Structure Formation

Non-linear clustering

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) and there is mode coupling → 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

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

In addition, another equation is needed for v (terms with v were neglected before because they only appeared in higher-order terms). That equation is also found from a combination of Euler, continuity and Poisson equations.

Now, perturbation theory assumes that it is possible to expand the density and velocity fields around the linear solution:

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

This defines the density contrast (and the same for the velocity perturbation) in several orders: n=0 (homogeneous), n=1 (linear), n=2 (quadratic) ,…, up to infinity.

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

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

In PT the equation of motion is solved recursively:

Insert the linear solution $\delta^{(1)}(a)$ in the non-linear terms (A, B), while the low-order terms are written with $\delta^{(2)}$. This provides an equation of motion for $\delta^{(2)}$. (and the same for the equation for v).

Solving the system finds a solution $\delta^{(2)}(a)$, which can then be inserted in A, B, to get an 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:

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

In the matter-dominated epoch

the linear solution is: $\delta^{(1)}(a) \sim a$ in the case of EdS model, single component)
(or $\delta^{(1)}(a) \sim a^f$ in the case of DM+DE)

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:

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

We see that:

- the growth of δ for a given scale k and in a given order is a power of "a" times a factor that might in principle depend on all scales.

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

- the Dirac delta tells us which scales contribute

For example in second order

the Dirac delta shows that the $\delta^{(2)}$ contribution to the growth of the density contrast of a scale k is determined by the mode coupling of the linear field $\delta^{(1)}$ at all pairs of scales q_1 and q_2 such that q_{12} (i.e. $q_1 + q_2$) is equal to k:

$$\delta^{(2)}(k,a) = a^2 \int d^3q_1 \int d^3q_2\, \delta_D(\vec{k} - \vec{q}_{12}) F_2(\vec{q}_1, \vec{q}_2) \delta^{(1)}_{q_1}(a) \delta^{(1)}_{q_2}(a)$$

The kernel for the second-order density contrast is given by

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

for each pair of scales (k_1, k_2) such that $k_1 + k_2 = k$

F_2 is a number

For example, **consider the growth of scale k = 10 h/Mpc:**

the contribution from parallel scales |k| = 8 h/Mpc and |k| = 2 h/Mpc is

$$F_2 = 5/7 + 0.5*2/8 + 0.5*8/2 + 2/7 = 3.125$$

while the contribution from parallel scales |k| = 6 h/Mpc and |k| = 4 h/Mpc is

$$F_2 = 5/7 + 0.5*2/8 + 0.5*8/2 + 2/7 = 2.08$$

and so

$$\delta_{10}^{(2)}(a) = [3.125\, \delta_8^{(1)}(a)\, \delta_2^{(1)}(a) + 2.08\, \delta_6^{(1)}(a)\, \delta_4^{(1)}(a) + \ldots.]\, a^2$$

But perturbation theory 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.**

However, we can gain some insight on the collapse and even build models to describe the non-linear distribution of matter by making some assumptions about collapsed halos.

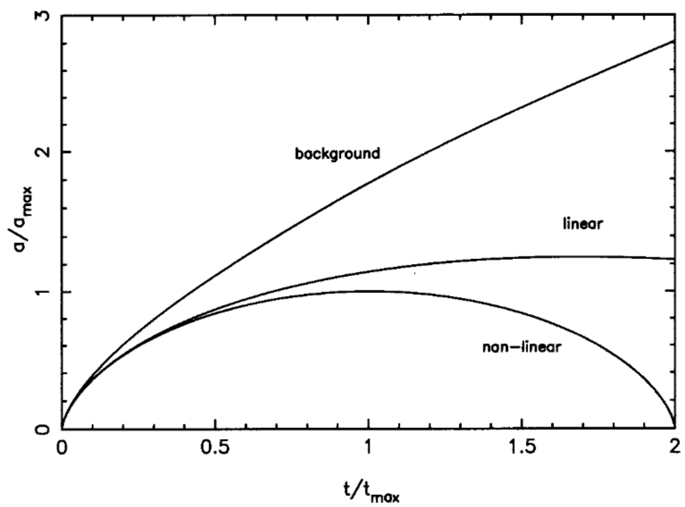
Halo properties

Spherical collapse model

Let us consider a spherical halo within the expanding flat background.

Assumption: **a DM halo can be treated as a mini-universe that expands within the Universe**.
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 to the linear evolution, as we saw earlier), **the mini-universe expands with a decreasing rate, stops 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. The **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 (a 2-component fluid with $\Omega_m > 1$, i.e., positive curvature and assuming negligible dark energy and radiation) cannot be written explicitly but can be given analytically as a parametric solution using an intermediate parameter θ:

(see homework)

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

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 $t = 2\,t_{max}$.

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 for the expansion of the mini-universe a(t):

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

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 expands $\sim a_{lin}^{-3}$, while the density of the universe decreases as its volume expands $\sim a_{back}^{-3}$

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

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

We have the $a_{lin}(t)$ solution

The background solution is $a_{back}(t) \sim t^{2/3}$, or more precisely, in this framework of the parametric variable is

$$a_{back} = \frac{1}{4}\left(6\pi \frac{t}{t_{\max}}\right)^{2/3} a_{max}$$

So we get:

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

We find that at **turnaround** ($t=t_{max}$), the linear density contrast is:

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

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 $a = a_{max}$, where $a_{back} = (6\,\pi)^{2/3} / 4$

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

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

which shows that 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 (**halos**) in the Press-Schechter theory.

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 half its maximum size.

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

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

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 $2^{2/3}=1.6$

→ **the background density decreased by a factor of $1.6^3 = 4$.**

So the final true (non-linear) density contrast increases by a factor of 8x4 = 32 from the value of 5.55 at turnaround, i.e.,

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

(see homework for an alternative derivation)

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

$\delta \sim 5$, at the decoupling from the comoving flow ($\delta_l \sim 1$)

δ - 178, for virialized structures ($\delta_l \sim 1.68$)

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 $\delta \sim 1.68$, we know we are dealing with a collapsed object.

Note that the derivation assumes the collapse occurs at z=0.
For higher redshifts, the threshold value is lower (there are less regions collapsed):

$$\delta_c^{lin} \sim 1.68 / (1+z)$$

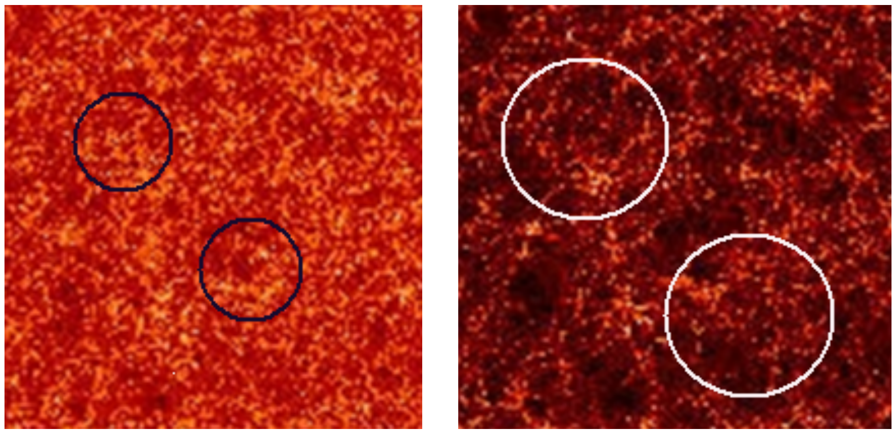
The derivation can also be made for different cosmological models.
The value of the typical virialized overdensity of dark matter in the presence of other dark energy models will be different.

Mass function

The mass function, the number of halos in mass bins, dN/dM, may be analytically derived under the following ***assumption***: **the probability that the smoothed δ, i.e. δ_R or δ_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.**

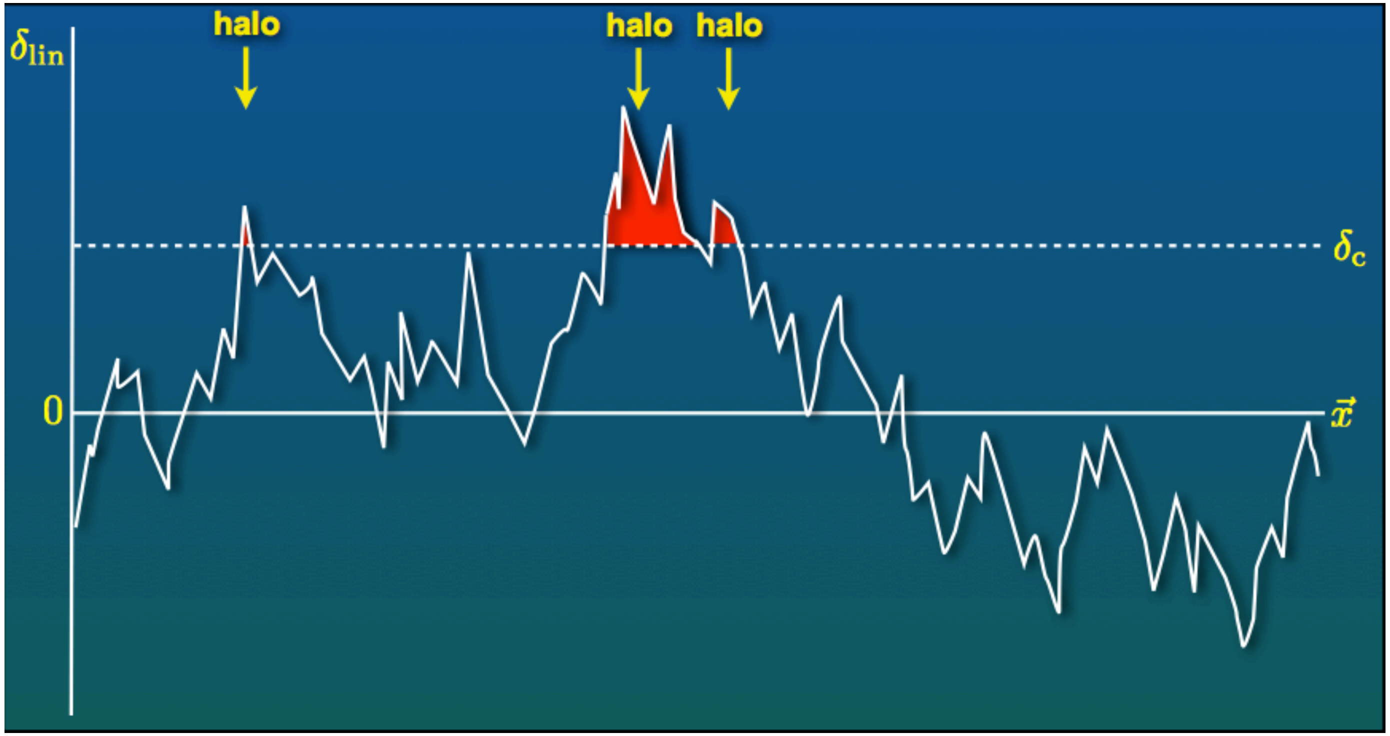
The idea and pioneering derivation is known as the Press-Schechter theory (1970s).

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, assume that the probability that the smoothed linear δ is above a threshold δ_c gives the fraction of mass contained in halos of mass larger than M.



δ_c : critical density for the non-linear collapse → δ_c = 1.68 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

We do not know the expression for the non-Gaussian distribution of the non-linear density contrast. But the assumption is made on the smoothed linear density contrast field δ_M , which is a Gaussian random field (like the linear density contrast δ)

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

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

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

(Note there is one such expression for each smoothing radius, i.e., for each variable δ_M)

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

Note 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**.

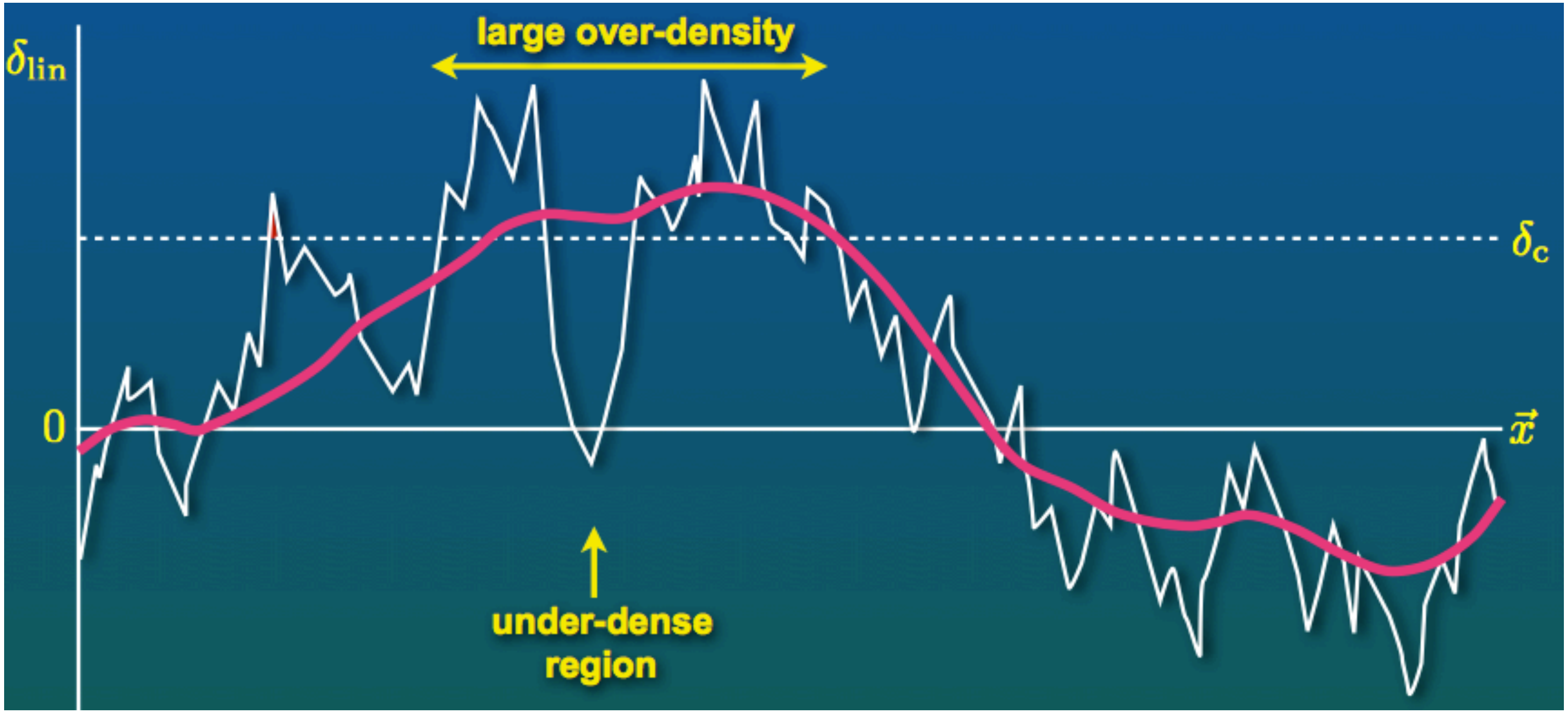
(the half on the positive side of the distribution)

Does this make sense?

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

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

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



So all matter is in one way or another inside a halo, and the PS assumption needs to be corrected. It is modified to:

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

correct fraction = 2 P ($\delta_M > \delta_c$)

From the fraction of total mass, we can write the

mass function $dn(M,t) = n(M,t)\, dM$, i.e., the number of halos with masses in the range M to M+dM per comoving volume:

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

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

(the minus sign meaning the number is a decreasing function of mass)

This is the integrand of the probability expression, times the factor two, and dividing by volume to get a number density → ($1/V = \rho/ M$)

and so the **Press-Schechter mass function** is

$$\frac{dn}{dM}(M,t) = -\left(\frac{2}{\pi}\right)^{1/2} \frac{\Omega_m \rho_c}{M} \frac{1}{\sigma_M} e^{(-\delta_c^2(t)/2\sigma_M^2)}$$

where the mean density ρ_0 is written as $\Omega_m\, \rho_{c;}$ the density contrast is fixed by the integration limit (i.e. we write δ_c); and its time-dependence is indicated explicitly.
(Notice that the time dependence could alternatively have been in σ_M (z) and the value of δ_c today used instead).

It is usual to change the integration variable to σ. In this case, we need to consider that

dP/dM = dP/dσ dσ/dM

The change of variable introduces a factor δ_c/σ in the integrand, and the mass function takes the alternative form

$$\frac{dn}{dM}(M,t) = -\left(\frac{2}{\pi}\right)^{1/2} \frac{\Omega_m \rho_c}{M} \frac{\delta_c(t)}{\sigma_M^2} \frac{d\sigma_M}{dM} e^{(-\delta_c^2(t)/2\sigma_M^2)}$$

It is also usual to write the mass function in logarithmic intervals of mass or logarithmic intervals of variance of the overdensity (σ), which introduces an extra 1/M or σ factors. For example,

$$\frac{dn}{d\ln M}(M,t) = -\left(\frac{2}{\pi}\right)^{1/2} \frac{\Omega_m \rho_c}{M^2} \frac{\delta_c(t)}{\sigma_M} \frac{d\ln \sigma_M}{d\ln M} e^{(-\delta_c^2(t)/2\sigma_M^2)}$$

The mass function depends on the **variance of the smoothed overdensity**.

Remember this is computed from a filtered integral of the power spectrum:

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

with $$W_R(k) = 3\frac{\sin kR - kR\cos kR}{(kR)^3}$$ (for a top-hat window)

Since M ~ R^3, there is a one-to-one relation between the size of the smoothed region R and its mass M, and **we can also define a σ_R, which is identical to σ_M.**

Note that if we integrate the power law power spectrum P(k) = A k^n using various filter sizes R and find the various σ_R, we can find **an expression relating the variance of the various scales.**

For example using R=8 Mpc/h as a reference, the result is:

$$\sigma_R^2 = \sigma_8^2 \left(\frac{R}{8\mathrm{h}^{-1}\mathrm{Mpc}}\right)^{-(3+n)}$$

and analogously for σ_M

$$\sigma_M^2 = \sigma_8^2 \left(\frac{M}{M_8}\right)^{-\frac{(3+n)}{3}}$$

where M_8 is the mass enclosed in the R=8 Mpc/h sphere

Notice that σ is monotonic in mass → it decreases with mass (a larger smoothing radius leads to a smaller variance)→ 1/σ is a **proxy** for mass.

This allows us to write the mass function in logarithmic bins of 1/σ instead of mass. It turns out the expression gets simpler, because the derivatives cancel:

$dn/d\ln\sigma^{-1} = - dn/dM \; dM/d\ln\sigma$

and the mass function becomes:

$$\frac{dn}{d\ln\sigma^{-1}}(M,z) = \frac{\Omega_m \rho_c}{M}\left(\frac{2}{\pi}\right)^{1/2} \nu\, e^{\frac{-\nu^2}{2}} = \frac{\Omega_m \rho_c}{M}\psi(\nu(M,z))$$

So the mass function has a simple form if written when using the inverse variance as an indicator of mass.

This expression also introduced the peak height : $\boldsymbol{\nu(z) = \delta_c(z) / \sigma_M}$ (which is the quantity that directly appears in the Gaussian)

It is also usual to define the **characteristic mass M_* - the mass of dark matter halos that virialize today.**

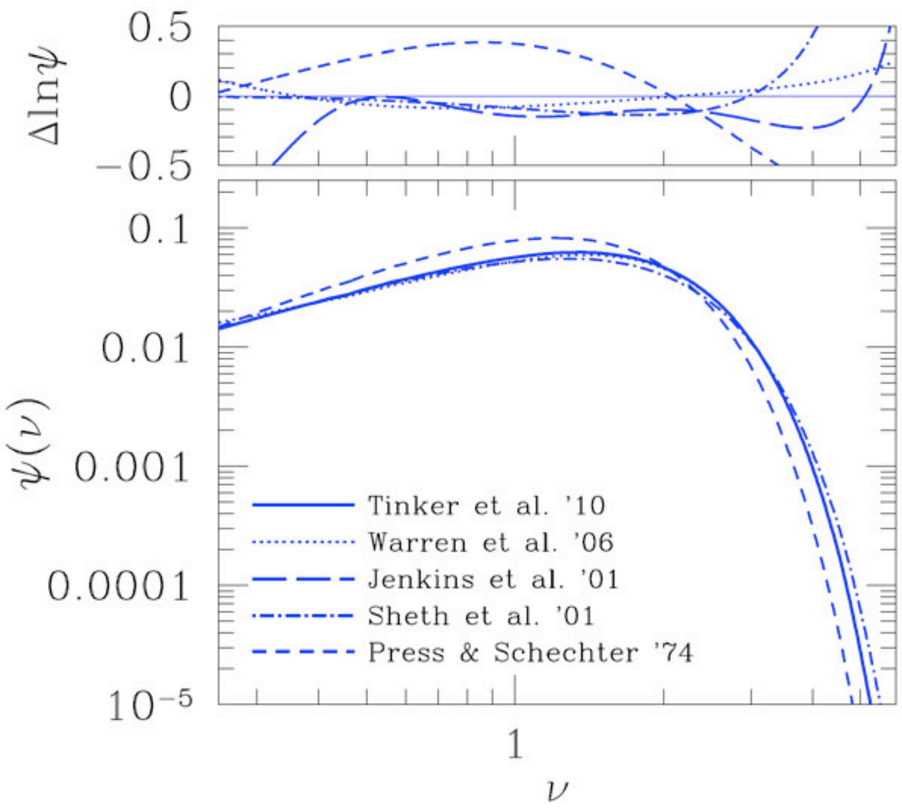
It corresponds to the scale that has $\nu = 1$ today, i.e., $\sigma_M = \delta_c\,(a=1)$

In **ΛCDM concordance model ($\sigma_8 = 0.9$, $n_s = 1$)**, the characteristic mass scale is
$M_* \sim 0.1\, M_8 \rightarrow R_* \sim 0.5\, R_8 \rightarrow$ half the radius of a typical cluster, it is still in the cluster range.

So larger clusters have linear $\delta \sim 1$ (start to collapse today), while smaller clusters have linear $\delta = \delta_c = 1.68$ today (are already virialized collapsed structures)

It is important to note that there is an **uncertainty on the theoretical mass function**.

For example, if assuming an ellipsoidal collapse model, instead of a spherical collapse, the resulting mass function is different (it is the **Sheth & Tormen mass function**). Many other mass functions were derived based on different assumptions, or based on fits to the halo abundance found in N-body simulations.



Chronological list of dark matter halo mass functions:

- Press& Schechter 1974
- Sheth & Tormen 1999
- Jenkins et al. 2001
- Reed et al. 2003
- Warren et al. 2006
- Reed et al. 2007
- Tinker et al. 2008
- Crocce et al. 2010
- Bhattacharya et al. 2011
- Courtin et al. 2011
- Angulo et al. 2012
- Watson et al. 2013
- Bocquet et al. 2016
- Despali et al. 2016
- Comparat et al. 2017
- Diemer 2020b
- Seppi et al. 2020

Halo model

The Halo model is a description of the non-linear inhomogeneous Universe based on the following ***assumptions***: **all collapsed dark matter is contained inside halos, and it can be inside halos of various scales.**

With this assumptions, the **non-linear power spectrum** (and correlation function) is determined from the **linear power spectrum** if we know three properties of the halos:

- 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 gives a weight function

- the **halo bias** → this determines the non-linear power spectrum from the linear power spectrum

In the halo model, **the mass is contained in N_h halos,**

and within a halo, the mass is discretized in **N_i points** forming a number density profile u(r)

The **mass density at a point r**, in the neighborhood of some mass points "i" , is then

$$\rho(r) = m \sum_{i=1}^{N} u(r - x_i)$$

The matter power spectrum is separated in two contributions:

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

- **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

1-halo term:

the power spectrum is the mean <δδ> (the overdensities ρ/<ρ>) over all halos, i.e. the sum of <mu mu> auto-correlation (same halo) over all halos, weighted by the number of halos in each mass bin (i.e., weighted by the mass function).

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

u(k|m) is the Fourier transform of the density profile of the halo of mass m

2-halo term:

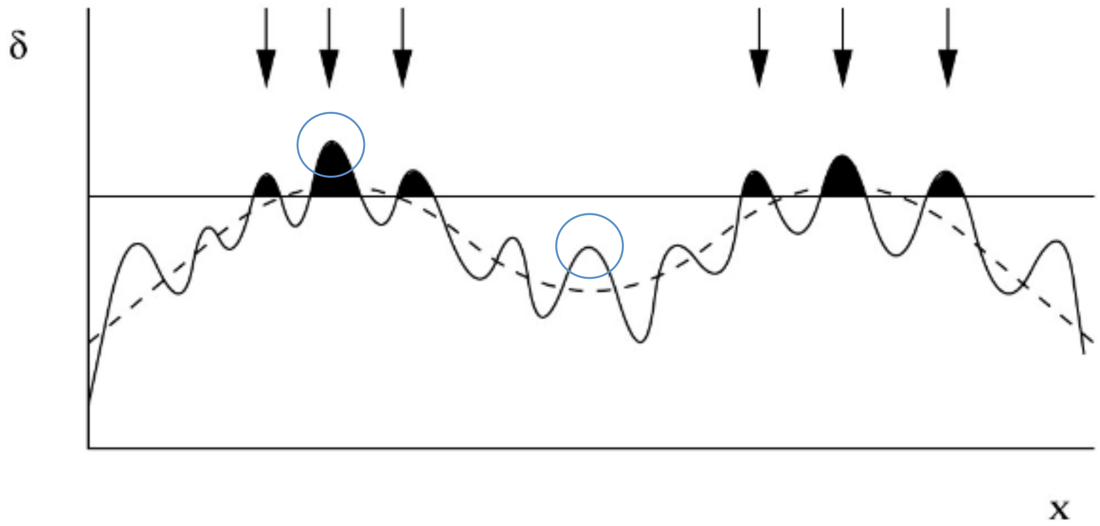
similarly, the power spectrum is also the mean <δδ> (the overdensities ρ/<ρ>) over all halos, but now it is the sum of <mu mu> correlation (between two halos), weighted by the mass function.

Since 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}).

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

Now, what is the **halo power spectrum** ?

Is it different from the matter power spectrum?



Consider two spatial locations with identical small scale σ_R overdensities, with respect to the local mean density.

Even though they have identical matter clustering properties (linear matter power spectrum), the one in the higher local environment will form a halo, and the other will not.

So, the halo density distribution is biased with respect to the (linear) matter density field → **halo bias** (there is also a bias between galaxy distribution and matter distribution → **galaxy bias)**

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

The bias is mass-dependent, larger halos are more biased b(m)

We can estimate the bias

by splitting the density field in small-scale and large-scale modes, i.e., writing at each location, $\delta = \delta_l + \delta_s$

A halo will form if the threshold is reached, i.e., when $\delta_s = \delta_c - \delta_l$. As expected, the halo formation is easier in locations with an already large value of δ_l

So, for a given mass M, **the bias can be defined as the difference between the number of halos that form in the presence of δ_l and the number of halos that would form if there were no underlying δ_l** (which would correspond to the matter power spectrum), i.e.

$$b(M) = \frac{n(\delta_l) - n(0)}{\delta_l}$$

This can be computed from the mass function

$$\frac{dn}{d\ln M}(M,t) = -\left(\frac{2}{\pi}\right)^{1/2} \frac{\Omega_m \rho_c}{M^2} \frac{d\ln\sigma_M}{d\ln M} \nu e^{(-\nu^2/2)}$$

where the **peak height** is now $\nu = \frac{\delta_c - \delta_l}{\sigma_M}$

i.e., **the mass function n in the presence of δ_l is a modification of the original mass function n(0)** that consists on changing the peak height value. Let us write it as a Taylor expansion around the original mass function:

$$n(\delta_l) \approx n(0) + \frac{dn}{d\delta_l}\delta_l$$

We need to compute the derivative of the mass function:

$$\frac{dn}{d\nu} = (1/\nu - \nu)\, n \qquad \frac{d\nu}{d\delta_l} = -1/\sigma_M \qquad \rightarrow \qquad \frac{dn}{d\delta_l} = \frac{\nu^2 - 1}{\nu\sigma_M} n$$

to first order, n in the derivative is ~ n(0) and we get the result,

$$n(\delta_l) \approx \left(1 + \frac{\nu^2 - 1}{\nu \sigma_M} \delta_l\right) n(0)$$

The halo bias is then given by

$$b(M, z) = 1 + \frac{\nu^2 - 1}{\delta_c(z)}$$

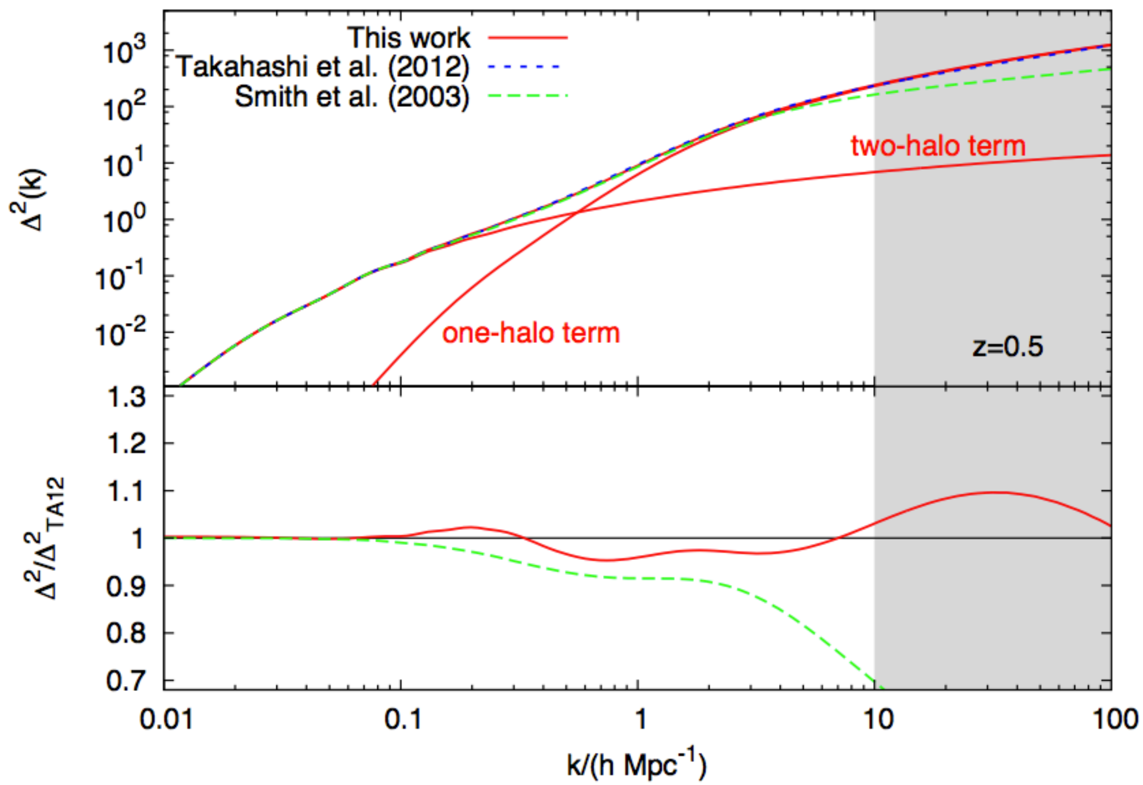
We see that the peak height is a determining quantity:

- halos with the characteristic mass M_* ($\nu = 1$) have no bias, their correlation function is representative of the underlying matter correlation

- halos with a larger mass (small σ_M , $\nu > 1$) have a larger bias, are more correlated than the underlying matter distribution

- halos with small mass ($\nu < 1$) have a bias $b < 1$

With an estimate for the halo bias, the mass function, the halo density profile *(to be seen later)*, and the linear matter power spectrum, the halo model predicts the following matter non-linear power spectrum:



The dimensionless non-linear matter power spectrum from the halo model is compared with other two non-linear Δ^2 that were computed with fitting functions from N-body simulations

The 1-halo term gives a good result on the smallest scales (deviating at most by 10%)

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

N-body dark matter simulations

The most complete way to compute the evolution of the density field is numerically, with N-body simulations.

On the resulting map, we can make a direct measurement of the non-linear matter power spectrum. In addition we can also find halos, and measure all their properties (mass function, density profile, halo bias).

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 individual DM 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.

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

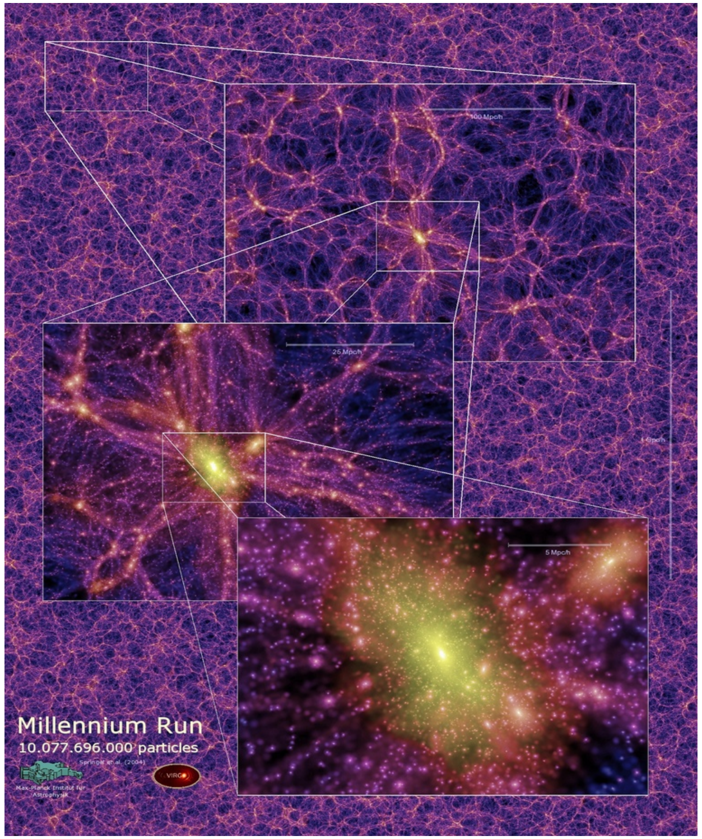
Larger volume:
z=18, L=500 Mpc /h

Lower resolution:
$M=9 \times 10^8\ M_{Sun}$

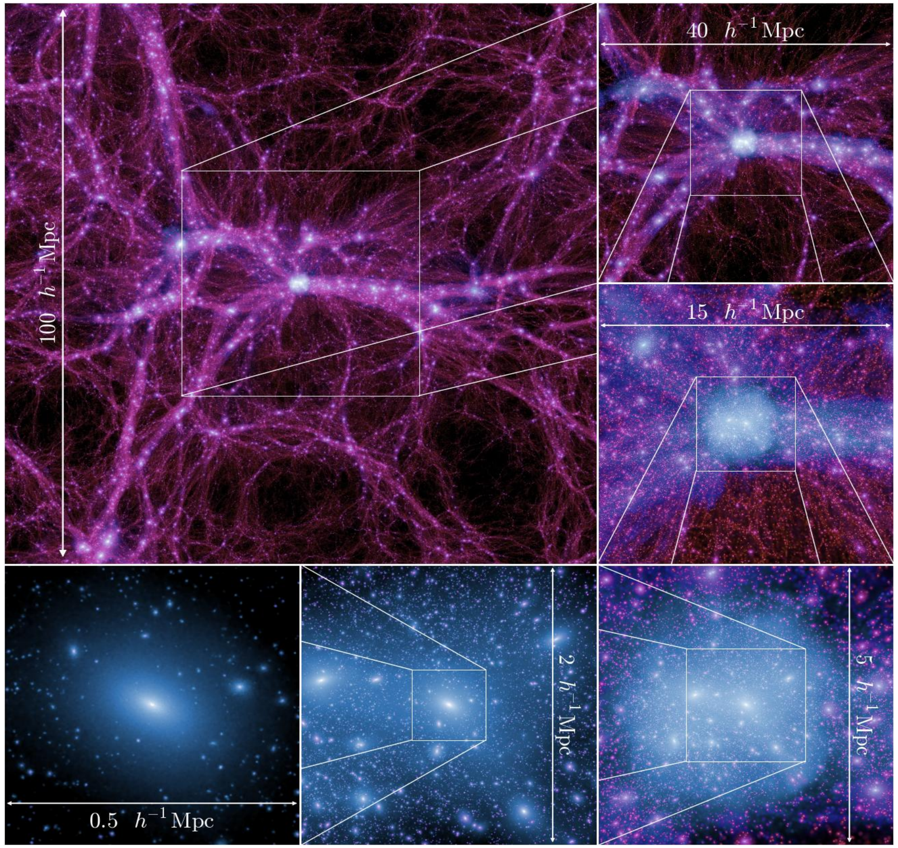
Millennium II

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

Higher resolution:
$M=7 \times 10^6\ M_{Sun}$



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

Iterative evolution

From the initial configuration, 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 is usual modified at small separation, deviating from 1/r, (by introducing a softening length ε - given by a fraction of the mean separation between two particles) that prevents strong collisions between the macroscopic particles.

This size will be a resolution limit for the simulation.

Then compute the **acceleration** on each particle from the **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 to describe the evolution of the scalar field (**Klein-Gordon equation**).

Then repeat for the next time step.

Grid

Because of the long-range of gravitational interactions, the potential at a given location depends on all particles → this standard process of **direct summation** is very slow, order N^2 computation.

To be faster, different algorithms were developed, where particles are placed on a grid → this enables an order logN computation.

From the initial conditions, the mass of each particle is distributed among the nearest cubic grid points.

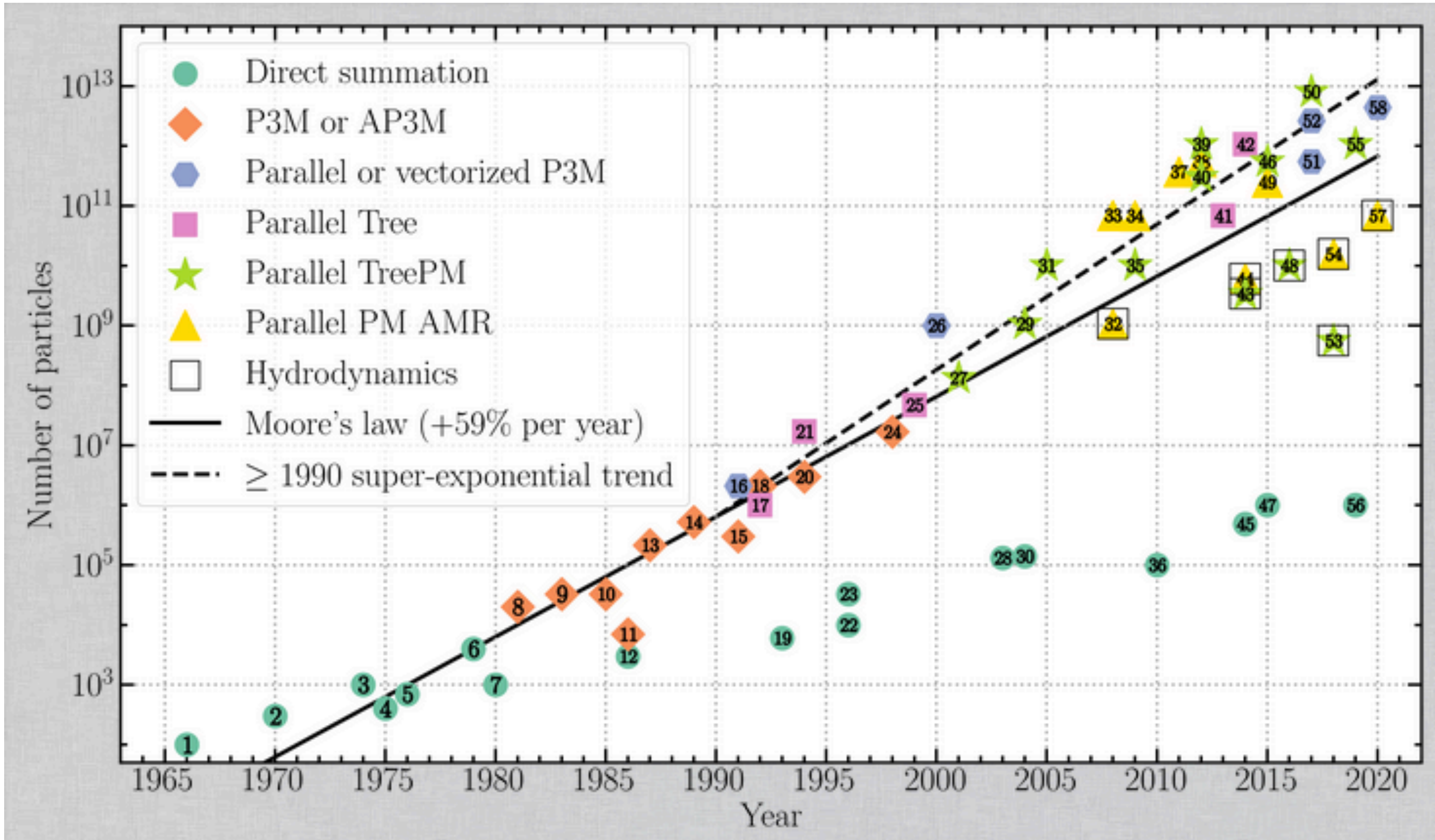
The acceleration field of this grid mass distribution is then computed with Fast Fourier Transform: FT of density → leads to the potential Φ_k through Poisson equation and acceleration $g_k = ik\, \Phi_k$ → acceleration field transformed back to the real space and used in the acceleration equation.

A simulation can contain several levels of grids, with different resolutions (**adaptive** mesh).

There are many different grid techniques. One example is the P^{3}M (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).



Post-processing of N-body dark matter simulations

Non-linear power spectrum

The power spectrum of the resulting density field can be measured in the output map.

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

From the measured non-linear power spectrum it was possible to develop **fitting functions that allow us to compute the non-linear power spectrum from the linear one.**

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

Peacock & Dodds prescription

This complicated expression fits the resulting NL dimensionless power spectrum from N-body simulations, with a complicated expression written as function of the linear dimensionless power spectrum and of a number of functions (that depend on the cosmological parameters and were fixed by measurements in the simulations).

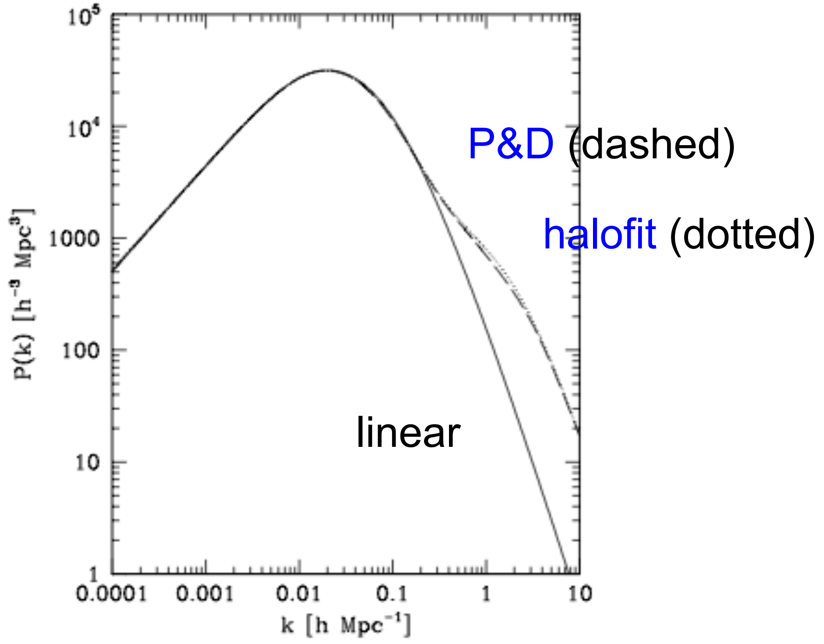
This is clearly not just a fit of a mathematical function. It is based on a physical assumption: **stable clustering**,

which assumes that the size r_L of a halo in linear theory (containing a density ρ_L) can be related to the true non-linear size r_{NL} through mass conservation: $r^3_L \rho_L = r^3_{NL} \rho_{NL}$

This defines a non-linear scale associated to each linear scale:

$$k_L = k_{NL} [1+ \Delta^2_{NL} (k_{NL})]^{-1/3}$$

The **halofit** is an approach that finds the halo bias (of the halo model) from N-body maps, and produces a P_L to P_{NL} function.



These and other fitting functions allow us 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).

Halo finders

A **halo** is a nonlinear peak of the matter density field, with its boundary defined by a certain density contrast.

There are different algorithms to identify halos.

A widely used one is to define a halo as a region around a local peak of density such that $\rho_{halo} > 200\ \rho_{cr}(z)$

Another popular method is the **Friends-of-Friends** algorithm:

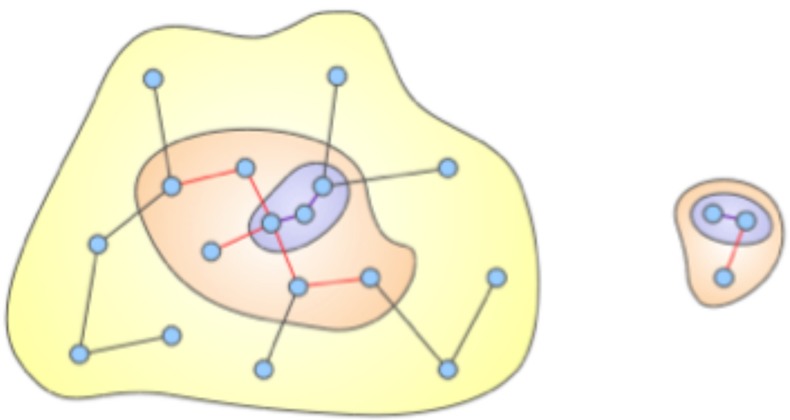
It is a purely geometric grouping of particles, where a halo is defined as a connected region within a density isosurface (threshold) - not depending on the halo being gravitationally bound or not.

The solution is determined by the choice of a single free parameter: the **linking length, b**

The algorithm computes the mean distance between particles: d_m
Then, it assigns to the same halo all particles such that $|r_i - r_j| < b\ d_m$

The result depends on the choice of the value of b. The value of b also depends on the cosmological model → a model that produces stronger clustering must have a larger threshold to define a halo.

For ΛCDM concordance model, b ~0.1-0.2



Using a hierarchy of linking lengths, we can also detect **substructures** (sub-halos inside a halo), which give origin to (baryonic) **satellites**.

For example **cluster galaxies** form in sub-halos of a cluster DM halo, while **field galaxies** form on autonomous galaxy halos. **Satellite galaxies** form in substructures of galaxy DM halos.

The mass of the various halos can be computed from the number of particles that belong to a halo. The volume and number of halos inside a volume can also be measured → can measure the halos **mass function**.

Halo density profile

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

It is 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:

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

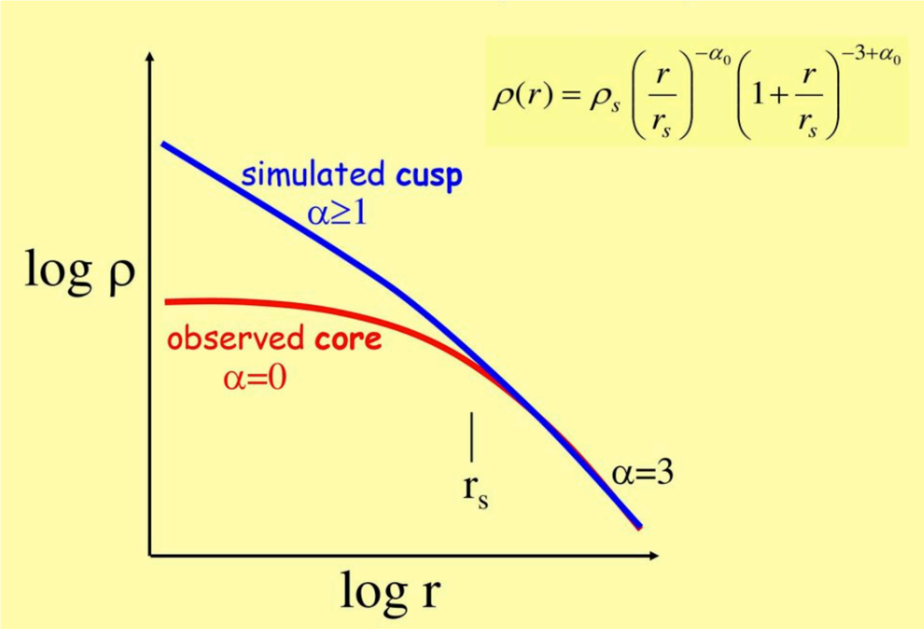
The **mean density of a halo** is: $\bar{\rho} = \frac{1}{V}\int_0^R \rho dV = 3\rho_s \int_0^1 \frac{dx\, x^2}{cx\,(1+cx)^2}$ $\quad x = r/R$

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?

Semi-analytic models (SAM) for galaxy formation

The cosmological model sets the initial conditions for **galaxy formation**.

In particular, the dark matter halos are the cradle of galaxy formation:

density contrast field → large bias → dark mater halo → small bias → astrophysical object (cluster or galaxy)

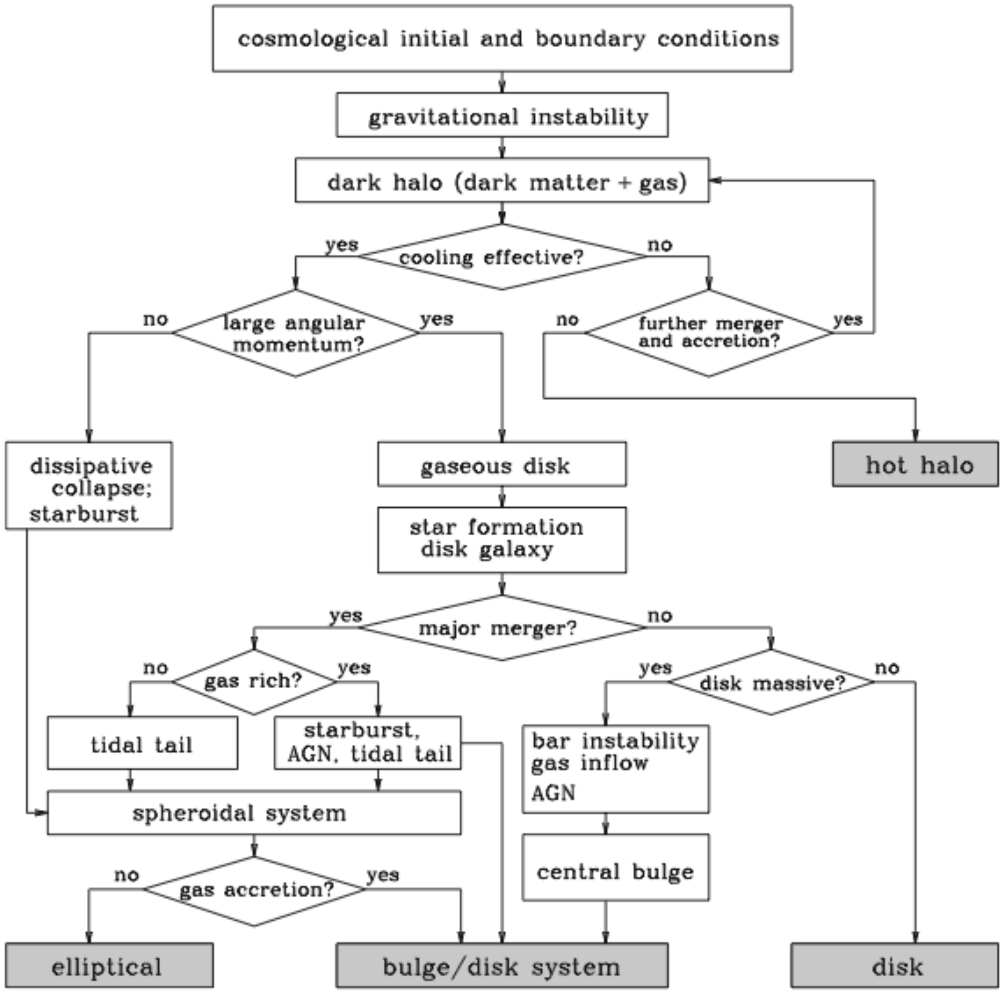
SAM populate dark matter halos with galaxies by considering the baryonic physical processes (also known as **gastrophysics**):

- **Gas cooling** - the gas cools down to condense, losing pressure, and falls into the center of the halo where it can form stars. Angular momentum conservation during the fall produces a disk → **spiral galaxies**

- **Star formation** - if the self-gravity of the gas dominates over the gravity of the dark matter, it collapses under its own gravity, and forms a baryonic collapsed object → **star**

- **Feedback** (from stars and AGNs) - the quantity of cold gas available decreases by influence of the environment

These and other physical processes impact the evolution of the four main baryonic components of a galaxy, which are : **hot gas, cold gas, stars** and a **supermassive black hole**.

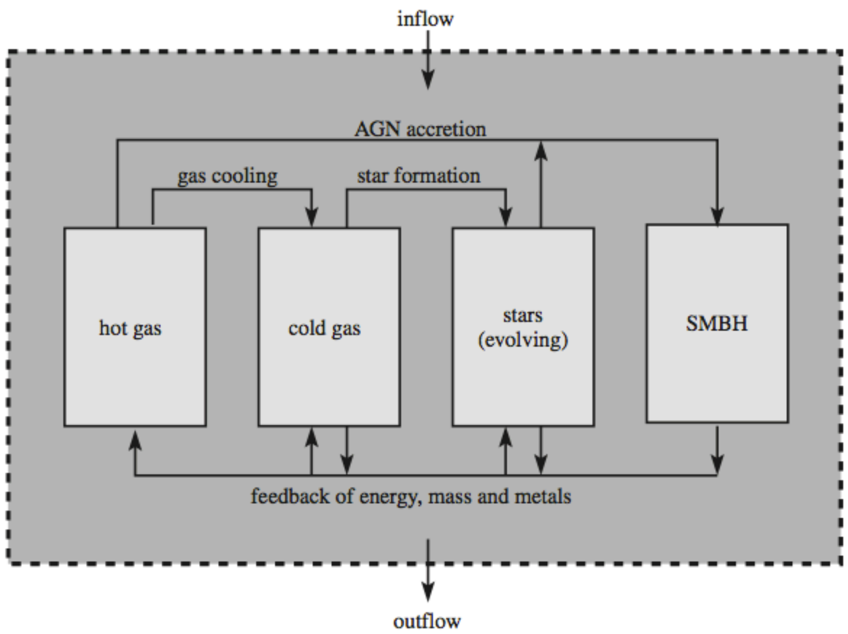


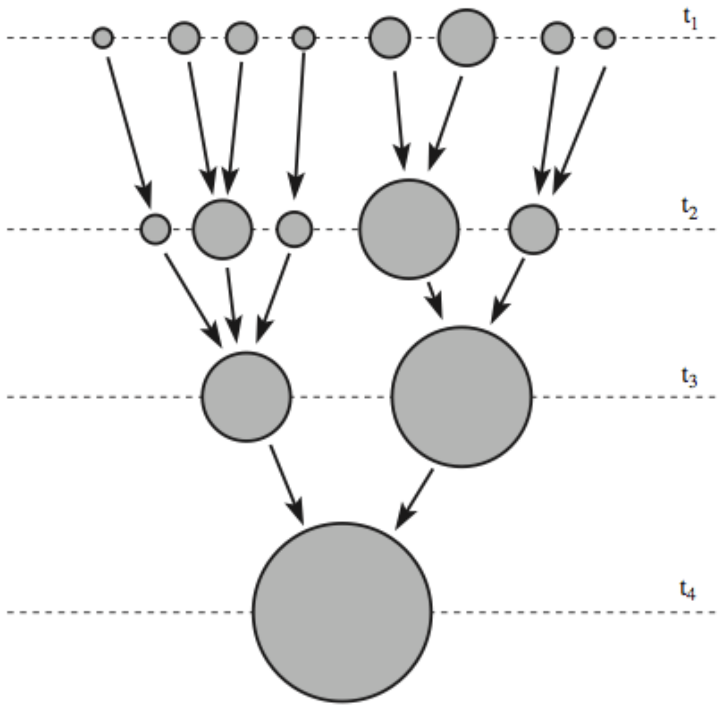
The various mechanisms are interconnected in a complex way

These mechanisms affect the baryonic matter in isolated halos. But the **merging** of halos during the DM evolution also has an important impact in the baryonic evolution

- **Mergers** - frequent interactions between halos lead to → **elliptical galaxies**

So SAM also need to consider the formation history of the dark matter halos and their abundance as function of redshift: the **merger trees**





Hydrodynamic simulations

Instead of populating the dark matter halos with baryonic matter, it is also possible to make N-body simulations from scratch with both dark matter particles and baryonic matter particles.

Baryonic particles, besides interacting with the gravitational potential (like dark matter particles), also experience several **radiative processes**: short-range forces described by **hydrodynamic equations.**

The basic set of equations are: continuity equation, Euler equation and the first law of thermodynamics, that are applied to the baryonic fluid, which is modelled as an ideal monoatomic gas with pressure $p = 2/3\ \rho\ u$

With similar techniques (than in DM N-body simulations) of defining a mesh, and discretizing the differential equations, the evolution of the two types of particles is simulated and

at each position and time step, the properties of the fluid are computed → overdensity, temperature, etc.

This evolution, however, does not model the full process.

The hydrodynamic equations also have to be complemented by the various astrophysical processes (gastrophysics) that shape the galaxy population:

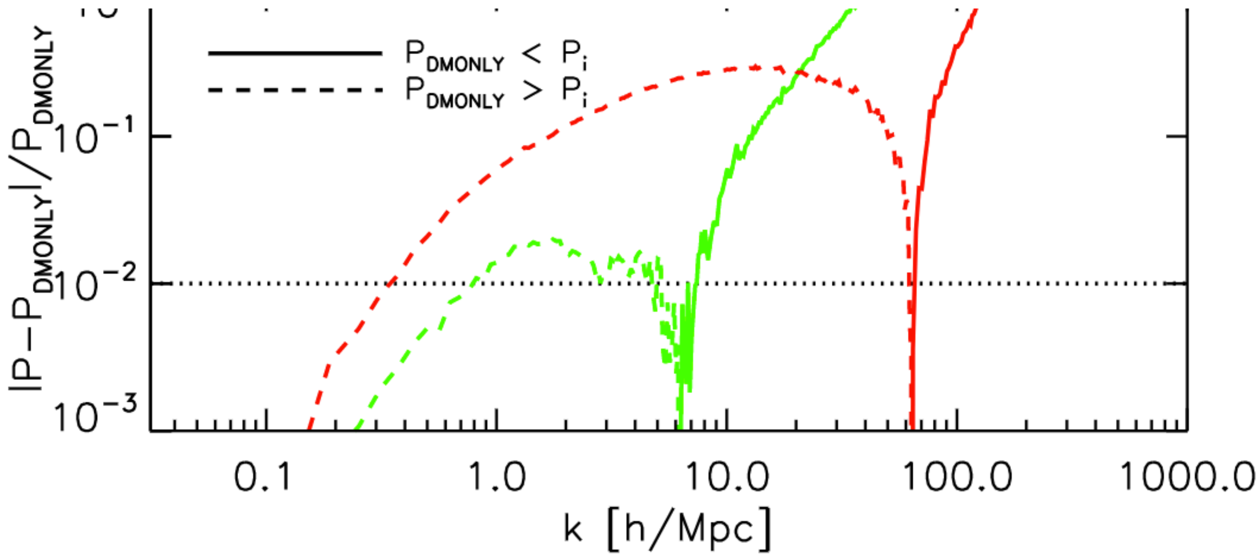
Gas cooling ; Interstellar medium; Magnetic fields; Cosmic rays; Star formation; Stellar feedback; AGN feedback; Supermassive black holes; Thermal conduction; Viscosity; Dust physics

The **result of a hydrodynamic simulation** is the evolution map of many more quantities than just the density → it allows us to explore: galaxy clustering as function of galaxy properties; gas distribution; stellar content of galaxies;

Hydrodynamic simulations also allows us to compute a more correct non-linear matter power spectrum that includes baryonic feedback:

- SN feedback + radiative cooling (green)

- AGN feedback (red)



In general:

heating and gas ejection smooth the density field → decrease of P_δ

radiative cooling helps the baryonic collapse → increase of P_δ

Relativistic N-body simulations

Newtonian dark matter N-body simulations are possible because the gravitational potential is weak on a wide range of scales.

But it becomes strong for $k < k_H$

Relativistic N-body simulations already exist and are useful to:

- compute structure formation on **large scales**

- take into account **horizon effects** relevant for large-scale interactions

- compute structure formation in **modified gravity** models

- take into account **dark energy clustering**

- compute the evolution of other **perturbations besides δ and Φ** → e.g. tensor perturbations (gravitational waves), anisotropic stress potential

This approach is also based on the evolution of a set of N particles, but the particles move along **geodesics** of the metric.

Then, the distribution of particle properties (density, velocity, anisotropic stress) is used to re-compute the energy-momentum tensor projected on a grid.

Then, a new metric is computed, **keeping all the 6 degrees-of-freedom of the perturbed metric:**

$$ds^2 = a^2 \left[-(1+2\Psi)dt^2 + B_i dx^i dt + ((1-2\Phi)\delta_{ij} + h_{ij})dx^i dx^j \right]$$

Notice that in the Newtonian limit B_i = 0 (2 dof), h_{ij} = 0 (2 dof), and Ψ = Φ (1 dof) → in that case there is only 1 metric perturbation (the gravitational potential Φ).

The N-body simulations compute the evolution of all 6 metric perturbations.

Example: resulting power spectra and gravitational waves map in Gevolution