Large-scale Structure

Lecture 2:
P(k) smoothing; galaxy bias; estimators
Baryon acoustic oscillations (BAO)
Redshift-space distortions (RSD)

Dragan Huterer

Bucharest 2025 PhD School
July 1-6, 2025

Last Time:

$$\Delta^2(k,a) = A\,\frac{4}{25}\,\frac{1}{\Omega_M^2}\left(\frac{k}{k_{\rm piv}}\right)^{n-1}\left(\frac{k}{H_0}\right)^4 [ag(a)]^2\,T^2(k)\,T_{\rm nl}(k)$$

A is the primordial amplitude (dimensionless, ~10^{-10})

Ω_M is the density of matter relative to critical

k_{piv} is some chosen pivot; modern convention is k_{piv}=0.05 Mpc^{-1}

H_0 is the Hubble constant

g(a) is the growth suppression factor

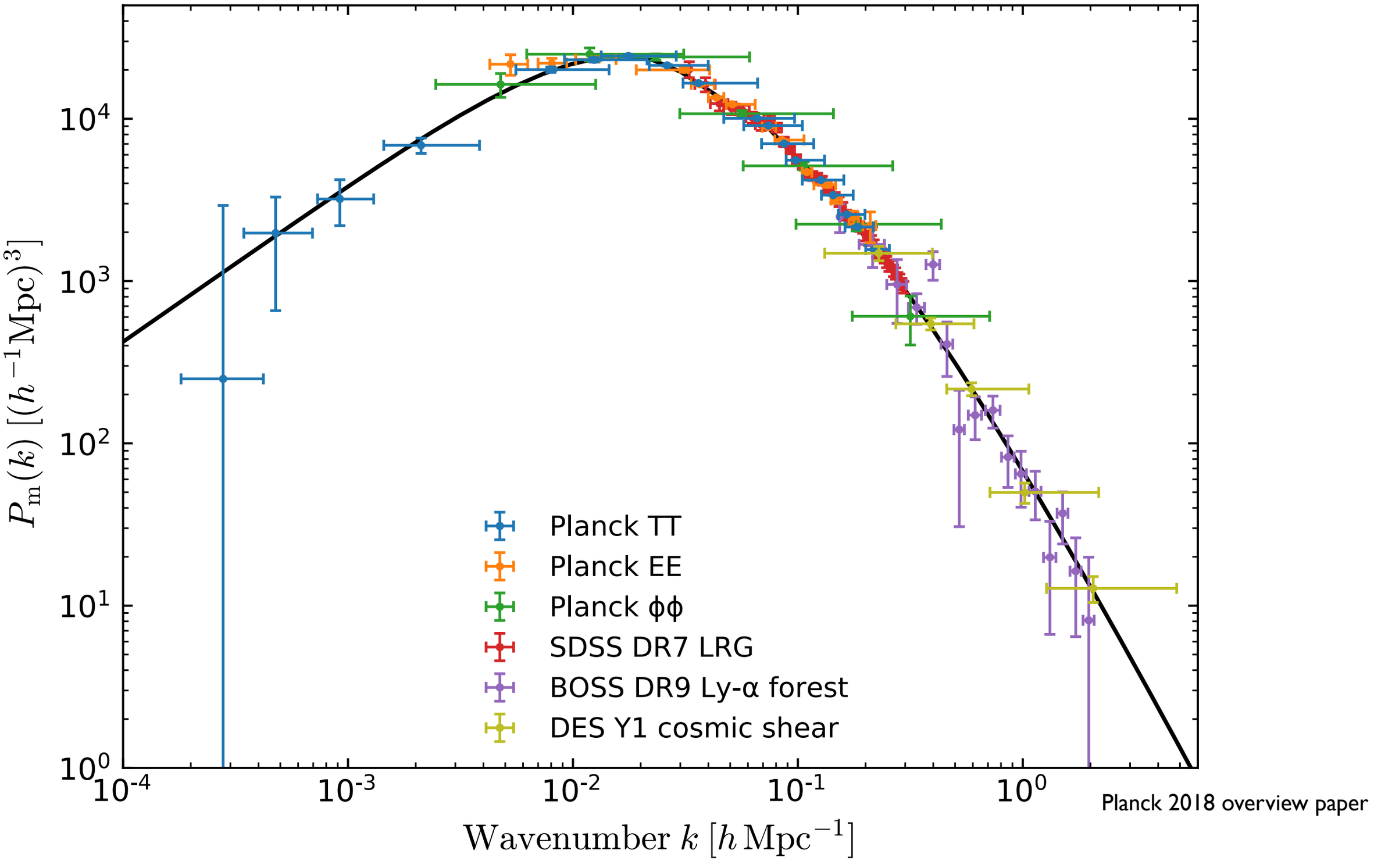
T(k) is the transfer function, accounts mainly for "turnover" in power spectrum due to radiation-matter transition.

$T_{nl}(k)$ is the prescription for nonlinear clustering; super important on scales $k \gtrsim 0.1$ h Mpc^{-1}

Power spectrum is a key quantity in cosmology

- Most LSS probes effectively measure it (usually integrated with some geometrical "kernel"); see next slide….
- and Inflation predicts it, so…
- it's a great "meeting place" between theory and data!
- In the limit of Gaussian LSS, contains **all** information (but ok, LSS is not Gaussian….)

Most* key LSS probes essentially measure P(k)



*Notable exceptions: SN Ia, cluster counts (≈ mass function)

A few more points about the power spectrum

- So far we have not assumed any **smoothing**. However in reality - theory or measurement - you have to assume that the density field is smoothed (why?).
- So far we talked about the matter power spectrum. Typically we measure the galaxy power spectrum. The two are not quite identical - they are related by **galaxy bias.**
- So far we have not discussed how you'd actually **measure** the galaxy power spectrum from some data.

We will next address these three points.

Smoothed overdensity

An "observed" delta
is necessarily smoothed

$$\delta(\vec{r}, R) = \int W(|\vec{r} - \vec{r}'|)\delta(\vec{r}')d^3\vec{r}'$$

$$\delta_{\vec{k}}(R) = W(k, R)\delta_{\vec{k}} \qquad W_{\rm TH}(k, R) = 3\frac{\sin(kR) - kR\cos(kR)}{(kR)^3} = \frac{3j_1(kR)}{kR}$$

Let us write the zero-lag correlation function with top-hat-smoothed field

$$\xi_{\rm TH}(0) = \int_0^\infty \Delta^2(k)|W_{\rm TH}(k, R)|^2 d\ln k$$

or, renaming it to agree with the literature, this is the
amplitude of mass fluctuations (squared) smoothed on scale R

$$\boxed{\sigma^2(R) = \int_0^\infty \Delta^2(k)\left(\frac{3j_1(kR)}{kR}\right)^2 d\ln k}$$

The famous sigma-eight

$$\sigma^2(R) = \int_0^\infty \Delta^2(k) \left(\frac{3 j_1(kR)}{kR} \right)^2 d\ln k$$

The amplitude of mass fluctuations σ(R)

is a derived quantity (power spectrum is "fundamental"), but very useful because it is a number summarizing the (square root of the) "amount of power" on a typical scale R

Can calculate it at any smoothing scale R (and any redshift z, suppressed in Eq above), but one choice is historically famous:

$$\sigma_8 \equiv \sigma(R = 8\,h^{-1}\mathrm{Mpc}, z = 0)$$

- σ_8 goes waaaay back to 1980s - was used as the measure of the overall amount of power/clustering on typical scales accessible in galaxy surveys
- In 1990s-2000s question of whether $\sigma_8 \approx 0.6$ or $\sigma_8 \approx 1.0$
- The answer is of course in the middle, $\sigma_8 \approx \mathbf{0.8}$, BUT
- Tension between CMB ($\sigma_8 \approx 0.78$) and grav lensing ($\sigma_8 \approx 0.82$) - at the forefront of research in cosmology today

So how do you <u>estimate</u> P(k) or ξ(r)?

The subject of estimators is science in its own right.
Many options… we discuss the simplest one.

Remember that ξ(r) is excess probability, $dP = n^2(1 + \xi(r_{12}))\, dV_1 dV_2$

Then how about a Peebles-Hauser (1974) estimator

$$\hat{\xi}_{\rm PH}(r) = \left(\frac{N_{\rm rand}}{N_{\rm data}}\right)^2 \frac{\rm DD(r)}{\rm RR(r)} - 1$$

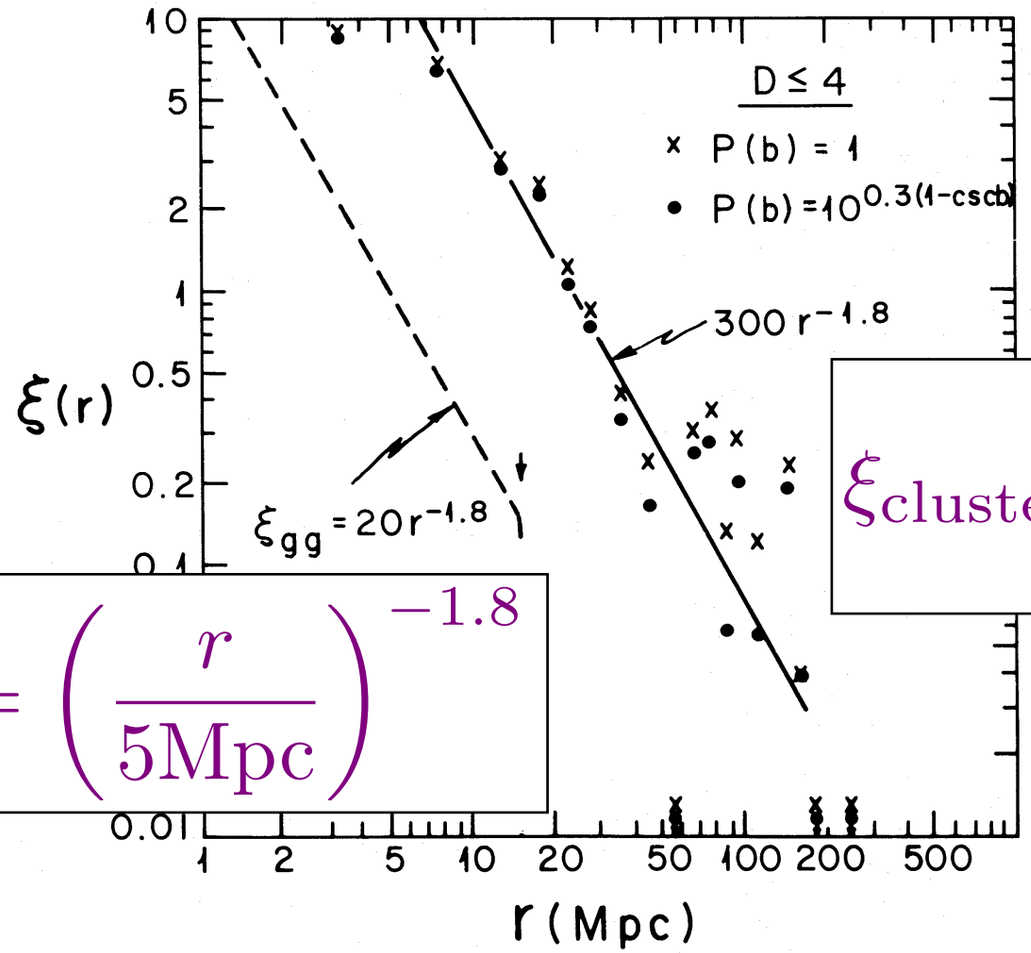
number of pairs separated by r±Δr in data (DD) and random (RR) catalogues

Better: (smaller variance): Landy-Szalay (1993) estimator

$$\hat{\xi}_{\rm LS} = \left(\frac{N_{\rm rand}}{N_{\rm data}}\right)^2 \frac{\rm DD(r)}{\rm RR(r)} - 2\frac{N_{\rm rand}}{N_{\rm data}}\frac{\rm DR(r)}{\rm RR(r)} + 1$$

Done naively, computation is N^2 process.
Can be drastically sped up by various clever algorithms.

(Galaxy/halo) bias



Bahcall & Soneira 1983

$$\xi_{\rm galaxies}(r) = \left(\frac{r}{5{\rm Mpc}}\right)^{-1.8}$$

$$\xi_{\rm clusters}(r) = \left(\frac{r}{25{\rm Mpc}}\right)^{-1.8}$$

Explained by Kaiser (1984):

$$\text{bias} \equiv \frac{\text{clustering of galaxies}}{\text{clustering of dark matter}} = \frac{\left(\frac{\delta\rho}{\rho}\right)_{\rm halos}}{\left(\frac{\delta\rho}{\rho}\right)_{\rm DM}}$$

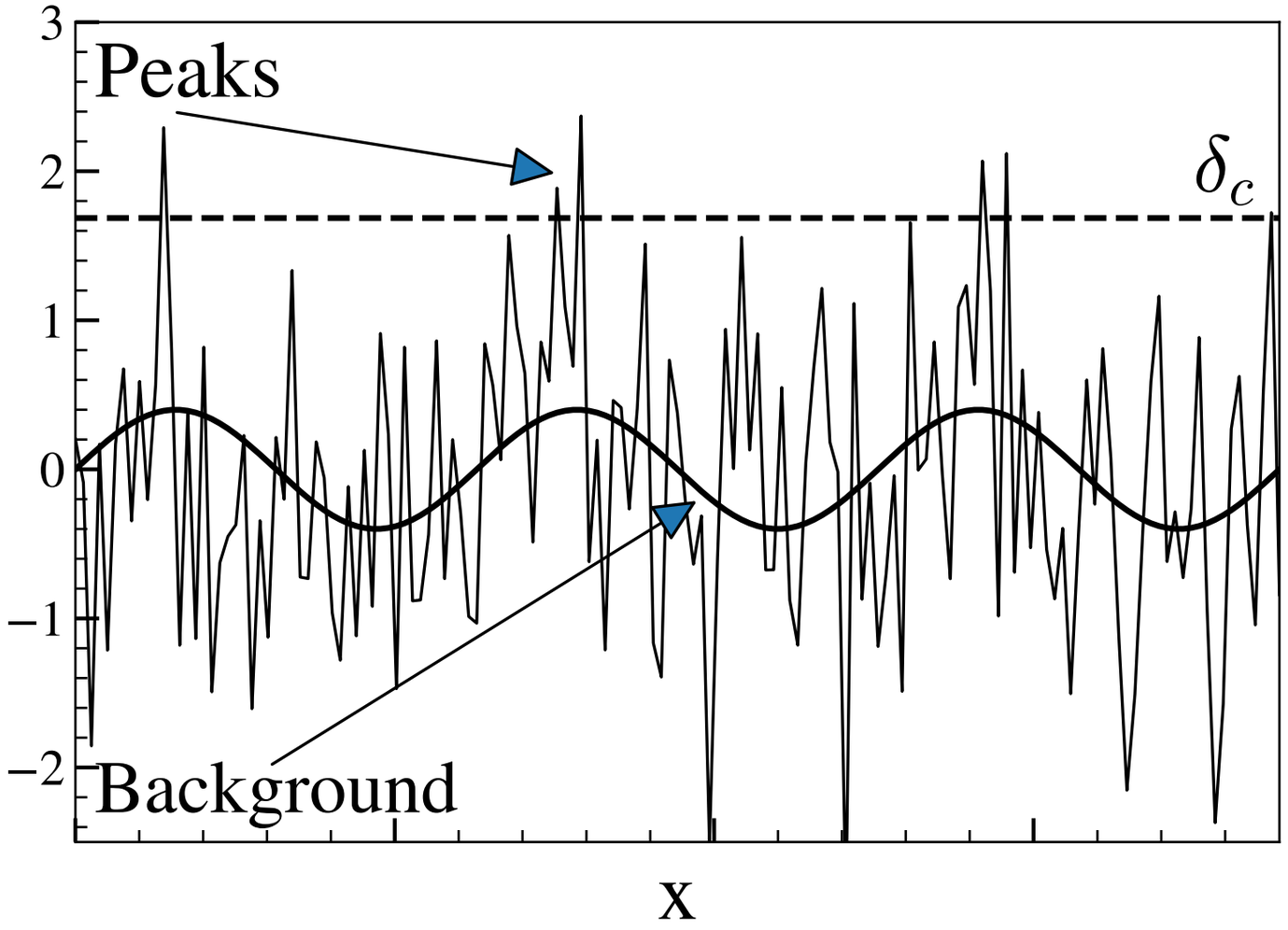
cosmologists measure

usually nuisance parameter(s)

theory predicts

Bias of galaxies (and DM halos)

$$\delta_h(k,z) = b(k,z)\delta_m(k,z) \qquad P_h(k,z) = \boxed{b^2(k,z)} P_m(k,z)$$

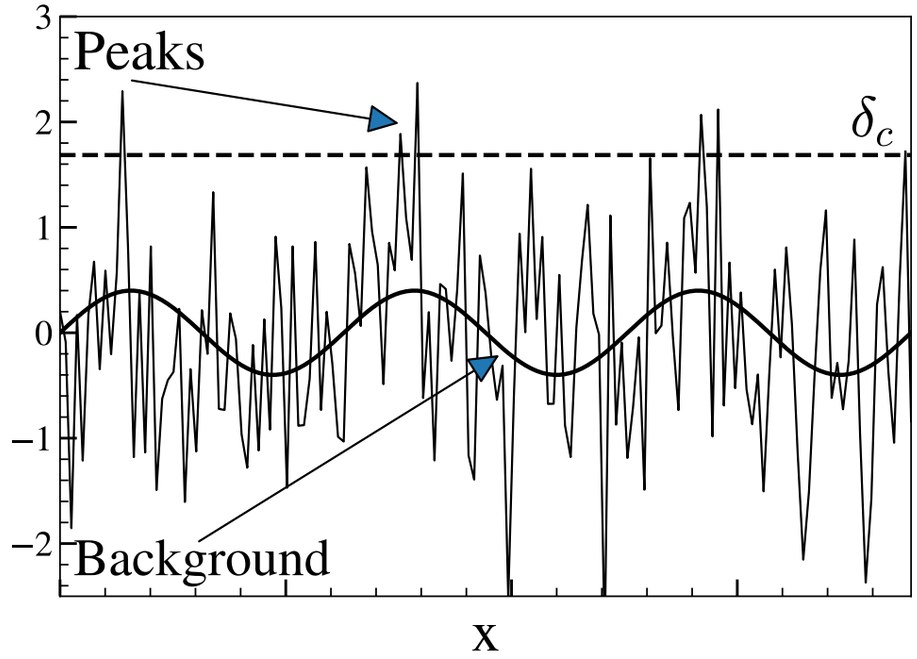


Fact of life: the density peaks simply cluster differently than the background field, even for a Gaussian field ⇒ bias (Bardeen, Bond, Kaiser, Szalay 1987)

Peak-background formula for bias

Imagine some long-wavelength perturbation δ_b (b=background) and short-wavelength one δ_p (p=peaks) ,with

$$\delta = \delta_b + \delta_p$$

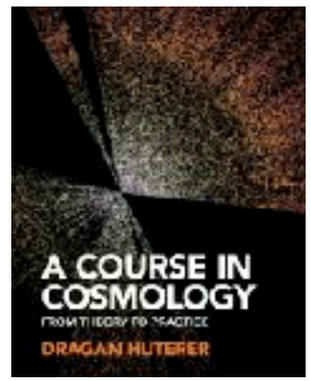


According to the spherical collapse model, to "become a halo" the density fluctuation needs too cross some threshold (e.g. $\delta_{crit} \approx 1.686$)

But because the short perturbation is "riding" on the long perturbation, it only needs to cross the threshold of $\delta_c = \delta_{crit} - \delta_b$.

Now adopt the Press-Schechter mass function

$$n(\nu) \propto \nu \exp(-\nu^2/2) \quad \text{where, recall} \quad \nu \equiv \delta_c/\sigma$$



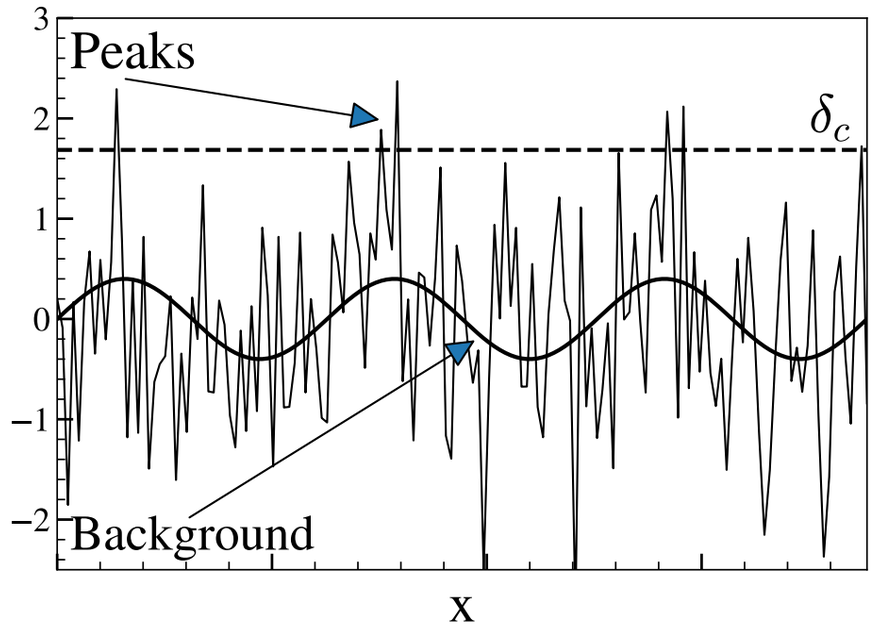
Peak-background formula for bias

Then:

$$n(\nu+\delta\nu) = n\left(\frac{\delta_c-\delta_b}{\sigma}\right) \approx n(\nu) + \frac{dn}{d\nu}\frac{d\nu}{d\delta}(-\delta_b)$$

$$= n(\nu) + \left(\frac{1}{\nu}-\nu\right) n(\nu)\left(-\frac{\delta_b}{\sigma}\right)$$

$$= n(\nu)\left[1+\frac{\nu^2-1}{\nu\sigma}\delta_b\right]$$



Thus $\delta n/n = (\nu^2-1)/(\nu\delta)\ \delta b$.

Finally, switch from Lagrangian bias (above) to Eulerian, $b_E=b_L+1$, to get

$$\boxed{b(M) \simeq 1 + \frac{\nu^2-1}{\delta_c}} \quad \text{(bias from peak} - \text{background split)}$$

Note: higher-peak/mass objects are more biased, as expected.

There are many extensions to the peak-background split theory,
including excursion-set formalism (see Zentner 2006 review).
However, one typically can't safely assume b(M, z) [or b(k, z)] is known -
except arguably at largest scales where b=constant as a function of M/k (but not z!).

Big-picture summary of LSS

Say someone gives you a big box (ok, a file) with 100 million galaxy (and a 100,000 cluster) positions - from either a simulation or real sky. What can you do?

1. Count them!

⇒ fine, but this would only work for clusters, as galaxies are "too complicated" and we can't model their abundance from first principles. Cluster mass fun is dn/dlnM(z).

2. Calculate their clustering, or 2pt function, $\xi(r)$ or $P(k)$!

⇒ super. Can do that via clustering of galaxies, galaxy shears (weak lensing), clusters of galaxies, etc. This has been the workhorse of cosmology since late 1970s!

3. Calculate higher-pt statistics, like $\zeta(r_1, r_2, r_3)$ or $B(k_1, k_2, k_3)$

⇒ interesting, but hard, both to calculate and (especially!) to theoretically predict.

4. Calculate alternative measure: count peaks, measures topology, etc

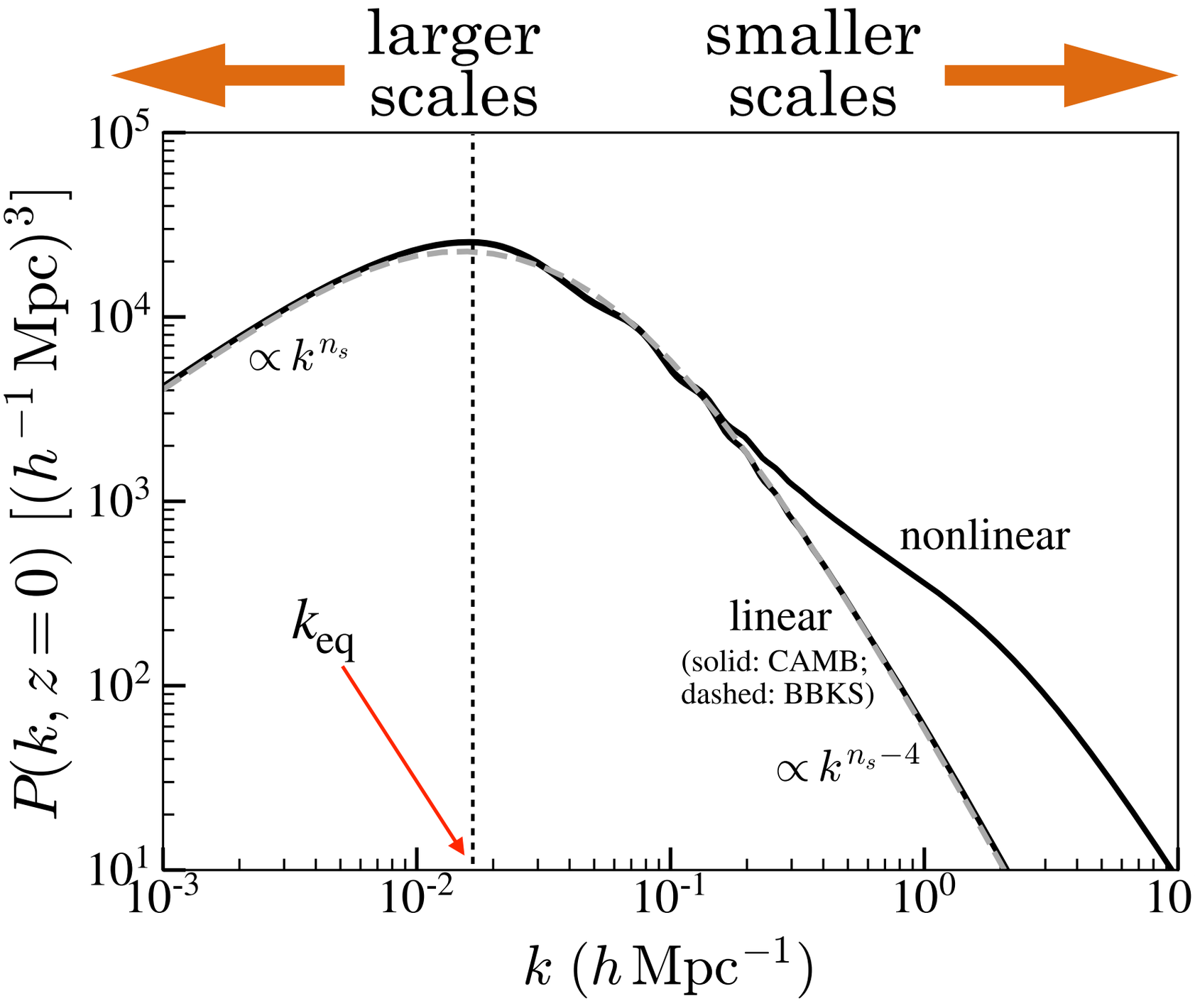
⇒ promising. It does typically contain the same or partial info as 1-pt, 2-pt and higher-pt statistics, but may be easier to calculate/model in practice.

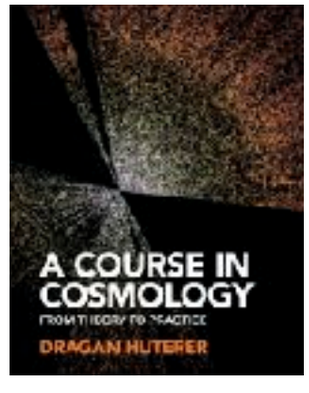
5. Study the internal structure of galaxies, star formation, etc

⇒ but now you are doing astrophysics and not cosmology, my dear friend.

Baryon Acoustic Oscillations (BAO)

Galaxy power spectrum

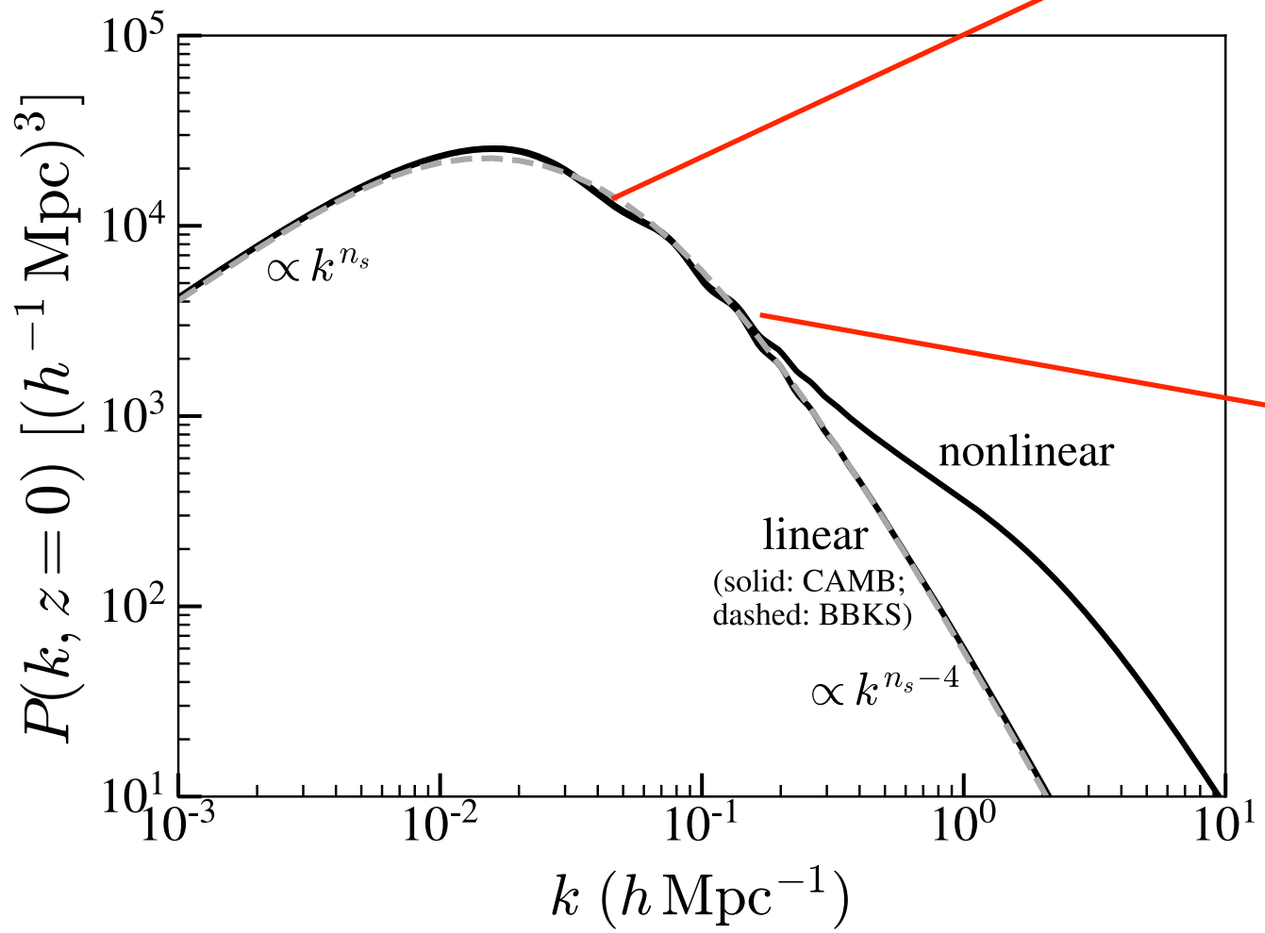


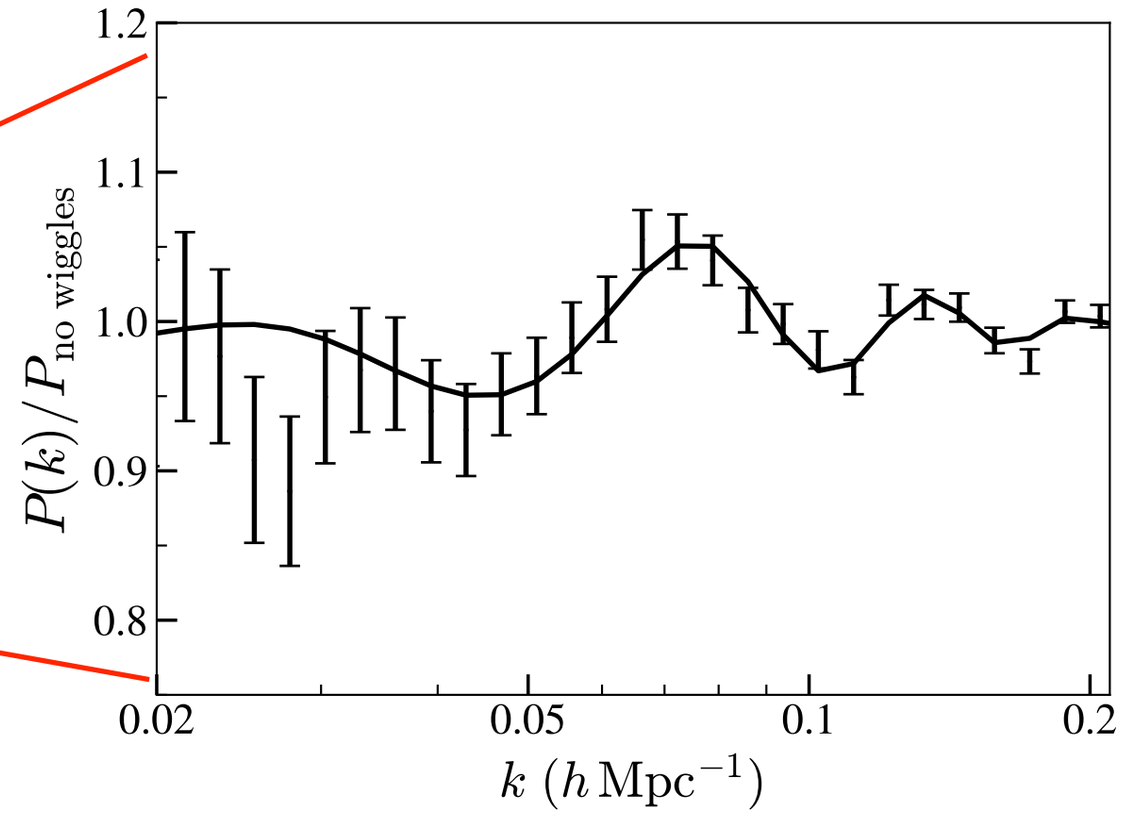


Matter power spectrum contains (almost!) all information at large scales in cosmology

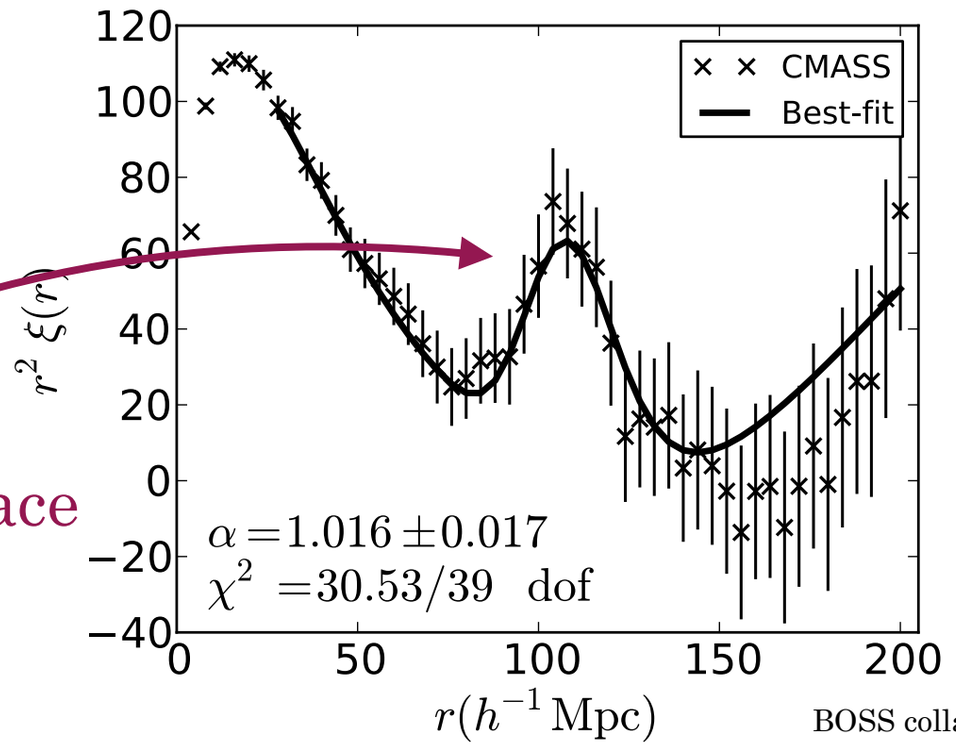
Baryon Acoustic Oscillations (BAO)

Multiple wiggles in Fourier space
(power spectrum)





…or one wiggle in configuration space
(2-point correlation function)

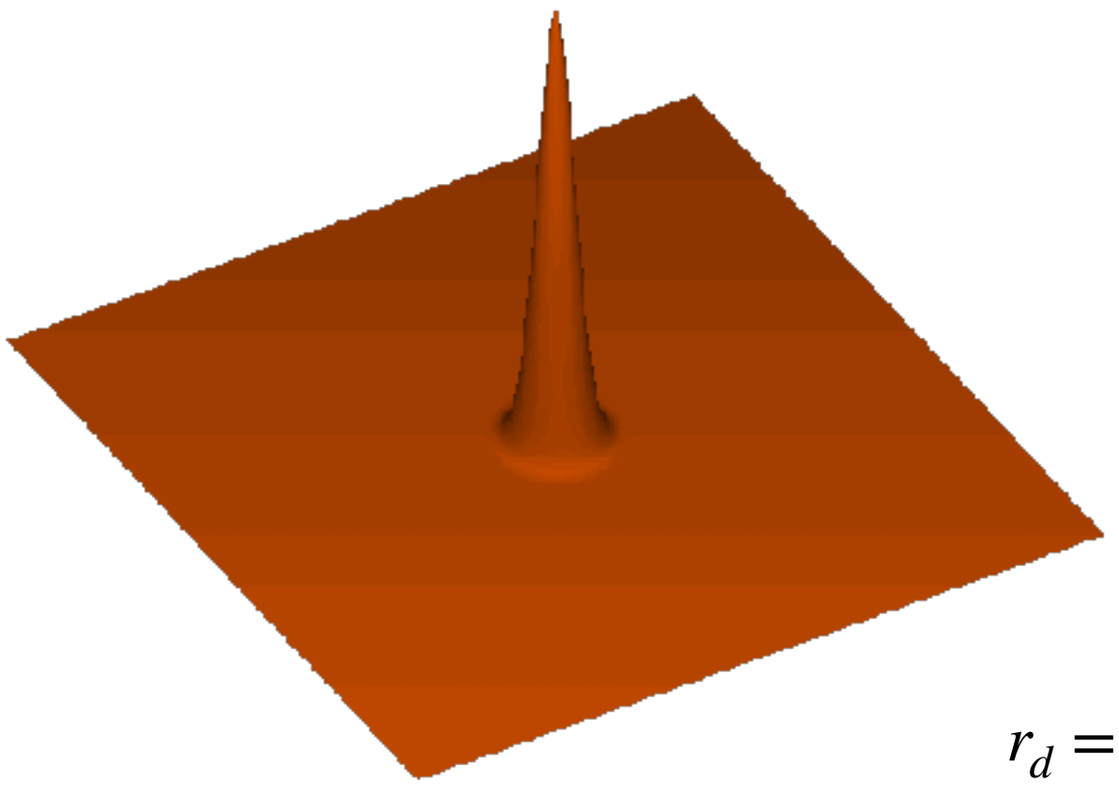


BOSS collaboration

First discussed in: Sunyaev & Zeldovich, 1972

How do the BAO wiggles come about?

At recombination (z_{drag}~1060, t~300,000 yrs)



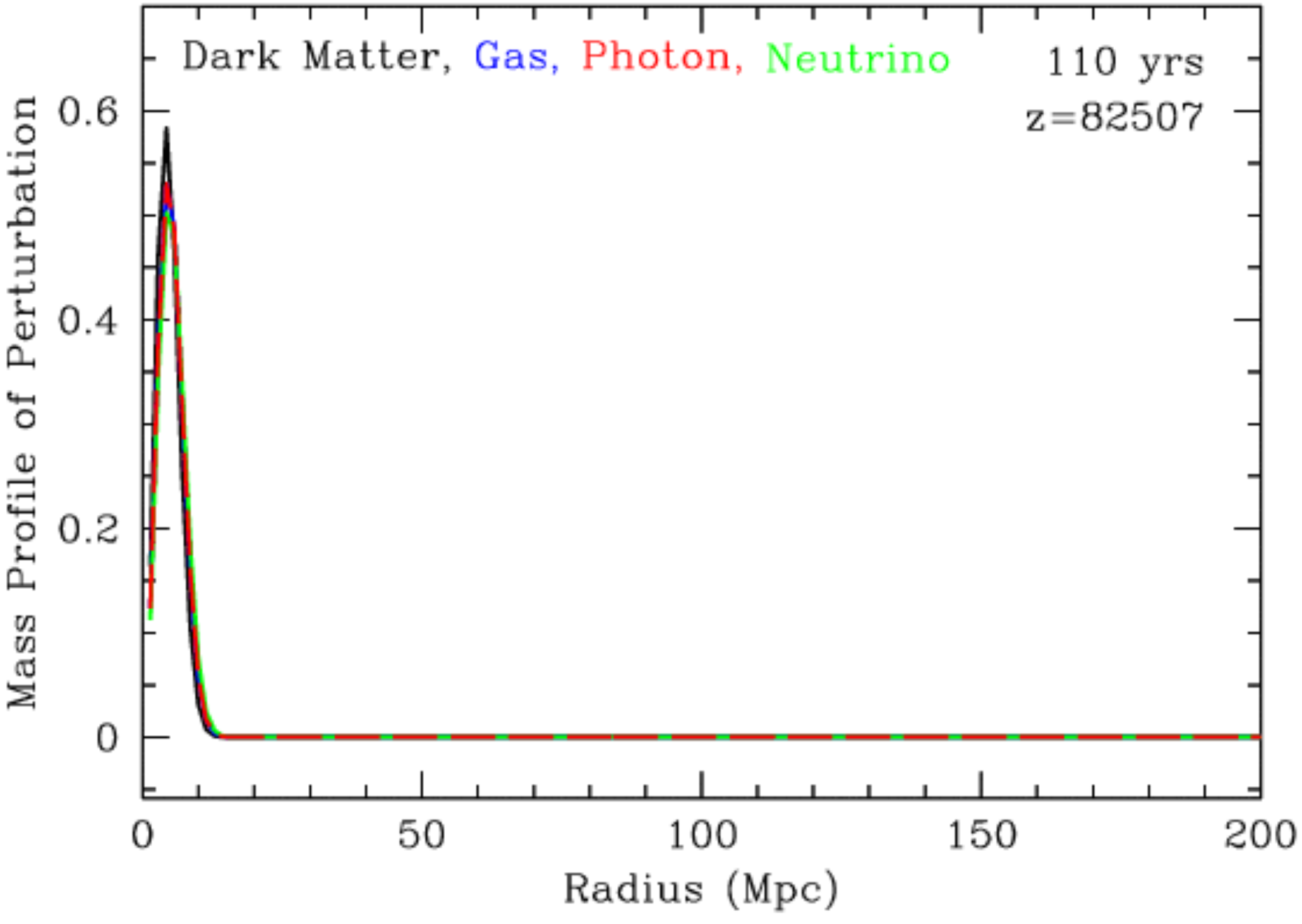
Eisenstein, Seo et al (2007)

- Plasma becomes optically thin
- Baryons decouple from photons
- Sound wave stalls

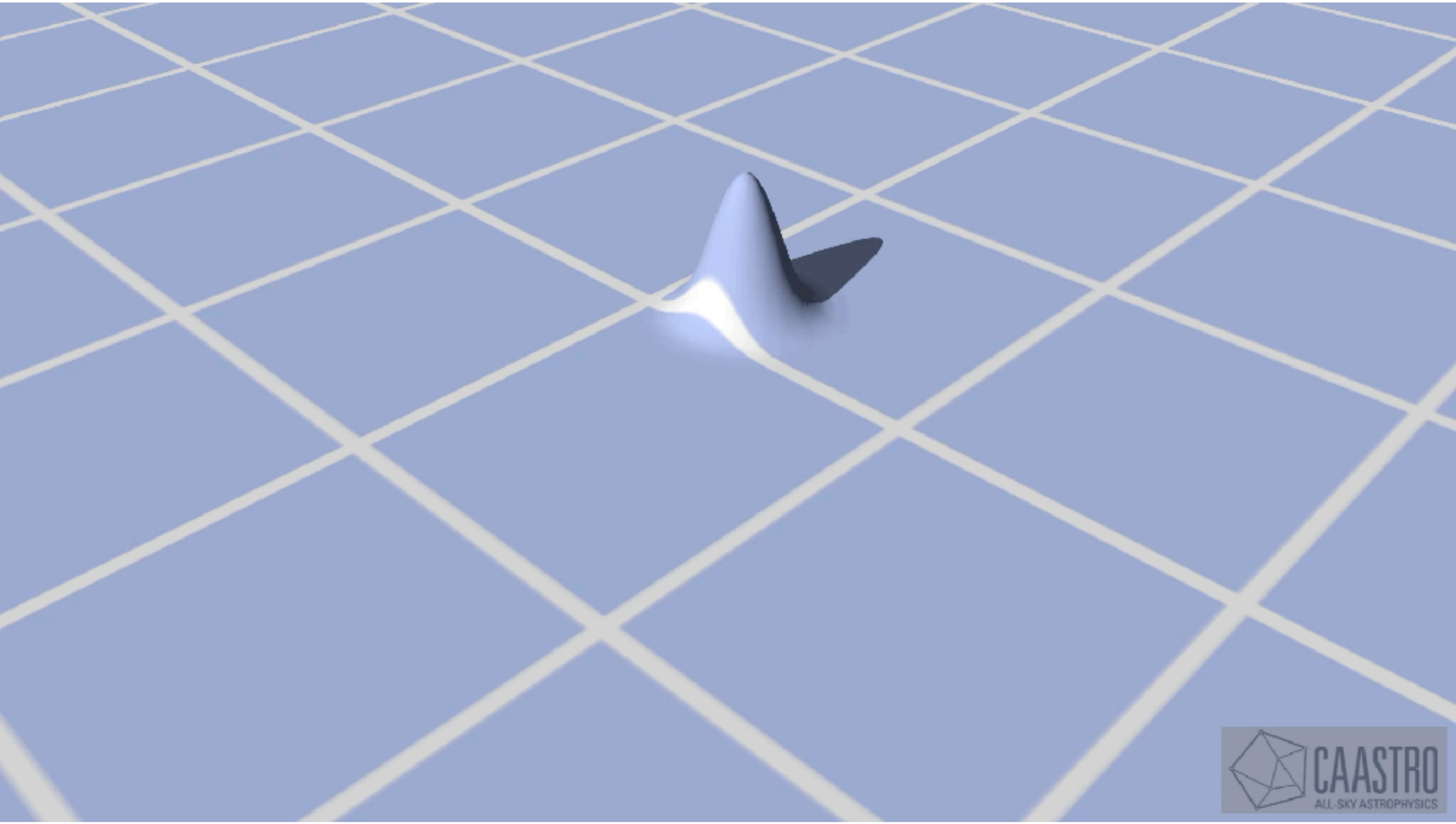
$$r_d = \frac{c}{\sqrt{3}} \int_0^{a_*} \frac{da}{a^2 H(a) \sqrt{1 + \frac{3\Omega_B}{4\Omega_\gamma} a}} \simeq 150\,\mathrm{Mpc}$$

A feature is imprinted the distance that the wave has traveled between the Big Bang and recombination

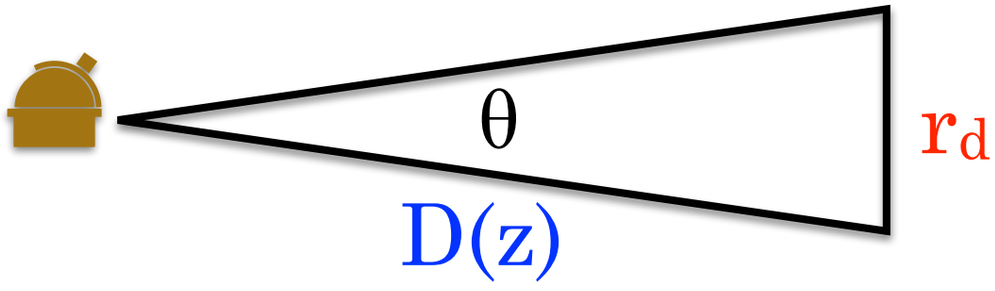
$\Longrightarrow$ **the sound horizon distance at recombination ($r_d \simeq 150\,\mathrm{Mpc}$)**



D. Eisenstein

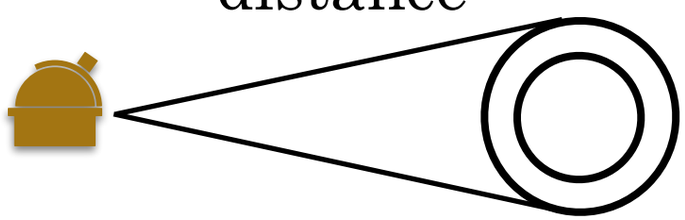

CAASTRO
ALL-SKY ASTROPHYSICS

Baryon Acoustic Oscillations

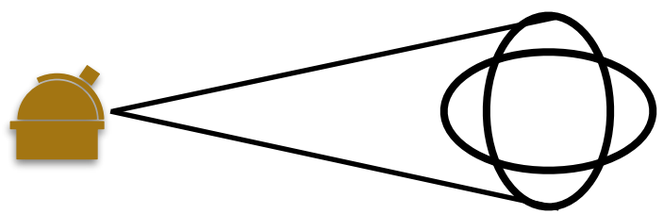


- Therefore, there is excess probability for galaxies having a neighbor at distance r_d — <u>excess probability for clustering</u>
- This imprints a preferred scale in clustering - the “standard ruler"
- The angle to the standard ruler gives $D(z)/r_d$
- Actually measure *two* kinds of distances: transverse or parallel to the line-of-sight; can be expressed as

Isotropic (“average”) distance



Ratio of transverse and line-of-sight distances

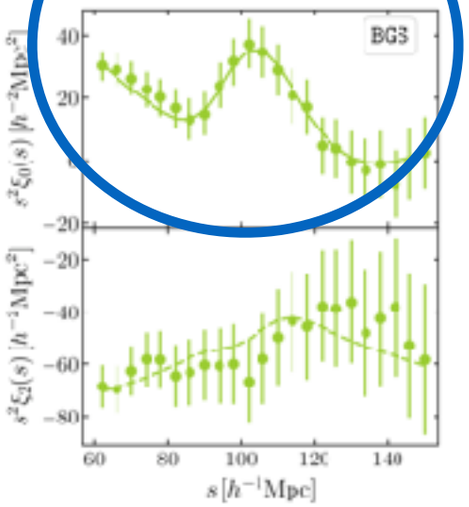


DESI DR2 Clustering Measurements

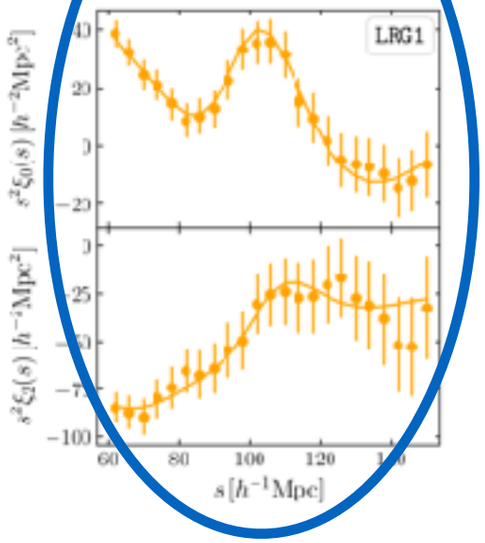
monopole

quadrupole

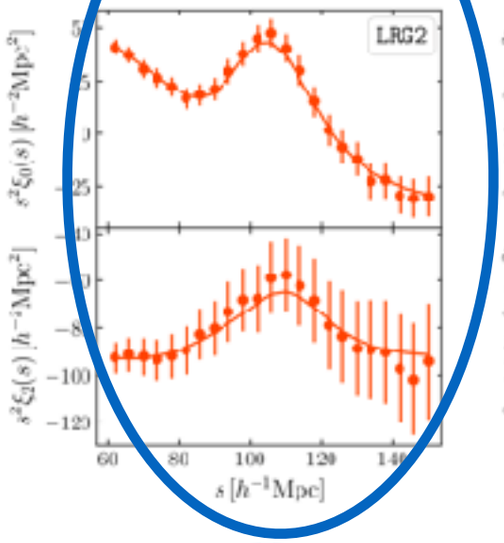
BGS



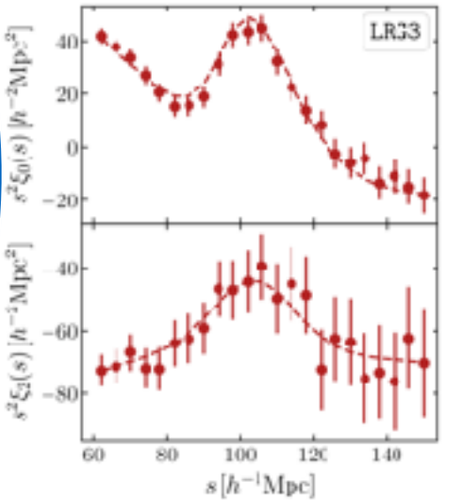
LRG1



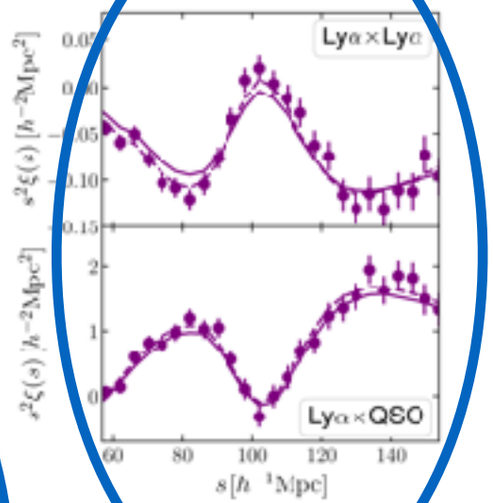
LRG2



LRG3

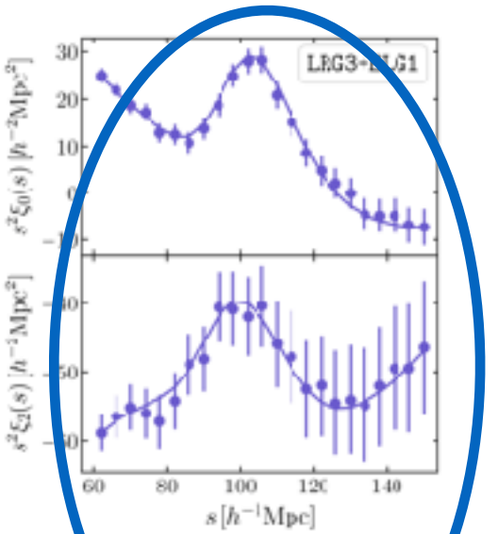


Lya × Lya

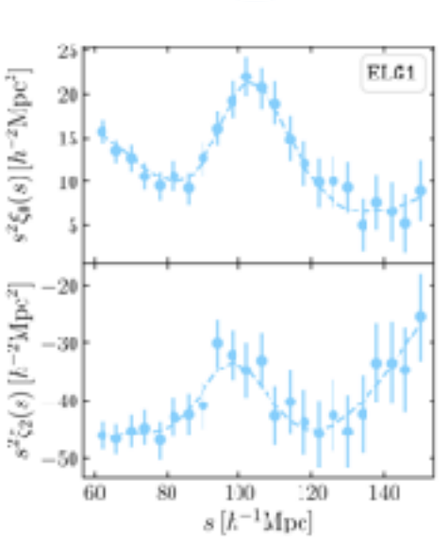


Lya × QSO

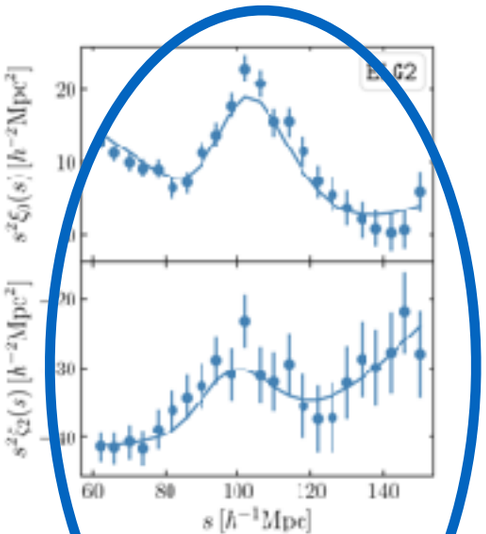
LRG3 × ELG1



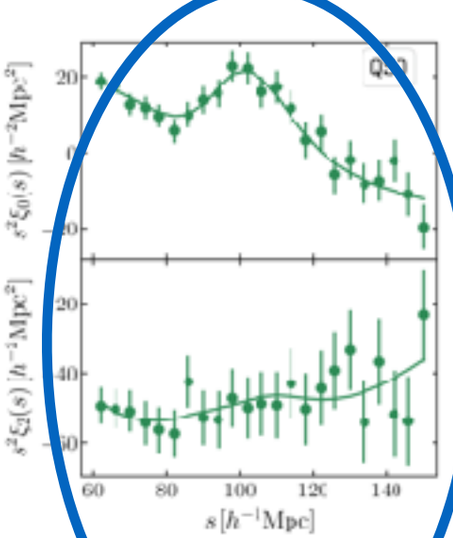
ELG1



ELG2



QSO

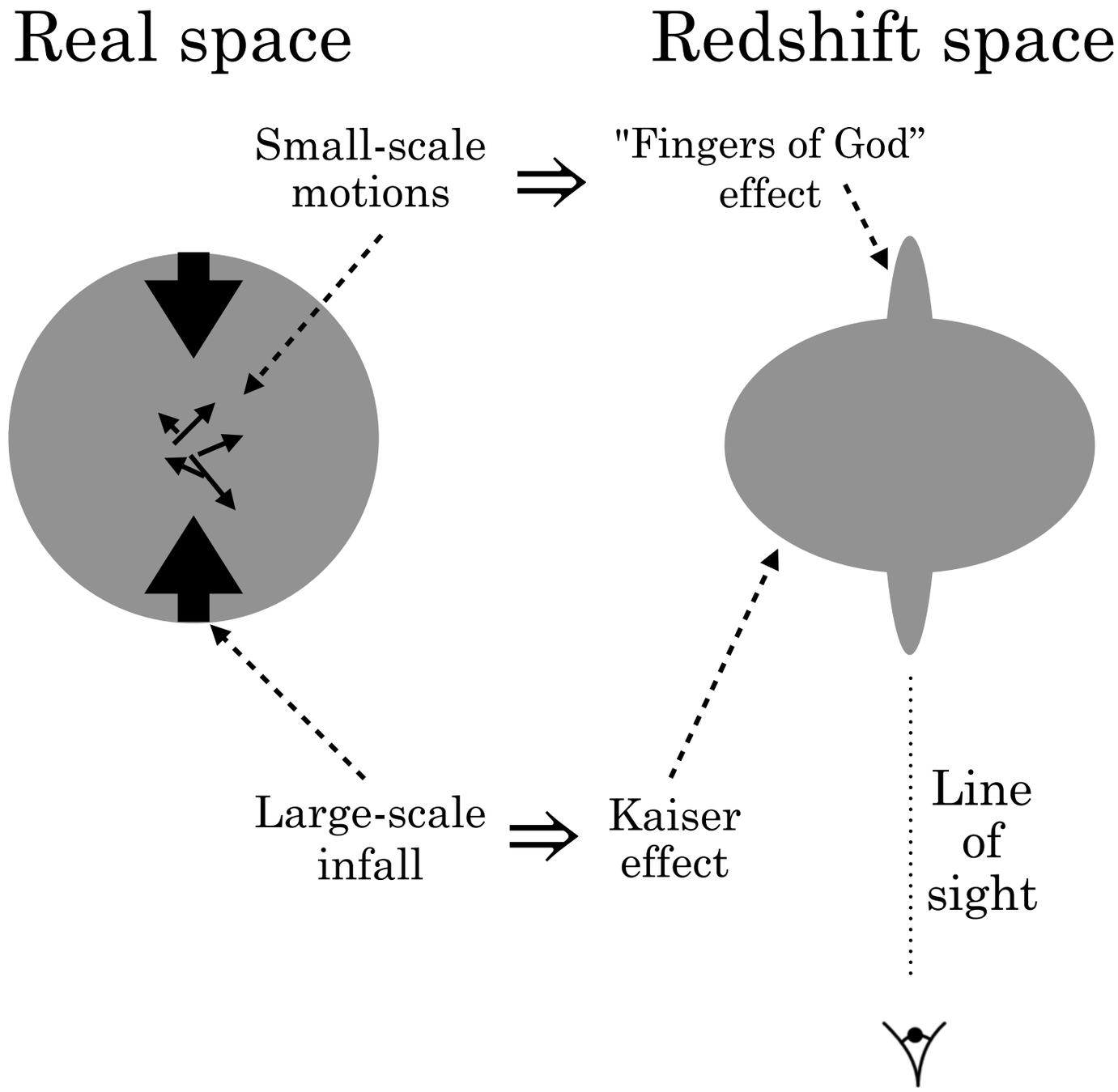


Used in DR2 analysis

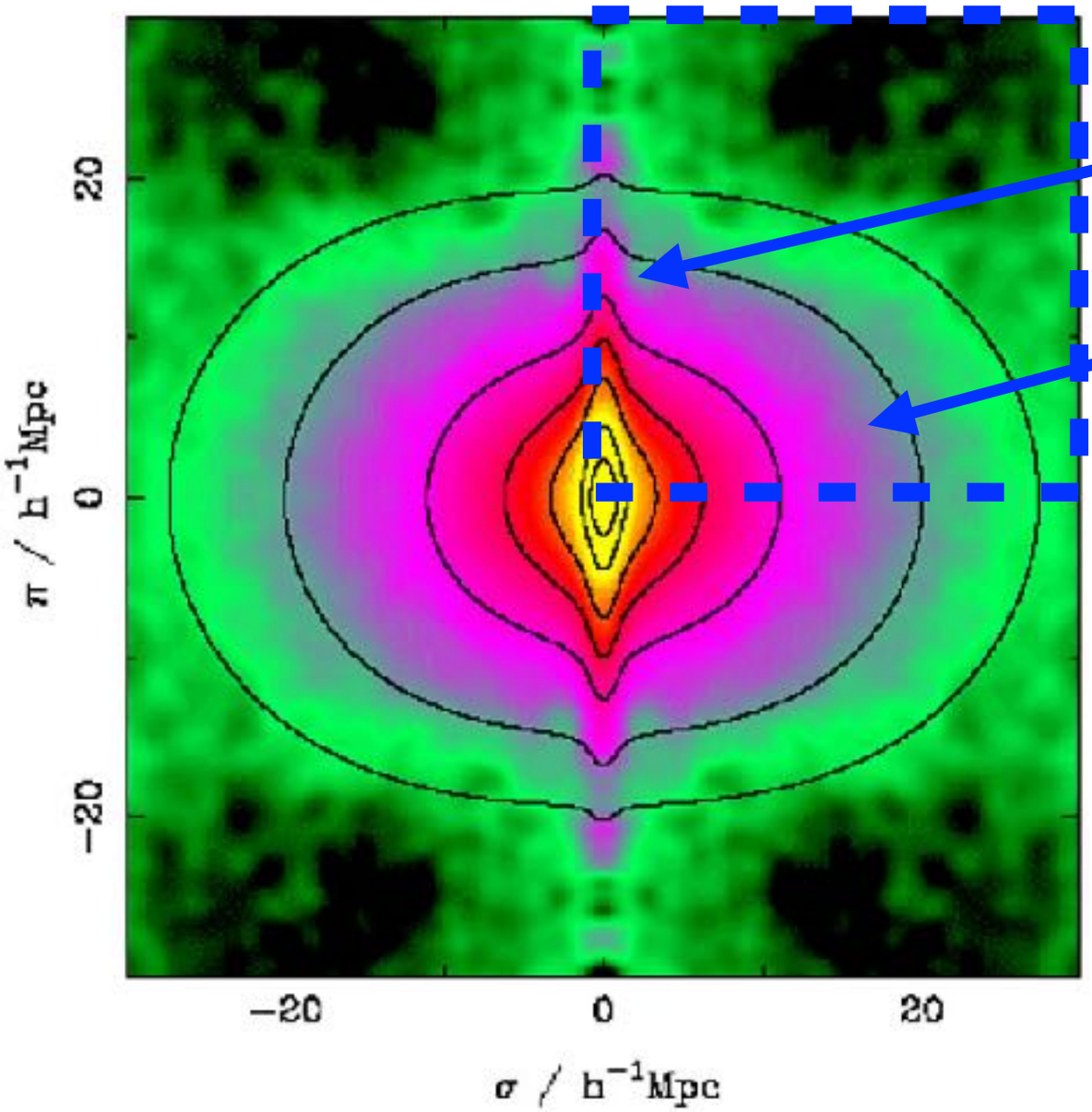
Redshift Space Distortions (RSD)

Redshift-space distortions

If galaxy positions are measured in redshift space (i.e. in a spectroscopic survey), then the correlation function takes on two new effects: “Fingers of God” (stupid name I know!), and “Kaiser effect”.



Redshift-space distortions



2dF data; Peacock et al (2001)

Kaiser (1987) formula:

$$P(\mathbf{k})^{(s)} = b^2 \left[1 + \beta\mu^2\right]^2 P(k)$$

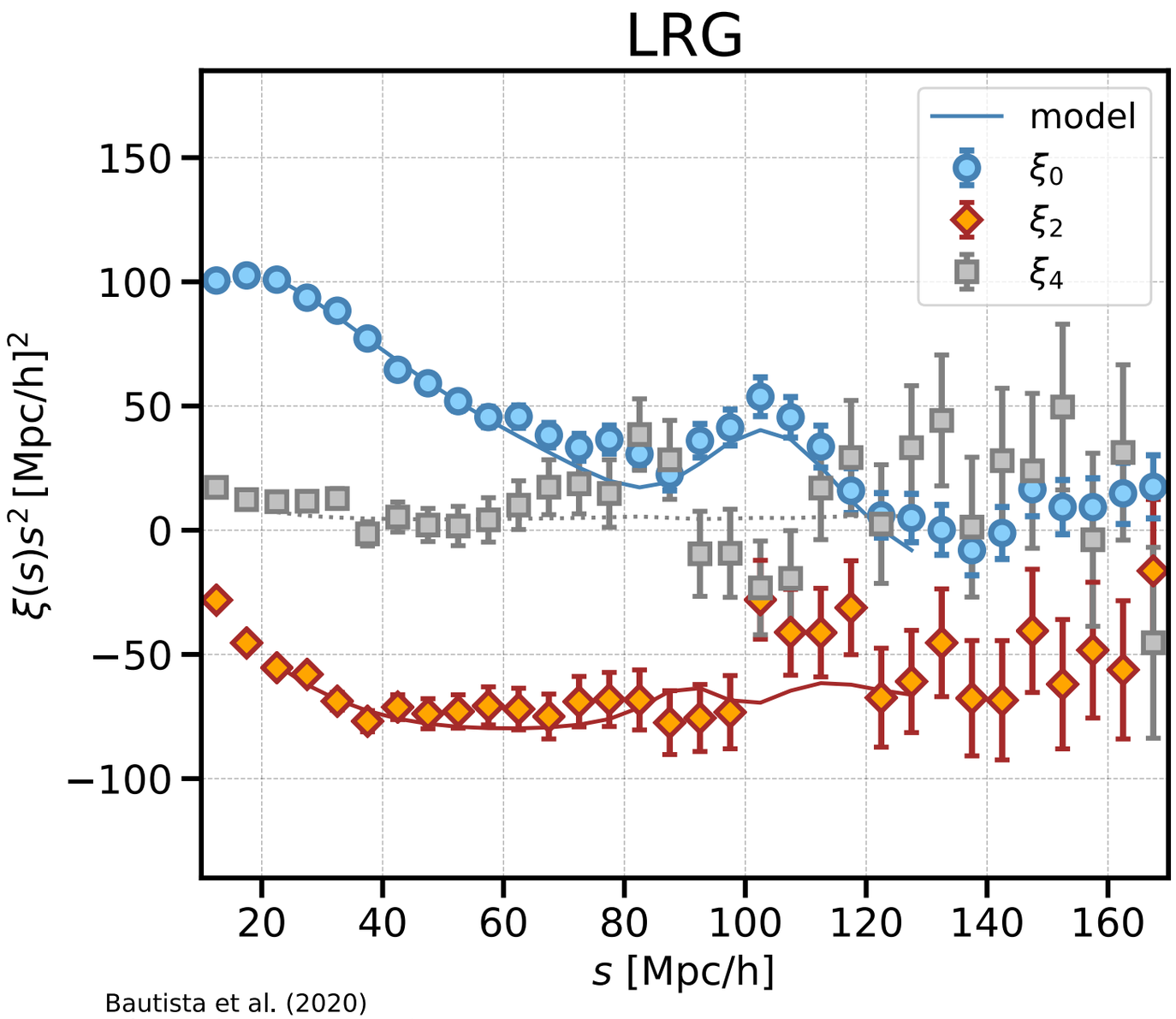
where $\beta \equiv \frac{f}{b}$; $\mu \equiv \hat{\mathbf{k}} \cdot \hat{r}_z = k_z/k$

growth rate!

Redshift Space Distortions

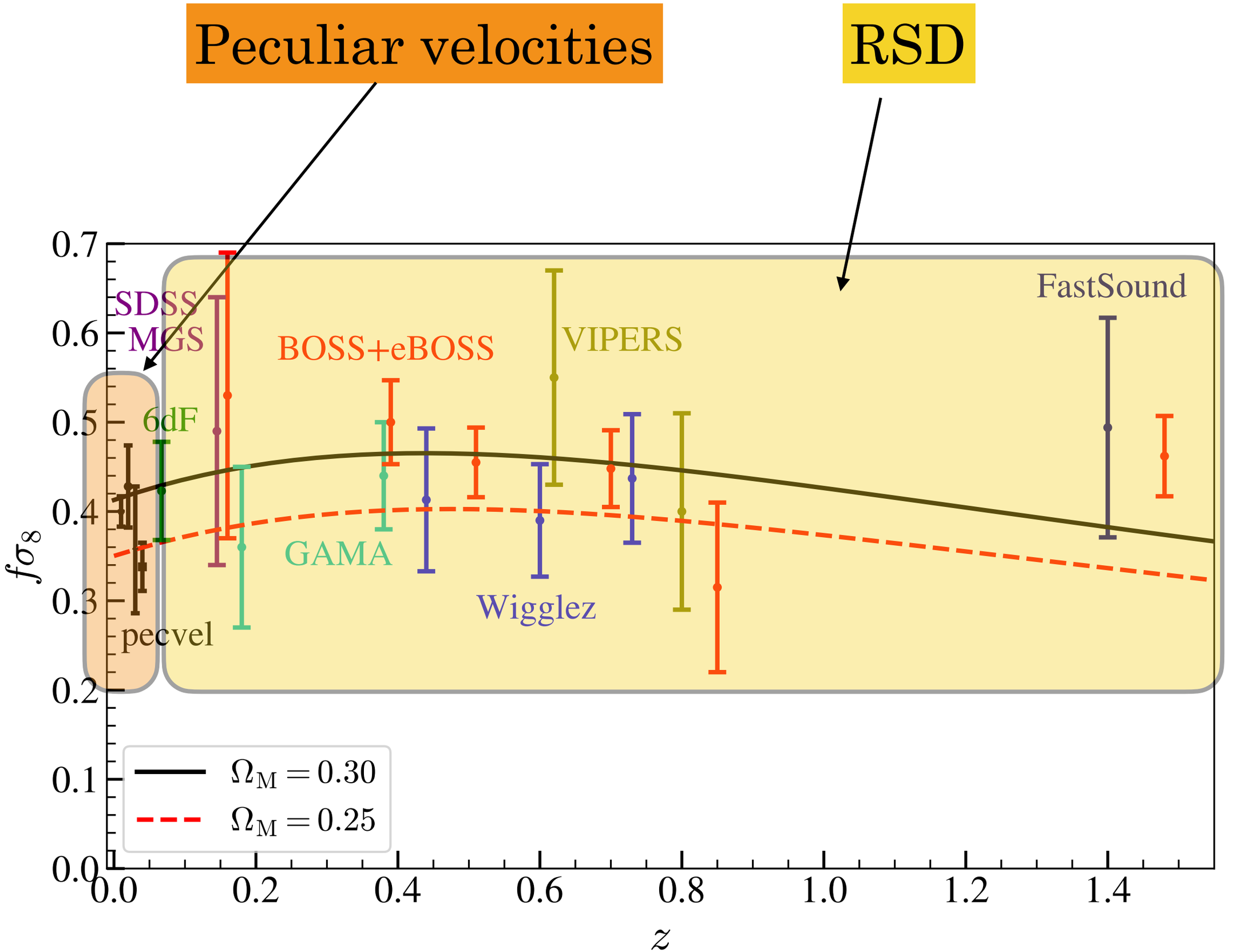
Kaiser RSD formula (roughly ok at large scales):

$$P(\mathbf{k})^{(s)} = b^2 \left[1 + \beta\mu^2\right]^2 P(k)$$



Bautista et al. (2020)
(plot by Jiamin Hou, MPE)

μ^0, μ^2, μ^4 terms become the monopole (l=0), quadrupole (l=2) and hexadecapole (l=4), respectively



Huterer, Astron. Astrophys. Rev. (2023)