Cosmological Structure Formation

Spatial statistics

STATISTICS OF RANDOM FIELDS

- Section 3.2-3.4 (p.32-38) in PT review (Bernardeau et al. 2002)
- Section 2.1 in Halo Model review (Cooray-Sheth 2002)

But first … some background

Continuous probability distributions

- $P(<x) = \int^x dx\ p(x)$
- m^{th} moment: $\langle x^m \rangle = \int dx\ p(x)\ x^m$
- Fourier transform: $F(t) = \int dx\ p(x) \exp(-itx)$
  - sometimes called Characteristic function
  - $d^mF/dt^m \sim i^m \langle x^m \rangle$, so F(t) is equivalent to knowledge of all moments
- If x>0, Laplace transform more useful:
- $L(t) = \int dx\ p(x) \exp(-tx)$

Distribution of sum of n independent random variates

- $p_2(s) = \int dx\, p(x) \int dy\, p(y)\ \delta_D(x+y = s)$

  $= \int dx\, p(x)\, p(s-x)$
- $F_2(t) = \int ds\, \exp(-its) \int dx\, p(x)\, p(s-x)$

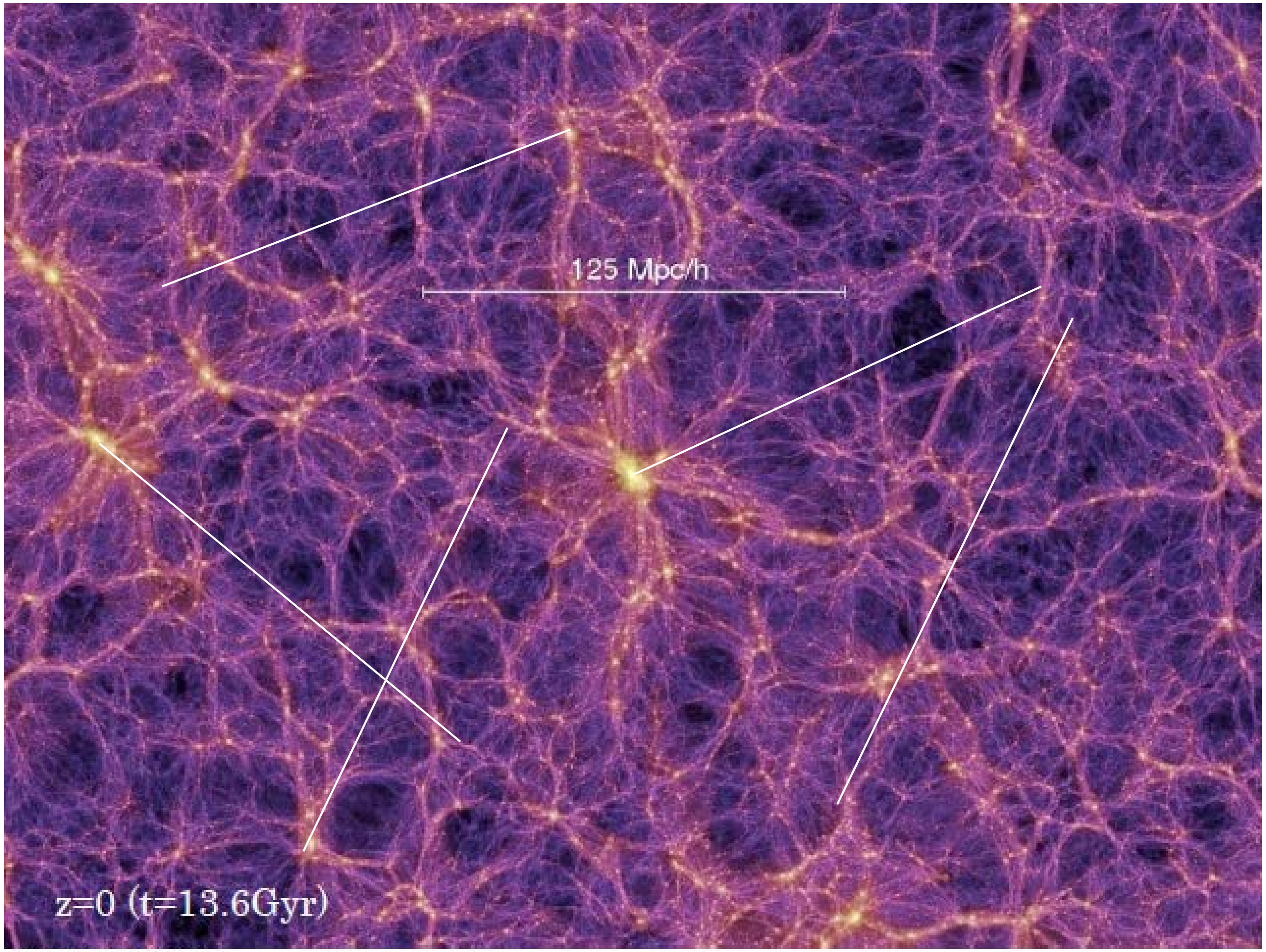
  $= \int ds \int dx\, p(x) \exp(-itx)\, p(s-x) \exp[-it(s-x)]$

  $= F_1(t)\, F_1(t)$
- $F_n(t) = [F_1(t)]^n$

= Convolve PDFs = Multiply CFs

Fourier transform exp(ikx) useful

- Convolutions become products
  - Smoothing on scale R: $\delta_R(\mathbf{x}) \rightarrow \delta(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}}\, W(kR)$
- Each derivative brings down a power of ik
  - Can transform differential equations into algebraic equations
- Integral brings 1/ik
  - divergence at k=0 ~ constant of integration



Tuesday, July 17, 2012

Quantify clustering by number of pairs compared to random (unclustered) distribution, triples compared to triangles (of same shape) in unclustered distribution, etc.

2pt spatial statistics

- $dP = \langle n_1 \rangle \, dV_1 \, \langle n_2 \rangle \, dV_2 \, [1 + \xi(\mathbf{r}_1, \mathbf{r}_2)]$
  $= \langle n \rangle^2 \, dV_1 \, dV_2 \, [1 + \xi(\mathbf{r}_1 - \mathbf{r}_2)]$ homogeneity
  $= \langle n \rangle^2 \, dV_1 \, dV_2 \, [1 + \xi(|\mathbf{r}_1 - \mathbf{r}_2|)]$ isotropy

Define: $\delta(\mathbf{r}) = [n(\mathbf{r}) - \langle n \rangle]/\langle n \rangle$

Then: $\xi(r) = \langle \delta(\mathbf{x}) \, \delta(\mathbf{x} + \mathbf{r}) \rangle$ ξ is the correlation function

Estimator: $\langle (D_1 - R_1)/R_1 \, (D_2 - R_2)/R_2 \rangle \sim (DD - 2DR + RR)/RR$

translational invariance isotropy

And FT is: $\langle \delta(\mathbf{k}_1) \, \delta(\mathbf{k}_2) \rangle = (2\pi)^3 \, \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \, P(|\mathbf{k}_1|)$

P(k) is the power spectrum

Vol. 21, No. 3 Publications of the Astronomical Society of Japan 1969

The Correlation Function for the Distribution of Galaxies

Hiroo TOTSUJI and Taro KIHARA

Department of Physics, Faculty of Science, University of Tokyo

(Received May 15, 1969; revised June 26, 1969)

Abstract

The correlation function for the spatial distribution of galaxies in the universe is determined to be $(r_0/r)^{1.8}$, r being the distance between galaxies. The characteristic length r_0 is 4.7 Mpc. This determination is based on the distribution of galaxies brighter than the apparent magnitude 19 counted by SHANE and WIRTANEN (1967). The reason why the correlation function has the form of inverse power of r is that the universe is in a state of "neutral" stability.

$$\frac{\text{Number of data pairs with separation } r}{\text{Number of random pairs with separation } r} \quad \frac{DD(r)}{RR(r)} = 1 + \xi(r)$$

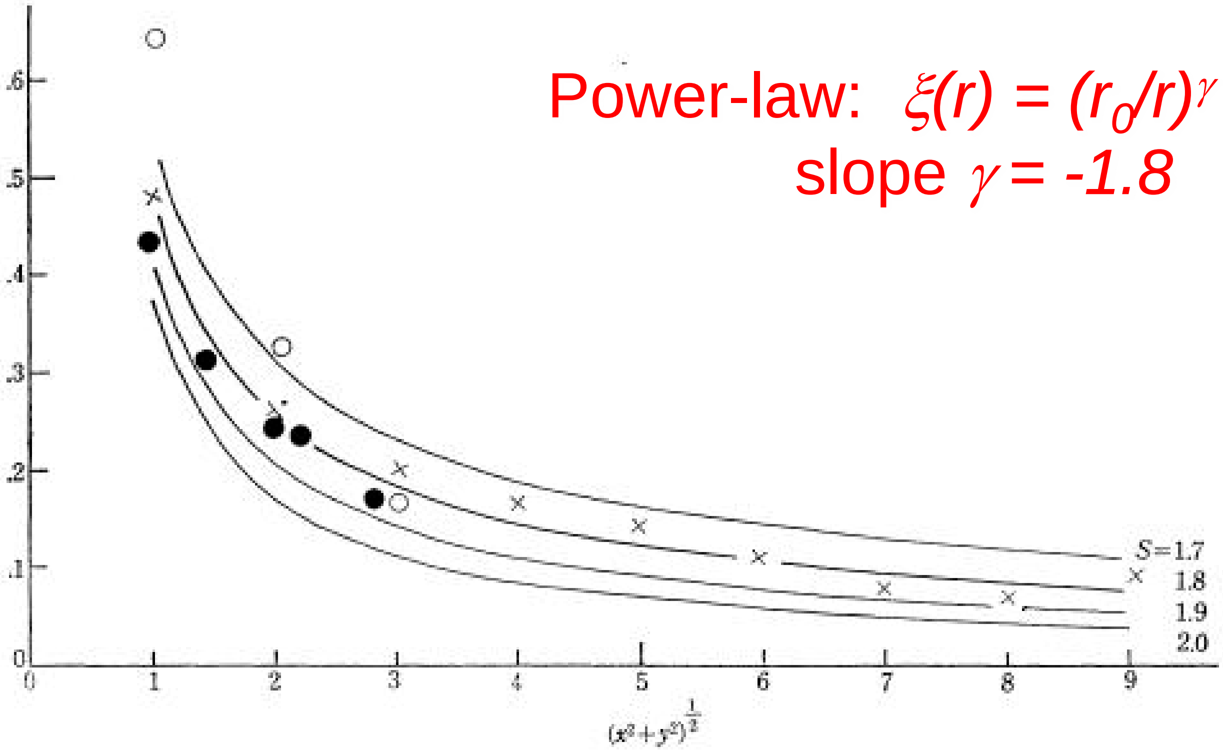
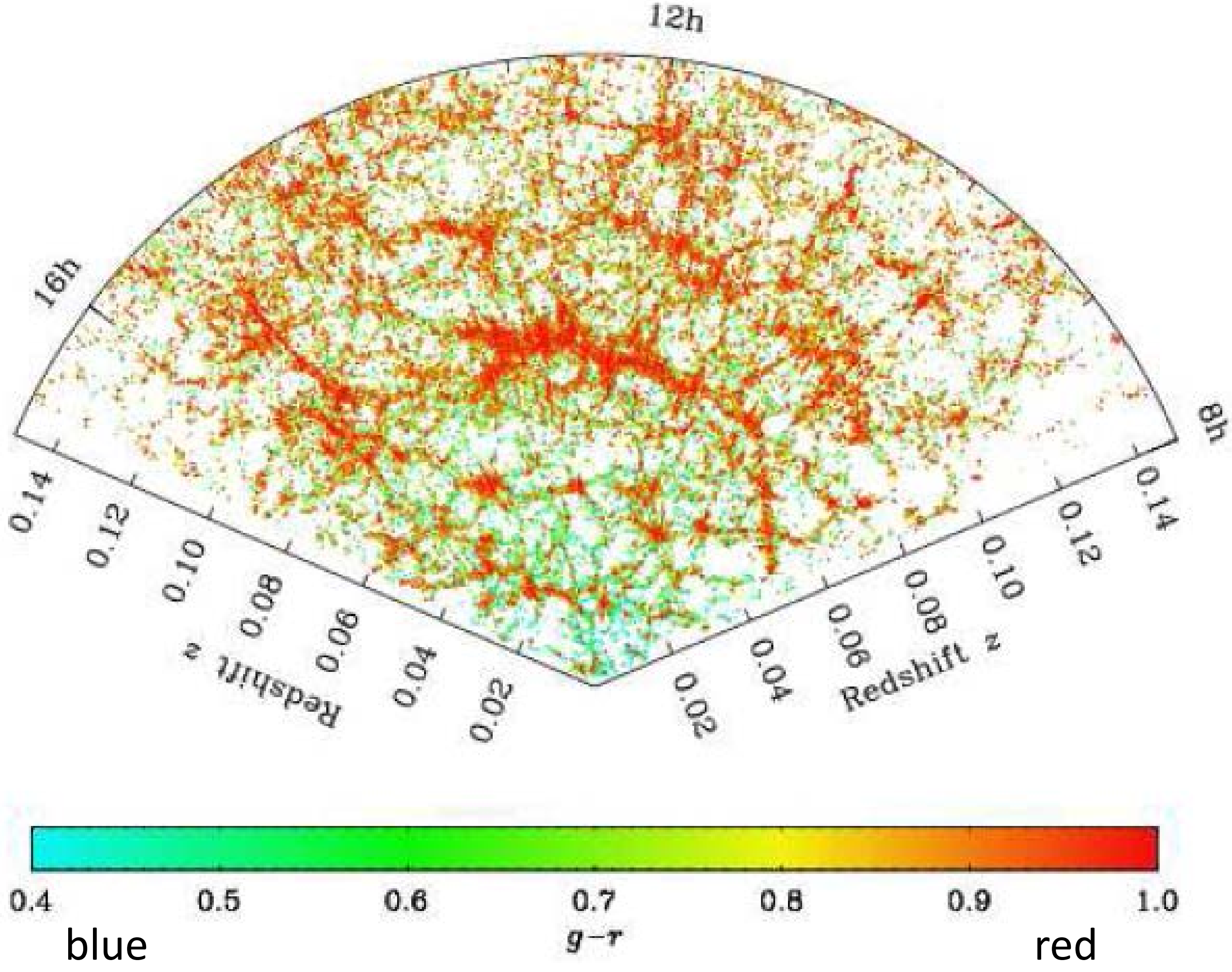


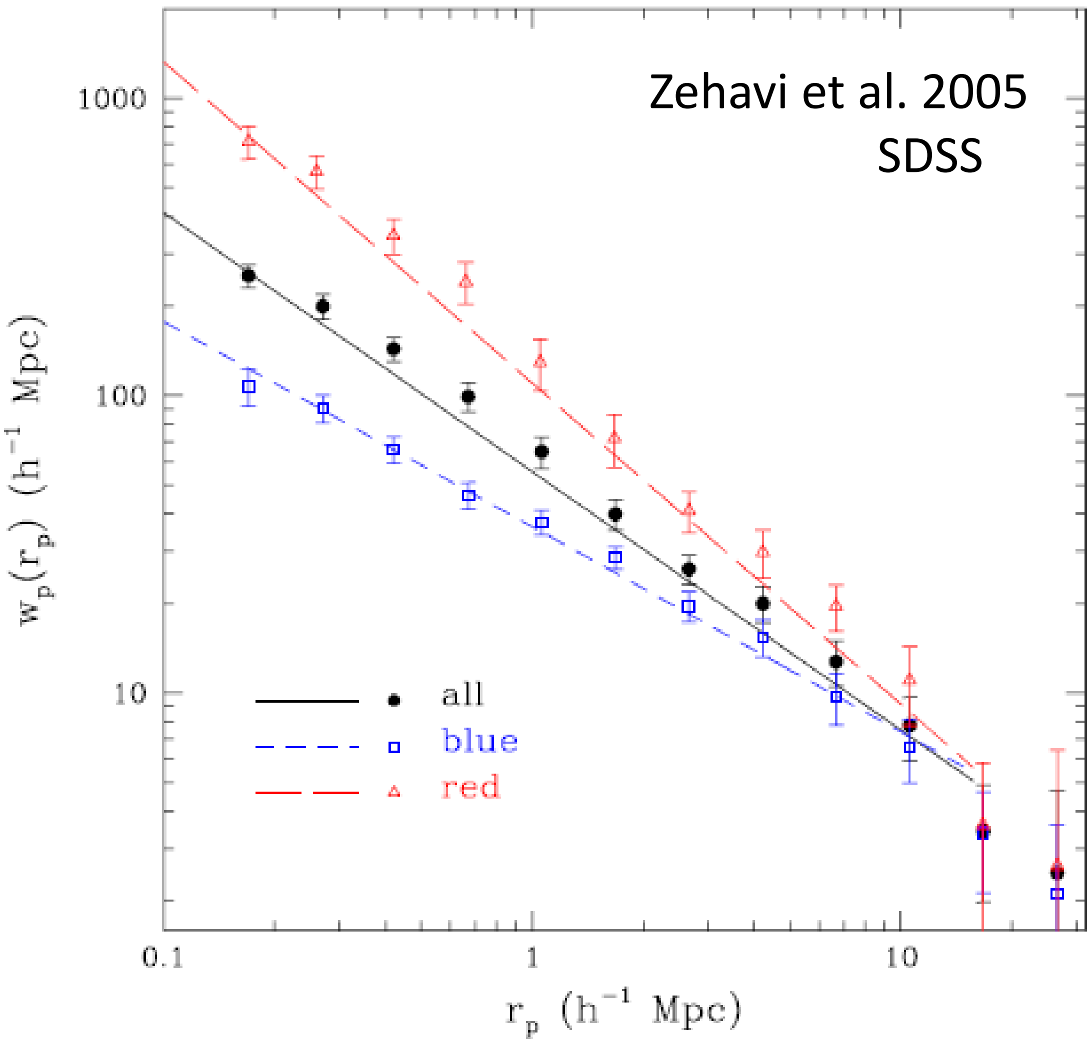
FIG. 2. Comparison of the empirical and theoretical values of $\frac{\langle\{N_1-\langle N\rangle\}\{N_2-\langle N\rangle\}\rangle}{\langle\{N-\langle N\rangle\}^2\rangle-\langle N\rangle}$.

The filled circles indicate the empirical values obtained by the authors, and the open circles and crosses by NEYMAN et al.; the unit solid angle is $1°\times1°$ for the circles and $10'\times10'$ for the crosses. The curves are theoretical values for $s=1.7$, 1.8, 1.9, and 2.0.


12h
16h
8h
0.14
0.12
0.10
0.08
0.06
0.04
0.02
Redshift z
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Redshift z
0.4
0.5
0.6
0.7
0.8
0.9
1.0
$g-r$
blue
red

Galaxy clustering depends on galaxy type: luminosity, color, etc.

(Final lectures use Halo Model to describe this.)



2pt spatial statistics

- $dP = \langle n_1 \rangle \, dV_1 \, \langle n_2 \rangle \, dV_2 \, [1 + \xi(\mathbf{r}_1, \mathbf{r}_2)]$
  $= \langle n \rangle^2 \, dV_1 \, dV_2 \, [1 + \xi(\mathbf{r}_1 - \mathbf{r}_2)]$ homogeneity
  $= \langle n \rangle^2 \, dV_1 \, dV_2 \, [1 + \xi(|\mathbf{r}_1 - \mathbf{r}_2|)]$ isotropy

Define: $\delta(\mathbf{r}) = [n(\mathbf{r}) - \langle n \rangle]/\langle n \rangle$

Then: $\xi(r) = \langle \delta(\mathbf{x}) \, \delta(\mathbf{x} + \mathbf{r}) \rangle$ ξ is the correlation function

Estimator: $1 + \xi(r)$ = data-pairs/random-pairs = DD(r)/RR(r)

$= \sum_{I,j}^{Ndata} 1 \text{ (if } r_{ij} = r) / \sum_{I,j}^{Nrandom} 1 \text{ (if } r_{ij} = r \text{ in same volume)}$

or $\langle (D_1 - R_1)/R_1 \, (D_2 - R_2)/R_2 \rangle \sim (DD - 2DR + RR)/RR$

And FT is: $\langle \delta(\mathbf{k}_1) \, \delta(\mathbf{k}_2) \rangle = (2\pi)^3 \, \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \, P(|\mathbf{k}_1|)$

P(k) is the power spectrum

(Better) Estimator

$\xi(r) = < \delta(\mathbf{x})\ \delta(\mathbf{x} + \mathbf{r}) >$

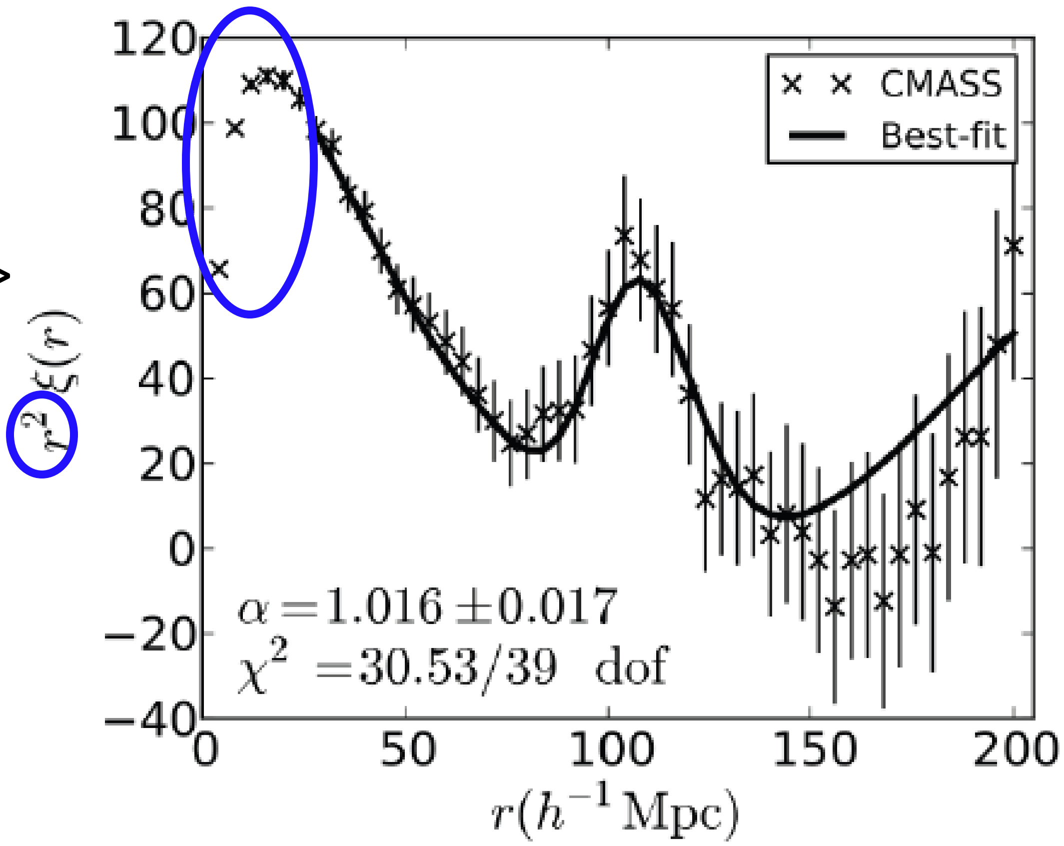
Since $\delta(\mathbf{r}) = [n(\mathbf{r}) - \langle n \rangle]/\langle n \rangle$

estimate using

$\xi = \langle (D_1-R_1)/R_1\ (D_2-R_2)/R_2 \rangle$

$\sim (DD-2DR+RR)/RR$

for pairs separated by r

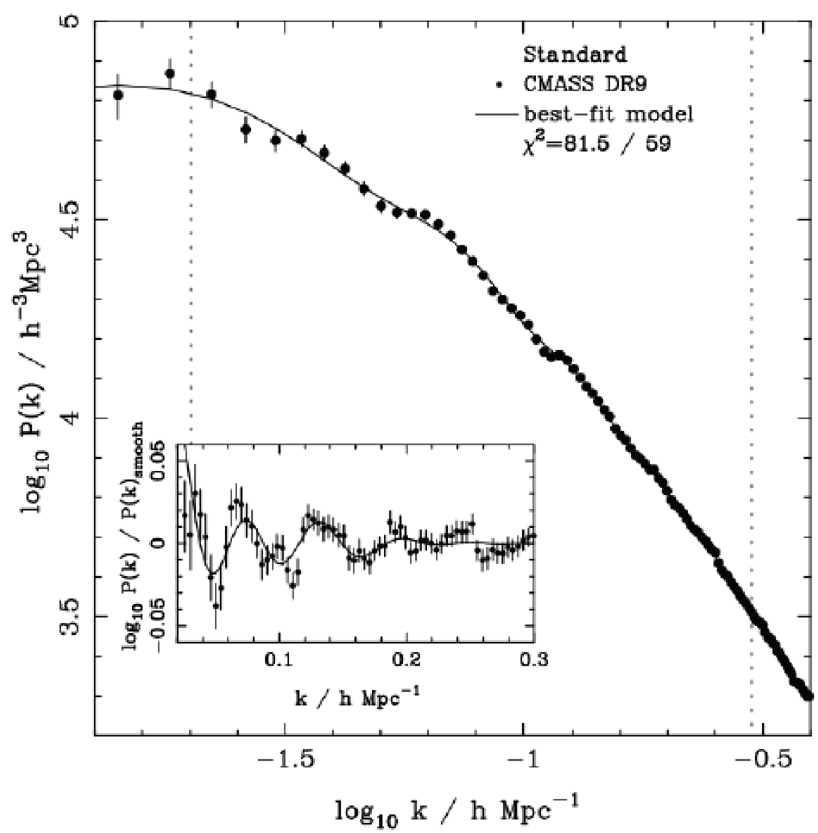


$$\begin{aligned}
\xi(r) &= \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \\
&= \lim_{V\to\infty} \frac{1}{V}\int_V \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{x}) \sum_{\mathbf{k}'} \delta^*_{\mathbf{k}'} \exp\left[-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})\right]\, d\mathbf{x} \\
&= \lim_{V\to\infty} \frac{1}{V}\int_V \sum_{\mathbf{k}} P(\mathbf{k}) \exp(-i\mathbf{k}\cdot\mathbf{r}) \\
&= \frac{1}{(2\pi)^3}\int P(k)\exp(-i\mathbf{k}\cdot\mathbf{r})d\mathbf{k} \\
P(k) &= \int \xi(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})d\mathbf{r}
\end{aligned}$$

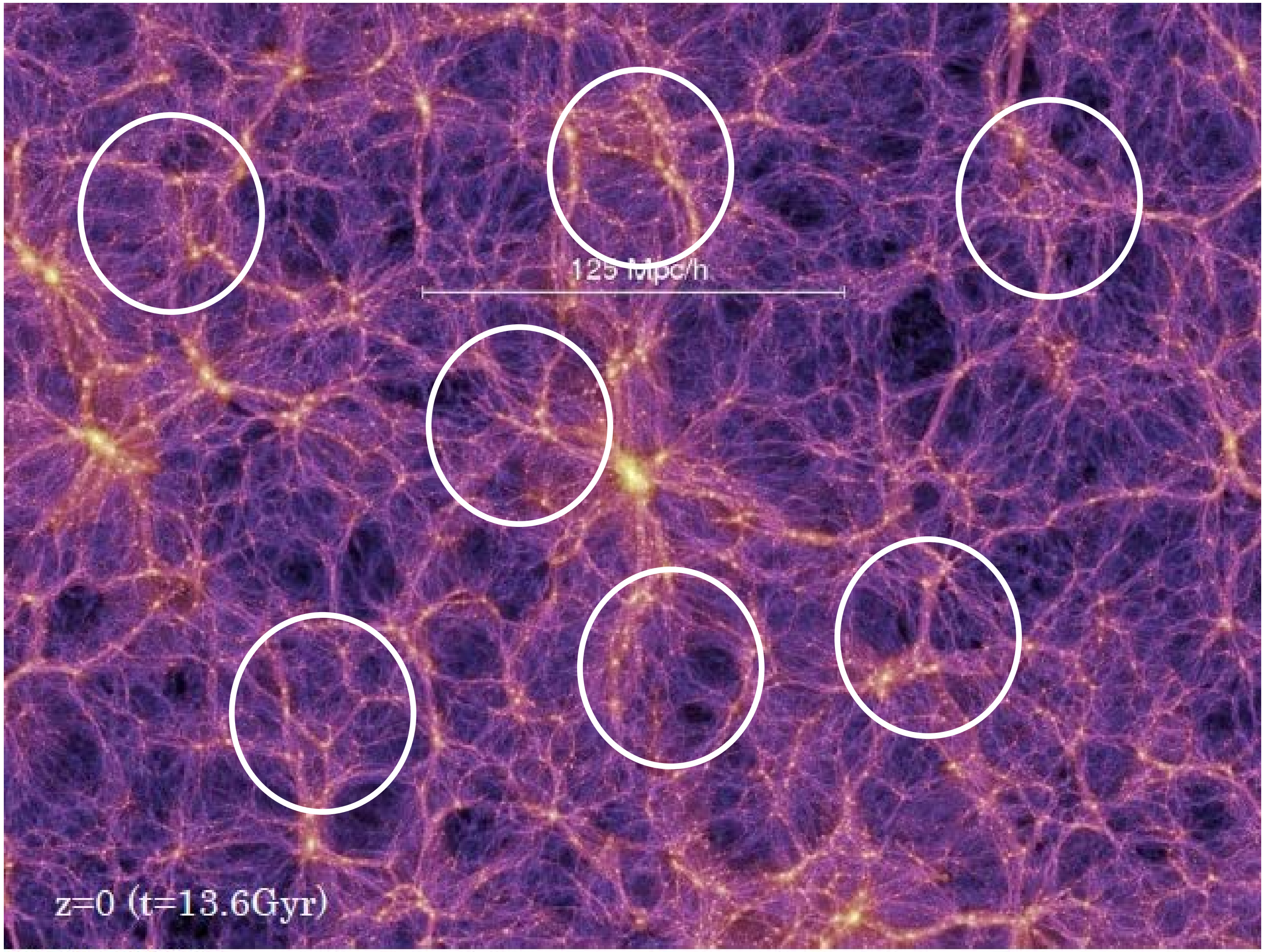
$$\int_\Omega \exp(-ikr\cos\theta)d\Omega = 4\pi\frac{\sin kr}{kr}$$

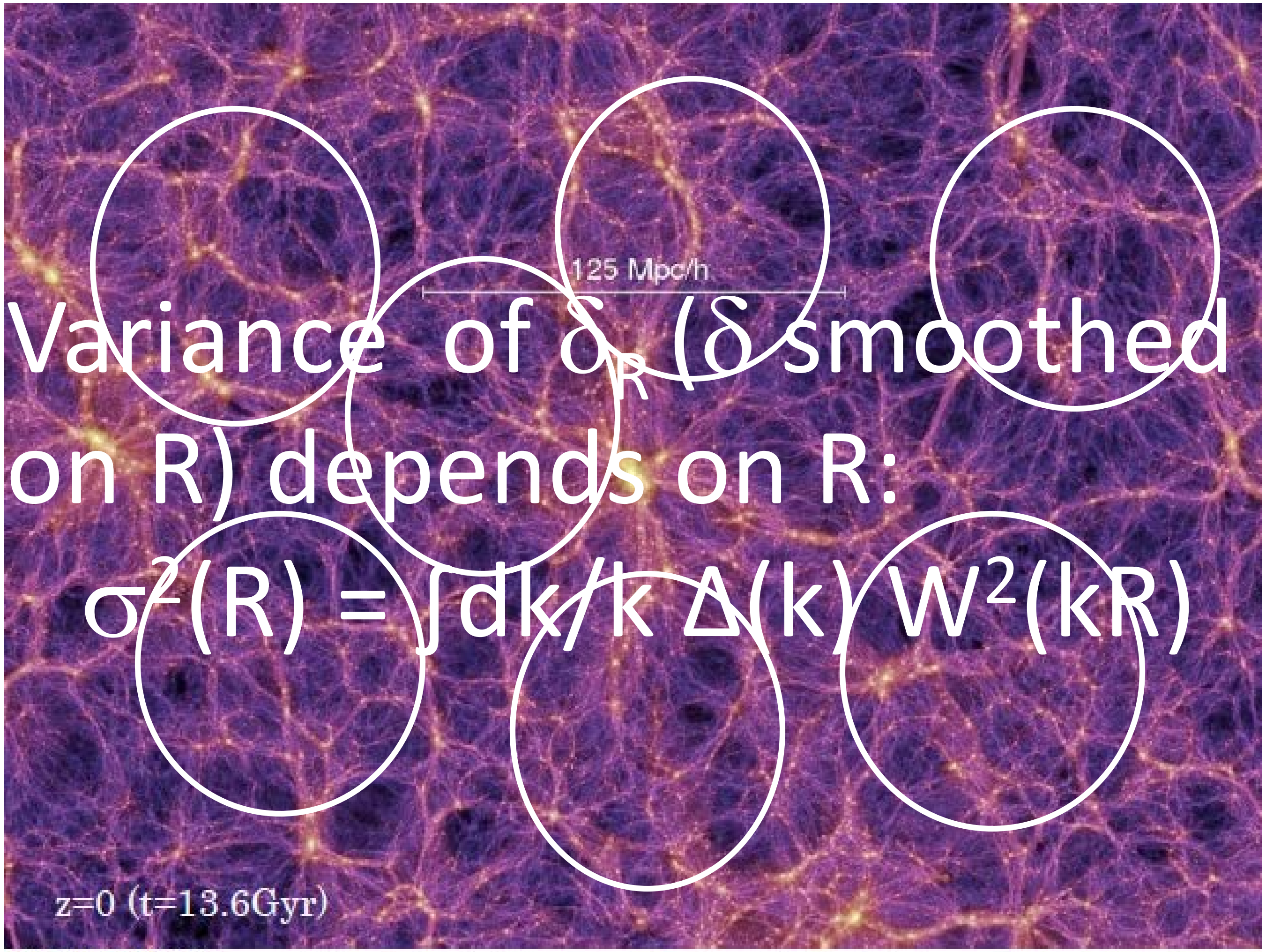
$$P(k) = \int_0^\infty \xi(r)\frac{\sin kr}{kr} r^2 dr$$

$$\xi(r) = \frac{1}{2\pi^2}\int_0^\infty P(k)\frac{\sin kr}{kr}k^2 dk$$



P(k) and ξ(r) are FT pairs


125 Mpc/h
z=0 (t=13.6Gyr)



Variance of δ_R (δ smoothed on R) depends on R:

$$\sigma^2(R) = \int dk/k\ \Delta(k)\ W^2(kR)$$

Variance of δ_R (δ smoothed on R):

$$\sigma^2(R) = \int dk/k\ \Delta(k)\ W^2(kR)$$

Correlations in smoothed field

$$\Delta(k) = k^3 P(k)/2\pi^2$$

$$\Delta_{R_1 R_2}(k) = \Delta(k)\ W(kR_1)\ W(kR_2)$$

$$\xi_{R_1 R_2}(r) = \int dk/k\ \Delta_{R_1 R_2}(k)\ j_0(kr)$$

E.g. Power-law P(k)

- $\xi(r) = \int dk/k\, [k^3 A k^n/2\pi^2]\, j_0(kr) \propto r^{-3-n}$ if $n>-3$
- $\sigma^2(R) = (A/2\pi^2) \int dk/k\, k^{n+3} \exp(-k^2R^2)$

  $= (A/2\pi^2)\, \Gamma[(n+3)/2]/2\, R^{-3-n}$
- $\xi_R(r) = (A/2\pi^2) \int dk\, k^{2+n} \exp(-k^2R^2)\, j_0(kr)$

  $= (A/2\pi^2)\, (\pi/2r)\, \mathrm{erf}(r/2R)$ if $n=-2$

  $\rightarrow \xi_0(r)$ when $r \gg R$

(smoothing irrelevant on large scales? BAO …)

Gaussian PDF

- $p(x) = \exp[-(x-\mu)^2/2\sigma^2]/\sigma\sqrt{2\pi}$
- $F(t) = \exp(it\mu)\ \exp(- t^2\ \sigma^2)$
- $F_n(t) = \exp(it\ n\mu)\ \exp(- t^2\ n\sigma^2)$
- Distribution of sum of n Gaussians is Gaussian with mean $n\mu$ and variance $n\sigma^2$
- In general, PDFs are not ‘scale invariant’

Gaussian field

- $p(\mathbf{x}) = \exp(- \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}/2)/ (2\pi)^{n/2} \sqrt{\text{Det}[\mathbf{C}]}$

  where $\mathbf{x} = (x_1, \ldots x_n)$ with $x_1 = x(r_1) - \langle x(r_1)\rangle$

  and $\mathbf{C}_{ij} = \langle x_i x_j \rangle$

- HW: Show that Fourier Transform/CF is

  $F(\mathbf{t}) = \langle \exp(i\mathbf{t}.\mathbf{x})\rangle = \int d\mathbf{x}\, p(\mathbf{x}) \exp(i\mathbf{t}.\mathbf{x}) = \exp(i\mathbf{m}^T\mathbf{t} - \mathbf{t}^T\mathbf{C}\mathbf{t}/2)$

  where $\mathbf{m} = (\langle x(r_1)\rangle, \ldots , \langle x(r_n)\rangle)$

- For Gaussian field **C** may be much simpler (e.g. approximately band diagonal) than $\mathbf{C}^{-1}$.

Gaussian Random Fields

Rayleigh (2d Gaussian) amplitude + uniform random phase

In a Gaussian field Fourier modes are uncorrelated, by this we mean

$$\langle \delta(\mathbf{k}_1) \ldots \delta(\mathbf{k}_{2p+1}) \rangle = 0$$
$$\langle \delta(\mathbf{k}_1) \ldots \delta(\mathbf{k}_{2p}) \rangle = \sum_{\text{all pair associations}} \prod_{p \text{ pairs (i,j)}} \langle \delta(\mathbf{k}_i)\delta(\mathbf{k}_j) \rangle$$

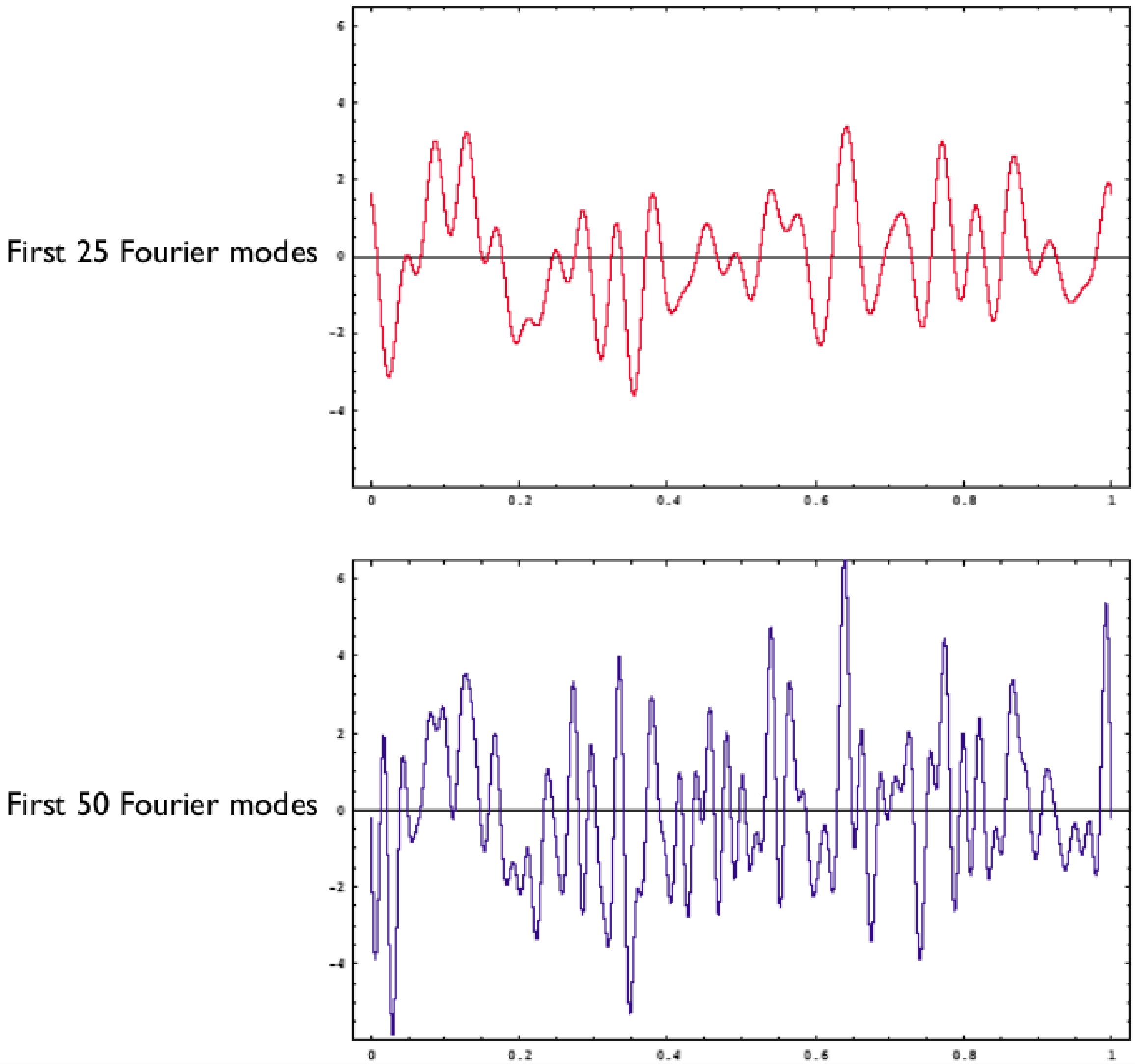
remember that $\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \delta_D(\mathbf{k}+\mathbf{k}')\, P(k)$

These properties are usually summarized by saying that connected moments of order larger than 2 are zero,

$$\langle \delta(\mathbf{k}_1) \ldots \delta(\mathbf{k}_N) \rangle_c = 0, \quad N > 2$$

Thus, to generate a Gaussian field, just draw 2 random numbers per mode...

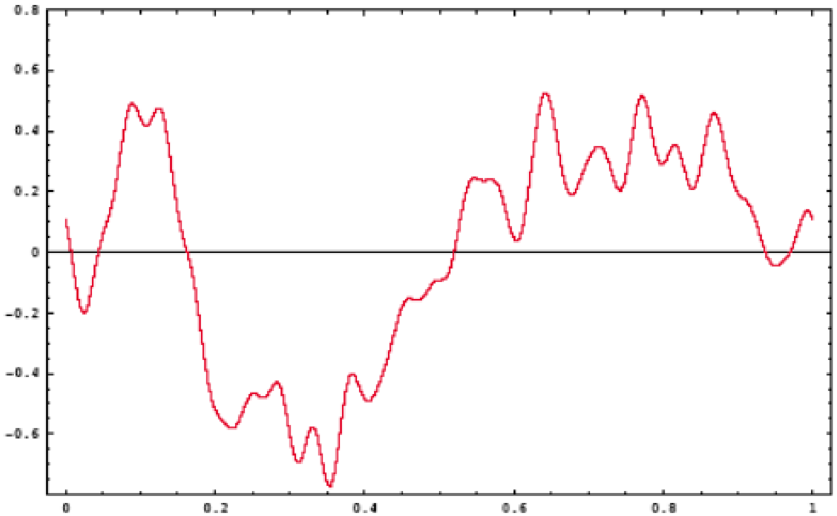
One realization of n=0 one-dimensional Gaussian Random Field



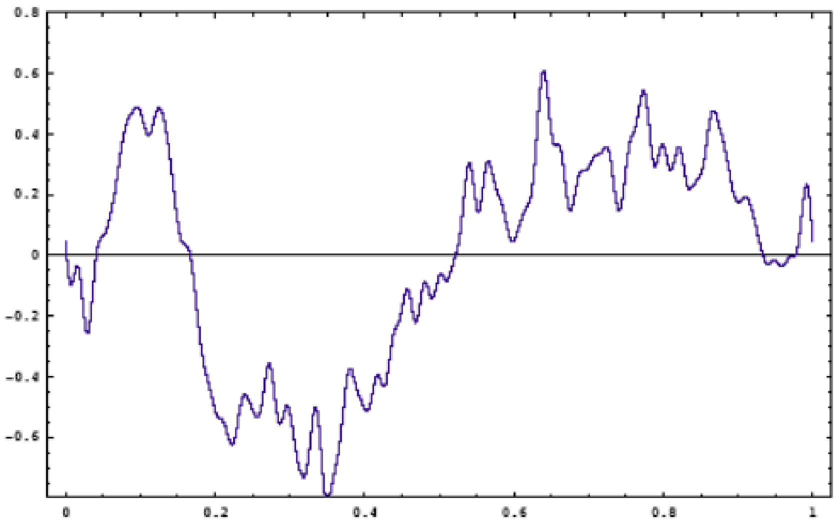
In a GRF, k-modes are indepen-dent

One realization of n=-2 one-dimensional Gaussian Random Field

First 25 Fourier modes



First 50 Fourier modes

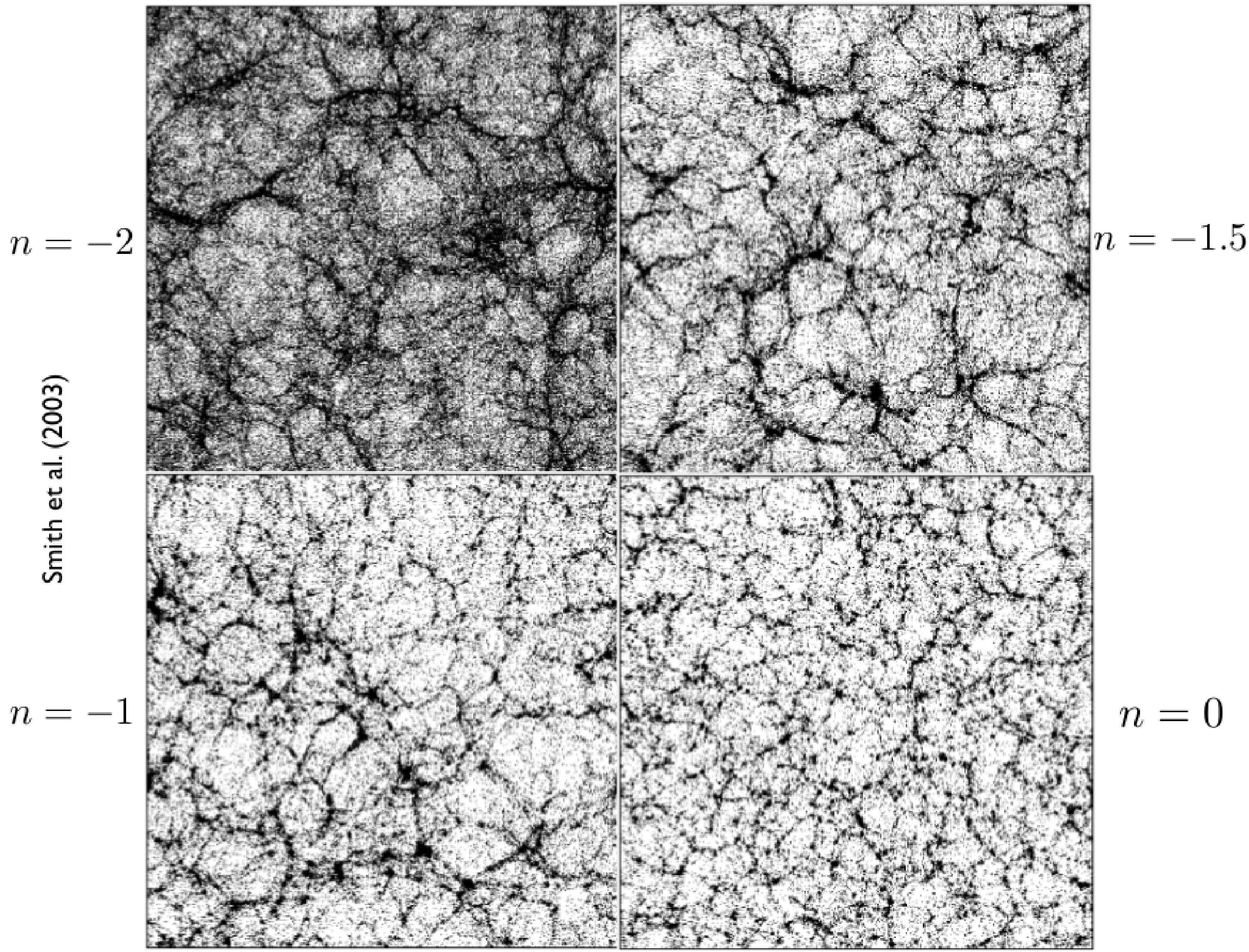


Gaussian field (contd.)

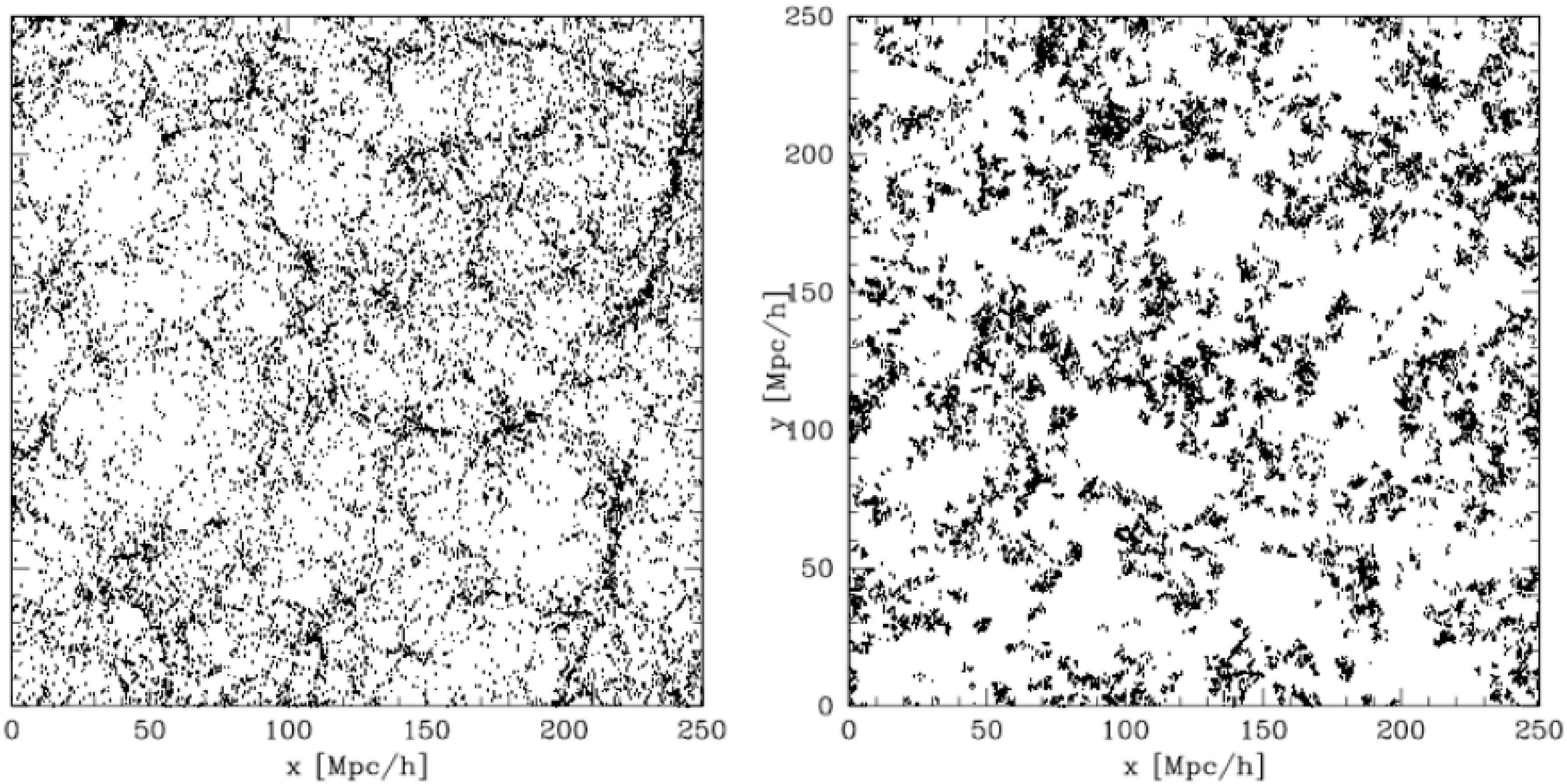
- Let $\delta(\mathbf{k})$ be multivariate Gaussian with zero-mean and diagonal covariance $\mathbf{P}_{ij} = <\delta(k_i)\,\delta(k_j)>$.
- Let $\delta(\mathbf{x}) = \Sigma\, \delta(\mathbf{k}) \exp(i\mathbf{k}.\mathbf{x}) = F_{ij}\, \delta(\mathbf{x}_j)$
- Since $\delta(\mathbf{x})$ is sum of (zero-mean) Gaussians for each **x**, it is itself a Gaussian number. So joint distribution of all $\delta(\mathbf{x})$ is multivariate Gaussian.

- HW1: Show that distribution of $\delta(\mathbf{x})$ is multivariate Gaussian with zero mean and covariance $\boldsymbol{\xi} = \mathbf{F}^T\, \mathbf{P}\, \mathbf{F}$. In general, ξ will not be diagonal:

Structure formation for P(k) ∝ k^n

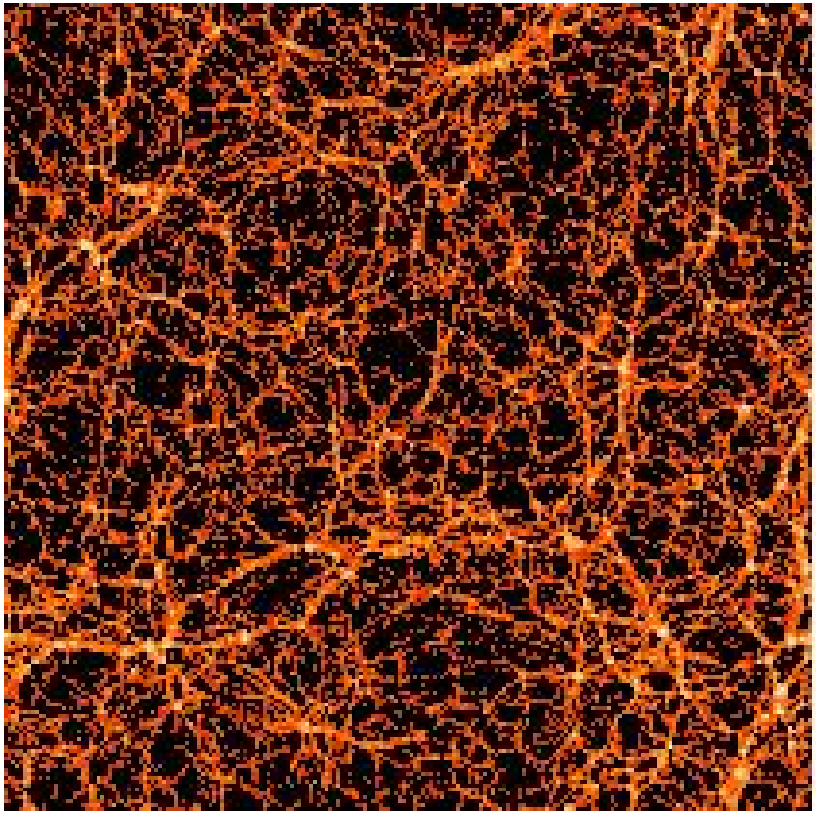


GRFs are special. In general there's (much) more to it than 2pt statistics

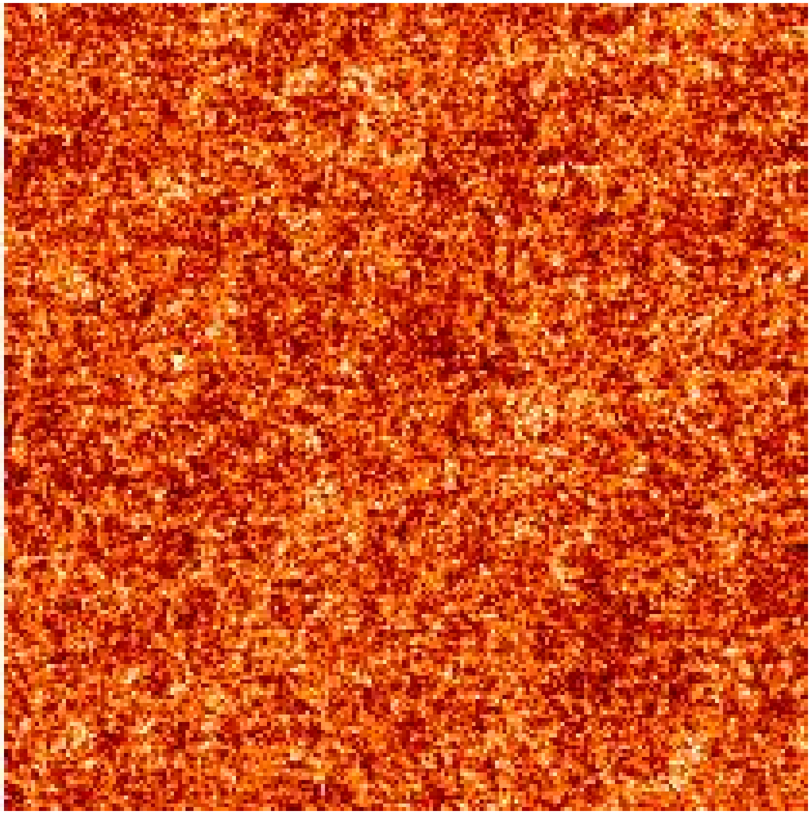


Both distributions have ~ same Pk

Full N-body



Scrambled phases



Both distributions have same P(k)