STATISTICAL PROPERTIES OF THE LARGE SCALE STRUCTURES:
CLUSTERING

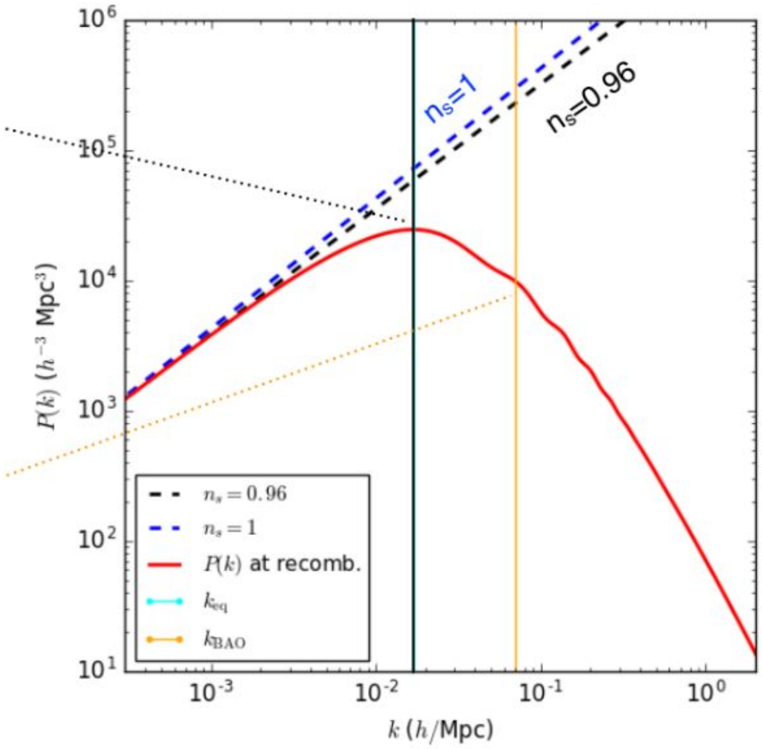
For a review on structure formation:
https://sites.astro.caltech.edu/~george/ay127/kamionkowski-perturbations-notes.pdf
https://people.ast.cam.ac.uk/~pettini/Intro%20Cosmology/Lecture14.pdf
For a review on BAO: https://arxiv.org/pdf/0910.5224.pdf
For a review on RSD: https://arxiv.org/pdf/astro-ph/9708102.pdf

EVOLUTION OF DENSITY PERTURBATIONS

Inflation generates primordial perturbations through the amplification of quantum fluctuations, which are stretched to astrophysical scales by the rapid expansion.

The simplest models of inflation predict that the initial fluctuations constitute a Gaussian random field, with an almost purely adiabatic primordial perturbations with a near scale-invariant power spectrum. In these models the primordial power spectrum is often described in terms of a spectral index n_s and an amplitude of the perturbations A_s as (k_p = 0.05 Mpc^{-1} = pivot scale):

$$P(k) = A_s \left(\frac{k}{k_p}\right)^{n_s - 1}$$

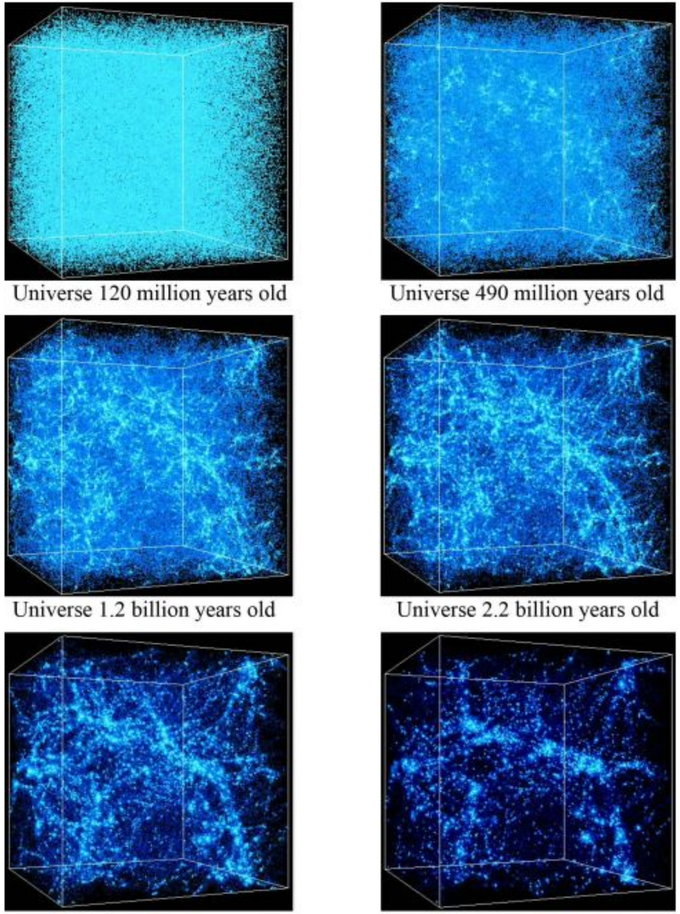


Large scale ⟷ **Small scale**

EVOLUTION OF DENSITY PERTURBATIONS

After the perturbations are created in the early Universe, they undergo a complex evolution which depends on the theory of gravity (GR), and the expansion history of the Universe.

- **Gravity is the dominant force that moves matter on the largest scales.**
- **The dark matter, which constitutes ~ 5/6 of the nonrelativistic matter in the Universe, is composed of “cold dark matter”, pressureless matter that interacts with everything else only gravitationally.**



Universe 120 million years old
Universe 490 million years old
Universe 1.2 billion years old
Universe 2.2 billion years old
Universe 6.0 billion years old
Universe 13.7 billion years old

EVOLUTION OF DENSITY PERTURBATIONS

- **Overdensity field:**

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle}$$

For CDM model:

	$z > z_{eq}$	$z < z_{eq}$
$R > R_H$	$\delta_r \propto \delta_m \propto a^2$	$\delta_r \propto \delta_m \propto a$
$R < R_H$	$\delta_{cdm} \sim const$ $\delta_{b+\gamma}\ oscillations$	$\delta_{cdm} \propto a$ $z < z_{rec}:\ \delta_b \rightarrow \delta_{cdm}$

Matter-radiation equality

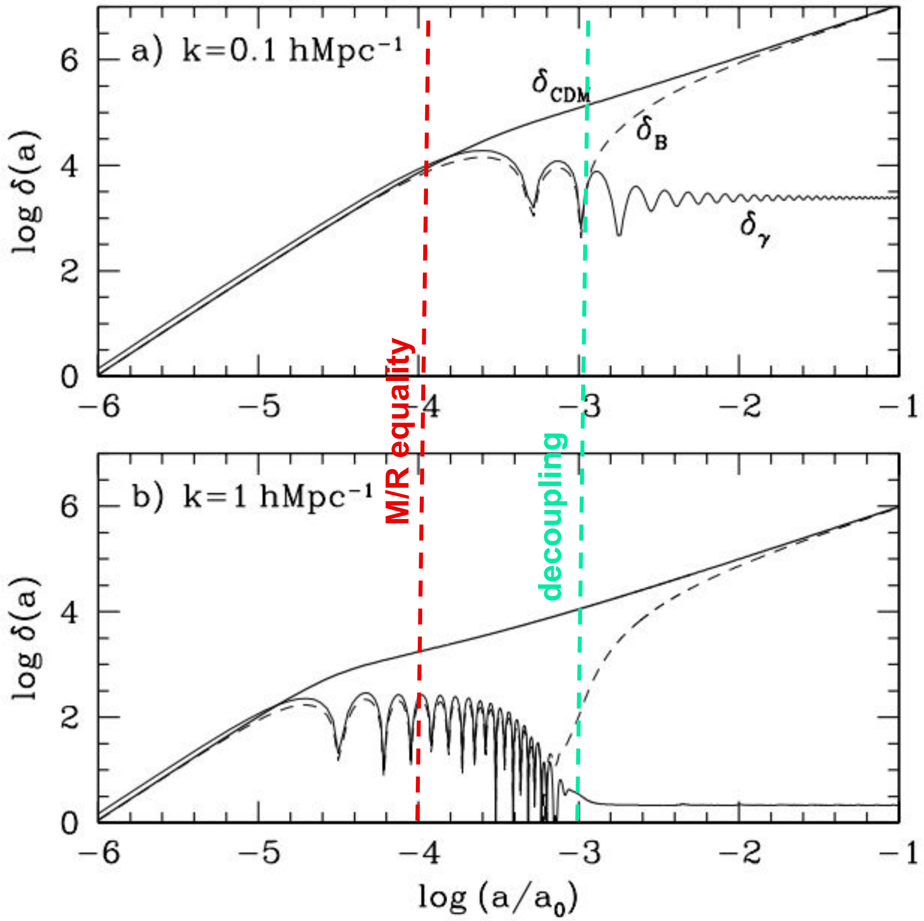
$$\frac{\rho_m(a_{eq})}{\rho_r(a_{eq})} = 1$$

$$1 + z_{eq} = \frac{a_0}{a_{eq}} \simeq 3250$$

Hubble radius

$$R_H(t) = \frac{c}{H(t)} \sim c\,t$$

EVOLUTION OF DENSITY PERTURBATIONS



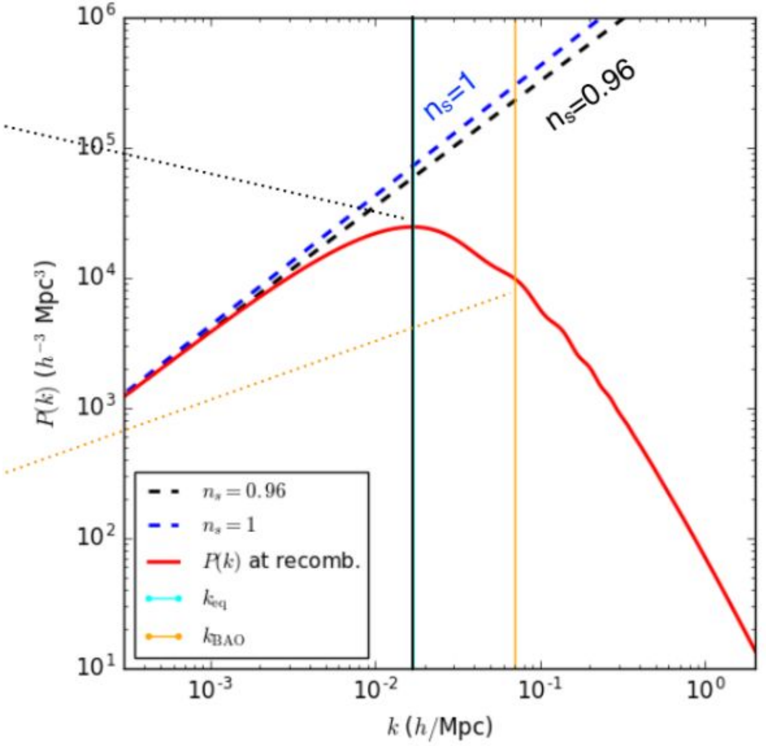
The evolution of adiabatic perturbations in a CDM universe with $\Omega_{\mathrm{m},0} = 1$, $\Omega_{\mathrm{B},0} = 0.05$, $h = 0.5$.
The scale factor is normalized at the present time.
Decoupling: $a \sim 10^{-3}$
Matter/radiation equality: $a \sim 10^4$

EVOLUTION OF DENSITY PERTURBATIONS

The primordial power spectrum of density fluctuations gets "processed" by the growth of density perturbations:

$$P(k,z) = P_{\rm primordial}(k)T^2(k,z)$$

where the transfer function T(k) takes into account the effects of gravitational amplification of density perturbation mode of wavelength *k*



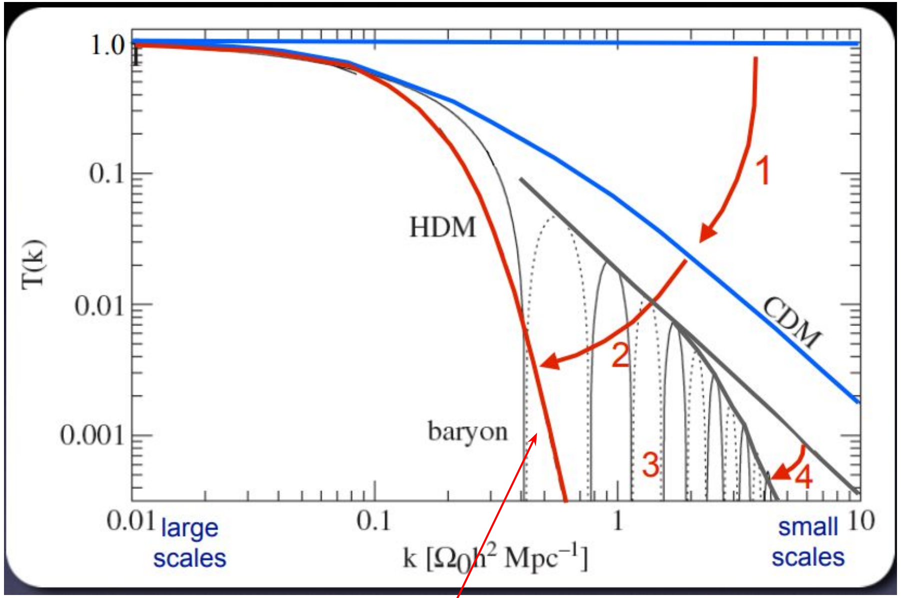
EVOLUTION OF DENSITY PERTURBATIONS

The primordial power spectrum of density fluctuations gets "processed" by the growth of density perturbations:

$$P(k, z) = P_{\rm primordial}(k) T^2(k, z)$$

where the transfer function T(k) takes into account the effects of gravitational amplification of density perturbation mode of wavelength *k*

Transfer function for different components



Note: The thermal velocity of DM particles determines the free streaming length below which structure formation is suppressed (free-streaming damping)

SIDE NOTE: HOT, WARM AND COLD DARK MATTER

Free-streaming cut-off scale:

$$k_{\rm FS} \sim 15.6 \frac{h}{\rm Mpc} \left(\frac{m_{\rm WDM}}{1{\rm keV}}\right)^{4/3} \left(\frac{0.12}{\Omega_{\rm DM} h^2}\right)^{1/3}$$

Cosmic Relics

Non-thermal Relics

not produced in thermal equilibrium (TE) with rest of Universe.

e.g., axions (FDM), monopoles, cosmic strings

Thermal Relics

are held in TE with other components of Universe until they `decouple', which happens when interaction rate $\Gamma = n\langle\sigma v\rangle$ drops below expansion rate $H(a)$

Warm Dark Matter

Hot Dark Matter

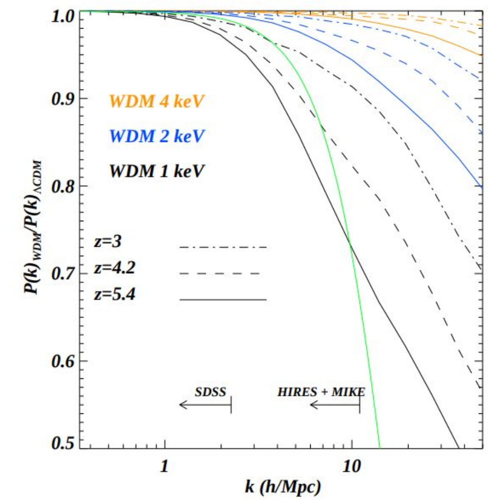
particles are still relativistic at decoupling, i.e., $3k_{\rm B}T_{\rm d} > m_{\rm X}c^2$

e.g., massive neutrinos

Cold Dark Matter

particles are non-relativistic at decoupling i.e., $3k_{\rm B}T_{\rm d} < m_{\rm X}c^2$

e.g., WIMPS

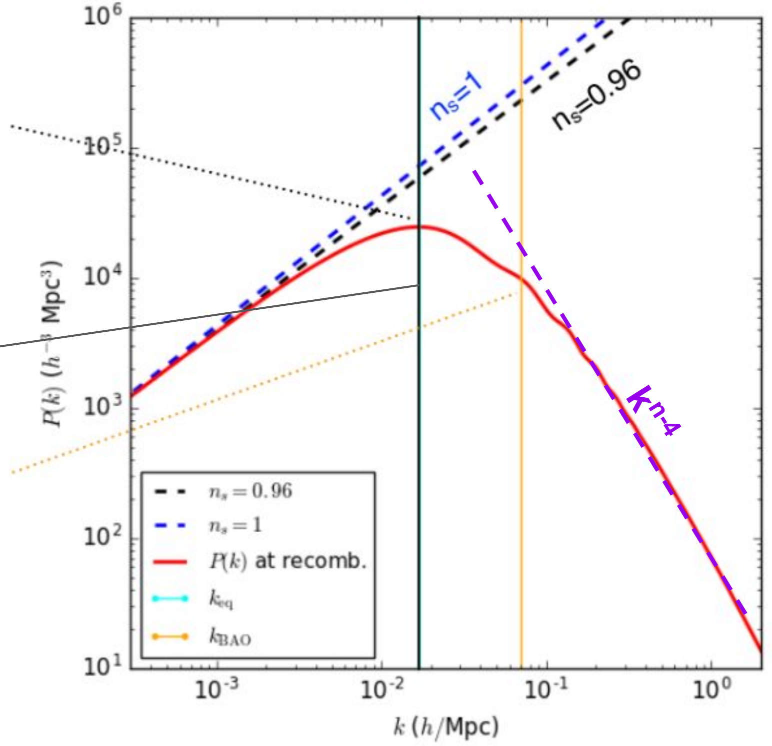


EVOLUTION OF DENSITY PERTURBATIONS

Transfer function (CDM):

$$T(k) \simeq \begin{cases} 1, & k \lesssim k_{\rm eq}, \\ (k/k_{\rm eq})^{-2}, & k \gtrsim k_{\rm eq}, \end{cases}$$

The characteristic length scale is the horizon-size at matter-radiation equality, $k_{eq} = 2\pi(ct_{eq})^{-1} \propto \Omega_{m,0}h^2$; before t_{eq}, below the horizon, matter fluctuations cannot growth.



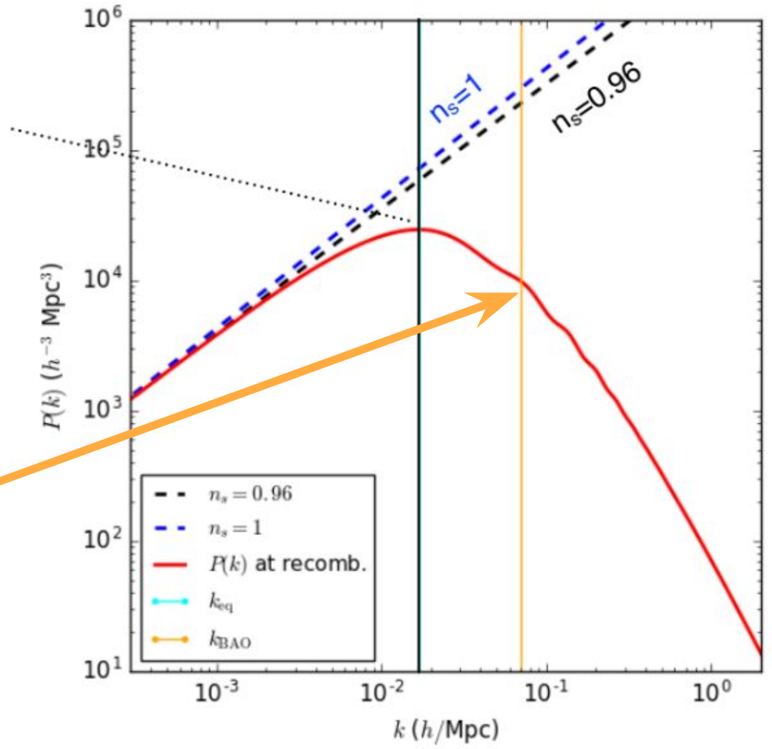
Thus, if $P_{primordial}(k) \propto k^n$, the processed power spectrum is:

- **$P(k) \propto k^n$ for $k \ll k_{eq}$**
- **$P(k) \propto k^{n-4}$ for $k \gg keq$**

EVOLUTION OF DENSITY PERTURBATIONS

Baryonic Acoustic Oscillations:

In the early, high-temperature Universe, baryons and photons were tightly coupled by Compton scattering, in a so-called photon-baryon fluid; the competing forces of radiation pressure and gravity set up oscillations in the photon-baryon fluid. As the Universe expands and cools down, atoms form (Recombination) and the interaction rate between baryons and photons decreases: photons begin to free-stream, leaving baryons in a shell with a radius approximately equal to the sound horizon at the time of decoupling. From that moment on, only the gravitational interaction between dark matter and baryonic matter remains. This characteristic radius is therefore imprinted as an overdensity and the power spectrum have an excess of power on this scale.

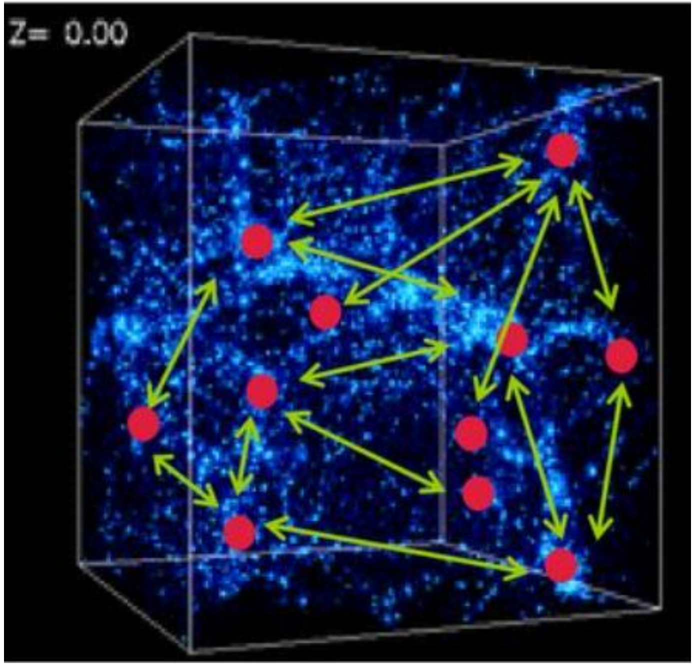


2PT CORRELATION FUNCTION

The most commonly used quantitative measure of large scale structure is the two-point correlation function. The two-point correlation function, ξ(*r*), is the excess probability (compared to an unclustered Poisson distribution) of finding two tracers of the matter density field (e.g. galaxies) separated by a distance *r* from each other:

$$dP = n(1 + \xi(r))dV$$

Where *n* is the mean number density of the tracer in question.



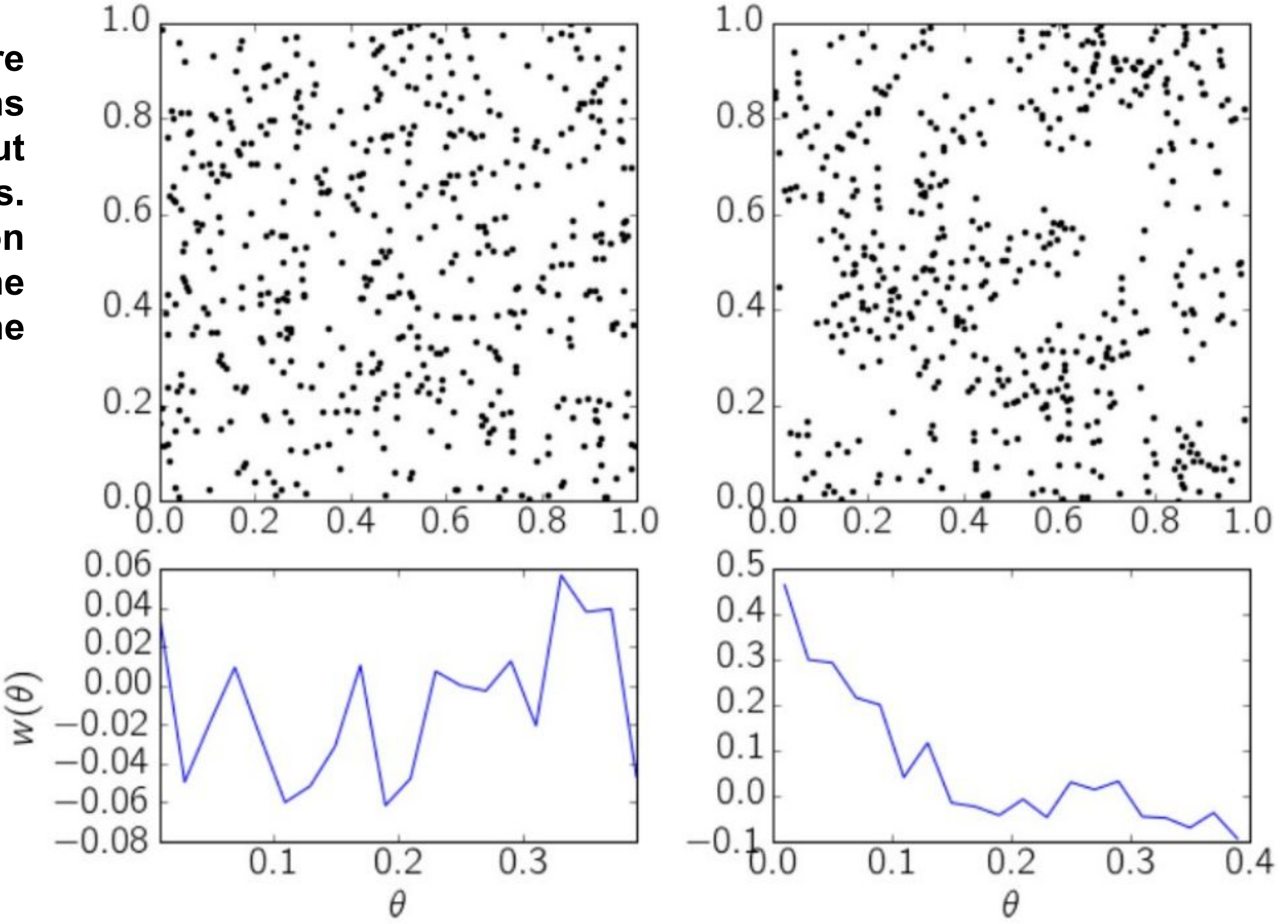
2 pt. Correlation Function Estimators:

$$\hat{\xi}(r) = \frac{DD(r) - RR(r)}{RR(r)}$$ (Peebles & Hauser, 1974)

$$\hat{\xi}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}$$ (Landy & Szalay, 1993)

2PT CORRELATION FUNCTION

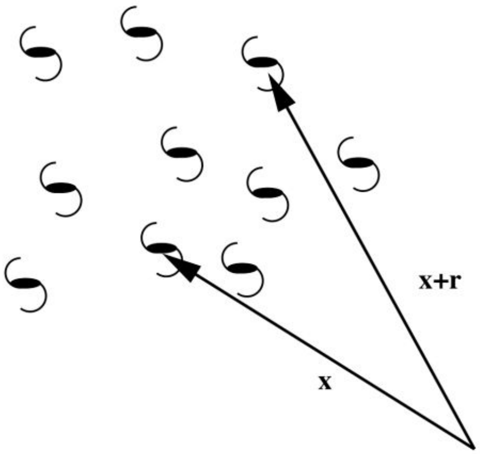
E.g.: the data points in Figure are drawn from distributions with equal densities, but different clustering properties. Using the two-point correlation function we can characterize the difference between the distributions.



2PT CORRELATION FUNCTION

Assuming the universe is isotropic, the correlation function is a function of a scalar distance. The two-point correlation function can then be written as

$$\xi(\vec{r}) := \langle \delta(\vec{x})\delta(\vec{x}+\vec{r}) \rangle$$



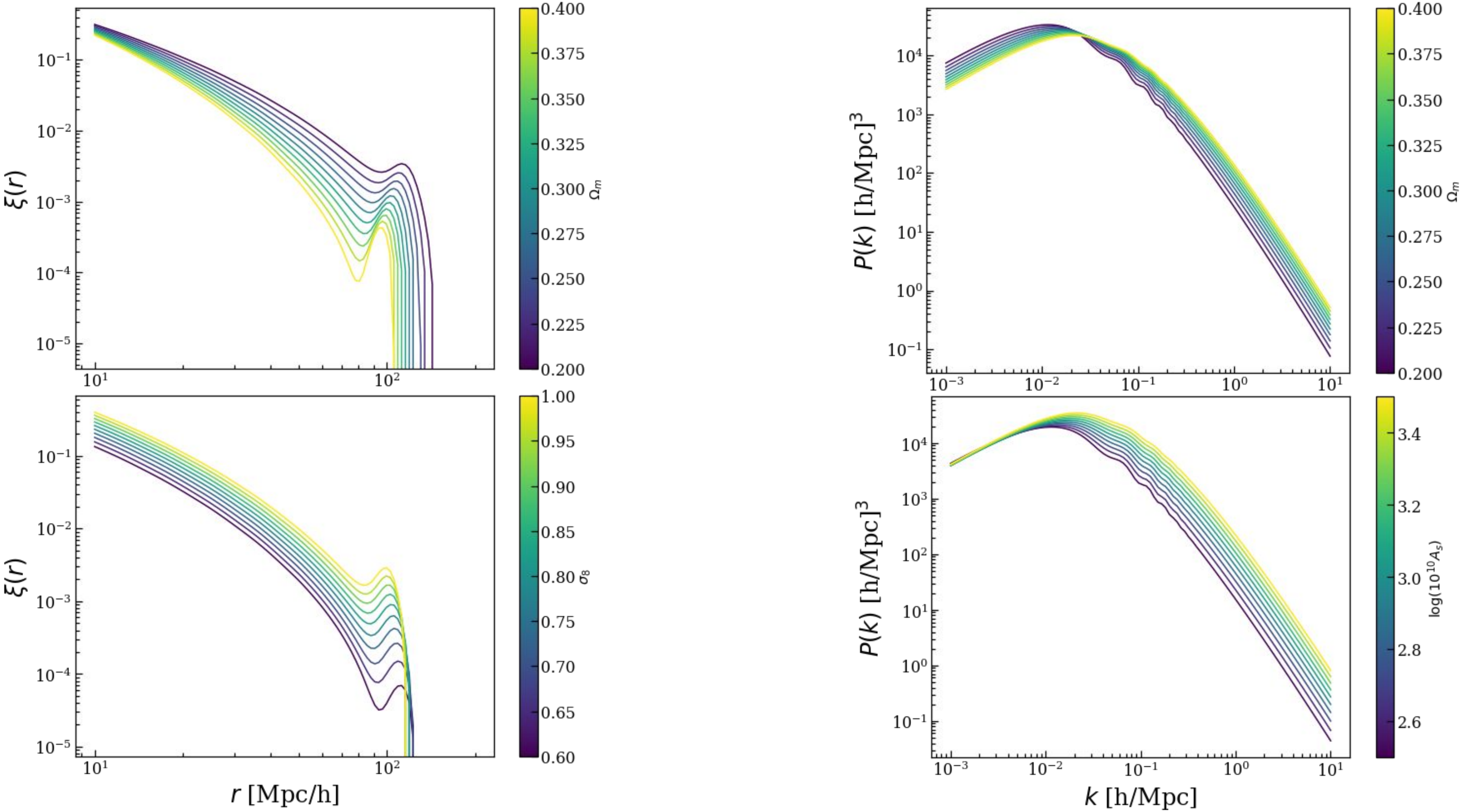
The 2PCF is closely connected to the power spectrum;
They form a Fourier transform pair:

$$\xi(\vec{r}) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3k P(\vec{k}) e^{i\vec{k}\cdot\vec{r}},$$

$$P(\vec{k}) = \int \mathrm{d}^3r \xi(\vec{r}) e^{-i\vec{k}\cdot\vec{r}},$$

A Gaussian random field is completely specified by either the two-point correlation function , or, equivalently, the power spectrum P(k)!

LINEAR 2PT CF AND POWER SPECTRUM

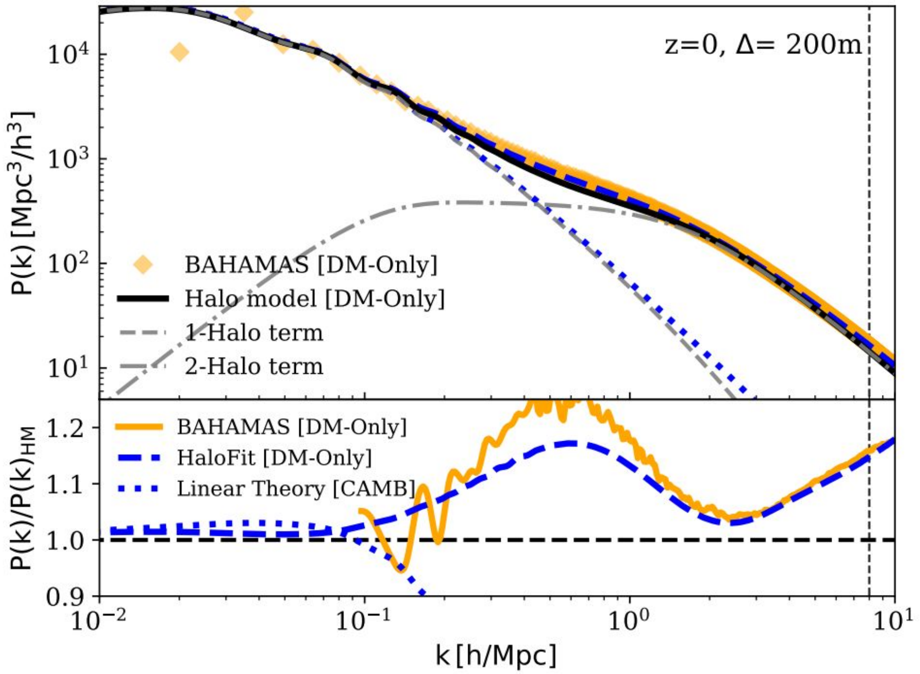


NON-LINEAR EVOLUTION

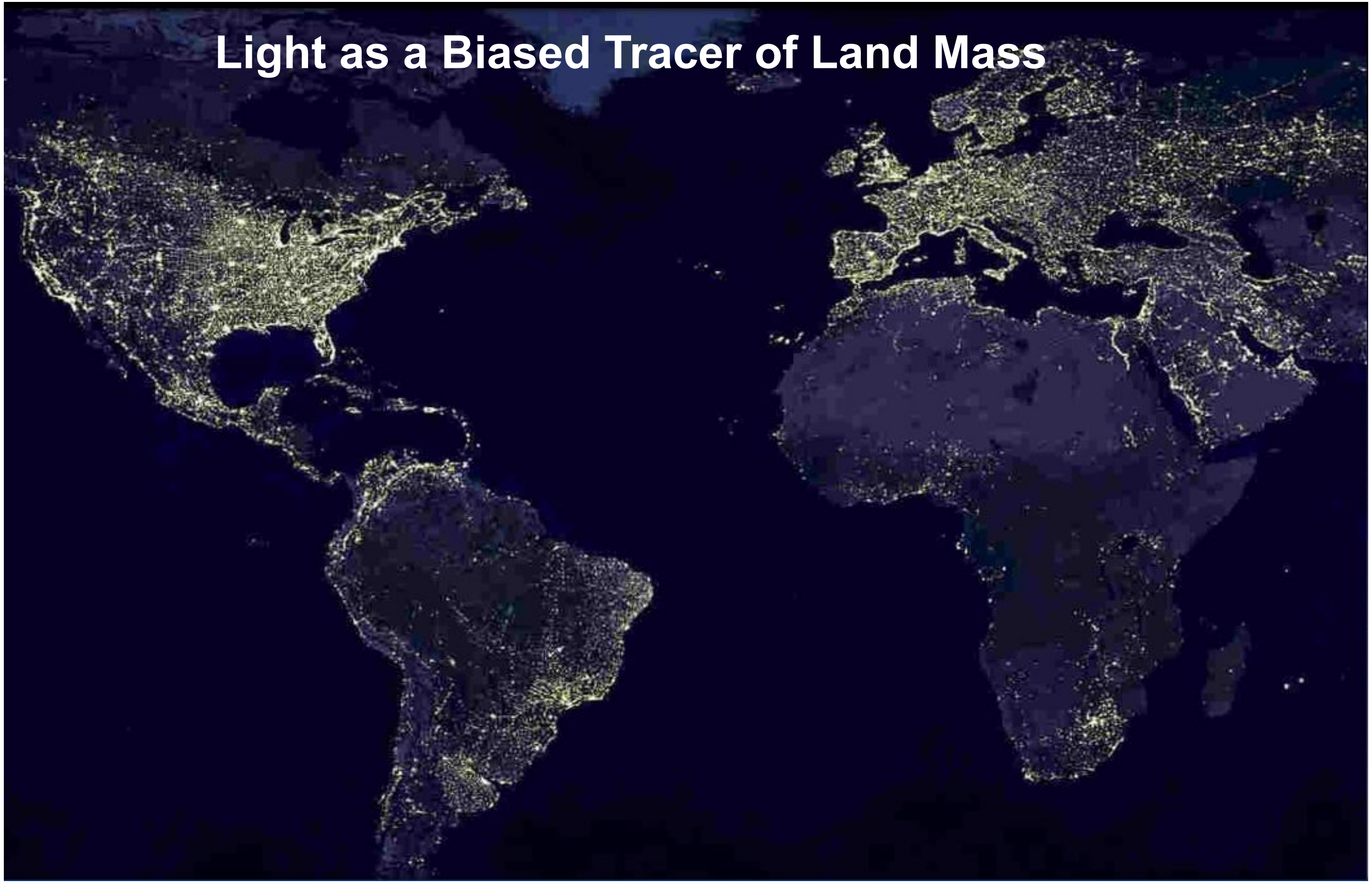
In the linear regime (δ<<1)we can calculate the evolution of a density field of arbitrary form using linear perturbation theory.

In the non-linear regime (δ>1) perturbation theory is no longer valid. Modes start to couple to each other, and one can no longer describe the evolution of the density field with a simple growth rate: in general, no analytic solutions exist and we must rely on simulations.

Because of this mode-coupling, the density field loses its Gaussian properties, i.e., in the non-linear regime, we no longer have a Gaussian random field. Hence, higher-order moments are required to completely specify density field.

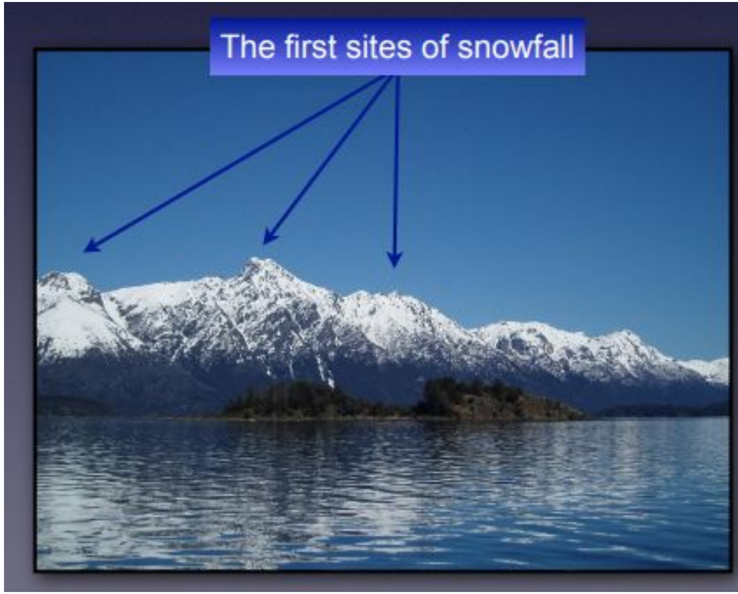


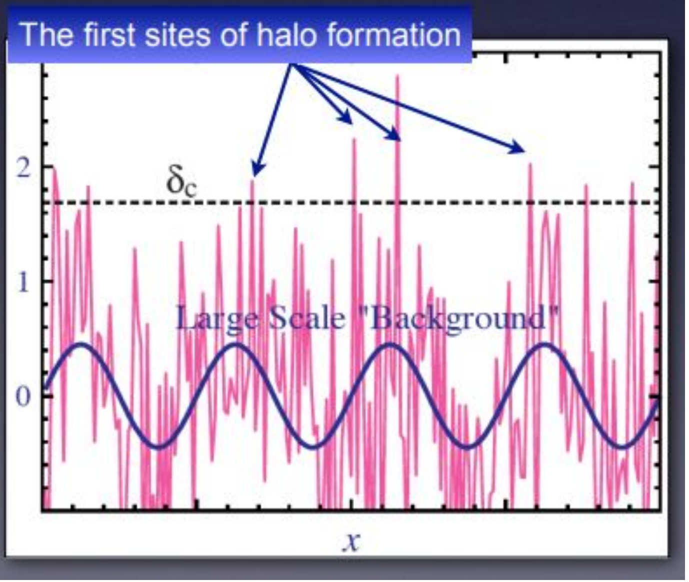
BIASED TRACERS OF THE DENSITY FIELD



HALO BIAS

Halo formation is not a random process; haloes are not a Poisson sampling of the matter field. Rather, they only form where the (smoothed) density field has a sufficiently high value: the critical overdensity for collapse. This `threshold' causes haloes to be biased tracers of the mass distribution. Because of the modulation of the small-scale density field by the long-wavelength modes, overdense regions (on large scales) contain enhanced abundance of dark matter haloes, so that these haloes display enhanced clustering.





2PT CORRELATION FUNCTION: BIAS

The collapsed structures we observe today, that is, galaxies and galaxy clusters, are formed at the locations peaks of the matter density field and more massive objects correspond to higher peaks. Thus, we expect their distribution to be a biased tracers of the underlying mass distribution and the bias grows with the mass of the objects.

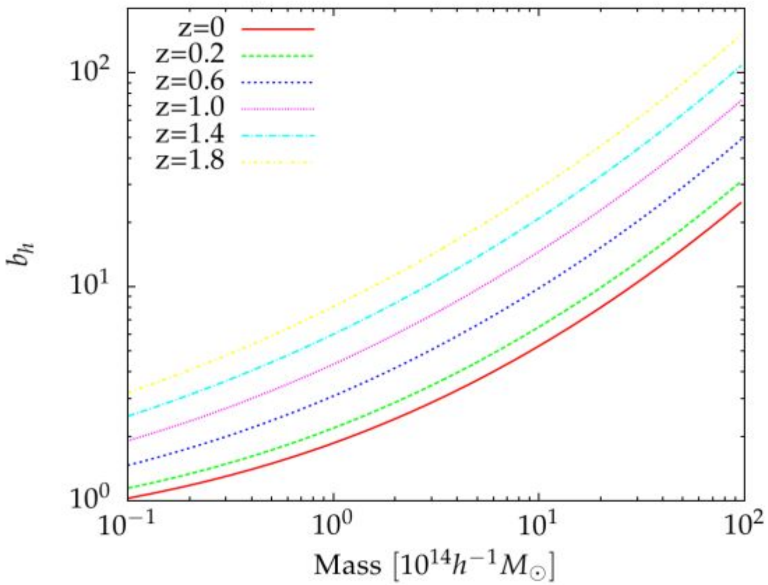
$$\delta_h = b_h \delta_m$$

$$\xi_{hh}(r) = \langle \delta_h(x)\delta_h(x+r)\rangle = b_h^2 \xi_{mm}$$

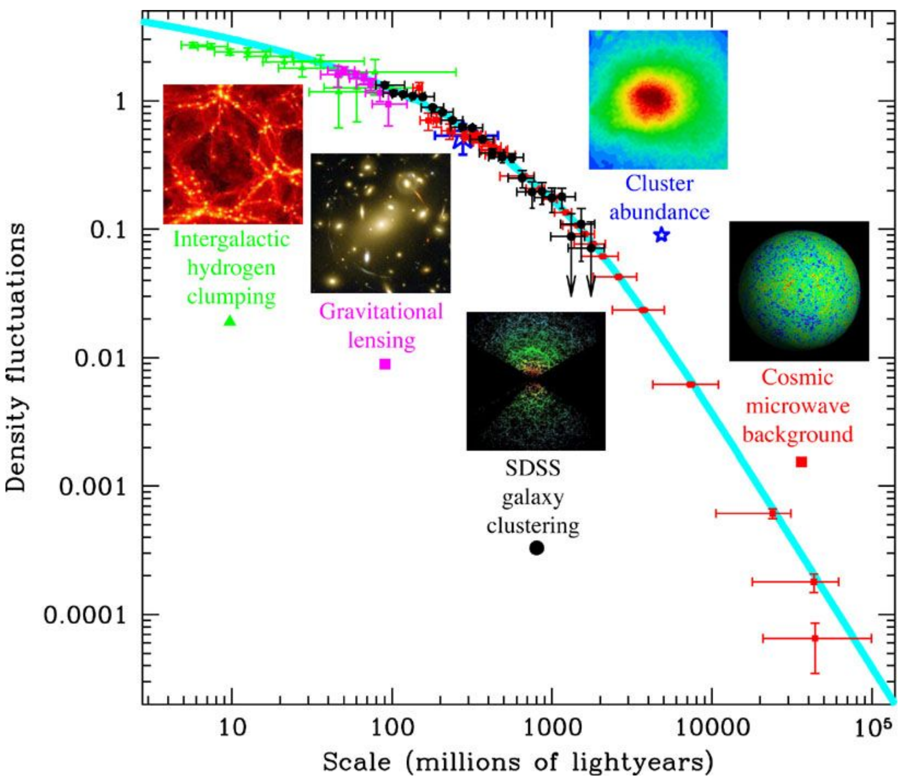
$$\mathscr{F} \downarrow \uparrow \mathscr{F}^{-1}$$

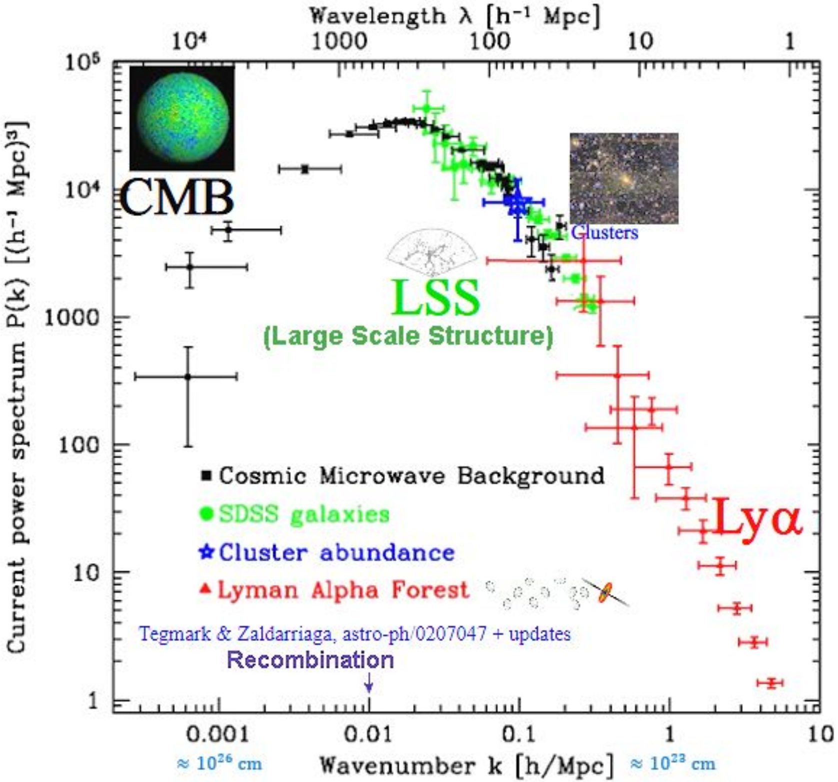
$$P_{hh}(k) = \langle \tilde{\delta_h}(k)\tilde{\delta_h}^*(k')\rangle = b_h^2 P_{mm}(k)$$

Halo bias vs mass and redshift



TRACERS OF THE MATTER DENSITY FIELD



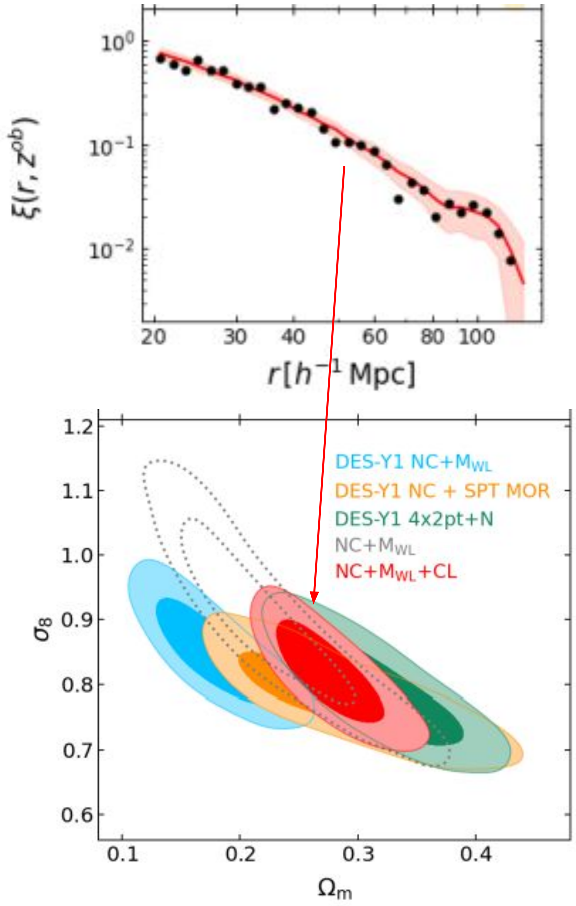


CLUSTERING OF CLUSTERS

Galaxy clusters, being the more massive structures of the Universe, have the larger bias. Moreover, their bias can be predicted from theory (or calibrated from simulations) from the halo mass function:

$$\boxed{\delta^{\mathrm{L}}_{\mathrm{halo}} = \frac{\mathcal{N}(M|\delta_0, S_0)}{(\mathrm{d}n(M)/\mathrm{d}M)V_0} - 1}$$

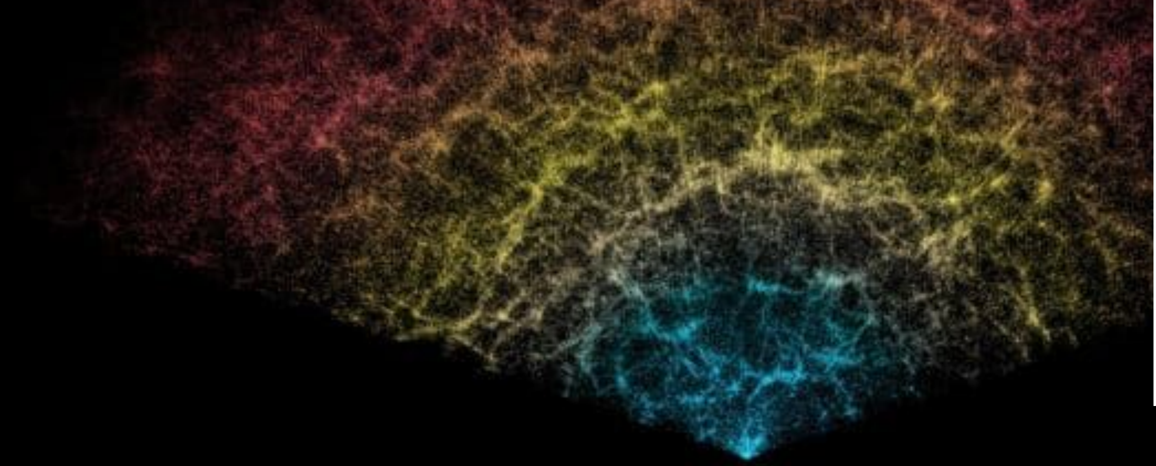
On the other hand, forming on the rare high density peaks of the matter density field, galaxy clusters are a sparse tracer of δ_m, and their 2pt CF signal is largely disturbed by shot noise.



https://arxiv.org/pdf/2310.09146.pdf

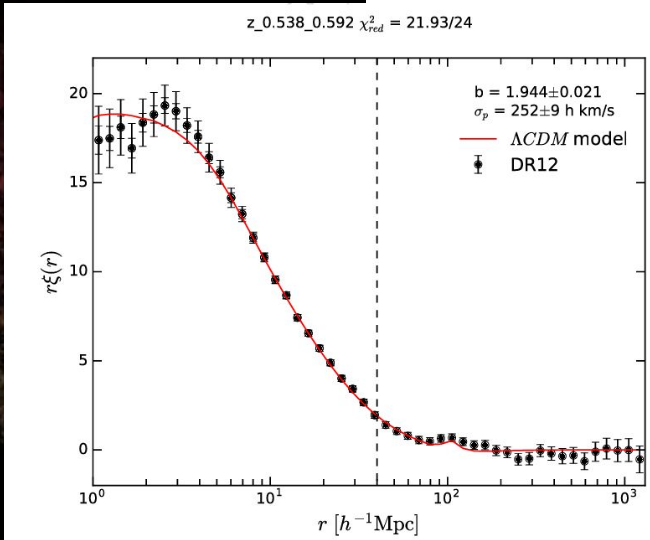
CLUSTERING OF GALAXIES

Galaxies provide a much wider sample ($\sim 10^8$) compared to clusters ($\sim 10^{3-4}$), thus greatly reducing the shot noise on 2pt CF measurements. On the other hand galaxy bias cannot be predicted from theory, and in general it depends on the specific galaxy sample (e.g. color and/or magnitude selection) employed for the analysis.



Galaxy map from Dark Energy Spectroscopic Instrument

The correlation function of the CMASS galaxies in the [0.538 − 0.592] redshift range.



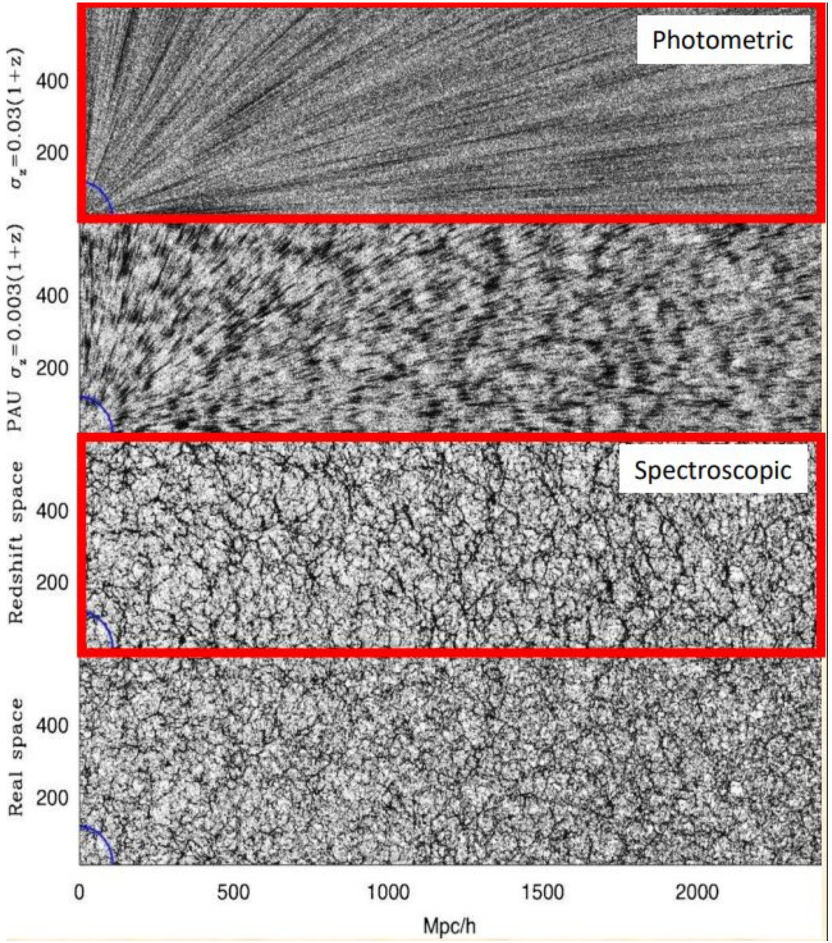
SPECTROSCOPIC VS PHOTOMETRIC SURVEYS

The clustering signal is 3D. However, while angular positions are in general easy to measure, the radial distance can be inferred only from redshift measurements, i.e. the shift to longer wavelengths of the light emitted by distant galaxies due to the recessional velocity ($v=H_0 \cdot d$).

This can be done in two ways:

1) Spectroscopic redshift [$\sigma_z \sim 0.001\times(1+z)$] : Measuring the spectrum of galaxies, and hence the shift of known emission/absorption lines.

2) Photometric redshift [$\sigma_z \sim 0.01\times(1+z)$]: Estimate from broadband photometry by fitting template spectra predicted from galaxy Spectral Energy Distribution or from observed galaxies with known spectroscopic redshift.



SPECTROSCOPIC VS PHOTOMETRIC SURVEYS

Photometric redshifts can be measured much faster than their spectroscopic counterparts. In spectroscopy, the light from the galaxy is separated into narrow wavelength bins a few angstroms across. Each bin then receives only a small fraction of the total light from the galaxy. Hence, to achieve a sufficiently high signal-to-noise ratio in each bin, long integration times are required, and only luminous sources can be targeted. For photometry, however, the bins are much larger, and it requires only a short exposure time to reach the same signal-to-noise ratio. Further, imaging detectors usually cover a greater area of the sky than multi-object spectrographs. This means that the redshifts of more objects can be measured simultaneously by using photometry than by spectroscopy. Thus in general:

Spectroscopic survey:

- **Accurate 3D reconstruction**
- **Medium depth**
- **Low density**
- **Strong selection effects**

Photometric survey:

- **Low resolution along the line-of-sight**
- **Usually deeper and larger area (i.e. larger volume)**
- **High density**
- **~ No selection effects**

SPECTROSCOPIC VS PHOTOMETRIC SURVEYS

Redshift errors perturbs the distance measurements along the line of sight affecting the clustering signal by reducing 2pt CF slope and smearing the BAO feature (see later)

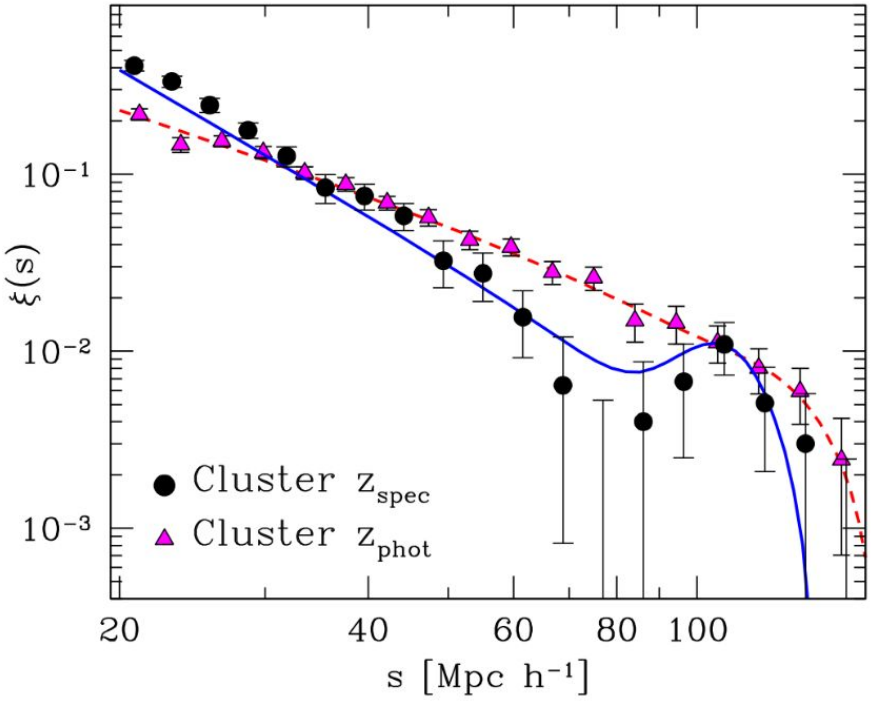
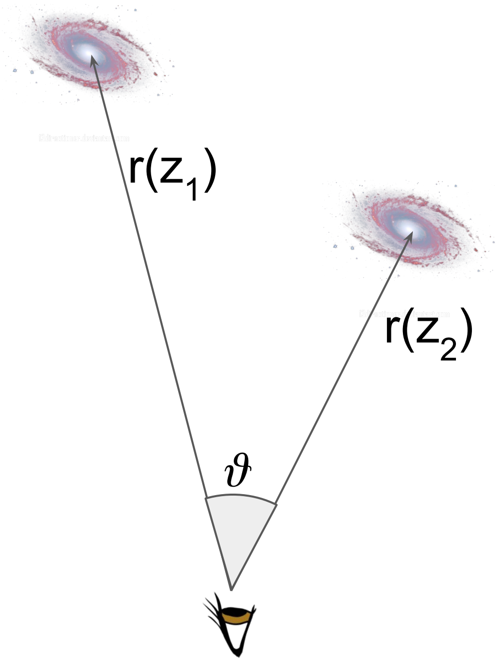


Figure 3.5: Comparison between the redshift-space two-point correlation functions of spectroscopic (black dots) and photometric (magenta triangles) cluster samples, selected from the Sloan Digital Sky Survey, with measured redshifts in the range $0.1 < z < 0.42$. The lines show the best-fitting empirical models obtained for spectroscopic (blue lines) and photometric clusters (dashed red lines). Credits to Veropalumbo et al. (2014).

ANGULAR 2PT CORRELATION FUNCTION

If the resolution along the line of sight is poor (e.g. in a photometric survey) or none (CMB last scattering surface), one can measure the angular correlation function. In analogy with ξ(r), we can define w(ϑ) as the excess probability of finding two tracers separated by an angular distance ϑ on the sky:

$$dP(\theta) = n(1 + w(\theta))d\Omega$$



w(ϑ) is related to the real space correlation function via a line-of-sight projection.

$$w(\theta) = \langle \delta(\hat{n})\delta(\hat{n} + \hat{\theta}) \rangle =$$

$$= \int\int dz_1 dz_2 \xi(\sqrt{r^2(z_1) + r^2(z_2) - 2r(z_1)r(z_2)cos\theta})$$

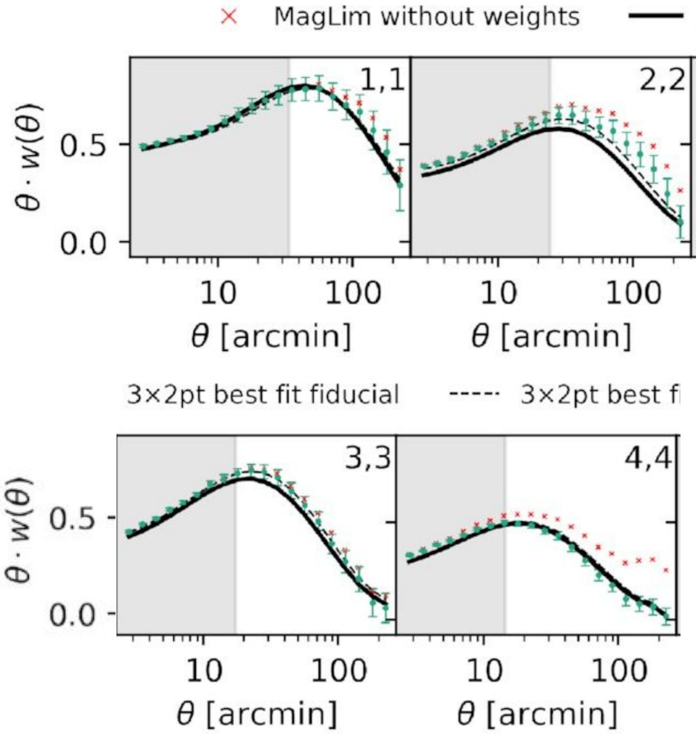
ANGULAR 2PT CF AND POWER SPECTRUM

Similarly to the real space 2pt CF we can relate the angular CF to the angular power spectrum, *C_ℓ*, via the multipole expansion:

$$w(\theta) = \sum_\ell \left(\frac{2\ell+1}{4\pi}\right) L_\ell(\cos\theta) C_\ell$$

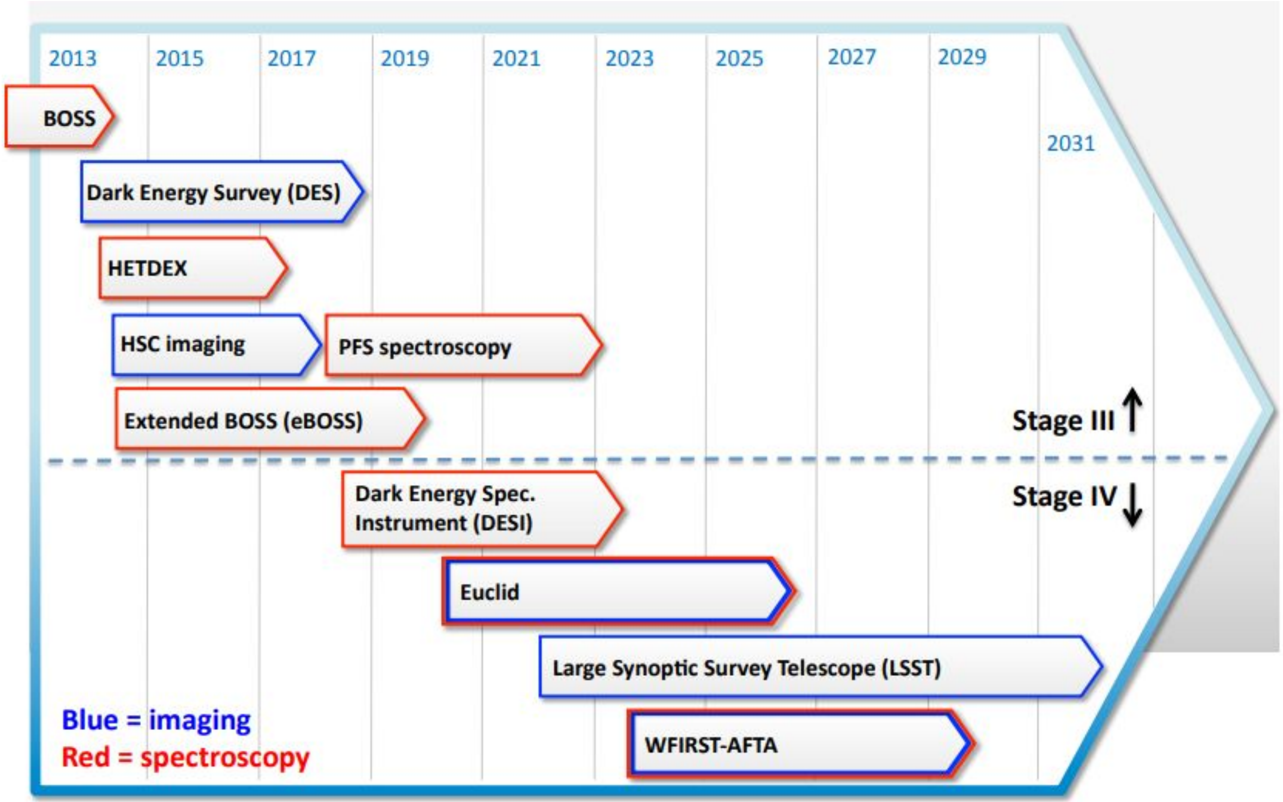
Legendre polynomial of degree ℓ

$$C_\ell = 2\pi n_\Omega \int_{-1}^{+1} w(\theta) L_\ell(\cos\theta) \mathrm{d}\cos\theta$$



DES Y3 angular galaxy clustering in 4 tomographic bins between 0.2<z<0.85

SPECTROSCOPIC VS PHOTOMETRIC SURVEYS



SPECTROSCOPIC VS PHOTOMETRIC SURVEYS

Galaxy surveys

Project	Dates	Area/deg2	Data	Redshift	Methods
BOSS	2008-2014	10000	Opt-S	0.3-0.7 (gal) 2-3.5 (Lyα Forest)	BAO/RSD
DES	2013-2018	5000	Opt-I	0.2-1.5	WL/CL/BAO/SN
eBOSS	2014-2020	7500	Opt-S	0.6-2.0 (gal/QSO) 2-3.5 (Lyα Forest)	BAO/RSD
SuMIRE	2014-2024	1500	Opt-I Opt-NIR-S	0.2-1.5 0.8-2.4 (gals)	WL/CL/ BAO/RSD
HETDEX	2014-2019	300	Opt-S	1.9-3.5 (gals)	BAO/RSD
DESI	2019-2024	14000	Opt-S	0-2 (gals) 2-3.5 (QSO/Lyα Forest)	BAO/RSD
LSST	2020-2030	20000	Opt-I	0.2-2	WL/CL/BAO/SN
Euclid	2020-2026	15000	Opt-I NIR-S	0.2-2 0.7-2.2 (gals)	WL/CL/BAO/RSD
WFIRST	2024-2030	2200	NIR-I NIR-S	1.0-3.0 (gals)	WL/CL/SN/BAO/RSD

From PDG 2016