
Parity Violation in the Large Scale Structure

Cristiano G. Sabiu · ++

University of Seoul · NSRI

2026/08/14 - Group Report, UoS

The correlation-function hierarchy

$$\underbrace{\xi(r) = \langle \delta_1 \delta_2 \rangle}_{\text{pairs}},$$

$$\underbrace{\zeta(r_{12}, r_{13}, r_{23}) = \langle \delta_1 \delta_2 \delta_3 \rangle}_{\text{triangles}},$$

$$\underbrace{\zeta_{\text{conn}}^{(4)} = \langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle_c}_{\text{tetrahedra (connected)}}$$

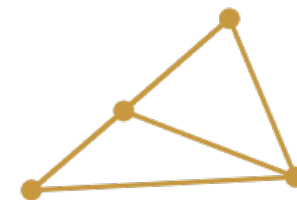
- ▶ Each rung adds a new geometric **shape** degree of freedom: separation $\rightarrow$ triangle $\rightarrow$ tetrahedron.
- ▶ Higher orders carry **non-linear growth**, **tidal fields**, and the **bias** expansion that ξ cannot see.
- ▶ **This talk:** treat them as a ladder $\xi \rightarrow +\zeta \rightarrow +\zeta^4_{\text{conn}}$ and count what each rung adds.



ξ : pair



ζ : triangle



$\zeta^{(4)}$: tetrahedron

The standard route is Fourier — the bispectrum

- ▶ The 3-point function's Fourier partner, the **bispectrum** $B(k_1, k_2, k_3)$, breaks the linear bias–amplitude degeneracy [Sefusatti+ 2006].
- ▶ Measured in BOSS/eBOSS [Gil-Marín+ 2017]; forecast extensively on Quijote.
- ▶ Landmark: the halo bispectrum roughly **doubles** the M_v constraint over $P(k)$ alone [Hahn+ 2020].
- ▶ A zoo of non-Gaussian summaries followed — marked $P(k)$, wavelet scattering, nearest-neighbour, voids.
- ▶ **Almost all of it is done in Fourier space.**

Configuration space is complementary

Signal is **local** in pair separation; BAO & scales cleanly localised.

Windows / masks / selection handled by the **random catalogue**.

Large scales connect **directly** to perturbation theory.

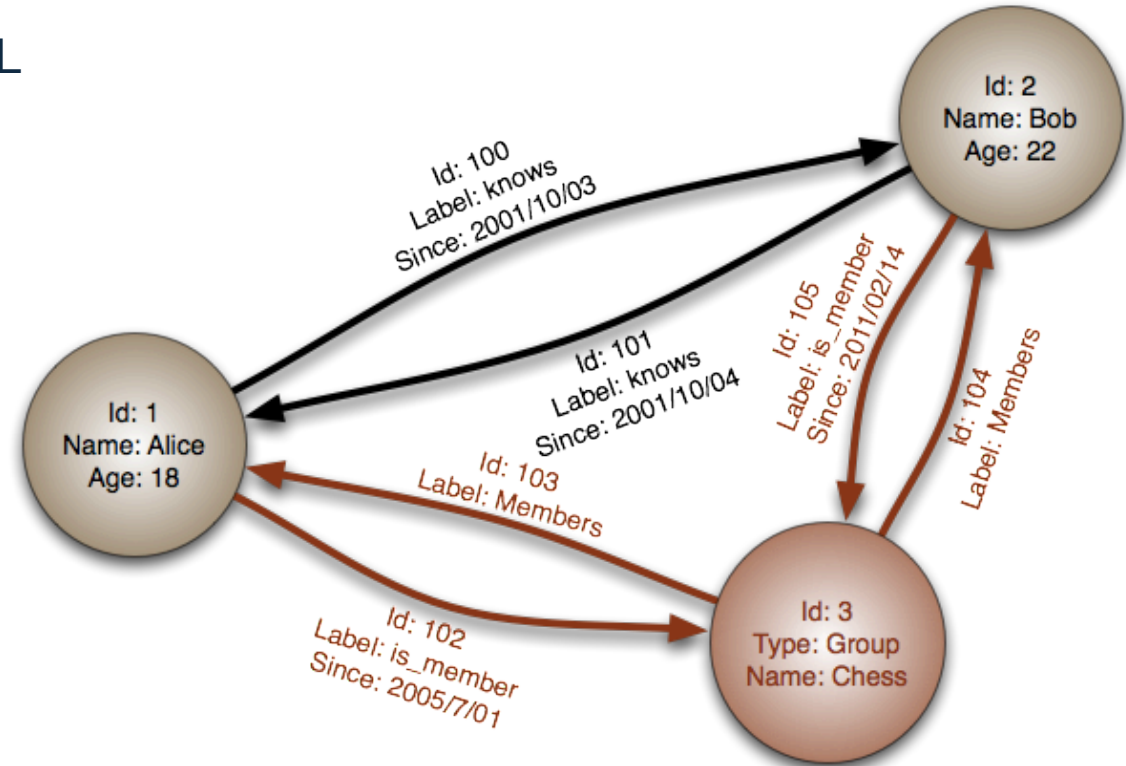
The obstacle was always speed: an N -point count scales as $O(N^n)$.

The engine: GRAMSCI on the GPU

Part 2 of the talk

Measuring it: statistics from a graph

- ▶ **The graph.** In 2009 Facebook gave up using their MySQL storage and moved to a graph structure
- ▶ A directed graph: relationships depend on direction
- ▶ **Nodes:** people, places, groups
- ▶ **Node properties:** name, age, etc
- ▶ **Relationships:** membership, known since

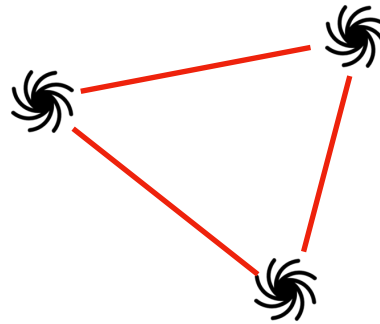


Measuring it: statistics from a graph

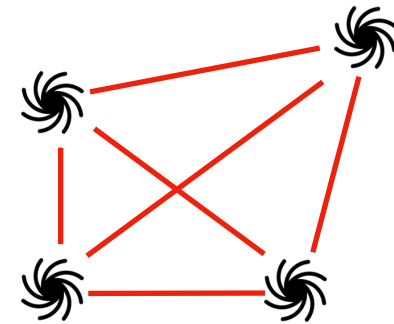
2-point



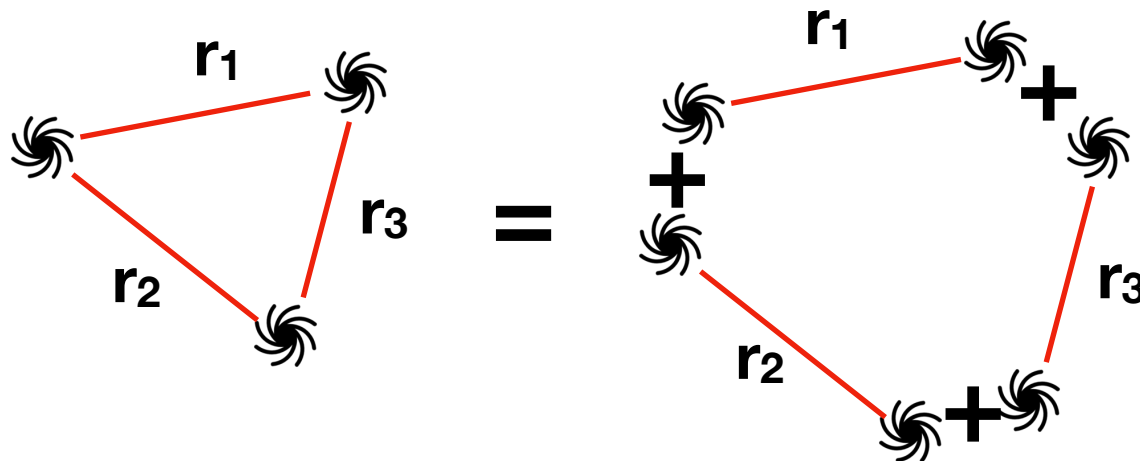
3-point



4-point



- ▶ The computationally expensive part of measuring higher order statistics is in making sure the distances between data points satisfy our specific criteria i.e. r_1 , r_2 , r_3 , etc
- ▶ However we notice that all higher order statistics are just complex combinations of pairs

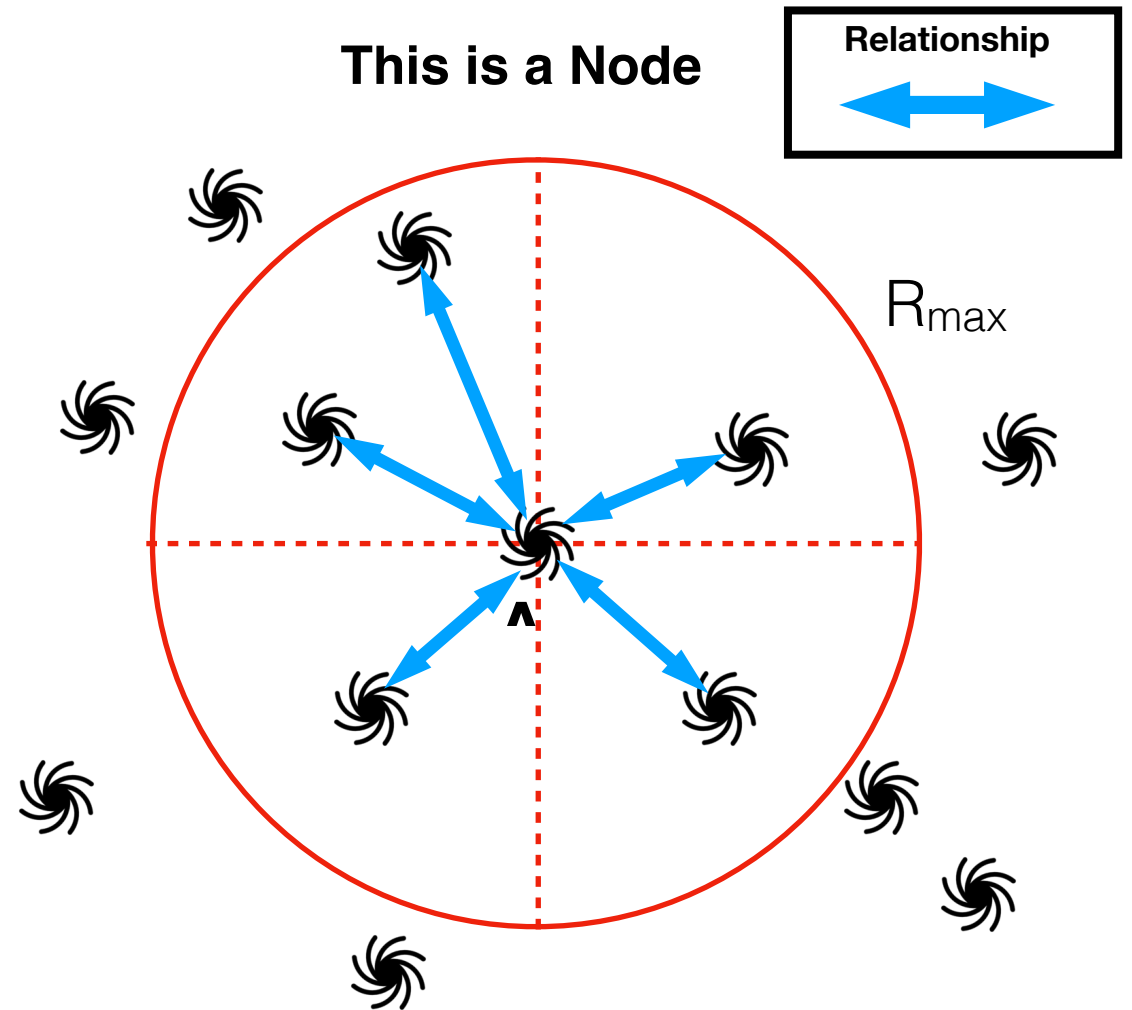


Key concept

So rather than treat data positions as the important feature, rather treat each data pair (relationship) as the main information!

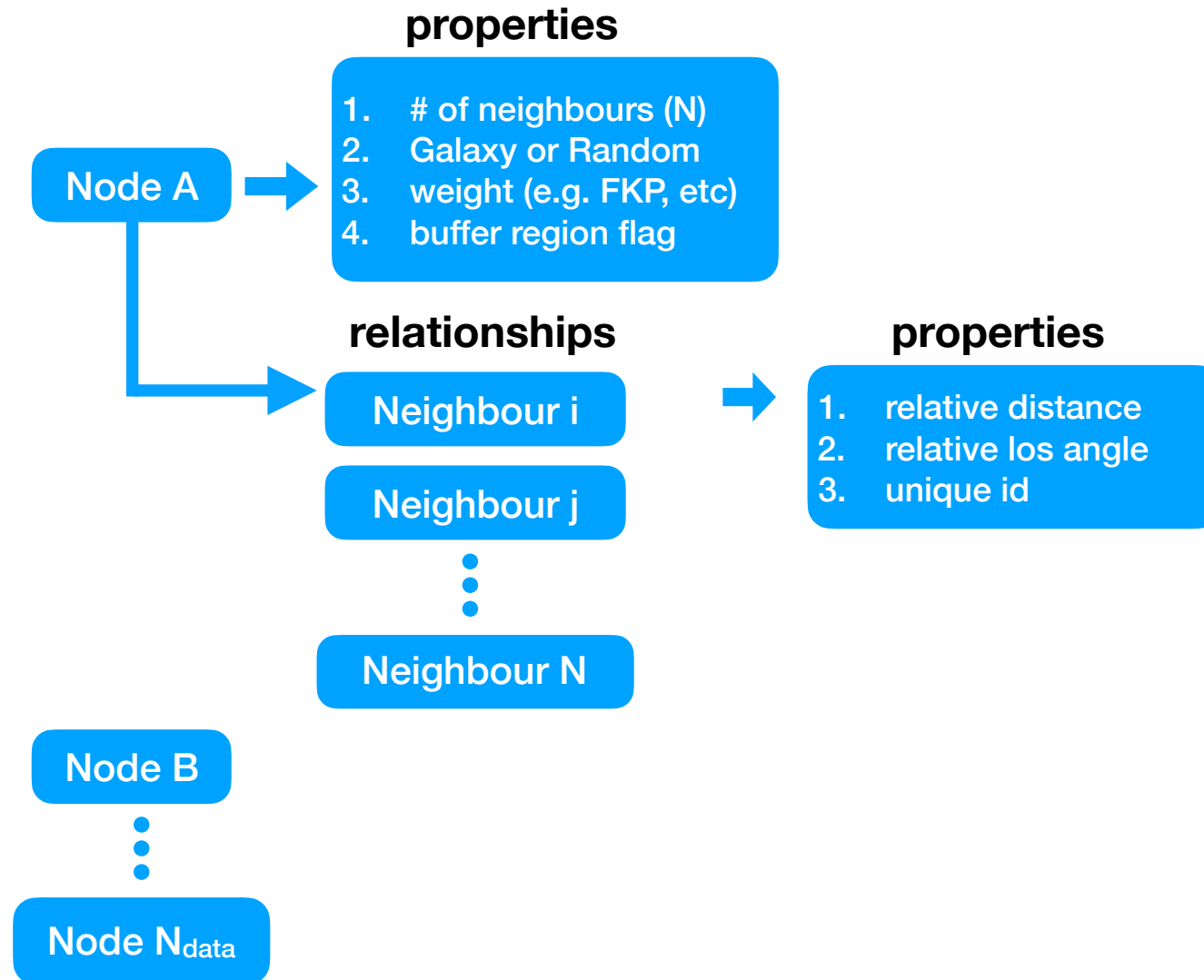
Measuring it: statistics from a graph

- ▶ Each galaxy (or random) point is a **node** which may have relationships to there nodes
- ▶ For our purposes the relationship information is the distance to neighbours within $R_{\max}$
- ▶ Neighbour list can be obtained quickly using a **kd-tree** algorithm



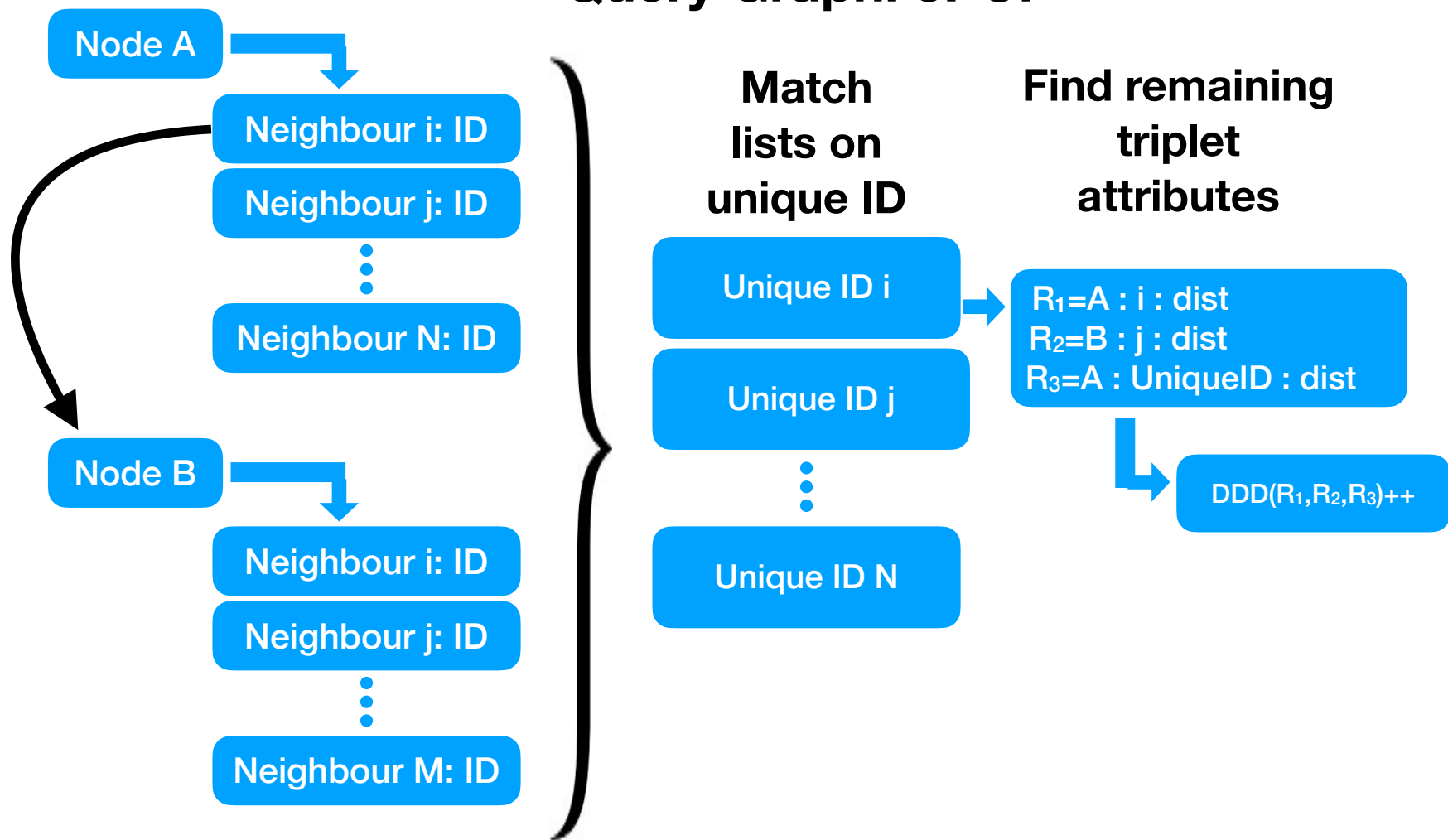
Measuring it: statistics from a graph

Graph Database Structure



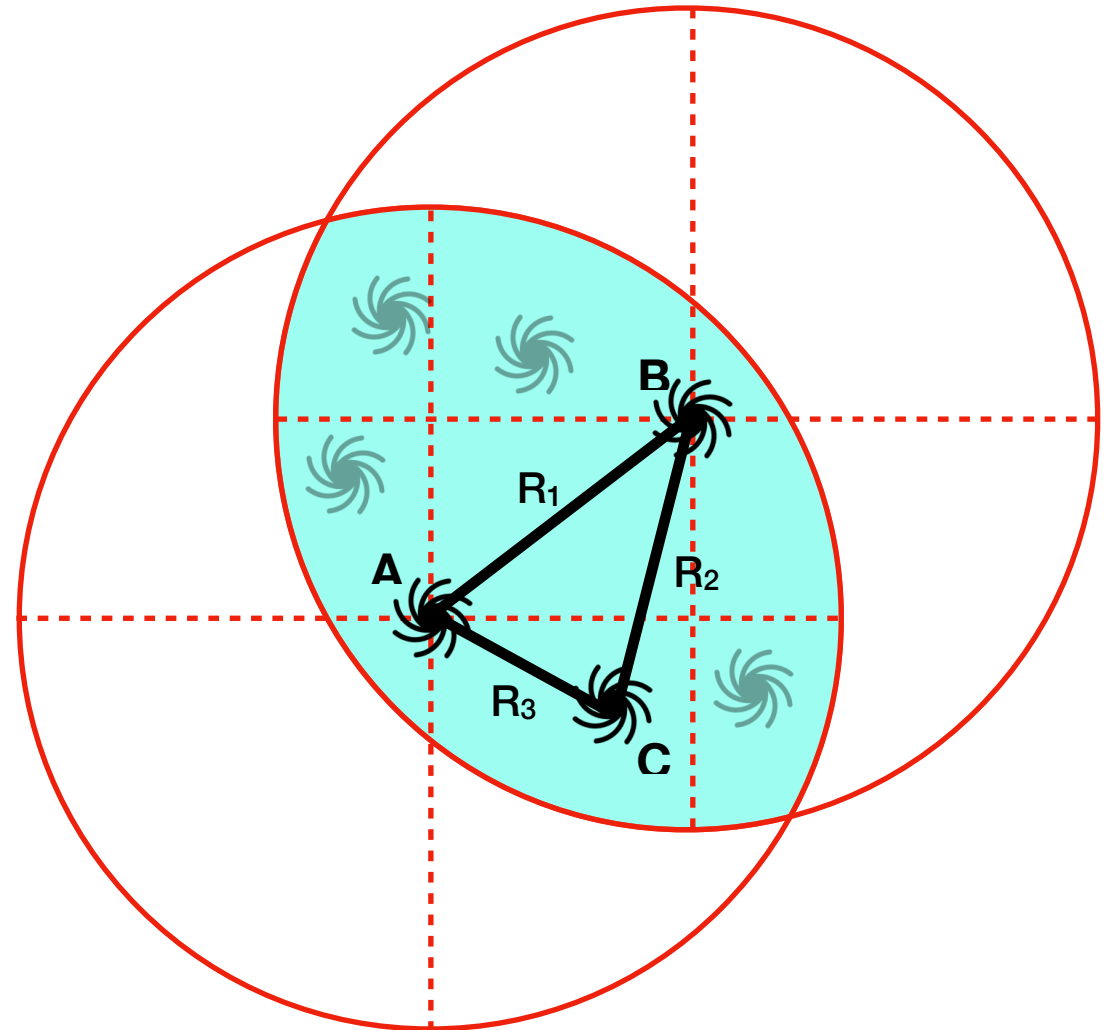
Measuring it: statistics from a graph

Query Graph: 3PCF



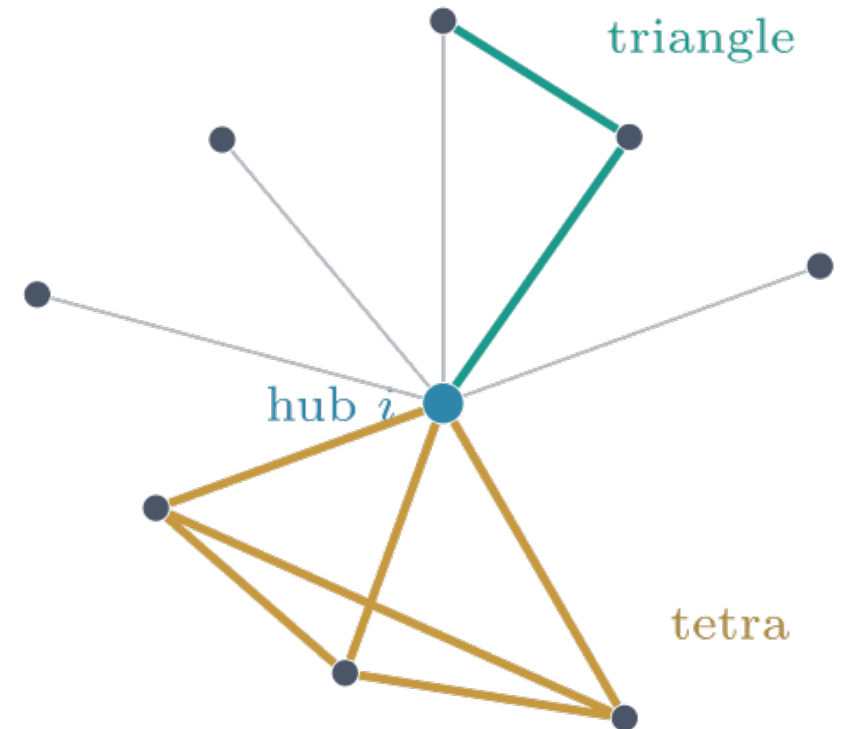
Measuring it: statistics from a graph

- ▶ Each galaxy (or random) point is a **node** which may have **relationships** to there nodes
- ▶ For our purposes the relationship information is the distance to neighbours within **$R_{\max}$**
- ▶ Neighbour list can be obtained quickly using a **kd-tree** algorithm

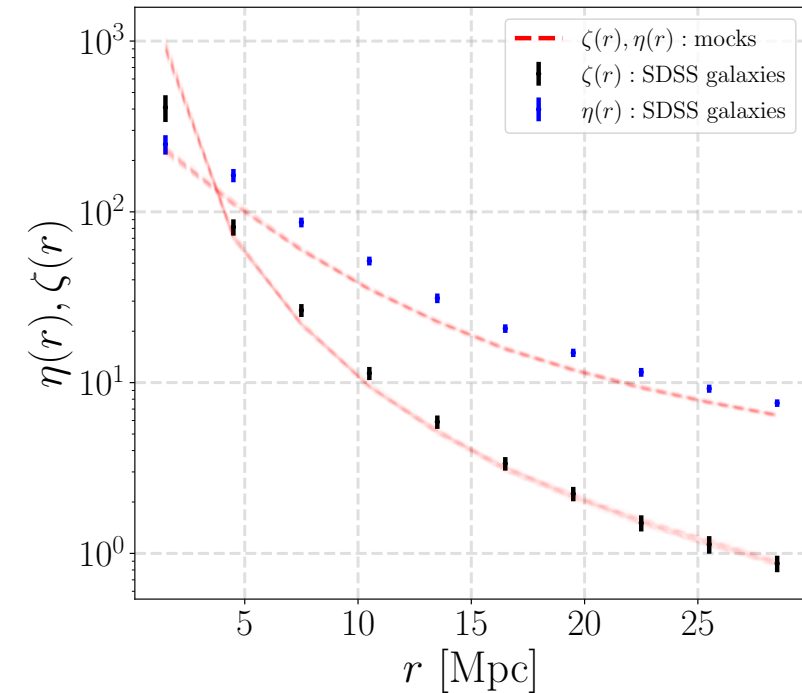
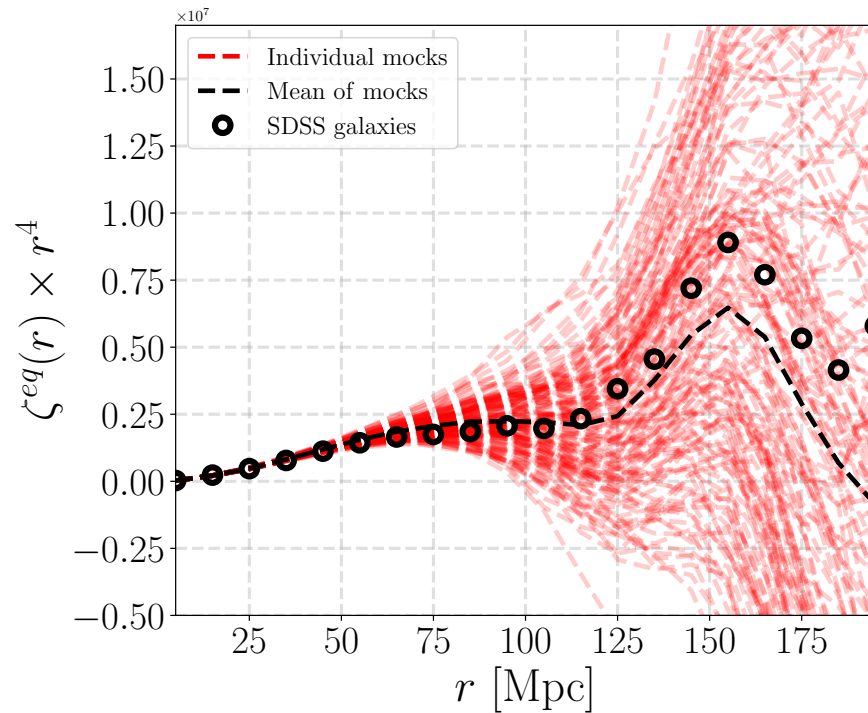


Measuring it: statistics from a graph

- ▶ **The estimator.** Minimum-variance, edge-corrected **Szapudi–Szalay** counts: data ($+1/N_D$) and random ($-1/N_R$) points carry **signed** weights.
- ▶ **The algorithm — GRAMSCI [Sabiu et.al., ApJS 2019].**
- ▶ Build a kd-tree **once**; store sorted adjacency lists — a graph.
- ▶ **Discard the coordinates.** All 2-/3-/4-point queries run on the graph.
- ▶ Enumerate sub-graphs: hub-and-spoke → triangles → tetrahedra. Memory $O(Nm)$, cost $O(Nm^{k-1})$ vs. naive $O(N^k)$.



Measuring it: statistics from a graph



- **Gramsci v1**: allowed the 3pcf in configuration space up to and beyond the BAO scale
- It also provided the first measurement of the 4pcf of SDSS galaxies
- I thought it could not be optimised further. Then Claude code came out with a new model... *Fable*

GRAMSCI v2 — five upgrades

1. **Merge-walk**: an $O(m)$ sorted-set intersection replaces binary search. CPU speedup 1.5–2× (3PCF), 2.8–3.8× (4PCF).
2. **Parity-decomposed 4PCF**: even/odd channels from the signed tetrahedron volume — a clean probe of **parity violation**.
3. **Connected 4PCF on the fly**: the disconnected part is subtracted **internally**, so ζ^4_{conn} is returned directly.
4. **Python/NumPy** interface to the Fortran engine.
5. **Full GPU port** (OpenACC) — **principal**.

Point (3) is the whole ladder

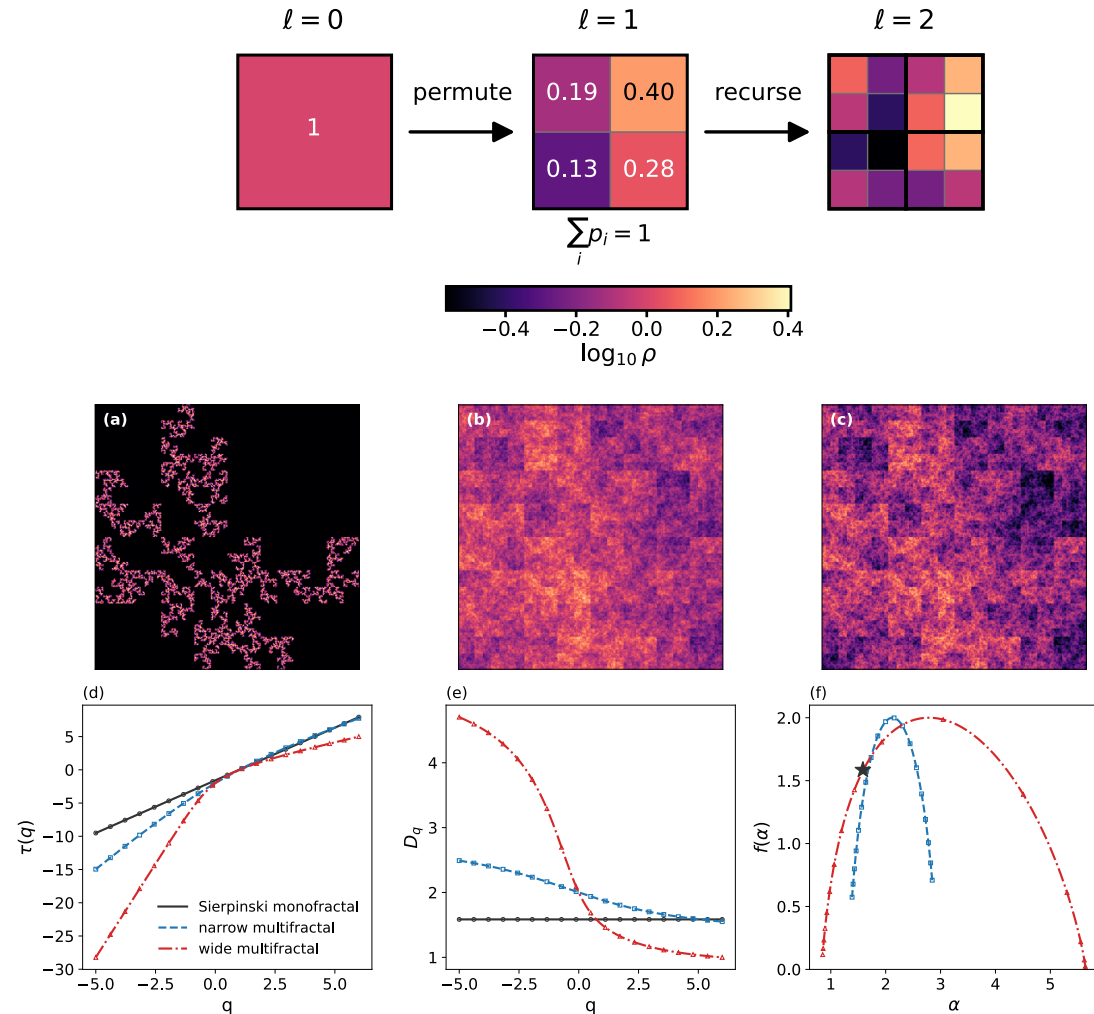
The on-the-fly connected 4PCF is **exactly** the ingredient Parts I & II rest on — ζ^4_{conn} isolated from its dominant Gaussian part, with no extra catalogue pass.

Validated

Bit-for-bit vs. the CPU reference; analytic injection tests; a parity null on Quijote; a BAO-scale DESI DR1 application.

GRAMSCI v2 — analytic test

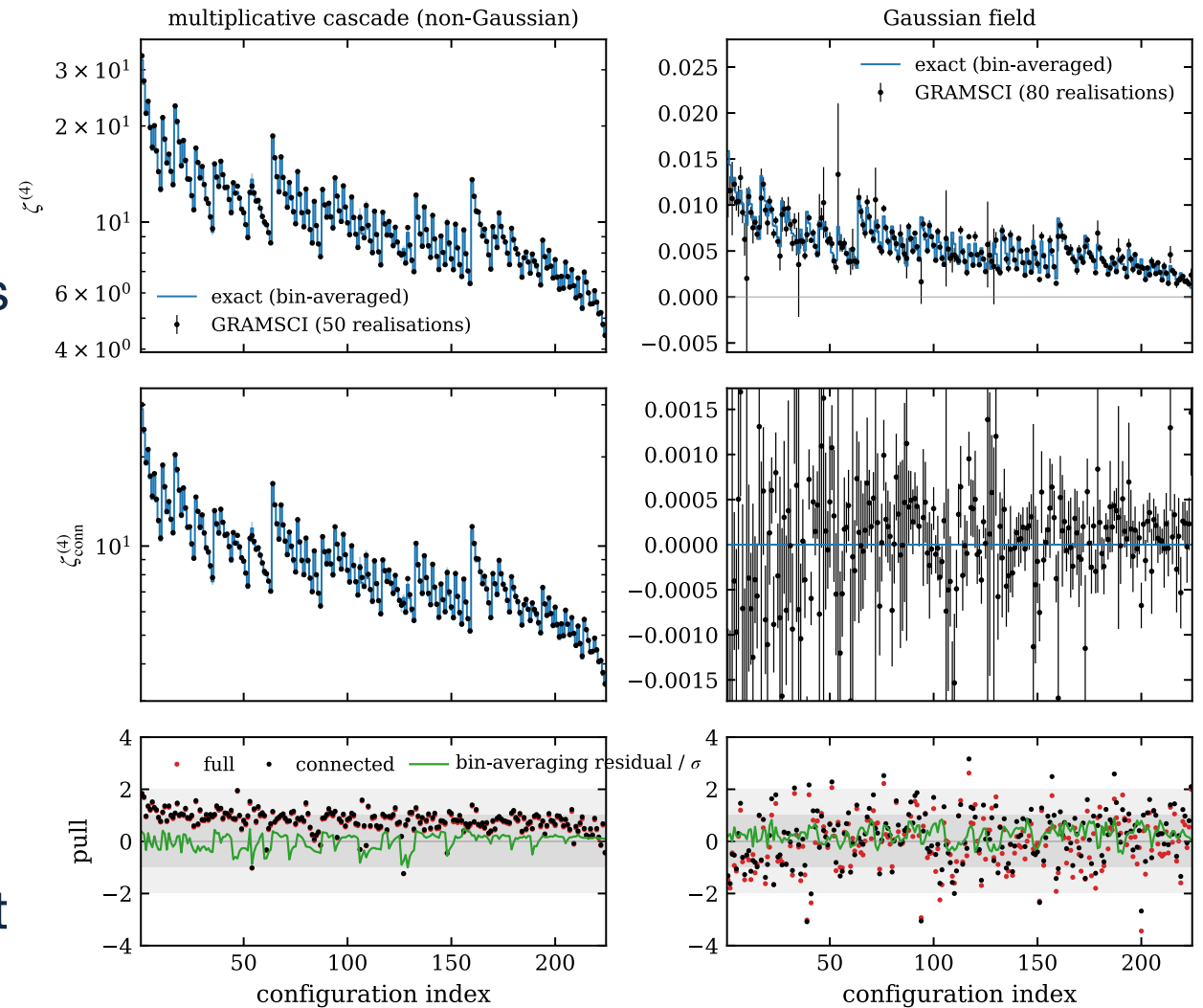
- ▶ There are very few ‘interesting’ distributions with known analytic N -point functions
- ▶ The trivial gaussian case for example has zero connected 4pcf
- ▶ The **multiplicative cascade**, canonical example from turbulence theory, has an exact analytic solution for the generalised Renyi dimensions and multifractal spectrum
- ▶ Greiner et.al. 1998 introduced a transfer operator that can act on the ancestral tree of the cascade providing the N -point functions



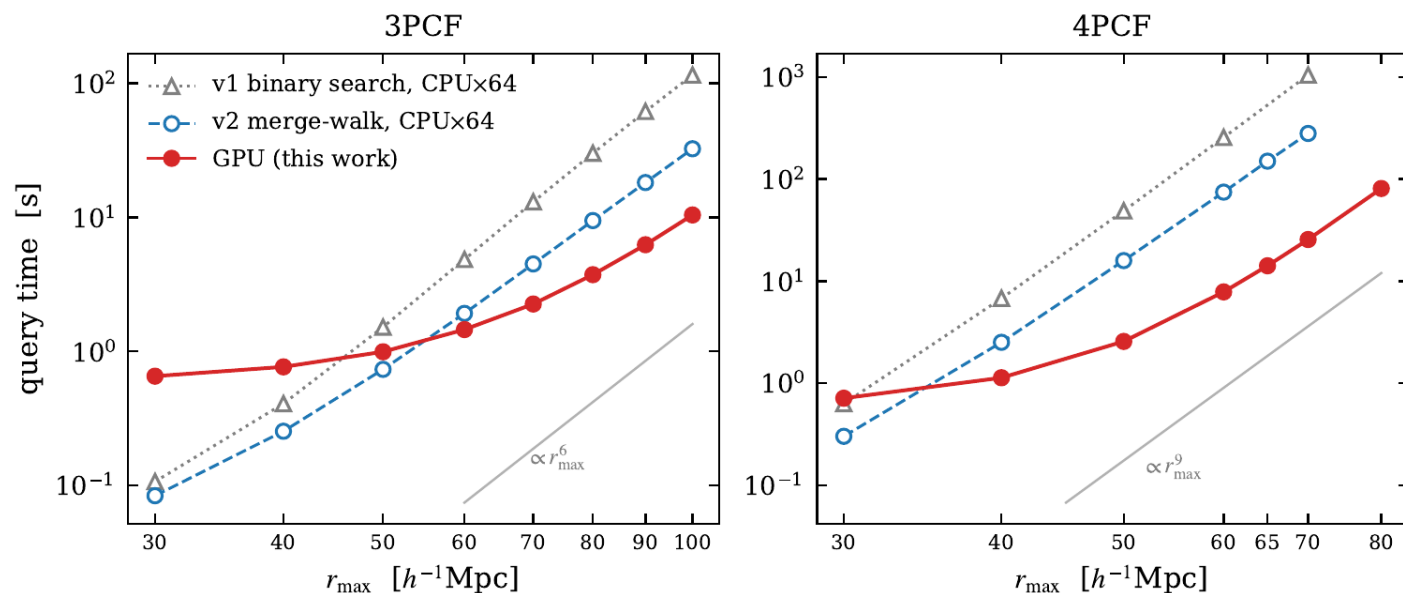
https://github.com/csabiu/multiplicative_cascade

GRAMSCI v2 — 4pcf correctness test

- ▶ There are very few ‘interesting’ distributions with known analytic N -point functions
- ▶ The trivial gaussian case for example has zero connected 4pcf
- ▶ The **multiplicative cascade**, canonical example from turbulence theory, has an exact analytic solution for the generalised Renyi dimensions and multifractal spectrum
- ▶ Greiner et.al. 1998 introduced a transfer operator that can act on the ancestral tree of the cascade providing the N -point functions



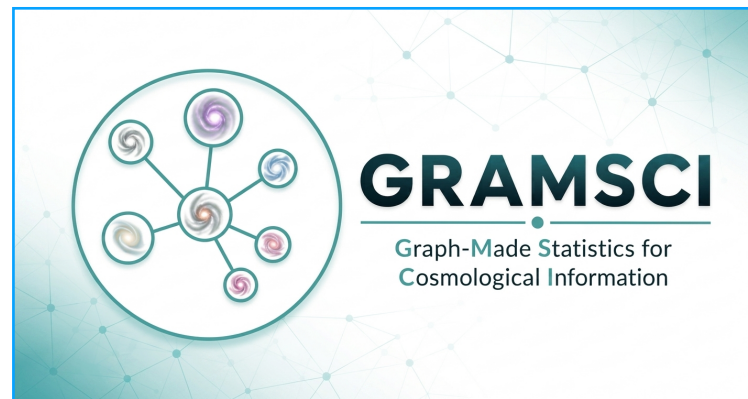
The GPU port — what made this campaign feasible



Three generations: v1 binary-search → v2 merge-walk → GPU (identical counts, CPU×64).

The 2019 GRAMSCI algorithm is my own; the v2 merge-walk and **GPU port** — turning months of CPU time into days — were developed by Claude's **Fable**.

- ▶ GPU speedup **2.6× (3PCF) to 9× (4PCF)** over a 64-thread node; one GPU $\approx$ 95–100 CPU cores.
- ▶ Cumulative v1 → GPU: **41×** (4PCF, $r_{\max}=70$: 1044 → 280 → 26 s).
- ▶ **Out-of-core** tiling: a 9×10^9 -edge (45 GB) BAO 3PCF on a 24 GB card, $\sim 20\%$ overhead.



<https://arxiv.org/abs/2607.06604>

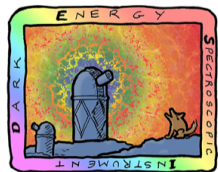
Sabiu 2026 ApJS, *under review*

Download: [GitHub.com/csabiu/Gramsci](https://github.com/csabiu/Gramsci)

The Dark Energy Spectroscopic Survey

Part 5 of the talk

DESI - tracer galaxy classes

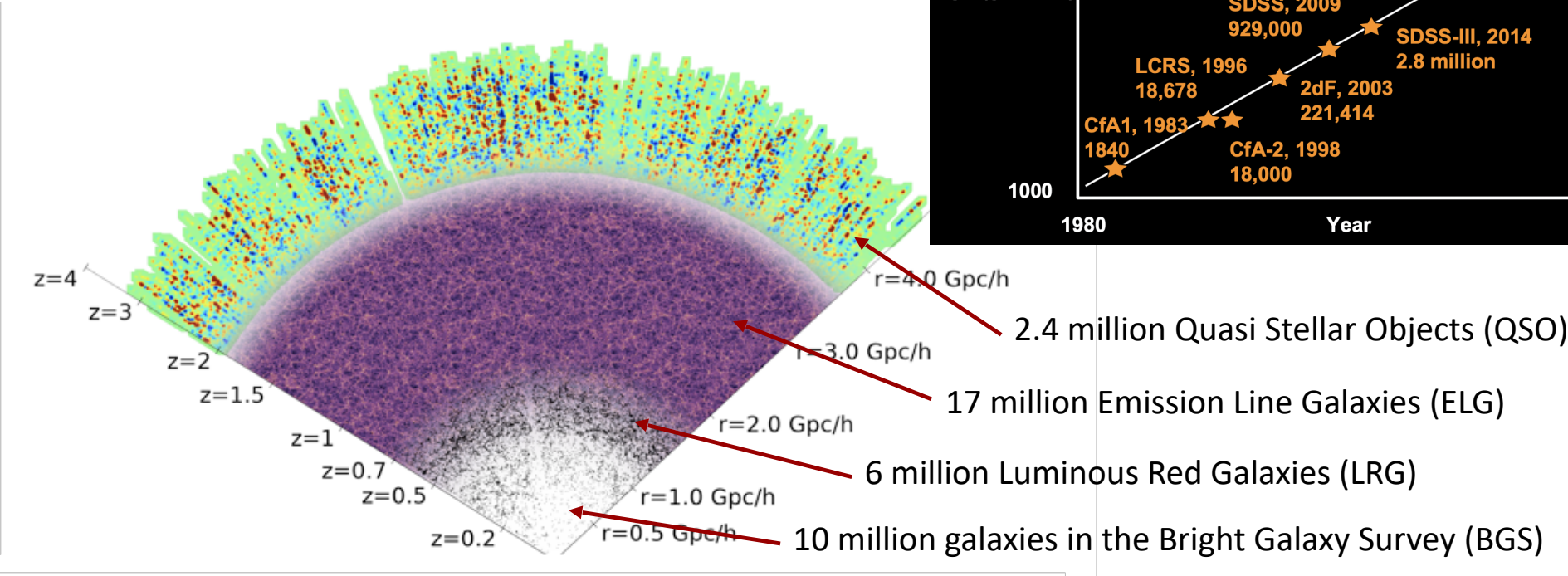
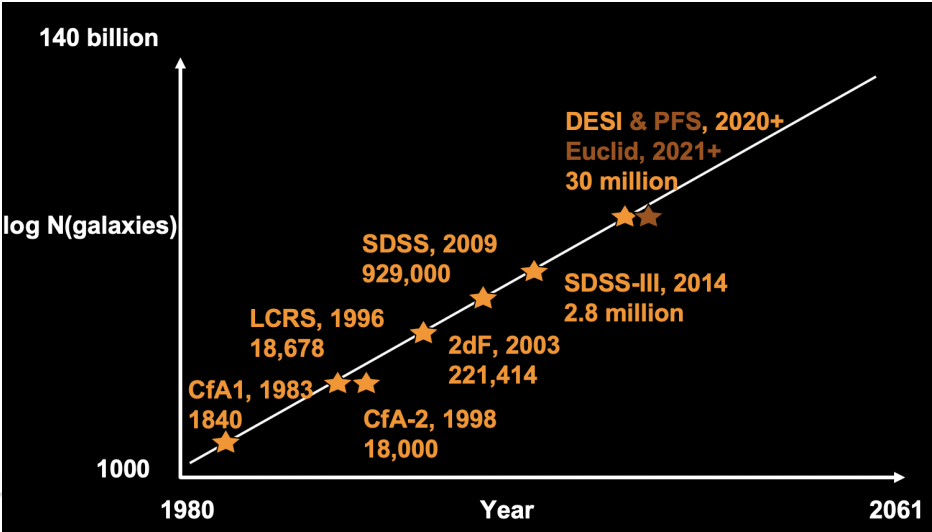


DARK ENERGY
SPECTROSCOPIC
INSTRUMENT

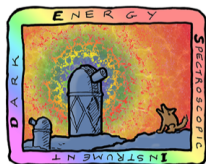
U.S. Department of Energy Office of Science

DESI will measure 35 million spectra

DESI will measure 10x more galaxies than previous surveys



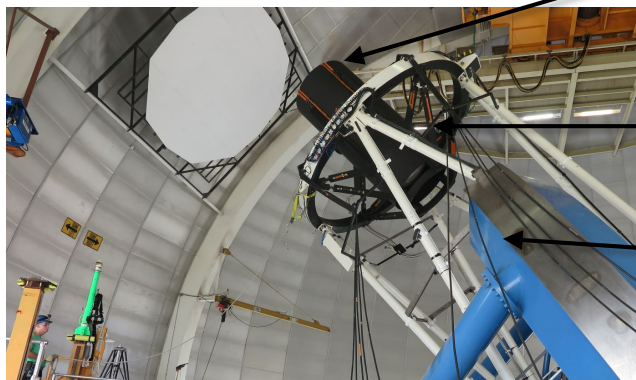
DESI - Telescope



DARK ENERGY
SPECTROSCOPIC
INSTRUMENT

U.S. Department of Energy Office of Science

DESI installed on the Mayall Telescope



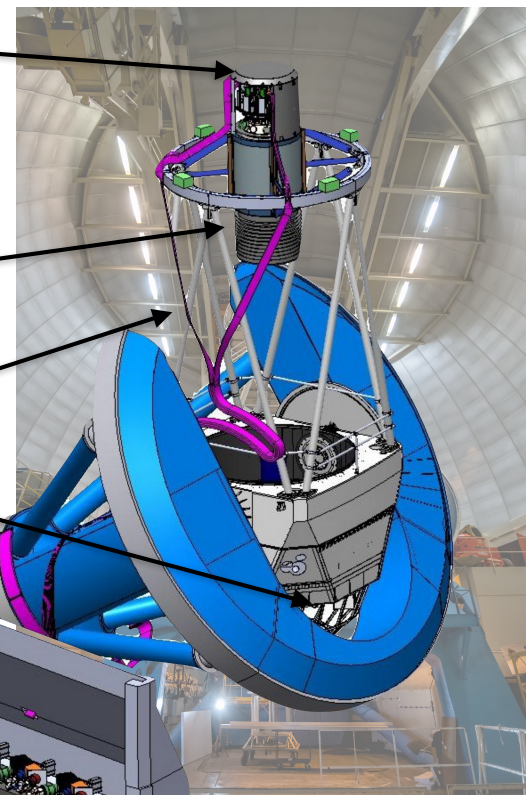
Focal plane assembly with
5000 fiber positioners

New 6-lens 8 deg²,
wide field corrector
in hexapod

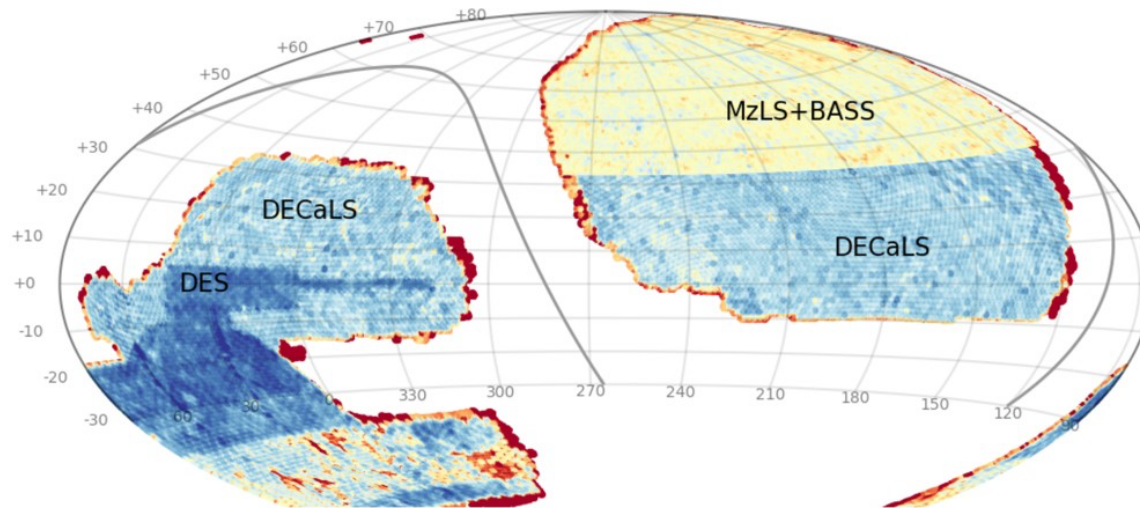
47.5m, 10-cable fiber
run

Fiber View Camera

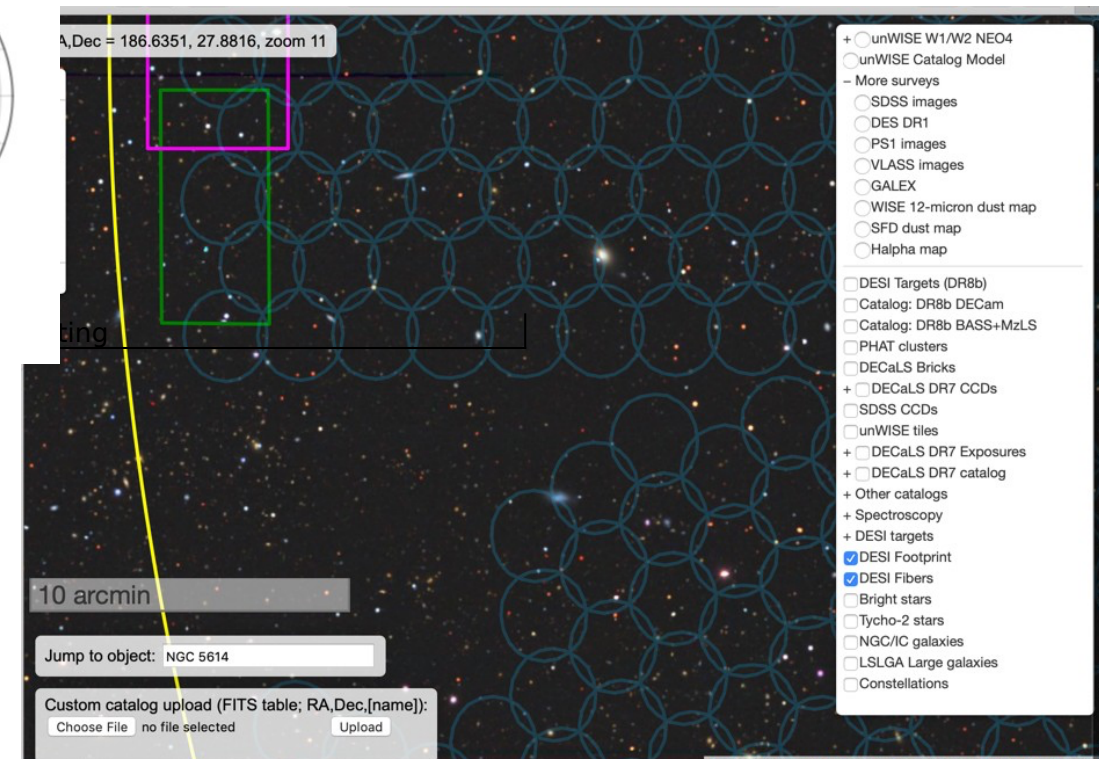
Ten, thermally-
controlled 3-channel
spectrographs
360-980nm



DESI - Pre-Imaging

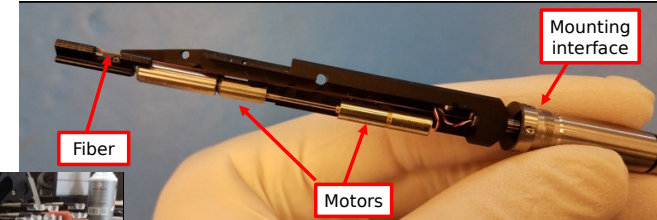
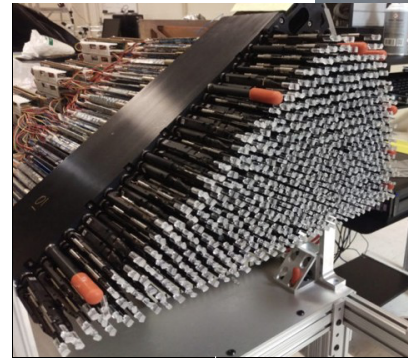
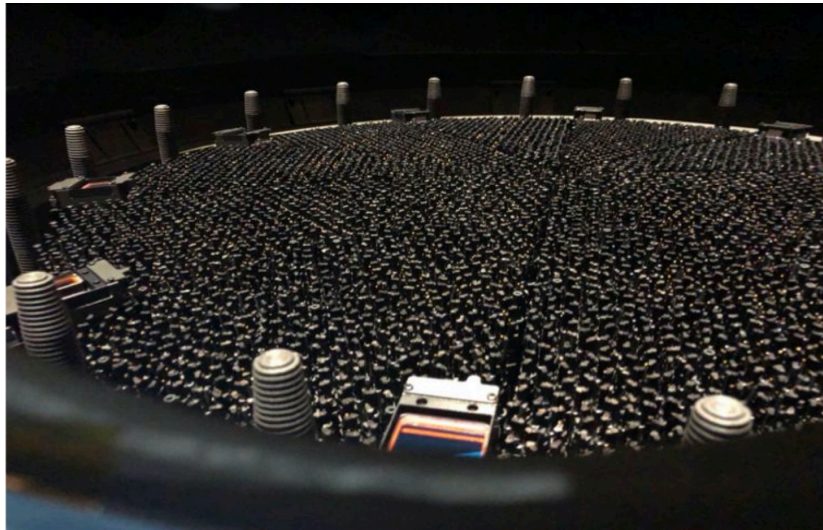


- ▶ **Pre-Imaging:** Before we can observe spectra we have to know where the targets (galaxies) are on the sky
- ▶ DESI imaging data is fully public
- ▶ DESI survey uses Data Release 9
- ▶ Data can be easily viewed: legacysurvey.org



DESI - pre-imaging

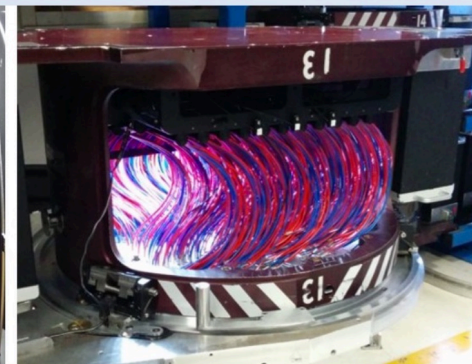
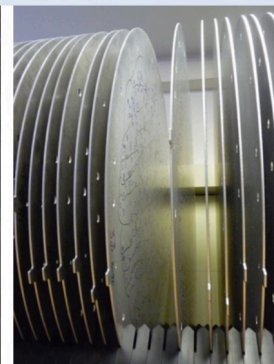
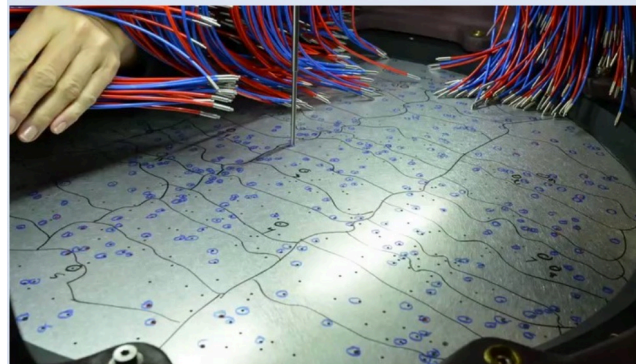
10 'petals' fill the focal plane



Each positioner has a motor to move within a limited patrol radius

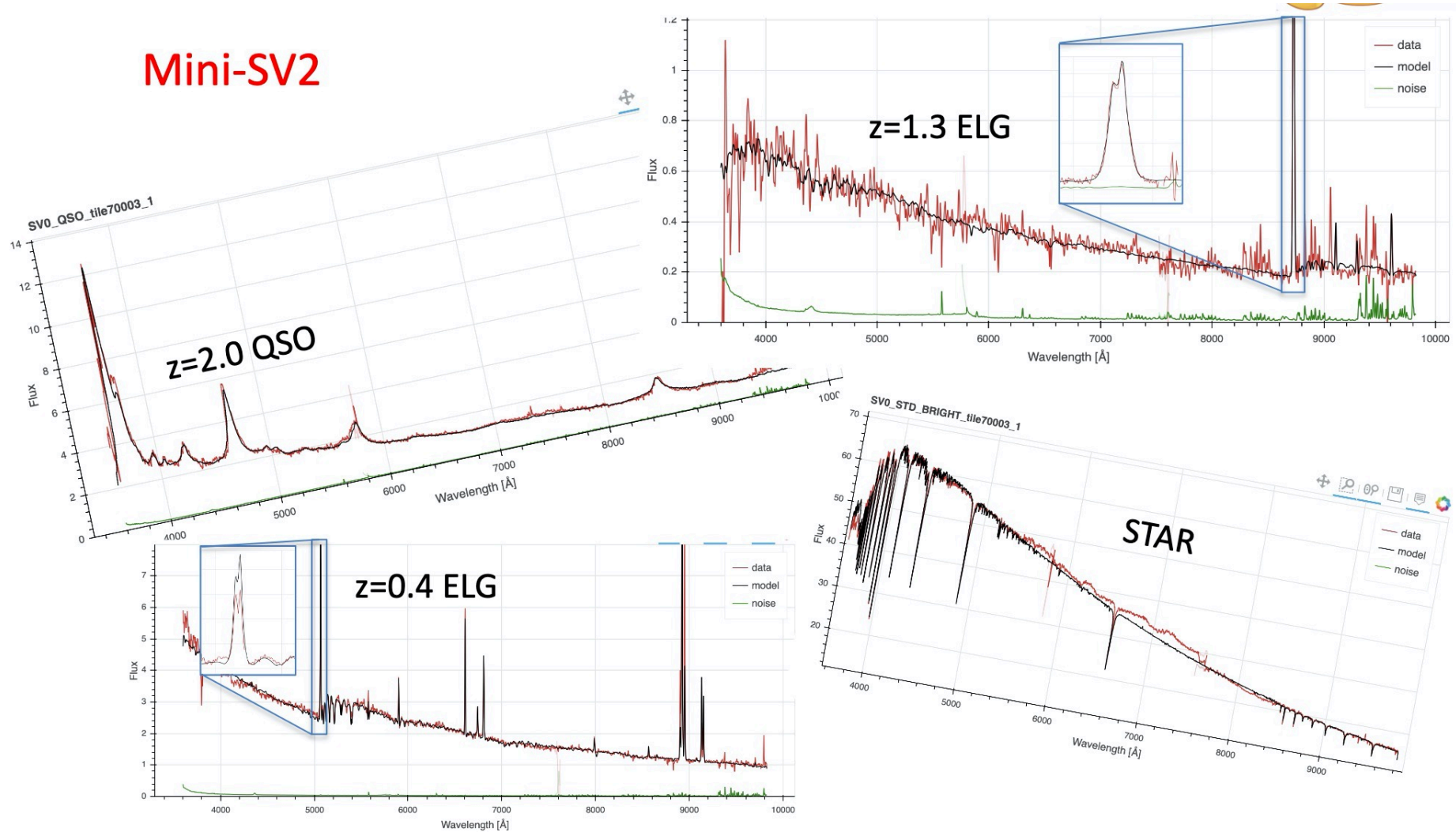
500 positioners per petal

No more of this!

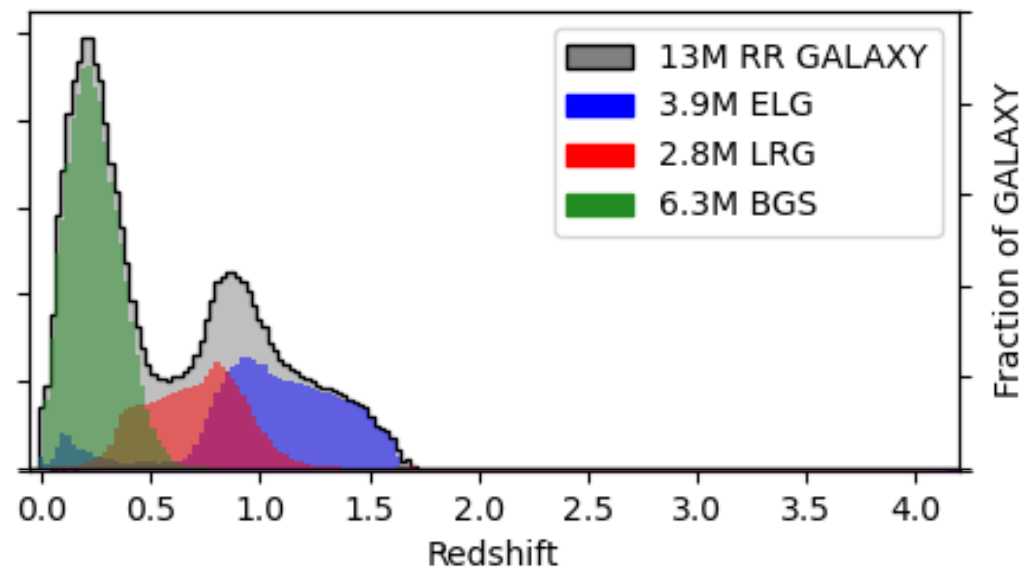
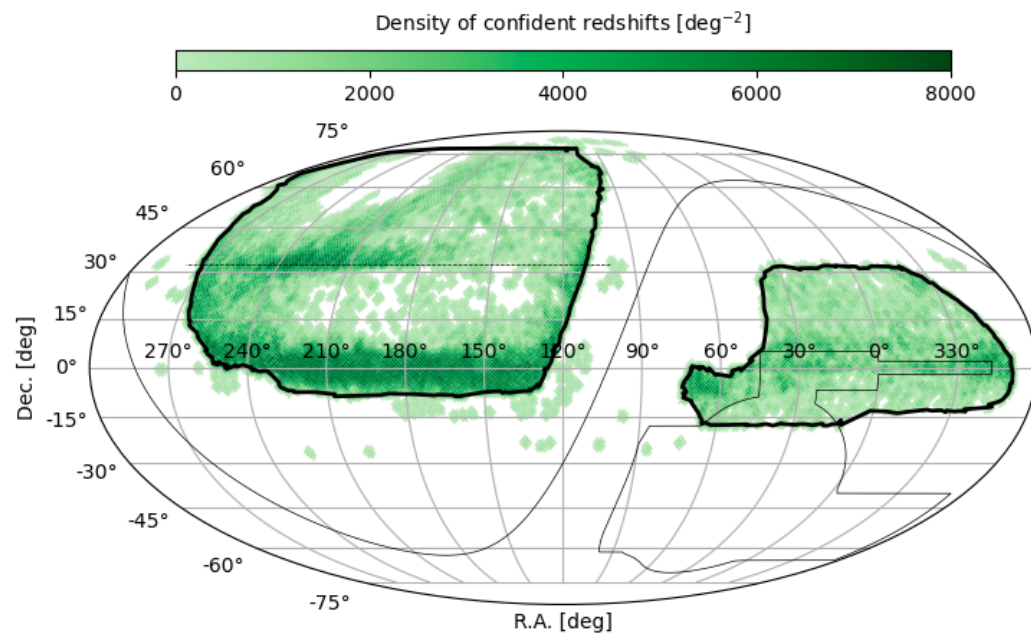


DESI - redshifts

Mini-SV2



DESI - redshifts

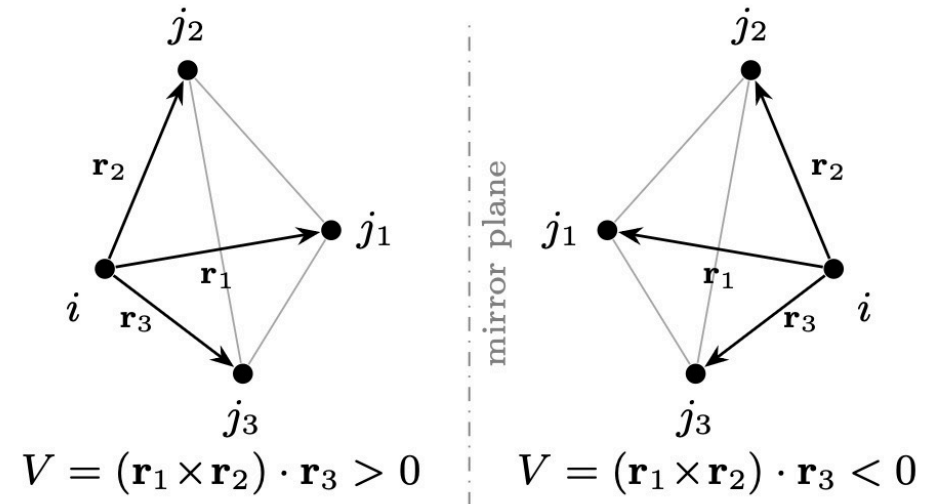


Parity Violation

Part 8 of the talk

Mirror symmetry in the Universe

- ▶ The 4-point function is the lowest order statistic that can directly differentiate between two shapes where their side lengths are the same.
- ▶ This makes the 4pcf a direct **parity violation** test for the large scale distribution of galaxies.
- ▶ The late Universe, large-scale physics is set by gravity and hydrodynamics, which conserve parity, but interesting violations could occur during inflation...
 - ▶ **Ghost Condensate**: ghost inflation, higher-derivative operators can generate a parity-odd scalar trispectrum
 - ▶ **Cosmological Collider**: chemical potential (parity-violating coupling), the exchange of a massive spin-1 particle generates a parity-odd contribution to the scalar trispectrum
- ▶ Other outside bets could include **Chern-Simons modified gravity**



Mirror symmetry

These two sets of 4 points have the same side lengths. But one cannot be rotated into the other.

Dichotomy

Inflation or systematic, highly unlikely to be known physics that's difficult to model ie astrophysics or non-linear gravitational evolution

The BOSS roller-coaster (2022–2024)

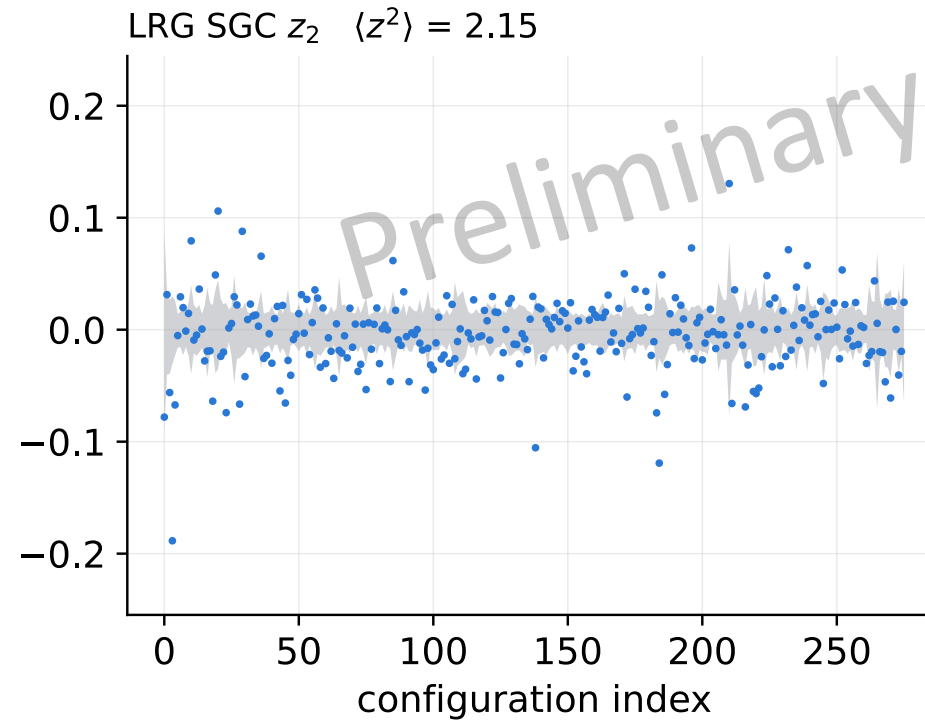
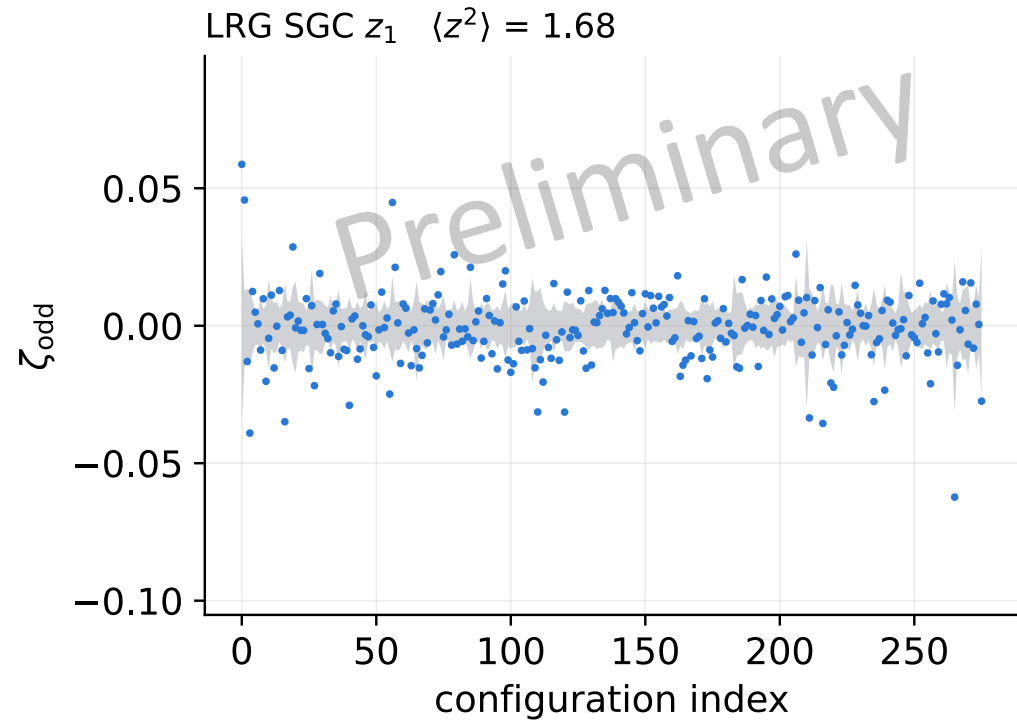
- ▶ **2022**: parity-odd 4PCF of BOSS CMASS+LOWZ: 2.9σ – 7.1σ “detections”. Philcox 2022, PRD 106, 063501; Hou, Slepian & Cahn 2023, MNRAS 522, 5701
- ▶ **2024**: the significance is controlled by how well mocks reproduce the parity-even 8-point function that sets the odd-channel covariance. Separating signal from covariance-bias: 0 – 2.5σ . “No evidence for parity violation in BOSS.” Krolewski, May, Smith & Hopkins 2024, JCAP 08, 044
- ▶ Independent composite-field tests: null. Gas et. al., arXiv:2509.13207, arXiv:2604.06021
- ▶ Lesson: the parity-odd channel has no cosmic-variance-free null — the mock covariance IS the measurement.

DESI DR1: the conflict this work arbitrates

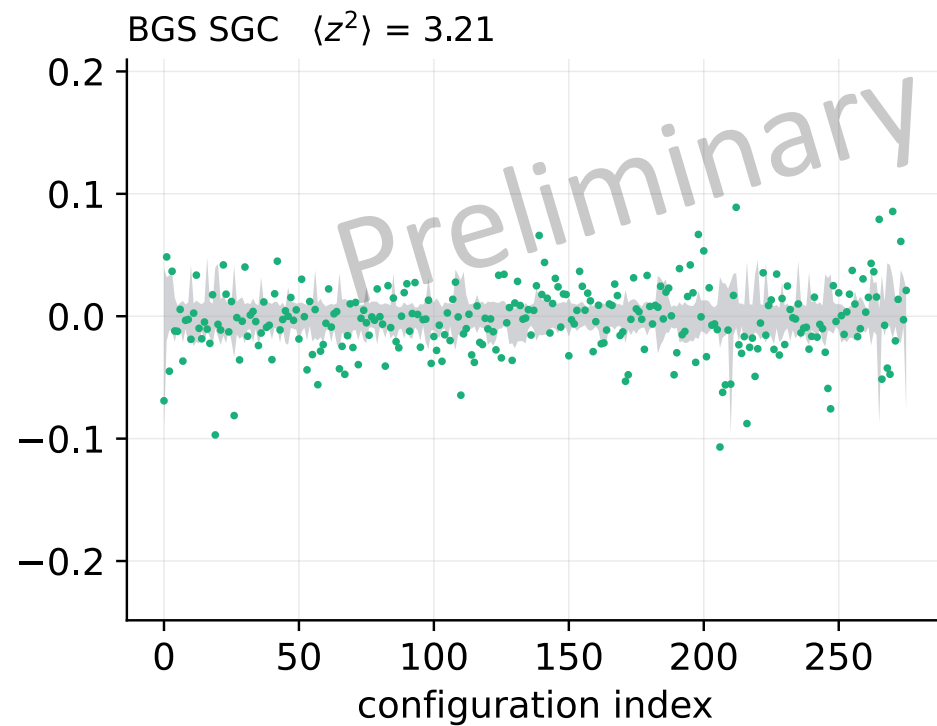
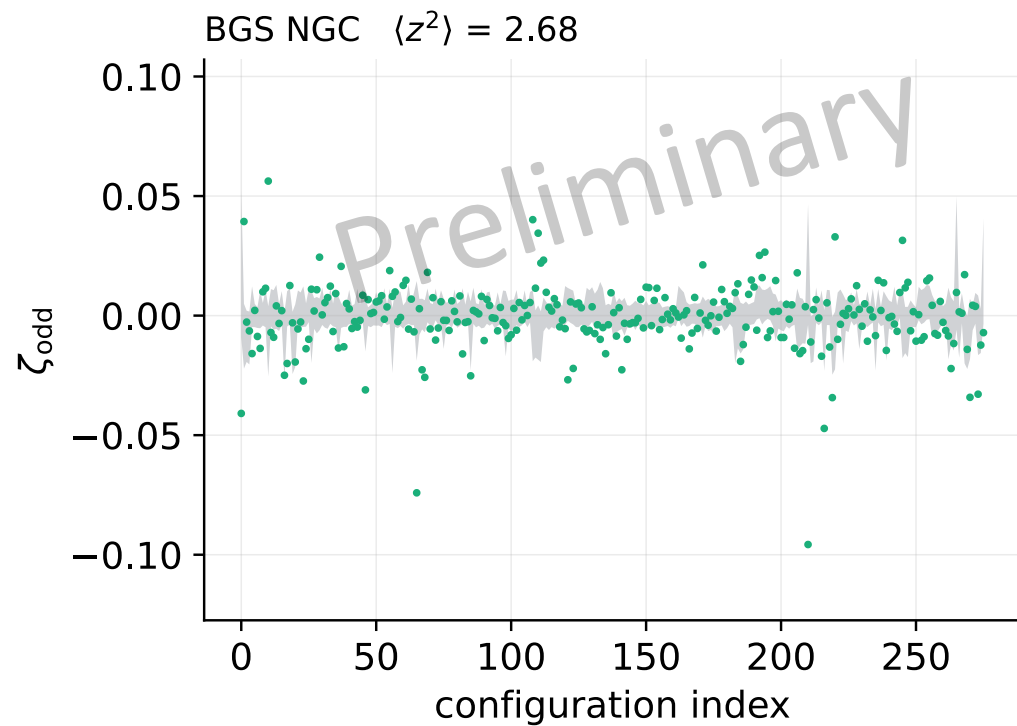
- ▶ Slepian et al. 2025 arXiv:2508.09133
 - ▶ harmonic-basis 4PCF, DR1 LRGs
 - ▶ 4–10 σ excess over parity-conserving mocks (auto-correlation)
 - ▶ . . . but no confirmation in cross-correlations between independent sky patches
- ▶ Hou et al. 2026 PRD 113, 103502
 - ▶ same data, same estimator family
 - ▶ uncorrected: up to 4.3 σ ; after fiber-assignment and covariance-mismatch corrections: –1.2 σ
 - ▶ data–mock covariance mismatch ~16% (odd) vs ~10% (even)

Both use the same harmonic-basis estimator lineage. An **independent estimator** with an **independent systematics budget** can arbitrate.

A preliminary search in DESI DR1

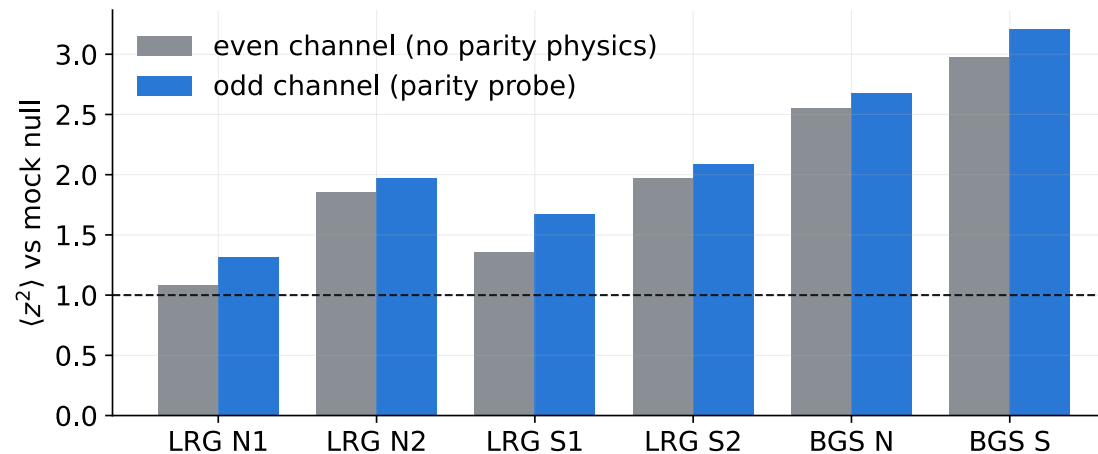


A preliminary search in DESI DR1



Parity odd - what is the excess?

- ▶ Data weights? Imaging-systematics or z-failure weights
- ▶ Fiber assignment? DESI has a complicated fiber assignment algorithm
- ▶ Approximate mocks? EZmock vs Abacus N-body covariance
- ▶ Completeness? BGS (high completeness, fiber assignment unlikely driver) shows the largest excess — kills the naive expectation.



- ▶ The even channel carries no parity physics — yet it is inflated by nearly the same factor everywhere. Whatever inflates the odd channel is probably not parity.

Parity - Next steps

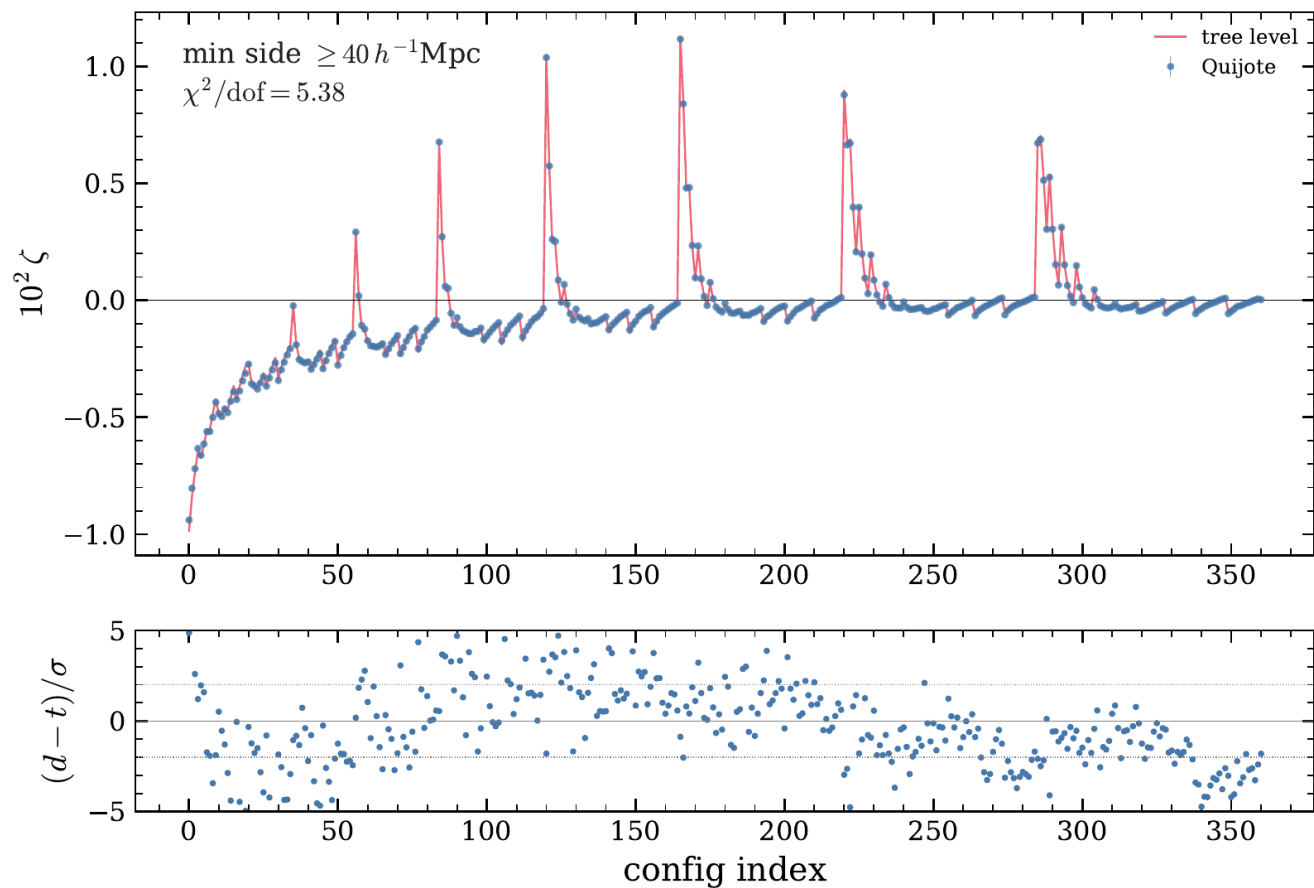
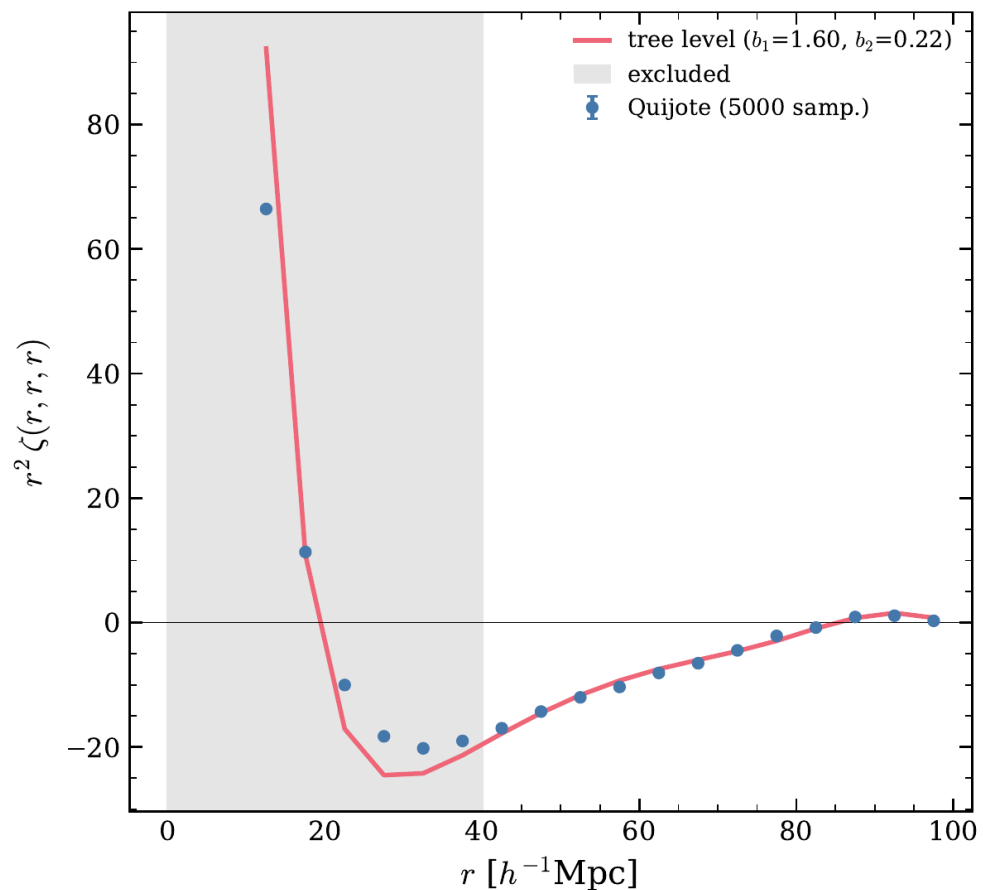
- ▶ I'm now computing the 4pcf of the larger **ELG sample**
- ▶ I will test fibre assignment more thoroughly
- ▶ Testing with real N -body rather than approx sims
- ▶ Use a template based search - need some help on the theory side!

Conclusions

- ▶ Parity violation search is difficult, will require more
 - ▶ realistic mocks
 - ▶ better systematic control
 - ▶ theoretical templates — need help!
- ▶ DR2 will be released soon. Increased volume!

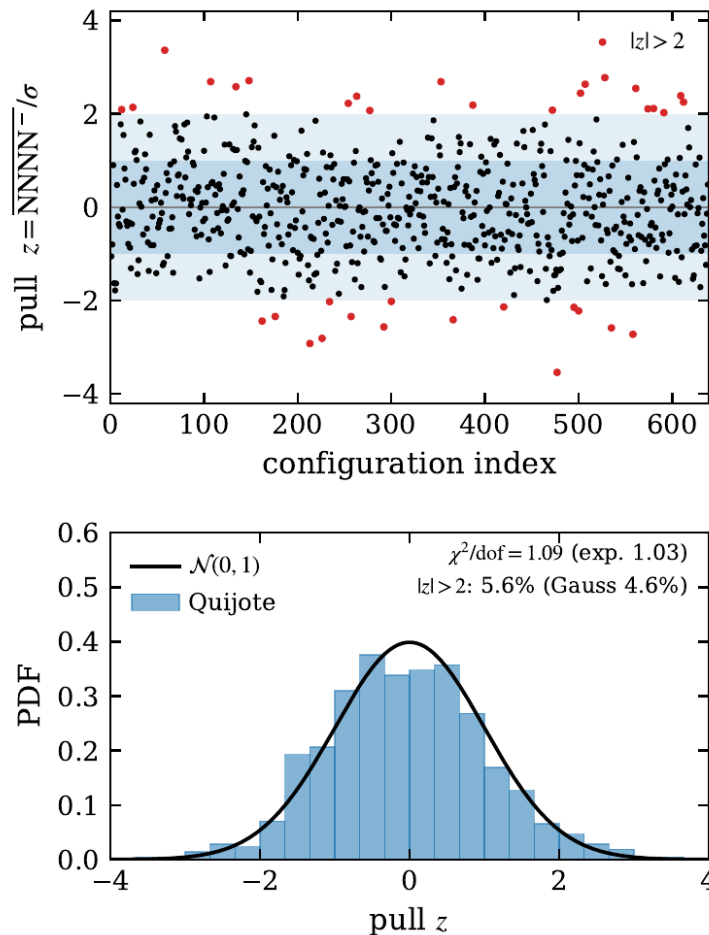
Backup slides

Backup: 3PCF vs tree-level, redshift space (Part I)



Backup: GRAMSCI parity validation

The parity-odd 4PCF — a clean probe of fundamental symmetries, for free



► Tetrahedron handedness from the signed volume V ; degenerate ($|V| < 10^{-9}$) tetrahedra get **zero** odd-channel weight (deterministic — no compiler/hardware noise).

► **Injection test**: recovers an injected chiral asymmetry to 0.3%.

► **Null test** (shown): gravity is parity-symmetric $\Rightarrow$ odd-4PCF must vanish. On Quijote, $\chi^2/\text{dof} = 1.10$ (Student-t expectation 1.03).

The same signed-weight pipeline yields the **parity-odd** 4PCF alongside the parity-even ζ^4 conn of Parts I & II.