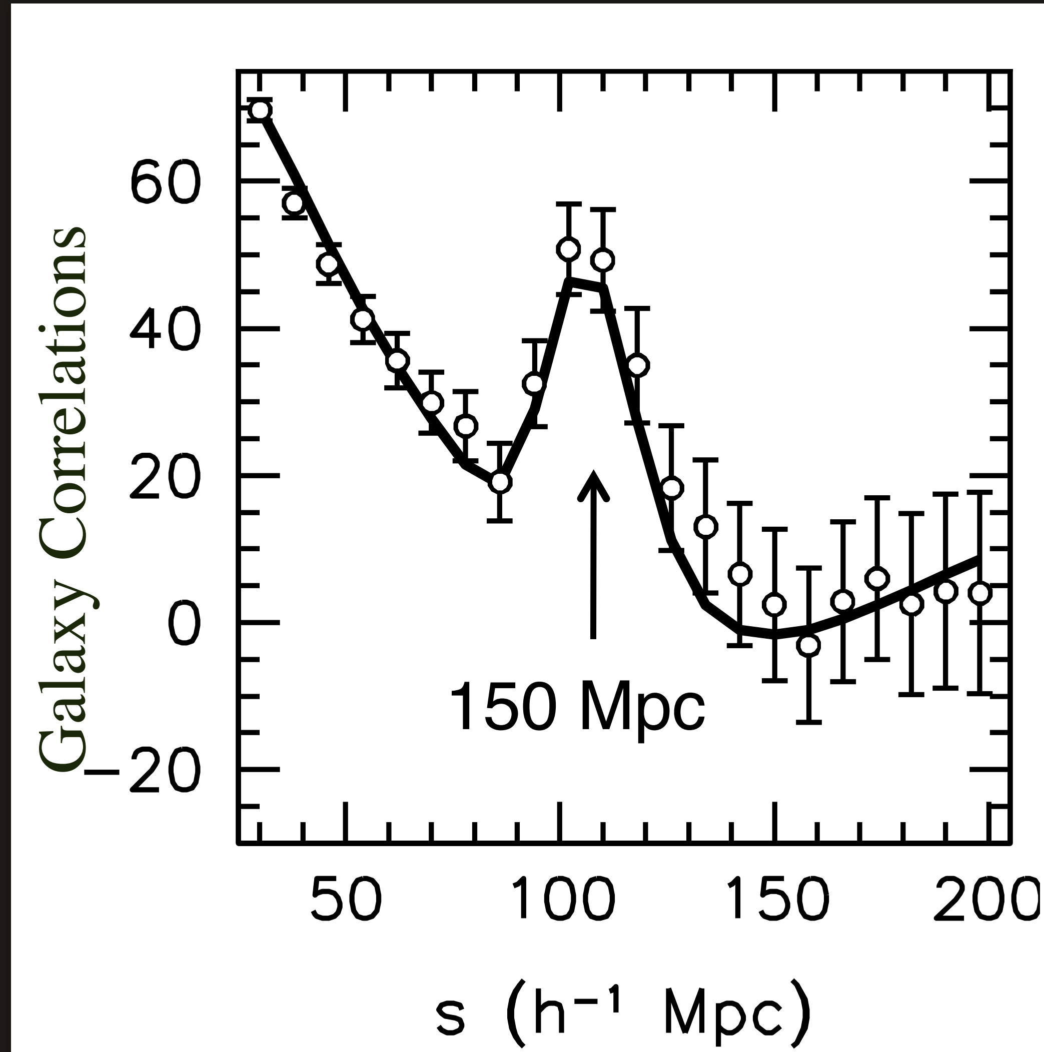
- Clustering Analysis of the BOSS galaxy sample has produced the world's best detection of the late-time acoustic peak.



Anderson et al. 2014;
Vargas, Ho et al. 2014;
Tojeiro et al. 2014

BAO in BOSS Galaxies

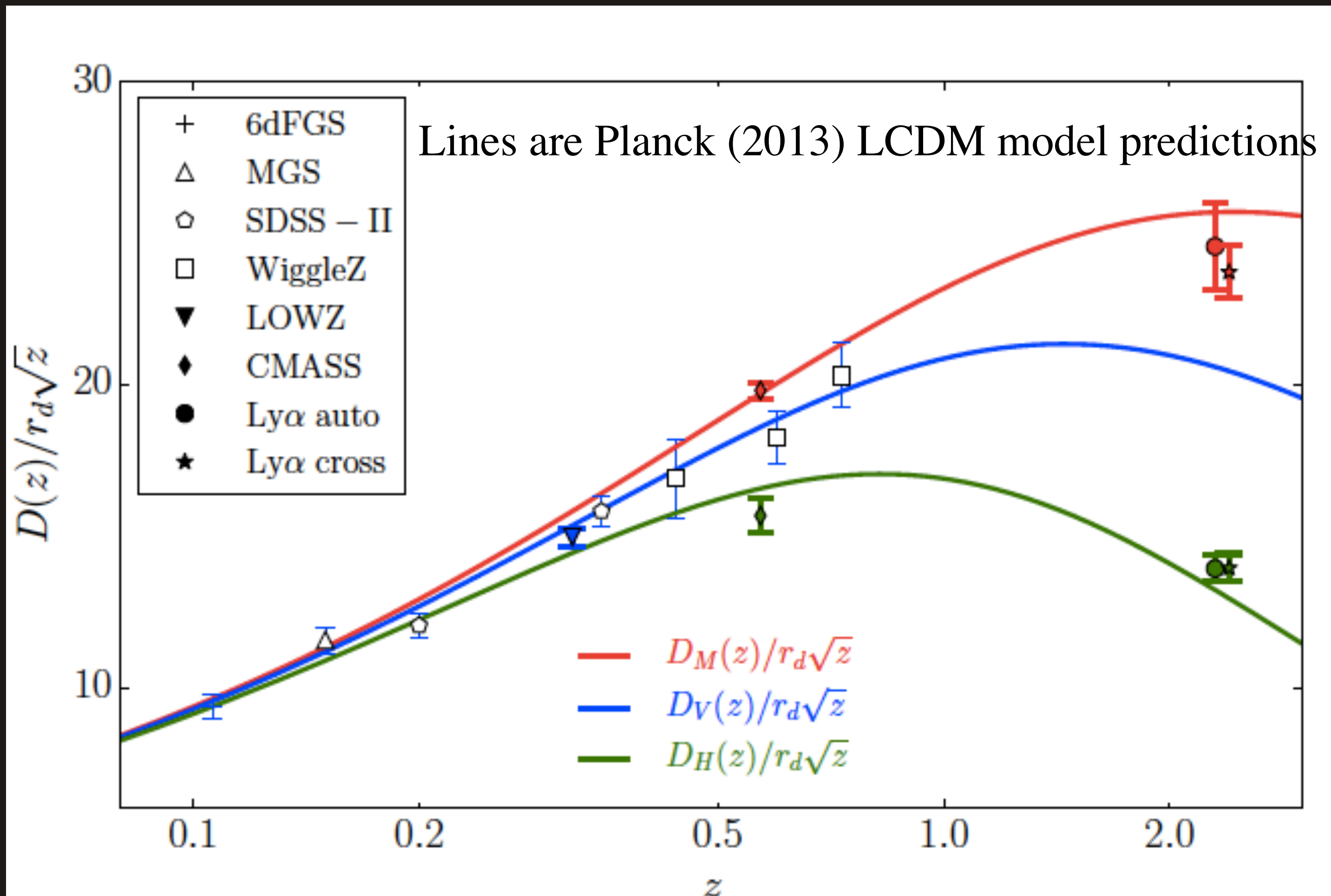
- The peak location is measured to 1.0% in our $z = 0.57$ sample and 2.1% in our $z = 0.32$ sample



Anderson et al. 2014;
Vargas, Ho et al. 2014;
Tojeiro et al. 2014

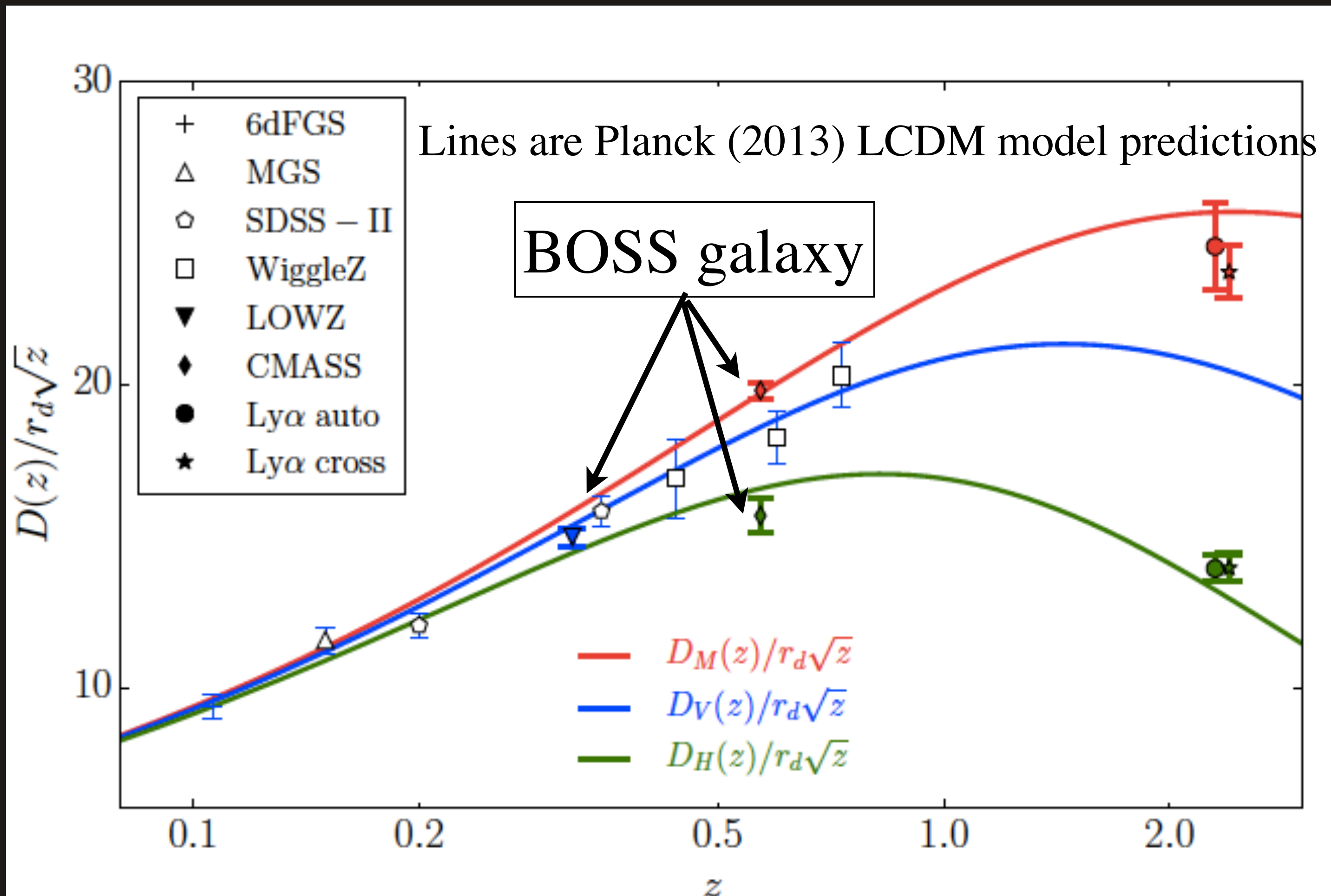
SDSSIII

Combining all BAO measurements



BOSS collaboration 2014
Anderson et al. 2014;
Vargas, Ho et al. 2014;
Tojeiro et al. 2014;

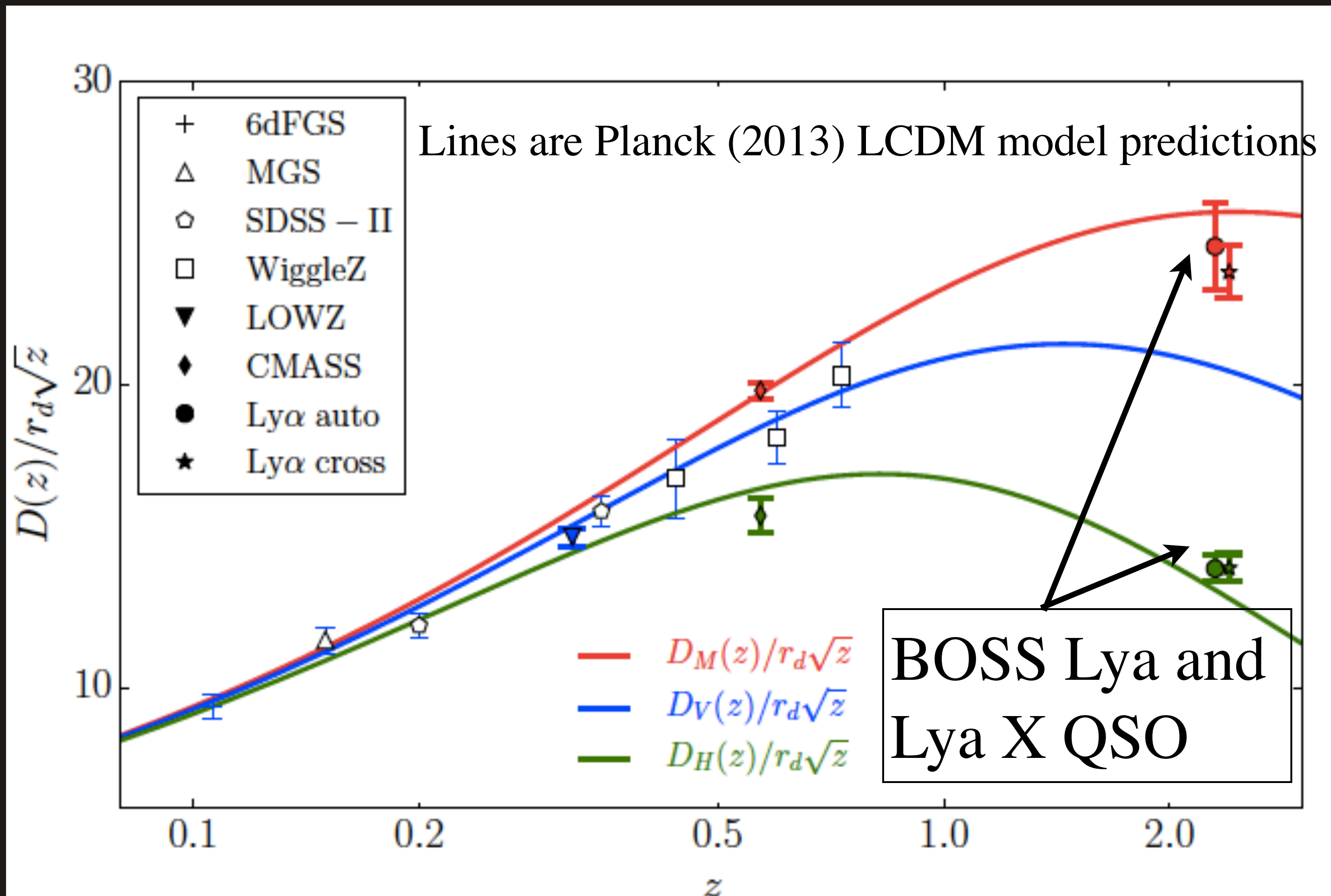
Combining all BAO measurements



BOSS collaboration 2014
Anderson et al. 2014;
Vargas, Ho et al. 2014;
Tojeiro et al. 2014;

SDSSIII

Combining all BAO measurements



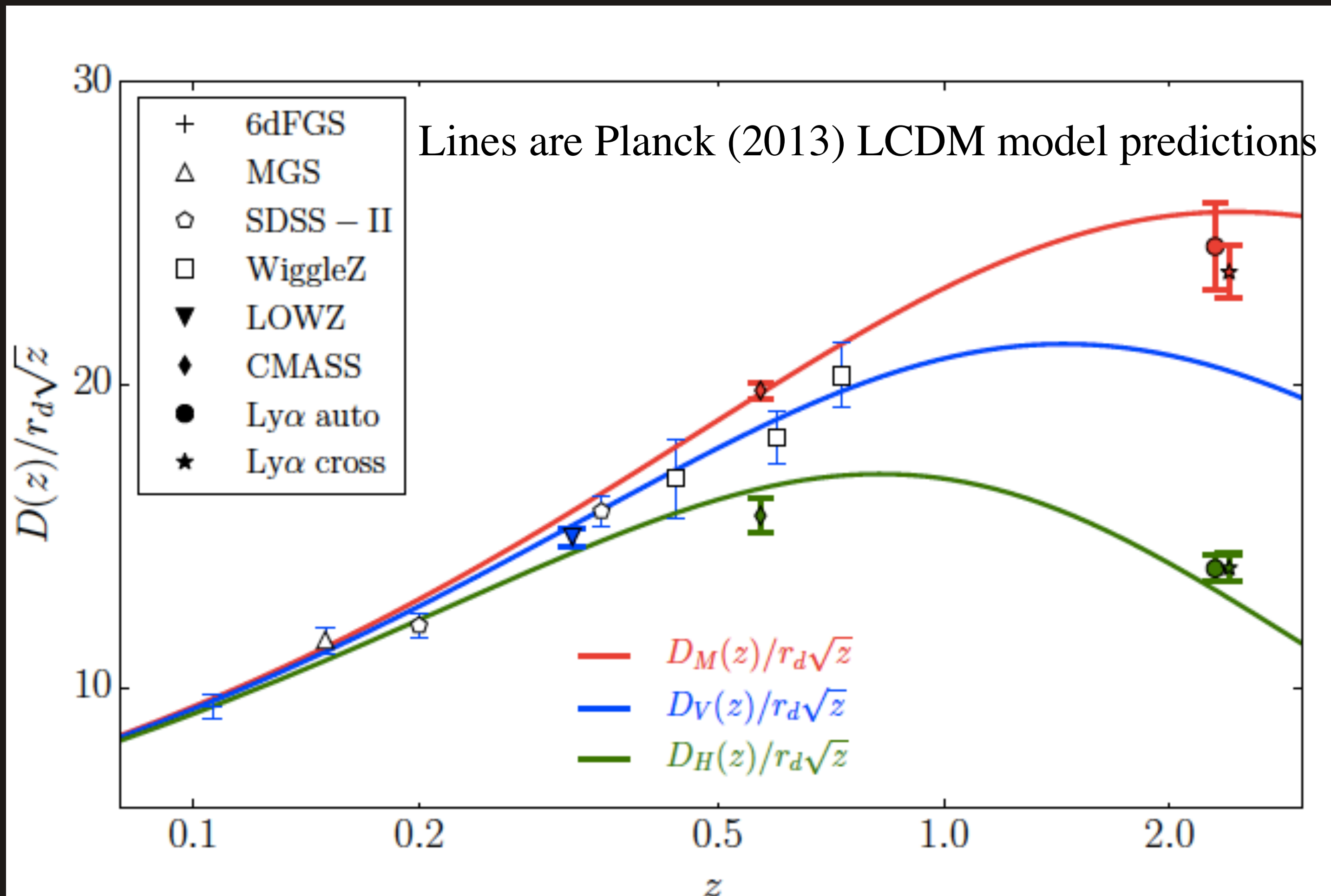
BOSS collaboration 2014

Font-Riberia et al. 2014

Busca et al. 2014

SDSSIII

Combining all BAO measurements



BOSS collaboration 2014
Anderson et al. 2014;
Vargas, Ho et al. 2014;
Tojeiro et al. 2014;

SDSSIII

Inverse Distance Ladder

- Absolute distances from BAO
- Do not use CMB except for sound horizon scale (standard radiation background calc.)
- SN data can shift up and down
- Hubble constant is robust to a variety of Dark energy models and spatial curvature.

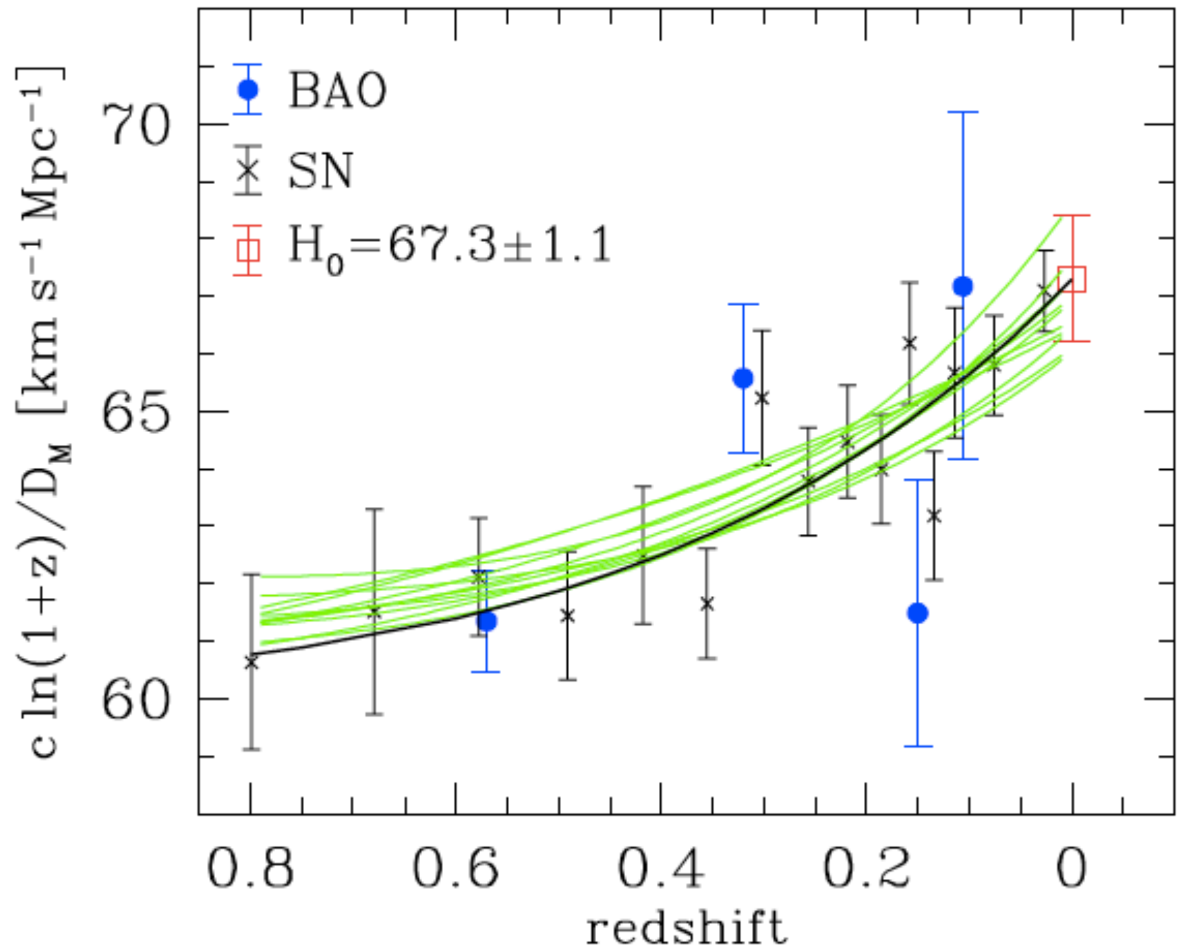
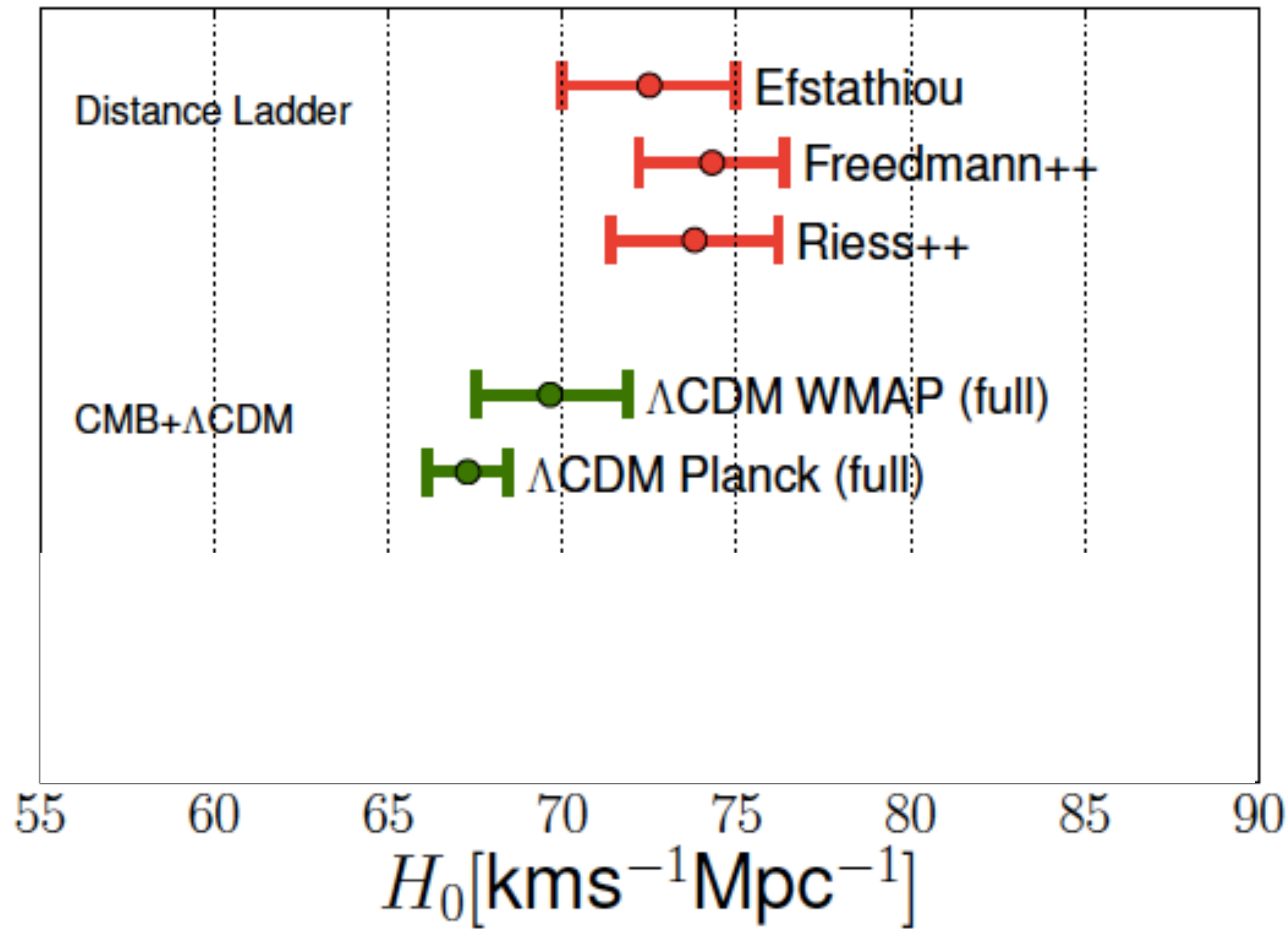


Figure 5. Determination of H_0 by the "inverse distance ladder" combining BAO absolute distance measurements and SNIa relative distance measurements, with CMB data used to calibrate the sound horizon scale r_d. The quan-

BOSS collaboration 2014

Shirley Ho, Planck 2014, Ferrara

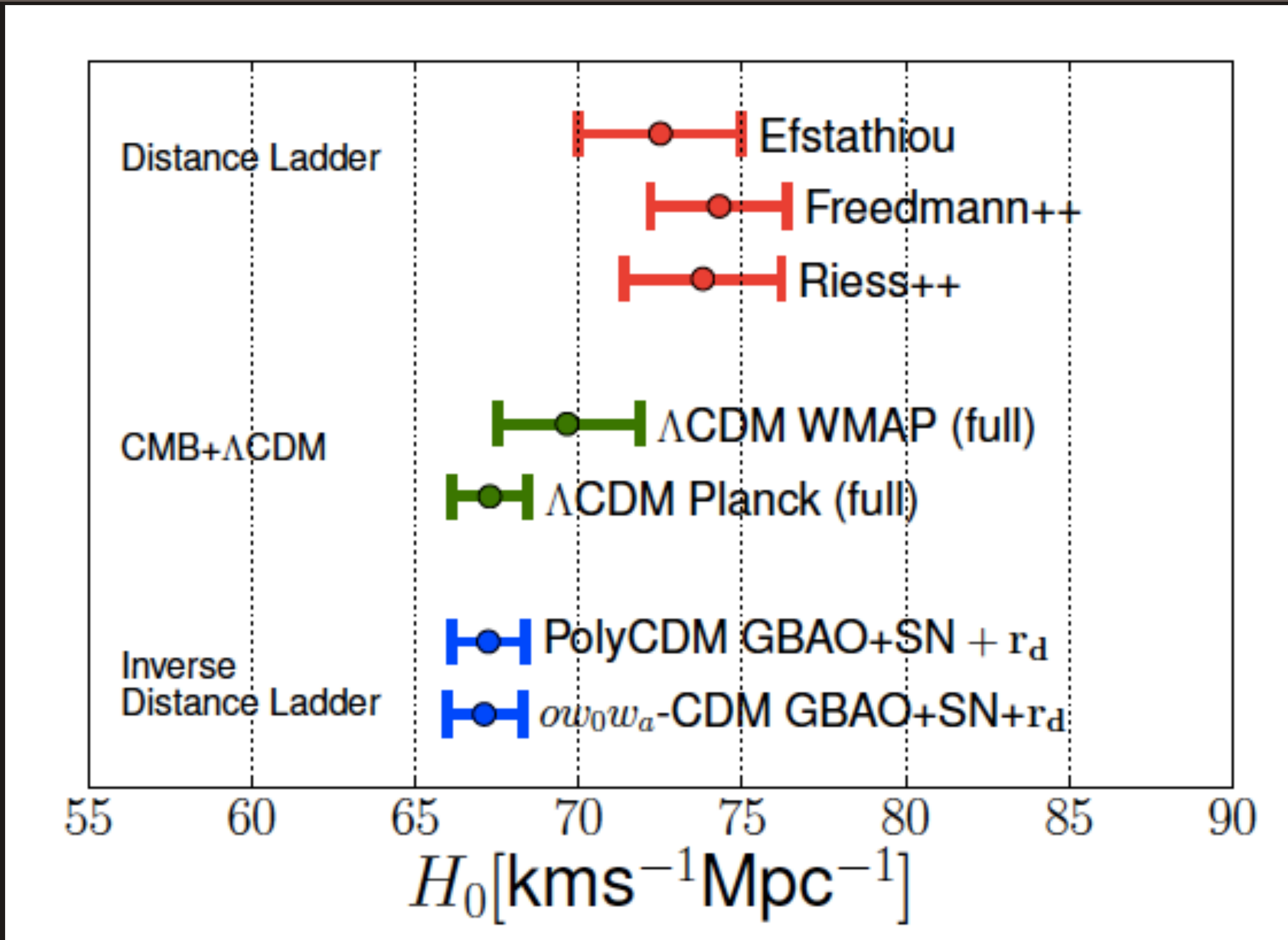
Hubble Constant comparison



BOSS collaboration 2014

Shirley Ho, Planck 2014, Ferrara

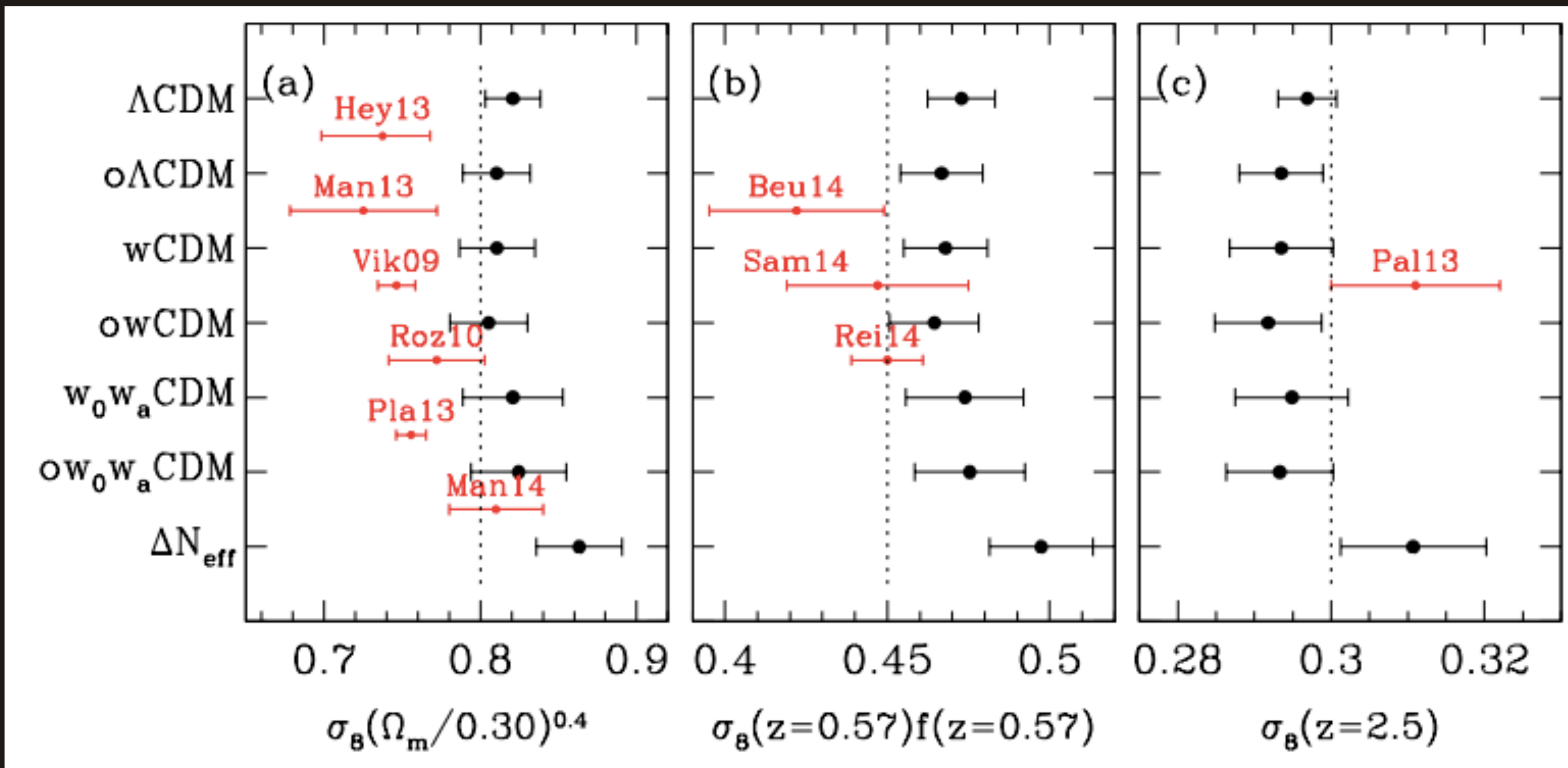
Hubble Constant comparison



BOSS collaboration 2014

Comparison with other LSS probes

Black: Planck +BAO + SN



Lensing, clusters

Redshift Space Distortions

Lya 1D P(k)

BOSS collaboration 2014

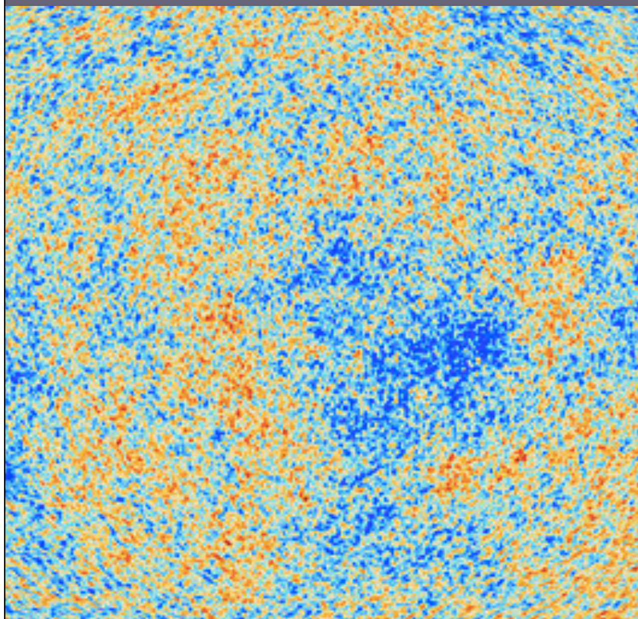
SDSS III

What are the challenges we learnt from BOSS ?

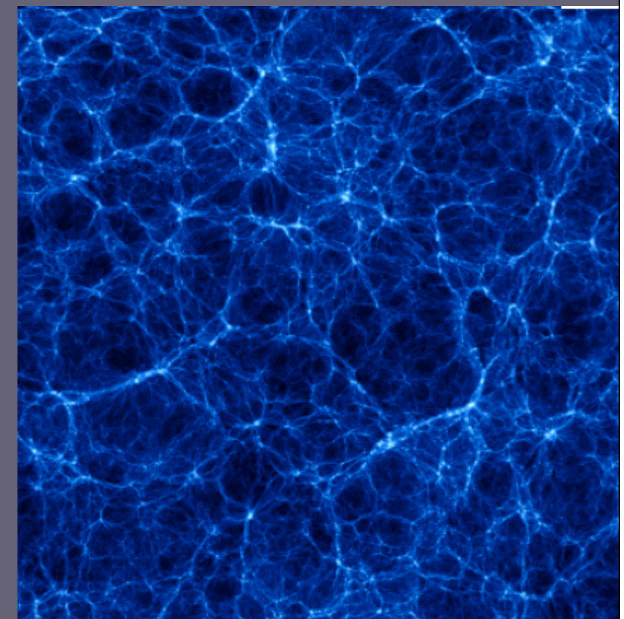
- Full blown analysis of Clustering of the Universe is hard, but definitely doable.
- Things we can improve:
  - Understanding the systematics from both observations and astrophysics (Ross, Percival et al. 2014, Vargas, Ho et al. 2014)
  - Making much better predictions of the Universe
  - Making more accurate measurements of the Universe

Making Better Predictions of the Universe

Initial condition of the Universe



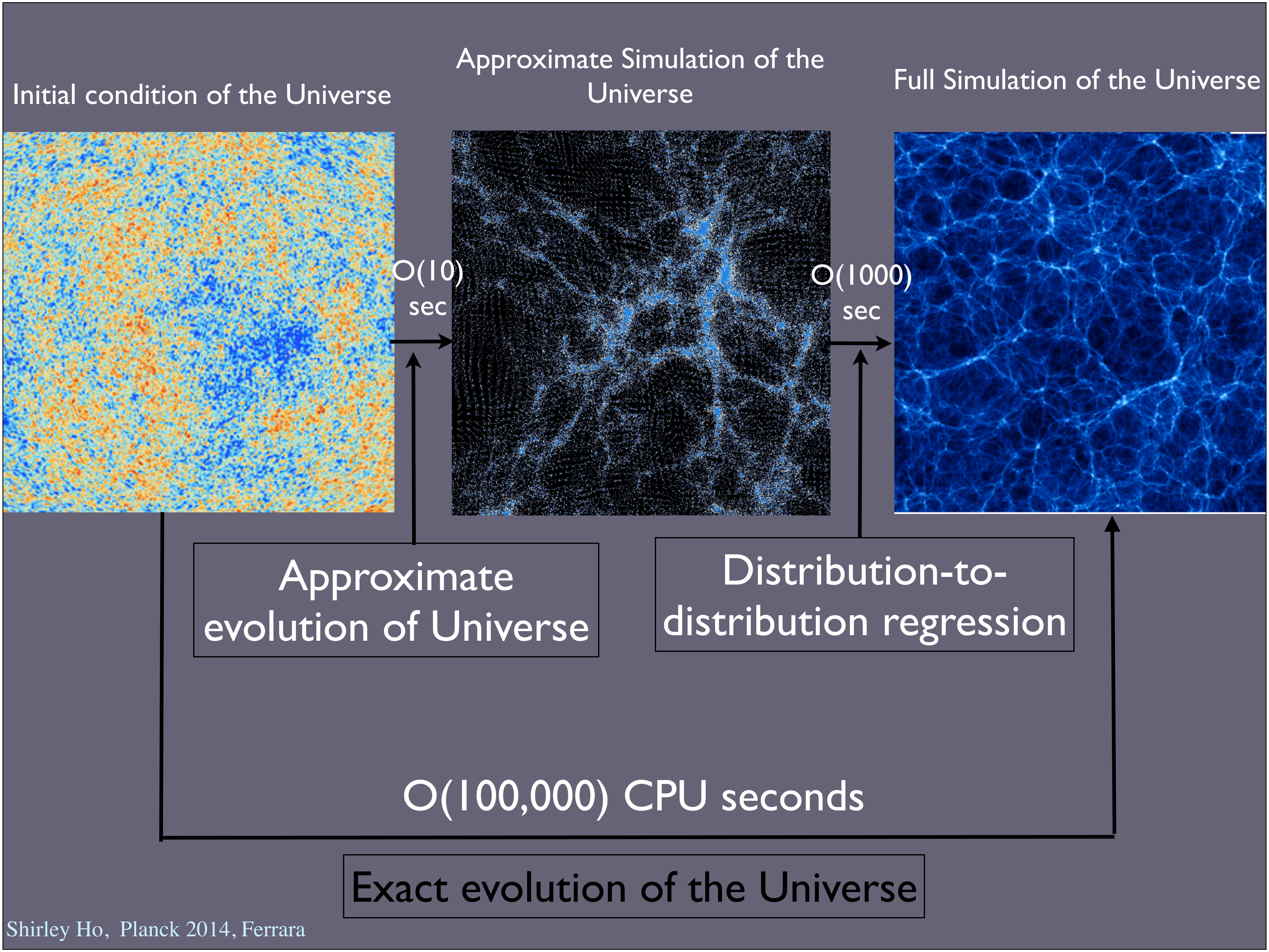
Full Simulation of the Universe



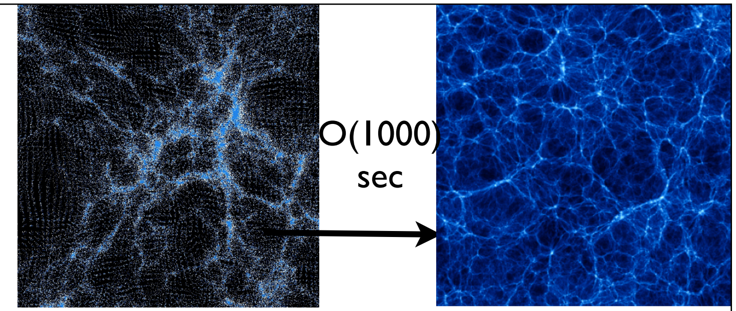
O(100,000) CPU seconds

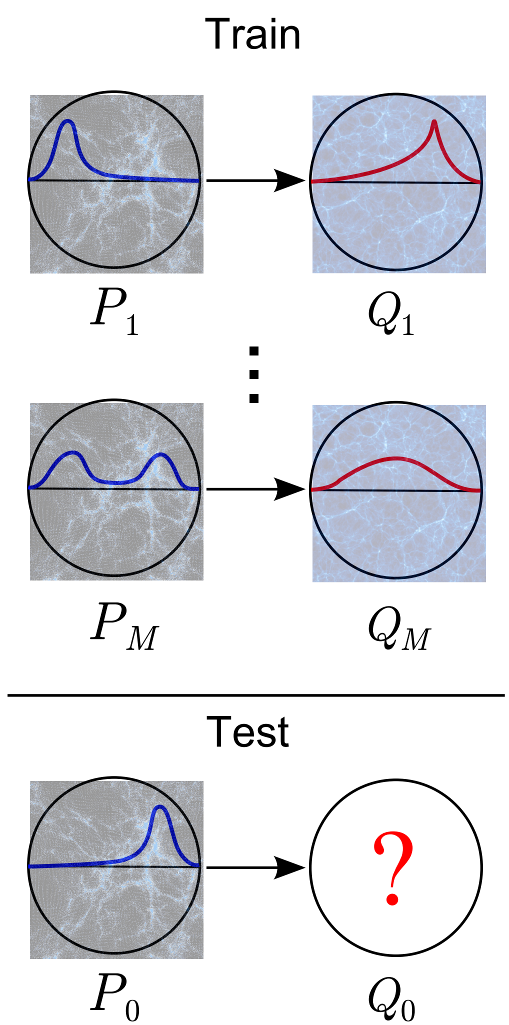
Exact evolution of the Universe

Shirley Ho, Planck 2014, Ferrara


Initial condition of the Universe
Approximate Simulation of the Universe
Full Simulation of the Universe
O(10) sec
O(1000) sec
Approximate evolution of Universe
Distribution-to-distribution regression
O(100,000) CPU seconds
Exact evolution of the Universe
Shirley Ho, Planck 2014, Ferrara

Distribution-to-distribution regression

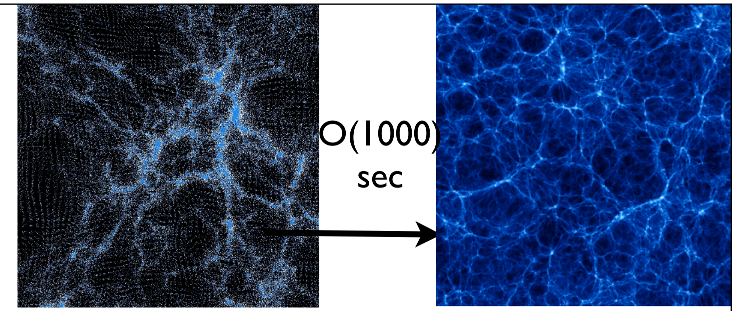


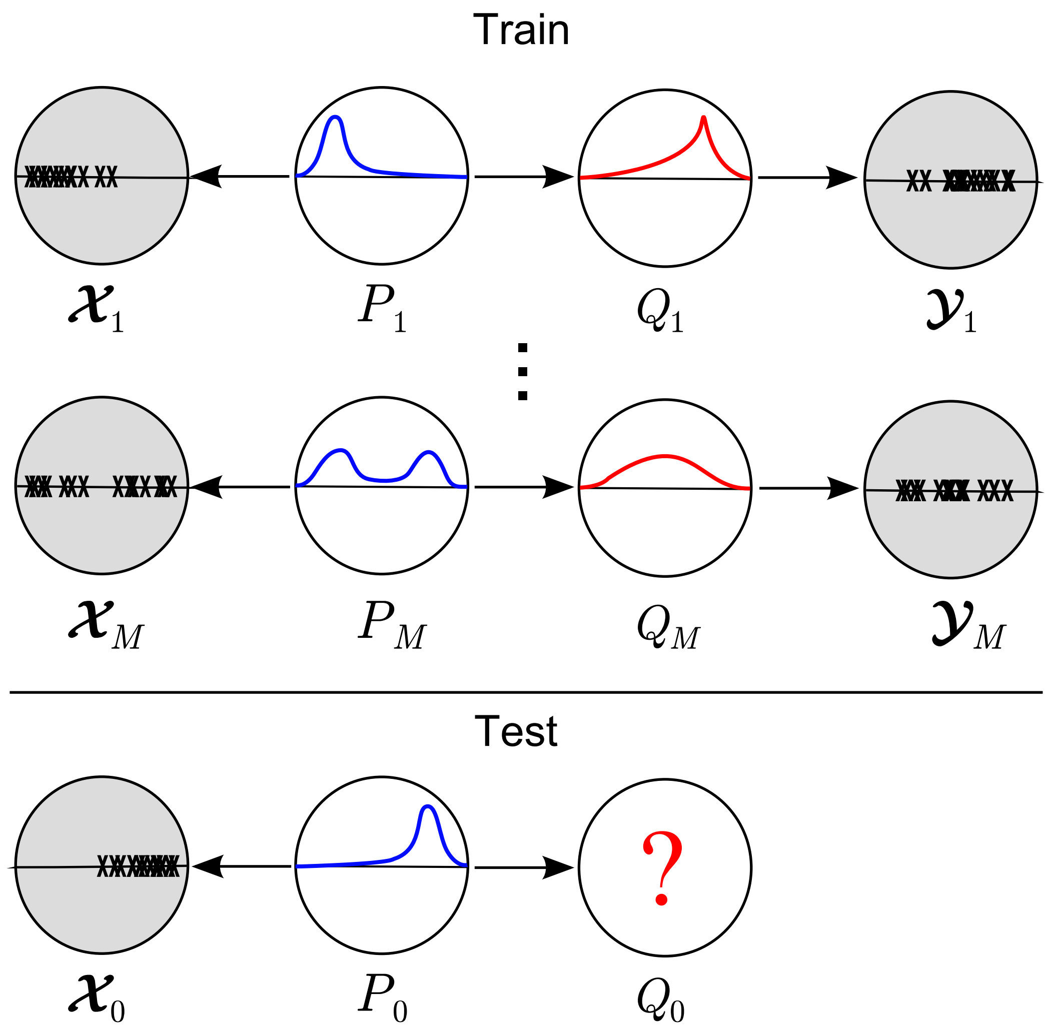


Shirley Ho, Planck 2014, Ferrara

Oliver, Poczos, Schneider et al. 2013

Distribution-to-distribution regression

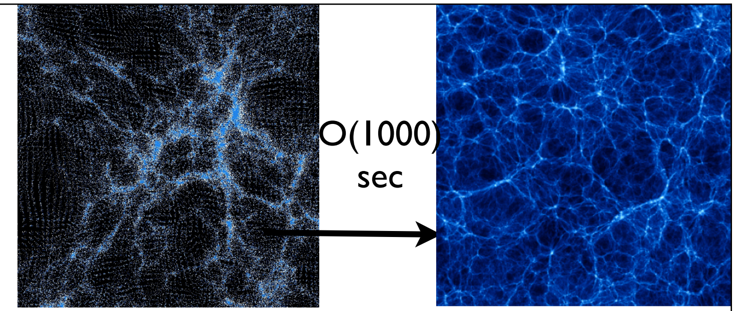


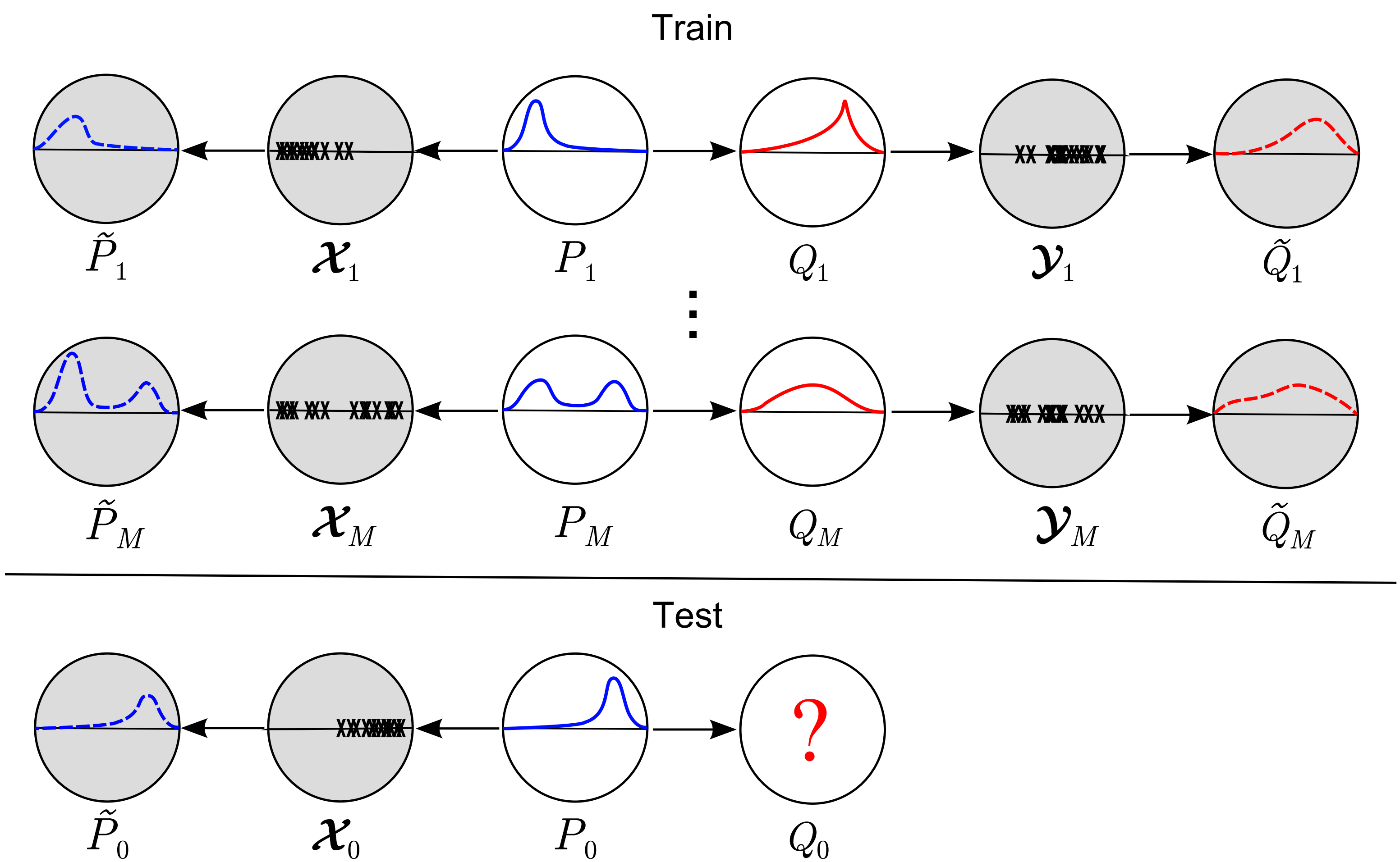


Shirley Ho, Planck 2014, Ferrara

Oliver, Poczos, Schneider et al. 2013

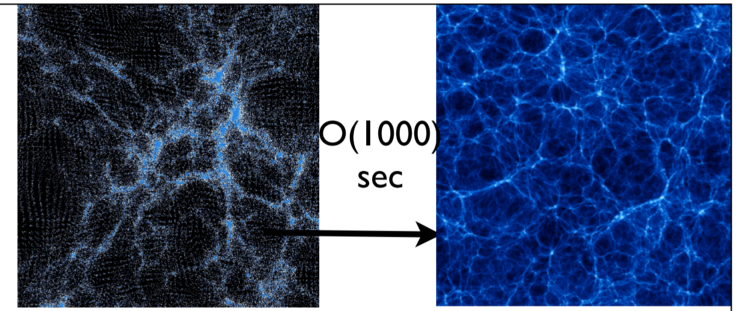
Distribution-to-distribution regression

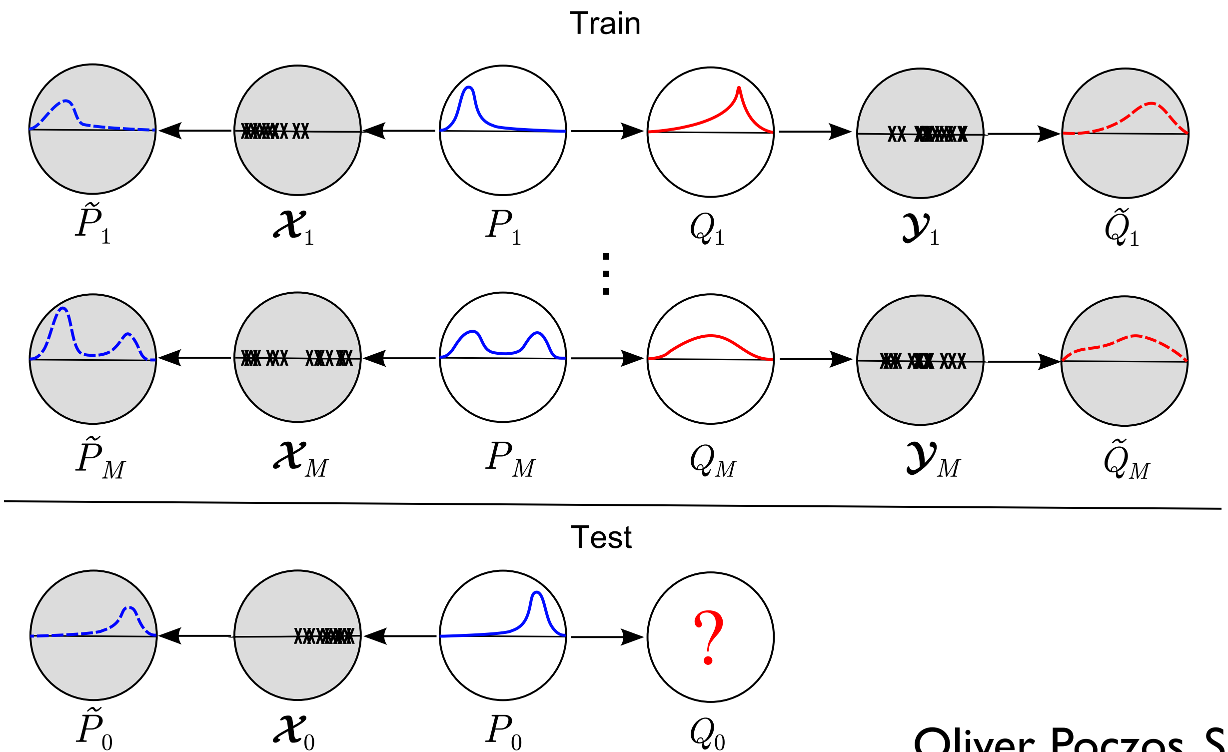




Oliver, Poczos, Schneider et al. 2013

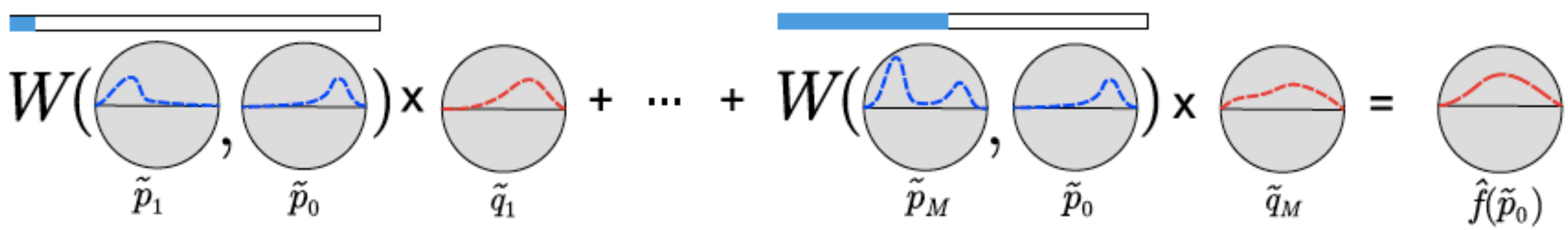
Distribution-to-distribution regression





Oliver, Poczos, Schneider et al. 2013

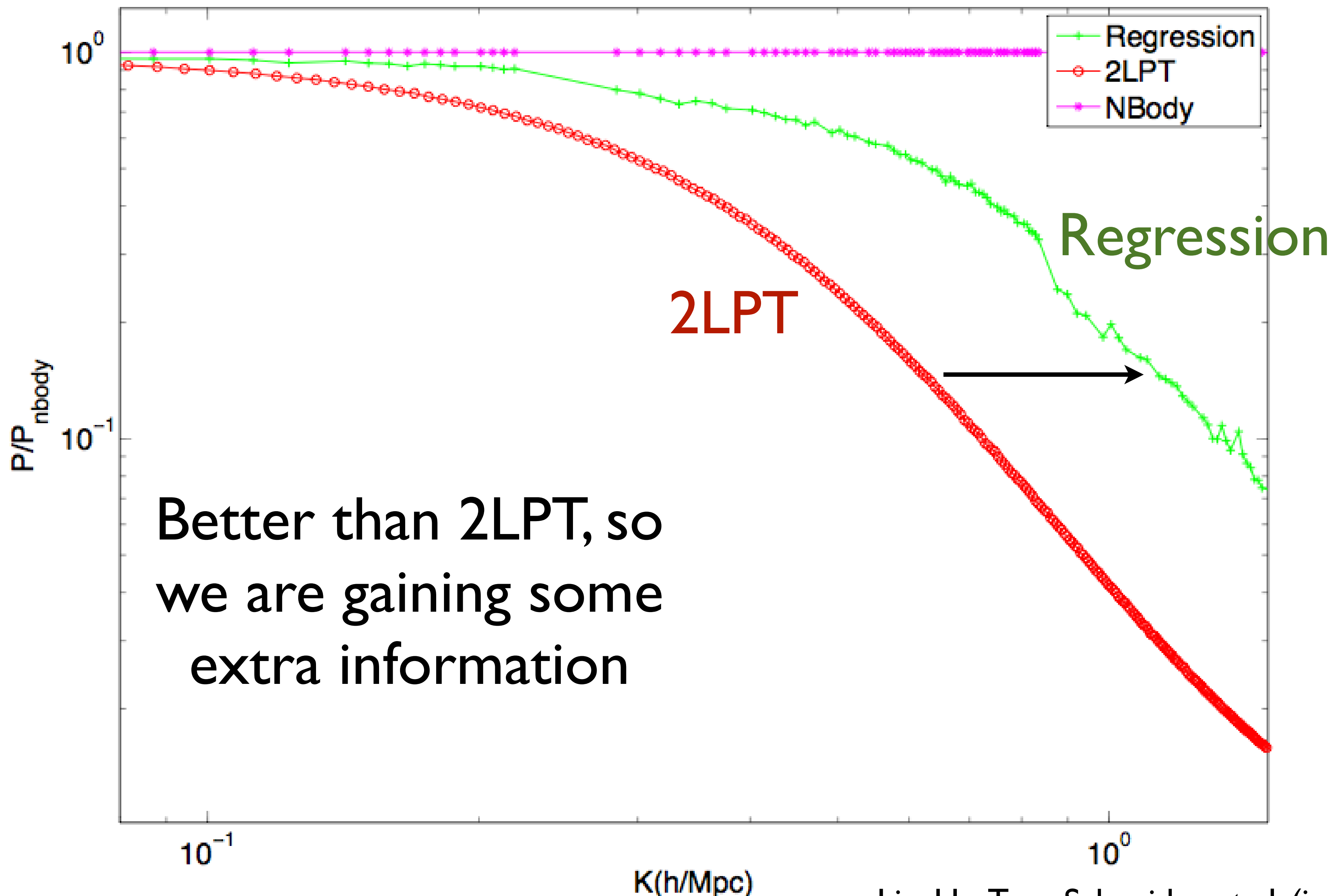
Find the inner product between test input and each training input



Quick Checks : Ratio of power-spectra

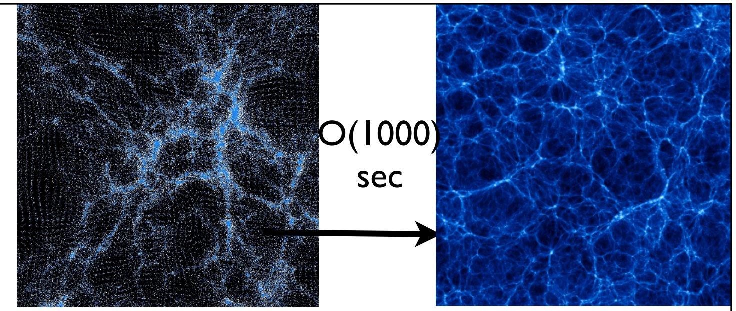
Take the power-spectrum of N-body boxes,
2LPT boxes and Regressed boxes,
and divide them by the N-body P(k)

Ratio of power-spectra



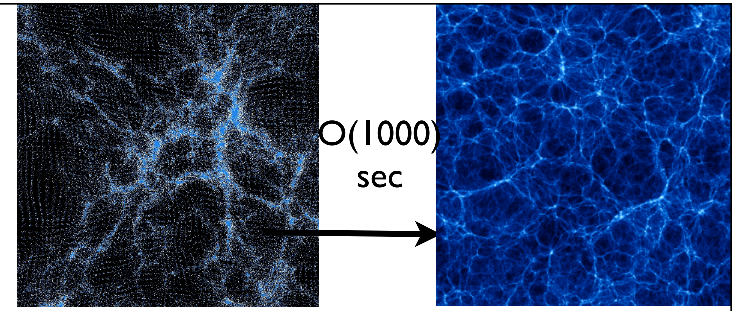
Liu, Ho, Trac, Schneider et al. (in prep)

Distribution-to-distribution regression



- After regression, we only have probability density distribution, not the particles.
- We need to "put the particles back"
  - Randomly select particles
  - Check its Probability density distribution, and compare against a random draw between [0,1]

Distribution-to-distribution regression



Halo Mass Function

FoF with same linking length

Theory

N-body

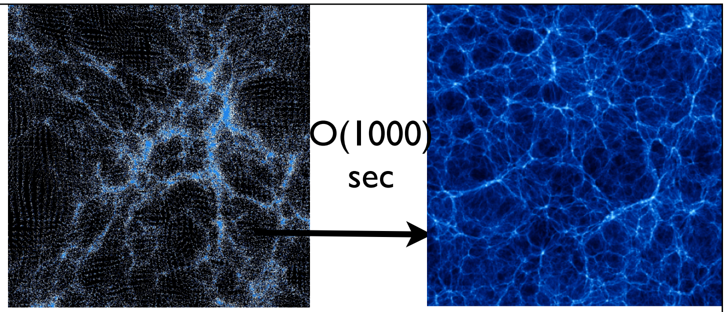
2LPT

First try

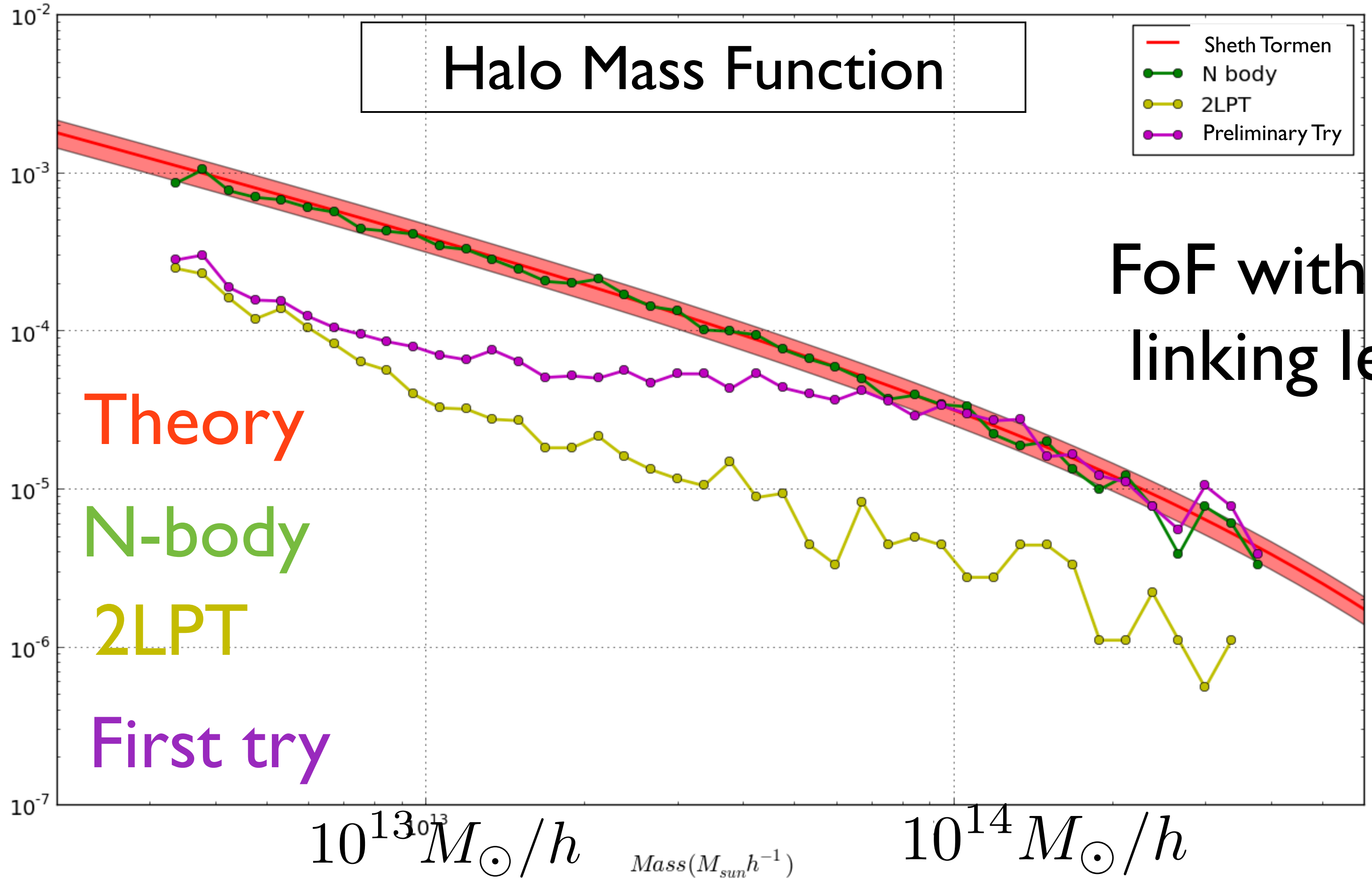
$10^{13} M_\odot/h$ $10^{14} M_\odot/h$

Liu, Ho, Trac, Schneider et al. (in prep)

Distribution-to-distribution regression

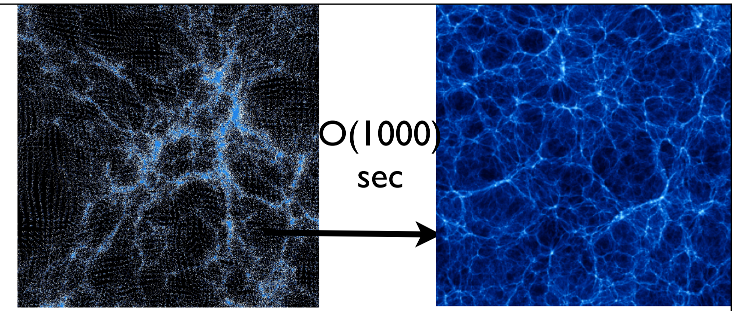


dn/dln(M) vs M at z =0



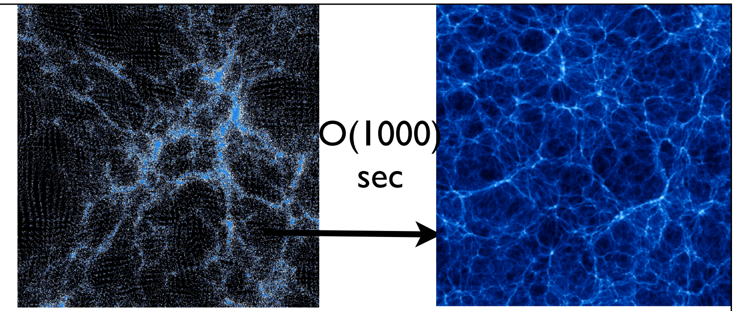
Liu, Ho, Trac, Schneider et al. (in prep)

Distribution-to-distribution regression



- After regression, we only have Probability density distribution, not the particles.
- We need to "put the particles back"
- First Try : Uses rejection method with randomly chosen points.
- New Two steps:
  - Set up random grid to sample the distribution
  - Add extra noise to positions of placed particles

Distribution-to-distribution regression



Halo Mass Function

Theory

N-body

2LPT

First try

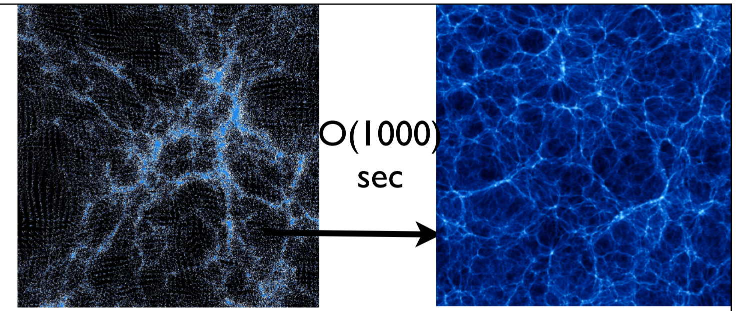
Reducing number of basis

Regression, with random sampling grid

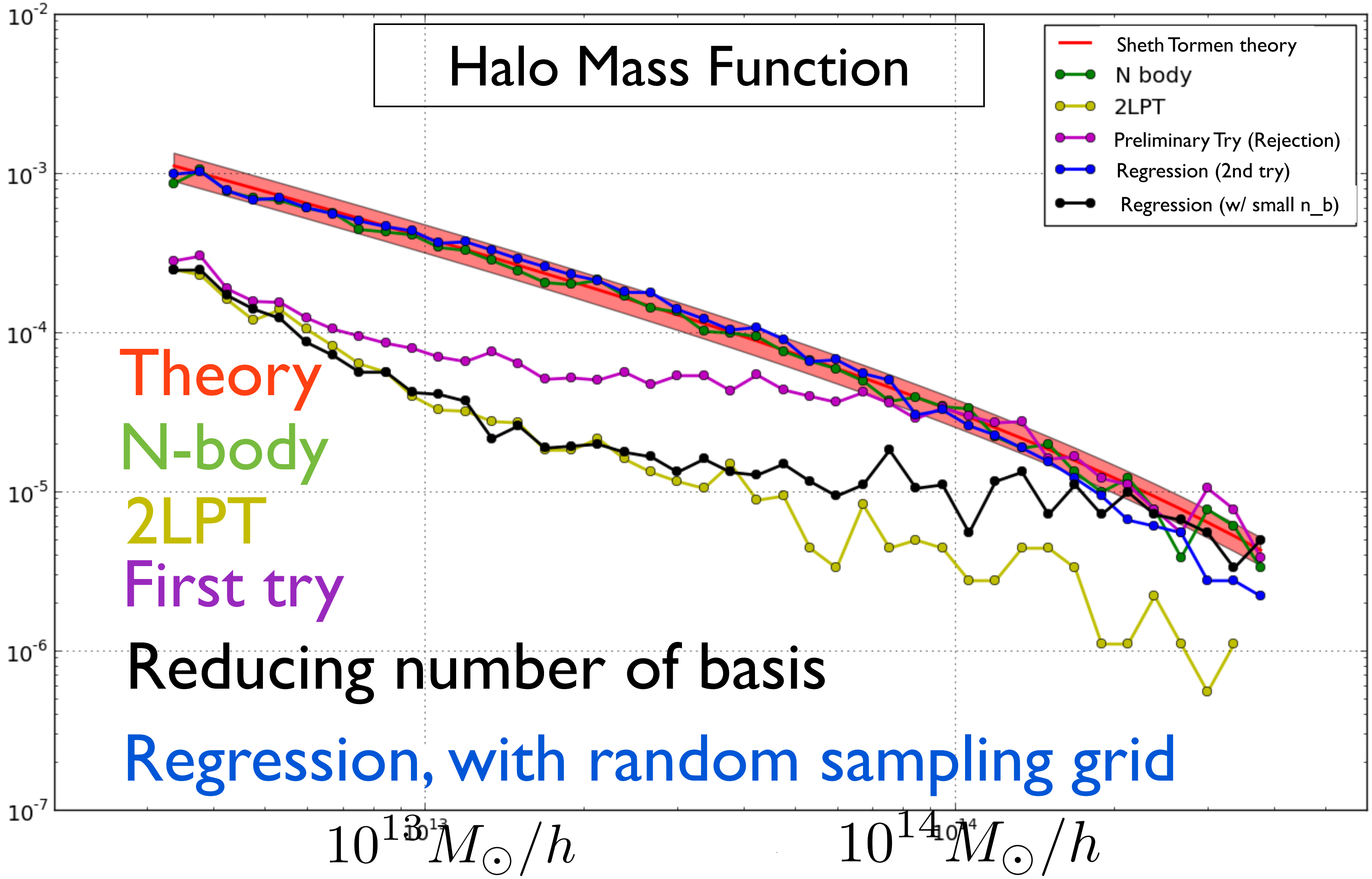
$10^{13} M_{\odot}/h$ $10^{14} M_{\odot}/h$

Liu, Ho, Trac, Schneider et al. (in prep)

Distribution-to-distribution regression



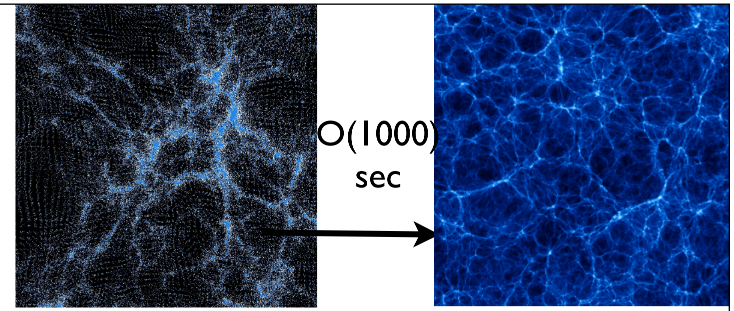
dn/dln(M) vs M at z =0



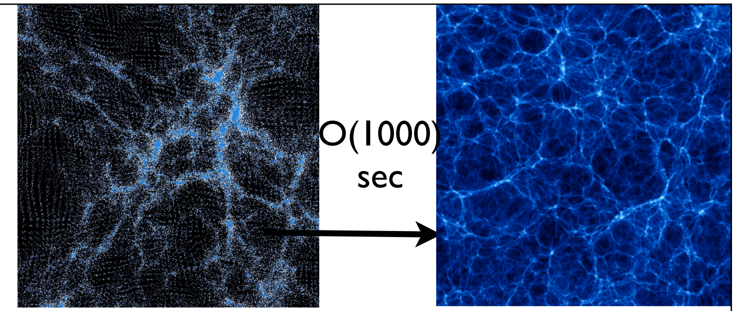
Liu, Ho, Trac, Schneider et al. (in prep)

The harder questions after regression

AKA: *The night (of graduate students) is dark and full of terror*

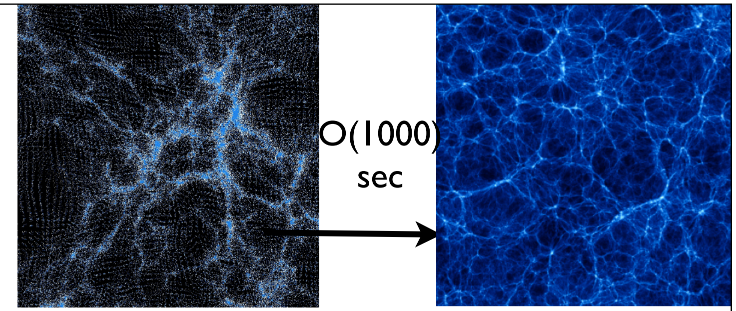


The harder questions after regression



- There must be tuning parameters, what are they?
- How different are we from other fast N-body codes?
- What else can we do to improve this?

The tuning parameters



- 3 Tuning parameters:
  - Sub-boxes Sizes
  - Random Grid Resolution used for repopulating the particles in each sub-boxes
  - Amplitude of gaussian noise added

The tuning parameters:
Nbody sub-box sizes

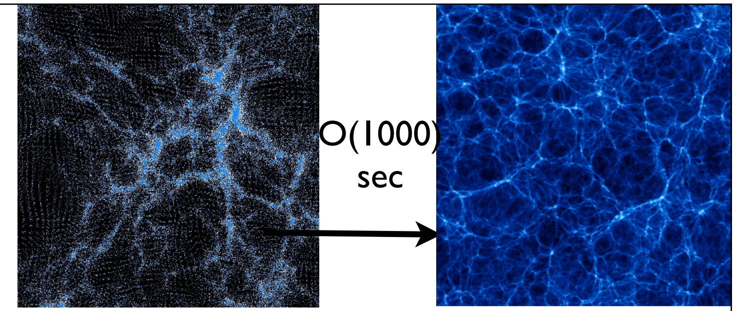


Table 6. Error Table for Changing N body bin size

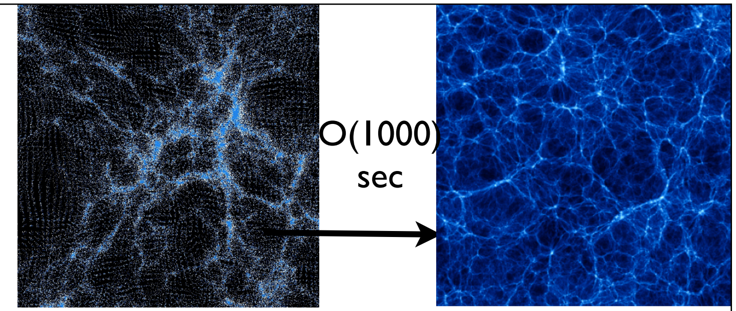
N-body Bin size	Prediction Error
1/2	7.79114
3/4	3.6540
7/8	2.3742

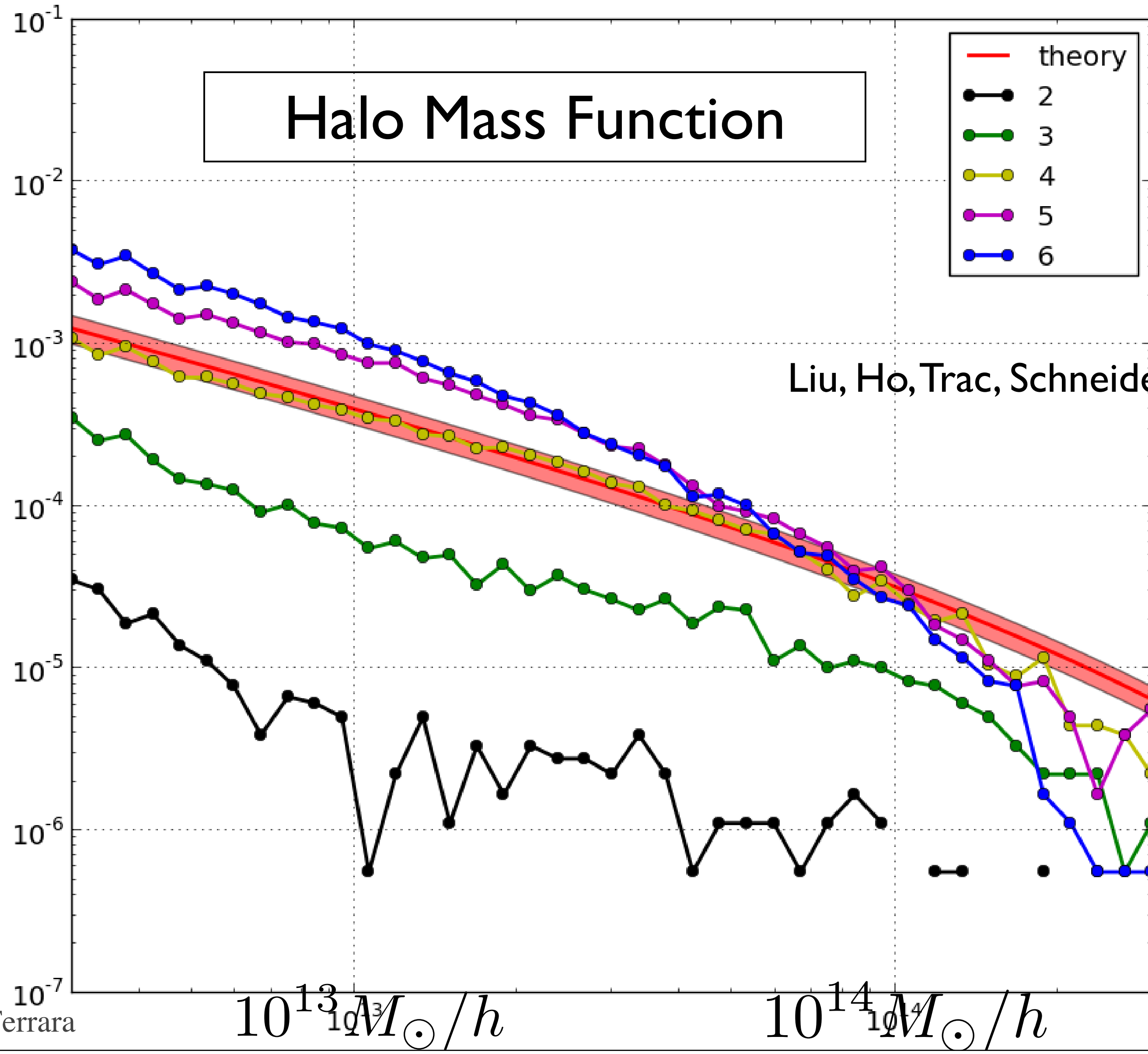
increasing sub-box sizes

Error = sum of the mean square error between the predicted and the true.

Liu, Ho, Trac, Schneider et al. (in prep)

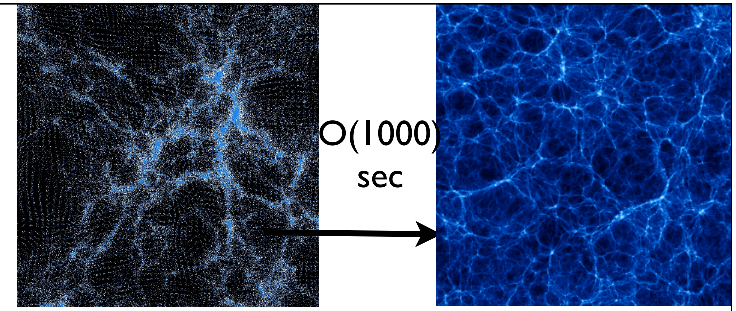
The tuning parameters: grid resolution





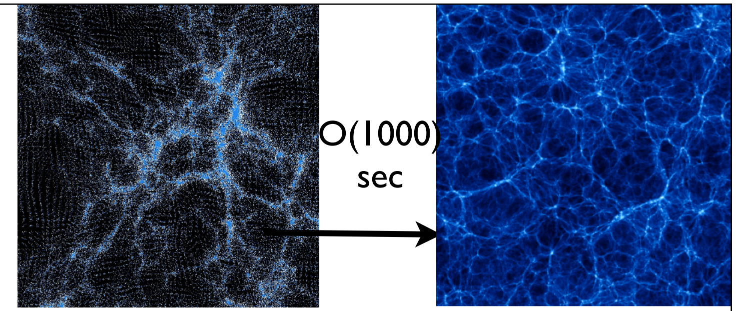
Liu, Ho, Trac, Schneider et al. (in prep)

How different are we from Other Fast Nbody codes?

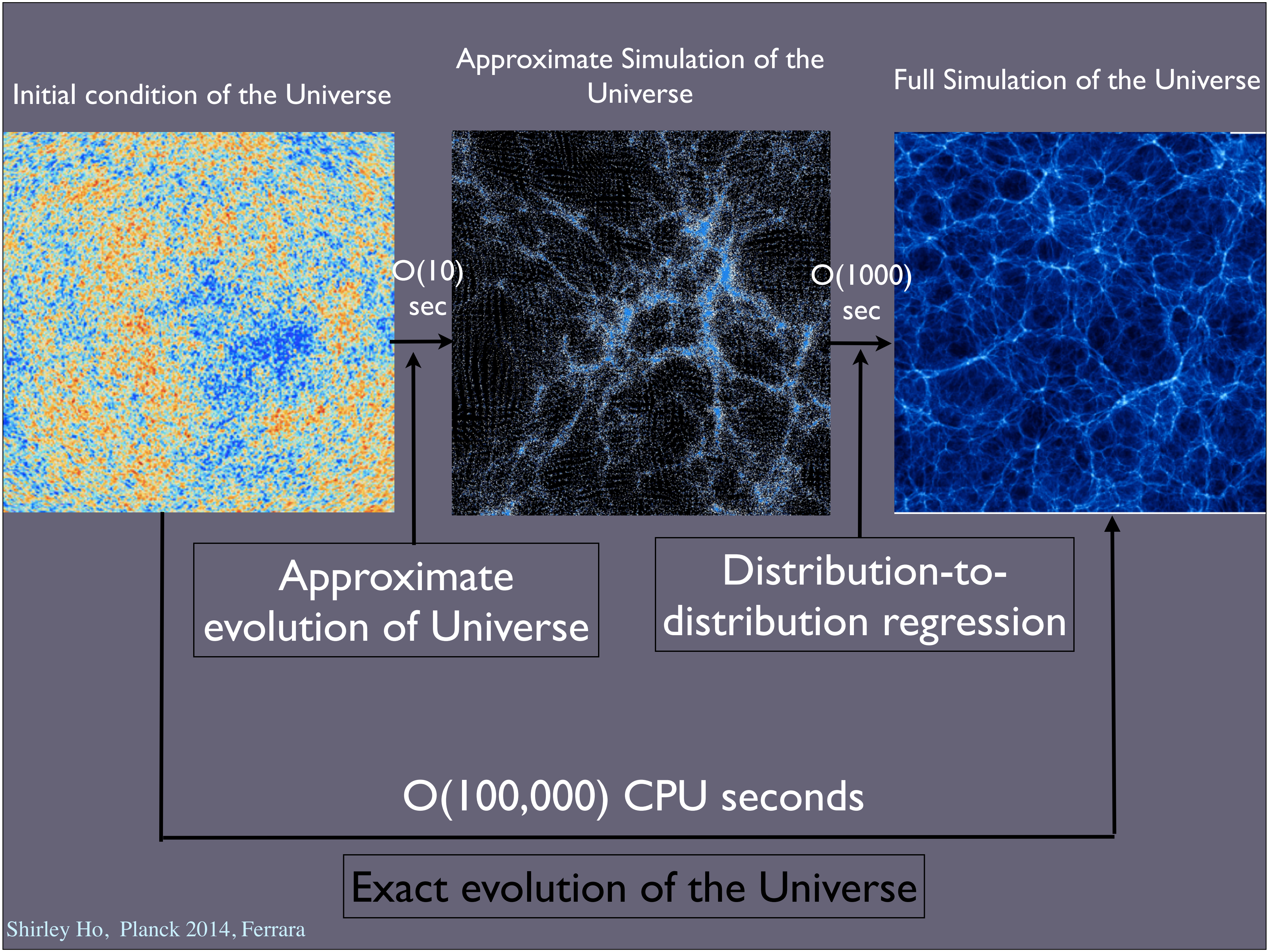


- Cosmological dependence comes from large scales.
- Small scale cosmological dependence are "inherited" from the large scales. No cosmological or astrophysical assumptions needed in the regression. We can predict simulations with cosmological parameters different from our training sets.
- All tuning parameters are either not or only very slightly cosmologically dependent
- Fast! 2 CPU hours for box size of 1 Gpc/h, 2048^3 particles.

Other things we can do to improve

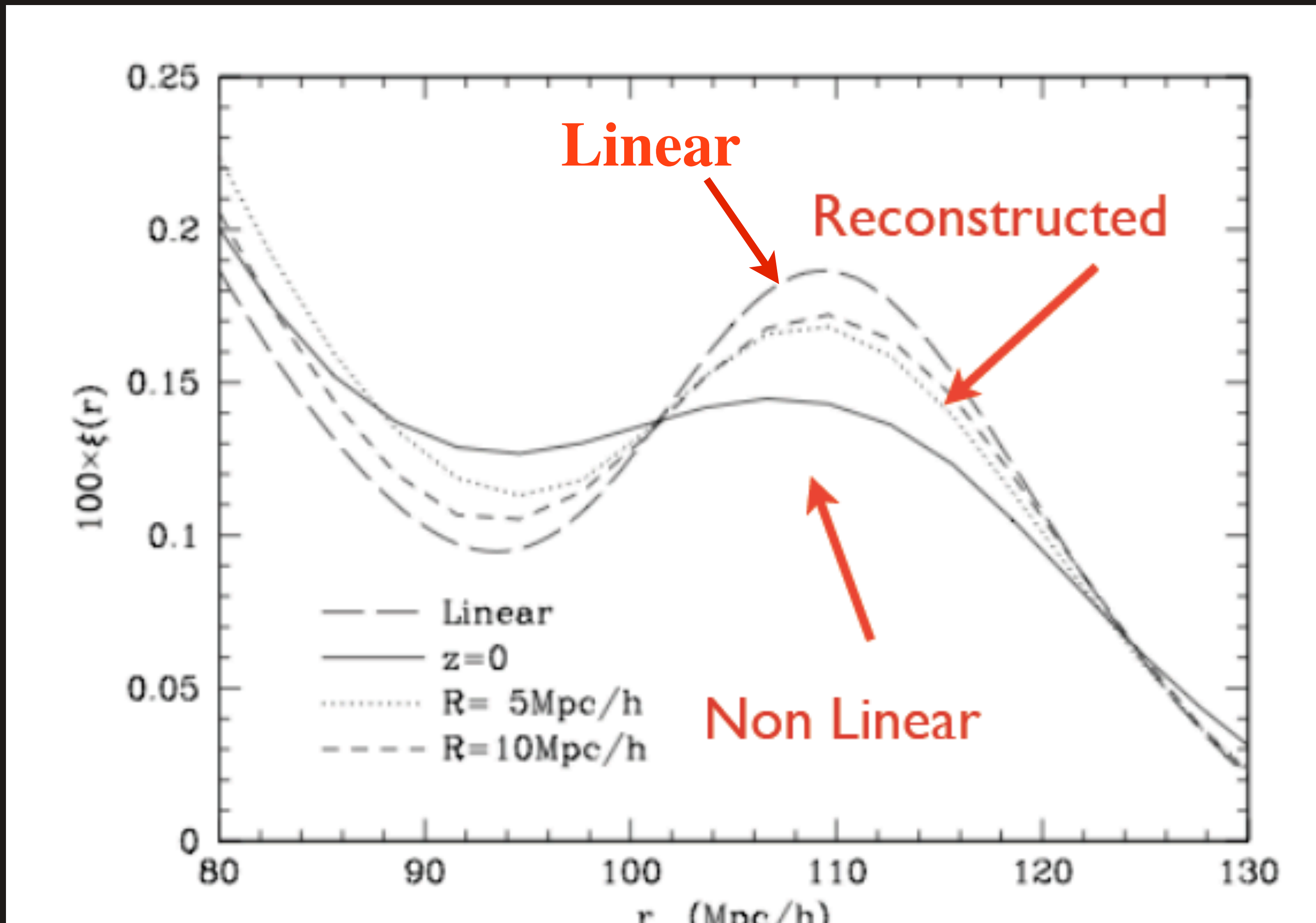


- Adding in velocities as part of the regression
- Investigate exactly how strong is the cosmological dependence for the regression, how different can the cosmological contents of the training set and the test set be?
- Investigate our higher point functions and other summary statistics.


Initial condition of the Universe
Approximate Simulation of the Universe
Full Simulation of the Universe
O(10) sec
O(1000) sec
Approximate evolution of Universe
Distribution-to-distribution regression
O(100,000) CPU seconds
Exact evolution of the Universe
Shirley Ho, Planck 2014, Ferrara

Making Better Measurement ?

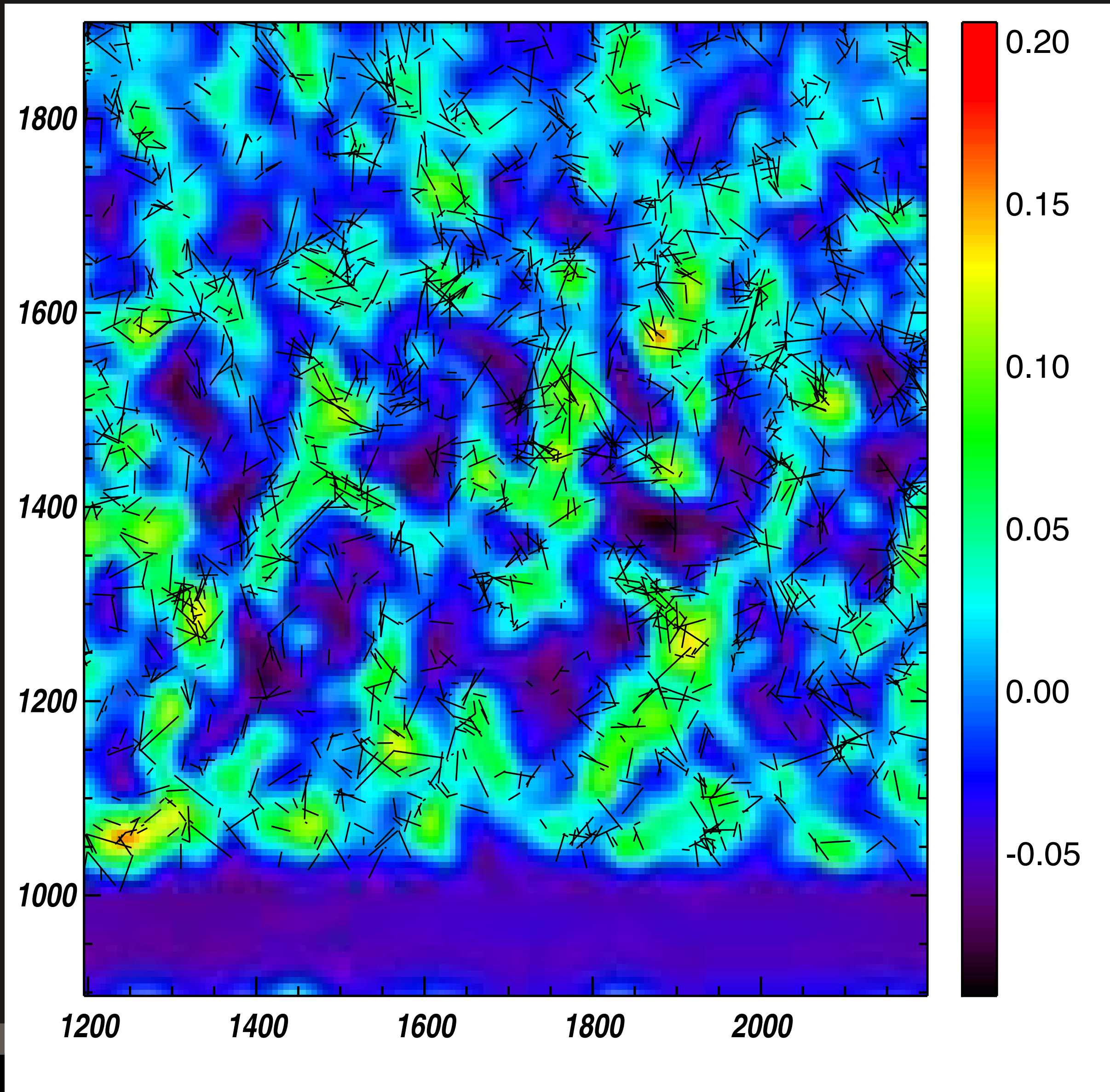
Going back in time!
to make BAO easier to detect



Eisenstein, Seo, Sirko, Spergel, 2007;
Noh, White & Padmanabhan 2009; Padmanabhan et al. 2012

SDSS III

Simulations for BOSS

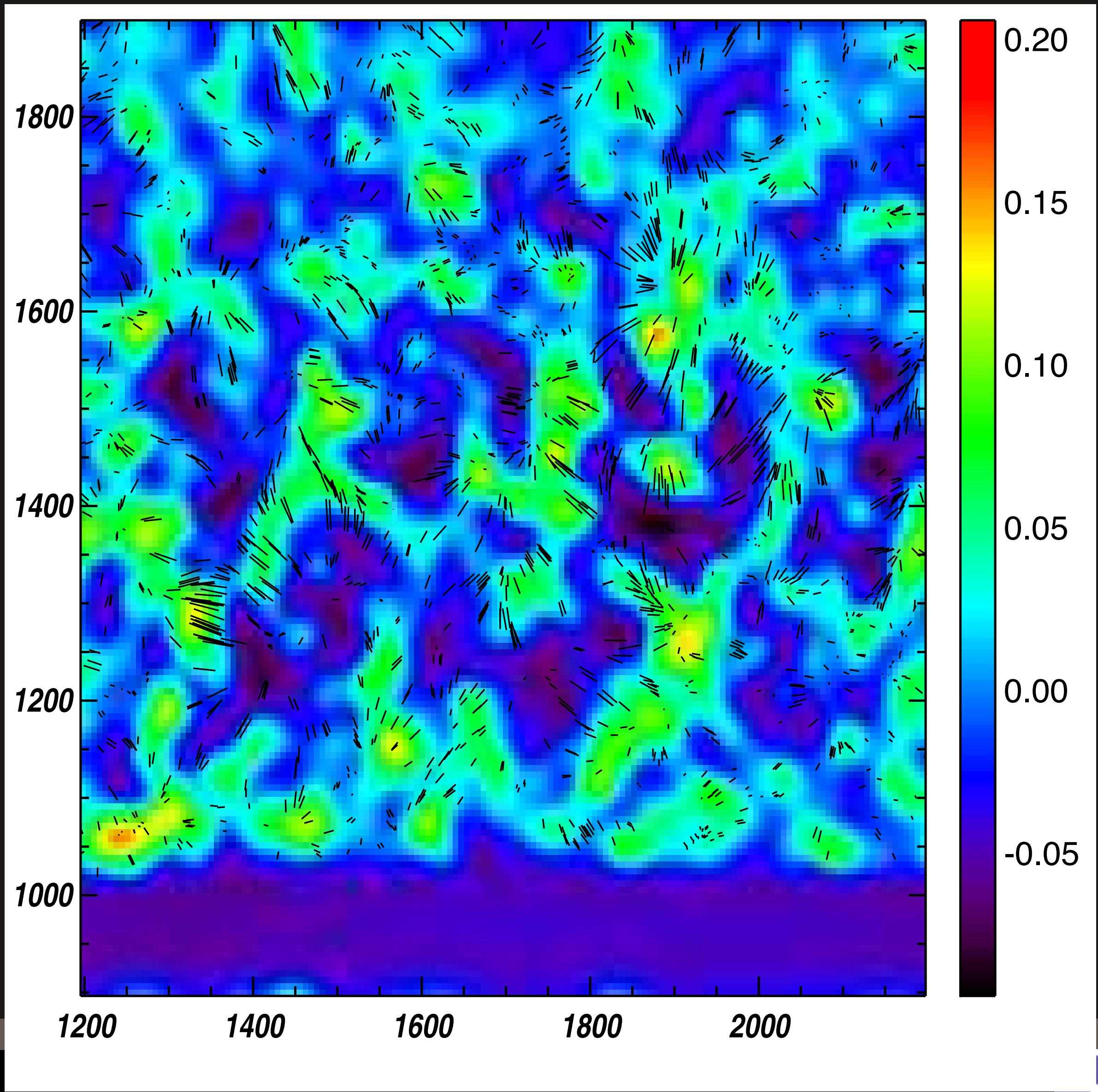


true velocities in simulations depends on information from many scales

Simulation by Martin White, Jeremy Tinker, Cameron McBride

Vargas, Ho, + BOSS collaboration 2014

Standard reconstruction for BOSS

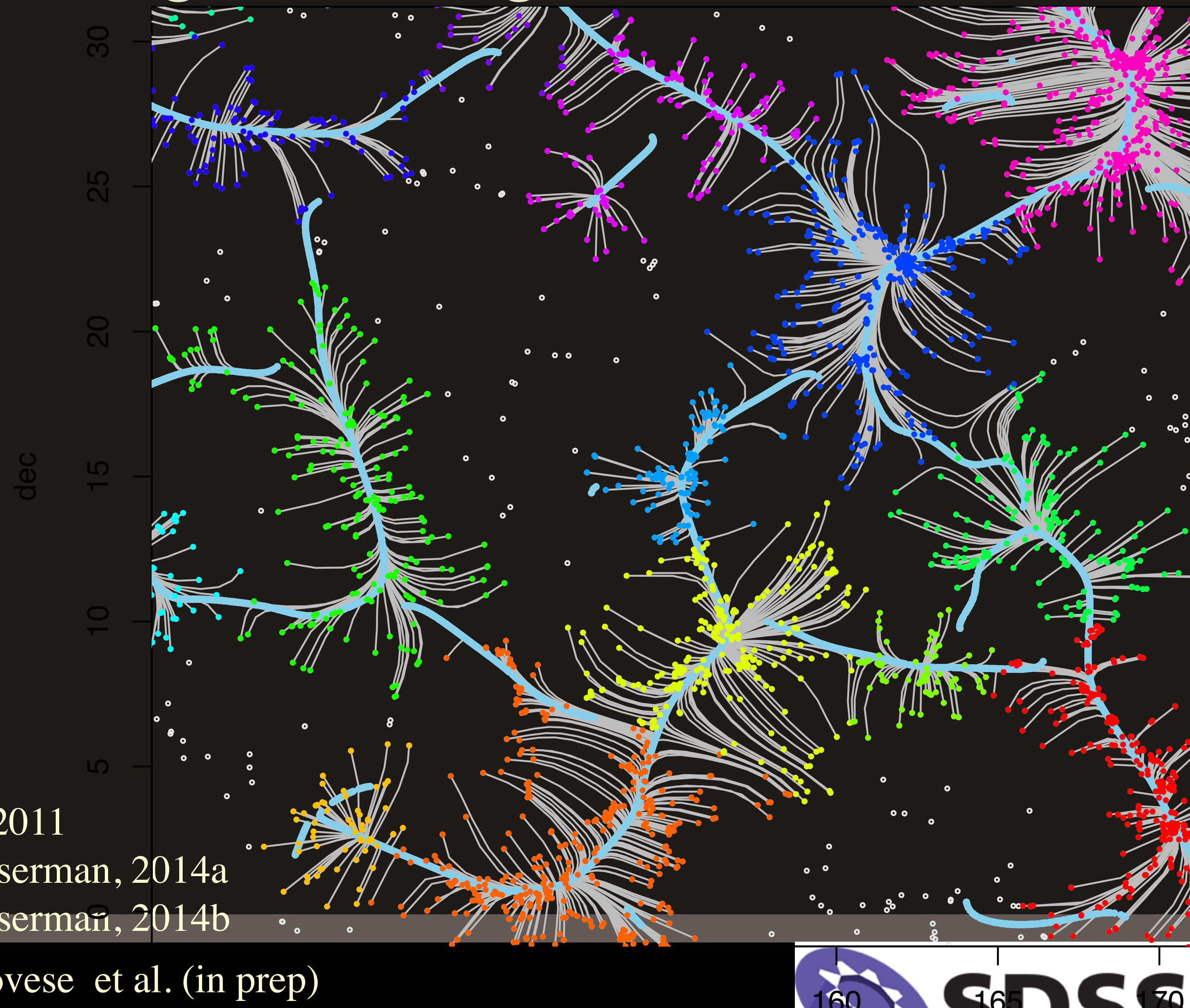


Captures the semi non-linear to linear scale.

Simulation by Martin White, Jeremy Tinker, Cameron McBride

Vargas, Ho, + BOSS collaboration 2014

New Method: SuRF
Subspace Ridge Finding

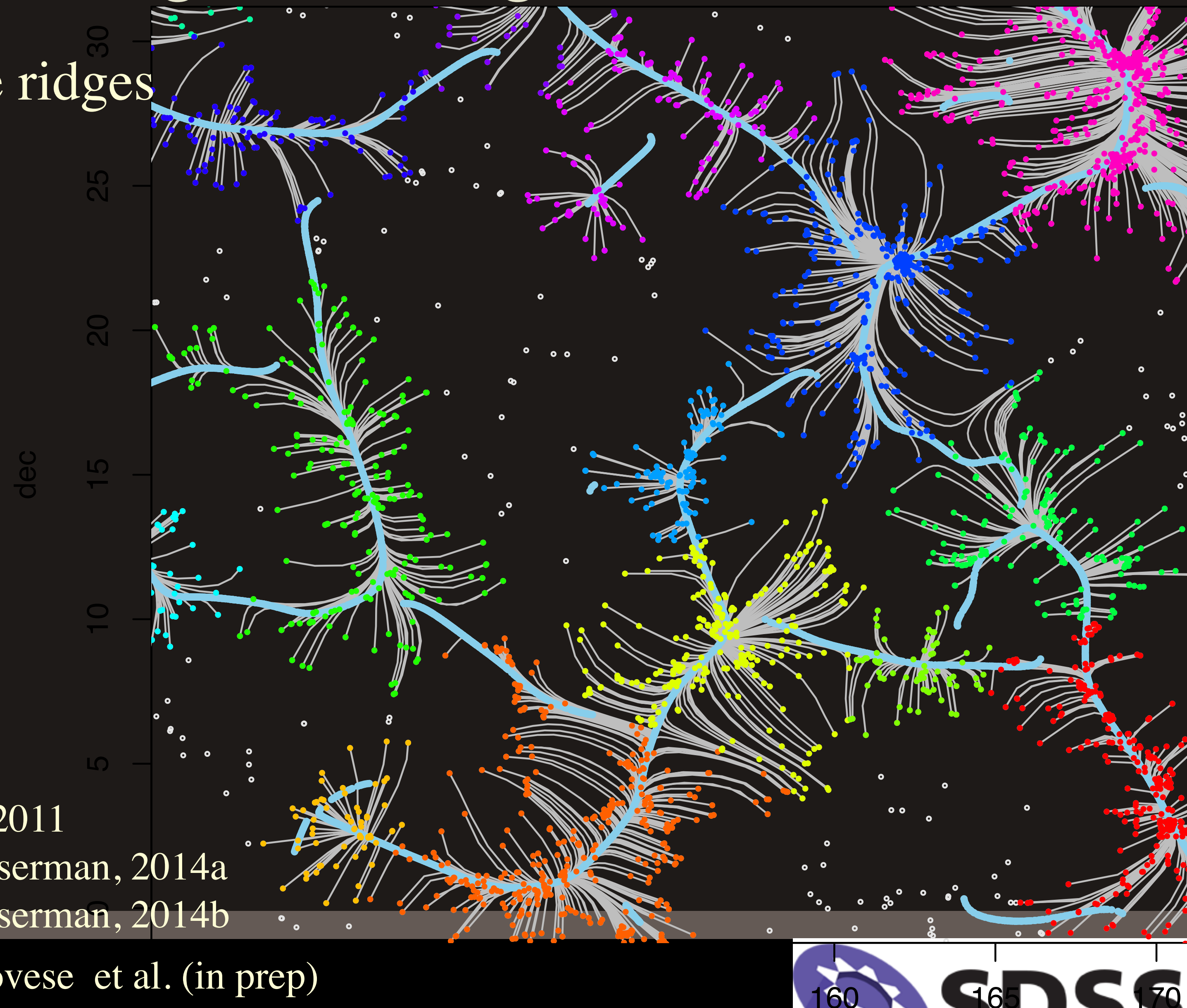


Ozertem & Erdogmus, 2011
Chen, Genovese & Wasserman, 2014a
Chen, Genovese & Wasserman, 2014b

Vargas, Ho, Chen, Genovese et al. (in prep)

New Method: SuRF
Subspace Ridge Finding

- A way to locate ridges in density field



Ozertem & Erdogmus, 2011
Chen, Genovese & Wasserman, 2014a
Chen, Genovese & Wasserman, 2014b

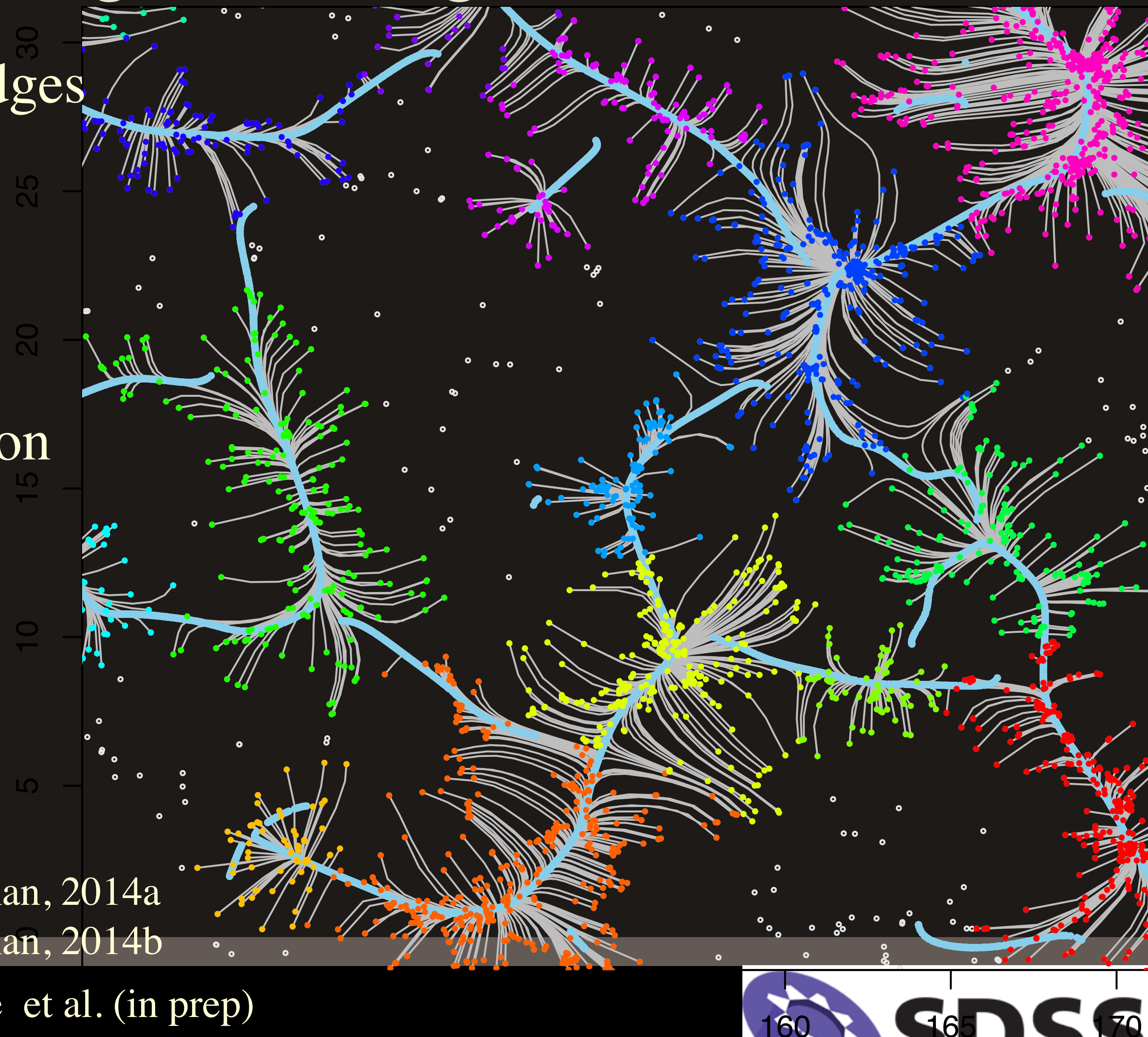
Vargas, Ho, Chen, Genovese et al. (in prep)

SDSS III

New Method: SuRF
Subspace Ridge Finding

- A way to locate ridges in density field

- Effectively find the way of steepest ascent in configuration space



Ozertem & Erdogmus, 2011
Chen, Genovese & Wasserman, 2014a
Chen, Genovese & Wasserman, 2014b

Vargas, Ho, Chen, Genovese et al. (in prep)

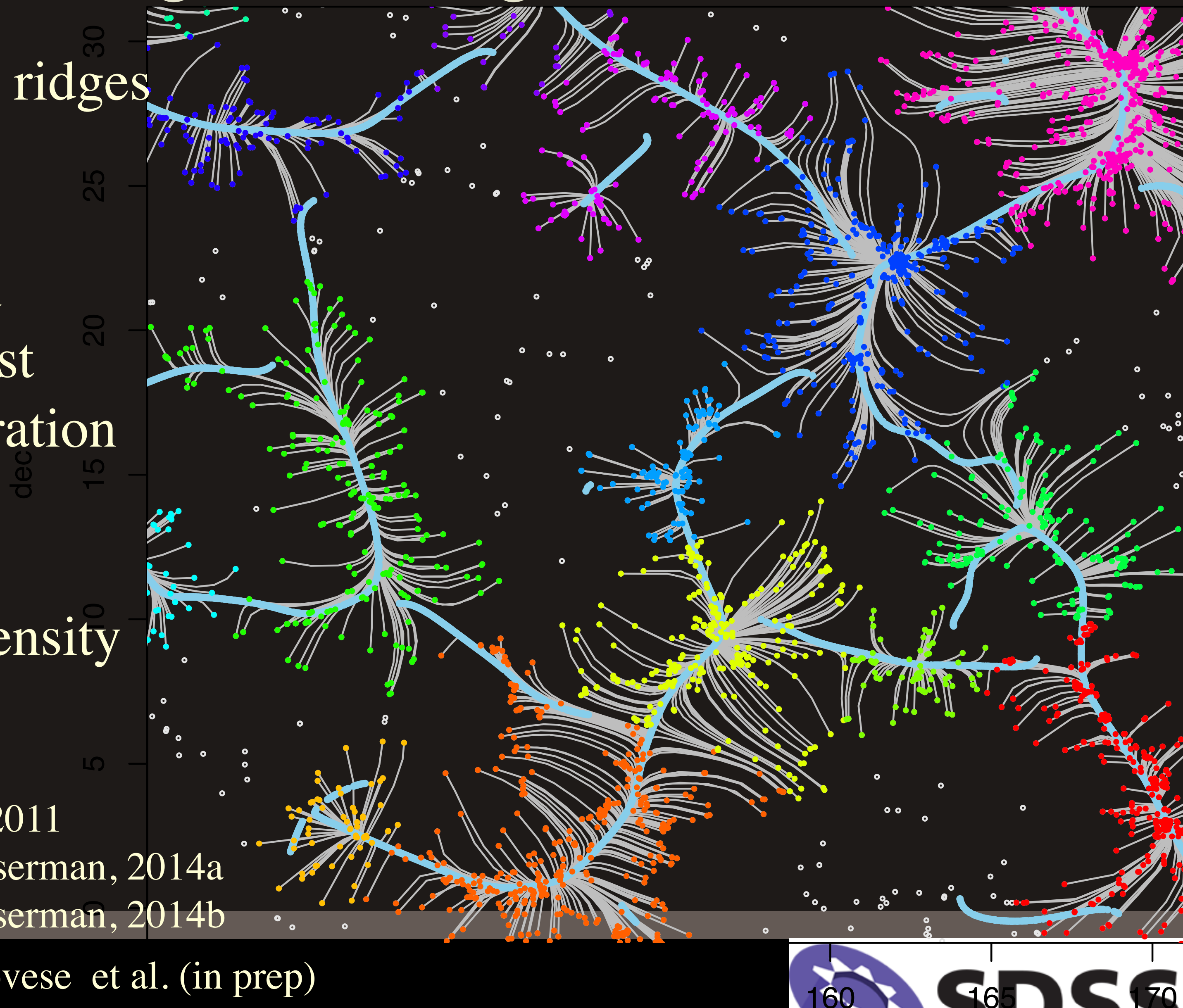
SDSS III

New Method: SuRF
Subspace Ridge Finding

- A way to locate ridges in density field

- Effectively find the way of steepest ascent in configuration space

- Good at high density region.



Ozertem & Erdogmus, 2011
Chen, Genovese & Wasserman, 2014a
Chen, Genovese & Wasserman, 2014b

Vargas, Ho, Chen, Genovese et al. (in prep)

SDSS III

Combining the standard and new

- Since the SuRF method is good at high density region, but not so good at low density region.
- We combine the standard reconstruction method with the new method by weighting the two methods to optimize the “similarity” between the predicted trajectory and the true trajectory of the particles.
- Important as velocity field by itself is a useful probe of the potential of the Universe. Eg. kSZ effect when combined with CMB.

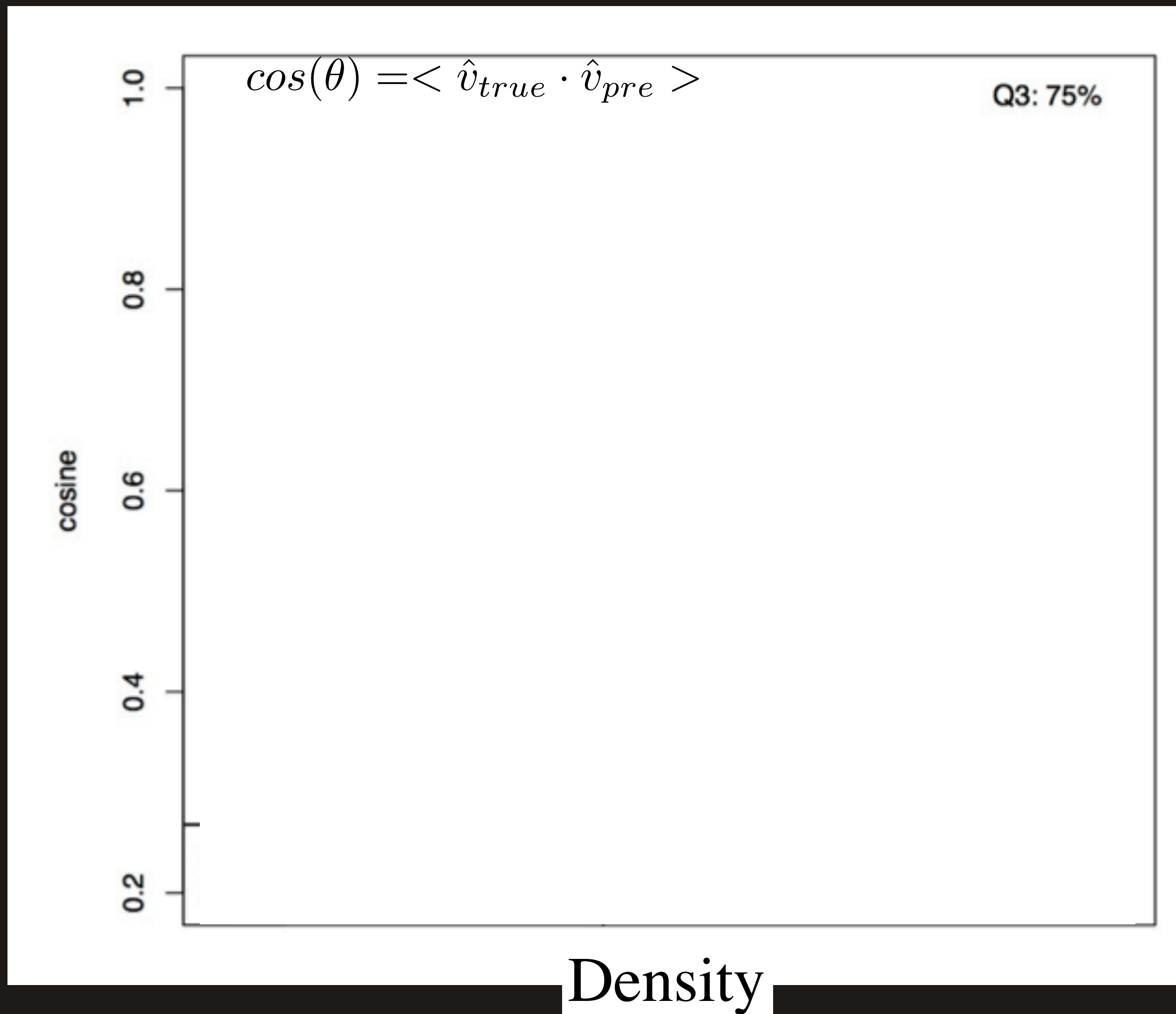
Combining the standard and new

- Since the SuRF method is good at high density region, but not so good at low density region.
- We combine the standard reconstruction method with the new method by weighting the two methods to optimize the "similarity" between the predicted trajectory and the true trajectory of the particles.
- Important as velocity field by itself is a useful probe of the potential of the Universe. Eg. kSZ effect when combined with CMB.

Combining the standard and new

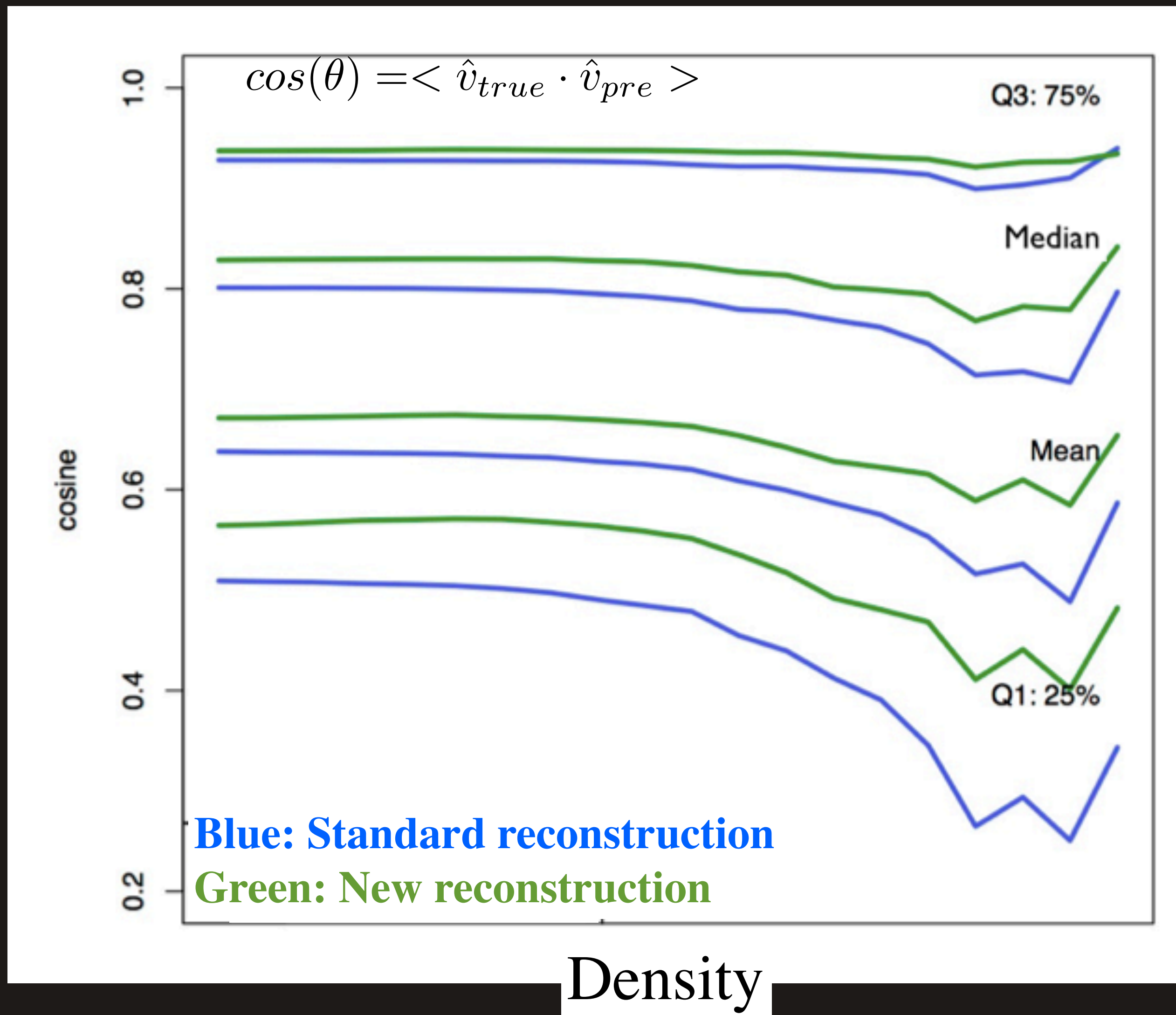
- Since the SuRF method is good at high density region, but not so good at low density region.
- We combine the standard reconstruction method with the new method by weighting the two methods to optimize the “similarity” between the predicted trajectory and the true trajectory of the particles.
- Important as velocity field by itself is a useful probe of the potential of the Universe. Eg. kSZ effect when combined with CMB.

Preliminary results:
Inner Product between True and Predicted velocities



Vargas, Ho, Chen, Genovese et al. (in prep)

Preliminary results:
Inner Product between True and Predicted velocities



Vargas, Ho, Chen, Genovese et al. (in prep)

SDSS III

Conclusion

- BAO has come of age, we can make 1% distance measurement using BAO at multiple redshifts
- This allows us to make quantitative statement of our cosmology AND
- There are many interesting fronts that we can improve upon:
  - Making better predictions of the non-linear Universe
  - Making better measurement of the Universe OR Making better predictions of the velocity field of the Universe

SDSS III