Title: Challenges for large-scale structure theory and data analysis

Speakers: Mikhail Ivanov

Series: Cosmology & Gravitation

Date: November 24, 2020 - 1:00 PM

URL: http://pirsa.org/20110062

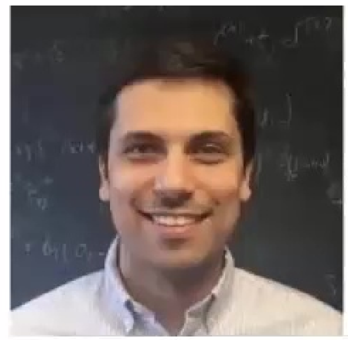
Abstract: In the first part of the talk I will review some recent progress in large-scale structure theory and show how it can be used to measure cosmological parameters from current and future redshift surveys. Then I will discuss some ongoing challenges in the modeling of galaxy clsutering data and covariance matrices. Finally, I will present a systematic calculation of the probability distribution function for the dark matter density field and discuss its potential as a cosmological probe.

Challenges in LSS theory and data analysis

Misha Ivanov (NYU)

O. Philcox

A. Chudaykin



M. Simonovic

M. Zaldarriaga

1907.06666,
1909.05277,
1912.08208,
2002.04035,
2003.08277,
2004.10607,
2006.11235, ++

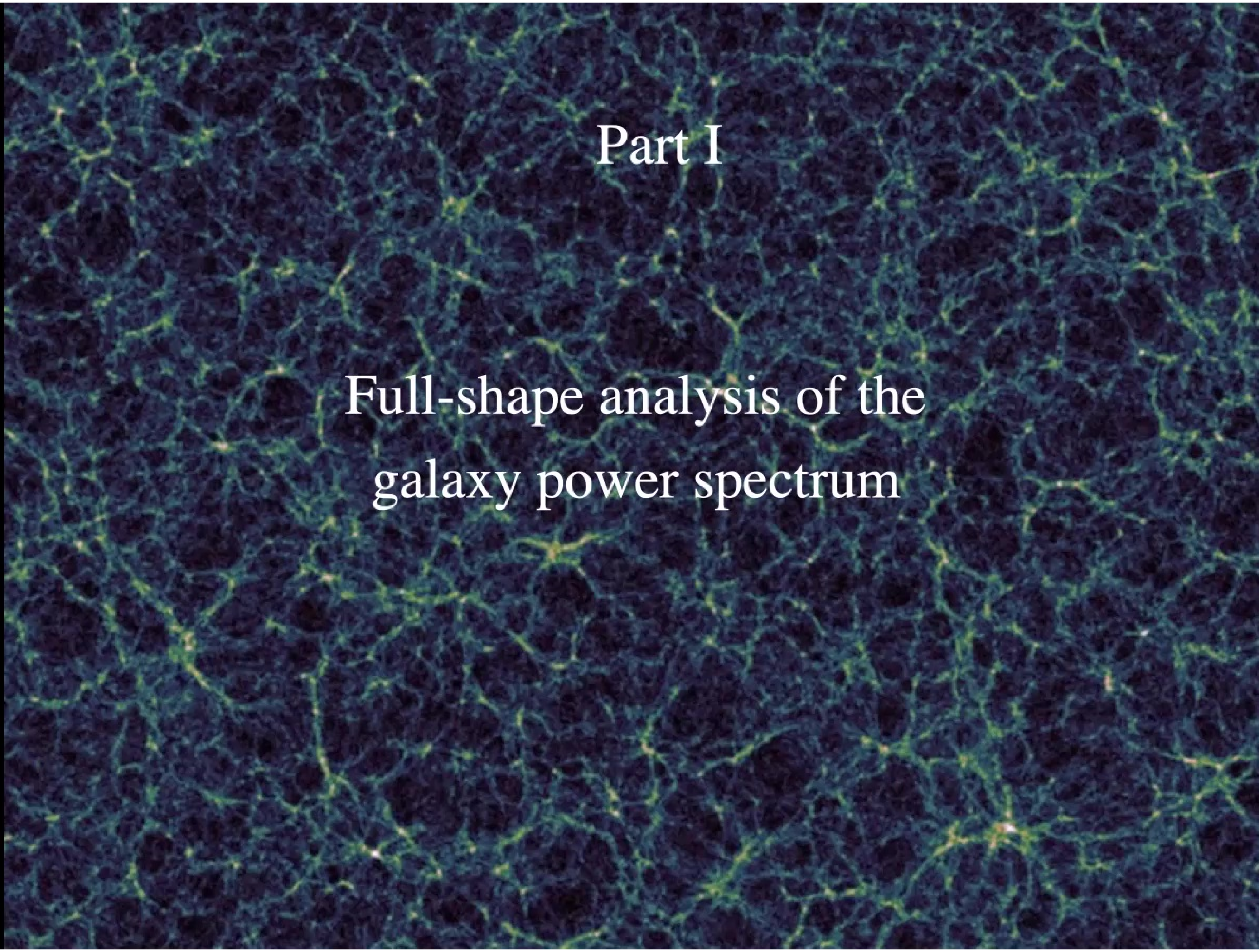
+ M. Schmittfull
+ M. Takada, T. Nischimichi,
+ J. Wadekar, R. Scoccimarro,
+C. Hill, E. McDonough, M. Toomey, S. Alexander,
+L. Senatore, P. Zhang, G. d'Amico
+ A. Kaurov, S. Sibiryakov

In this talk

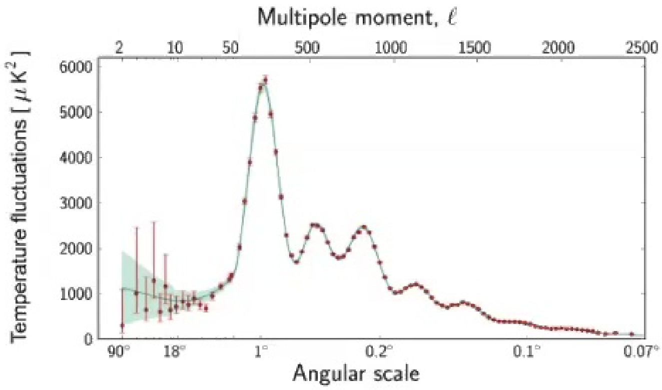
measuring
cosmo. parameters
from BOSS
and Euclid/DESI

current
challenges:
bispectrum,
post-recon BAO,
covariance matrices

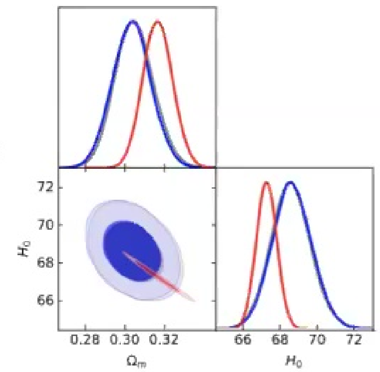
more challenges:
counts-in-cells
statistics


Part I
Full-shape analysis of the
galaxy power spectrum

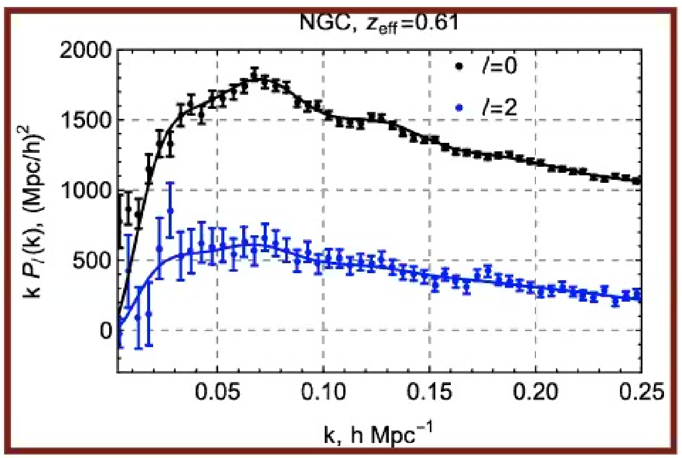
Cosmology from LSS



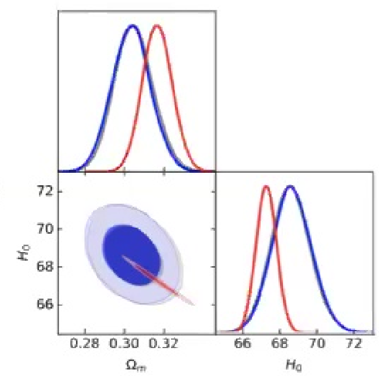
$$+\ C_\ell(H_0, \Omega_m, \ldots) \quad =$$



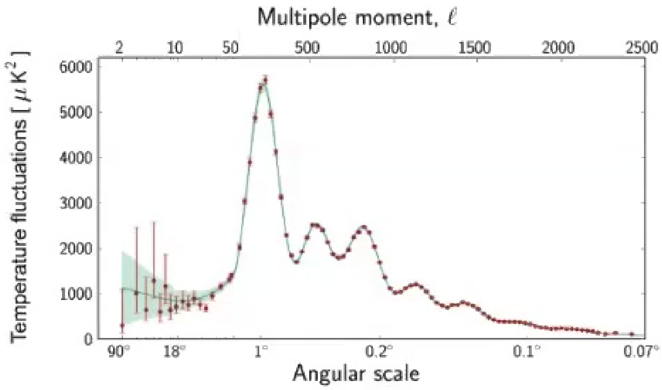
CLASS, CAMB



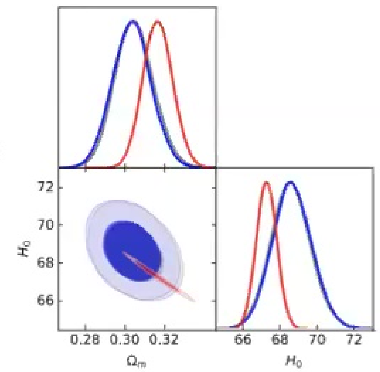
$$+\ P_{\vec{k}}(H_0, \Omega_m, \ldots) =$$



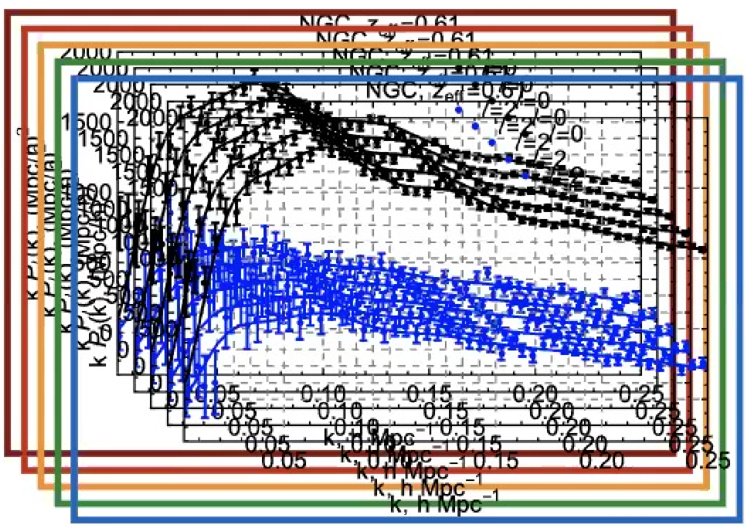
Cosmology from LSS



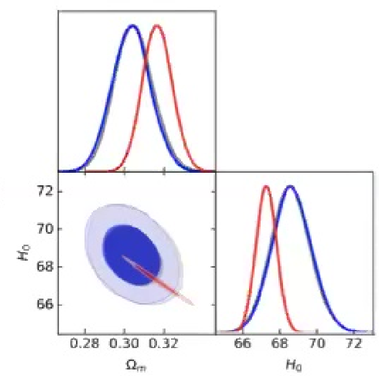
$$+\, C_\ell(H_0, \Omega_m, \ldots) \quad =$$



CLASS, CAMB



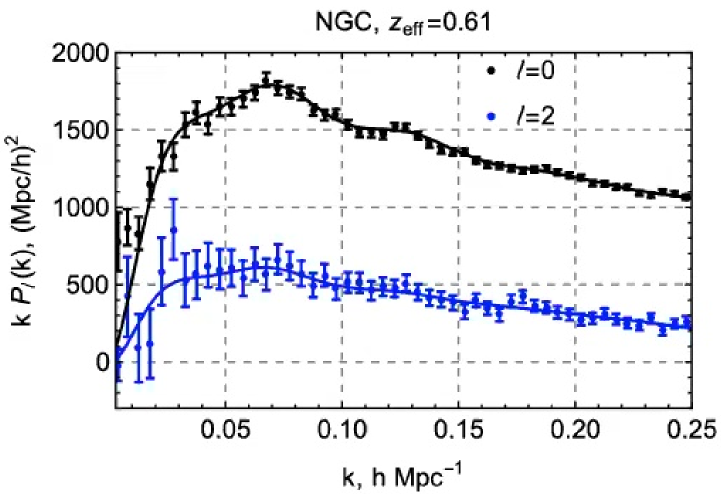
$$+\, P_{\vec{k}}(H_0, \Omega_m, \ldots) =$$



Our pipeline in a nutshell

I. Consistently recompute power spectrum as we vary cosmology (CMB style)
using the full non-linear PT model

Baumann, Nicolis, Senatore, Zaldarriaga'12
++



II. MCMC analysis thanks to FFTLog

McEwen, Fang, Hirata, Blazek (2016)

Schmittfull, Vlah, McDonald (2016)

Simonovic, Zaldarriaga et al. (2017)

III. `CLASS-PT` + Montepython

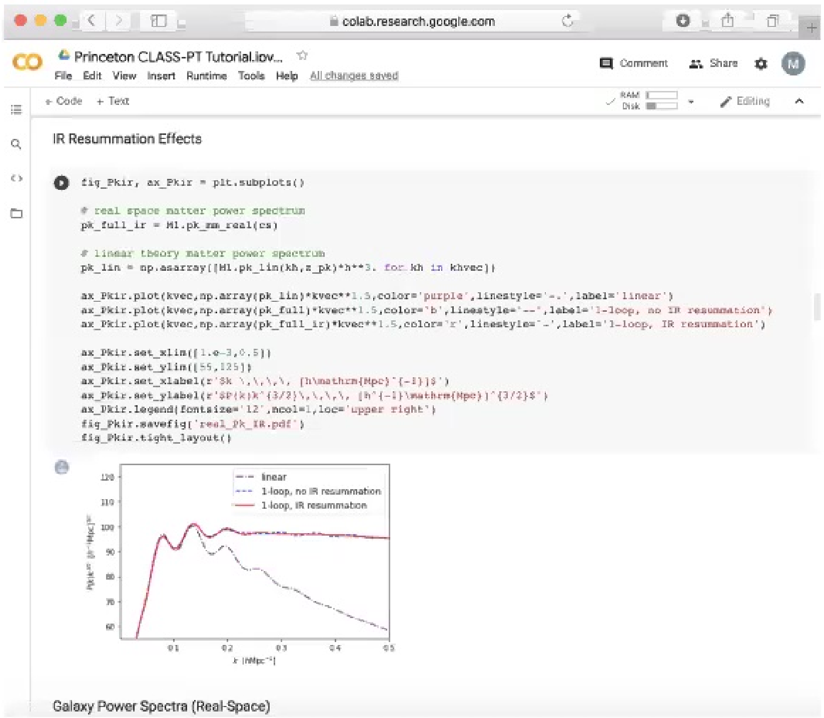
1) User friendly & works out-of-the box
2) Easy scales with # of parameters
3) No hard coding !

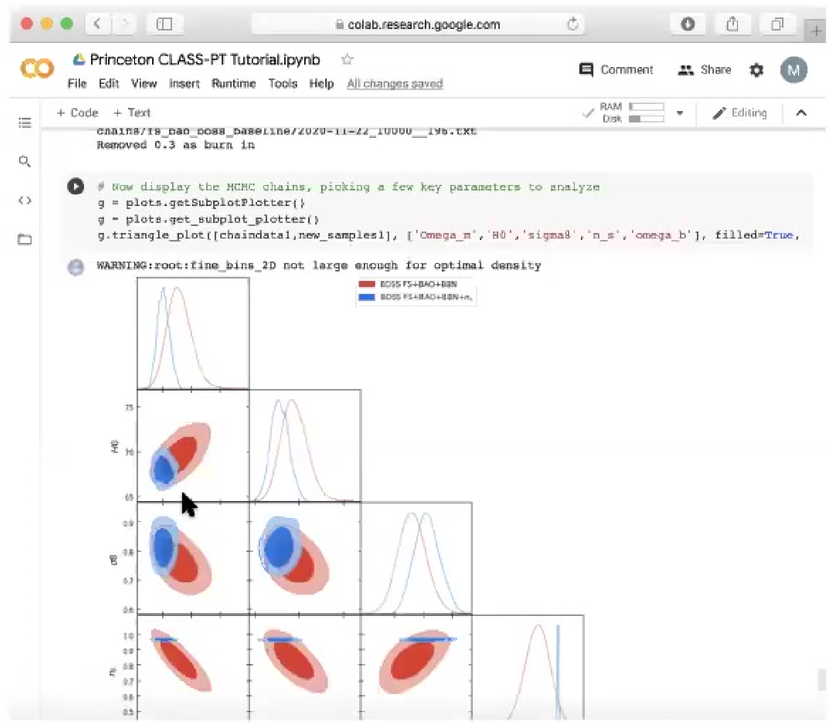
2004.10607

https://github.com/Michalychforever/CLASS-PT

Our pipeline in a nutshell

https://github.com/Michalychforever/CLASS-PT





Why the EFT (PT) ?

- Build on solid physical principles and starts from the microscopic description
  different flavors: Eulerian, Lagrangian, TSPT,...
- allows one to quickly scan over cosmo. params
- natural setup for nuisance param. marginalization
- tests at the level of summary statistics (PT challenge)
- tests at the field level

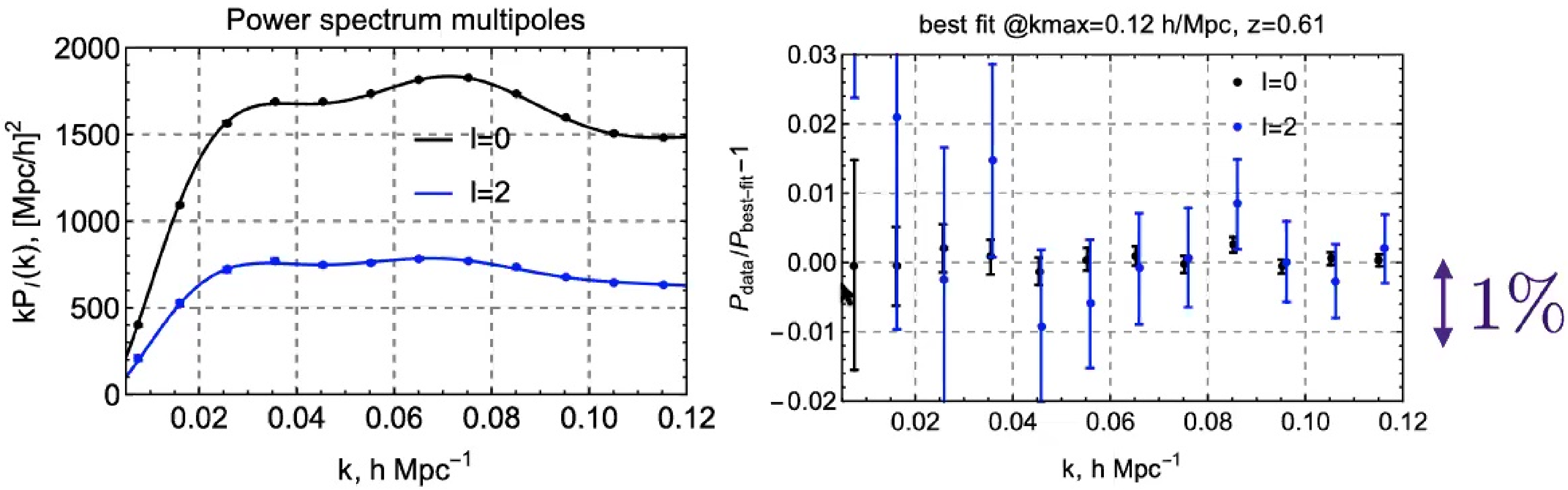
PT Challenge

Large N-body sims $\sim 600\ (\mathrm{Gpc}/h)^3$ =100x BOSS = 10x DESI

http://www-utap.phys.s.u-tokyo.ac.jp/~nishimichi/data/PTchallenge/

w/ M. Takada, T. Nishimichi, + G. D'Amico, L. Senatore, P. Zhang

2003.08277

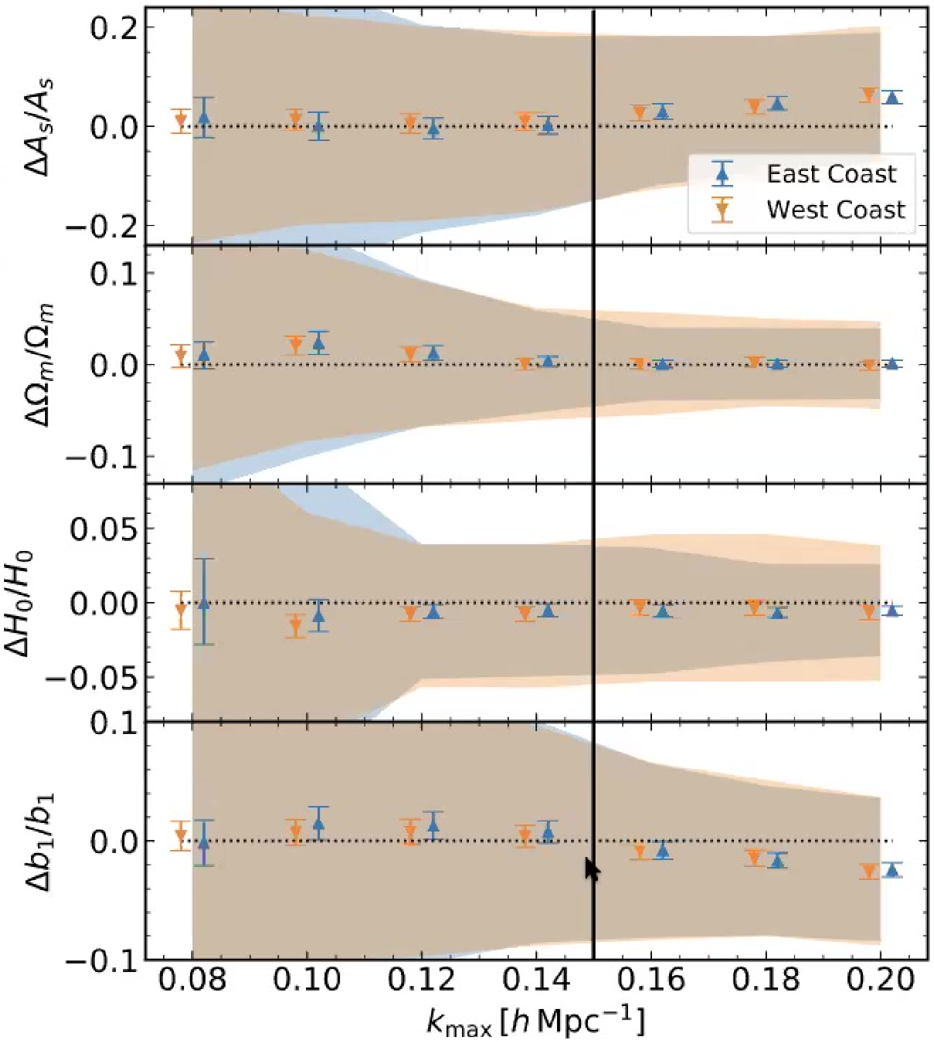


True cosmology recovered with ~0.5% accuracy

Test of simulations too!   Different halos, HOD, etc.

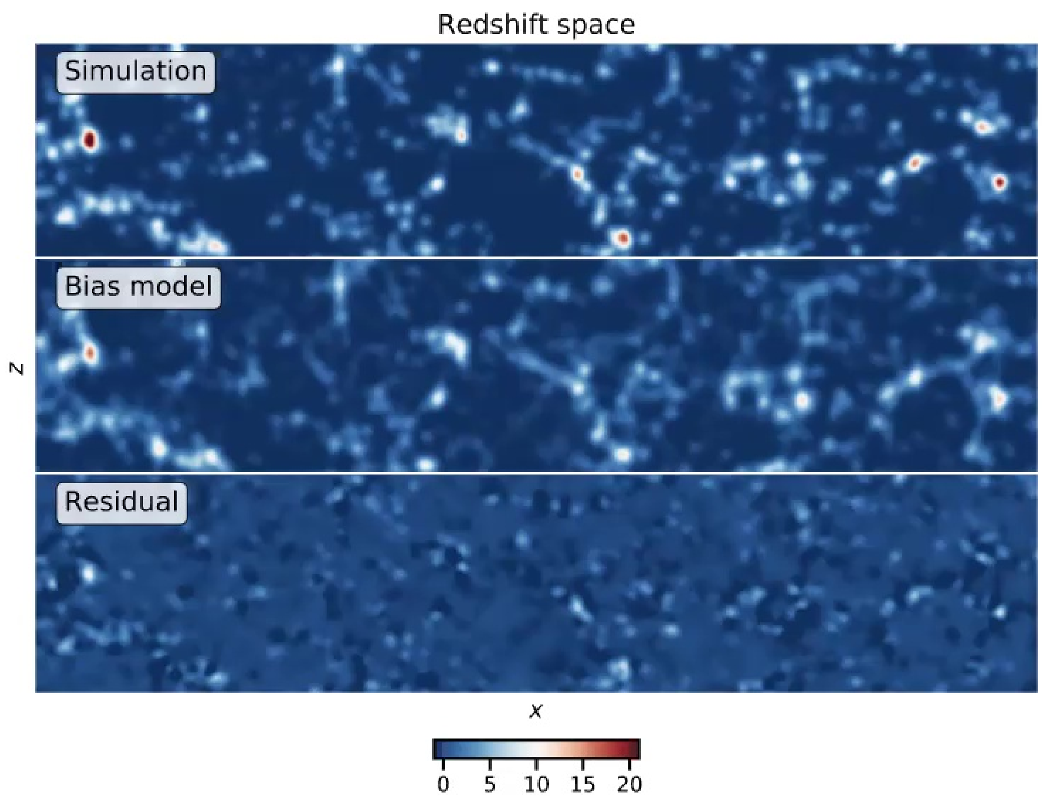
PT Challenge

no bias



bias < 1sigma

Field level comparison



Dark matter

Halos + galaxies

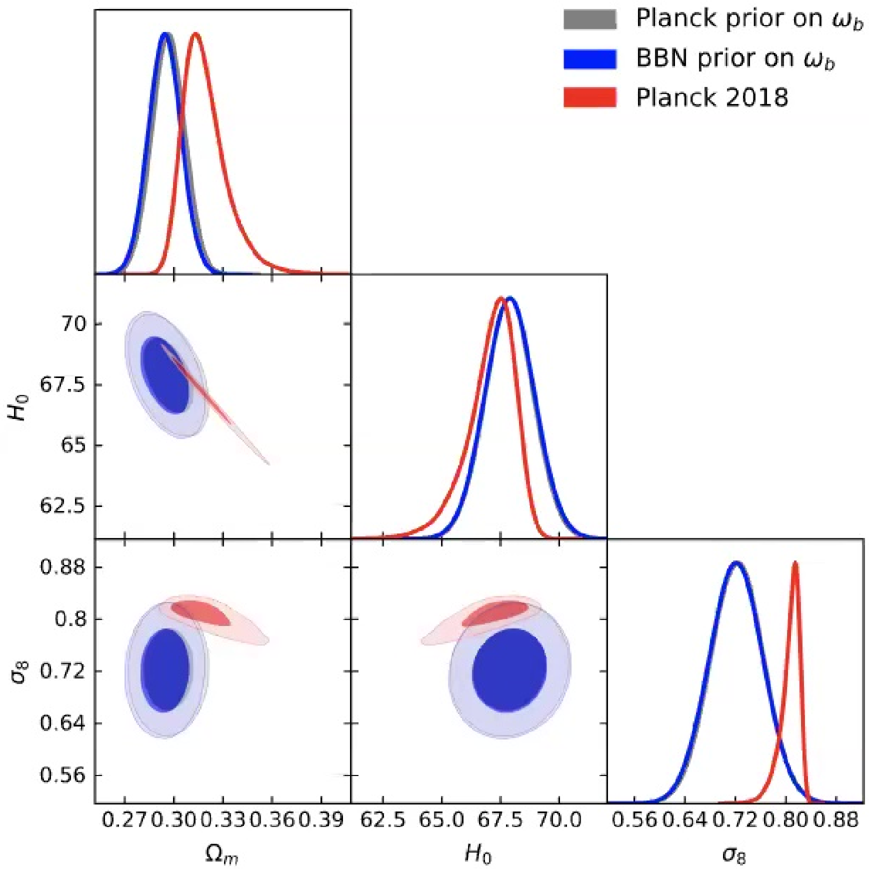
Halos in redshift space

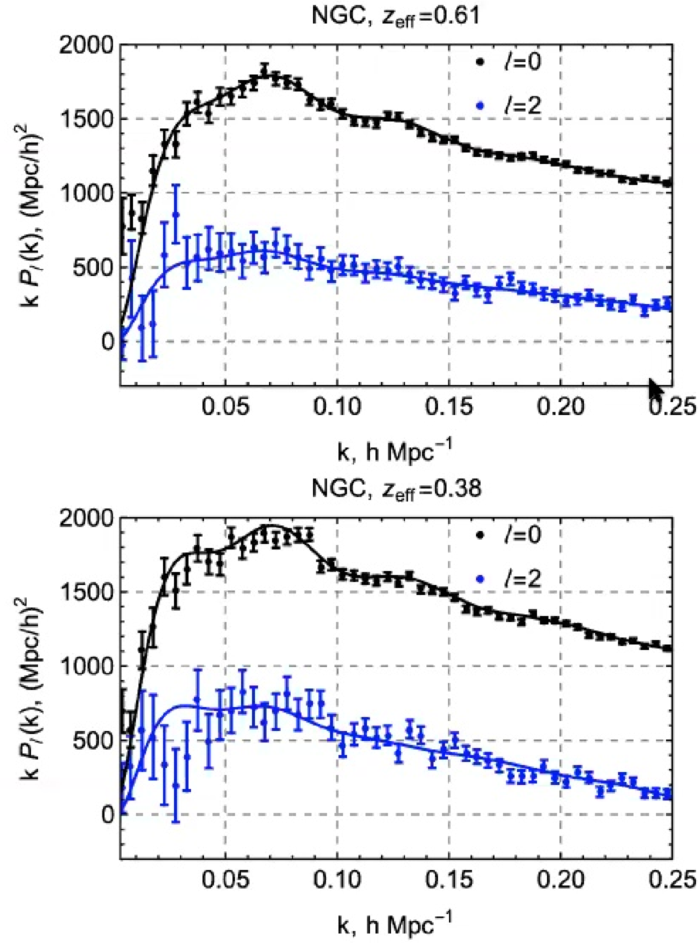
Baldauf et al'15

Schmittfull et al'18

Schmittfull, Simonovic, MI, Philcox, Zaldarriaga, '20

Reanalysis of the BOSS data (LCDM)

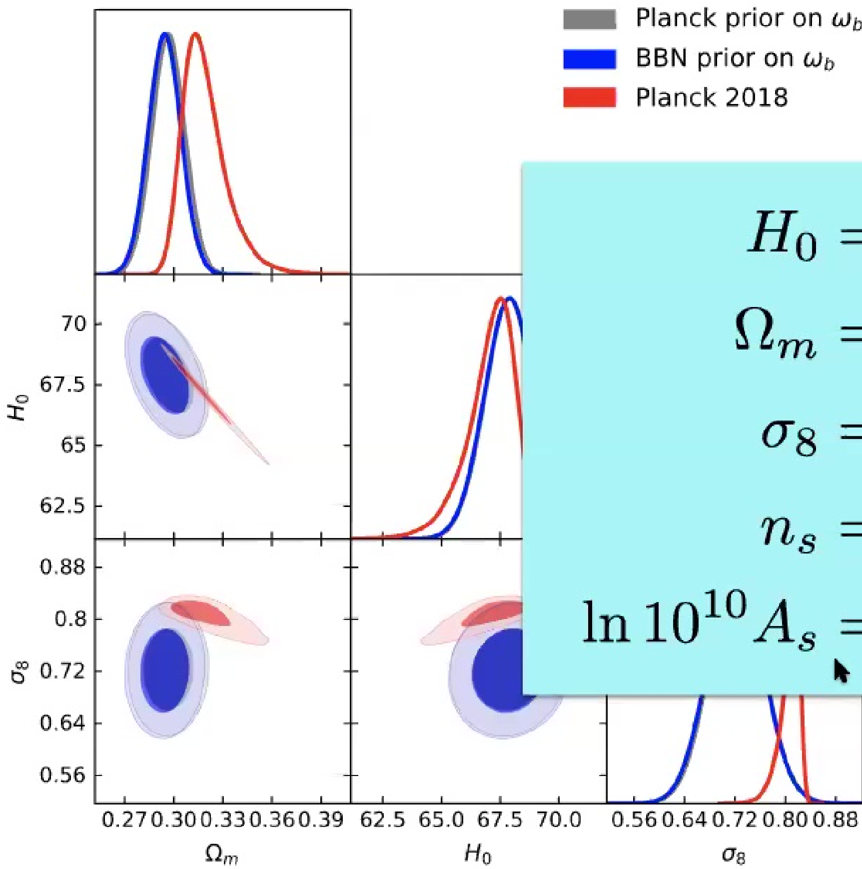




$z_{\rm eff} = 0.38,\ 0.61$ (SGC + NGC)

Nuisance params: $b_1, b_2, b_{\mathcal{G}_2}, P_{\rm shot} + c^{(0)}_{\nabla^2\delta}, c^{(2)}_{\nabla^2_z\delta}, c^{(0)+(2)}_{\nabla^4_z\delta}$ BBN prior on ob!

Reanalysis of the BOSS data (LCDM)



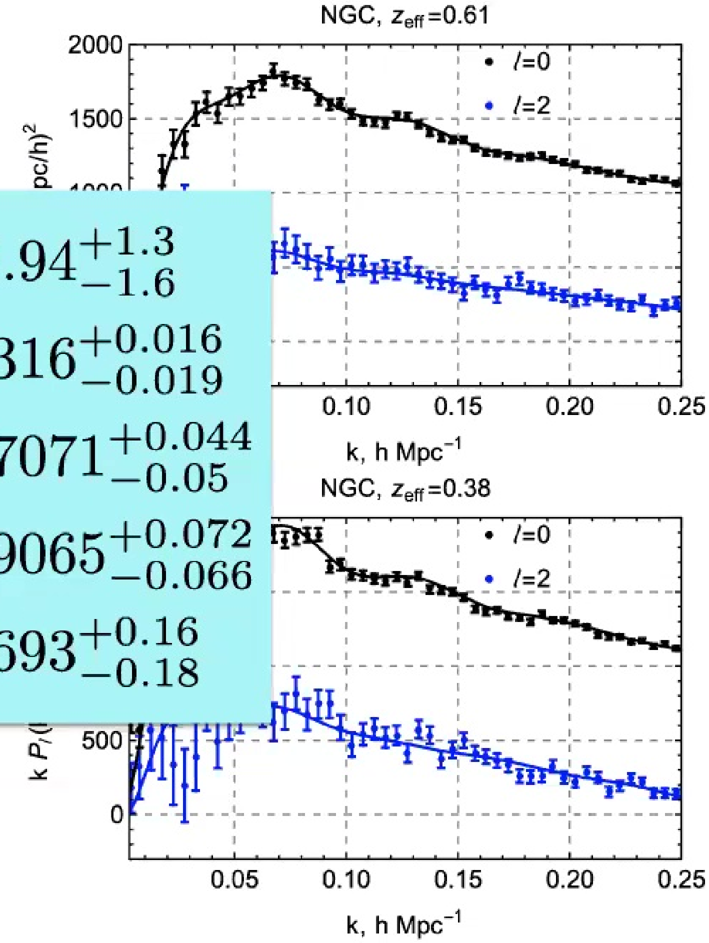
$$H_0 = 69.94^{+1.3}_{-1.6}$$

$$\Omega_m = 0.316^{+0.016}_{-0.019}$$

$$\sigma_8 = 0.7071^{+0.044}_{-0.05}$$

$$n_s = 0.9065^{+0.072}_{-0.066}$$

$$\ln 10^{10} A_s = 2.693^{+0.16}_{-0.18}$$



$z_{\rm eff} = 0.38,\ 0.61$ (SGC + NGC)

Nuisance params: $b_1, b_2, b_{\mathcal{G}_2}, P_{\rm shot} + c^{(0)}_{\nabla^2\delta}, c^{(2)}_{\nabla^2_z\delta}, c^{(0)+(2)}_{\nabla^4_z\delta}$ BBN prior on ob!

Few comments

- ★ Not including the selection bias !

  Obuljen, Percival, Dalal '20

- ★ LCDM is explicitly assumed
- ★ The BBN prior can be removed - constraints on H0 without the sound horizon

  Philcox, Sherwin, Farren, Baxter '20

- ★ Can also add the post-recon BAO

  Philcox, MI, Simonovic, Zaldarriaga '20

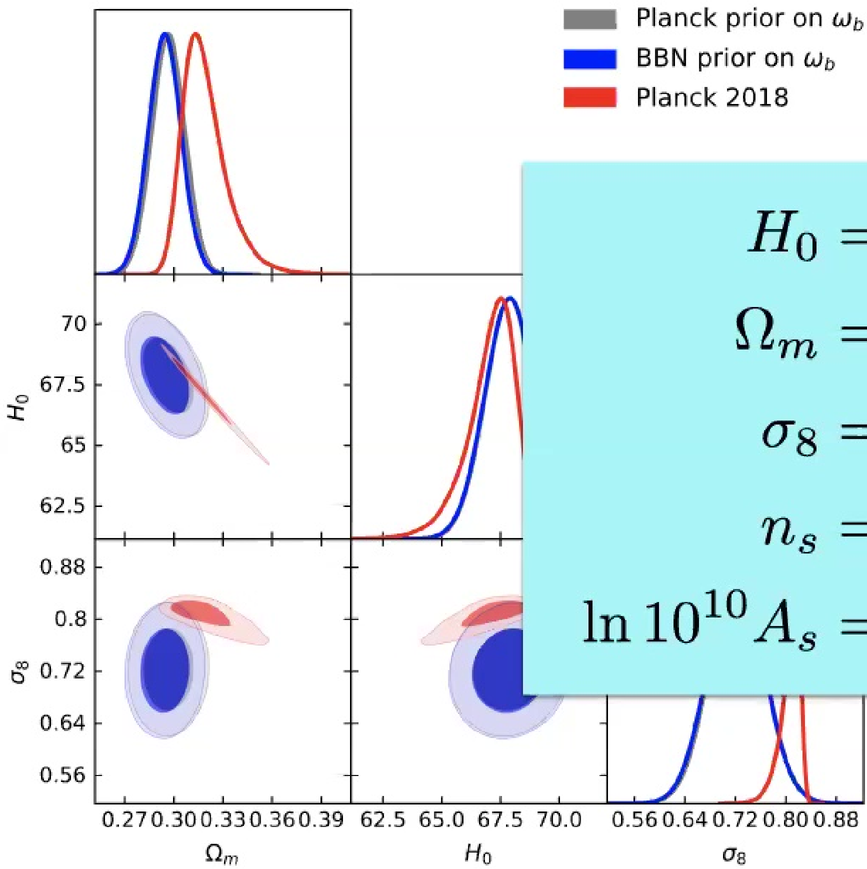
- ★ Beyond LCDM

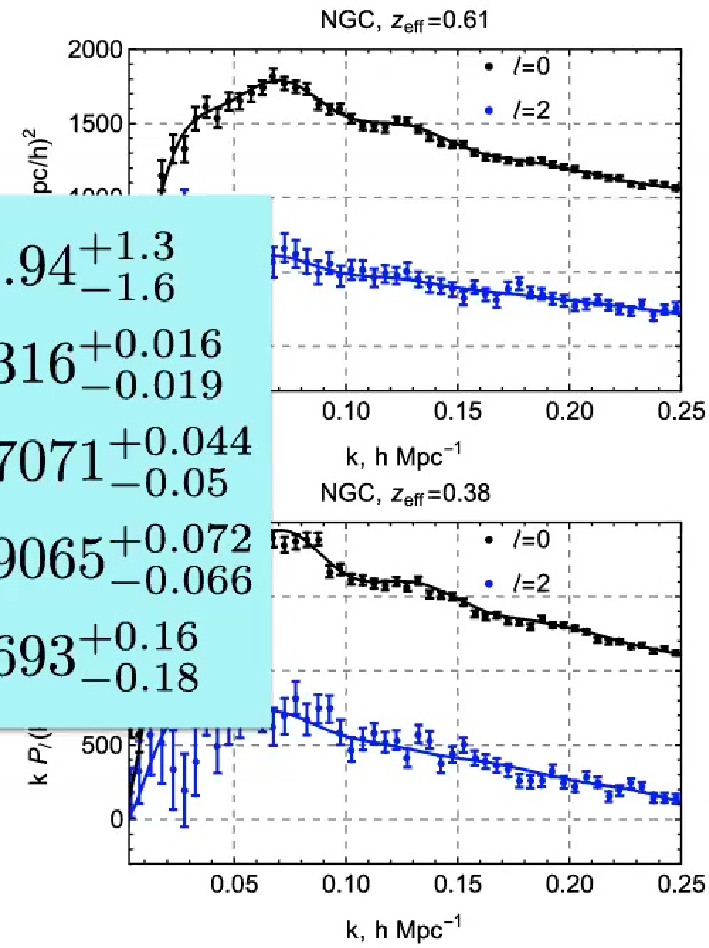
  Chudaykin, Dolgikh, MI'20
  D'Amico, Senatore, Zheng'20

- ★ Similar results by

  D'Amico, Kokron, Gleyzes, Markovich, Zheng, Beutler, Gil-Marin, Senatore'20

Reanalysis of the BOSS data (LCDM)





$$H_0 = 69.94^{+1.3}_{-1.6}$$
$$\Omega_m = 0.316^{+0.016}_{-0.019}$$
$$\sigma_8 = 0.7071^{+0.044}_{-0.05}$$
$$n_s = 0.9065^{+0.072}_{-0.066}$$
$$\ln 10^{10} A_s = 2.693^{+0.16}_{-0.18}$$

$z_{\rm eff} = 0.38,\ 0.61$ (SGC + NGC)

Nuisance params: $b_1, b_2, b_{\mathcal{G}_2}, P_{\rm shot} + c^{(0)}_{\nabla^2\delta}, c^{(2)}_{\nabla^2_z\delta}, c^{(0)+(2)}_{\nabla^4_z\delta}$ BBN prior on ob!

Few comments

★ Not including the selection bias !

Obuljen, Percival, Dalal '20

★ LCDM is explicitly assumed

★ The BBN prior can be removed - constraints on H0 without the sound horizon

Philcox, Sherwin, Farren, Baxter '20

★ Can also add the post-recon BAO

Philcox, MI, Simonovic, Zaldarriaga '20

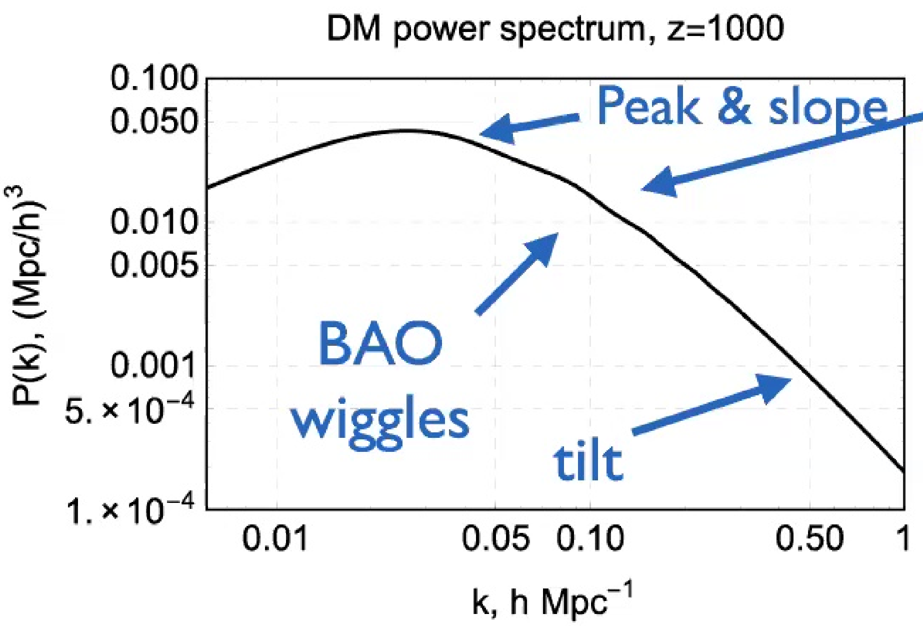
★ Beyond LCDM

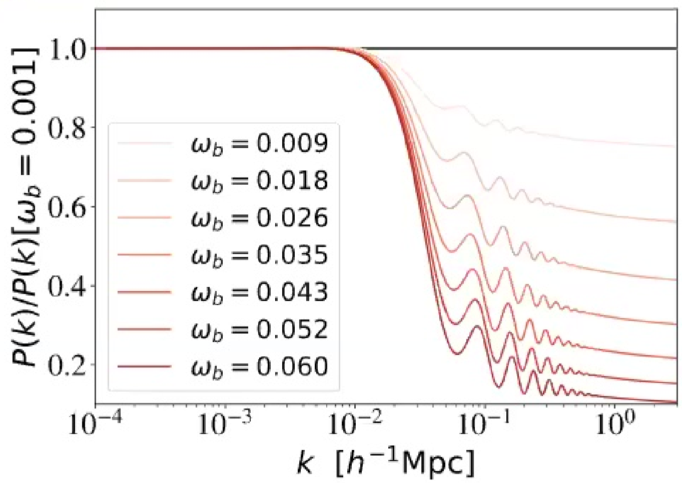
Chudaykin, Dolgikh, MI'20
D'Amico, Senatore, Zheng'20

★ Similar results by

D'Amico, Kokron, Gleyzes, Markovich, Zheng, Beutler, Gil-Marin, Senatore'20

Information





- ✔ Shape $(\omega_b, \omega_{cdm}, n_s)$
- ✔ Distance $D_V \to H_0$
- ✔ Amplitude $f\sigma_8 \to \sigma_8$
- ✘ AP effect $D_A, H \to \text{nothing}$

Forecast for Euclid/DESI - like survey

1907.06666 *w/ A. Chudaykin*

What if you gave me the data right now?

- ★ MCMC using the same pipeline w/ full non-linear model
- ★ Marginalize over all necessary nuisance params
- ★ Same data cuts as we use now

The future

Euclid consortium

DARK ENERGY SPECTROSCOPIC INSTRUMENT

LSST

PAU

SKA
SQUARE KILOMETRE ARRAY

Forecast for Euclid/DESI - like survey

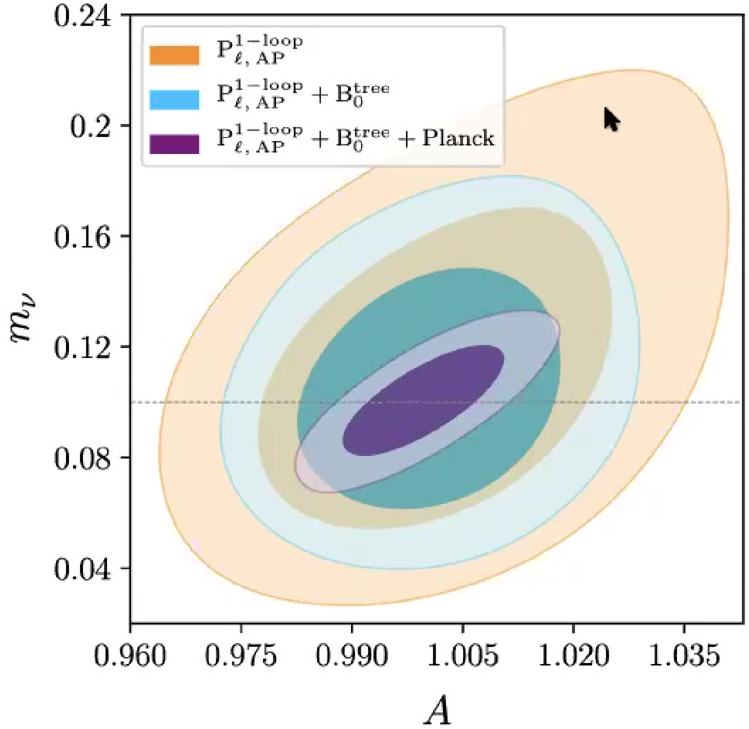
1907.06666 *w/ A. Chudaykin*

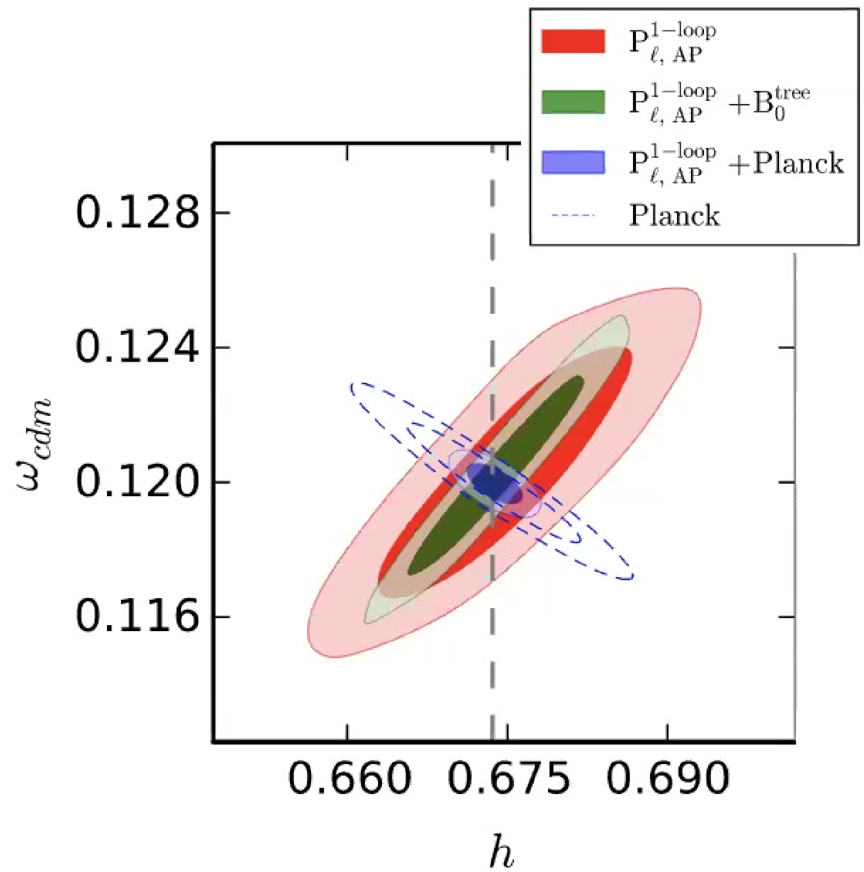
What if you gave me the data right now?

- ★ MCMC using the same pipeline w/ full non-linear model
- ★ Marginalize over all necessary nuisance params
- ★ Same data cuts as we use now

MCMC forecast for Euclid/DESI survey

$$\sigma(m_\nu) = 13 \text{ meV}$$





1907.06666 *w/ A. Chudaykin*

Part II

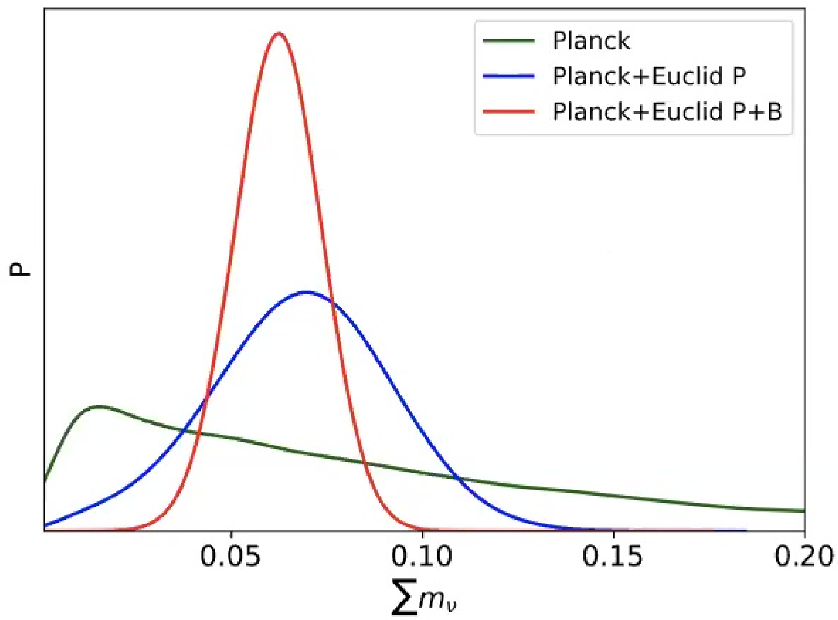
Current challenges:

Bispectrum

Post-reconstruction BAO

Covariance matrices & data compression

Why the bispectrum ?



Done:

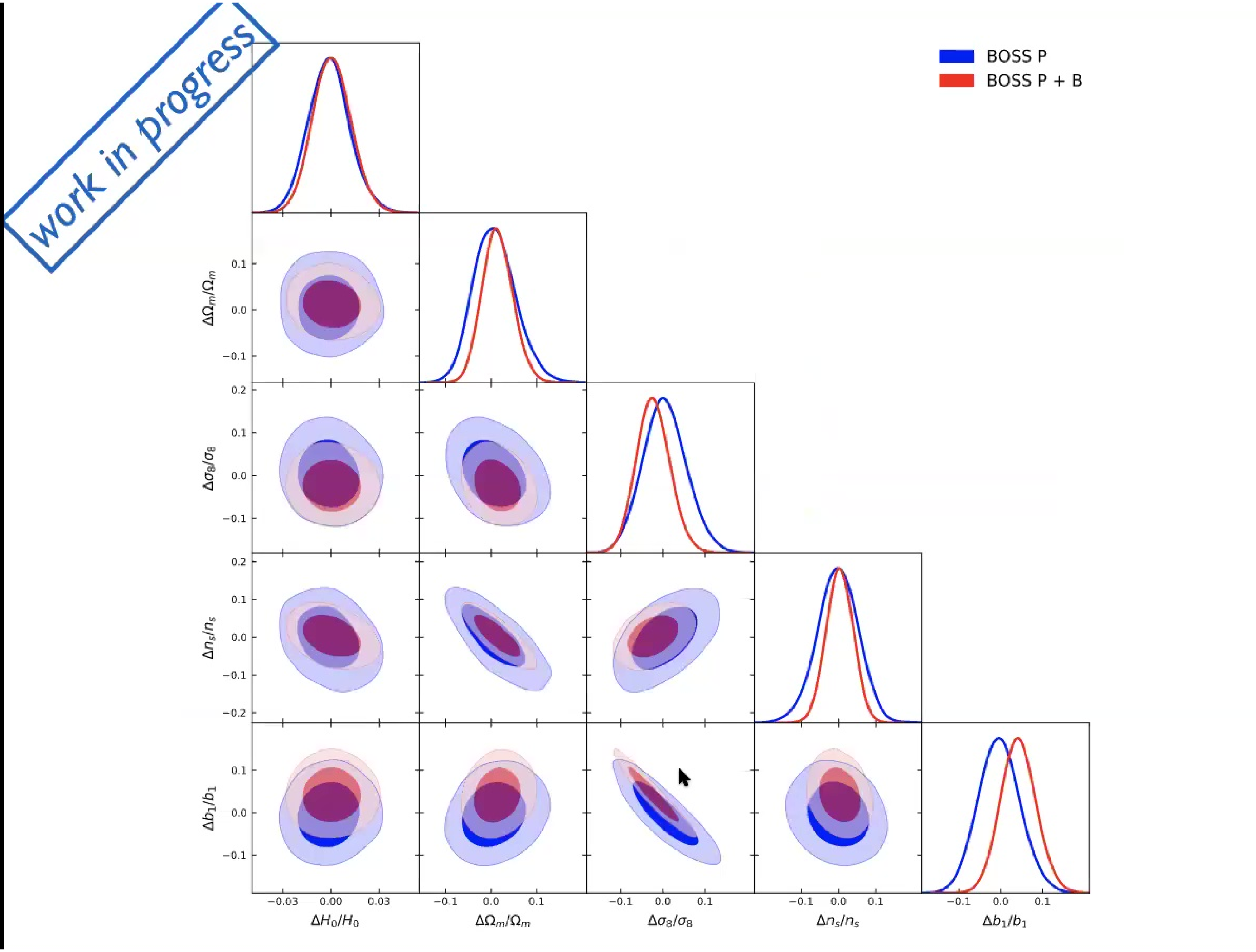
- tree-level theory
- IR resummation
- data compression
- factorization

Technical challenges:

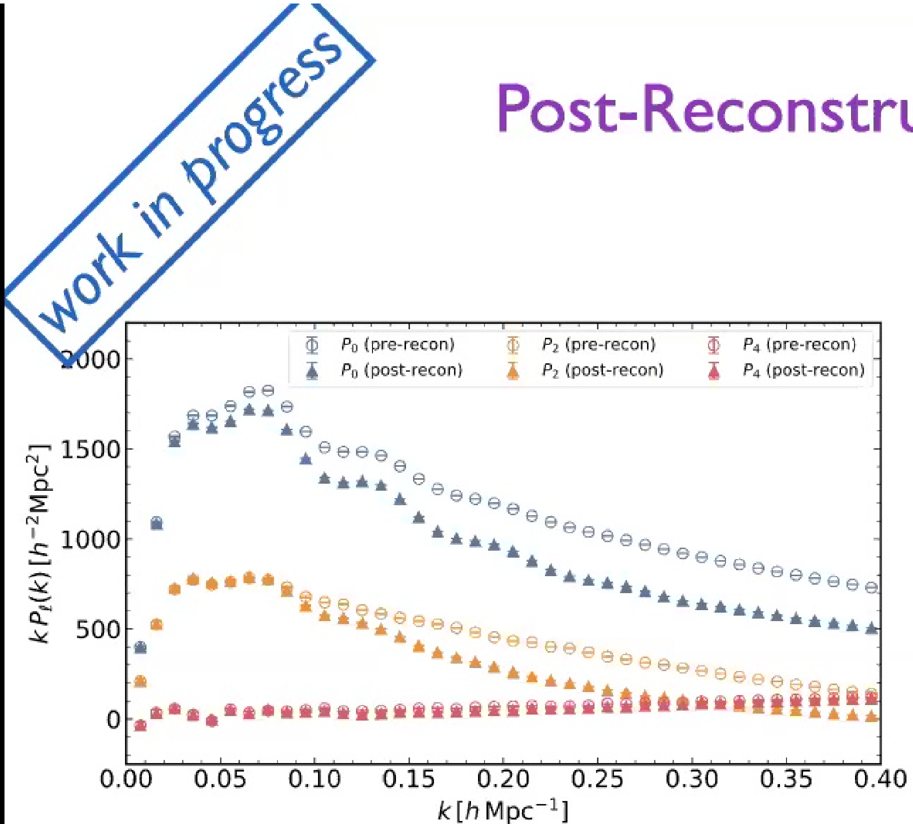
- binning
- IR - resum. implement.
- survey mask
- large datavector

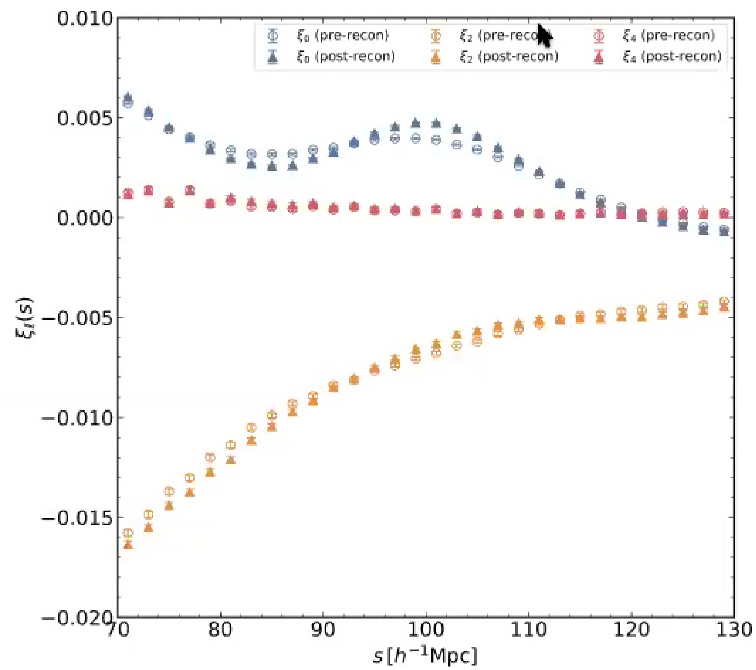
More to do:

- analytic covariance
- one-loop


work in progress
BOSS P
BOSS P + B
$\Delta\Omega_m/\Omega_m$
$\Delta\sigma_8/\sigma_8$
$\Delta n_s/n_s$
$\Delta b_1/b_1$
$\Delta H_0/H_0$
$\Delta\Omega_m/\Omega_m$
$\Delta\sigma_8/\sigma_8$
$\Delta n_s/n_s$
$\Delta b_1/b_1$

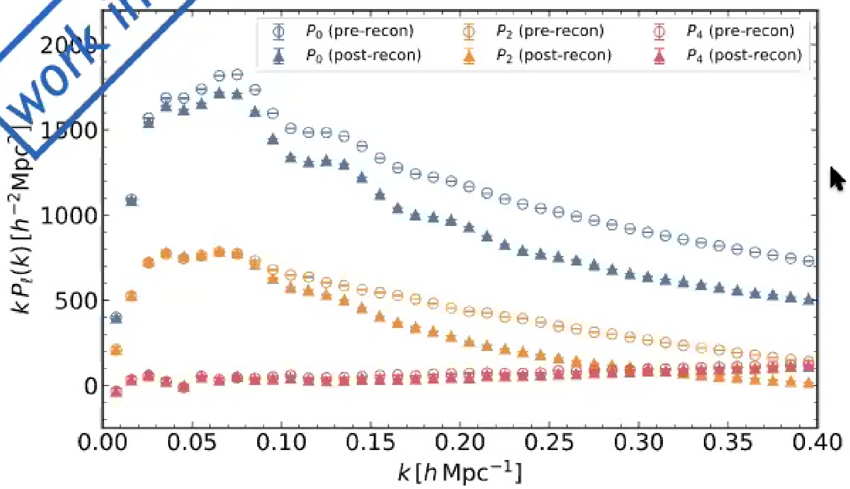
Post-Reconstructed BAO



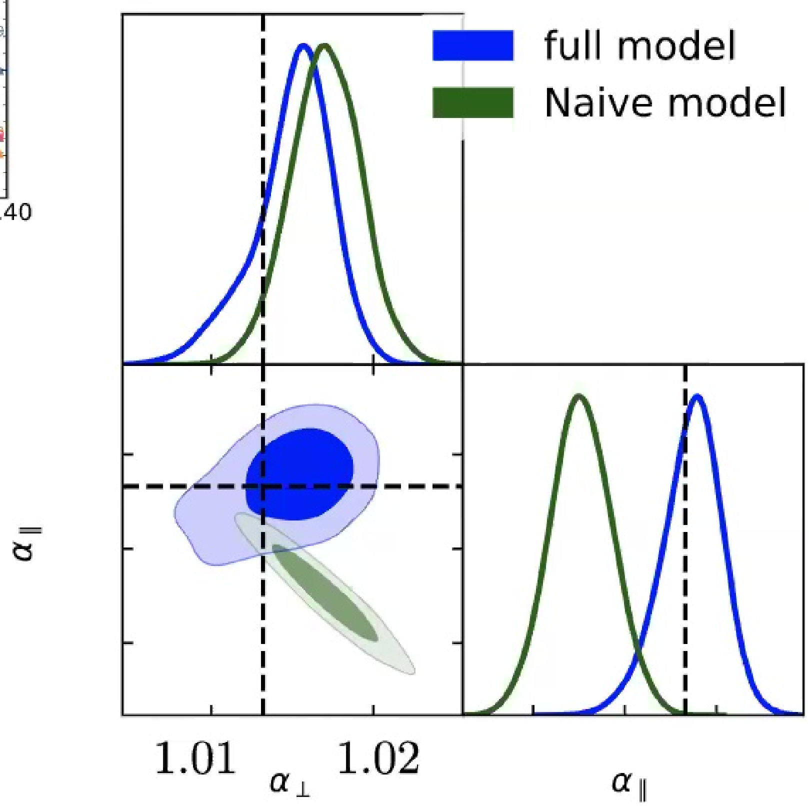


Post-Reconstructed BAO

work in progress



data by T. Nischimichi



$$P_{\rm rec} = P_{\rm nw} + P_{\rm w} e^{-\Sigma^2 k^2}$$

correct modeling of post-recon BAO is crucial

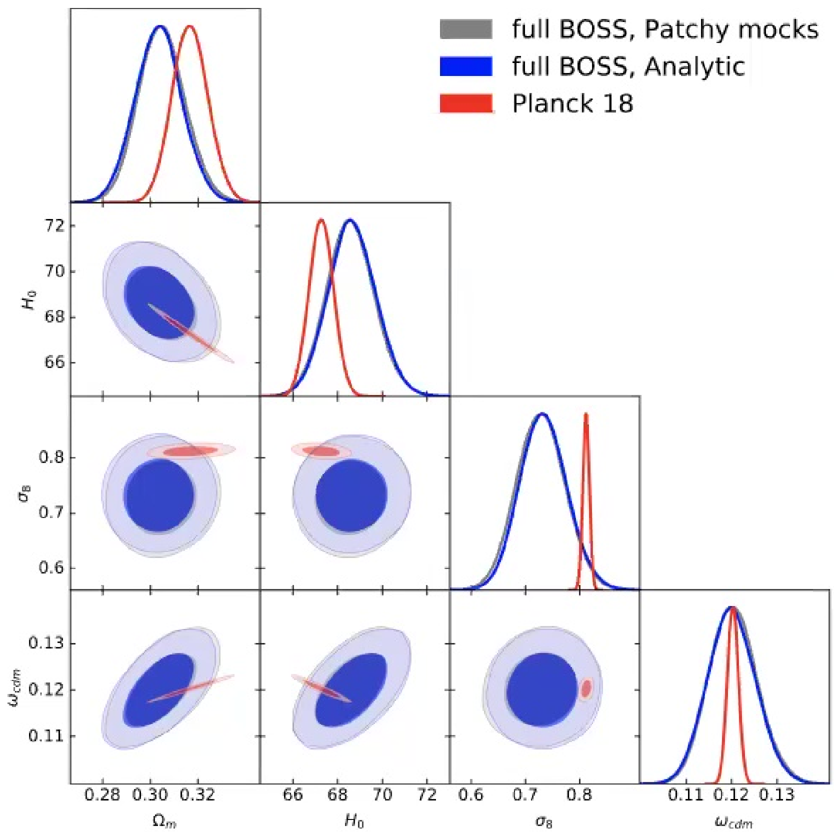
Hikage, Koyama, Heavens'17
Cheng, Vlah, White'19

Covariance matrices

Accurate parameter estimation with few mocks or without them

Analytic covariance

allows one to save millions of CPU hours



Wadekar, Scoccimarro, 1910.02914

Wadekar, MI, Scoccimarro' 2009.00622

see Jay Wadekar's talk

Covariance matrices: subspace projections

Philcox, MI, Zaldarriaga, Simonovic, Schmittfull' 2009.03311

$$\sum_{i,j=1}^{N_{\rm bins}} P(k_i)P(k_j)C^{-1}_{ij} = \sum_{i=1}^{N_{\rm bins}} \lambda_i Q^2(k_i) \approx \sum_{i=1}^{\bar{N}} \lambda_i Q^2(k_i)$$

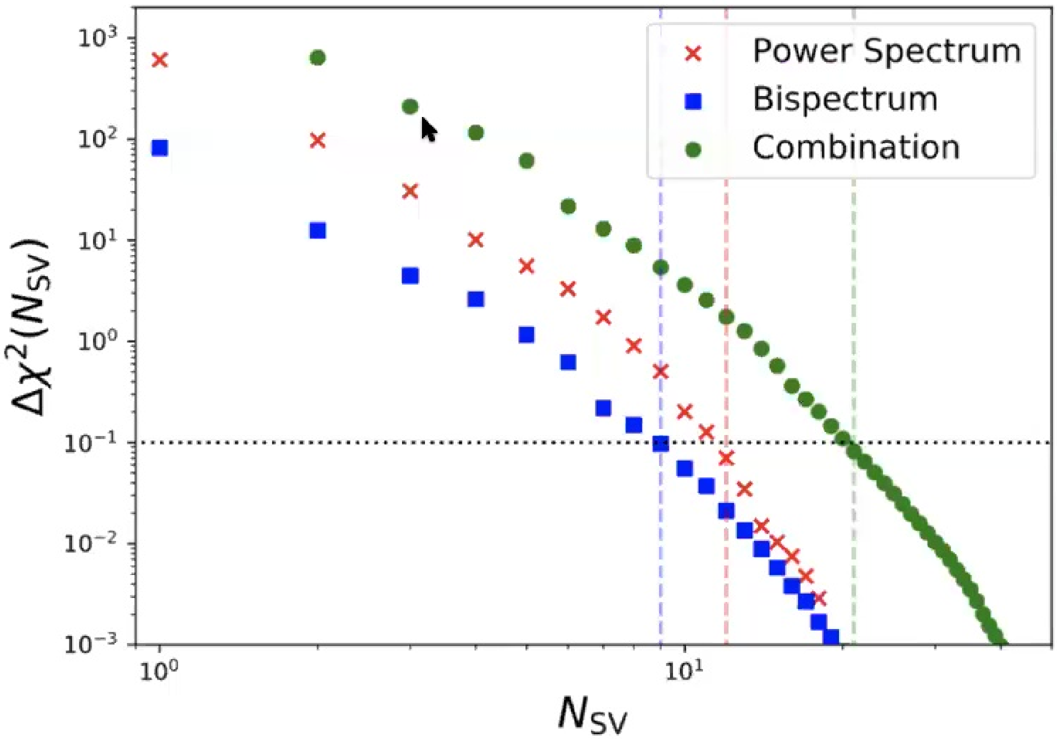
Taylor et al'13, Scoccimarro'00 ,++

$$\{\theta_j\} \to \{P_j(k_i)\} \to \chi^2_{\rm eff} = \sum_{a=1}^{N_{\rm bank}} \sum_{i,j}^{N_{\rm bins}} (P^{(a)} - \bar{P})\hat{C}^{-1}(P^{(a)} - \bar{P})$$

$$X_i^{(a)} \equiv \hat{C}_{ij}^{-1/2} \Delta P_j^{(a)} \qquad X_i^{(a)} = \sum_{\alpha=1}^{N_{\rm bins}} U^a_\alpha D_\alpha V^a_\alpha \approx \sum_{\alpha=1}^{N_{\rm SV}} U^a_\alpha D_\alpha V^a_\alpha$$

$$\bar{\chi}^2 \equiv \frac{1}{N_{\rm bank}} \chi^2_{\rm eff} = \frac{1}{N_{\rm bank}} \sum_{\alpha}^{N_{\rm SV}} D^2_\alpha$$

Covariance matrices: subspace projections



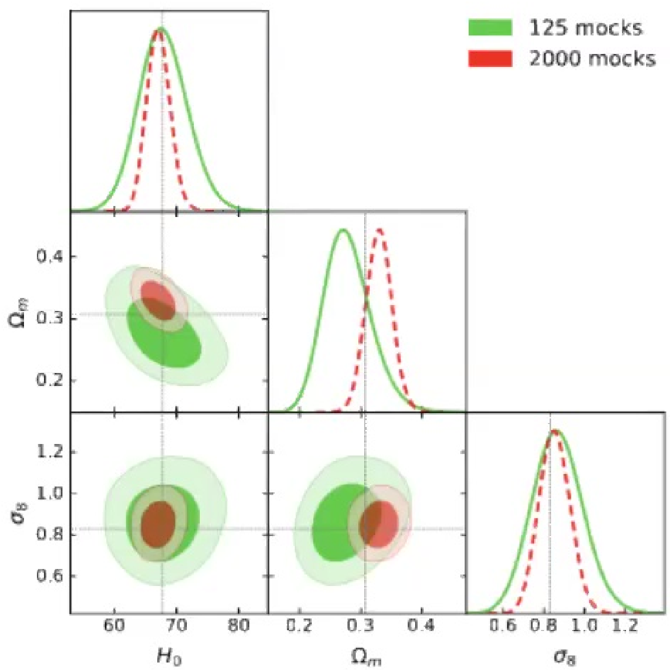
BOSS:

$$N_{\rm bins}(P) = 100 \qquad N_{\rm SV}(P) = 12$$
$$N_{\rm bins}(B) = 2300 \qquad N_{\rm SV}(B) = 9$$

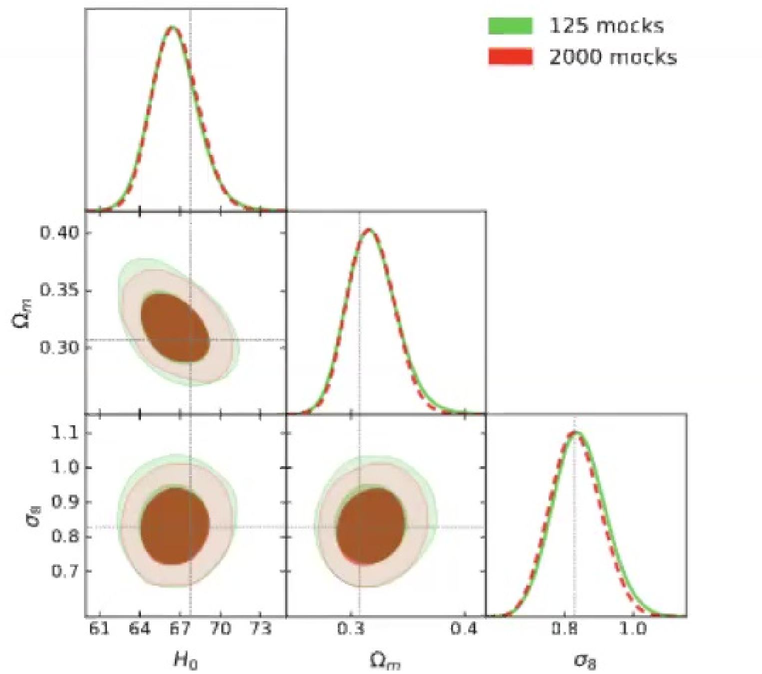
Euclid/DESI:

$$N_{\rm SV}(P) = 16$$

Covariance matrices: subspace projections



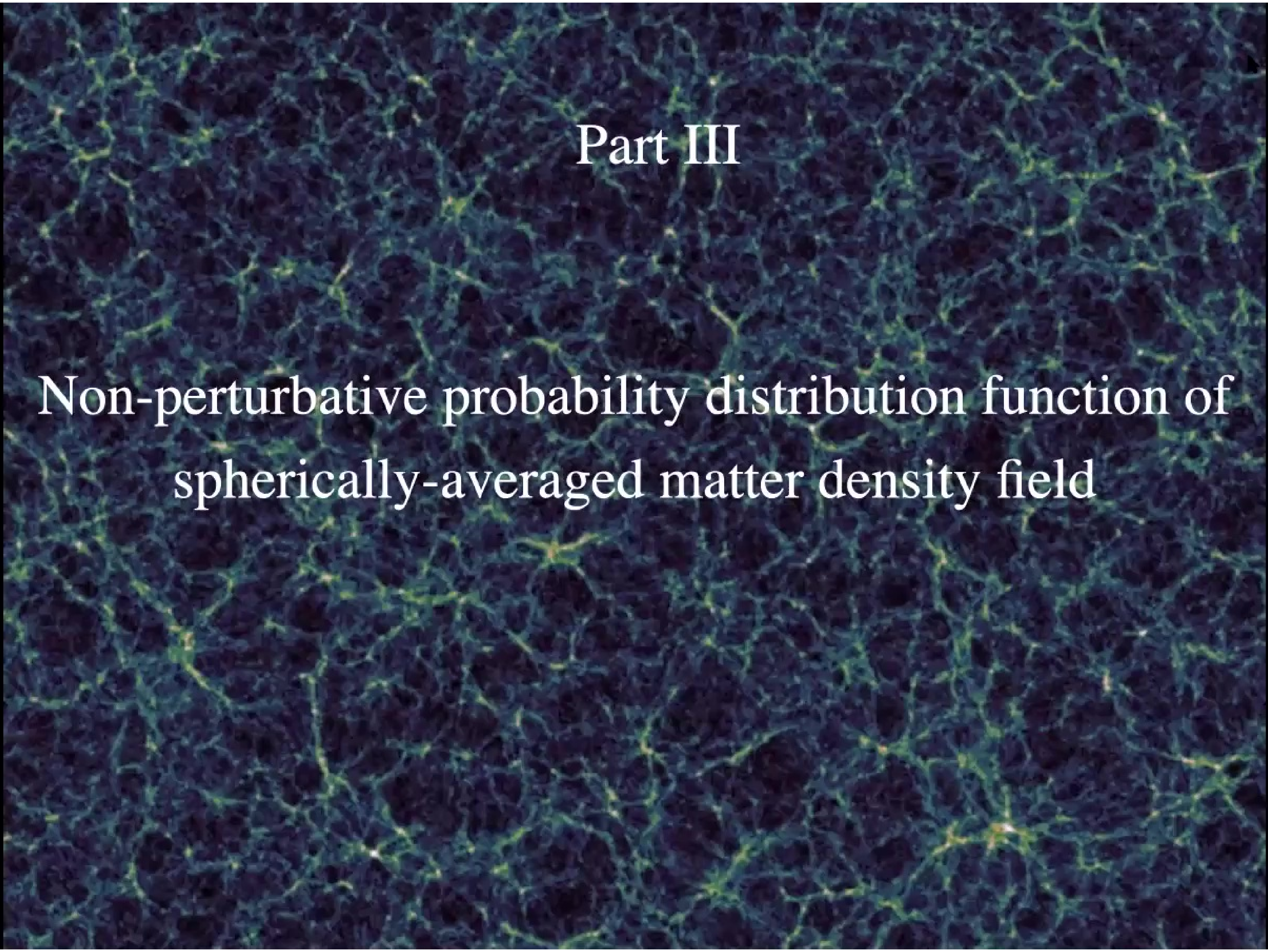
(a) 96-bin Power Spectrum



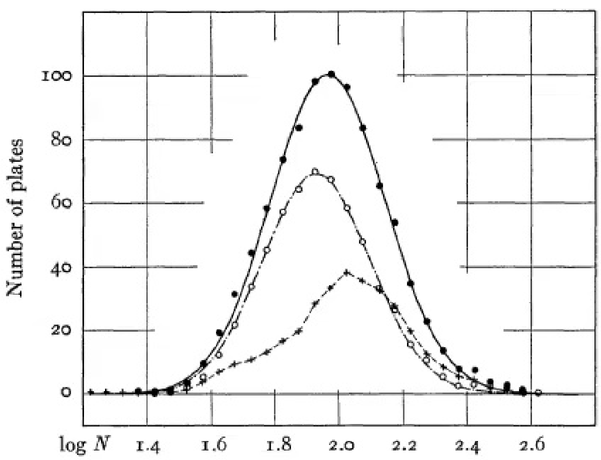
(c) 12 Subspace Coefficients

Percival, Ross et al' 13
Dodelson, Schneider'13

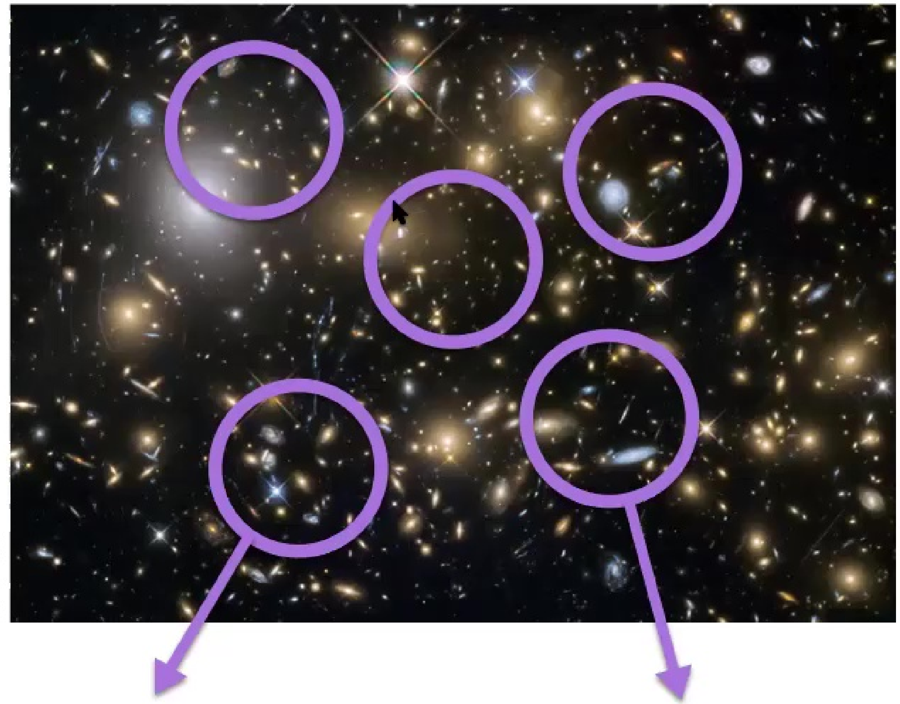
Philcox, MI, Zaldarriaga, Simonovic, Schmittfull'20


Part III
Non-perturbative probability distribution function of
spherically-averaged matter density field

Counts in Cells



IG. 8.—Frequency distribution of log N reduced to the galactic poles. Data from le XIV. Crosses represent the 331 extra-survey fields; circles, the 587 survey fields; s, the combined data, 918 fields. The smooth curves through the survey fields and combined data (the two upper curves) are normal error-curves adjusted to the ts.



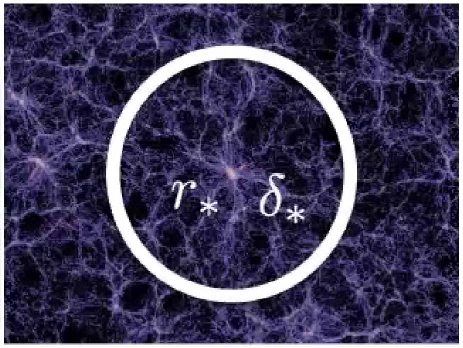
$N_{gal} = 40$ $N_{gal} = 10$

E. Hubble, 1934

Reliable theoretical model from first principles?

Theory

probability $\mathcal{P}(\delta_*)$ that a cell of radius r_* contains averaged density contrast δ_*



$\hbar \sim \sigma_R^2 \ll 1$ $\qquad$ $\sigma_R^2 = \langle \delta^2 \rangle_R$

$$\mathcal{P} = \exp\left\{ -\frac{1}{\hbar} \left(a_0 + \hbar a_1 + \ldots \right) \right\}$$

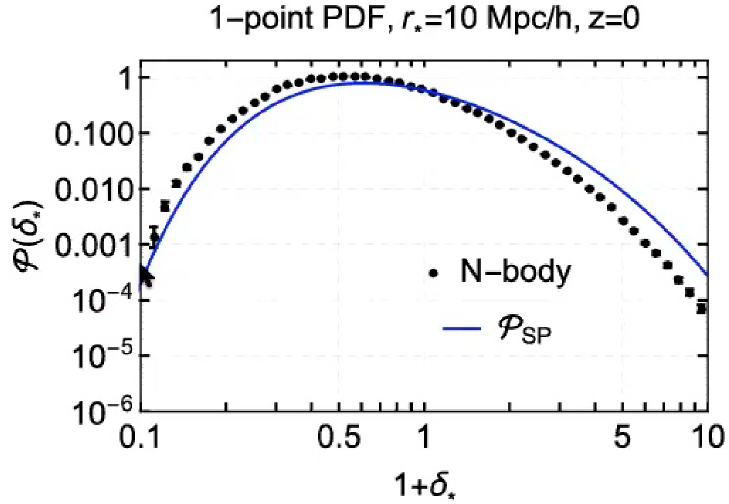
Saddle-point solution
(= 'instanton')
defined by spherical collapse

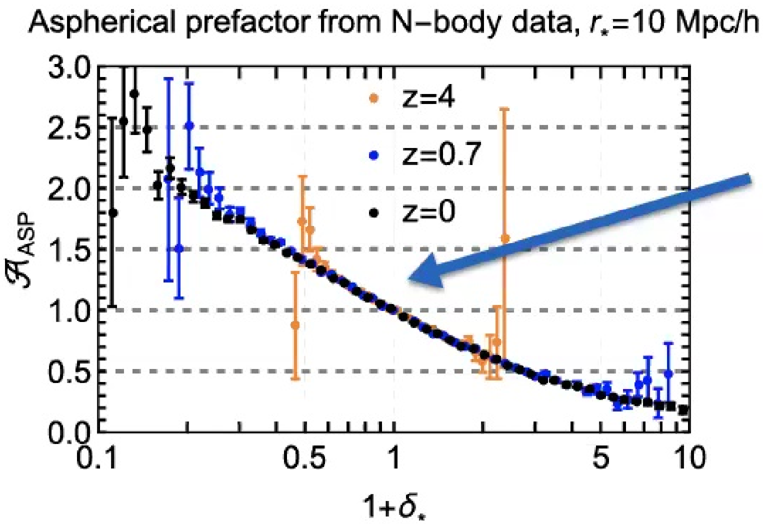
Valageas'02, Uhlemann et al.'16

Prefactor due to fluctuations
(='determinant')
contains aspherical corrections

Aspherical prefactor

$$\mathcal{P} = \mathcal{P}_{SP} \times \mathcal{A}_{ASP}$$



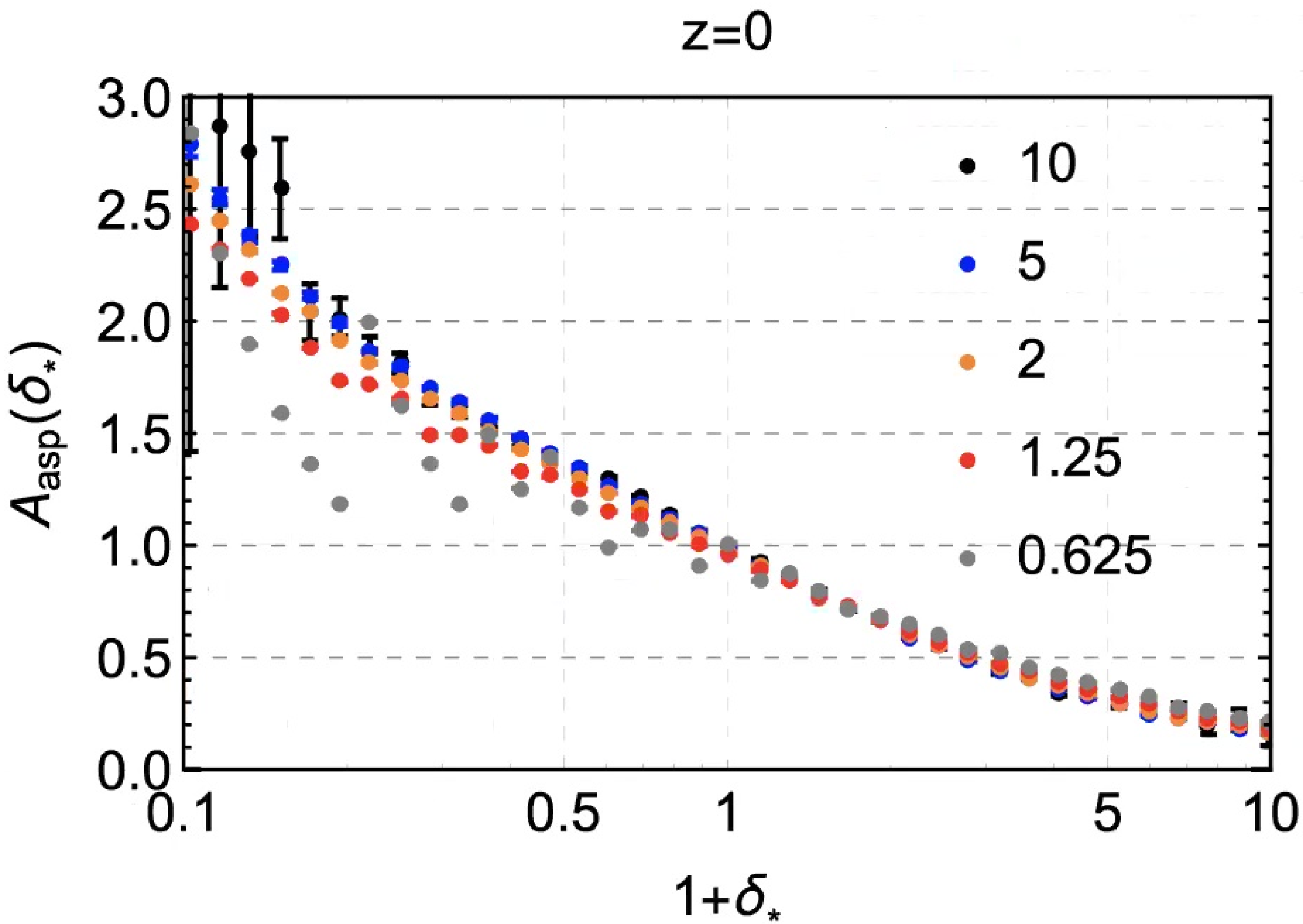


Value at the origin controlled by unitarity

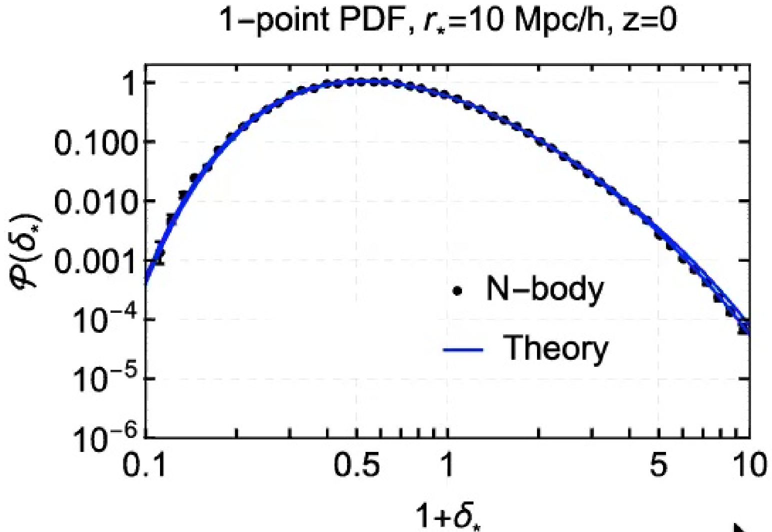
Slope at the origin controlled by translation invariance

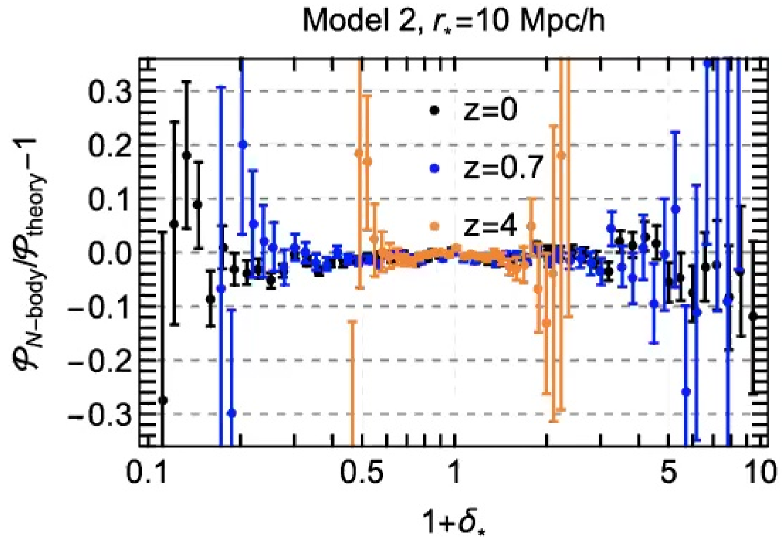
N.B. Quasi-classical scaling is valid !

Semiclassical scaling



Aspherical prefactor: EFT in curved background





Consistency and Predictivity

Consistency:

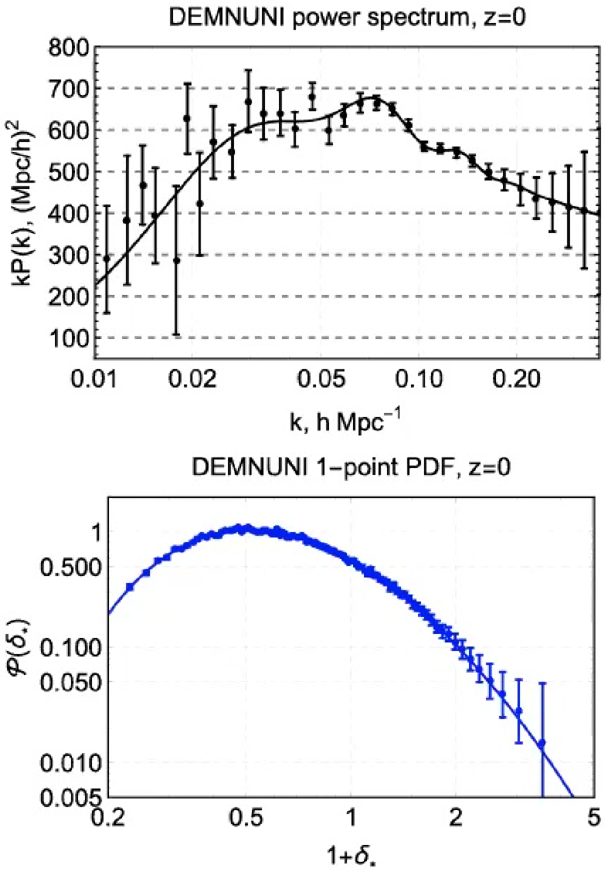
- Controlled calculation: not just O(1), but NLO corrections as well
- Unitarity is automatic
- Translation invariance retained - ensures the correct mean

Predictions:

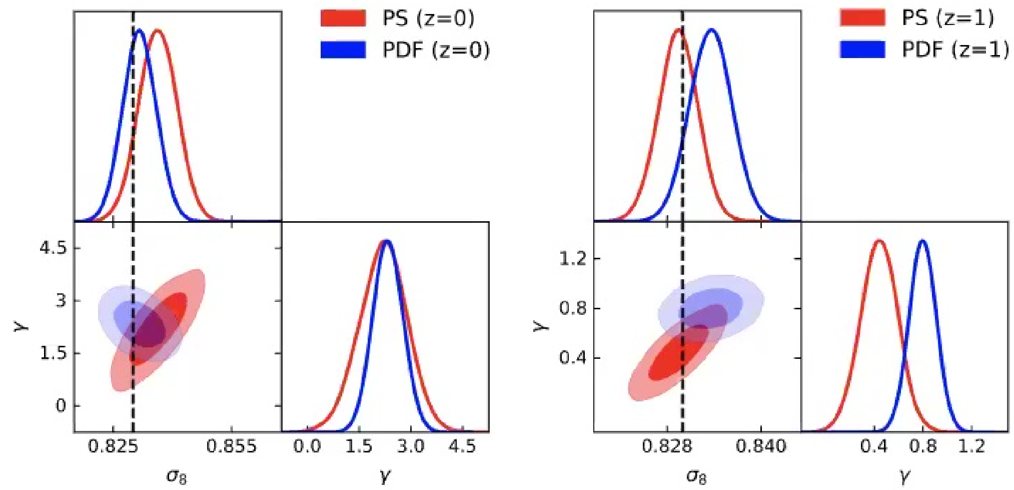
- Semiclassical scaling
- Prefactor's shape
- Cosmology-dependence
- Dark matter sound speed (ctr.) the same as in PS
- Average profiles of cosmic density

Cosmological constraints - ?

Uhlemann et al'19



covariance is a bottleneck



work in progress

Summary

- LSS (full-shape) is a powerful probe,
  PT is robust & precise
- Cosmology similar or better than Planck with
  DESI/Euclid + neutrino masses @5sigma
- Bispectrum and post-recon BAO models
  have to improve
- Analytic covariances + data compression:
  saving time and making life easier
- PDF: intriguing non-perturbative relations
  and potentially an interesting probe