The 30th Jerusalem Winter School in Theoretical Physics

Lecture 1
Non-Linear Structure Formation

Frank van den Bosch
Yale University, Jan 2013

Background Material

Lecture 1 (Mon 31/12)

Structure Formation: from linear to non-linear

Lecture 2 (Tue 1/1)

(Extended) Press-Schechter Theory

Lecture 3 (Wed 9/1)

The Structure of Dark Matter Halos

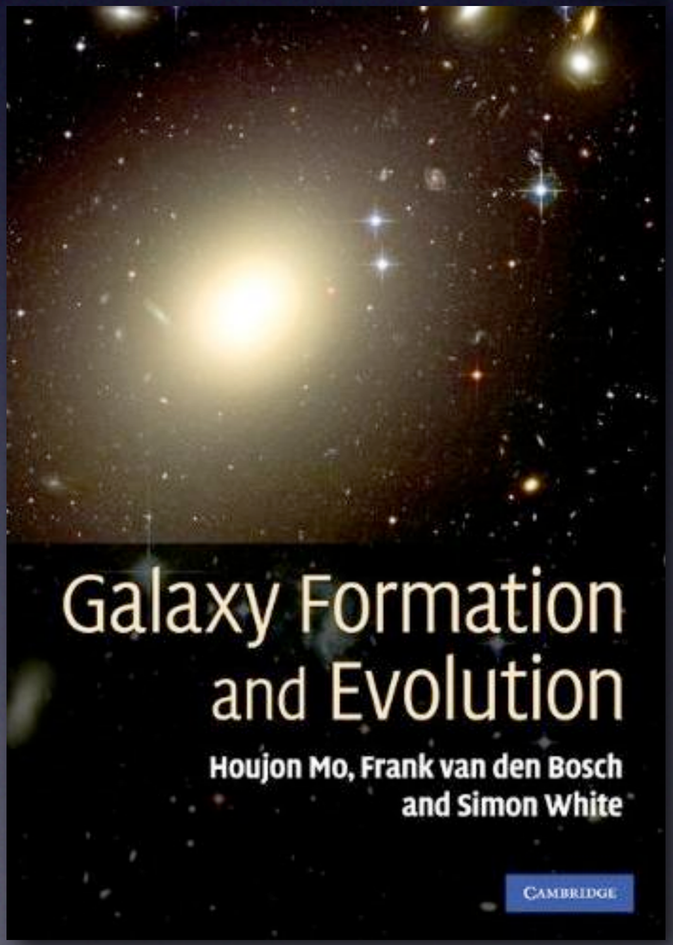
Lecture 4 (Thu 10/1)

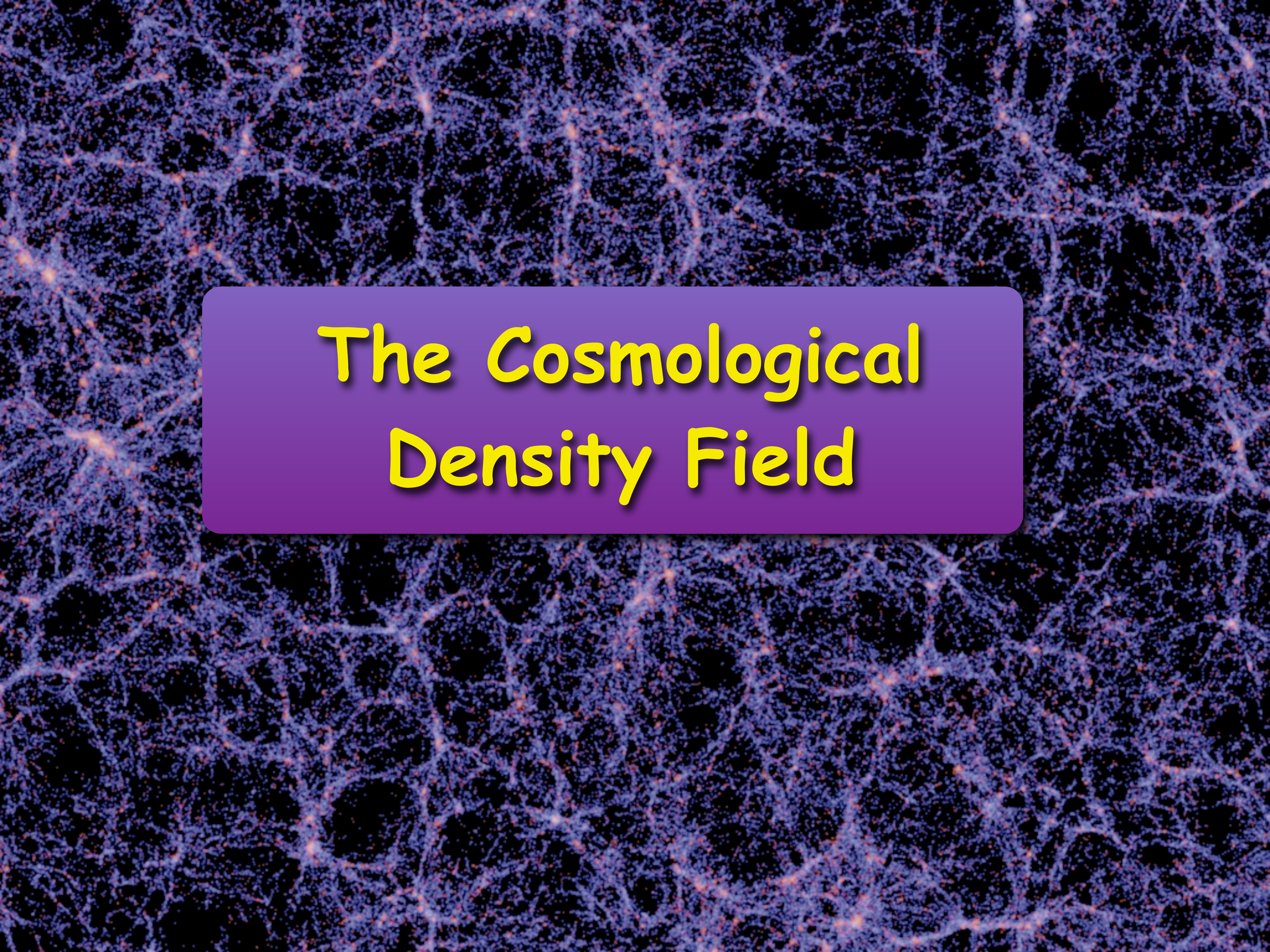
Semi-Analytical Models of Galaxy Formation

The slides of these lectures are available (in PDF) at:

http://www.astro.yale.edu/vdbosch/teaching.html

More detailed treatment of material covered in these lectures can be found in the textbook Galaxy Formation and Evolution available from Cambridge University Press





The Cosmological Density Field

The Cosmological Density Field

How can we describe the cosmological (over)density field, $\delta(\vec{x}, t)$, without having to specify the actual value of δ at each location in space-time, $(\vec{x}, t)$?

Since $\delta(\vec{x})$ is believed to be the outcome of some random process in the early Universe (i.e., quantum fluctuations in inflaton), our goal is to describe the probability distribution

$$\mathcal{P}(\delta_1, \delta_2, ..., \delta_N)\, \mathrm{d}\delta_1\, \mathrm{d}\delta_2 \ldots \mathrm{d}\delta_N$$

where $\delta_1 = \delta(\vec{x}_1)$, etc. For now we will focus on the cosmological density field at some particular (random) time. We will address it's time evolution later in this lecture.

This probability distribution is completely specified by the moments

$$\langle \delta_1^{l_1} \delta_2^{l_2} ... \delta_N^{l_N} \rangle = \int \delta_1^{l_1} \delta_2^{l_2} ... \delta_N^{l_N} \mathcal{P}(\delta_1, \delta_2, ..., \delta_N)\, \mathrm{d}\delta_1\, \mathrm{d}\delta_2 \ldots \mathrm{d}\delta_N$$

NOTE: $\langle ... \rangle$ denotes an ensemble average. For instance, $\langle \delta(\vec{x}) \rangle$ means the average overdensity at $\vec{x}$ for many realizations of the random process...

The Ergodic Hypothesis

PROBLEM: Theory specifies ensemble average, but observationally we have only access to one realization of the random process....

Ergodic Hypothesis: Ensemble average is equal to spatial average taken over one realization of the random field...

First Moment

$$\langle\delta\rangle = \int \delta\,\mathcal{P}(\delta)\,\mathrm{d}\delta = \frac{1}{V}\int_V \delta(\vec{x})\,\mathrm{d}^3\vec{x} = 0$$

Essentially, the ergodic hypothesis requires spatial correlations to decay sufficiently rapidly with increasing separation so that there exists many statistically independent volumes in one realization....

QUESTION: How many moments do we need to completely specify the cosmological density field?

In principle infinitely many. However, there are good reasons to believe that the initial cosmological density field is special, in that it is a Gaussian random field...

Gaussian Random Fields

A random field $\delta(\vec{x})$ is said to be Gaussian if the distribution of the field values at an arbitrary set of N points is an N-variate Gaussian:

$$\mathcal{P}(\delta_1, \delta_2, ..., \delta_N) = \frac{\exp(-Q)}{[(2\pi)^N \det(\mathcal{C})]^{1/2}}$$

$$Q \equiv \frac{1}{2} \sum_{i,j} \delta_i \, (\mathcal{C}^{-1})_{ij} \delta_j$$

$$\mathcal{C}_{ij} = \langle \delta_i \delta_j \rangle \equiv \xi(r_{ij})$$

where we have defined the two-point correlation function $\xi(\vec{r}) = \langle \delta(\vec{x}) \, \delta(\vec{x} + \vec{r}) \rangle$

As you can see, for Gaussian random field the N-point probability function $\mathcal{P}(\delta_1, \delta_2, ..., \delta_N)$ is completely specified by the two-point correlation function.

In particular, the one-point distribution function of the field is

$$\mathcal{P}(\delta) \, \mathrm{d}\delta = \frac{1}{\sqrt{2\pi} \, \sigma} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) \mathrm{d}\delta$$

where $\sigma^2 = \langle \delta^2 \rangle = \xi(0)$ is the variance of the density perturbation field.

The two-point correlation function

Second Moment

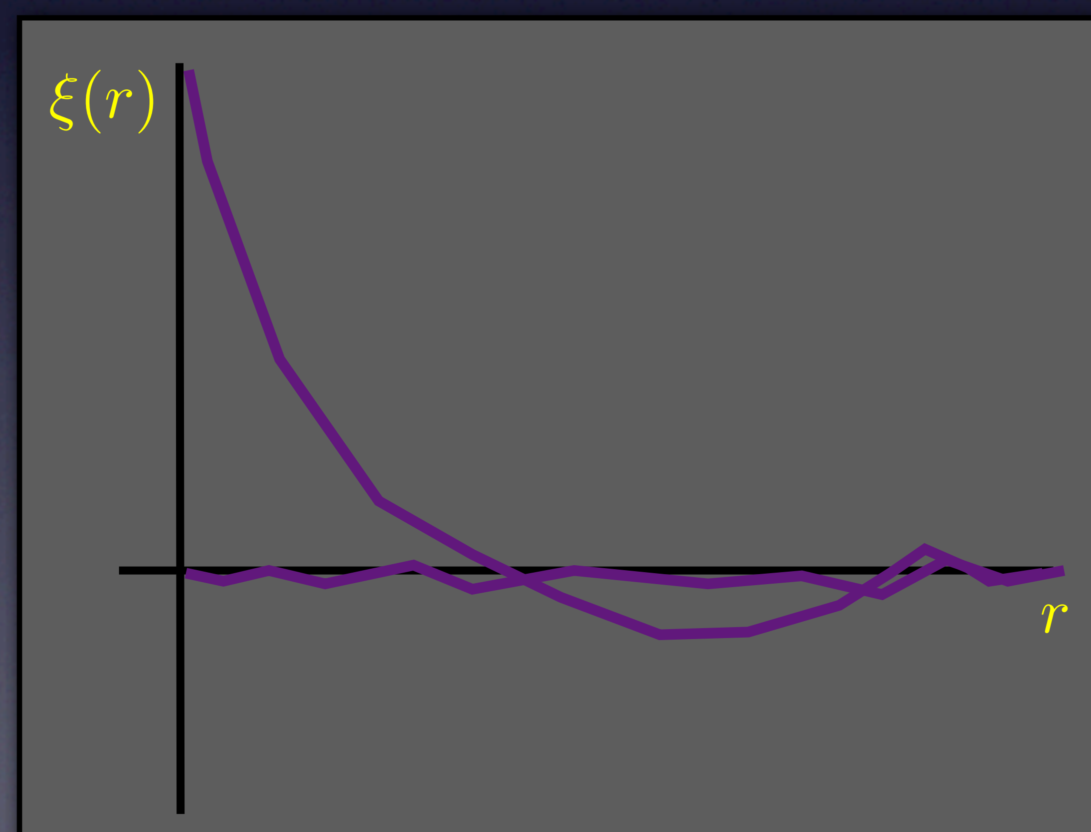
$$\langle \delta_1 \, \delta_2 \rangle \equiv \xi(r_{12}) \qquad r_{12} = |\vec{x}_1 - \vec{x}_2|$$

$\xi(r)$ is called the two-point correlation function

Note that this two-point correlation function is defined for a continuous field, $\delta(\vec{x})$. However, one can also define it for a point distribution:

$$1 + \xi(r) = \frac{n_{\rm pair}(r \pm {\rm d}r)}{n_{\rm random}(r \pm {\rm d}r)}$$

Poisson distribution / Clustered distribution



Higher-Order Correlation Functions

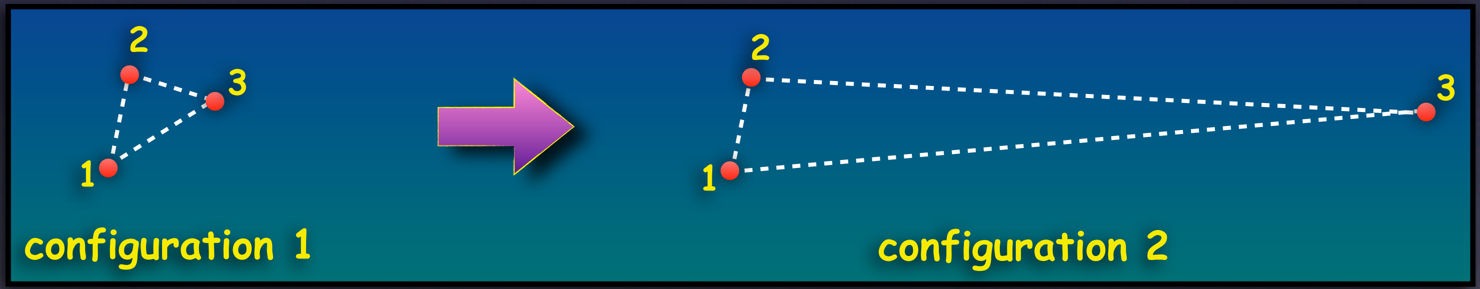
The n-point correlation function is defined as $\xi^{(n)} \equiv \langle \delta_1 \, \delta_2 \, ... \, \delta_n \rangle$

The reduced (or irreducible) n-point correlation function is defined as

$$\xi^{(n)}_{\rm red} \equiv \langle \delta_1 \, \delta_2 \, ... \, \delta_n \rangle_{\rm c}$$

where $\langle ... \rangle_{\rm c}$ is the cumulant or connected moment.

These reduced (or irreducible) correlation functions express the part of the n-point correlation functions that cannot be obtained from lower-order reduced correlation functions:



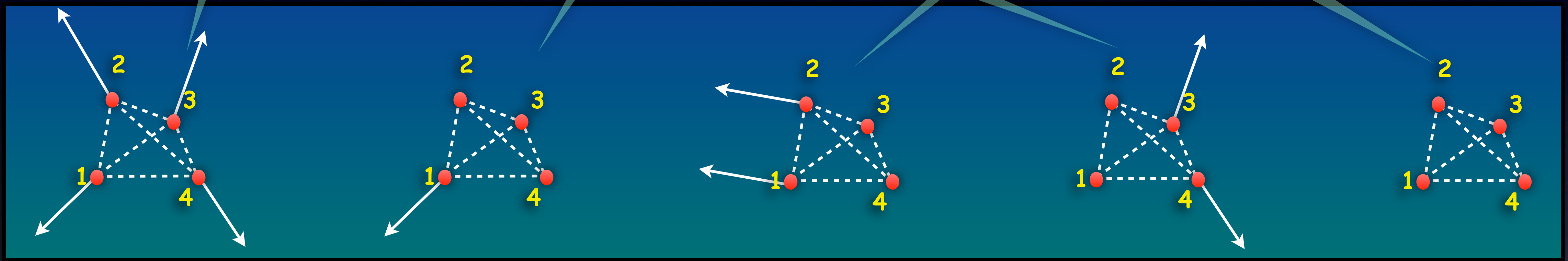
In the limit where r_{13} goes to infinity, the correlation between the three points in configuration 2 is entirely due to that between points 1 and 2. The reduced correlation function subtracts the correlations due to these configurations from the total correlation function. ...

Higher-Order Correlation Functions

As an example, consider the four point correlation function:

$$\langle\delta_1\delta_2\delta_3\delta_4\rangle = \langle\delta_1\rangle_c\,\langle\delta_2\rangle_c\,\langle\delta_3\rangle_c\,\langle\delta_4\rangle_c + \langle\delta_1\rangle_c\,\langle\delta_2\delta_3\delta_4\rangle_c\,(4\,\text{terms}) + \langle\delta_1\delta_2\rangle_c\,\langle\delta_3\delta_4\rangle_c\,(3\,\text{terms})$$
$$+\langle\delta_1\delta_2\rangle_c\,\langle\delta_3\rangle_c\,\langle\delta_4\rangle_c\,(6\,\text{terms}) + \langle\delta_1\delta_2\delta_3\delta_4\rangle_c$$

Using similar diagrams we can understand the origin of each of these terms



Here ●⟶ means: "this point moving to infinity"

Since $\langle\delta\rangle_c = \langle\delta\rangle = 0$ we have that

$$\langle\delta_1\delta_2\delta_3\delta_4\rangle_c = \xi^{(4)}_{1234} - \xi^{(2)}_{12}\,\xi^{(2)}_{34} - \xi^{(2)}_{13}\,\xi^{(2)}_{24} - \xi^{(2)}_{14}\,\xi^{(2)}_{23}$$

For a Gaussian random field, all connected moments (=reduced correlation functions) of $n > 2$ are equal to zero: ⇨ One can use higher-order reduced correlation functions to test whether a density field is Gaussian or not...

The Power Spectrum

Often it is very useful to describe the matter field in Fourier space:

$$\delta(\vec{x}) = \sum_k \delta_{\vec{k}}\, e^{+i\vec{k}\cdot\vec{x}} \qquad \delta_{\vec{k}} = \frac{1}{V}\int \delta(\vec{x})\, e^{-i\vec{k}\cdot\vec{x}}\, \mathrm{d}^3\vec{x}$$

Here V is the volume over which the Universe is assumed to be periodic.

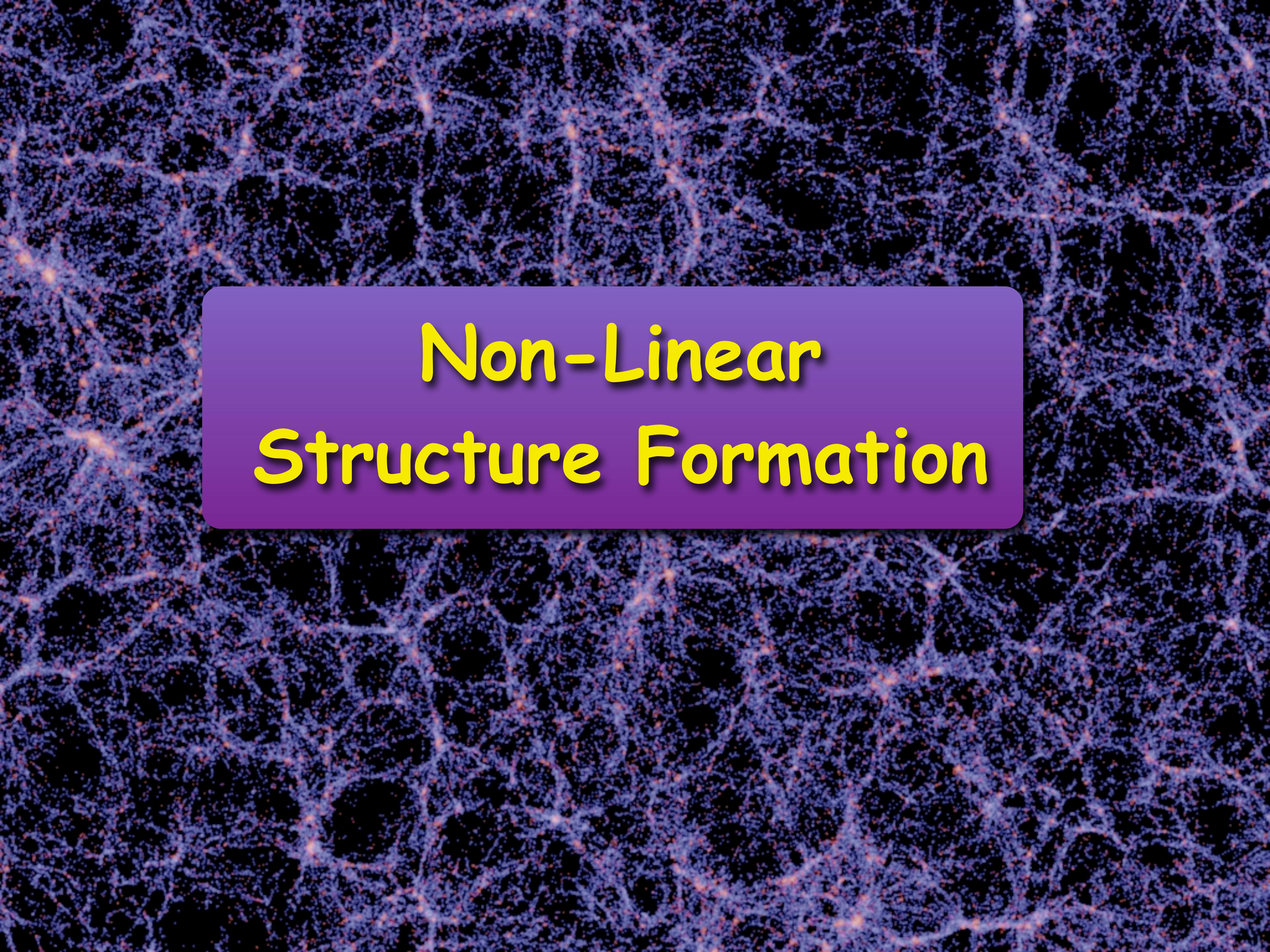
Note: the perturbed density field can be written as a sum of plane waves of different wave numbers k (called `modes')

The Fourier transform (FT) of the two-point correlation function is called the power spectrum and is given by

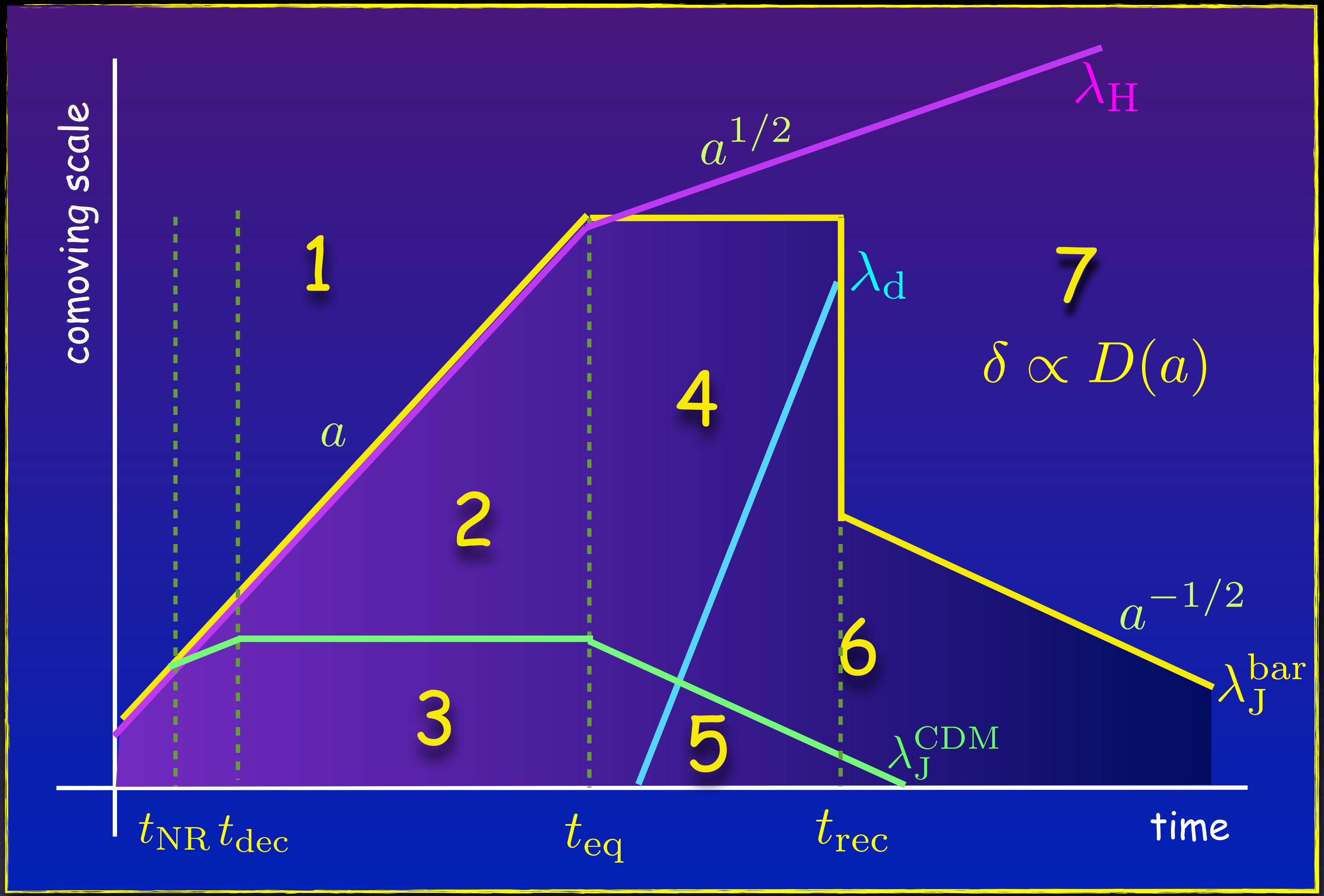
$$\begin{aligned} P(\vec{k}) &\equiv V\langle|\delta_{\vec{k}}|^2\rangle \\ &= \int \xi(\vec{x})\, e^{-i\vec{k}\cdot\vec{x}}\, \mathrm{d}^3\vec{x} \\ &= 4\pi \int \xi(r)\, \frac{\sin kr}{kr}\, r^2\, \mathrm{d}r \end{aligned}$$

Note: P(k) has units of volume!

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


Non-Linear
Structure Formation

Linear Perturbation Growth in a Nutshell



	Baryons	CDM
1	growth	growth
2	oscillations	stagnation
3	oscillations	damping
4	oscillations	growth
5	damping	damping
6	damping	growth
7	growth	growth

After recombination the growth of linear perturbations on our scales of interest ($10^6 M_\odot < M < M^{15} M_\odot$) is governed by the linear growth rate; $D(a)$

In this linear regime, all modes k evolve similarly and independently: $\delta_{\vec{k}} \propto D(a)$

$$P(k,t) = P_{\rm i}(k)\, T^2(k)\, D^2(t)$$

Once perturbations become of order unity, structure formation becomes non-linear....

Non-Linear Evolution

In the linear regime ($\delta \ll 1$) we can calculate the evolution of a density field of arbitrary form using linear perturbation theory.

In the non-linear regime ($\delta > 1$) 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...

Because of this mode-coupling, the density field looses 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.

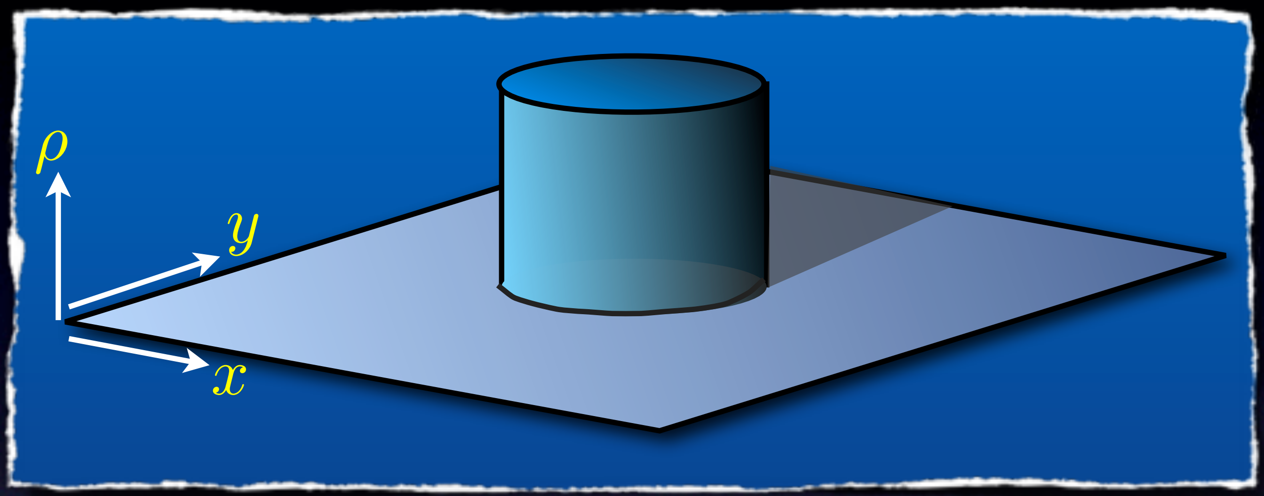
How to proceed?

- Oversimplified, but insightful, analytical model (this lecture)
- Higher-order perturbation theory (see MBW §4.1.7)
- Numerical simulations (see MBW §5.6.2)
- The Halo Model (see MBW §7.6)

Top-Hat Spherical Collapse

In order to gain insight into the non-linear evolution of density perturbations, we now consider the highly idealized case of Top-Hat Spherical Collapse.

- Universe is homogeneous, except for a single, top-hat, spherical perturbation.
- Universe is in matter-dominated phase, after recombination...
- Collisionless fluid ⇨ treatment is only valid for collisionless Dark Matter.
- Einstein-de Sitter (EdS) cosmology



$$\Omega_{\rm m}(t) = 1 \qquad H(t)\cdot t = \frac{2}{3}$$

$$\bar{\rho} = \frac{1}{6\pi G t^2} \qquad D(a) = a \propto t^{2/3}$$

NOTE:

Although the following treatment is only valid for an EdS cosmology, similar models can be constructed for other cosmologies as well, including ΛCDM (see MBW §5.1.1 + 5.1.2)

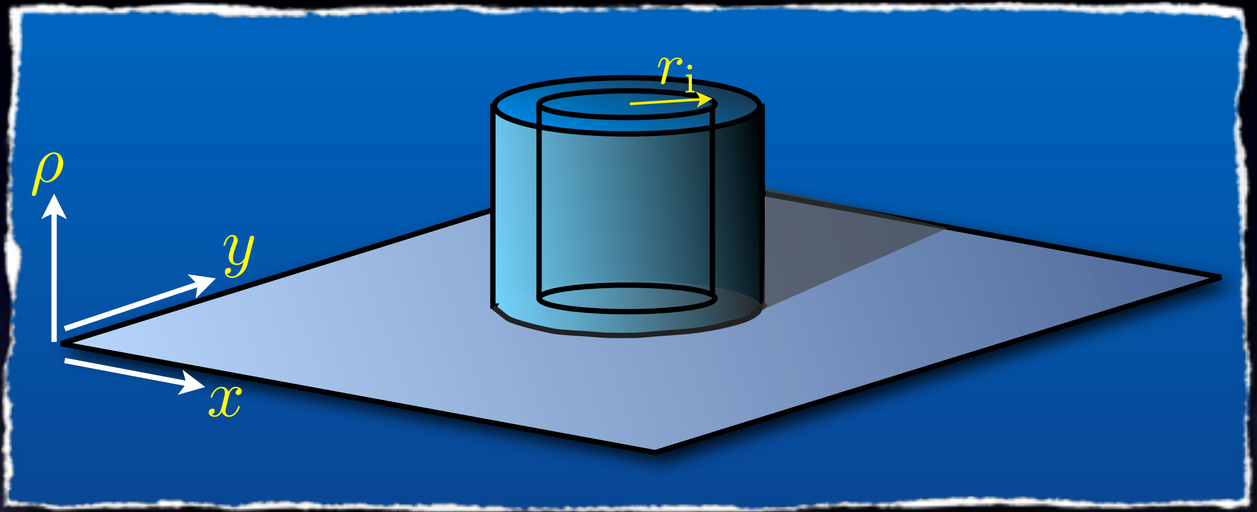
Furthermore, since all cosmologies behave similar to EdS at early times, this treatment is always good approximation at high z....

Top-Hat Spherical Collapse

Consider our spherical top-hat perturbation: Let r_i denote the radius of some mass shell inside the top-hat at some initial time, t_i, and let δ_i and $\bar{\rho}_\mathrm{i}$ denote the top-hat overdensity and the back-ground density at that same time.

The mass enclosed by the shell is

$$M(< r) = \frac{4}{3}\pi r_\mathrm{i}^3\,\bar{\rho}_\mathrm{i}\,[1+\delta_\mathrm{i}]$$
$$= \frac{4}{3}\pi r^3(t)\,\bar{\rho}(t)\,[1+\delta(t)]$$



where the second equality expresses mass conservation: because of spherical symmetry, the mass inside the shell is conserved, but only up to shell crossing !!!

Newton's first Theorem:

a spherically symmetric matter distribution outside a sphere exerts no force on that sphere

→

Equation of motion

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{GM}{r^2}$$

Top-Hat Spherical Collapse

Integrating the equation of motion once yields

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 - \frac{GM}{r} = E$$

where the integration constant E is clearly the specific energy of our shell.

For $E < 0$, mass shell is bound, and solution can be written in following parametric form:

$$r = A\,(1 - \cos\theta) \qquad \theta \in [0, 2\pi]$$
$$t = B\,(\theta - \sin\theta)$$

$$A = \frac{GM}{2|E|} \qquad B = \frac{GM}{(2|E|)^{3/2}} \quad \Rightarrow \quad A^3 = GMB^2$$

This solution implies the following evolution for our mass shell:

- shell expands from $r = 0$ at $\theta = 0$ $(t = 0)$
- shell reaches a maximum radius $r_{\rm max}$ at $\theta = \pi$ $(t = t_{\rm max} = \pi B)$
- shell collapses back to $r = 0$ at $\theta = 2\pi$ $(t = t_{\rm coll} = 2t_{\rm max})$

The time of maximum size is often called the turn-around time, $t_{\rm ta} = t_{\rm max}$, while the time of collapse is also called the virialization time $t_{\rm vir} = t_{\rm coll} = 2t_{\rm ta}$

Top-Hat Spherical Collapse

Now let us focus on the evolution of the actual overdensity:

The mean density of the top-hat is $\rho = \frac{3M}{4\pi r^3} = \frac{3M}{4\pi A^3}(1-\cos\theta)^{-3}$

The mean density of the background is $\bar{\rho} = \frac{1}{6\pi G t^2} = \frac{1}{6\pi G B^2}(\theta - \sin\theta)^{-2}$

Hence, the actual overdensity of our spherical top-hat region, according to the spherical collapse (SC) model, which in general will be non-linear, is

$$1 + \delta = \frac{\rho}{\bar{\rho}} = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1-\cos\theta)^3}$$

where we have used that $A^3 = GMB^2$.

Before we examine this SC model in some detail, we first compare it to predictions from linear theory....

Top-Hat Spherical Collapse

For a number of reasons (in particular for use in EPS theory), it is also useful to compare this SC overdensity model to what linear theory predicts for $\delta(t)$.

According to linear theory, perturbation in EdS cosmology evolve as

$$\delta_{\rm lin} \propto D(a) \propto a \propto t^{2/3}$$

In order to use the correct initial conditions (ICs), we have to use our parametric solution of $r(t)$ in the limit $\theta \ll 1$. Using a Taylor series expansion of $\sin\theta$ and $\cos\theta$ one can show that:

$$\delta_{\rm i} = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t_{\rm i}}{t_{\rm max}}\right)^{2/3} \qquad (\delta_{\rm i} \ll 1)$$

NOTE: this implies that since $\delta(r) =$ constant inside the top-hat, each mass shell that is part of the top-hat will turn-around (reach maximum expansion) at the same time....

Top-Hat Spherical Collapse

For a number of reasons (in particular for use in EPS theory), it is also useful to compare this SC overdensity model to what linear theory predicts for $\delta(t)$.

According to linear theory, perturbation in EdS cosmology evolve as

$$\delta_{\rm lin} \propto D(a) \propto a \propto t^{2/3}$$

In order to use the correct initial conditions (ICs), we have to use our parametric solution of $r(t)$ in the limit $\theta \ll 1$. Using a Taylor series expansion of $\sin\theta$ and $\cos\theta$ one can show that:

$$\delta_{\rm i} = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t_{\rm i}}{t_{\rm max}}\right)^{2/3} \qquad (\delta_{\rm i} \ll 1)$$

Combining the above, we have that, according to linear theory:

$$\delta_{\rm lin} = \delta_{\rm i}\left(\frac{t}{t_{\rm i}}\right)^{2/3} = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_{\rm max}}\right)^{2/3}$$

Turn-Around & Collapse

Spherical Collapse (SC) model:

$$1 + \delta = \frac{\rho}{\bar{\rho}} = \frac{9}{2} \frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}$$

Linear Theory

$$\delta_{\rm lin} = \delta_{\rm i} \left(\frac{t}{t_{\rm i}}\right)^{2/3} = \frac{3}{20}(6\pi)^{2/3} \left(\frac{t}{t_{\rm max}}\right)^{2/3}$$

Turn-Around: $(t_{\rm ta} = t_{\rm max}; \theta = \pi)$

SC model: $1 + \delta(t_{\rm ta}) = \frac{9\pi^2}{16} \simeq 5.55$

linear theory: $\delta_{\rm lin}(t_{\rm ta}) = \frac{3}{20}(6\pi)^{2/3} \simeq 1.062$

Collapse (shell crossing) $(t_{\rm coll} = 2t_{\rm ta})$

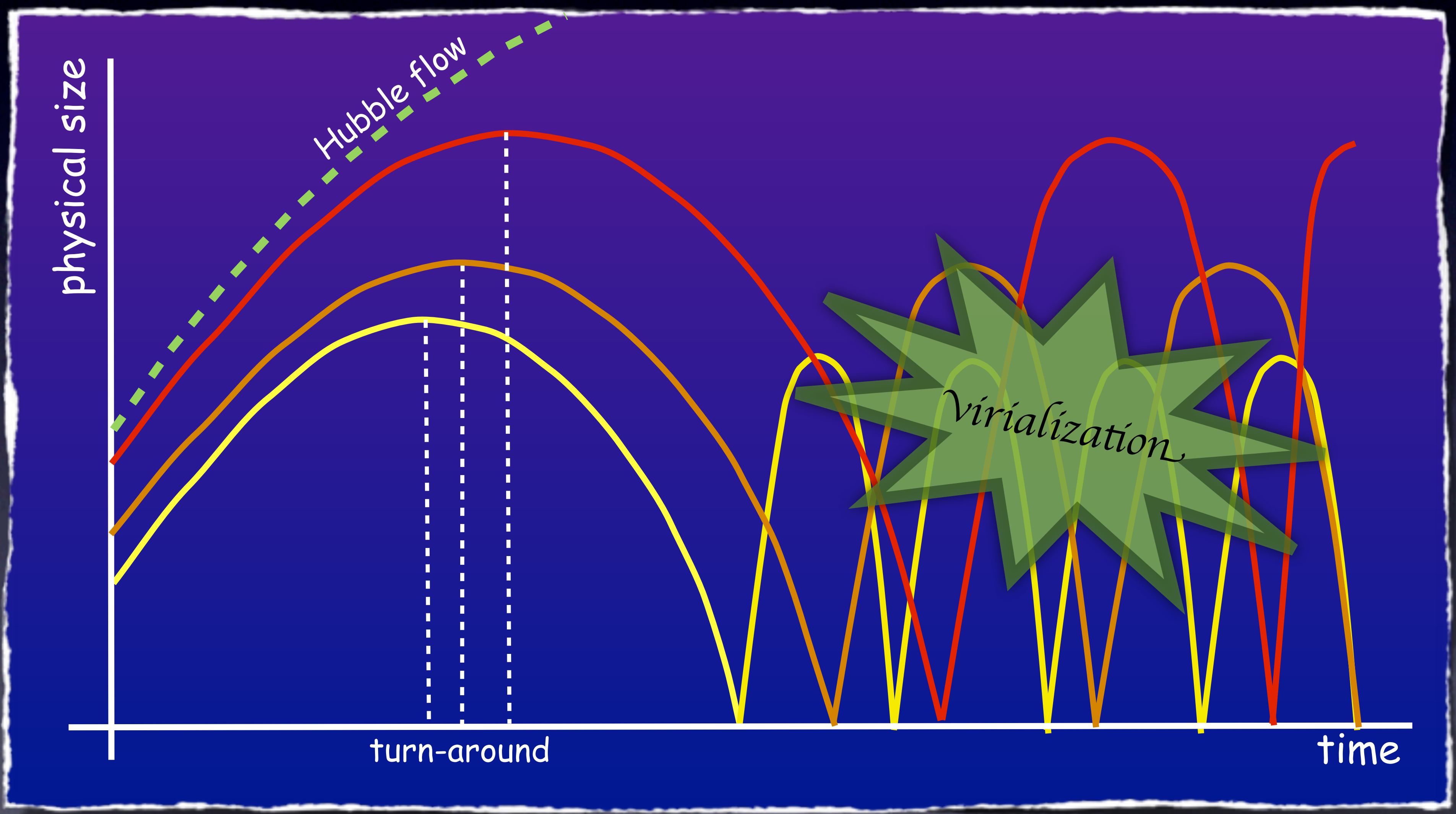
SC model: $\delta(t_{\rm coll}) = \infty$

linear theory: $\delta(t_{\rm coll}) = \frac{3}{20}(12\pi)^{2/3} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \simeq 1.686$

Shell Crossing & Virialization

The SC model discussed above is only valid up to the point of shell crossing. Afterall, after shell crossing M(r) is no longer a conserved quantity!

According to the SC model, $\delta(t_{\rm coll}) = \infty$, which would result in the formation of a black hole. However, in reality, the collapse is never perfectly spherical.



Individual oscillating shells interact gravitationally, exchanging energy (virializing). This process, to be described in more detail below, results in a virialized dark matter halo

Final Density of a Collapsed Dark Matter Halo

Virialization means that the system relaxes towards virial equilibrium:

We can use the virial theorem to make a simple estimate of the final density of our collapsed & virialized dark matter halo:

Virial Equilibrium: $2\,K_{\rm f} + W_{\rm f} = 0$

Energy conservation: $E_{\rm f} = K_{\rm f} + W_{\rm f} = E_{\rm i} = E_{\rm ta}$

$$E_{\rm ta} = W_{\rm ta} = -\frac{GM}{r_{\rm ta}}$$

$$E_{\rm f} = W_{\rm f}/2 = -\frac{GM}{2r_{\rm vir}}$$

$$\Rightarrow \quad r_{\rm vir} = r_{\rm ta}/2$$

A mass shell is expected to virialize at half its turn-around radius.

Hence, after virialization, the average density of the material enclosed by the mass shell is 8 times denser than at turn-around....

Final Density of a Collapsed Dark Matter Halo

We now compute the average overdensity of a virialized dark matter halo:

$$1 + \Delta_{\rm vir} \equiv 1 + \delta(t_{\rm coll}) = \frac{\rho(t_{\rm coll})}{\bar{\rho}(t_{\rm coll})}$$

NOTE: for consistency with many textbooks and journal articles, we use the symbol $\Delta_{\rm vir}$, rather than $\delta_{\rm vir}$ to indicate the virialized overdensity....

Using that $\bar{\rho} \propto a^{-3} \propto t^{-2}$ (EdS), and that $t_{\rm coll} = 2 t_{\rm ta}$ we have that

$$1 + \Delta_{\rm vir} = \frac{8\,\rho_{\rm ta}}{\bar{\rho}(t_{\rm ta})/4} = 32\,(1 + \delta_{\rm ta}) = 18\pi^2 \simeq 178$$

For non-EdS cosmologies, the virial overdensities are well approximated by

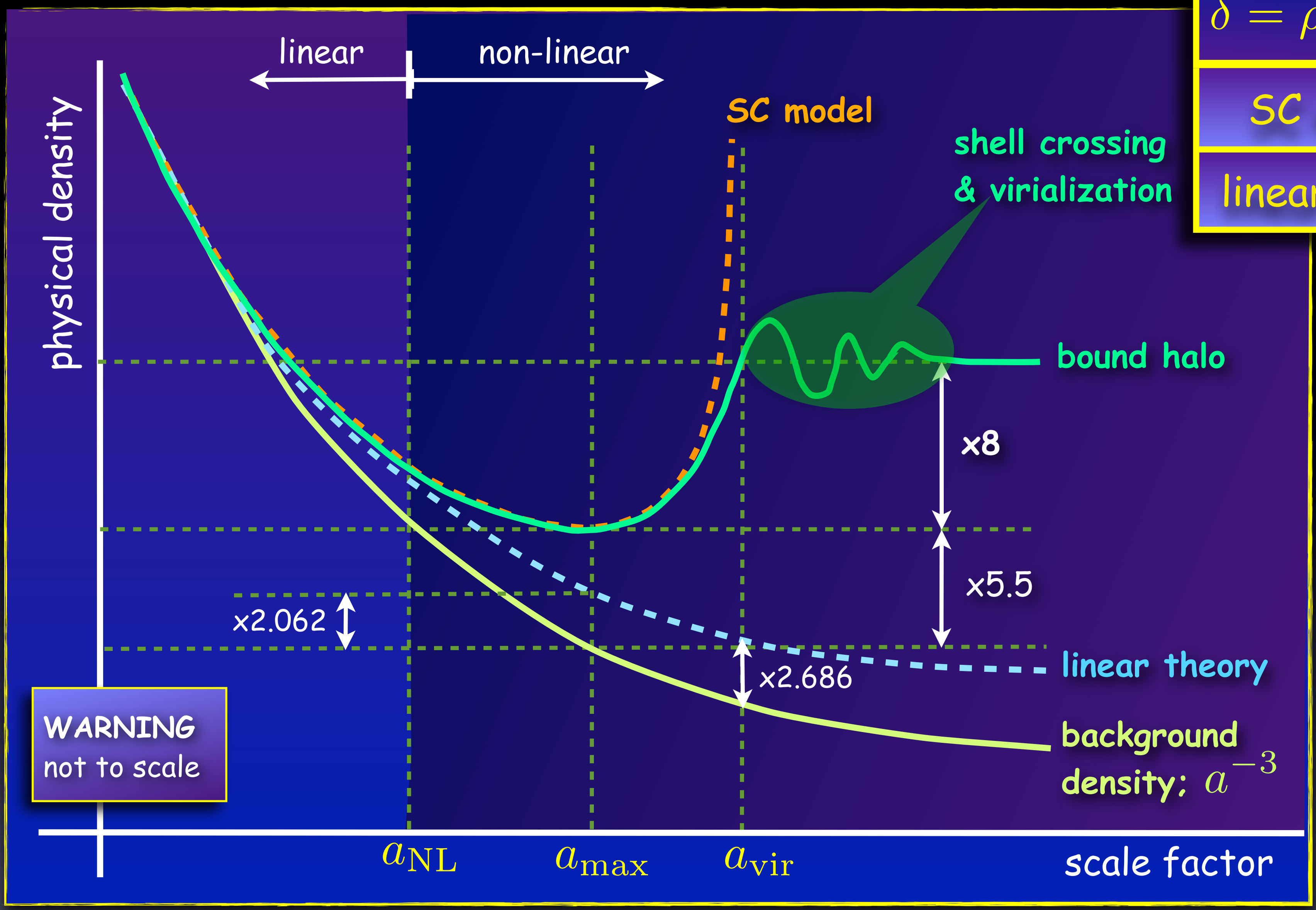
$$\Delta_{\rm vir} \approx (18\pi^2 + 60\,x - 32\,x^2)/\Omega_{\rm m}(t_{\rm vir}) \qquad (\Omega_\Lambda = 0)$$

$$\Delta_{\rm vir} \approx (18\pi^2 + 82\,x - 39\,x^2)/\Omega_{\rm m}(t_{\rm vir}) \qquad (\Omega_\Lambda \neq 0)$$

(Bryan & Norman 1998)

Here $x = \Omega_{\rm m}(t_{\rm vir}) - 1$. These equations are often used to `define' dark matter haloes in N-body simulations or in analytical models....

Summary: The Spherical Collapse (SC) Model



$\delta = \rho/\bar{\rho} - 1$	turn-around	collapse
SC model	4.55	∞
linear model	1.062	1.686

Although SC model becomes inaccurate (brakes down) shortly after turn-around it is still useful for identifying important epochs in linearly evolved density field...

The linearly extrapolated density field collapses when $\delta_{\rm lin} = \delta_{\rm c} \simeq 1.686$

Virialized dark matter haloes have an average overdensity of $\Delta_{\rm vir} \simeq 178$

The Zel'dovich Approximation

So far we considered perturbations in Eulerian (`grid') coordinates. Individual overdensities stay at a fixed (comoving) position and grow or decay in amplitude....

We now switch to Lagrangian description, which follows motion of individual particles. This gives insights into dynamics of structure formation process, and, unlike its Eulerian counterpart, remains (fairly) accurate in the mildly non-linear regime...

It is easy to see that Eulerian description brakes down in mildly non-linear regime: Once overdensities ($\delta_{\rm i} > 0$) reach amplitudes of order unity, the underdensities ($\delta_{\rm i} < 0$) have grown to $\delta < -1$, which would imply a negative (=unphysical) density...

Zel'dovich (1970) came up with a Lagrangian formalism that is based on the following approximation (known as Zel'dovich Approximation, ZA):

particles continue to move in the direction of their initial displacement

$$\Rightarrow \quad \vec{x}(t) = \vec{x}_{\rm i} - c(t) \cdot \vec{f}(\vec{x}_{\rm i})$$

$\vec{x}_{\rm i}$ initial (Lagrangian), comoving coordinates

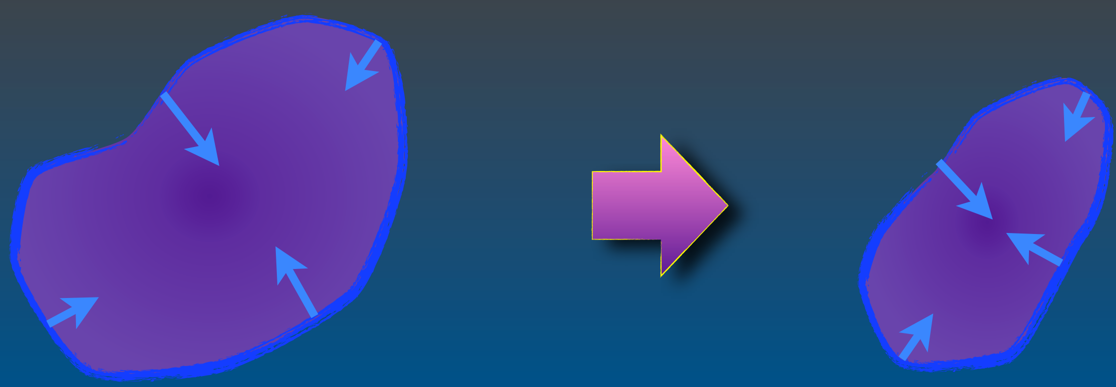
$c(t)$ function of time, to be determined below

$\vec{f}(\vec{x}_{\rm i})$ vector function of initial coordinates, specifying direction of velocity

The Zel'dovich Approximation

ZA: $\vec{x}(t) = \vec{x}_{\rm i} - c(t) \cdot \vec{f}(\vec{x}_{\rm i})$

Note: the ZA is exact if perturbation is a 1D sheet in an otherwise homogeneous universe; in that case direction of velocity remains fixed...



Mass conservation

$$\rho(\vec{x},t)\, a^3(t)\, {\rm d}^3\vec{x} = \rho_{\rm i}(\vec{x}_{\rm i})\, {\rm d}^3\vec{x}_{\rm i}$$

Here $a(t)$ is the scale-factor normalized to unity at the initial time $t_{\rm i}$: the scaling with $a^3(t)$ is required since $\vec{x}$ are comoving coordinates. The equation of mass conservation is valid (up to orbit crossing) for any geometry; no spherical symmetry is required!!

Using Linear Algebra:

$$\rho(\vec{x},t) = \rho_{\rm i}(\vec{x}_{\rm i})\, a^{-3} \left\| \frac{{\rm d}\vec{x}}{{\rm d}\vec{x}_{\rm i}} \right\|^{-1}$$

Here $\|A\| = \det(A) = \prod_i \lambda_i$ with λ_i the eigenvectors of the matrix A

The Zel'dovich Approximation

Using that the tensor $\left(\frac{\mathrm{d}\vec{x}}{\mathrm{d}\vec{x}_\mathrm{i}}\right)_{jk} = \delta_{jk} - c(t)\frac{\partial f_j}{\partial x_k}$ we have that

$$\rho(\vec{x},t) = \rho_\mathrm{i}(\vec{x}_\mathrm{i})\, a^{-3} \frac{1}{(1-c\lambda_1)\,(1-c\lambda_2)\,(1-c\lambda_3)}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the eigenvalues of the deformation tensor $\partial f_i / \partial x_j$

Using that $\rho_\mathrm{i}(\vec{x}_\mathrm{i}) = \bar{\rho}_\mathrm{i}\,[1 + \delta_\mathrm{i}(\vec{x}_\mathrm{i})] \simeq \bar{\rho}_\mathrm{i}$ and that $\bar{\rho}(t)\, a^3 = \bar{\rho}_\mathrm{i}\, a_\mathrm{i}^3$ this yields

$$1 + \delta(\vec{x},t) = \frac{\rho(\vec{x},t)}{\bar{\rho}(t)} = \frac{1}{(1-c\lambda_1)\,(1-c\lambda_2)\,(1-c\lambda_3)}$$

(recall that $a_i = 1$)

We can gain some useful insight from this equation (using that $c(t) > 0$) :

- if $\lambda_i > 0$ this implies collapse in the direction of the i^{th} eigenvector.
- if $\lambda_i < 0$ this implies expansion in the direction of the i^{th} eigenvector.
- if $c(t) = 1/\lambda_i$ `shell' crossing happens along the direction of the i^{th} eigenvector.
- as long as $c\lambda_1 \ll 1$ the perturbation is still in the linear regime.

The Zel'dovich Approximation

Linearization of the equation for the density perturbation yields

$$1 + \delta(\vec{x},t) = \frac{1}{1 - c\,(\lambda_1 + \lambda_2 + \lambda_3)} \simeq 1 + c\,(\lambda_1 + \lambda_2 + \lambda_3)$$

Hence, we have that, in the linear regime $\delta(\vec{x},t) = c(t)\,\mathrm{Tr}(\partial f_i/\partial x_j) = c(t)\,\vec{\nabla}\cdot\vec{f}$

If we compare this to the fact that, in the linear regime, $\delta(\vec{x},t) = D(t)\,\delta_{\rm i}(\vec{x}_{\rm i})$ we see that $c(t) = D(t)$ and $\vec{\nabla}\cdot\vec{f} = \delta_{\rm i}$.

Using the Poisson equation, according to which $\delta_{\rm i} = \nabla^2\Phi_{\rm i}/4\pi G\bar{\rho}_{\rm i}$ (recall that $a_{\rm i} = 1$) and the fact that $\nabla^2\Phi = \vec{\nabla}\cdot\vec{\nabla}\Phi$, we finally see that $\vec{f} = \vec{\nabla}\Phi_{\rm i}/4\pi G\bar{\rho}_{\rm i}$

$$\vec{x}(t) = \vec{x}_{\rm i} - \frac{D(a)}{4\pi G\bar{\rho}_{\rm i}}\,\vec{\nabla}\Phi_{\rm i}$$

Zel'dovich Approximation

Zel'dovich Pancakes

The ZA describes the non-linear evolution of density perturbations. It has two important advantages over the spherical collapse model:

- it makes no oversimplified assumptions about geometry
- it remains accurate well into the quasi-linear regime

To understand why the ZA is more accurate in the quasi-linear regime (brakes down at a later stage), have a look at its predicted evolution for an overdensity:

$$1 + \delta(\vec{x}, t) = \frac{\rho(\vec{x}, t)}{\bar{\rho}(t)} = \frac{1}{(1 - c\lambda_1)\,(1 - c\lambda_2)\,(1 - c\lambda_3)}$$

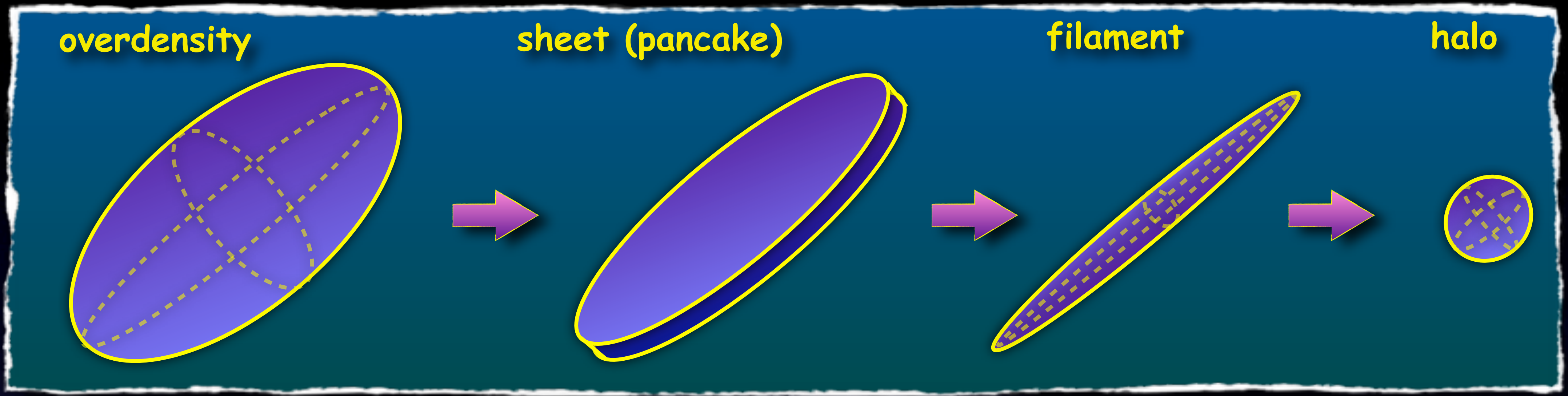
It is clear from this equation that collapse happens first along the axis associated with the first (largest) eigenvalue, λ_1 → gravity accentuates asphericity!

Hence, collapse leads to flattened structures, called (Zel'dovich) pancakes. The ZA approximation is so accurate simply because, as mentioned above, it becomes exact in the limit of planar perturbations...

Because ZA is so accurate, it is often used in setting up the initial conditions for N-body simulations..

Ellipsoidal Collapse

As is evident from the ZA, in general density perturbations will collapse according to:



For a uniform, ellipsoidal overdensity in homogeneous universe (ellipsoidal top-hat) one can obtain analytical approximations for time evolution of its 3 principal axes (see MBW §5.3).

This can be used to compute the critical overdensity for collapse (of the longest axis = `halo formation') in linear theory. The result can be obtained by solving

$$\frac{\delta_{\rm ec}}{\delta_{\rm sc}} \approx 1 + 0.47 \left[5(e^2 \pm p^2) \frac{\delta_{\rm ec}^2}{\delta_{\rm sc}^2} \right]^{0.615}$$

Sheth, Mo & Tormen (2001)

Here $\delta_{\rm ec} = \delta_{\rm ec}(e, p)$ is the critical overdensity for ellipsoidal collapse, $\delta_{\rm sc} = \delta_{\rm c} \simeq 1.686$ is the critical overdensity for spherical collapse, and the plus (minus) sign is used if p is negative (positive)....

Ellipsoidal Collapse

$$\frac{\delta_{\rm ec}}{\delta_{\rm sc}} \approx 1 + 0.47 \left[5(e^2 \pm p^2)\,\frac{\delta_{\rm ec}^2}{\delta_{\rm sc}^2}\right]^{0.615}$$

Ellipsoidal collapse

The parameters e and p characterize the asymmetry of the initial tidal field:

$$e \equiv \frac{\lambda_1 - \lambda_3}{2(\lambda_1 + \lambda_2 + \lambda_3)} \qquad p \equiv \frac{\lambda_1 + \lambda_3 - 2\lambda_2}{2(\lambda_1 + \lambda_2 + \lambda_3)}$$

Note that for a spherical system $\lambda_1 = \lambda_2 = \lambda_3$ → $e = p = 0$ → $\delta_{\rm ec} = \delta_{\rm sc} \simeq 1.686$

In general, however, $\lambda_1 > \lambda_2 > \lambda_3$ which results in $\delta_{\rm ec} > \delta_{\rm sc}$, which implies that structures collapse later under ellipsoidal collapse conditions (more realistic) than under spherical collapse conditions.

As a final remark, as we will see later, less massive structures are more strongly influenced by tides and therefore more ellipsoidal...This has important implications....

Relaxation & Virialization

Relaxation & Virialization

More details in MBW §5.4

Relaxation: the process by which a physical system acquires equilibrium or returns to equilibrium after a disturbance. Often, but not always, relaxation erases the system's "knowledge" of it's initial conditions.

Virialization: the process by which a physical system settles in virial equilibrium

Virial Equilibrium: A system is said to be in virial equilibrium if

$$2K + W + \Sigma = 0$$

K = kinetic energy
W = potential energy
Σ = work done by surface pressure

Often, Σ can be ignored, in which case virial equilibrium implies that $E = -K = W/2$

Two-body relaxation time: the time required for a particle to change its kinetic energy by about its initial amount due to two-body interactions

As you learn in Galactic Dynamics, the two-body relaxation time,

$$t_{\rm relax} \simeq \frac{N}{10 \ln N} t_{\rm cross}$$

Here N is the number of particles and $t_{\rm cross} \sim R/v$ is the system's crossing time.

For almost all collisionless systems of interest to us (galaxies, dark matter haloes) it is easy to show that $t_{\rm relax} \gg t_{\rm Hubble} \simeq 1/H_0$

Relaxation & Virialization

PUZZLE: if galaxies and haloes have two-body relaxations times that are orders of magnitude larger than the Hubble time, how can galaxies (and haloes) appear relaxed?

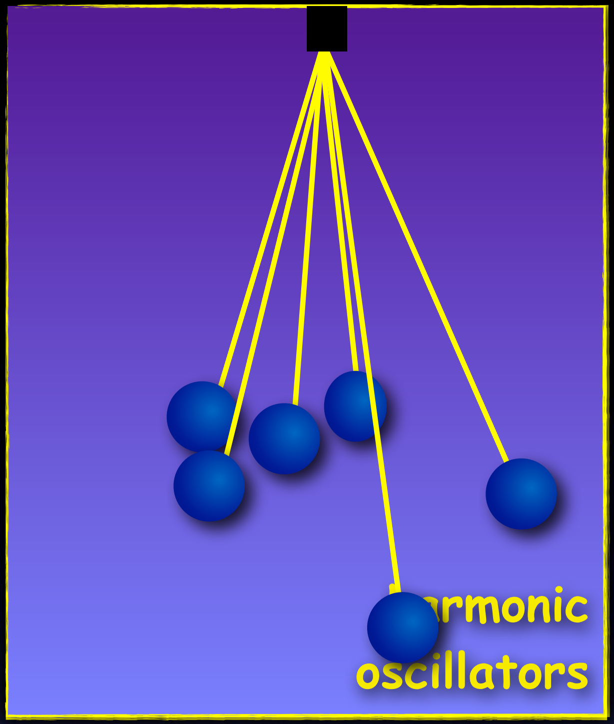
Collisionless systems such as galaxies and dark matter haloes do not relax via two-body interactions, but rather by a combination of four other mechanisms:

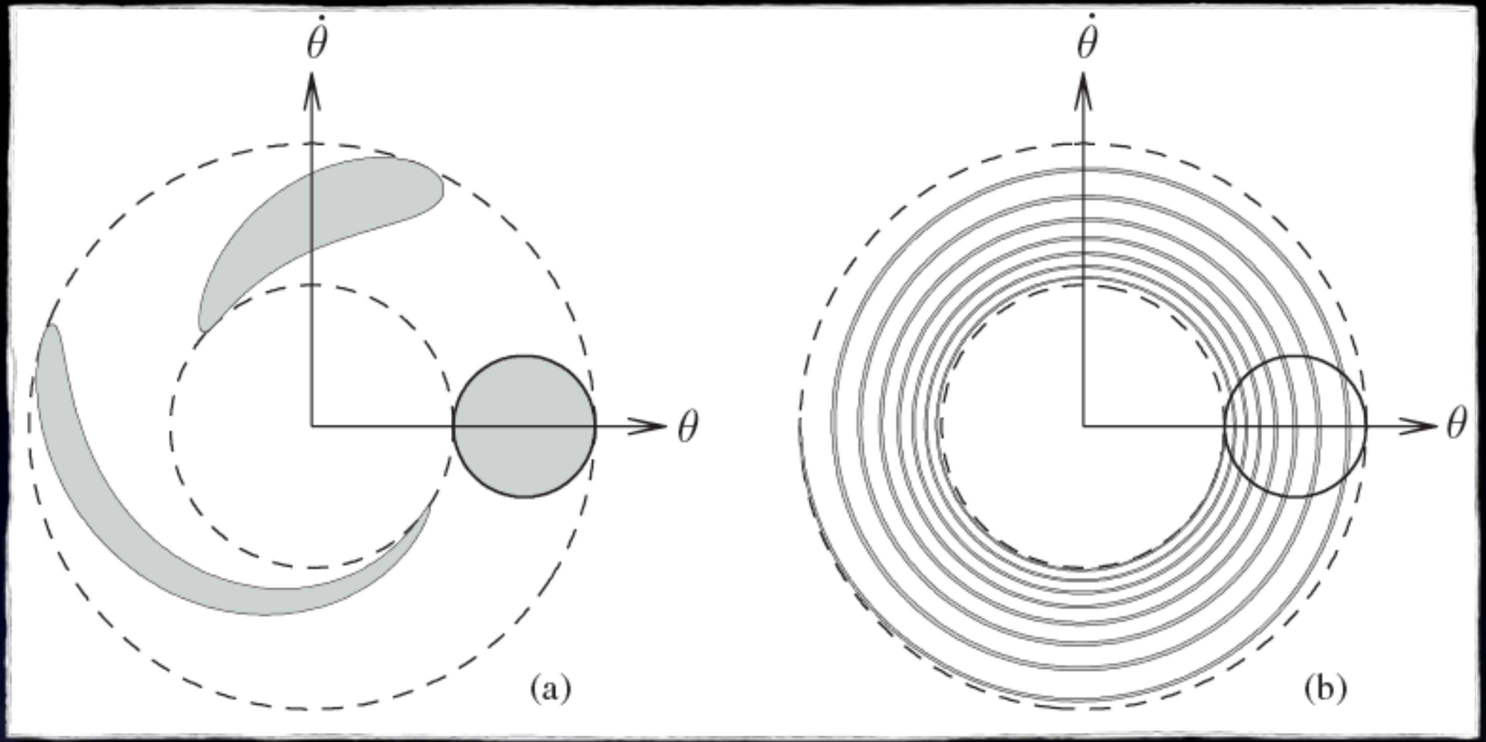
Phase-mixing	the spreading of neighboring points in phase-space due to the difference in frequencies between neighboring orbits
Chaotic mixing	the spreading of neighboring points in phase-space due to the chaotic nature of their orbits
Violent Relaxation	the change in energy of individual particles due to changes in the overall potential
Landau damping	the damping and decay of perturbations due to decoherence between particles and waves (recall free streaming)

In what follows, we briefly discuss phase mixing and violent relaxation. Chaotic mixing and Landau damping will not be covered due to time constraints (but see MBW §5.5).

Phase Mixing

More details in MBW §5.5





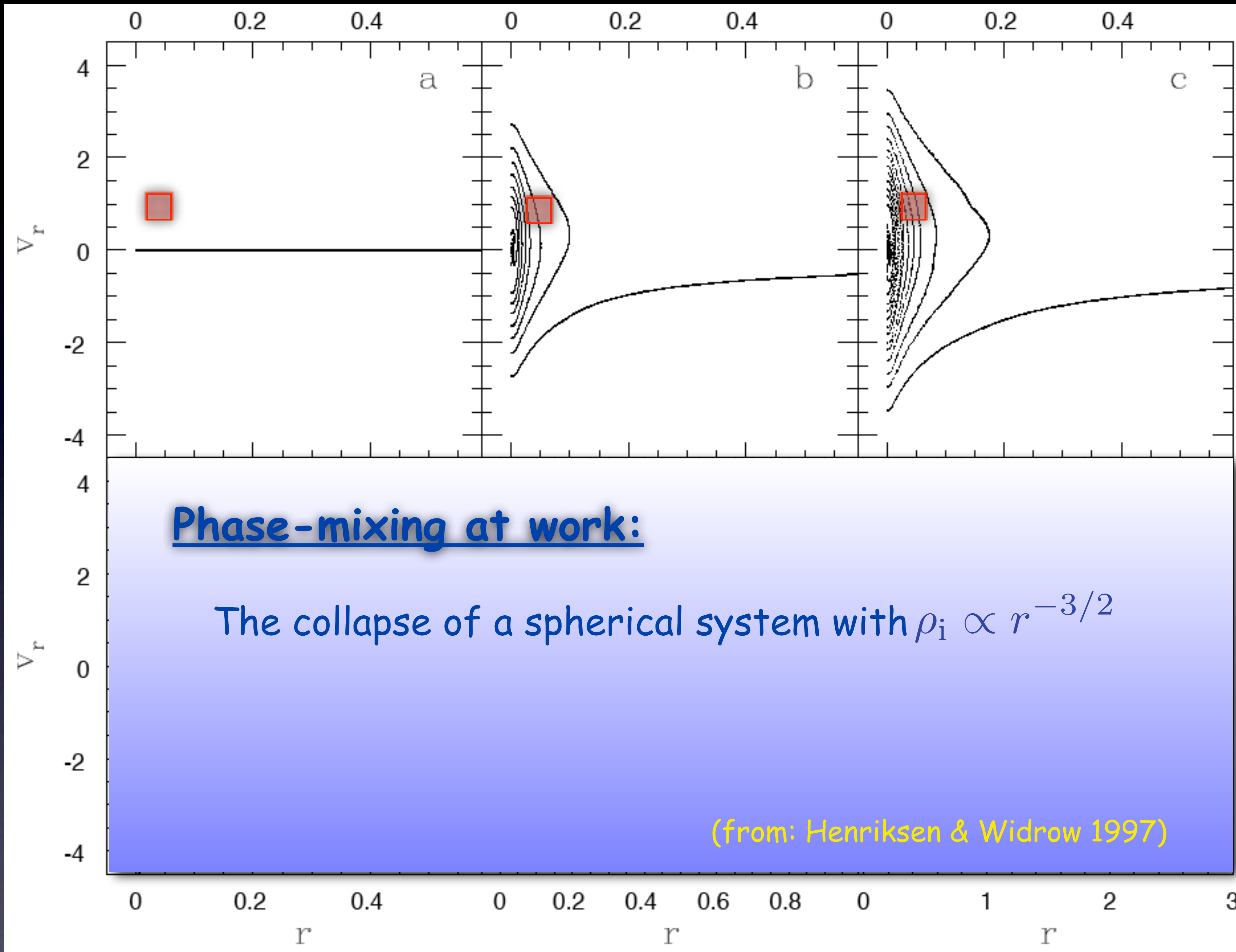
Consider a large number of harmonic oscillators, all with slightly different frequencies (i.e, with slightly different sling-lengths). If they are close to each other initially, they will, over time, phase-mix (the overall system appears more relaxed).

Let ϕ_i and ω_i be the phase and frequency of oscillator i, then oscillators i and j separate at a rate $(\Delta\phi)_{ij}(t) = 2\pi(\Delta\omega)_{ij}t$: ⇒ phase mixing scales linearly with time.

According to the collisionless Boltzmann equation, the (fine-grained) DF $f(\vec{x},\vec{v})$ remains constant. However, the coarse-grained DF, $f_c(\vec{x},\vec{v})$, measured at the initial region of phase-space, decreases as a function of time, as more and more "vacuum" is mixed in.

Phase Mixing

More details in MBW §5.5



Phase-mixing at work:

The collapse of a spherical system with $\rho_{\rm i} \propto r^{-3/2}$

(from: Henriksen & Widrow 1997)

Phase-mixing of dark matter particles in a numerical N-body simulation. The particles are initially placed in a stratified sphere with zero-velocities. Collapse rapidly phase mixes the particles

Note how the number of particles in the red box, representing the coarse-grained DF, f_c, becomes more and more similar to that of neighboring boxes; the system is relaxing...

Note that phase-mixing is a relaxation process that does not cause any loss of information: at the fine-grained level, phase-mixing is perfectly reversible and preserves all knowledge of the initial conditions....

Violent Relaxation

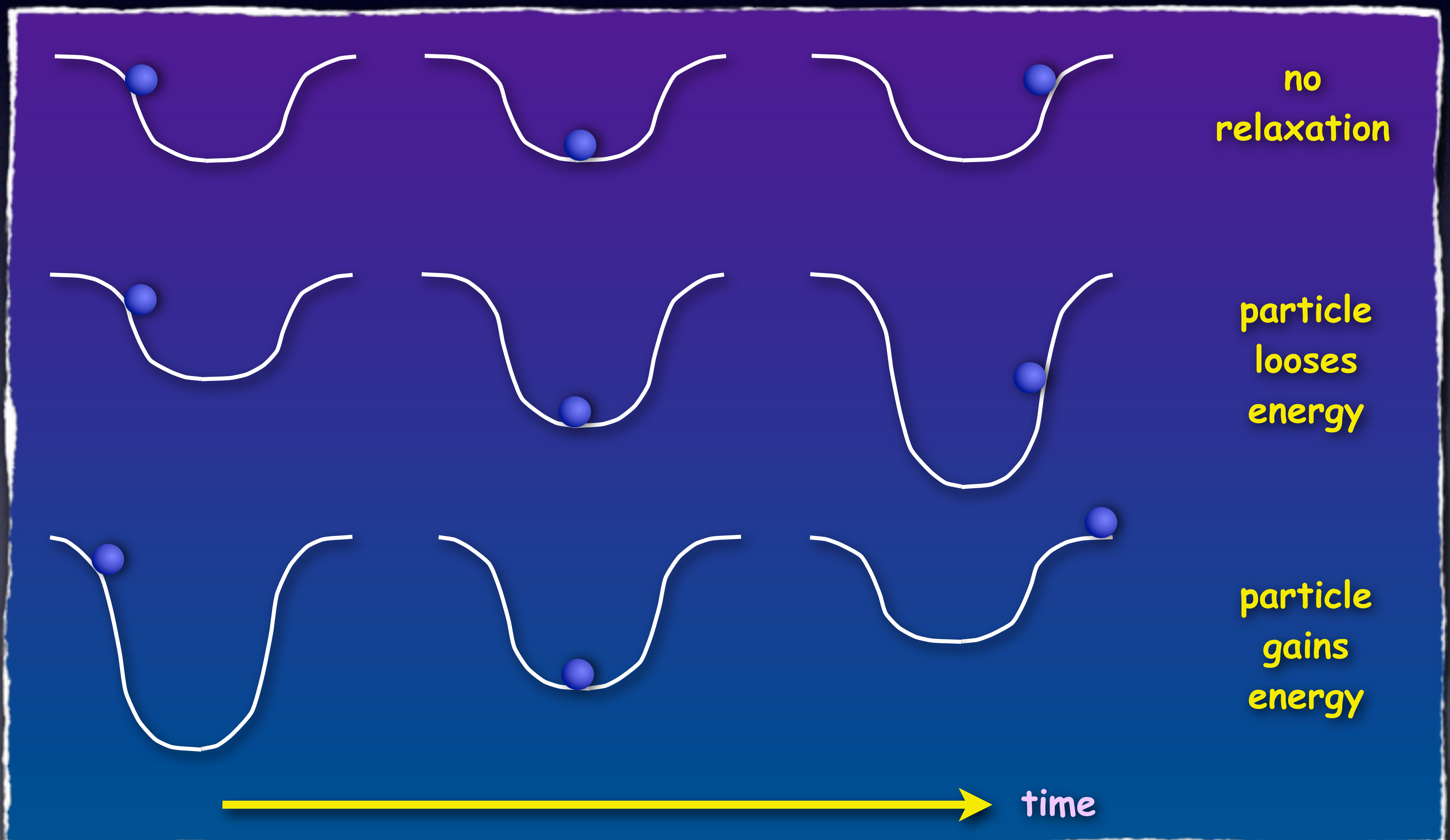
More details in MBW §5.5

Since $E = v^2/2 + \Phi$ and $\Phi = \Phi(\vec{x}, t)$ we have that:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\partial E}{\partial \vec{v}}\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} + \frac{\partial E}{\partial \Phi}\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\vec{v}\cdot\vec{\nabla}\Phi + \frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\vec{v}\cdot\vec{\nabla}\Phi + \frac{\partial \Phi}{\partial t} + \frac{\partial \Phi}{\partial \vec{x}}\cdot\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = \frac{\partial \Phi}{\partial t}$$

Thus we see that the only way in which a particle's energy can change in a collisionless system is by having a time-dependent potential.

Exactly how a particle's energy changes due to violent relaxation depends in a complex way on the particle's initial position and energy: particles can gain or loose energy. Overall, however, violent relaxation increases the width of the energy distribution...



Violent Relaxation

More details in MBW §5.5

A few remarks about violent relaxation:

- Note that dE/dt is independent of particle mass; hence, violent relaxation has no tendency to segregate particles by mass (in fact, it will undo any pre-existing segregation). This is very different from collisional relaxation, where momentum exchange during collisions drives system towards equipartition of kinetic energy: more massive particles end up with lower velocities ➡ mass segregation.

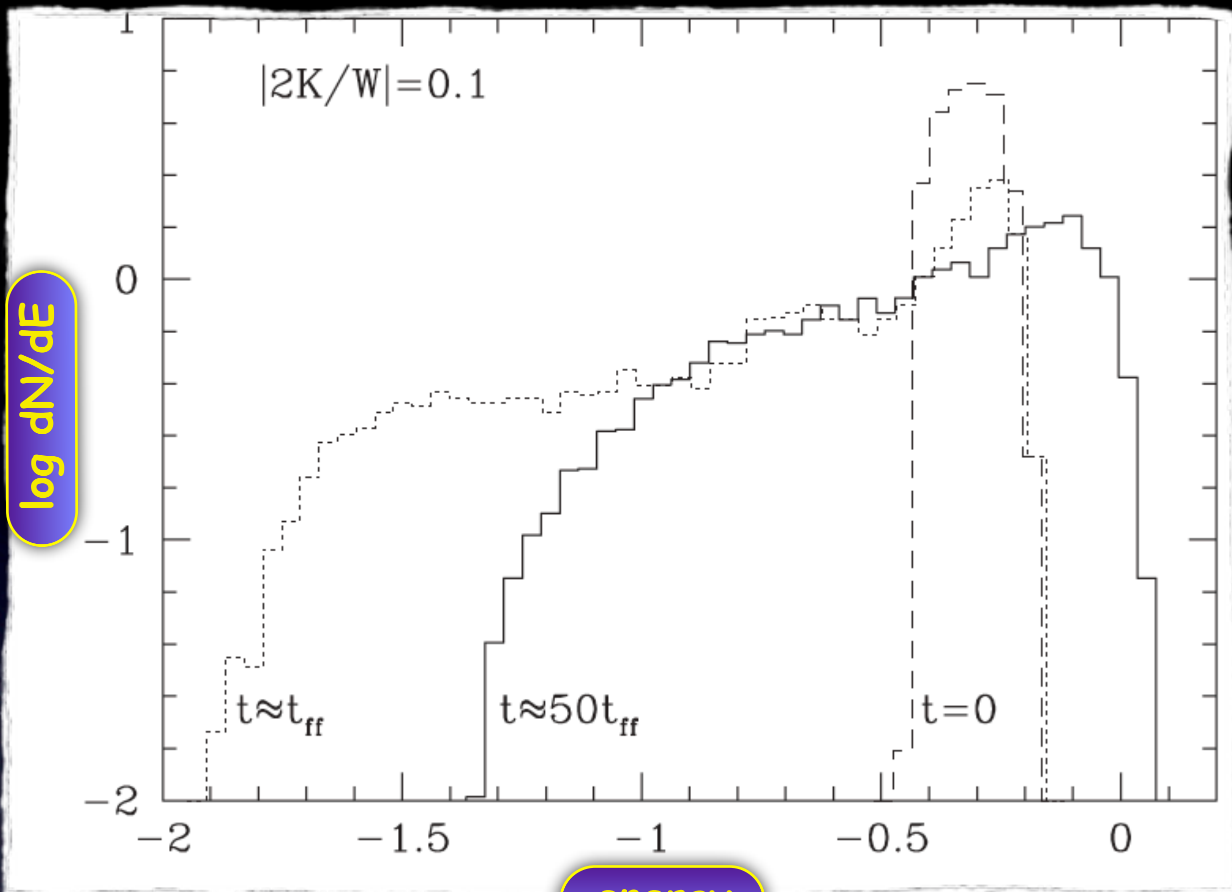
- During collapse of a collisionless sytem the CBE is still valid, i.e., the fine-grained DF does not evolve $\mathrm{d}f/\mathrm{d}t = 0$ ➡ violent relaxation only mixes at the coarse-grained level. Note, though, that unlike for a steady-state system, $\partial f/\partial t \neq 0$

- The time scale for violent relaxation is of order the time scale on which the potential changes by its own amount. This is basically the collapse time scale (≅free fall time) ➡ violent relaxation is very fast, hence its name

- Violent relaxation is self-limiting: as soon as a system approaches any equilibrium, the large-scale potential fluctuations vanish; the mixing due to violent relaxation destroys the coherence that drives potential fluctuations ➡ violent relaxation does not run to completion; not all knowledge of initial conditions is erased

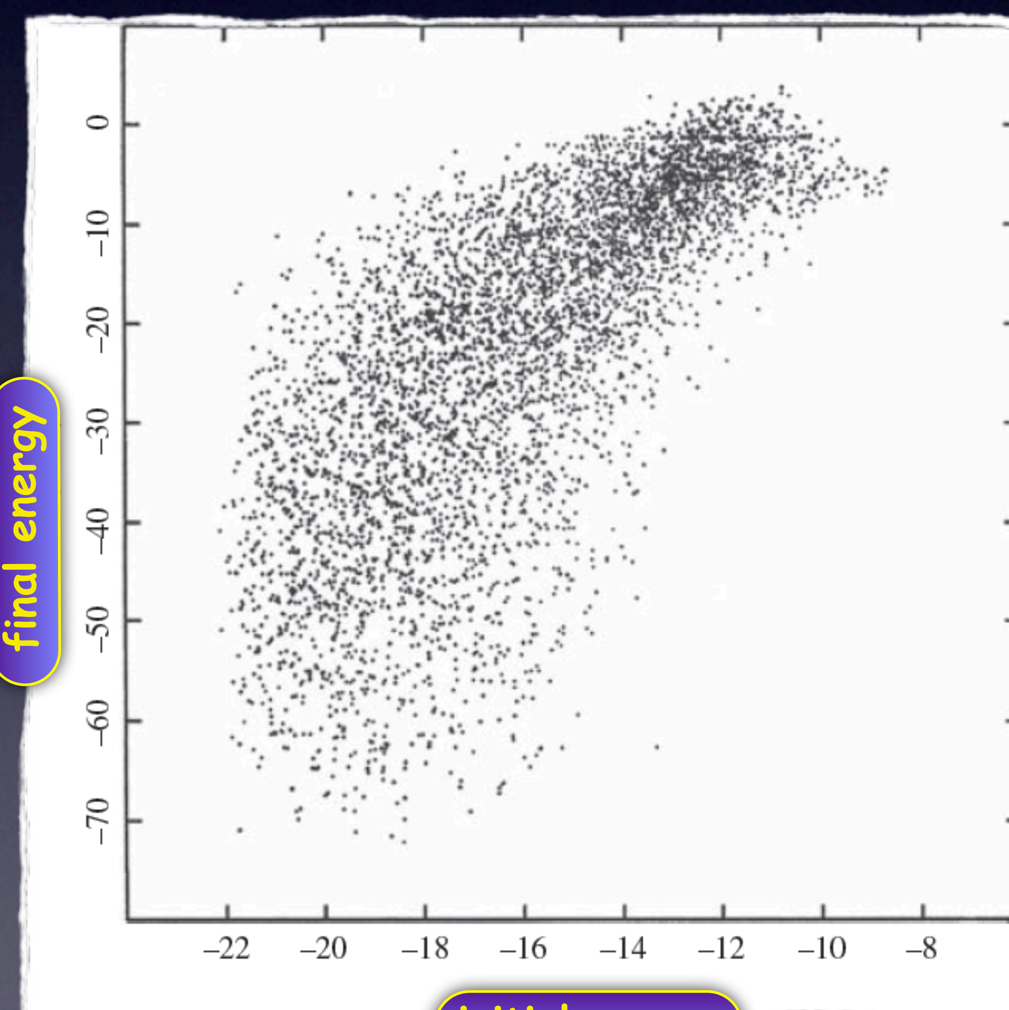
Violent Relaxation

More details in MBW §5.5



Differential energy distributions of particles in Nbody simulation of gravitational collapse. Note how violent relaxation broadens the energy distribution with time.

(from: van Albada 1982)

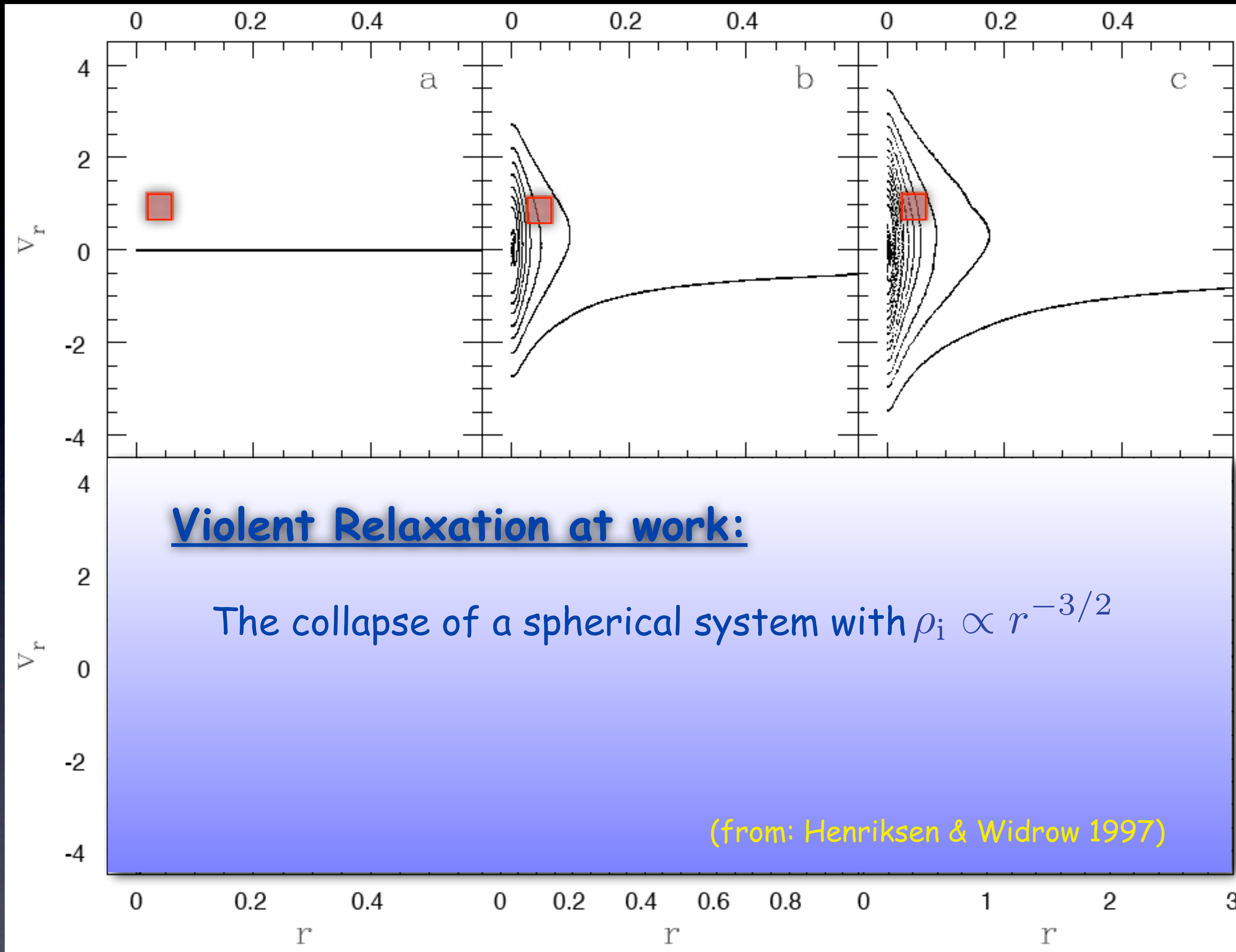


Scatter plot of final vs initial energies of the particles in the above Nbody simulation. Note that the correlation is significant, indicating that violent relaxation has not completely erased memory of the system's initial conditions.

(from: van Albada 1982)

Violent Relaxation

More details in MBW §5.5



Violent Relaxation at work:

The collapse of a spherical system with $\rho_{\rm i} \propto r^{-3/2}$

(from: Henriksen & Widrow 1997)

Note how phase-mixing is the dominant relaxation mechanism during the initial phases of the collapse.

After some time there is a transition to a more "erratic" flow: due to the time-varying potential phase-space streams start to undergo complicated bends and wiggles. This is violent relaxation at work!

Note how the number of particles in the red box, representing the coarse-grained DF, $f_{\rm c}$, becomes more and more similar to that of neighboring boxes; the system is relaxing...

Violent relaxation leads to efficient coarse-grain mixing of the DF and erases the system's memory of its initial conditions in a non-reversible way.