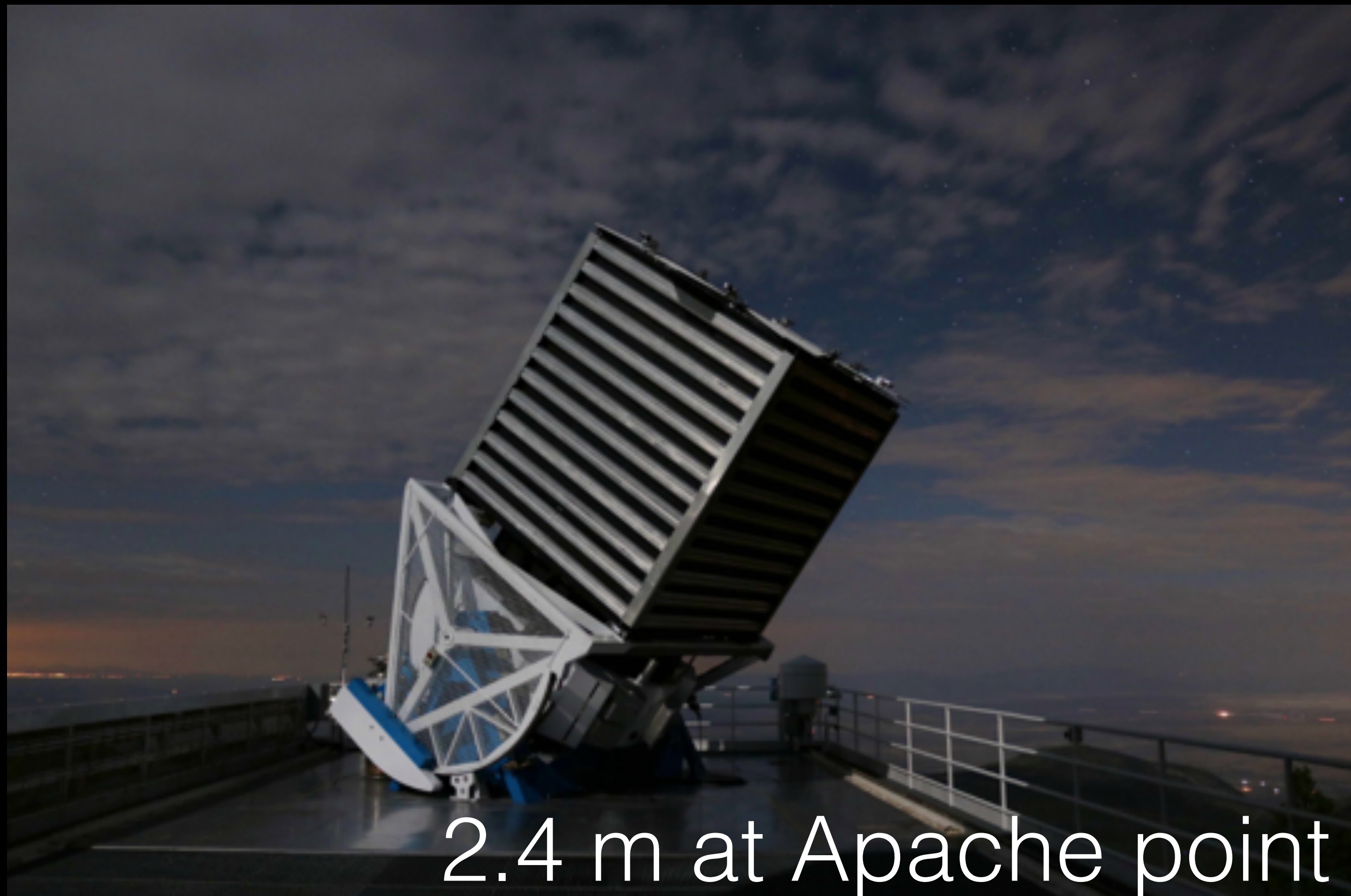
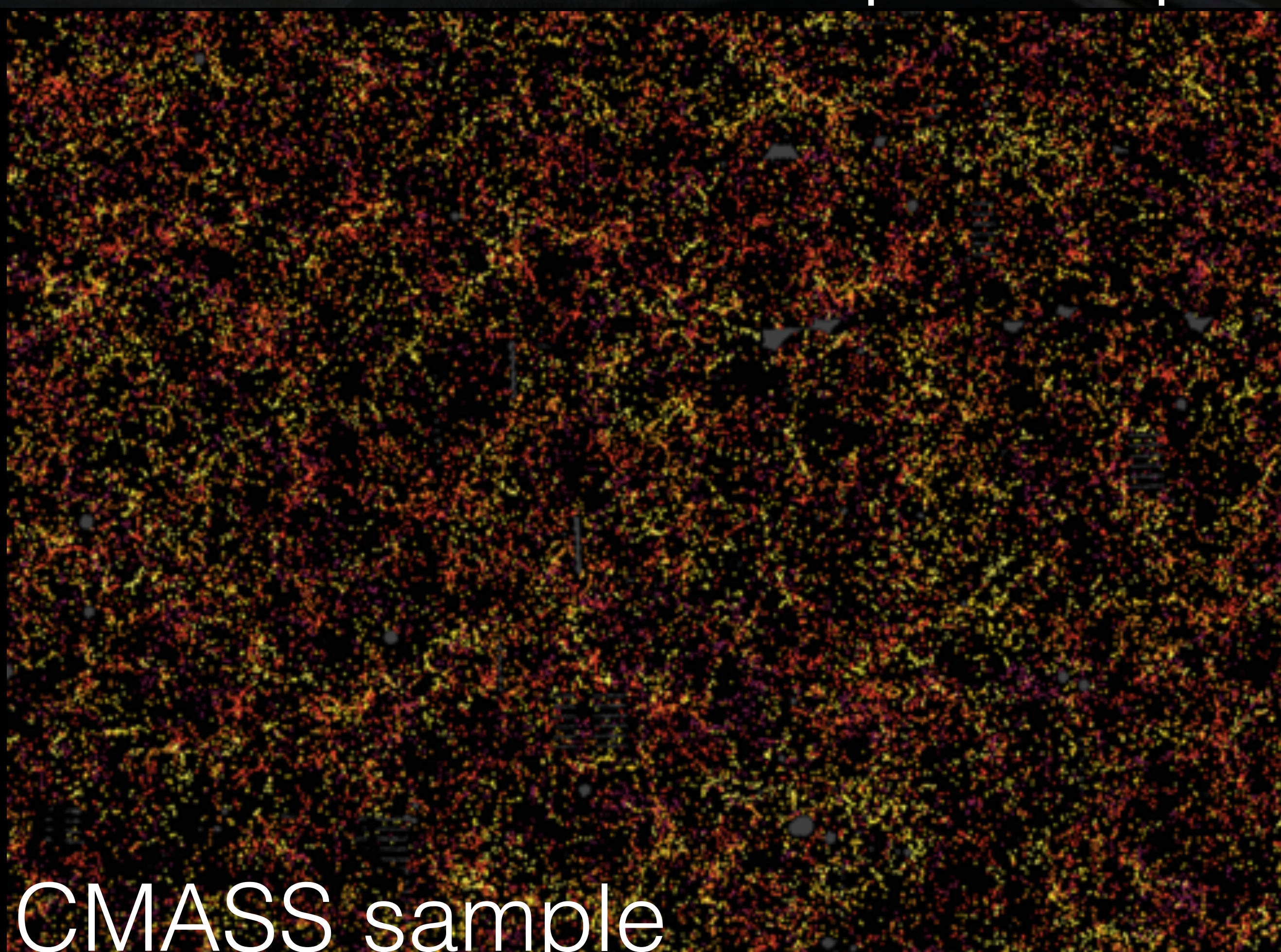
ZACHARY SLEPIAN
EINSTEIN (CHAMBERLAIN) FELLOW, LBNL

# MAXIMIZING INFORMATION-EXTRACTION FROM NEXT GENERATION SURVEYS

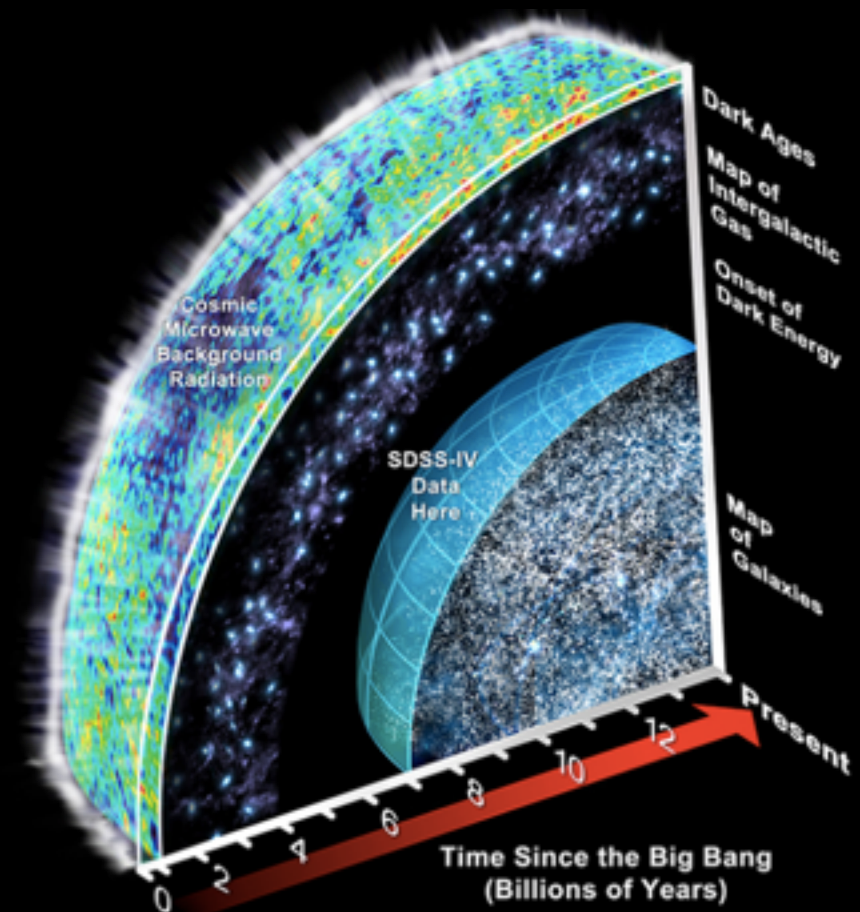
TEVPA
9 AUGUST 2017



2.4 m at Apache point



CMASS sample




SDSS

http://astro.uchicago.edu/cosmus/projects/sloangalaxies/ http://www.sdss.org/press-releases/astronomers-map-a-record-breaking-1-2-million-galaxies-to-study-the-properties-of-dark-energy/ http://www.sdss.org/wp-content/uploads/2016/07/sdss_gaulme1.jpg

# Why make a map?

https://www.pinterest.com/pin/79235274667115008/

# To enable discovery
# To find patterns

https://www.pinterest.com/pin/462322717978863523/

http://www.freakingnews.com/Angel-Butterflies-Design-by-Escher-Pictures-26905.asp

# To enable discovery
# To find patterns

**Homogeneous**

**Isotropic**

https://www.pinterest.com/pin/462322717978863523/

http://www.freakingnews.com/Angel-Butterflies-Design-by-Escher-Pictures-26905.asp

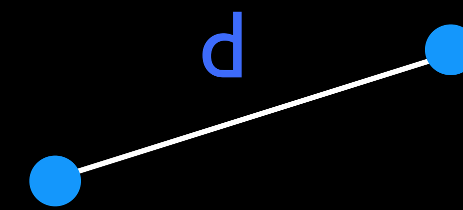
**Laws of physics: same everywhere and in all directions**
**This is what we *mean* by laws.**

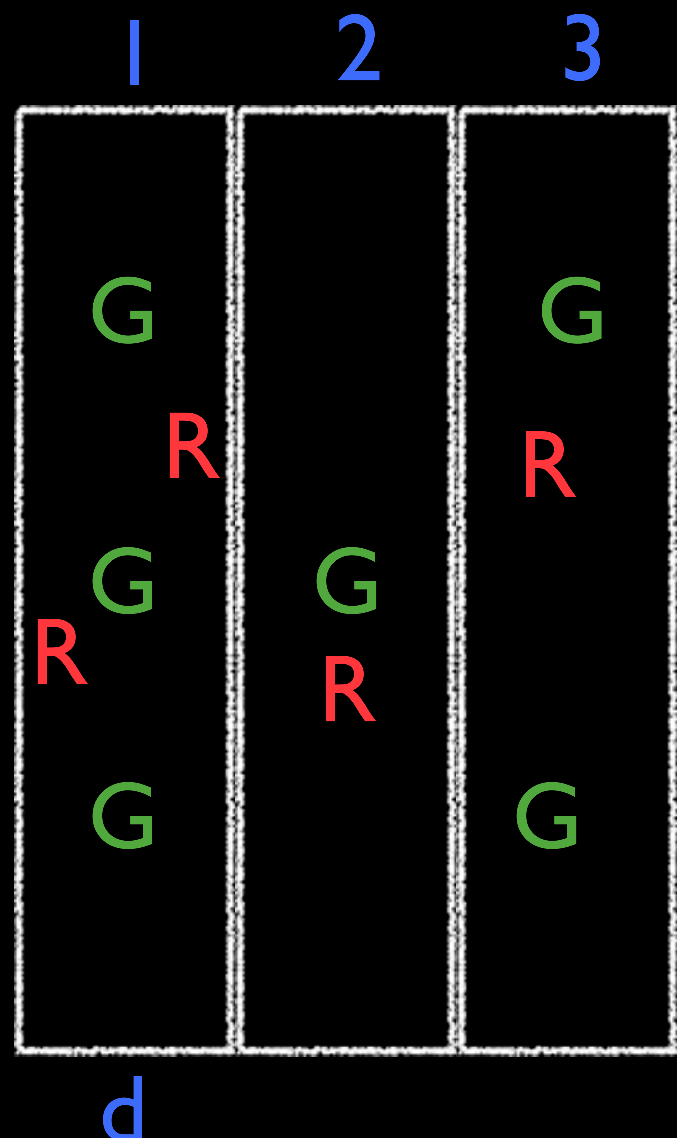
**Galaxy clustering will be so too.***

**Need a way to *mod out* translations and rotations.**

# CORRELATION FUNCTIONS

2-point correlation function (2PCF): count excess pairs of galaxies over random

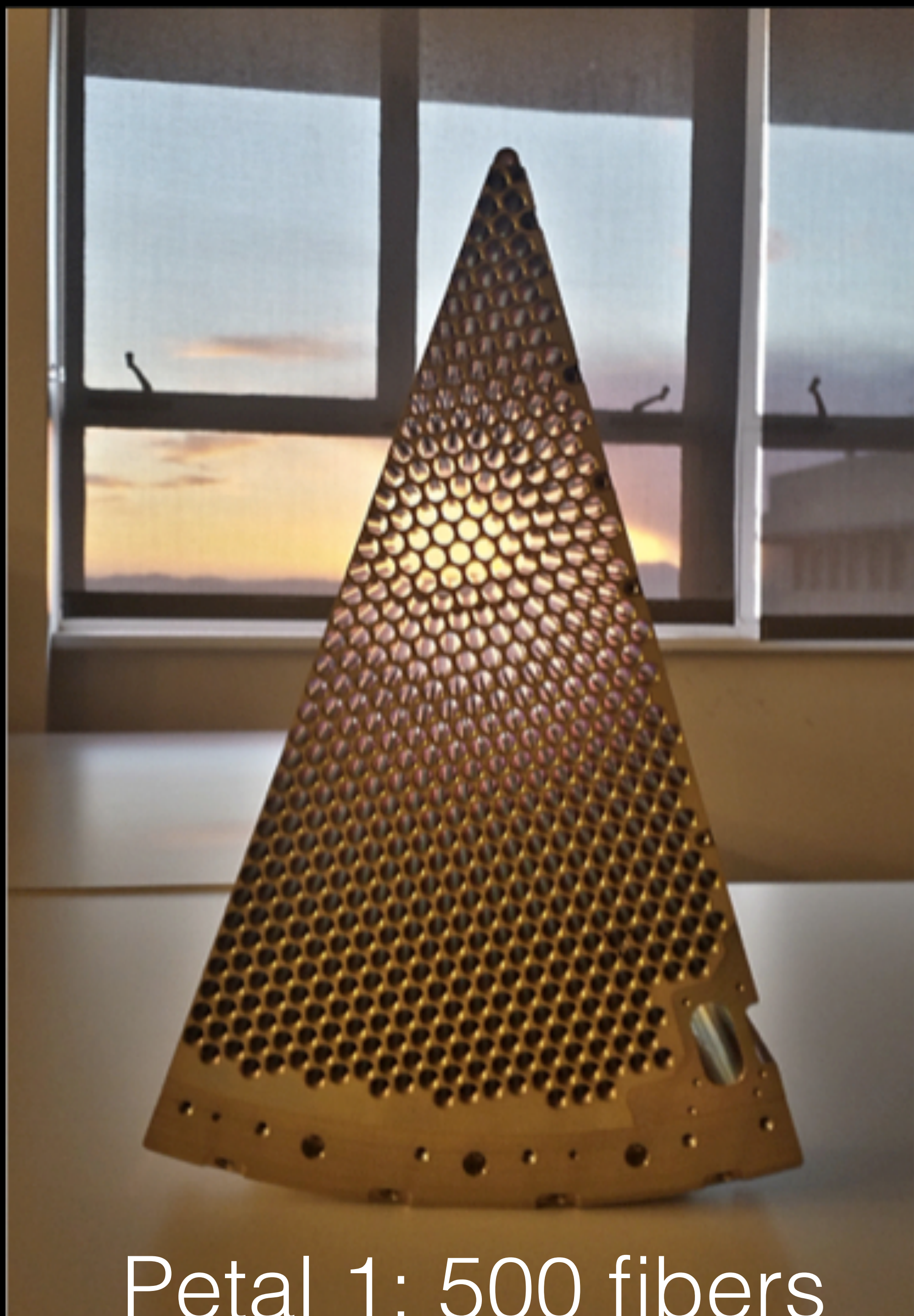


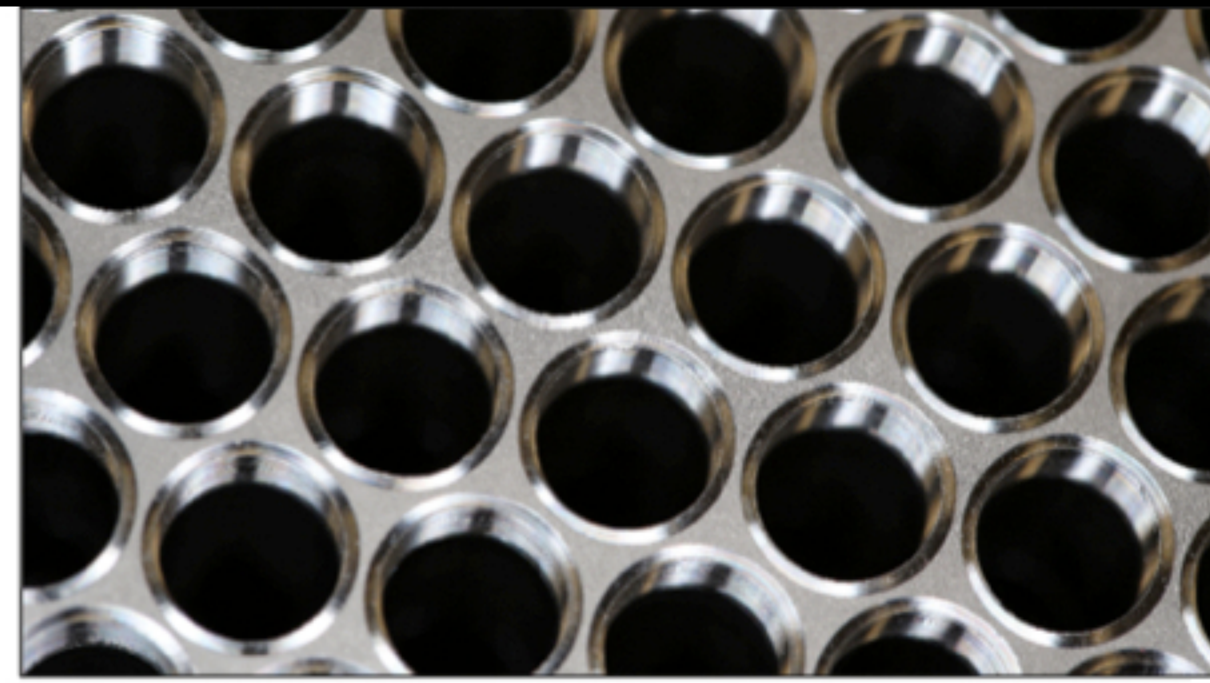



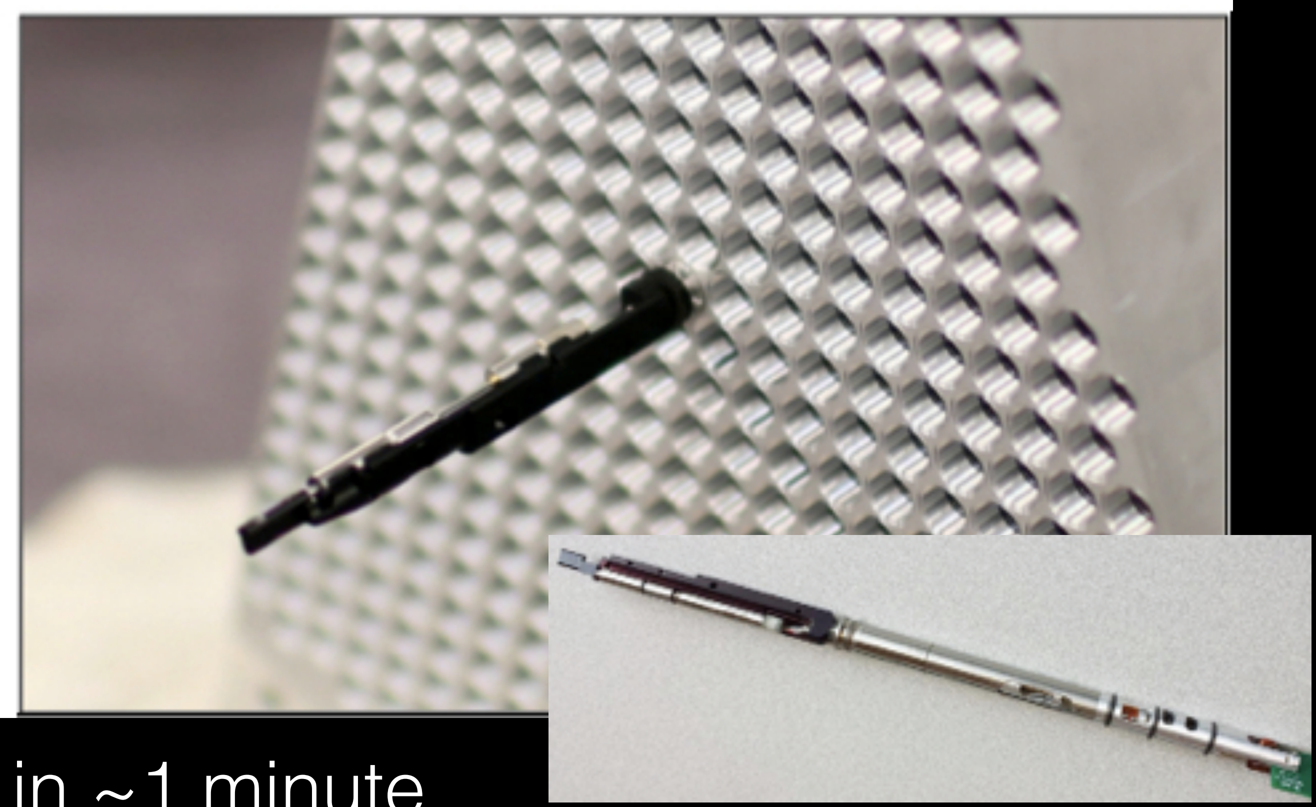

| | d | 2d |
|---|---|---|
| 2PCF | 2/3 | 2 |
| total | 3 5 | 2 6 |
| 13 | | 2 6 |
| 23 | 1 2 | |
| 12 | 2 3 | |

$$2PCF = (GG - RR)/RR$$

# DARK ENERGY SPECTROSCOPIC INSTRUMENT





5k robotic positioners: reconfigure in ~1 minute

http://newscenter.lbl.gov/wp-content/uploads/sites/2/2016/08/first-petal-2.jpg

http://www.symmetrymagazine.org/sites/default/files/images/standard/proto10mm_pos.jpg

**DESI will get 30 million spectra: ELGs, LRGs, Quasars**

**What will we do with them?**

# 2PCF and anisotropic 2PCF

Slepian & Eisenstein 1506.04746—rotating line of sight anisotropic 2PCF with FTs

Slepian & Eisenstein 1510.04809—combining rotating line of sight measurements to cancel wide angle effects

Hand Li Slepian Seljak 1704.02357—using FT method to high multipoles to remove plane of sky systematic

# How much are we leaving on the table?

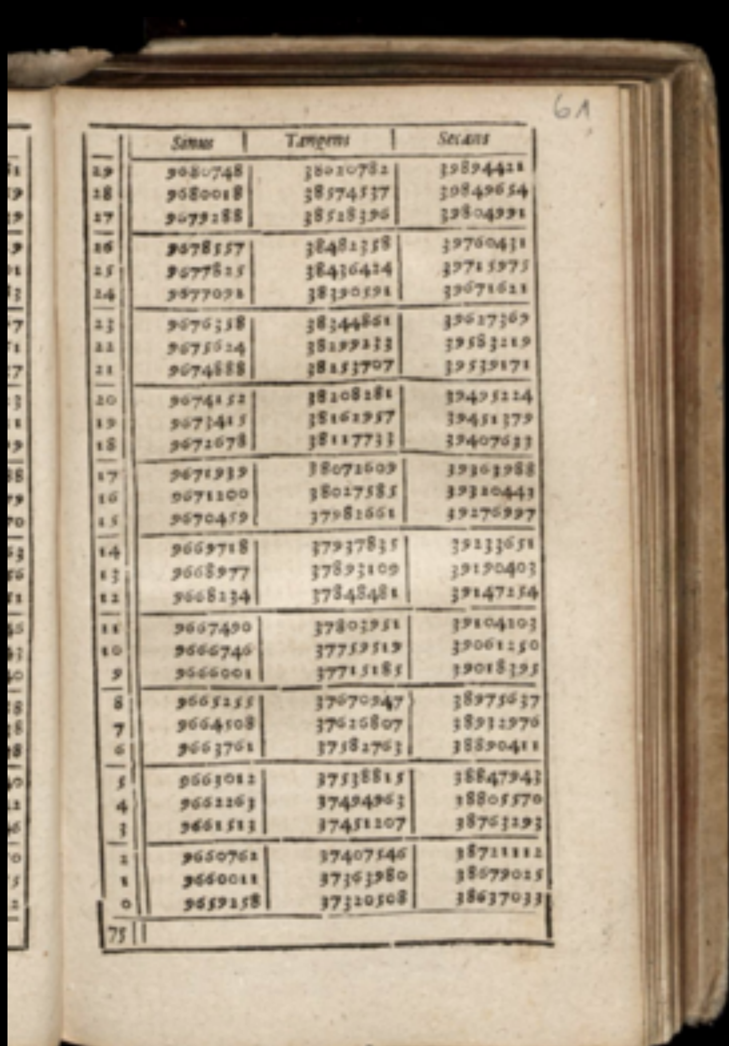
https://www.pinterest.com/pin/165296248795775708/

**How do we extract *all* the information from galaxy positions?**

for a Gaussian Random Field, 2PCF would do so, but much of the interesting part of clustering is exactly in *deviation from GRF*

# Consider an easier question
# How do we specify all the information in the function f(x) = sin x?

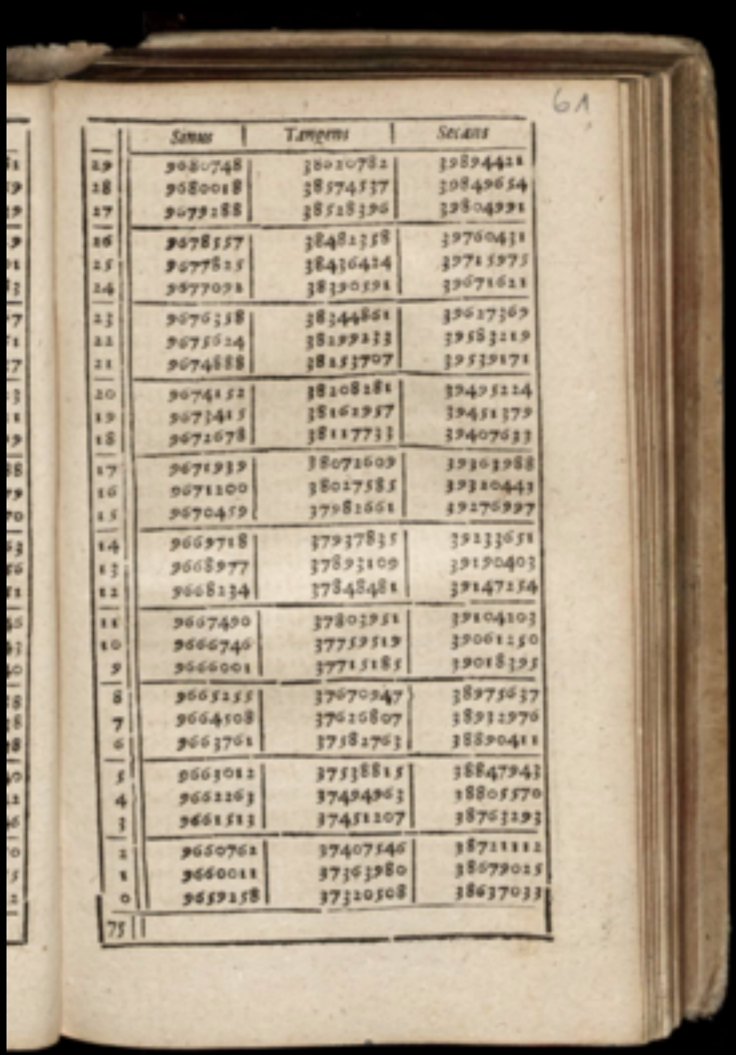
## Table?




https://upload.wikimedia.org/wikipedia/commons/a/a9/Bernegger_Manuale_137.jpg?

# Consider an easier question
# How do we specify all the information in the function f(x) = sin x?

## Table?




https://upload.wikimedia.org/wikipedia/commons/a/a9/Bernegger_Manuale_137.jpg?

## Series!

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

The Edgeworth expansion is the analog for a probability distribution

For almost any distribution, it contains all the information

**The correlation functions contain all the information, and are manifestly translation and rotation-invariant**

Suppose we want to measure *all* of them

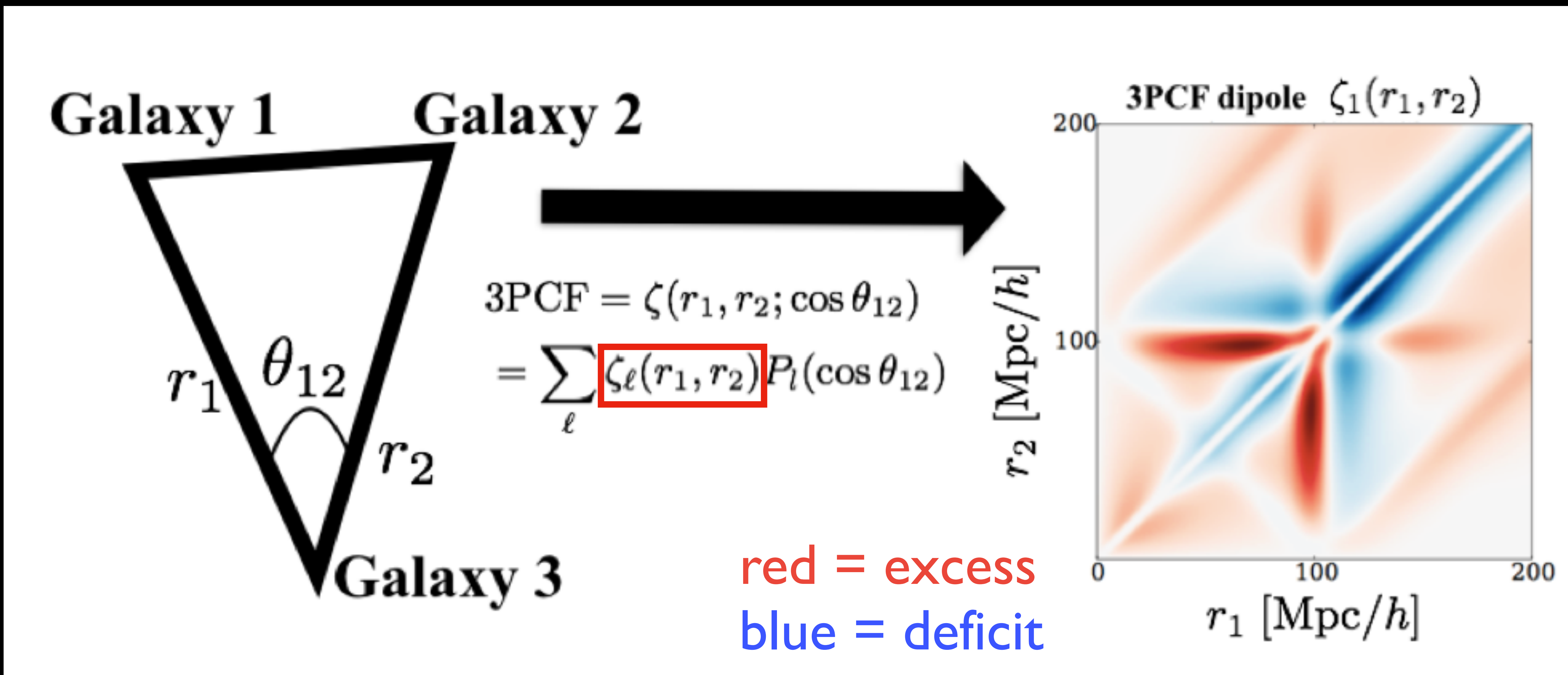
Challenges:
-measurement
-model
-covariance

-What has already been done with data
-Algorithms being developed
-Solutions to these challenges

Fry 1984 suggested using 4PCF

# ISOTROPIC 3PCF ALGORITHM

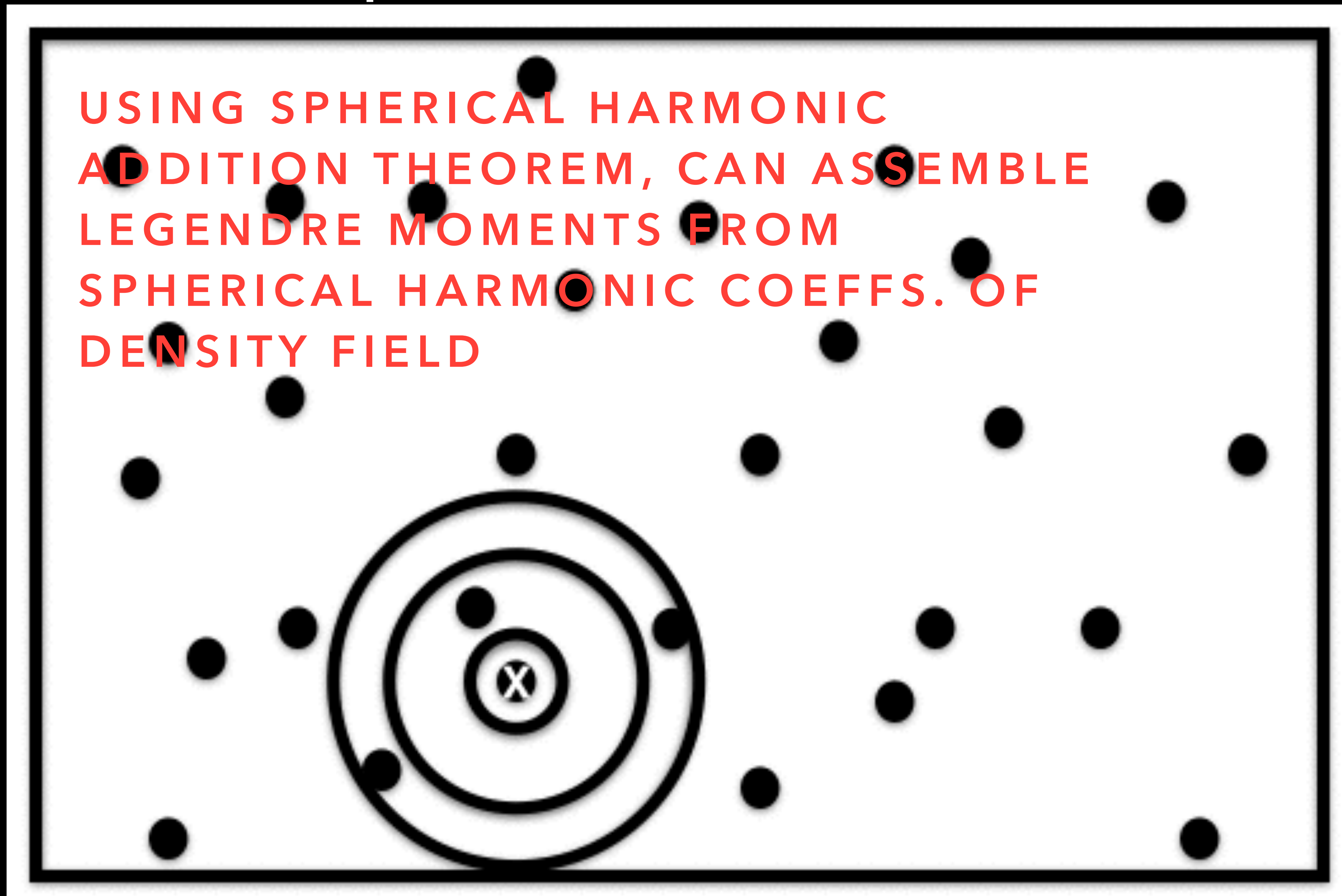
3-point correlation function (3PCF): excess triangles over random




Color represents # of triangles with given side lengths; in this panel, angle dependence is projected onto $P_l$

3PCF basis: Szapudi 2004, also Verde+
Algorithm: SE1506.02040, SE1506.04746

# Around each galaxy, compute $a_{lm}$ in spherical shells/radial bins



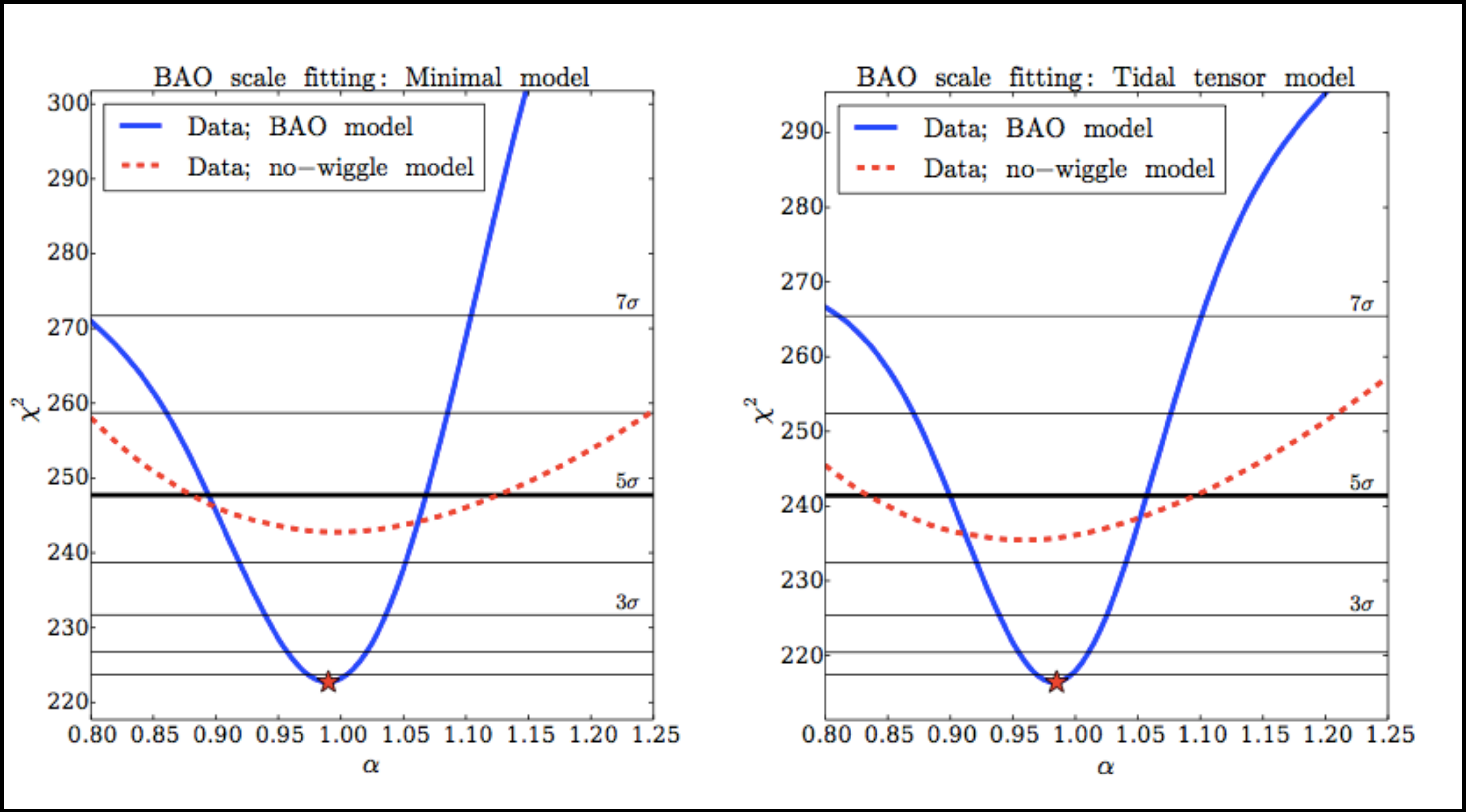

$$a_{lm}(r;\vec{s}) = \sum_{\text{gals } j \text{ in bin}} Y^*_{lm}(\hat{r}_j)$$

NOW ORDER N *ABOUT EACH* GALAXY, SO N$^2$ OVERALL

CAN BE EVEN FASTER WITH FOURIER TRANSFORMS

SE1506.02040, SE1506.04746

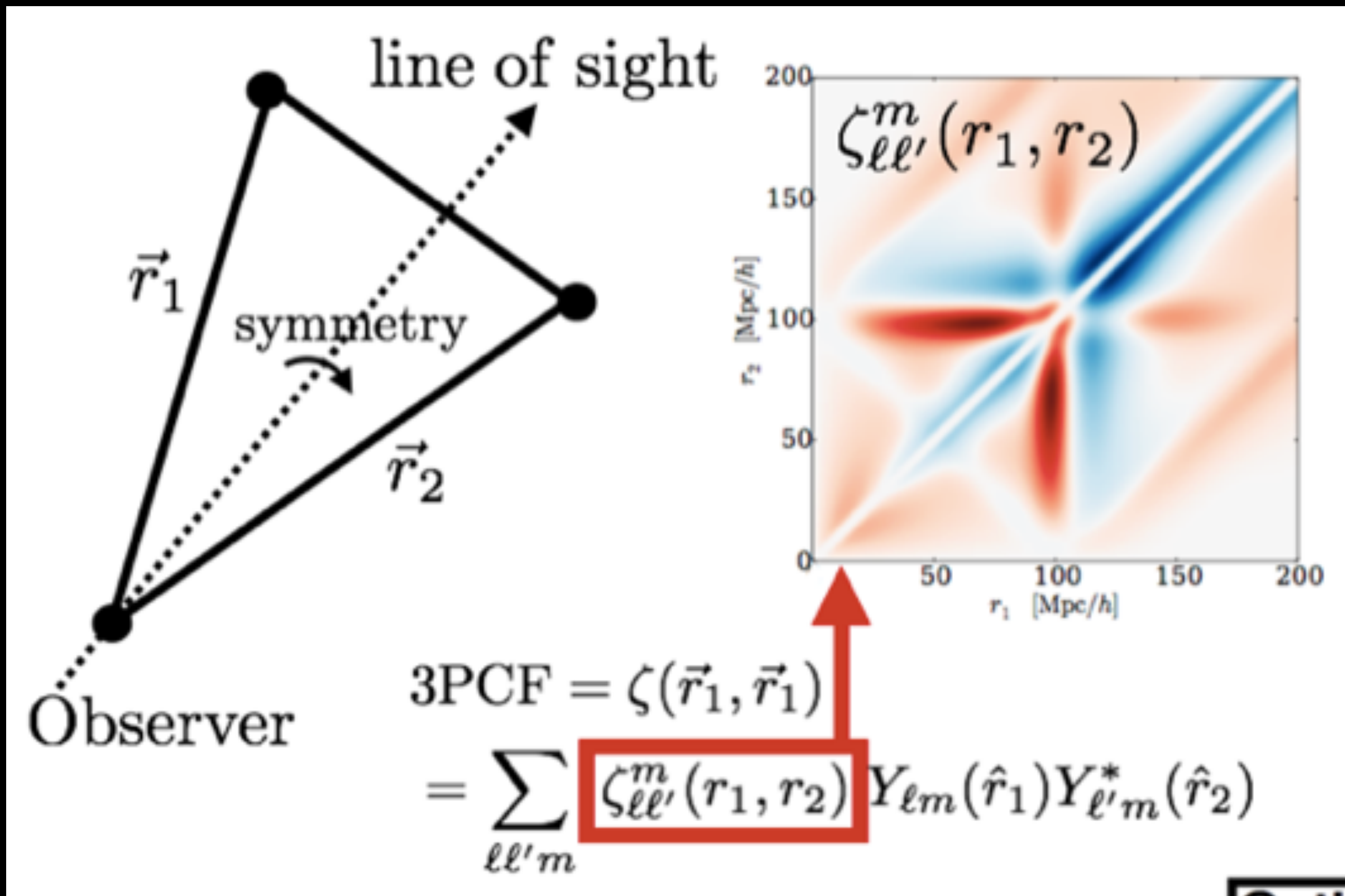
# THE FIRST HIGH-SIGNIFICANCE BAO DETECTION IN THE 3PCF



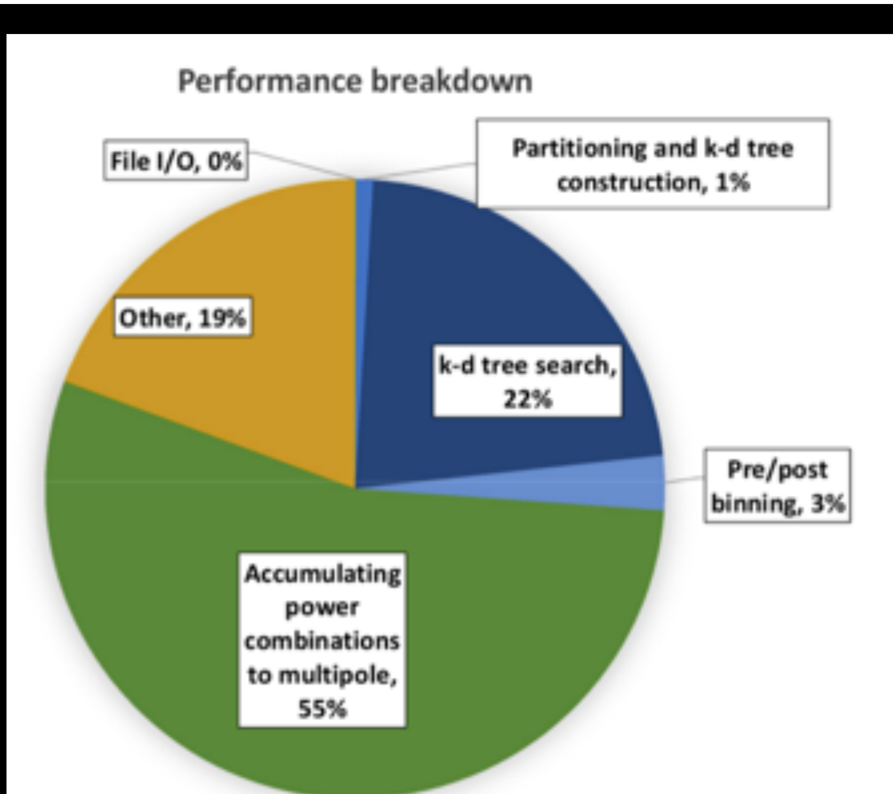

Measure distance to $z = 0.57$ (6 billion years in the past) with 1.7% precision: first use of BAO method in 3PCF

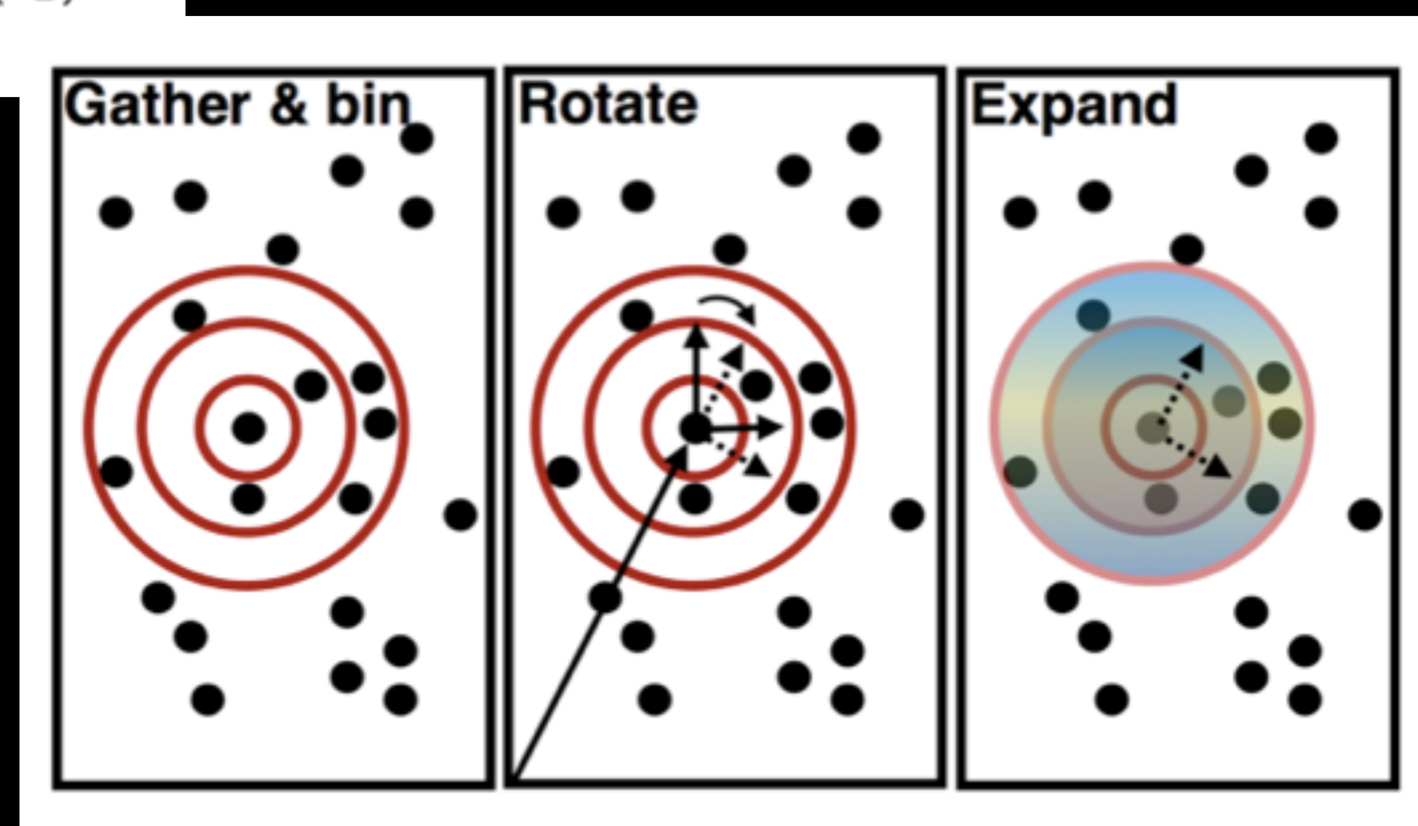
Data: Slepian+1607.06097, 1512.02231,1607.06098
Model: SE1607.03109

# PROBING REDSHIFT-SPACE DISTORTIONS: ANISOTROPIC 3PCF




-Ran at scale on CORI
-9,600 nodes, 5.6 sustained PF
-80% peak for instruction mix
-Obtain harmonic coefficients with matrix algebra libraries
**3PCF for 2 billion haloes in 20 minutes**




Algorithm: SE17 in prep.; Implementation: Friesen+(incl. ZS) 17, accepted ACM: SC'17 conference; First use on data: Nugent+17 in prep.

# GOING TO N POINTS

Take not just two harmonic coefficients, but N-1 of them: form all rotation-invariant combinations

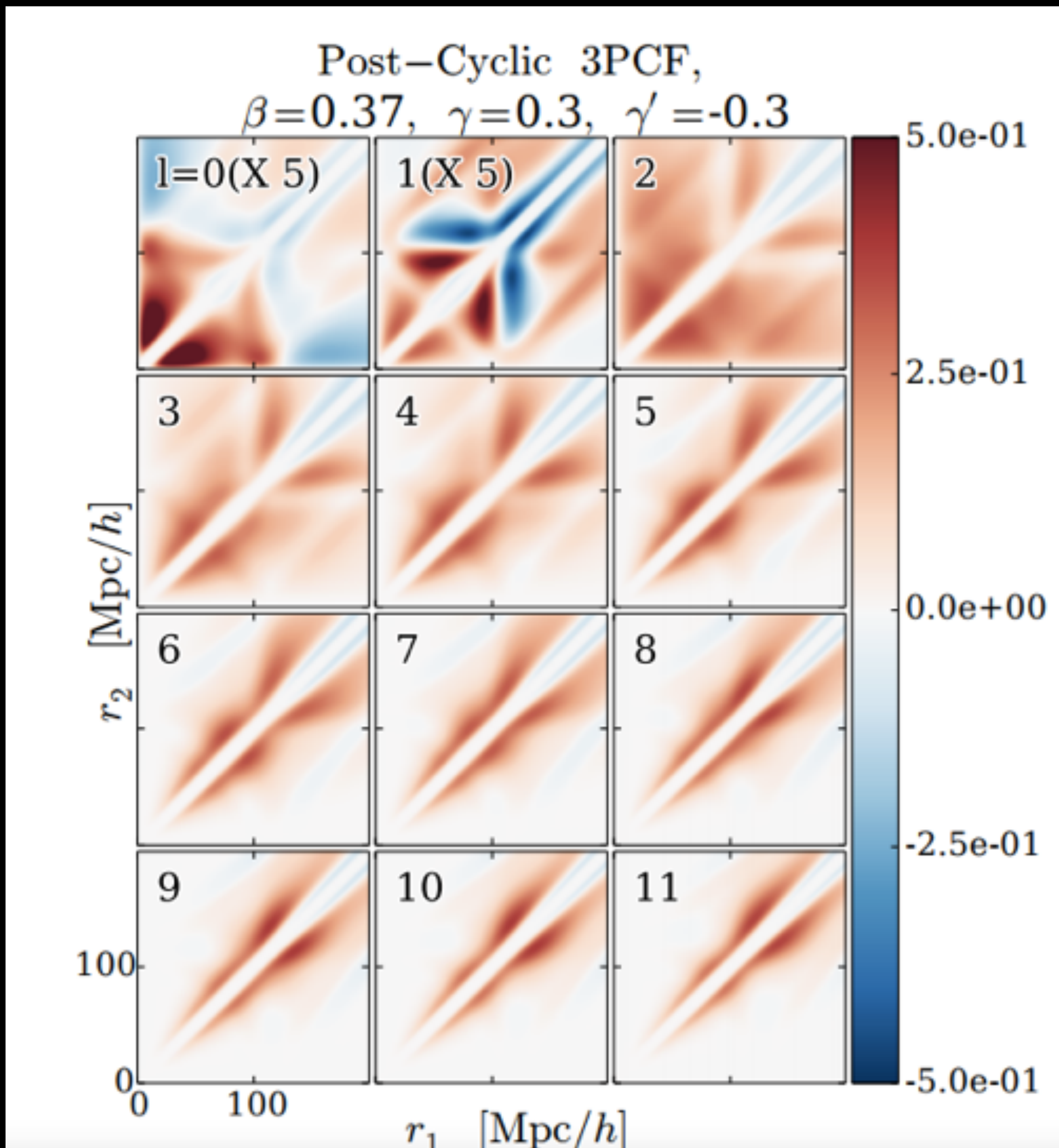
Sum over spins because they pick a preferred direction

$$\text{4PCF harmonics} = \sum_m \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{l_1 m_1}(r_1) a_{l_2 m_2}(r_2) a_{l_3 m_3}(r_3)$$

Weights are just integrals of Wigner D-matrices (describing rotations of spherical harmonics)

Used for CMB bispectrum for a long time (Luo 1993, Spergel & Goldberg 2000, Bartolo+2010), CMB trispectrum (Hu 2001)

# MODELING




-Use perturbation theory
-Isotropic 3PCF: Scoccimarro +1999, Gil-Marin+16
-Anisotropic 3PCF: Rampf & Wong 2012

-NPCF: Tassev 2013, Bertolini+2016

-Use EFT

-Use simulations

SE1607.03109, SE17 in prep.

# AUTO COVARIANCE

$$\mathrm{Cov}_{l_1 l_2 m, l_1' l_2' m'}(r_1, r_2; r_1', r_2') = \frac{(4\pi)^{3/2}}{V}(-1)^{m+m'}(-i)^{l_1+l_2+l_1'+l_2'}$$
$$\times \int r^2 dr \sum_{l_q l_p l_k} \frac{1}{\sqrt{(2l_q+1)(2l_p+1)(2l_k+1)}}$$
$$\times \sum_{J_1 J_2 J_3} \mathcal{D}_{J_1 J_2 J_3} C_{J_1 J_2 J_3} \begin{pmatrix} J_1 & J_2 & J_3 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times \Bigg\{ \xi_{l_k}(r) \Big[ w_1 f^{l_q}_{J_1 l_1 l_1'}(r; r_1, r_1') f^{l_p}_{J_2 l_2 l_2'}(r; r_2, r_2')$$
$$+ w_2 f^{l_q}_{J_1 l_1 l_2'}(r; r_1, r_2') f^{l_p}_{J_2 l_2 l_1'}(r; r_2, r_1') \Big] + \begin{pmatrix} J_1 & J_2 & J_3 \\ S_1 & S_2 & S_3 \end{pmatrix}$$
$$\times \Big\{ f^{l_q}_{J_1 l_1}(r; r_1) \Big[ w_3 f^{l_p}_{J_2 l_2 l_2'}(r; r_2, r_2') f^{l_k}_{J_3 l_1'}(r; r_1') \delta^K_{S_1 -m, S_3 -m'}$$
$$+ w_4 f^{l_p}_{J_2 l_2 l_1'}(r; r_2, r_1') f^{l_k}_{J_3 l_2'}(r; r_2') \delta^K_{S_1 -m, S_3 m'} \Big]$$
$$+ f^{l_p}_{J_2 l_2}(r; r_2) \Big[ w_5 f^{l_q}_{J_1 l_1 l_2'}(r; r_1, r_2') f^{l_k}_{J_3 l_1'}(r; r_1') \delta^K_{S_2 m, S_3 -m'}$$
$$+ w_6 f^{l_q}_{J_1 l_1 l_1'}(r; r_1, r_1') f^{l_k}_{J_3 l_2'}(r; r_2') \delta^K_{S_2 m, S_3 m'} \Big] \Big\} \Bigg\}.$$

$$f^l_{nm}(r; r_i) = \int \frac{k^2 dk}{2\pi^2} P_l(k) j_n(kr) j_m(kr_i)$$
$$f^l_{nmj}(r; r_i, r_j') = \int \frac{k^2 dk}{2\pi^2} P_l(k) j_n(kr) j_m(kr_i) j_j(kr_j').$$

-Gaussian Random Field piece is always leading contribution to covariance, so can always compute in terms of 1 and 2-D angular momentum-weighted integrals of linear power spectra (or their Kaiser-formula multipoles)

-Quite fast to compute (hours on laptop)

-Tested against mocks and works well on large scales (> 20 Mpc)

-Can also compute directly or by jack-knifing

Isotropic 3PCF: SE1506.02040, Anisotropic (above): SE17 in prep.

# CROSS COVARIANCE

How do you handle e.g. 3PCF X 4PCF covariance?

Compress down to parameters you measure (e.g. biases, BAO scale, *f*, $\sigma_8$) and use mock catalogs to obtain this low-*d* matrix

Since it is low-*d*, don’t need many mocks to well-determine its inverse

# SUMMARY

Use higher point correlation functions to systematically extract all the information there is from galaxy surveys

There are challenges but they are solvable

Let’s not leave any information on the table

Thanks!