MAXIMIZING INFORMATION EXTRACTION FROM NEXT GENERATION SURVEYS
2.4 m at Apache point

CMASS sample
Why make a map?
To enable discovery
To find patterns
To enable discovery
To find patterns

Homogeneous

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

2PCF = (GG - RR)/RR
DARK ENERGY SPECTROSCOPIC INSTRUMENT

Petal 1: 500 fibers

5k robotic positioners: reconfigure in ~1 minute
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?
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?
Consider an easier question
How do we specify all the information in
the function $f(x) = \sin x$?

**Table?**

**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 $P_I$

**Equation:**

$$3PCF = \zeta(r_1, r_2; \cos \theta_{12}) = \sum_\ell \zeta_\ell(r_1, r_2) P_I(\cos \theta_{12})$$

**Legend:**
- Red = excess
- Blue = deficit

3PCF basis: Szapudi 2004, also Verde+
Algorithm: SE1506.02040, SE1506.04746
Around each galaxy, compute $a_{lm}$ in spherical shells/radial bins

Using spherical harmonic addition theorem, can assemble Legendre moments from spherical harmonic coeffs. of density field

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

Now order $N$ about each galaxy, so $N^2$ overall can be even faster with Fourier transforms
The first high-significance BAO detection in the 3PCF

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

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

$$4\text{PCF harmonics} = \sum_{m} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{l_1m_1}(r_1)a_{l_2m_2}(r_2)a_{l_3m_3}(r_3)$$

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)
- Use perturbation theory
- Isotropic 3PCF: Scoccimarro +1999, Gil-Marín+16
- Anisotropic 3PCF: Rampf & Wong 2012
- NPCF: Tassev 2013, Bertoloni+2016
- Use EFT
- Use simulations
**AUTO COVARIANCE**

- 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

\[
\text{Cov}_{l_1 l_2 m, l'_1 l'_2 m'}(r_1, r_2; r'_1, r'_2) = \frac{(4\pi)^{3/2}}{V} (-1)^{m + m'} (-i)^{l_1 + l_2 + l'_1 + l'_2} \\
\times \int r^2 dr \sum_{l_q l_p l_k} \frac{1}{\sqrt{(2l_q + 1)(2l_p + 1)(2l_k + 1)}} \\
\times \sum_{J_1 J_2 J_3} D_{J_1 J_2 J_3} C_{J_1 J_2 J_3} \left( \begin{array}{ccc}
J_1 & J_2 & J_3 \\
0 & 0 & 0
\end{array} \right) \\
\times \left\{ \xi_{l_1 l_2 l_3} \left[ w_1 f_{J_1 l_1 l_2}^l(r; r_1, r'_1) f_{J_2 l_2 l_3}^l(r; r_2, r'_2) + w_2 f_{J_1 l_1 l_2}^l(r; r_1, r'_2) f_{J_2 l_2 l_3}^l(r; r_2, r'_1) \right] + \left( \begin{array}{ccc}
J_1 & J_2 & J_3 \\
S_1 & S_2 & S_3
\end{array} \right) \\
\times \left\{ f_{J_1 l_1}^l(r; r_1) \left[ w_3 f_{J_2 l_2 l_3}^l(r; r_2, r'_2) f_{J_3 l_3}^l(r; r'_1) \delta^K_{S_1 - m, S_3 - m'} + w_4 f_{J_2 l_2 l_3}^l(r; r_2, r'_1) f_{J_3 l_3}^l(r; r'_2) \delta^K_{S_1 - m, S_3 m'} \right] \\
+ f_{J_2 l_2}^l(r; r_2) \left[ w_5 f_{J_3 l_3}^l(r; r_1, r'_1) f_{J_2 l_2 l_3}^l(r; r'_2) \delta^K_{S_2 m, S_3 - m'} + w_6 f_{J_3 l_3}^l(r; r_1, r'_2) f_{J_2 l_2 l_3}^l(r; r'_1) \delta^K_{S_2 m, S_3 m'} \right] \right\} \right\}.
\]

\[
f_{n m}^l(r; r_i) = \int \frac{k^2 dk}{2\pi^2} P_l(k) j_n(kr) j_m(kr_i) \\
f_{n m j}^l(r; r_i, r'_j) = \int \frac{k^2 dk}{2\pi^2} P_l(k) j_n(kr) j_m(kr_i) j_j(kr'_j).
\]

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!