The Inhomogeneous Universe

Dark matter non-linear clustering

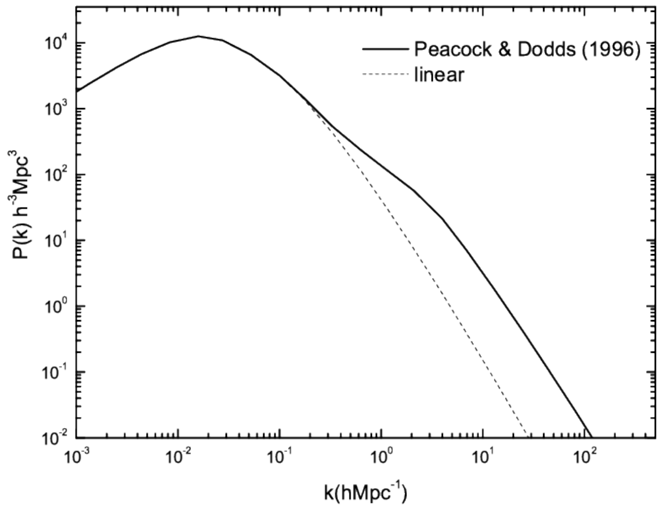
When the density contrast reaches δ ~1, we can no longer linearize the Einstein and the conservation equations.

There are three main approaches to study the **non-linear clustering** (the collapse into gravitationally bound strutures):

- **Perturbation theory**

- **N-body simulations**

- **The halo model**



Here we will just give a short overview of these methods.

But before thurning to these items, we can have a look in more detail into the collapsing process of a halo, **using zero-order equations,** in a **mini-universe approach :**

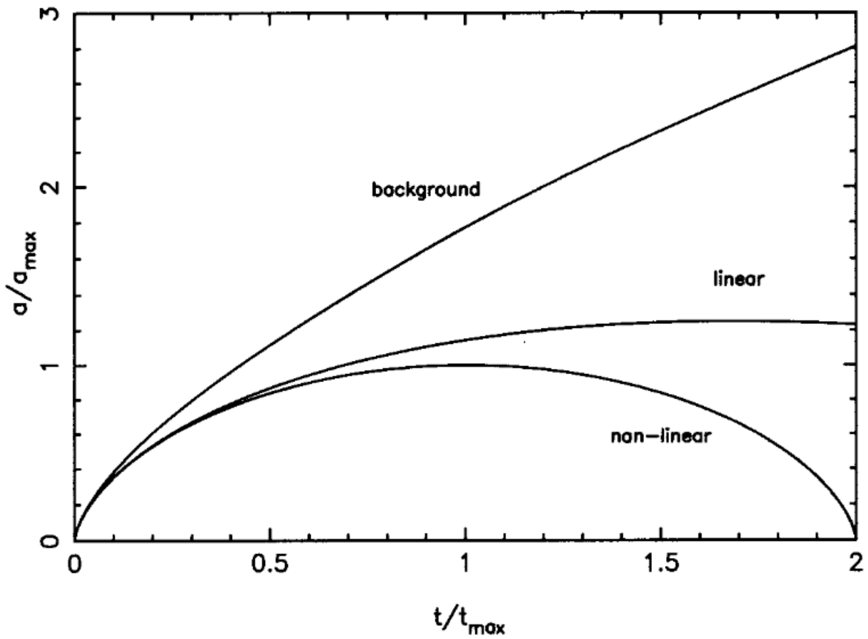
this is called the **spherical collapse** model

Spherical collapse

Let us consider a spherical halo within the expanding background.

The halo expands with a slower rate than the expanding Universe, meaning it is clustering with respect to the comoving flow.

Extending this process to the non-linear collapse (and not only in the linear evolution), **the expansion rate will start decreasing reaching zero at a time t_{max} and then contracts.**



So the halo mini-universe needs to be a closed universe, with curvature, in order to be able to expand and then contract.

Its solution a(t) is found from the Friedmann equation, it grows from 0 to a_{max} (the turnaround point) and contracts to 0.

Its **scale factor** a(t) corresponds to the **size** of the halo r(t) of mass M.

The **solution** a(t) for the closed mini-universe cannot be written explicitly but only as a parametric solution:

$$\frac{a(t)}{a_{\max}} = \frac{1}{2}(1 - \cos\theta), \qquad \frac{t}{t_{\max}} = \frac{1}{\pi}(\theta - \sin\theta)$$

where θ goes from 0 to π (when a = a_{max}) to 2π (when a=0). The process is symmetric, i.e., the maximum expansion occurs at $t=t_{max}$ and the full collapse ends at $t=2t_{max}$.

In the beginning of the process (for small θ) we can insert the parametric functions and **expand** to low order in θ, to find an explicit solution a(t):

$$\frac{a_{\rm lin}(t)}{a_{\rm max}} \simeq \frac{1}{4}\left(6\pi\frac{t}{t_{\rm max}}\right)^{2/3}\left[1-\frac{1}{20}\left(6\pi\frac{t}{t_{\rm max}}\right)^{2/3}\right]$$

Notice the first term, with $a(t) \sim t^{2/3}$ is the expansion in a matter-dominated flat Universe, while the square bracket term gives the correction from the presence of curvature.

The density of the halo (the mini-universe) decreases as its volume a_{lin}^3 expands, and the density of the universe decreases as tis volume a^3 expands.

The **linear density contrast** of the halo with respect to the mean density of the universe is the ratio:

$$1+\delta_{\rm lin} = \frac{a_{\rm back}^3}{a_{\rm lin}^3}$$

From the $a_{lin}(t)$ solution and a(t), we get:

$$\delta_{\text{lin}} = \frac{3}{20}\left(6\pi \frac{t}{t_{\text{max}}}\right)^{2/3}$$ (using $(1 - A)^{-3} \sim 1 + 3A$)

At **turnaround** ($t=t_{max}$), the linear density contrast is then:

$$\delta_{\text{lin}}^{\text{turn}} = \frac{3}{20}(6\pi)^{2/3} \doteq 1.06$$

Notice that the approximate expression of a(t) for small θ is not valid at the turnaround point where θ = π.

So $a_{lin}(t_{max})$ is not the actual value of the scale factor of the mini-universe. It is the value it would have if the linear regime was still valid (this is why it was called the linear density contrast).

The true value of the scale factor of the mini-universe at turnaround is of course $a = a_{max}$,

and the true **non-linear density contrast** is simply:

$$1 + \delta_{\rm nonlin}^{\rm turn} = \frac{a_{\rm back}^3}{a_{\rm max}^3} = \frac{(6\pi)^2}{4^3} = 5.55$$

showing the density contrast is already quite large at the turnaround point.

After contracting, at the **end of the collapse**,
i.e., at $\theta = 2\pi$ (or $t=2t_{max}$), the linear density contrast is

$$\delta_{\rm lin}^{\rm coll} = \frac{3}{20}(12\pi)^{2/3} = 1.686$$

and this is the value chosen for the threshold that defines collapsed regions in the Press-Schechter theory *(to be seen later)*

Again, the true (non-linear) density contrast is much larger than this.

In fact, if the collapse goes to zero, it would even be infinite.

In reality the collapsing dark matter particles in the halo deviate from exact radial trajectories, dissipative physics convert their kinetic energy into random motions → the random motions make the collapse to relax leading to an equilibrium state: **virialization** instead of complete collapse.

It is usually assumed (supported by N-body simulations) that during the contraction phase (from $t=t_{max}$ to $t=2t_{max}$), the halo virializes →

stabilizing at a size equal to its maximum size.

So its density at virialization is $\sim (a_{max} / 2)^3$, i.e.,

halo density increases a factor of 8 since the turnaround (and not a factor of infinite as if there would be total collapse)

On the other hand at that time ($t=2t_{max}$), the background scale factor (that grows with $t^{2/3}$) increased by a factor of $2^{2/3}=1.6$

background density decreases by a factor of $1.6^3 = 4$.

So the final true (non-linear) density contrast increases by a factor of 32 from the value of 5.55 at turnaround, i.e.,

$$1 + \delta_{\rm nonlin}^{\rm vir} \simeq 178$$

Even though the details of the non-linear clustering cannot be studied in this approach, and the non-linear power spectrum cannot be computed, the spherical collapse method allows us to compute typical values of the overdensities:

δ ~ 5, at the decoupling from the comoving flow

δ - 178, for virialized structures

The spherical collapse method also provides a **threshold** value to recognize fully collapsed (virialized) halos, based only on linear density fields.

This means that if we extrapolate calculations in the linear regime and find a value of $\delta \sim 1.69$, we know we are dealing with a collapsed object.

Note that the derivation assume the collapse occurs at z=0.
For higher redshifts, the value is lower:

$$\delta_c^{lin} \sim 1.69 / (1+z)$$

The derivation can also be made for other models. The value of the typical virialized overdensity of dark matter in the presence of different dark energy models will be different.

Perturbation theory

When the density contrast reaches δ ~1, we can no longer linearize the equations.

The evolution equation up to second-order may be written as,

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \frac{1}{a^2}\nabla\,.\,(1+\delta)\nabla\Phi + \frac{1}{a^2}\sum_{i,j}\frac{\partial^2}{\partial x^i\partial x^j}[(1+\delta)v^i\,v^j]$$

The multiplications of perturbations become convolutions in Fourier space, and so the spatial dependence is no longer local (i.e., each mode can no longer be treated as independent), but there are mode couplings → the evolution equation in Fourier space to second-order becomes,

$$\frac{d^2\delta_k}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_k}{dt} = 4\pi\,G\bar{\rho}\delta_k + A - B$$

with

$$A = 2\pi\, G\bar{\rho} \sum_{k'} \left[\frac{k.k'}{k'^2} + \frac{k.(k-k')}{|k-k'|^2} \right] \delta_k\, \delta_{k'}$$

the evolution of a scale depends on all the other scales

$$B = \int (1+\delta) \left(\frac{k.v}{a} \right)^2 \mathrm{e}^{-ik.x}\, d^3x$$

Perturbation theory assumes that it is possible to expand the density (and velocity) fields around the linear solution:

$$\delta(x,t) = \sum_n \delta^{(n)}(x,t)$$

which defines the density contrast in several orders, n=0 (homogeneous), n=1 (linear), n=2 (quadratic) ,…, up to infinity.

$$\delta^{(0)} = 0$$

$\delta^{(1)}$ is the linear solution

The equation of motion is solved recursively:

Inserting the linear solution $\delta^{(1)}(a)$ in the non-linear terms (A, B), provides the equation of motion for $\delta^{(2)}$.

Its solution $\delta^{(2)}(a)$ can then be inserted in A, B, to get the equation for $\delta^{(3)}$,

and so on ….

Summing all orders up to n provides the result δ(a) up to order n in perturbation theory:

$$\delta(x,t) = \sum_n \delta^{(n)}(x,t)$$

Matter-dominated epoch (a single component: Einstein - de Sitter)

The linear solution is: $\delta^{(1)}(a) \sim a$

Applying the recursion, **the growing solutions for the different orders are found to be proportional to a^n** → growth becomes very fast in the non-linear regime.

The solution for each order is:

$$\delta^{(n)}(k,a) = a^n \int d^3\mathbf{q}_1 \cdots \int d^3\mathbf{q}_n \, \delta_D(\mathbf{k} - \mathbf{q}_{1\cdots n}) F_n(\mathbf{q}_1, \ldots, \mathbf{q}_n) \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n)$$

The growth of a scale (in a given order) is a power of "a" times a factor that depends on all scales.

The **kernel F_n** encapsulates the mode couplings - it is an n-point quantity, i.e., for each order, F_n is a coupling between n scales.

For example, for n=2, F_2 couples pairs of scales:

$$F_2 = \frac{5}{7} + \frac{1}{2}\frac{k_1 . k_2}{k_1^2} + \frac{1}{2}\frac{k_1 . k_2}{k_2^2} + \frac{2}{7}\frac{(k_1 . k_2)^2}{k_1^2 \, k_2^2}$$

and so $\delta^{(2)}$ is obtained by integrating over the variables k_1 and k_2

However, perturbation theory also fails quite soon.

Indeed, once collapsed bound structures form, there will be multiple streams of matter, i.e., matter no longer defines a unique flow in a perturbed region, and the idea of a global cosmological fluid breaks down, and the expansion in orders of perturbations, with solutions found recursively is no longer a good description.

A better way to compute the evolution of the density field is numerically, with N-body simulations.

N-body simulations

The general approach to simulate the space and time evolution of the dark matter density field is

to discretize the system in a set of N particles, of mass M (they are not microscopic particles),

within an evolving comoving volume of at least side L = 200 Mpc/h (to be able to contain large-scale structures),

and introduce periodic boundary conditions (particles leaving the cube, re-enter from another side) in order to simulate the force from particles outside the cube.

The DM particles are colisionless, they do not interact with each other.

They evolve because they feel the gravitational field.

Initial conditions

The first step is to set the **initial conditions:**

Place the particles in spatial initial positions (at the starting redshift of the simulation), such that the density (computed from that configuration) is a realization of the initial power spectrum.

Iterative evolution

Then the system can start to **evolve iteratively** (in time steps):

First compute the **potential** for that discrete configuration (Poisson equation):

$$\Phi(\vec{x}) = -G \sum_{i=1}^{N} \frac{m_i}{[(\vec{x}-\vec{x}_i)^2 + \epsilon^2]^{1/2}}$$

Note that the potential may be modified at small separation (by a softening length ε - given by a fraction of the mean separation between two particles) to prevent spurious collisions between the macroscopic particles.

Then compute the **acceleration** on each particle from Euler equation:

$$\ddot{\vec{x}} = -\vec{\nabla}\Phi(\vec{x})$$

Additional sources of acceleration may be included, like gradient of pressure, or the effect of a dark energy scalar field. In that case, we need an additional equation of motion for the field (Klein-Gordon equation).

Then repeat for the next time step.

Grid

This standard process is slow → order N^2 computation.

To be faster, particles may be put on a **grid**.
From the initial conditions, the mass of each particle is distributed among the nearest cubic grid points.

The acceleration field of this grid mass distribution is then computed with Fast Fourier Transform:

FT of density → leads to the potential Φ_k through Poisson equation and acceleration $g_k = ik\,\Phi_k$ → acceleration field transformed back to the real space and used in the acceleration equation.

There are many different grid techniques,

one example is the P^3M (particle-particle particle-mesh) where

particle-mesh (grid) interactions are used on larger separations and

particle-particle interactions are used on smaller separations for better spatial resolution (need to subtract these contributions from the Fourier calculation).

Volume and resolution

The simulations are usually a trade-off between **volume and resolution** (i.e., size of the simulations vs. mass of the particles).

For example, two versions of the Millenium simulations contain 10^{10} particles in two different configurations:

Millennium

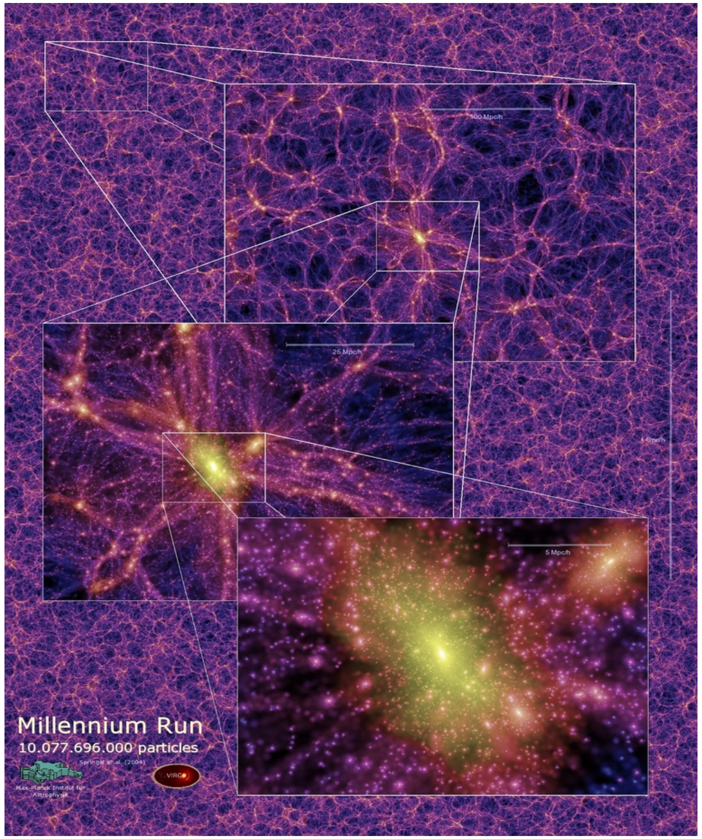
Larger volume:
z=18, L=500 Mpc /h

Lower resoution:
M=9 x 10^8 M_{Sun}

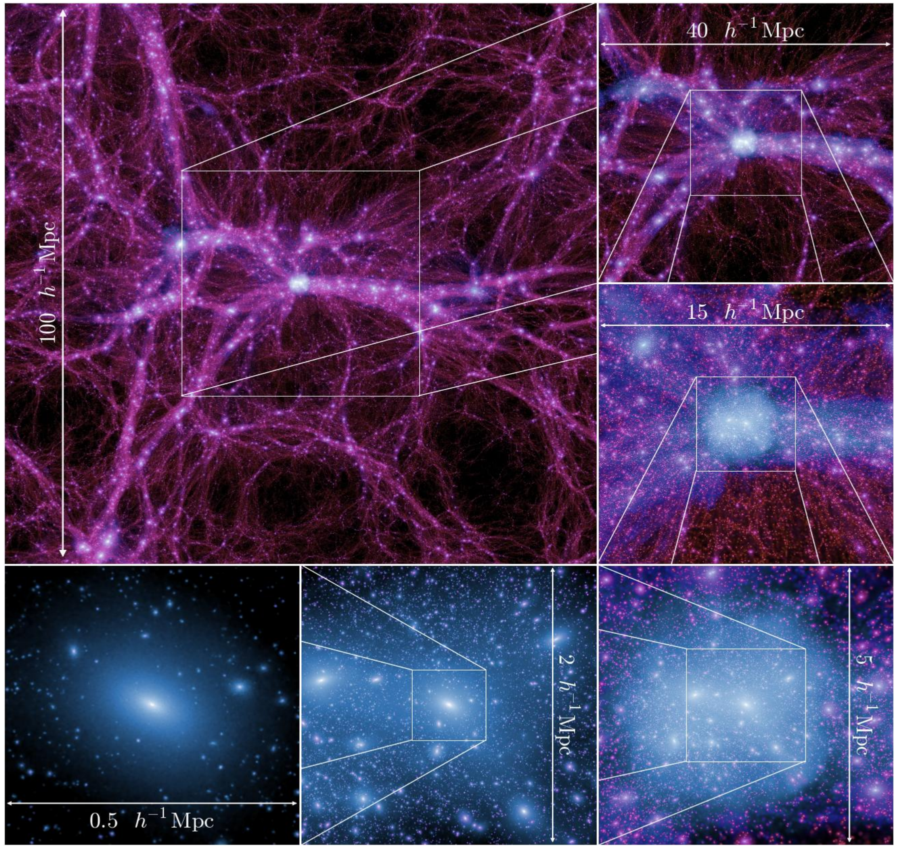
Millennium II

Smaller volume:
z=6, L=100 Mpc /h

Higher resoution:
M=7 x 10^6 M_{Sun}



sides of inner images are (Mpc/h): 500, 100, 25, 5



sides of inner images are (Mpc/h): 100, 40, 15, 5, 2, 0.5

Results

The **power spectrum of the resulting density field** can be computed from the resulting configuration.

It is naturally the **non-linear power spectrum**.

There are many other important aspects of post-processing or different types of simulations, such as:

light-cone projections (**ray-tracing**) to obtain angular 2D maps - more relevant to compare with observations

hydrodynamic simulations: including baryons (gas) with radiative processes, described by hydrodynamic equations (continuity, Euler, first law of thermodynamics)

semi-analytical models and **halo occupation distribution**: to populate dark matter halo merger trees with galaxies

Fitting functions

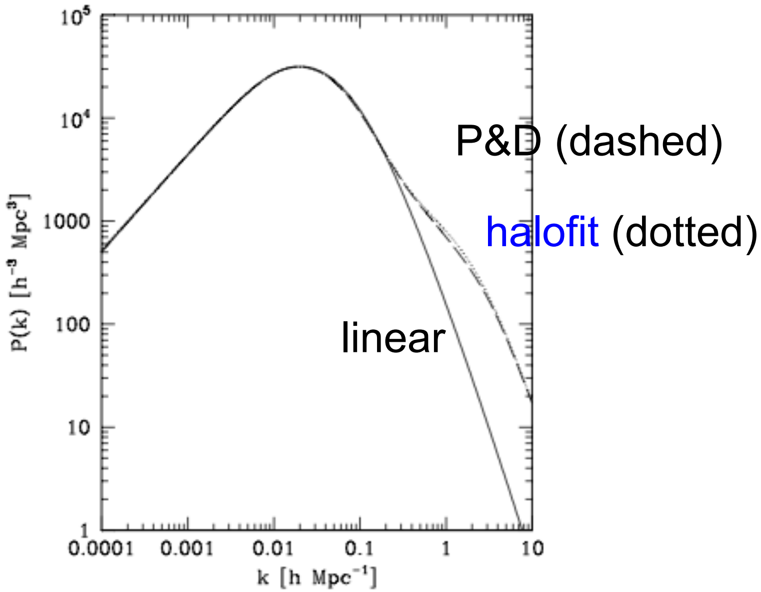
From the non-linear power spectrum obtained, it was possible to develop **fitting functions capable of reproducing the non-linear power spectrum from the linear one.**

$$f(\Delta_L^2) = \Delta_L^2(k_L)\left[\frac{1+B(n)\beta(n)\Delta^2(k_L)+[A(n)\Delta^2(k_L)]^{\alpha(n)\beta(n)}}{1+([A(n)\Delta^2(k_L)]^{\alpha(n)}g^3(\Omega,\Lambda)/[V(n)\Delta^2(k_L)^{1/2}])^{\beta(n)}}\right]^{1/\beta(n)}$$

Peacock & Dodds

This complicated expression fits the resulting NL dimensionless power spectrum, with the linear dimensionless power spectrum and a number of functions (that depend on the cosmological parameters and were fixed by measuring in the simulations).

There are different fitting functions, because they are based on different physical assumptions, like stable clustering or the halo model, they are not only mathematical fits. Such functions allow to compute P_{NL} with good precision for a range of cosmological parameter values without the need to run a new simulation for each values of the parameters (needed for parameter constraints).



The Halo model

At the end of an N-body simulation (i.e., at z=0), the dark matter particles cluster in 3-dimensional **halos** → defined by a density threshold, found with halo finder algorithms, like the friends-of-friends.

Inside some of the gravitationally bound halos, there are other gravitationally bound smaller sub-halos, which give origin to satellites.

For example, sub-halos of a cluster halo develop in cluster galaxies, while field galaxies form on autonomous galaxy halos.
Sub-halos of a galaxy halo will develop in satellite galaxies.

A halo is a set of neighbouring simulation particles that remain localized, gravitationally bound, with a density larger than a certain threshold : usually defined as

$$\rho_{halo} = 200\ \rho_{cr}\ (z)$$

As they form, halos decouple from the expansion (are collapsed objects) and usually can be spherical or triaxial.

The Halo model is a description of the non-linear inhomogeneous Universe based on the assumption that **all collapsed dark matter is contained inside halos of various scales,** from cluster scales (the largest collapsed perturbation at z=0 in the standard model) to smaller scales.

With this assumption, the non-linear power spectrum (and correlation function) is fully determined by three properties:

- the distribution of matter inside the halos, i.e., their **density profiles** → this determines the non-linear power spectrum on the smallest scales.

- the **mass function** of the collapsed halos, i.e., the distribution of halos as function of scale → this determines the non-linear power spectrum on the largest (of the non-linear) scales.

- the **linear power spectrum** → this determines the linear power spectrum in the remaining scales.

Density profile

The halo density profile is found from fitting the halos densities found in the simulations.

In the simulations it was found that the density profile is independent of scale (it is the same for halos of all sizes and masses) → it follows a universal form, known as the NFW profile (Navarro, Frenk & White),

It is given by,

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}$$

The density profile introduces **two new cosmological parameters** :

- halo density **amplitude** ρ_s

- **characteristic radius** r_s

The characteristic scale is related to a third parameter: the halo **concentration**.

$$c = \frac{R}{r_s}$$

The profile of a halo of size R can alternatively be written using the concentration parameter:

$$\rho_s = \frac{200}{3}\rho_c(z)\,\frac{c^3}{\ln(1+c) - c/(1+c)}$$

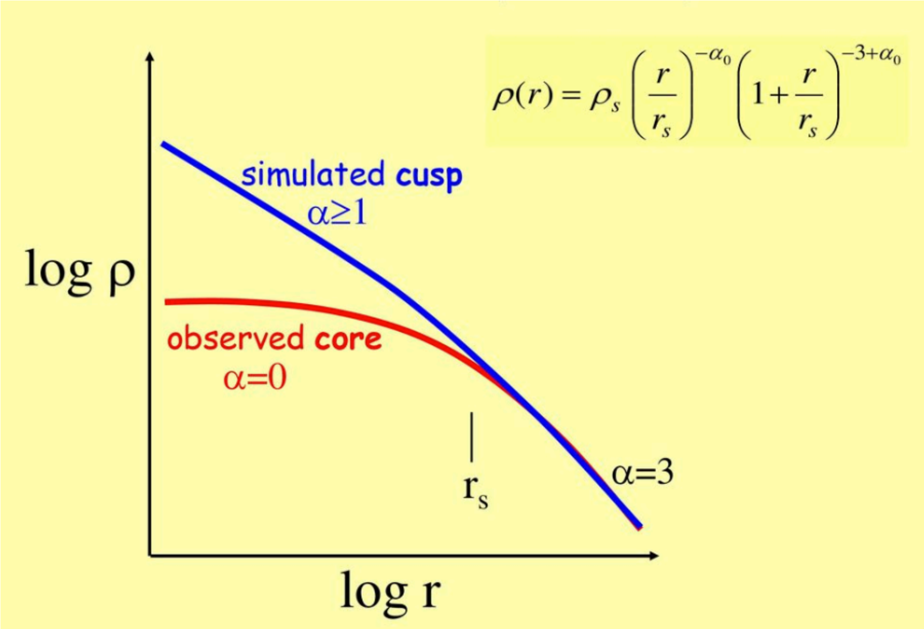
The **mean density of a halo** is: $\bar{\rho} = \frac{1}{V}\int_0^R \rho dV = 3\rho_s \int_0^1 \frac{dx\, x^2}{cx\,(1+cx)^2}$ $\qquad x = r/R$

The **mass within a halo** of size R is:

$$M = \bar{\rho}V = 200\rho_c(z)\,\frac{4}{3}\pi R^3$$

$$M = 100\,\frac{H^2(z)\,R^3}{G}$$

(halos at higher z are more compact).



NFW profile: The density goes with r^{-1} at the inner part of the halo and r^{-3} at the outer part.

→ the **cusp/core problem** of small-scale cosmology

In general, dark matter simulations have a steeper profile (a **cusp**) while observations (e.g. rotation curves of galaxies) have a flatter inner profile (a **core).**

Simulations with DM+baryons have in general flatter inner profiles due to **baryonic feedback** (SN and AGN gas outflows can change the gravitational potential)

Some DM models can predict flatter inner curves → failure of the ΛCDM description on small-scales?

Mass function

The mass function dN/dM may be measured in numerical simulations.

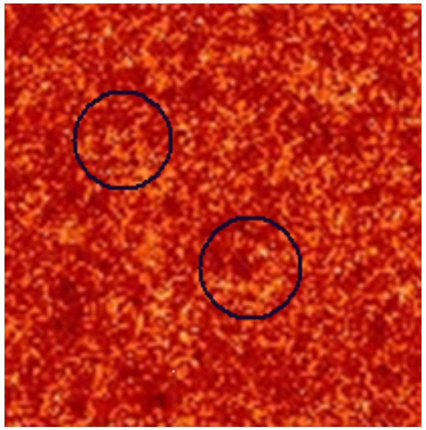
It may also be analytically derived under some description of the non-linear inhomogeneous Universe.

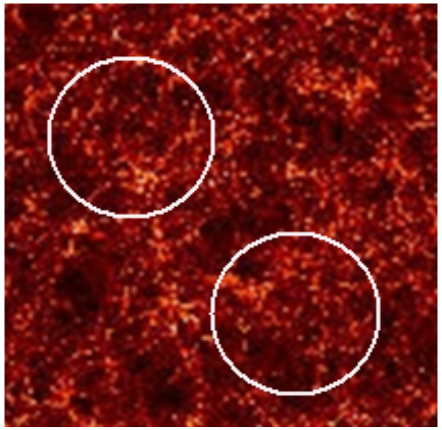
The pioneering derivation is the one presented by the Press-Schechter theory (1970s), based on a different description than the (more recent) halo model description.

The PS description of the non-linear inhomogeneous universe is based on the following assumption:

First, consider a density field and smooth it with a filter of a given scale R (associated to a mass scale M).

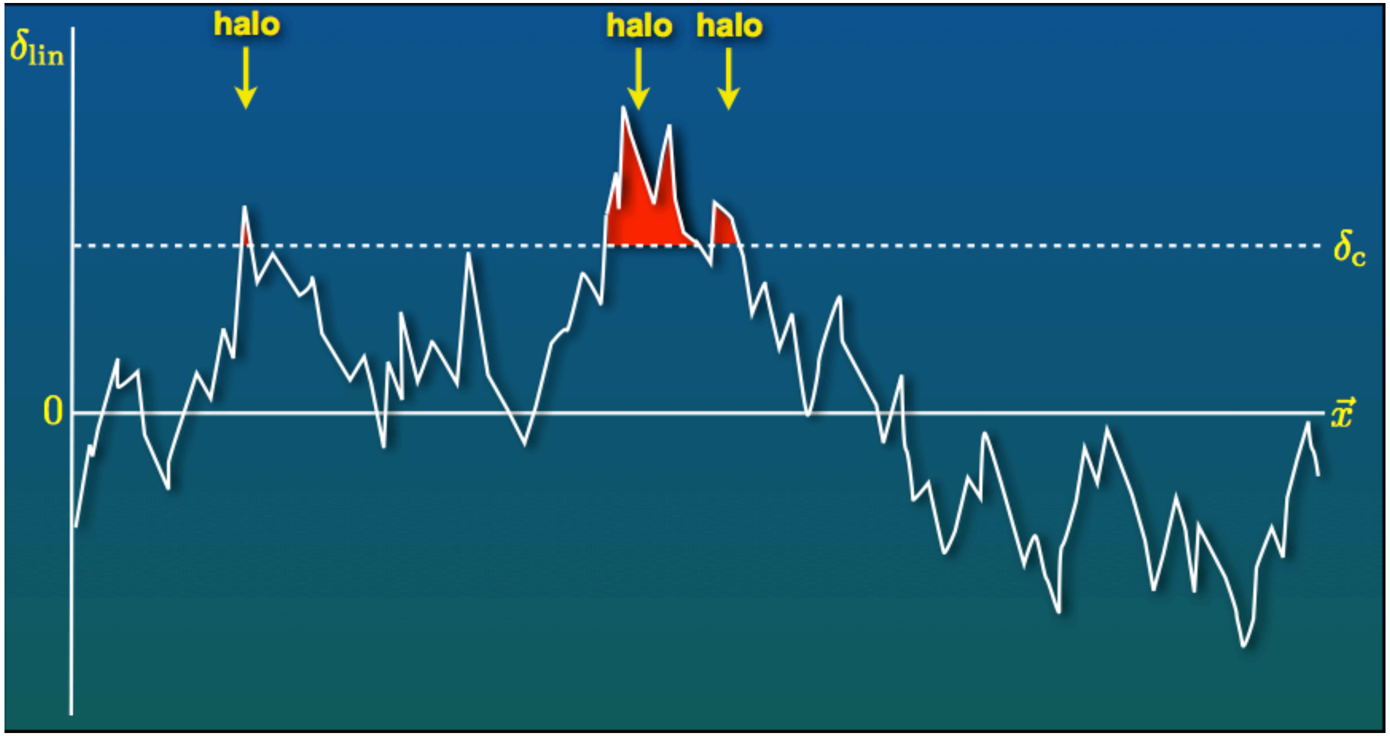
M_1





M_2

Then, the probability that the smoothed δ, i.e. $\delta_{M,}$ is above a threshold δ_c (critical overdensity), gives the fraction of mass contained in non-linear collapsed objects (halos) of mass larger than M.



δ_c : critical density for the non-linear collapse → δ_c = 1.69 according to the spherical collapse assumption, where it is the value of linear δ in virialized halos

Remember the overdensities grow as $\delta = a^f \delta_0 = D(t)\, \delta_0$ (D<1).
So a region that today has $\delta_0 > \delta_c / D(t)$, was already collapsed at time t

The smoothed density contrast field δ_M is a Gaussian random field (like the original density contrast δ_0)

$$P(\delta_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(\delta_1 - <\delta_1>)^2}{2\sigma_1^2}}$$

The probability of δ_M being above δ_c is the integral over the (tail of the) Gaussian, i.e., the **complementary error function**:

$$\mathcal{P}(\delta_M > \delta_c) = \frac{1}{\sqrt{2\pi}\,\sigma_M} \int_{\delta_c}^{\infty} \exp\left[-\frac{\delta_M^2}{2\sigma_M^2}\right] d\delta_M = \frac{1}{2}\mathrm{erfc}\left[\frac{\delta_c}{2\sigma_M}\right]$$

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du$$

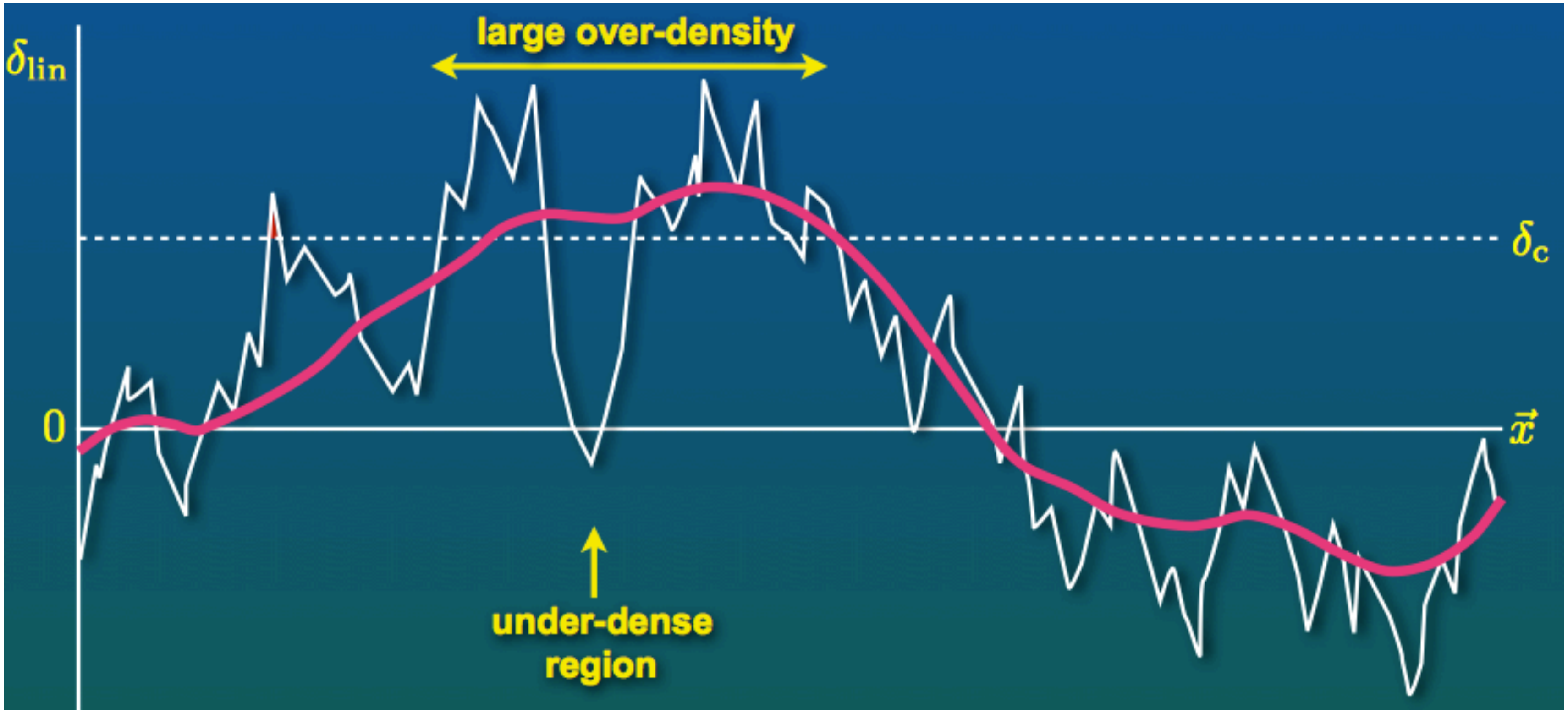
Notice that erfc(0) = 1 → **the PS assumption tells us that if the threshold was zero, only half of the mass would be collapsed in halos**.

Does this make sense?

- At first thought yes, because there are regions with δ < 0 (underdensities), they do not attract matter, do not form halos.

On average they account for half ot the total mass in the Universe.

- However, matter from underdense regions must fall towards the overdensities → they end up inside larger halos:



So all matter should be accounted for inside halos (like in the modern halo model), and the PS assumption must be modified to:

The probability that $\delta_{M,}$ is above a threshold δ_c gives half of the fraction of total mass contained in non-linear collapsed objects (halos) of mass larger than M:

$$P\ (\delta_M > \delta_c) = \mathrm{erfc}\ (\delta_c / 2\ \sigma_M)$$

From the fraction of total mass, we can write the mass function n(M,t) dM, i.e., the number of halos with masses in the range M to M+dM per comoving volume:

$$n\ (M,t) = dn/dM\ dM$$

i.e., $n\ (M,t) = 2\ dP/dM\ \ \rho_0 / M\ \ dM$

This is basically the integrand of our expression, with the difference that the probability was written for mass while we want now a number density → $(1/V = \rho / M)$

It is usual to write the mass function in logarithmic intervals of mass:

$$n(M,t) = \frac{\mathrm{d}n}{\mathrm{d}M} = \frac{1}{M}\frac{\mathrm{d}n}{\mathrm{d}\ln M}$$

which introduces an extra M in the denominator

and also to consider the variation of the probability with the variance of the overdensity (σ) instead of the mass,

i.e., to use dP/dlnM = dP/dlnσ dlnσ/dlnM

This is useful because structure formation gives predictions for σ.

The expression for the Press-Schechter mass function is then:

$$n\,(M,t)\,dM = \sqrt{\frac{2}{\pi}}\frac{\rho_0}{M^2}\frac{\delta_c}{\sigma}\frac{d\ln\sigma}{d\ln M}\exp(-\delta_C^2/2\sigma^2)$$

We can compute **the mass function for a given theoretical model**.
It depends on:

- Ω_m

- linear power spectrum at z=0 → from which we can compute σ for each scale M (or R)

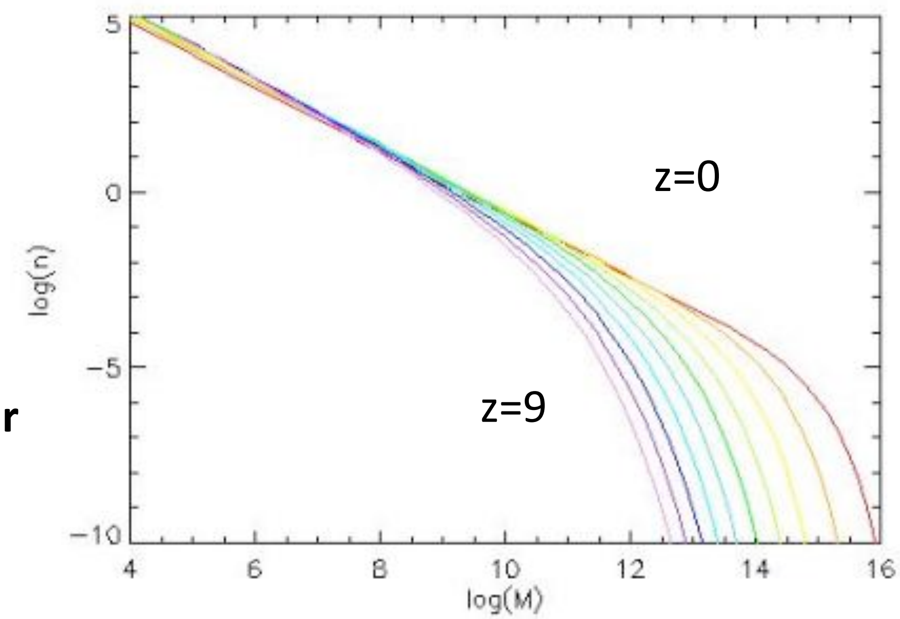
$$\sigma_R^2 = \left\langle \delta^2(k)\, W_R^2(k) \right\rangle = \frac{1}{(2\pi)^3} \int d^3k\, W_R^2(k) P(k)$$

$$W_R(k) = 3\frac{\sin kR - kR\cos kR}{(kR)^3}$$

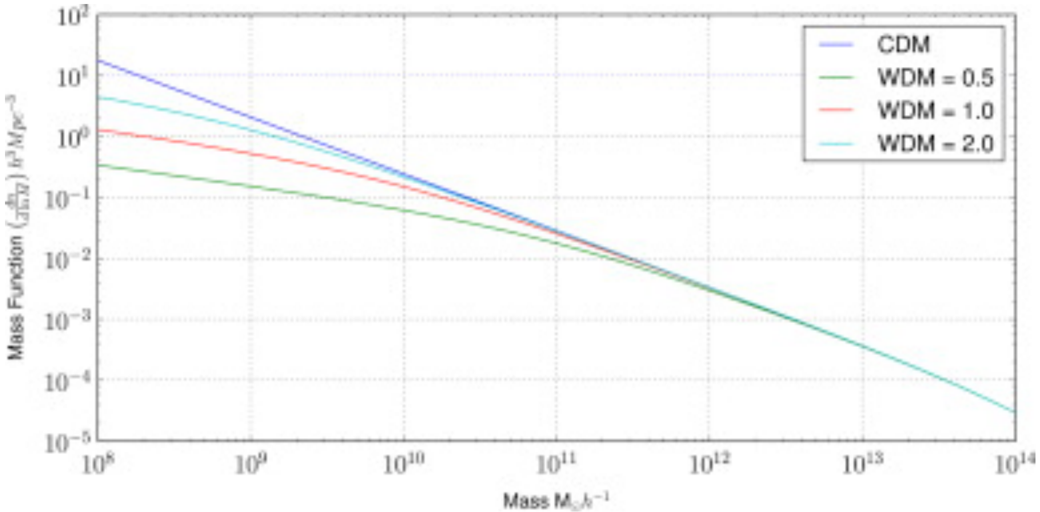
(for a top-hat window)

- δ_c (t): δ_c = 1.69 / D(t) → it increases with z (higher threshold at high z, meaning there are less collapsed structures then)

number of halos per comoving volume



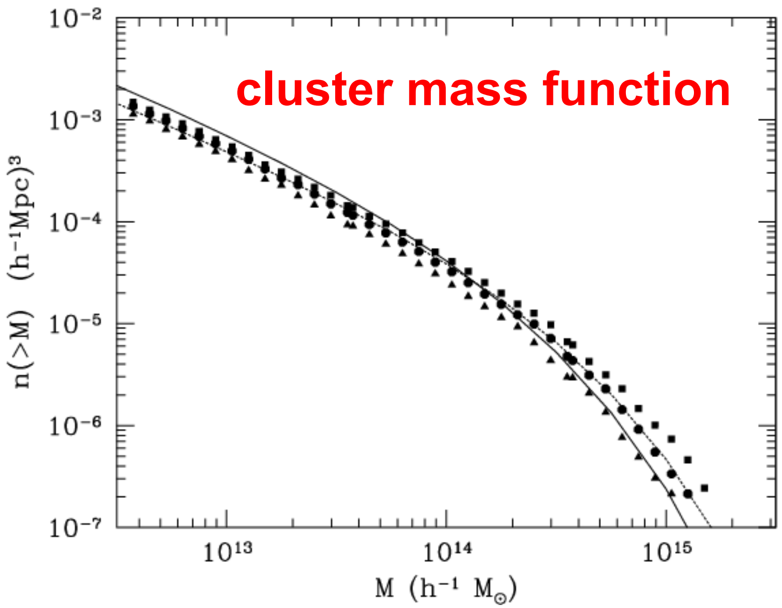
The number of halos per comoving volume:



decreases with mass (smaller σ, i.e., amplitude of the power spectrum) and redshift (the exponential is the dominating factor)

and it depends on the cosmological parameters

The number of clusters observed per mass bin **(cluster abundance)** **is a biased estimator of the number of halos in the cluster mass range →** comparing with the theoretical mass function allows us to constrain the cosmological parameters, especially σ_8 and Ω_m → it is a structure formation **cosmological probe**



The non-linear matter power spectrum

In the halo model, the mean density is written as an integral over all halos of different masses:

$$\bar{\rho} = \int dm\, n(m) m,$$

where n(m) is the number density of halos of mass m (i.e. the mass function)

The two-point correlation function of the density field is separated in two contributions:

$$\xi(\mathbf{x} - \mathbf{x}') = \xi^{1h}(\mathbf{x} - \mathbf{x}') + \xi^{2h}(\mathbf{x} - \mathbf{x}')$$

- **1-halo term**: for the smallest scales, when the two points belong to the same halo

- **2-halo term**: for larger scales, when the two points are in different halos

The power spectrum is then the sum of the two terms:

$$P(k) = P^{1h}(k) + P^{2h}(k)$$

The power spectrum is basically a weighted integral over <mm> (the overdensities times the mean density) - weighted by the mass function - and convolved with the density profile:

$$P^{1h}(k) = \int dm\, n(m) \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2$$

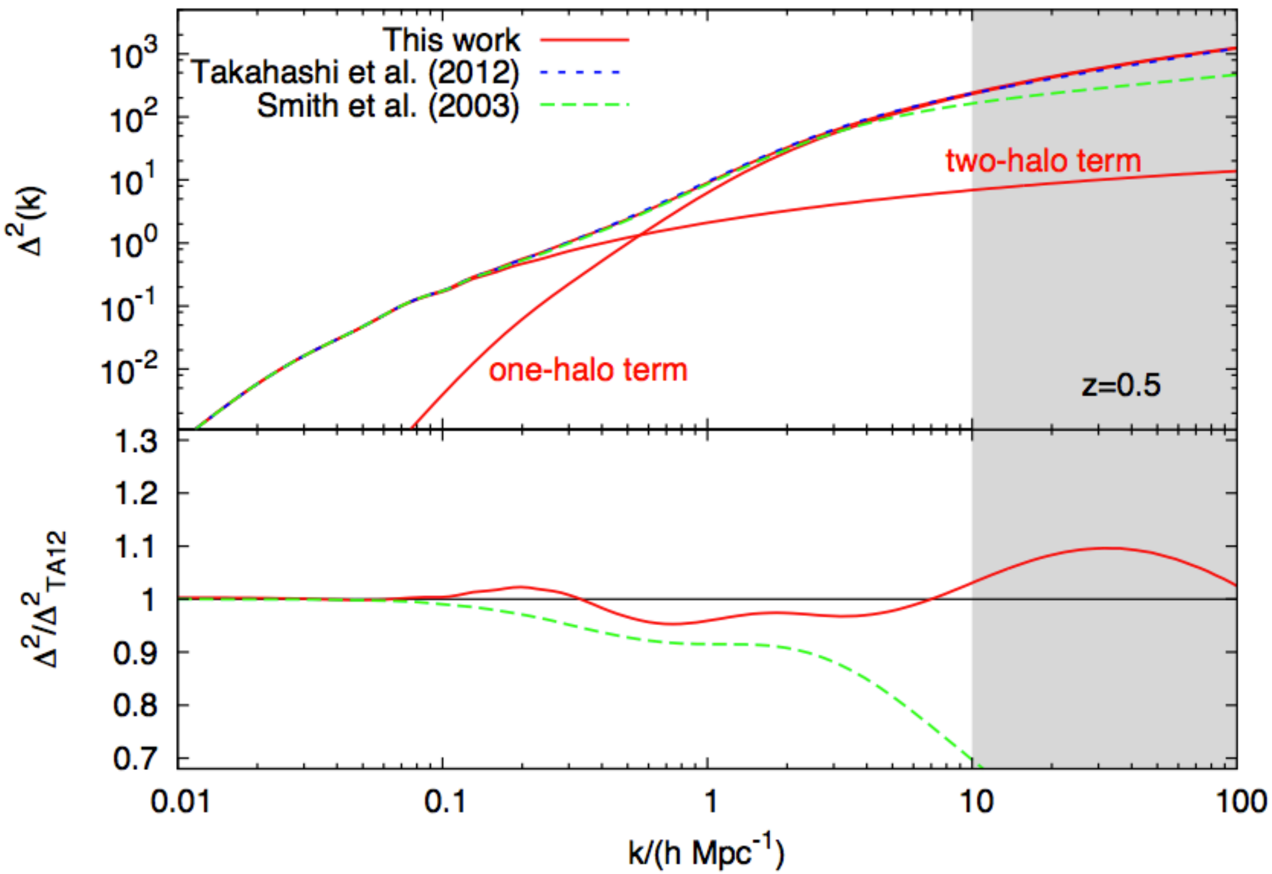
$$P^{2h}(k) = \int dm_1\, n(m_1) \left(\frac{m_1}{\bar{\rho}}\right) u(k|m_1) \int dm_2\, n(m_2) \left(\frac{m_2}{\bar{\rho}}\right) u(k|m_2)\, P_{hh}(k|m_1, m_2)$$

n(m) is the mass function - so besides being associated to a cosmological probe, the mass function is important to compute the P_{NL} in the halo model -
m_i is the mass of the halo i,
u(k|m) is the Fourier transform of the density profile of the halo of mass m]

The probability of finding a second halo, given a first one is not independent → a conditional probability is needed, i.e. a halo-halo correlation function (or in this case, in Fourier space, a halo-halo power spectrum P_{hh}):

$$P_{hh}(k|m_1, m_2) \approx \prod_{i=1}^{2} b_i(m_i) P^{\mathrm{lin}}(k)$$

(the halo-halo power spectrum is just a biased version of the linear matter power spectrum).



The dimensionless non-linear matter power spectrum computed in three ways (halo model and two fitting functions from N-body simulations):

The 1-halo term gives a good result on the smallest scales.
The 2-halo term gives a good result on the largest scales.

The halo model result and the Takahashi fit deviate at most by 10% on very small scales.