M1 Cosmology - 7 - Evolution of Cosmological Structure

Nick Kaiser

January 14, 2022

Contents

List of Figures

1 Linear perturbation theory

1.1 Perturbation growth in the matter dominated era

1.1.1 The spherical ‘top-hat’ (or ‘Swiss-cheese’) model

This is the simplest model for a density perturbation in a pressure free universe.

- In this model, we take a dust-filled (i.e. zero pressure) FLRW model and excise a sphere and replace it by a slightly smaller expanding sphere with the same gravitational mass
  - so the exterior is unperturbed
  - in the Newtonian limit (small binding energies) this means simply concentrating the same proper mass into a smaller volume
  - in the relativistic version, there is an excess of proper mass as the gravitational mass per unit proper mass is reduced by the gravitational potential
- here we will explore how the density fluctuation evolves with time - first in the Newtonian case
- there are, in general, two ways to set up such a perturbation (as illustrated in figure 1)
  - one is to keep the energy of the sphere fixed but to delay the ‘bang-time’
    - this generates a ‘decaying mode’ and is not very interesting
  - the other is to vary the energy, keeping the bang-time tixed
    - this generates the ‘growing mode’ – which is what we’ll consider here

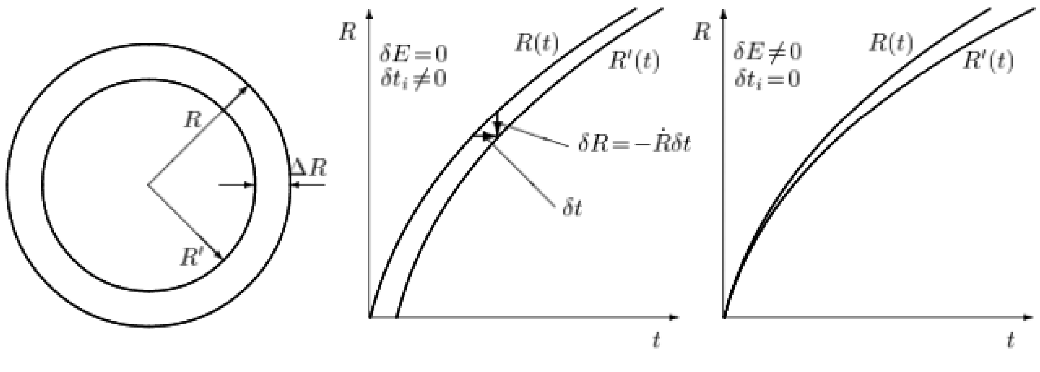

Figure 31.1: One can generate a perturbation of a dust-filled cosmology by excising a sphere of matter and replacing it with a smaller sphere of radius R'. The middle panel illustrates a decaying perturbation produced by a ‘delayed bang’. The right panel shows the more interesting growing perturbation that can be generated by perturbing the energy of the sphere. The space-time in the gap is Schwarzschild.

Figure 1: The ‘Swiss-cheese’ model for a density perturbation.

- the density perturbation is $\delta\rho/\rho = 3\delta R/R$
  - so to calculate $\delta\rho/\rho = \delta_\rho(t)$ we need to calculate $\delta R(t)$
- the perturbation to v^2 at constant R is independent of time

- it follows that
  - $v\delta v =$ constant so $\delta v = \delta\dot{R} \propto 1/v \propto \sqrt{R/GM} \propto t^{1/3}$
    - we'll assume here an Einstein-de Sitter universe for simplicity
  - so the velocity perturbation is growing with time $\delta\dot{R} \propto t^{1/3}$
  - these 'peculiar' motions are observable on large-scales using measurements of galaxy distances and redshifts (as in the 'Rubin-Ford effect').
- and integrating that gives $\delta R = \int dt \delta\dot{R} \propto t^{4/3}$ and hence, for $\delta\rho/\rho \sim \delta R/R$,
  - $\boxed{\delta\rho/\rho \propto t^{2/3}}$
- this is often called *'gravitational instability'* but that is a misnomer
  - the interior evolves conserving its total energy
  - the perturbations evolve preserving whatever binding energy they had originally
  - the perturbation to the gravitational potential is
  - $\delta\phi \sim -\delta(GM/R) = GM/R^2\delta R = (GM/R)(\delta R/R) = (1/3)(GM/R)(\delta\rho/\rho)$
  - but with $\delta\rho/\rho \propto t^{2/3} \propto R$ this is independent of time
  - and with $M \sim \rho R^3$ and $G\rho \sim H^2$ we have
  - $\delta\phi/c^2 \sim (HR/c)^2\delta\rho/\rho$
- This model can be made fully relativistic à la Oppenheimer and Snyder
  - the interior is taken to be part of a closed FLRW model
  - while the exterior may be a flat or open FLRW model
  - and space-time in the gap is Schwarzschild
  - one new feature that emerges from the relativistic analysis is that the amount of proper mass in the interior region is larger than that in the region that has been excised from the exterior (they have the same active gravitational mass – the Schwarzschild mass parameter – but that required more proper mass in the interior as the matter their has a greater (negative) binding energy)
  - at early times, the perturbation is "outside the horizon"
    - the rate of change of its radius $\dot{R} = HR$ is larger than the speed of light c
  - but its expansion velocity decreases with time and it "enters the horizon" with $HR \sim c$
    - at which time is has $\delta\rho/\rho \sim \delta\phi/c^2$

The key feature of the spherical top-hat model are:

1. there are two modes: growing $\delta \propto t^{2/3}$ and decaying $\delta \propto t^{-1}$
2. the growing mode has
   - an associated binding energy perturbation $\delta\phi \sim$ that is constant in time
   - peculiar velocities that obey $\delta v^2 = v\delta v \sim \phi$ (or $\delta v \sim \delta\phi/v$) which also grow with time (as $\delta v \propto 1/v \propto R^{1/2} \propto t^{1/3}$)

1.1.2 General (i.e. non-spherical) perturbations

- Consider the region of the universe around us within some modest redshift; maybe $z < 0.1$ for concreteness
  - which means we can treat it's dynamics using Newtonian gravity

  and let the density be $\rho(\mathbf{r}, t)$ and the velocity be $\mathbf{v}_{\rm phys}(\mathbf{r}, t)$ where $\mathbf{r}$ is the physical position.

1.1.2.1 The background model

- The 'background solution' is spatially constant density $\rho(\mathbf{r}, t) = \overline{\rho}(t)$ and pure Hubble flow $\mathbf{v}_{\rm phys} = H(t)\mathbf{r}$, which satisfy the Friedmann equation $H^2 = (8/3)\pi G\overline{\rho} +$ constant and the continuity equation $d\overline{\rho}/dt = -3H\overline{\rho}$

1.1.2.2 Comoving coordinates, density perturbation and peculiar velocity

- Define *comoving spatial coordinates*
  - $\boxed{\mathbf{x} = \mathbf{r}/a(t)}$
  - where the scale factor is such that $H = (da/dt)/a$
    * so a could be the distance between and arbitrarily chosen pair of (fictitious) fundamental observers who are moving like the matter in the background
  - it satisfies $da/a = Hdt$ so $\log a = \int dt\, H$ or $a \propto e^{\int dt\, H}$
- and define the density perturbation
  - $\delta(\mathbf{x}, t) = \delta\rho(\mathbf{x}, t)/\overline{\rho}(t) = (\rho(a\mathbf{r}, t) - \overline{\rho}(t))/\overline{\rho}(t)$
- and the *peculiar* velocity
  - $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_{\rm phys}(a\mathbf{x}, t) - H\mathbf{r}$
  - so $\mathbf{v}$ is the difference between the actual velocity in a lumpy universe as compared to what the Hubble velocity would be at that location in the absence of perturbations

1.1.2.3 Scalar and vector perturbation modes

- The results for the spherical perturbation suggest that we can decompose these into a growing and decaying mode with amplitudes (a function of $\mathbf{x}$ at some initial time)
  - $\delta(\mathbf{x}, t) = \delta^+(\mathbf{x}, t_i)(t/t_i)^{2/3} + \delta^-(\mathbf{x}, t_i)(t/t_i)^{-1}$
- in detail, it is a little more complicated than that:
  - if one decomposes the density as a sum of *comoving Fourier modes*
  - $\delta(\mathbf{x}, t) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}$
  - and similarly for the velocity
  - $\mathbf{v}(\mathbf{x}, t) = \sum_{\mathbf{k}} \mathbf{v}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}$
  - then for each $\mathbf{k}$-mode there are *four* degrees of freedom, not two, as there are 3 components for the velocity $\mathbf{v}_{\mathbf{k}}$
  - this gives extra modes that are not present in the spherical perturbation
    * for which, if we make this decomposition, the velocity for each mode is aligned with the wave-vector $\hat{\mathbf{v}}_{\mathbf{k}} = \hat{\mathbf{k}}$
    * which is like the velocity field in a sound wave – what is called a *longitudinal mode*
  - the extra modes have non-zero *transverse velocity* $\mathbf{v}_{\perp\mathbf{k}} = \mathbf{v}_{\mathbf{k}} - |\mathbf{v}_{\mathbf{k}}|\hat{\mathbf{k}}$
    * these are called *vector perturbations*
    * while the kind of modes present in the spherical model are called *scalar perturbations*
  - but the vector perturbations decay with time, so we usually ignore them and use the decomposition above
- the longitudinal velocity field determines the rate at which the density perturbation is changing (via the continuity equation)
- thus, given some initial density and (longitudinal) velocity perturbation we can solve for the amplitudes of the growing and decaying modes $\delta^{\pm}(\mathbf{x}, t_i)$

- the growing mode will come to dominate and so, in linear theory, the density perturbation grows with time simply as
  - $\delta(\mathbf{x}, t) = \delta^+(\mathbf{x}, t_i)(t/t_i)^{2/3}$
  - because all modes grow at the same rate, regardless of the wave-number

  these have associated potential perturbations $\delta\phi(\mathbf{x}, t)$ that are independent of time and peculiar velocities proportional to $-\boldsymbol{\nabla}\delta\phi$ growing like $t^{1/3}$.

1.1.2.4 Newtonian pressure-free perturbation theory

We will now formalise this a little – this will provide a useful basis on which we will build the description of more general perturbations.

A particle with peculiar velocity $\mathbf{v}$ will, in the absence of any peculiar gravitational acceleration or pressure gradient, suffer a *cosmic drag*:

$$d\mathbf{v}/dt = -H\mathbf{v}. \tag{1}$$

This looks like a friction, but it comes about simply because the particle is passing fundamental observers who are receding from one another[1].

Just as they see photons to have decreasing energy (or momentum) the momentum of a massive particle – as measured by fundamental observers it it passing – decreases as $m\mathbf{v} \propto 1/a$.[2]

Now this is the convective derivative $d\mathbf{v}/dt = \dot{\mathbf{v}} + (\mathbf{v} \cdot \boldsymbol{\nabla})\mathbf{v}$. But the velocity is of 1st order in the density perturbation. So in linear theory $\dot{\mathbf{v}}$ and $d\mathbf{v}/dt$ are the same.

Adding the effect of the density perturbations $\delta\rho = \overline{\rho}\delta$, which 'source' the *peculiar gravity* $\delta\varphi$ satisfying

$$\boxed{\nabla^2_{\mathbf{r}}\delta\varphi = 4\pi G\overline{\rho}\delta} \tag{2}$$

(where the subscript on the Laplacian shows it is Laplacian with respect to physical coordinates and $\delta \equiv \delta\rho/\overline{\rho}$) the equation of motion above becomes

$$\dot{\mathbf{v}} = -H\mathbf{v} - \boldsymbol{\nabla}_{\mathbf{r}}\delta\varphi. \tag{3}$$

It is convenient to define $\mathbf{u} = \mathbf{v}/a$ which measures the rate of change of comoving coordinate, which has time derivative $\dot{\mathbf{u}} = \dot{\mathbf{v}}/a - \mathbf{v}\dot{a}/a^2 = \dot{\mathbf{v}}/a - H\mathbf{u}$, and which, again in the absence of gravity and pressure, satisfies

$$\boxed{\dot{\mathbf{u}} = -2H\mathbf{u} - \frac{1}{a^2}\boldsymbol{\nabla}_{\mathbf{x}}\delta\varphi} \tag{4}$$

where we have switched the spatial partial derivative to $\boldsymbol{\nabla}_{\mathbf{x}} = a\boldsymbol{\nabla}_{\mathbf{r}}$. This is the Dmitriev-Zel'dovich equation used in N-body calculations. But in that context it is the Euler equation and $\dot{\mathbf{u}}$ is the total time derivative of particle velocities.

Next we use the equation of continuity. The convective version of this is $d\rho/dt = -\rho\boldsymbol{\nabla}_{\mathbf{r}} \cdot \mathbf{v}_{\mathrm{phys}}$ or

$$d\rho/dt = -\rho\boldsymbol{\nabla}_{\mathbf{r}} \cdot (H\mathbf{r} + \mathbf{v}) = -3H\rho - \rho\boldsymbol{\nabla}_{\mathbf{r}} \cdot \mathbf{v} \tag{5}$$

Replacing $\rho \Rightarrow \overline{\rho}(1+\delta)$, using the background continuity equation $d\overline{\rho}/dt = -3H\overline{\rho}$ and $\boldsymbol{\nabla}_{\mathbf{r}}\cdot\mathbf{v} = a^{-1}\boldsymbol{\nabla}_{\mathbf{x}}\cdot(a\mathbf{u}) = \boldsymbol{\nabla}_{\mathbf{x}} \cdot \mathbf{u}$ gives

$$d\delta/dt = -(1 + \delta)\boldsymbol{\nabla}_{\mathbf{x}} \cdot \mathbf{u}. \tag{6}$$

[1] One way to see this is to look at the motion of the particle from the perspective of a reference fundamental observer that it happens to be passing at time t_0. There is, by assumption, no peculiar gravity, so that observer is in free fall; it's an inertial observer. So, from its perspective the particle – also inertial – has a velocity that is constant (plus a tidally induced velocity change that – the tidal acceleration being proportional to the distance – is second order in $t - t_0$). After an interval dt, the particle has moved a physical distance $d\mathbf{r} = \mathbf{v}dt$ and is now passing another observer who is receding from the reference observer with a Hubble velocity $d\mathbf{v}_{\mathrm{H}} = Hd\mathbf{r} = H\mathbf{v}dt$. So the velocity of the particle is seen by these observers to be changing as $d\mathbf{v}/dt = -d\mathbf{v}_{\mathrm{H}}/dt = -H\mathbf{v}$.

[2] This means that from the perspective of fundamental observers the de Broglie wavelength of the particle is increasing in proportion to the scale factor. One might feel tempted to say that the expansion of the universe is stretching the wavelength of the particle's wave-function, or even that this is happening because of the coupling of the wave-function to the gravitational field of the expanding universe. But that is silly; what one has here is, in locally inertial coordinates, a wave with constant $\mathbf{k} = \mathbf{p}/\hbar$ being viewed by a sequence of observers who look at if from boosted frames that change continually with the time of the observation.

And linearising, we can replace $d\delta/dt \Rightarrow \dot{\delta}$ and drop the term involving both δ and $\mathbf{u}$, to give

$$\boxed{\dot{\delta} = -\boldsymbol{\nabla}_{\mathbf{x}} \cdot \mathbf{u}.} \tag{7}$$

Taking the divergence of (4) and using the above to eliminate $\mathbf{u}$, and Poisson's equation $a^{-2}\nabla^2_{\mathbf{x}}\delta\varphi = \nabla^2_{\mathbf{r}}\delta\varphi = 4\pi G\overline{\rho}\delta$, gives a single second order equation for the density perturbation $\delta(t)$:

$$\boxed{\ddot{\delta} + 2H\dot{\delta} - 4\pi G\overline{\rho}\delta = 0} \tag{8}$$

As a sanity check, if we assume an Einstein-de Sitter model, so $4\pi G\overline{\rho} = H^2/2$ and $a \propto t^{2/3}$ so $H = 2/3t$, and, as an ansätz, $\delta_{\mathbf{k}} \propto t^{\alpha}$, this becomes the algebraic equation

$$3\alpha^2 + \alpha - 2 = 0 \tag{9}$$

which, as expected, has solutions $\alpha = 2/3$ and $\alpha = -1$.

Things were relatively straightforward since the evolution of the perturbations is independent of their size. When we include the effects of pressure, we will find that the growth rate is wavelength dependent. To deal with that, we need to write the density, potential and velocity perturbations as Fourier syntheses: $\delta(\mathbf{x}, t) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}}$, where reality of $\delta(\mathbf{x}, t)$ imposes the symmetry $\delta_{-\mathbf{k}}(t) = \delta^{\star}_{\mathbf{k}}(t)$, and similarly $\delta\varphi(\mathbf{x}, t) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}}$. The longitudinal nature of the velocity field is imposed by writing $\mathbf{u}(\mathbf{x}, t) = \sum_{\mathbf{k}} u_{\mathbf{k}}(t)\hat{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}$.

Equations (4), (7) and (2) are all linear and so become equations for the Fourier mode amplitudes, and the spatial derivatives in these become algebraic: $\boldsymbol{\nabla}_{\mathbf{x}}e^{i\mathbf{k}\cdot\mathbf{x}} = i\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{x}}$. The equation of motion (4) is

$$\dot{u}_{\mathbf{k}} = -2Hu_{\mathbf{k}} - ik\varphi_{\mathbf{k}}/a^2 \tag{10}$$

Poisson's equation (2) is

$$|\mathbf{k}|^2\varphi_{\mathbf{k}} = -4\pi G\overline{\rho}a^2\delta_{\mathbf{k}}. \tag{11}$$

and the continuity equation (7) is

$$\dot{\delta}_{\mathbf{k}} = -i|\mathbf{k}|u_{\mathbf{k}}. \tag{12}$$

Using the time derivative of the last and (11) in (10) gives equation for the mode amplitude $\delta_{\mathbf{k}}(t)$:

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\overline{\rho}\delta_{\mathbf{k}} = 0 \tag{13}$$

where the subscript $\mathbf{k}$ is rather redundant as nothing here depends on $\mathbf{k}$.

The above results are valid in the matter dominated era and also in the presence of a cosmological constant or dark energy (provided it is sufficiently 'stiff', as would be the case for a scalar field with low mass). Next we consider the evolution of perturbations prior to recombination when the pressure of the radiation needs to be taken into account.

1.2 Perturbations with non-vanishing pressure

The foregoing analysis is adequate to describe density perturbations when they are 'outside the horizon' when they behave essentially like frozen spatial curvature perturbations – think of the embedding diagram of an overdensity as a bowl-shaped depression – and for perturbations of comoving wave-number (or spatial frequency) $\mathbf{k}$ sufficiently small that they enter the horizon (at the time when $\lambda_{\rm phys} \sim a(t)/|\mathbf{k}| \sim ct$) in the matter dominated era.

That limits the domain of applicability to super-cluster scales or larger.

Modes with wavelengths corresponding to galaxies and clusters of galaxies enter the horizon before recombination and behave like sound waves because there is a restoring force from the pressure gradient associated with the density fluctuation, at least if we assume that the 'entropy per baryon' is constant; these are so called 'adiabatic' or 'isentropic' perturbations. We will have more to say about this later.

We will start by reviewing the theory of 'linearised' sound-waves in a non-expanding framework. This introduces the sound-speed and relates it to $\delta P/\delta\rho$. We then consider waves in the radiation dominated era, making the simplifying assumption that the the gravity of the perturbations is negligible (a very good approximation once the waves are a good deal smaller than the horizon size). Then we consider sound waves in the matter dominated era. Now including the effect of self-gravity; this introduces the concept of the 'Jeans length' which separates large-scale perturbations that grow with time – essentially because their sound speed has become imaginary – and smaller ones that oscillate (and they do so, as we shall see, with diminishing amplitude).

1.2.1 The sound speed in a relativistic plasma

Let's consider sound waves of small amplitude in a non-expanding plasma modelled as a perfect fluid. We can think of it as being confined by a box and we will neglect gravity entirely.

1.2.1.1 The continuity equations

We can obtain the equations of motion from the laws of conservation of energy and momentum, as expressed in ${T^{\mu\nu}}_{,\mu} = 0$. The stress energy tensor for the fluid is

$$T^{\mu\nu} = c^{-2}(\mathcal{E} + P)U^\mu U^\nu + \eta^{\mu\nu} P \tag{14}$$

where $U^\mu = (\gamma c, \gamma \mathbf{v})$ is the 4-velocity of (an observer moving along with) the fluid in whose frame the stress energy tensor has components $T^{\mu\nu} = \mathrm{diag}(\mathcal{E}, P, P, P)$, where, as usual, $\mathcal{E}$ is the energy density and P is the pressure (flux density of momentum). We will be interested here in the case where the pressure is that of the radiation (assumed to be tightly coupled to the plasma).

The general form of the continuity equations are then

$$\begin{aligned} 0 &= (\mathcal{E} + P)_{,\mu} U^\mu U^\nu + (\mathcal{E} + P)({U^\mu}_{,\mu} U^\nu + U^\mu {U^\nu}_{,\mu}) + c^2 P^{,\nu} \\ &= U^\nu \frac{d(\mathcal{E} + P)}{d\tau} + (\mathcal{E} + P)\left(c U^\nu \boldsymbol{\nabla} \cdot \mathbf{v} + dU^\nu/d\tau\right) + c^2 P^{,\nu} \end{aligned} \tag{15}$$

where $\boldsymbol{\nabla} \cdot \mathbf{v} = {U^\mu}_{,\mu}$ is the volume expansion rate of the fluid as measured in the frame of the fluid[3]. And these, for $\nu = 0$ and $\nu = i$ respectively, are

$$\begin{aligned} \frac{d\mathcal{E}}{d\tau} &= -(\mathcal{E} + P)\boldsymbol{\nabla} \cdot \mathbf{v} \\ \frac{d\mathbf{v}}{d\tau} &= -\frac{c^2}{\mathcal{E} + P} \boldsymbol{\nabla} P \end{aligned} \tag{16}$$

Q: The appearance of the enthalpy $\mathcal{E}+P$ in the energy conservation equation is relatively straightforward; the energy goes down for two reasons – one is its dilution as the volume increases, the other is that the fluid does PdV work. But why does enthalpy appear in the second? Forget about expansion for a minute; the pressure is the flux of momentum, so its divergence $\boldsymbol{\nabla} P$ is the rate at which momentum is changing locally. So this equation says that the rate of change of momentum is $\rho + P/c^2$ times the acceleration $\mathbf{a} = d\mathbf{v}/d\tau$. Why not just $\rho \mathbf{a}$?

1.2.1.2 The linearised wave equation

These admit 'zeroth order' solutions where the energy density $\mathcal{E}(\mathbf{x}, t) = \mathcal{E}_0$ is constant in space and time, as is the pressure $P(\mathbf{x}, t) = P_0$, and where the fluid velocity field vanishes everywhere: $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_0 = \mathbf{0}$.

Writing $\mathcal{E} = \mathcal{E}_0 + \mathcal{E}_1(\mathbf{x}, t) + \ldots$ and similarly for the pressure and velocity, it is clear that the spatial derivatives on the right hand sides are 1st order quantities, so, working at linear order, we can take the multiplying factors on the right to be the zeroth order quantities.

On the left we will have convective derivatives with respect to proper time of 1st order quantities $\mathcal{E}_1$ and $\mathbf{v}_1$. But then we can take these to be partial derivatives with respect to coordinate time, since $d/d\tau = \gamma d/dt = (1 + \mathcal{O}(|\mathbf{v}_1|^2/c^2))d/dt$ and $d/dt = \partial_t + \mathbf{v}_1 \cdot \boldsymbol{\nabla}$, so we can take $d/d\tau = \partial_t$ on the left, so

$$\begin{aligned} \dot{\mathcal{E}}_1 &= -(\mathcal{E}_0 + P_0)\boldsymbol{\nabla} \cdot \mathbf{v}_1 \\ \dot{\mathbf{v}}_1 &= -\frac{c^2}{\mathcal{E}_0 + P_0} \boldsymbol{\nabla} P_1. \end{aligned} \tag{17}$$

which can be combined to

$$\ddot{\mathcal{E}}_1 = c^2 \nabla^2 P_1 \tag{18}$$

[3]To obtain this we have used the fact that, with $U^\mu = dx^\mu/d\tau = c\gamma dx^\mu/dx^0$ its 4-divergence is ${U^\mu}_{,\mu} = c\partial_\mu(\gamma dx^\mu/dx^0) = c d\gamma/dx_0 + c\gamma\partial_\mu(1, \mathbf{v})$. Evaluating this scalar in the instantaneously co-moving rest-frame of the fluid, the first term vanishes (since $d\gamma = d(1 - \mathbf{v} \cdot \mathbf{v}/c^2)^{-1/2} = \mathbf{v} \cdot d\mathbf{v}/\gamma^3$ and $\mathbf{v} = 0$ in this frame) while the second is $c\partial_i v^i = c\boldsymbol{\nabla} \cdot \mathbf{v}$.

And if the perturbation to the pressure associated with a perturbation to the energy density $\delta\mathcal{E} = c^2\delta\rho$ is δP, we have a linear, dispersionless wave equation

$$\ddot{\mathcal{E}}_1 = c_{\rm s}^2 \nabla^2 \mathcal{E}_1 \tag{19}$$

which allows travelling wave solutions moving at the sound speed $c_{\rm s}$ whose square is

$$c_{\rm s}^2 = \frac{\delta P}{\delta \rho}. \tag{20}$$

1.2.1.3 The comoving sound horizon scale

The pressure, we will assume, comes entirely from the radiation and is $P = \frac{1}{3}\rho_{\rm r} c^2$. In the limit that the radiation dominates the density, we have $\delta P/\delta\rho = c^2/3$ and waves travel at $c_{\rm s} = c/\sqrt{3}$. More generally, if a volume element undergoes a fractional volume change $\delta V/V$ the change in the density of baryons (assumed non-relativistic) is $\delta\rho_{\rm b}/\overline{\rho}_{\rm b} = -\delta V/V$ while that of the radiation – with 'adiabatic index' $-4/3$ – is $\delta\rho_{\rm r}/\overline{\rho}_{\rm r} = -(4/3)\delta V/V$ and so $\delta\rho = \delta\rho_{\rm b} + \delta\rho_{\rm r} = -(\overline{\rho}_{\rm b} + \frac{4}{3}\overline{\rho}_{\rm r})\delta V/V$ while $\delta P = \frac{1}{3}\delta\rho_{\rm r} c^2 = -(4/9)\overline{\rho}_{\rm r}\delta V/V$ so

$$c_{\rm s}^2 = \frac{c^2}{3}\frac{4\overline{\rho}_{\rm r}}{4\overline{\rho}_{\rm r} + 3\overline{\rho}_{\rm b}}. \tag{21}$$

Just as we defined the comoving (light) horizon scale as $\lambda_{\rm H} \equiv c/Ha \sim ct/a$ – whose physical implication is that it is the comoving distance the light can travel in one expansion time $t_{\rm exp} \equiv 1/H$, and which limits the range over which any causal physical influence can be propagated – one can define the comoving sound horizon scale to be $\lambda_{\rm s} \equiv c_{\rm s}/Ha$.

In the radiation dominated era, $c_{\rm s} = c/\sqrt{3}$ and the sound horizon tracks the light horizon. So it is only just a little after a wavelength enters the horizon that sound waves can propagate across it.

In the matter era, in contrast, $c_{\rm s} \simeq (2/3)\sqrt{\overline{\rho}_{\rm r}/\overline{\rho}_{\rm b}} \propto 1/\sqrt{a}$ while $H \propto \sqrt{\rho} \propto a^{-3/2}$, so $c_{\rm s}/Ha \propto a^0$. Thus the comoving sound horizon tracks the light horizon until $t_{\rm eq}$ – both growing as a – but thereafter is constant. This is only valid as long as the photons remain coupled to the baryons; i.e. up until $t_{\rm rec}$ when the hydrogen (re)combines. At that point, the sound speed for the baryons drops precipitously, and pressure becomes unimportant. While the 'plateau' of constant comoving sound horizon only lasts a limited time, it is important since it introduces quite a strong feature in the spectrum of density fluctuations emerging after recombination.

1.2.2 Sound waves in the radiation era

A general treatment of perturbations in the radiation era is somewhat complicated. Broadly speaking, while outside the horizon the pressure gradients are not important and the spatial curvature perturbations are 'frozen in' (though the precise definition of this is a little subtle as the spatial curvature depends on the hypersurfaces on which on measures the curvature). At horizon entry the pressure gradient acceleration is similar to the gravitational acceleration. But, very soon after, the pressure gradients dominate the evolution, and what we have, as long as the photons remain tightly coupled, are waves whose frequency is changing – because of the increase of the wavelengths as they expand – and which evolve adiabatically and, one might think, would undergo some kind of secular evolution. In fact, the amplitude of the waves does not change, as we now show.

To understand the damping, is is sufficient to consider a plasma in the Milne model, with waves that are standing or travelling waves in comoving coordinates. We will assume that $\overline{\rho}_{\rm r} \gg \overline{\rho}_{\rm b}$, so $P = \mathcal{E}/3$ and the continuity equations are

$$\begin{aligned} \frac{d\mathcal{E}}{d\tau} &= -\frac{4}{3}\mathcal{E}\boldsymbol{\nabla}\cdot\mathbf{v} \\ \frac{d\mathbf{v}}{d\tau} &= -\frac{c^2}{4}\frac{\boldsymbol{\nabla}\mathcal{E}}{\mathcal{E}}. \end{aligned} \tag{22}$$

where all quantities are as measured by the observer who sees $T^{\mu\nu} = \mathcal{E}\,{\rm diag}(1, 1/3, 1/3, 1/3)$.

We will take as the zeroth order solution a velocity field which is that of fundamental Milne observers. These see $\boldsymbol{\nabla}\cdot\mathbf{v}_0 = 3H = 3/\tau$. And, unsurprisingly, the 1st equation then tells us that $d\mathcal{E}_0/\mathcal{E}_0 = -4d\tau/\tau$ with solution $\mathcal{E}_0(\tau) \propto \tau^{-4}$ (i.e. going like a^{-4} since the scale factor is proportional to τ in the Milne model).

We now want to introduce perturbations, writing $\mathcal{E}(\mathbf{x},\tau) = \mathcal{E}_0(\tau)+\mathcal{E}_1(\mathbf{x},\tau)$, and with associated velocity field with divergence $\boldsymbol{\nabla}\cdot\mathbf{v} = \boldsymbol{\nabla}\cdot(\mathbf{v}_0+\mathbf{v}_1) = 3/\tau+\boldsymbol{\nabla}\cdot\mathbf{v}_1$. The spatial gradients here are with respect to physical distance. The waves, however, are postulated to have wavelengths that expand in proportion to τ, thus, for a wave $\mathcal{E}_1 = \frac{1}{2}(\mathcal{E}_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} + \text{c.c.})$, so $\mathbf{k}$ is a comoving wavenumber, and with $\lambda_{\text{phys}} = 2\pi c\tau/|\mathbf{k}|$ small compared to the curvature scale, we have $\boldsymbol{\nabla}\mathcal{E}_1 = \frac{1}{2}(i\mathbf{k}\mathcal{E}_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} + \text{c.c.})/c\tau$, and similarly $\boldsymbol{\nabla}\cdot\mathbf{v}_1 = \frac{1}{2}(i\mathbf{k}\cdot\mathbf{v}_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} + \text{c.c.})/c\tau$

The energy continuity equation is

$$\dot{\mathcal{E}}_0 + \dot{\mathcal{E}}_1 = -\tfrac{4}{3}(\mathcal{E}_0+\mathcal{E}_1)\boldsymbol{\nabla}\cdot(\mathbf{v}_0+\mathbf{v}_1) \tag{23}$$

whose 1st order components are

$$\dot{\mathcal{E}}_1 = -\tfrac{4}{3}(\mathcal{E}_1\boldsymbol{\nabla}\cdot\mathbf{v}_0 + \mathcal{E}_0\boldsymbol{\nabla}\cdot\mathbf{v}_1) = -4\mathcal{E}_1/\tau - \tfrac{4}{3}\mathcal{E}_0\boldsymbol{\nabla}\cdot\mathbf{v}_1. \tag{24}$$

For a single wave, with a longitudinal velocity field, the energy continuity equation becomes

$$kv_\mathbf{k} = -\frac{3c\tau}{4\mathcal{E}_0}(\dot{\mathcal{E}}_\mathbf{k} + 4\mathcal{E}_\mathbf{k}/\tau) \tag{25}$$

with time derivative

$$k\dot{v}_\mathbf{k} = -\frac{15c}{4\mathcal{E}_0}(\dot{\mathcal{E}}_\mathbf{k} + 4\mathcal{E}_\mathbf{k}/\tau) - \frac{3c\tau}{4\mathcal{E}_0}(\ddot{\mathcal{E}}_\mathbf{k} + 4\dot{\mathcal{E}}_\mathbf{k}/\tau - 4\mathcal{E}_\mathbf{k}/\tau^2) \tag{26}$$

while the momentum continuity equation tells us that

$$k\dot{v}_\mathbf{k} = -\frac{k^2\mathcal{E}_\mathbf{k}}{4c\tau\mathcal{E}_0}. \tag{27}$$

Putting these together (and dropping terms like $\mathcal{E}_\mathbf{k}/\tau^2$ as compared to $\dot{\mathcal{E}}_\mathbf{k}/\tau$ and $k^2\mathcal{E}_\mathbf{k}/\tau^2$) gives

$$\boxed{\ddot{\mathcal{E}}_\mathbf{k} + 9\dot{\mathcal{E}}_\mathbf{k}/\tau - \tfrac{1}{3}(k^2/\tau^2)\mathcal{E}_\mathbf{k} = 0,} \tag{28}$$

which is a damped oscillator equation. Making the usual substitution $\mathcal{E}_\mathbf{k} = \chi/\tau^{9/2}$ eliminates the damping term. The frequency is $\omega \propto 1/\tau$ and adiabaticity says that $\dot{\chi}^2 = \omega^2\chi^2 \propto \omega$ so $\chi \propto \sqrt{\tau}$ and hence $\mathcal{E}_\mathbf{k} \propto 1/\tau^4$. But that is the same variation as $\mathcal{E}_0$, and hence the amplitude of the wave remains constant.

You might think there would be a (much) easier way to see this, and indeed there is. If we think about a certain comoving volume, the mass is $M = V\mathcal{E}/c^2 \propto 1/a$, which is oscillating at a frequency $\omega \propto 1/a$ and has some velocity v. The energy is $E \sim Mv^2$ and adiabaticity says this should vary in proportion to ω, so $v^2 \propto \omega/M \propto a^0$ so the amplitude of the velocity is unchanging. That means that the physical displacement is $\delta r \sim vT/2\pi = v/\omega \propto a$, but that means the comoving dispacement $\delta r/a \propto a^0$ and is also unchanging.

1.2.3 Sound waves in the matter era

We will now consider evolution of sound waves after t_{eq}. Of interest are waves that are below the horizon scale – so we can use the Newtonian approximation developed above (with gravity incorporated) – but may be above or below the sound horizon scale.

1.2.3.1 The equation of motion and the Jeans length

To include the pressure gradient in the equation of motion for δ we just need to augment the gravitational acceleration $-\boldsymbol{\nabla}_\mathbf{r}\delta\varphi$ by the pressure gradient acceleration $-\rho^{-1}\boldsymbol{\nabla}_\mathbf{r}P$ which, for linear perturbations, is $-\overline{\rho}^{-1}\boldsymbol{\nabla}_\mathbf{r}\delta P$ or, with $\delta P = c_s^2\delta\rho$ is $-c_s^2\boldsymbol{\nabla}_\mathbf{r}\delta$. The result is

$$\boxed{\ddot{\delta}_\mathbf{k} + 2H\dot{\delta}_\mathbf{k} - (4\pi G\overline{\rho} - c_s^2k^2/a^2)\delta_\mathbf{k} = 0} \tag{29}$$

The effect of pressure radically changes the solutions. If $k \ll k_{\text{J}}$ where the *comoving Jeans wave-number* is

$$k_{\text{J}} \equiv \sqrt{4\pi G\overline{\rho}a^2/c_s^2} \tag{30}$$

we can ignore the effect of pressure, and we have growing perturbations as before.

But if $k > k_{\text{J}}$ the sign of the coefficient of $\delta_\mathbf{k}$ becomes positive and the result is that we get perturbations that oscillate with time.

1.2.3.2 Adiabatic damping of sound waves

They are however damped oscillations. As usual, to understand the adiabatic damping, we make a change of variables: $\delta_{\mathbf{k}} = \Delta_{\mathbf{k}} t^{\alpha}$. In terms of $\Delta_{\mathbf{k}}$ the equation of motion becomes

$$\ddot{\Delta}_{\mathbf{k}} + 2\left(\frac{\alpha}{t} + H\right)\dot{\Delta}_{\mathbf{k}} + \left(\frac{c_s^2 k^2}{a^2} - \frac{3}{2}H^2 + \frac{2H\alpha^2(\alpha-1)}{t^3}\right)\Delta_{\mathbf{k}} = 0 \tag{31}$$

so, if we take $\alpha = -2/3$ the coefficient of the damping term vanishes and we have

$$\boxed{\ddot{\Delta}_{\mathbf{k}} + \left(\frac{c_s^2 k^2}{a^2} - H^2\right)\Delta_{\mathbf{k}} = 0} \tag{32}$$

which is an undamped oscillator equation for the auxiliary variable: $\ddot{\Delta}_{\mathbf{k}} = -\omega_{\mathbf{k}}^2 \Delta_{\mathbf{k}}$ with time varying frequency.

For wavelengths less than the Jeans length, we can neglect H^2 as compared to $c_s^2 k^2/a^2$ and we have that the frequency varies as $\omega_{\mathbf{k}} \simeq c_s k/a \propto a^{-3/2}$. Adiabaticity tells us that the envelope obeys $\omega_{\mathbf{k}}^2 \Delta_{\mathbf{k}}^2 \propto \omega_{\mathbf{k}}$ so $\Delta_{\mathbf{k}} \propto 1/\sqrt{\omega_{\mathbf{k}}} \propto a^{3/4}$. But with $\alpha = -2/3$, $\delta_{\mathbf{k}} = \Delta_{\mathbf{k}} t^{\alpha} = \Delta_{\mathbf{k}}/a \propto a^{-1/4}$.

We therefore infer that the amplitude will decrease with time (but not very rapidly) as

$$\delta_{\mathbf{k}} \propto (1+z)^{1/4}. \tag{33}$$

As with sounds waves in the radiation era, we could have reached this conclusion more simply by saying that the matter in some fixed comoving volume has mass M – which is now independent of a – which is oscillating with velocity v, and with frequency $\omega \sim c_{\mathrm{s}}\lambda_{\mathrm{phys}} \propto a^{-3/2}$, so its energy is $E \sim Mv^2$ and for this to vary in proportion to the frequency we must have $v \propto \sqrt{\omega}$. The physical displacement is $\delta r \sim v/\omega \propto 1/\sqrt{\omega} \propto a^{3/4}$ while the physical wavelength is increasing as $\lambda_{\mathrm{phys}} \propto a$ and therefore the amplitude of the wave – which must be on the order of $\delta r/\lambda_{\mathrm{phys}}$ – varies with the scale factor as $a^{-1/4}$.

1.2.4 Isentropic vs isocurvature fluctuations

We have assumed above that the *'entropy per baryon'* is a universal constant. For thermal radiation the entropy density is just, to order of magnitude k_{B} times the number density of photons. The entropy also has a contribution from other relativistic species like neutrinos. Such perturbations are often called, for obvious reasons, *isentropic*. They are also often called *adiabatic* fluctuations as they are the kind of perturbation you generate if you take all the contents of some region of space and compress or rarify it, without and heat flowing in or out of the volume (so adiabatic in the sense of 'no heat flow').

An alternative type of fluctuation, though one that is considered less these days, is the kind that would be generated if, for example, *baryogenesis* – i.e. whatever unknown physics is responsible for the fact that there is an excess of matter over anti-matter at late times but, judging from the fact that there are $\sim 10^9$ photons per baryon, a small difference at early times – acted in a way that was somewhat inhomogeneous.

This, or other entropy generating processes happening in an inhomogeneous way, could generate initial conditions where the total energy density is initially unperturbed and where any excess of the density of baryons is compensated for by a deficit of radiation density and vice verse. If so, there would be no associated curvature fluctuations, leading to the terminology *isocurvature perturbations.* These possibilities are illustrated in figure 2.

Such perturbations would, once inside the horizon, oscillate as sound waves, but about an asymmetric offset (in which the temperature would be uniform in space – hence the alternative nomenclature of *'isothermal'* perturbations.

One recent application of isocurvature perturbations is in relation to the (controversial) *'dark flows'* claimed from measurements of the *kinematic Sunyaev Zel'dovich* (KSZ) effect with clusters of galaxies. These suggest that either there are substantial *'bulk-flows'* on surprisingly large scales or that the conventional interpretation of the dipole anisotropy of the CMB as being due to our motion with respect to the frame of rest of distant matter is incorrect. One way to accomplish the latter would be for there to be an isocurvature component with a very large wavelength (larger than the present horizon) that generates a dipole. Such models are sometimes called *'tilted'* cosmologies. The possibility that the 'conventional wisdom' regarding the CMB dipole is false was first pointed out by Jim Gunn.

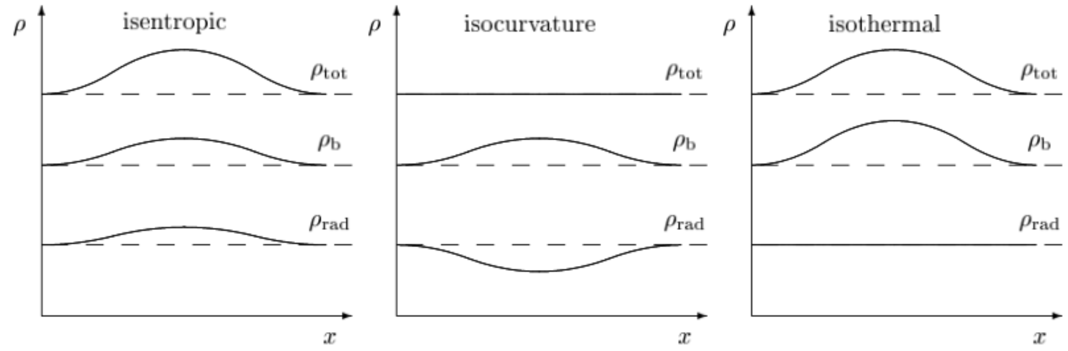

Figure 31.2: The left hand panel shows an 'adiabatic' or 'isentropic' perturbation of the kind we have been discussing above. In such a perturbation we crush the matter and radiation together. Such perturbations have a net density inhomogeneity and consequently have non-zero curvature or potential perturbations. A very natural alternative is to generate a perturbation in which the initial perturbation in the baryon density is cancelled by a corresponding under-density in the radiation. Such perturbations have, initially, no net density perturbation and therefore no associated curvature perturbation, and are called 'isocurvature'. For super-horizon scale perturbations the curvature is frozen in, but there is a non-zero pressure gradient, and once the perturbations enter the horizon this becomes effective and will act to annul the pressure gradient. In the example shown in the center panel, there is an inward directed pressure gradient which will act to erase the under-density in radiation, but in doing so will enhance the over-density in the baryons. The radiation density will over-shoot and one will have an oscillation about a state in which the radiation is uniform. The equilibrium state about which these oscillations will occur is shown in the right panel and is known as a 'isothermal' perturbation, since the radiation density, and therefore also the temperature, are constant.

Figure 2: Isentropic and isocurvature perturbations.

2 Scenarios for structure formation

2.1 The adiabatic, baryon dominated universe

- the first scenario to be explored in detail was the baryon dominated model with 'adiabatic' – what we would now call isentropic – initial conditions

- these evolve conserving the curvature perturbation while 'outside the horizon', but then oscillate like sound waves once they 'enter the horizon'

- this results in what, in the Soviet Union were known as *'Sakharov oscillations'* in the emergent power spectrum

  - calculating this involves a slight subtlety called the *'velocity overshoot effect"*.
    - recombination is quite rapid; the width of the last-scattering surface is about 10% of the horizon size.
    - if we were to model it as instantaneous, then what we would do is match the density and velocity of the sound wave, at the instant the pressure 'switches off', to a growing and decaying mode (and then calculate the output spectrum using only the growing mode)
    - the velocity in the sound wave is $v \sim c_{\rm s}\delta \sim (\lambda_{\rm J}/t)\delta$ while that in a growing mode is $v^+ \sim H\lambda\delta^+ \sim (\lambda/t)\delta^+$
    - so, for $\lambda < \lambda_{\rm J}$, the velocity associated with a sound wave of amplitude δ is greater than the growing mode velocity for the same amplitude
    - this argument suggests that, for $\lambda \ll \lambda_{\rm J}$, the sound wave is a combination of growing and decaying modes δ^+ and δ^- that nearly cancel
    - and that the final growing mode amplitude is obtained by matching the velocity (and it is actually waves that are at a temporal node of the density that couple most strongly to the growing modes at later times; hence the velocity overshoot terminology)
    - this analytic argument gives a sense of what is involved, but should not be taken too seriously, as, in reality, there isn't a great range of wavelengths below the maximum Jeans length for

which one can model recombination as instantaneous and one must rely on the numerical calculations (the first ones being done by Peebles and Yu).

- these are the *'baryonic acoustic oscillations'* discussed earlier and the evolution such sound waves is illustrated in figure 3

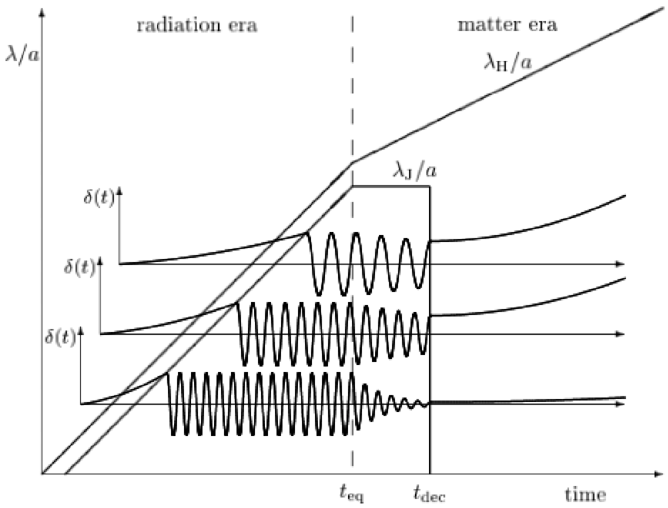

Figure 31.3: Evolution of initially adiabatic (or isentropic) perturbations is shown schematically for perturbations of three different wavelengths. The perturbation passes through three phases. First, when outside the horizon, the perturbation amplitude grows as $\delta \propto a^2$ in the radiation era and as $\delta \propto a$ in the matter dominated era. Perturbations which enter the horizon before t_{eq} oscillate at constant amplitude until t_{eq}. For $t_{eq} < t < t_{dec}$ the amplitude decays adiabatically as $\delta \propto 1/a^{1/4}$. Short wavelength perturbations are, in addition, subject to diffusive damping, and are strongly attenuated. Perturbations which persist to t_{dec} then couple to growing and decaying perturbations in the now pressure-free neutral gas.

Figure 3: Evolution of perturbations in a baryon dominated universe with adiabatic/isentropic initial conditions.

- The fact that the comoving Jeans length is unevolving during the interval $t_{eq} < t < t_{rec}$ (the latter also often called t_{dec} for *'decoupling'*) means that there is a 'step' in the power spectrum as modes with k bigger than the maximum Jeans frequency get damped (slowly) while those with slightly smaller k enjoy continual growth.

- This feature happens at a length scale comparable to that of super-clusters.

- The other feature in the power spectrum is the damping at higher k. This was invoked as a way to explain the mass of galaxies.

The figures here are schematic only. The details were worked out by Peebles and Yu and by Wilson and Silk in the 70s. The result is the *'transfer function'* giving the amplitude for modes emerging after decoupling relative to the initial value at horizon crossing. The latter was usually assumed to have the *'Harrison-Zel'dovich'* spectrum $P(k) \propto k$, for which the amplitude at horizon crossing – and therefore also the gravitational potential fluctuations (or the curvature fluctuations) – are scale invariant.

2.2 The 'hot dark matter' (HDM) scenario

The evidence for copious amounts of DM ('missing mass') in clusters had been around from the 30s with the work of Zwicky. But the late 70s and early 80s saw a strengthening of the evidence from rotation curves of galaxies and from relative motions of pairs of galaxies. Big bang nucleosynthesis suggested a small baryon density, and the idea that the universe may be dominated by *'non-baryonic'* dark matter – and that this might resolve the flatness problem – gathered strength.

One possibility for non-baryonic DM is a massive neutrino. The number density of neutrinos is known from the number of species (3 in the standard model) and from them being in thermal equilibrium in the

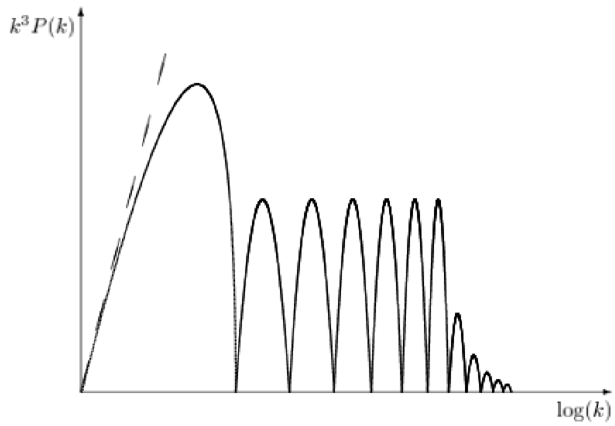

Figure 31.4: Power spectrum in the adiabatic-baryonic model (schematic). The dashed line indicates the initial power spectrum. The main peak is at a scale just larger than the maximum Jeans length, where the perturbations underwent continuous growth. Shorter waves entered the horizon before $z_{\rm eq}$ and subsequently oscillated, so their amplitude is suppressed. The nodes in the output spectrum are those wavelengths which have zero amplitude in the growing mode at the time of decoupling. The cut-off in the power spectrum at high k is due to diffusive damping.

Figure 4: Power spectrum of perturbations in a baryon dominated universe with adiabatic perturbations

early universe (so each species would be roughly as abundant as photons). In order for one of the species to give closure density requires a mass of about 30eV, and they would become non-relativistic close to $z_{\rm eq}$. Interestingly, there was a claimed measurement made in the Soviet Union that gave around this value (though this has subsequently been debunked) and which spurred much interest in this hypothesis.

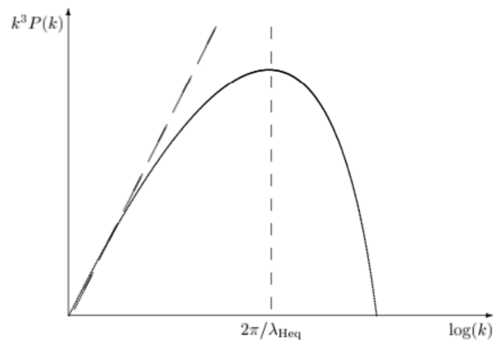

Figure 31.5: Power spectrum in the hot dark matter (HDM) model (schematic). The long-dashed line indicates the initial power spectrum. The vertical dashed line indicates the horizon size at the time the neutrinos become non-relativistic at $z \simeq z$eq. In the HDM model the first structures to form are super-cluster scale, and smaller scale-structures must form by fragmentation.

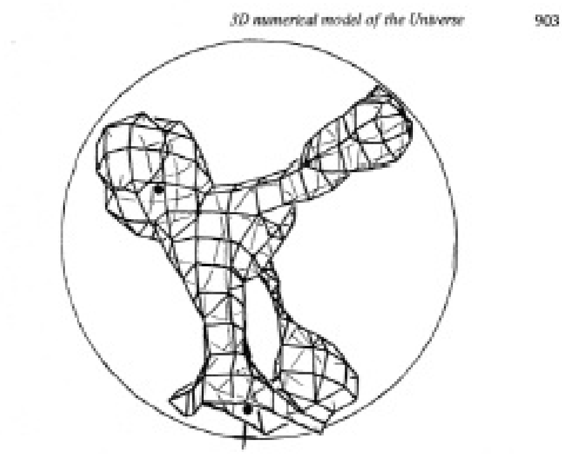

Figure 4. A surface of constant density level is plotted for the same region as that in Fig. 3.

Figure 5: Transfer function in the HDM scenario (left). Pioneering numerical simulation by Sergei Shandarin on right. This figure was dubbed the 'cosmic chicken'.

The evolution of perturbations in the HDM model (worked out by Dick Bond and George Efstathiou and others) is very different: perturbations entering the horizon when the neutrinos are still relativistic are washed out by free-streaming of the perturbations. There would still be some surviving contribution from the baryonic component on small scales, but these are very weak. The result is sketched in figure 5.

The formation of structure in this scenario is described as *'top-down'*, with the first structures being of supercluster scale. As pointed out by Zel'dovich, these would form, at first as large 'pancakes'. In the *'Zel'dovich approximation* particles are assumed to move essentially ballistically, but with velocities growing with time as $t^{1/3}$, and the emergence of structure is analogous to the pattern of caustics that form on the bottom of a swimming pool on a sunny day). Smaller scale structures, such as galaxies, were then assumed to have formed by fragmentation of these supercluster-scale pancakes (or blinis).

One very interesting observation was made by Tremaine and Gunn. The phase space density of the neutrinos is given initially by the Fermi-Dirac distribution and cannot increase (it may *decrease* in a *'coarse-grain average'* sense). They pointed out that the phase space density observed in the cores of clusters is uncomfortably high for this model.

2.3 The 'cold dark matter' (CDM) scenario

The next scenario to be explored in detail was the *cold dark matter* (CDM) model. Here the DM is some particle much heavier than the hypothetical massive neutrino. Common candidates are things like the super-symmetric (fermionic) partner to the graviton the graviton, the idea being that the lightest such particle would be stable as there is nothing it can decay to.

In the CDM model it is assumed that any thermal velocities of the particles are negligible; that they are initially on a 3-dimensional sub-space of the 6-dimensional phase space.

In the radiation dominated era these particles are only a tiny component of the density. As perturbations – assumed to be isentropic – enter the horizon, the dominant component (the radiation, tightly coupled to the baryons) starts to oscillate, and the gravitational potential fluctuations – which up to that point had been constant – rapidly diminishes and growth of the density perturbations of the CMB 'stagnates'. But it is not washed out, as in the HDM scenario, rather the amplitude of the perturbations remains, broadly speaking, equal to its value at horizon crossing. Once the CDM comes to dominate over the radiation density, growth recommences. The resulting 'transfer function' is sketched in figure 6.

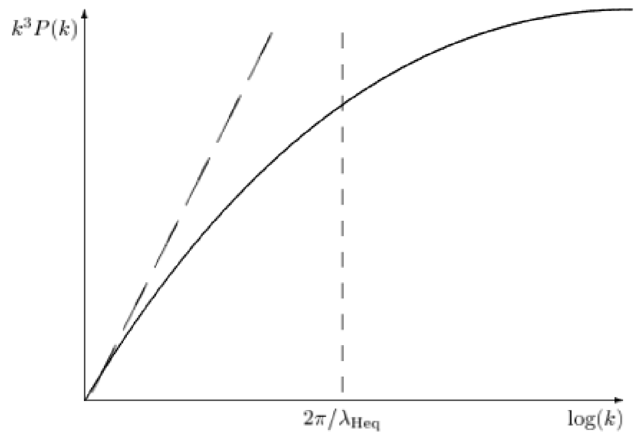

Figure 31.6: Power spectrum in the cold dark matter (CDM) model (schematic). The long-dashed line indicates the initial power spectrum. The vertical dashed line indicates the horizon size at the time the neutrinos become non-relativistic at $z \simeq z$eq.

Figure 6: Transfer function for density perturbations in a universe dominated by cold dark matter.

With the canonical 'Harrison-Zel'dovich' scale invariant $n = 1$ initial spectrum the output spectrum is $n = 1$ for wavelength much larger than the horizon size at $z_{\rm eq}$ but $n = -3$ on small scales. With the asymptotic small-scale $n = -3$ index, the initial linear theory variance is growing logarithmically with wave-number. That means that the first objects to turn around and collapse and virialise will be small 'mini-haloes'. But the expectation is for a rapid increase in the mass of collapsed structures until one reaches scales where n is significantly greater than -3, at which point one expects a 'hierarchical' growth of structure with the small haloes being incorporated into successively larger haloes.

But the transition is rather gradual. As was already appreciated at the time – particularly by Richard Gott and Martin Rees – the observed clustering of galaxies (on scales of $\sim 1-10$ Mpc) was in good agreement with what would arise from a post-decoupling spectrum with $n \simeq -1$. Detailed calculations – again Dick Bond and George Efstathiou were highly influential – showed that CDM delivers the goods.

An essential difference between the scenarios described above is the difference in the strength of the fluctuations at large scale required in order to provide formation of galaxies and structure as observed. It is lowest in the CDM model. This proved to be the undoing of the alternative scenarios. But in the early 80s this was not yet known.

2.3.1 The baryonic wiggles

The transfer function sketched in figure 6 is what one would get for a universe with only CDM and without any, or with only a negligible amount of, baryons. But in fact, the baryons are not negligible; they constitute about 20% of the mass density. So the emergent spectrum after decoupling will contain, in addition to the fluctuations in the CDM density that of the baryons which, as we have seen, have Sakharov oscillations[4].

The baryon density being sub-dominant, and the sound wave having damped slightly compared to its horizon crossing value, the baryonic perturbations may give an enhancement or a diminution of the total $\delta\rho = \delta\rho_c + \delta\rho_b$ depending on the temporal phase of the sound wave at decoupling.

The resulting wiggles in the spectrum are weak, but are important as they can be measured in the spatial distribution of galaxies and they provide a 'standard ruler' of known comoving length (essentially the maximum comoving Jeans length) and can be used to measure $D_a(z)$ as we describe later.

3 Non-Linear Cosmological Structure Models

In chapter **??** we explored the evolution of small amplitude perturbations of otherwise homogeneous cosmological models. This provides an accurate description of the evolution of structure from very early times. On sufficiently large scales, the structure is still in the linear regime today, but small scale structures have reached the point where $\delta\rho/\rho \gtrsim 1$ and have gone *non-linear*.

When dealing with the development of non-linear structure we can usually neglect radiation pressure and assume that the structures are much smaller than the horizon scale, so a Newtonian treatment is valid. However, the equations of motion are still relatively complicated and it is hard to find exact solutions except in highly idealized models such as spherical or planar 1-dimensional collapse. One approach to non-linear structure growth is to attempt to evolve the initial conditions forward from the linear regime numerically using either *N-body simulations*, to evolve the collisionless Boltzmann equation, or *hydrodynamical simulations* to evolve the Euler, energy and continuity equations. The former is adequate to describe the evolution of collisionless dark matter matter, but the latter is required if one also wants to treat the baryonic matter. Another possibility is to extend perturbation theory beyond linear order. This is an area where there has been much activity by theorists in recent years. These calculations typically assume a Gaussian initial density field, and then compute the emergence of non-Gaussianity, e.g. the skewness, or the kurtosis of the density distribution. Such results are limited to the 'quasi-linear' regime; i.e. density contrasts $\delta \lesssim 1$. This is a rather limited range of validity. Also, since most interest is in theories with 'hierarchical' initial fluctuation spectra, when one scale is just going non-linear, there are smaller scale structures which will be highly non-linear. Usually such calculations deal with this by assuming some smoothing of the initial δ-field, but the validity of this is questionable.

Here I shall describe a number of approximate methods and models that directly address the 'quite-strongly non-linear' regime. These models are typically quite idealized, but they are still useful as they provide insight into the way structure has evolved, and is evolving today.

3.1 A simple model for the formation of galaxy clusters

- We can estimate what we would expect for the density, and *density contrast*, of a recently virialised cluster using the following simple model:
  - we assume there is a 'background' cosmology which, for simplicity, is Einstein-de Sitter, so $r \propto t^{2/3}$
  - in that background we carve out a sphere of mass M
    - which, in the background, would have been marginally bound to itself
  - and we replace it by a sphere of the same mass with negative binding energy that will expand to some maximum radius an then recollapse

- at the time $t_{\rm max}$ of maximum expansion the kinetic energy K was zero and the potential energy was $U(t_{\rm max}) \sim GM/r_{\rm max}$ with some coefficient determined by the shape
- the virial theorem tells us that, after it has collapsed and virialised, it will have $2K + U = 0$
- with $E = K + U = U(t_{\rm max})$, and $K = -U/2$, this implies that the final binding energy must be

[4] In the context of the CDM model, there is not expected to be an extended period before decoupling with constant Jeans length. The baryon density in these models is $\Omega_b \simeq 0.05$ (consistent with BBN results) while the CMB radiation has present $\Omega_r \simeq 5 \times 10^{-5}$. That means that at $z = 1000$, $\rho_b \simeq \rho_r$ and the sound speed was still $c_s \simeq c$. This means that the maximum comoving sound horizon is essentially independent of $\Omega_b \simeq 0.05$.

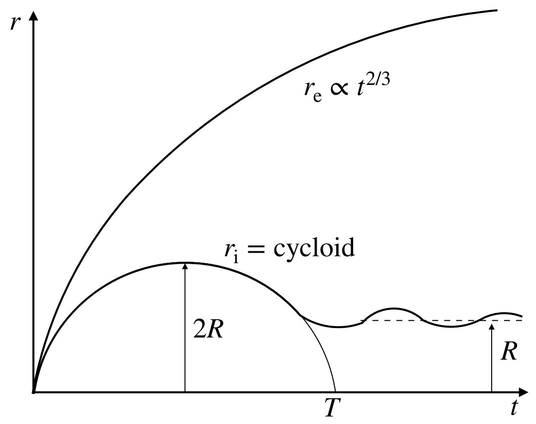

Figure 7: A simple model for the formation of a cluster is that it was initially a uniform density sphere 'carved out' of a uniform density 'background' universe but with a lower total energy, so it was gravitationally bound to itself and therefore doomed to expand only to some maximum radius and then collapse. The virial theorem tells us that, in order to generate enough kinetic energy to satisfy $2K+U=0$, it must collapse by about a factor 2 from its maximum size. This gives the Kepleresque relation between radius, mass and time of collapse: $R = \sqrt[3]{GM/4\pi^2}\, T^{2/3}$. If the background universe were of marginally bound Einstein-de Sitter form, it would have $r_e(T) = \sqrt[3]{9GM/2}\, T^{2/3}$. It follows that the recently virialised object should have a density contrast $\rho/\overline{\rho} = (r_e/R)^3 = 18\pi^2 \simeq 200$.

  - $\boxed{U(t_{\rm vir}) = 2U(t_{\rm max})}$

- so, since $U \propto 1/r$, it must have collapsed by a factor 2 in order to generate the kinetic energy required to stabilise itself against further collapse
- the equations of motion are
  - $$\begin{array}{rl} \text{exterior}: & v^2 = 2GM/r \\ \text{interior}: & v^2 = 2GM/r - \text{constant} \end{array}$$
- the solution for the exterior is
  - $r_{\rm e} = \alpha t^{2/3}$
  - with α a constant, which implies, for the velocity
  - $v = dr/dt = (2/3)\alpha t^{-1/3} \Rightarrow v^2 = (4/9)\alpha^2 t^{-2/3} = (4/9)\alpha^3/r$
  - and which, with the equation of motion, implies $\alpha^3 = 9GM/2$ and so
  - $\boxed{r_{\rm e}(t) = \sqrt[3]{9GM/2}\, t^{2/3}}$
- the solution for the interior is the cycloid
  - $$\begin{array}{r} r = R(1-\cos\eta) \\ t = (T/2\pi)(\eta - \sin\eta) \end{array}$$
  - which we can verify by computing $v = dr/dt = (dr/d\eta)/(dt/d\eta) = (2\pi R/T)\sin\eta/(1-\cos\eta)$
  - which, with a little algebra, implies $v^2 = (2\pi R/T)^2(2/(1-\cos\eta)-1)$, or
  - $v^2 = 8\pi^2 R^3/T^2 r - 4\pi R^2/T^2$
  - and which, with the equation of motion, implies
  - $\boxed{R = \sqrt[3]{GM/4\pi^2}\, T^{2/3}}$
- comparing these we can estimate the density of a recently virialised object with respect to that of a critical density of the same age as
  - $\boxed{\rho/\overline{\rho} = (r_e(t=T)/R)^3 = 18\pi^2 \simeq 200}$
- while a crude model – it neglects completely the effect of dark energy, for instance – this is borne out by numerical simulations
  - these show that, if we consider a sphere around a simulated cluster, or 'halo', within which the density contrast is about 200 then this delineates quite well the exterior 'infall region' from the virialised interior where we have multiple streams of matter
- it also turns out, for rich clusters, with velocity dispersions of about 1000km/sec, to be about the same as the 'Abell radius' (1.5Mpc/h) that George Abell arrived at empirically

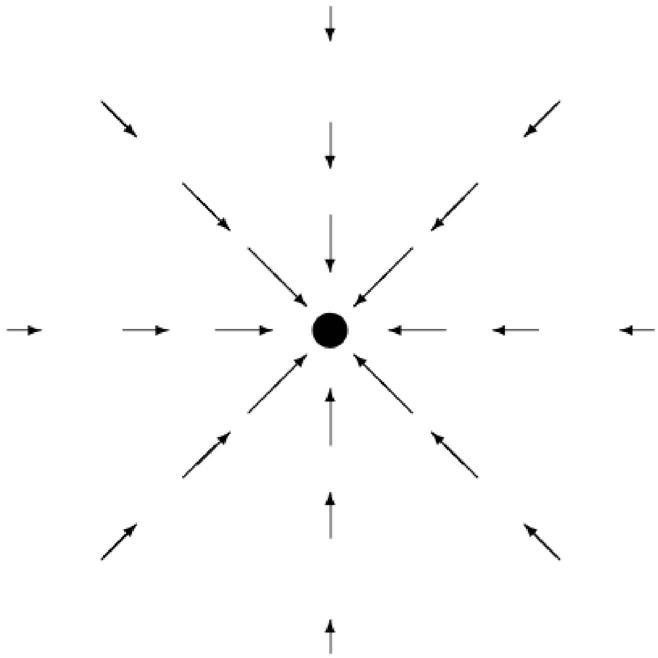

Figure 8: Illustration of the kind of divergence free flow pattern induced by a point mass. The peculiar velocity falls off as $\delta v \propto 1/R^2$, with the consequence that the flux of matter through any a shell of radius R is independent of R. For $R > 0$ the density remains unperturbed, but there is a net accumulation of mass at the origin.

3.2 Gunn-Gott Spherical Accretion Model

Another very illuminating model is that of Gunn and Gott (19??) who considered what happens if one introduces a point-like 'seed' of mass M_0 into an otherwise uniform Einstein -de Sitter universe.

First consider the linear theory. At large radii, the mass will induce a peculiar acceleration at physical distance R of

$$g = \frac{GM_0}{R^2}. \tag{34}$$

Acting over a Hubble time $t \sim 1/H$ this will generate a peculiar infall velocity

$$\delta v \sim gt \sim \frac{GM_0}{HR^2}. \tag{35}$$

This kind of $\delta v \propto 1/R^2$ flow (see figure 8) is 'divergence free', so there is no change in the density at large distances (think of two concentric comoving shells; the flux of matter across a surface is the velocity times the area and is independent of radius, so there is no build up except at the center). The amount of mass convected across a shell in one Hubble time is $\delta M \sim \overline{\rho} R^2 \delta v t$ which, with (35) and $H^2 = 8\pi G\overline{\rho}/3$, gives $\delta M \sim M_0$. Thus the seed induces, after one expansion time, a growing mode density perturbation $(\delta\rho/\rho)_i \sim M_0/M$, and this subsequently grows with time as $\delta\rho/\rho = \delta M(t)/M \propto a(t)$. This is assuming, for simplicity, an Einstein - de Sitter background.

The amount of mass accumulated in the center is therefore

$$\delta M(t) \sim M_0 \frac{a}{a_i} \propto a(t). \tag{36}$$

This mass represents a density contrast of order unity at a physical radius R such that $\delta M \sim \overline{\rho} R^3$, and slightly inside will lie the turnaround radius

$$R_{\rm turn} \sim (\delta M/\overline{\rho})^{1/3} \propto a^{4/3}. \tag{37}$$

As time goes on, progressively larger shells will turn around, collapse and virialize in some complicated way with shell crossing etc. However, we may reasonably expect that the final specific binding energy of a shell of a certain mass will be equal, modulo some factor of order unity, to its initial specific binding energy $\delta\phi$. Now the initial binding energy is a power law in radius:

$$\delta\phi \sim GM_0/R \propto 1/R \propto M^{-1/3}, \tag{38}$$

whereas the final binding energy as a function of the final radius R_f is

$$\delta\phi \sim GM(R_f)/R_f \tag{39}$$

where $M(R_f)$ is the mass within radius R_f. Equating these gives the scaling law

$$M(R_f) \propto R_f^{3/4}. \tag{40}$$

If the mass within radius R_f is a power law in R_f then so also is the density:

$$\rho(R_f) \sim M(R_f)/R_f^3 \propto R_f^{-9/4}. \tag{41}$$

This analysis then tells us that the virialized system should have a power law density profile. What is interesting about this result is that it is very close to the $\rho(R) \sim R^{-2}$ density run for a flat rotation curve halo, and also similar to the profile of clusters of galaxies, which are also often modeled as 'isothermal spheres'.

3.3 The Zel'dovich Approximation

In linear theory, and for growing perturbations in an Einstein - de Sitter model, particles move with peculiar velocity

$$\mathbf{v}(\mathbf{r}, t) = (t/t_0)^{1/3}\mathbf{v}_0(\mathbf{r}) \tag{42}$$

where now $\mathbf{r}$ is a comoving spatial coordinate and $\mathbf{v}_0$ is the peculiar velocity at some initial time. This says that the peculiar velocity field just grows with time at the same rate $\mathbf{v} \propto t^{1/3}$ at all points in space.

The physical displacement of a particle in time dt is $d\mathbf{x} = \mathbf{v}dt$, so the comoving displacement is

$$d\mathbf{r} = \frac{d\mathbf{x}}{a} = \frac{\mathbf{v}}{a}dt \equiv \mathbf{u}dt \tag{43}$$

where the comoving peculiar velocity is $\mathbf{u} = \mathbf{v}/a$. The rate of change of comoving position with scale factor a is then

$$\frac{d\mathbf{r}}{da} = \frac{\mathbf{v}}{a}\frac{dt}{da}, \tag{44}$$

but with $\mathbf{v} \propto t^{1/3}$ and $a \propto t^{2/3}$, so $da/dt \propto t^{-1/3}$, this says that

$$\frac{d\mathbf{r}}{da} \propto t^0. \tag{45}$$

Therefore, if we define a new 'time' $\tau \propto a$, then the particles move ballistically in comoving coordinate space: $d\mathbf{r}/d\tau =$ constant. Zel'dovich's approximation is to assume that this ballistic motion continues into the non-linear regime.

The result is a *Lagrangian mapping* resulting in formation of *caustics*, or surfaces of infinite density. This is very analogous to the formation of caustics on the swimming pool floor, which we explored in our study of geometric optics in chapter **??**. There the horizontal deflection of the rays — a 2-dimensional vector displacement — increases linearly with distance from the surface, and here the 3-dimensional comoving displacement increases linearly with 'time' τ. We can write the actual comoving position or *Eulerian coordinate* $\mathbf{x}$ as a function of the initial or *Lagrangian coordinate* $\mathbf{r}$ as

$$\mathbf{x}(\mathbf{r}) = \mathbf{r} + \tau\mathbf{U}(\mathbf{r}) \tag{46}$$

where $\mathbf{U}$ is a suitably scaled version of $\mathbf{u}$.

Until caustics form, the density is

$$\rho \propto \left|\frac{\partial x_i}{\partial r_j}\right|^{-1} \tag{47}$$

where $|\partial\mathbf{x}/\partial\mathbf{r}|$ is the Jacobian of the transformation from Lagrangian to Eulerian coordinate. This is just conservation of mass: $dM = \rho_L d^3r = \rho_E d^3x$, with ρ_E and ρ_L the densities in Eulerian and Lagrangian space respectively. Now from (46), $\partial x_i/dr_j = \delta_{ij} + \tau\partial U_i/\partial r_j$, so we can also write the density as

$$\rho \propto \frac{1}{(1+\tau\lambda_1)(1+\tau\lambda_2)(1+\tau\lambda_3)} \tag{48}$$

where the λ_i are the eigenvalues of the *deformation tensor* $\Phi_{ij} = \partial U_i / \partial r_j$.

If there is a negative eigenvalue $\lambda_1 < \lambda_2, \lambda_3$, then the density of a small comoving volume of matter will become infinite with collapse along the appropriate *principle axis* when $\tau = -1/\lambda_1$. *Pancakes* — or perhaps we should call then *blinis* — form with a multi-stream region sandwiched between the caustic surfaces. These pancakes grow rapidly and intersect to form a cellular network of walls or pancakes intersecting in lines with the matter in these lines draining into the nodes where all three of the eigenvalues of $\partial U_i / \partial r_j$ go negative.

The Zel'dovich approximation seems to give a good picture of formation of structure in the HDM model, but continued unaccelerated motion of particles after shell crossing is clearly unrealistic. A useful modification of Zel'dovich's approximation is to assume that particles move ballistically until shell-crossing, at which point they stick together. This is described by *Burger's equation*, and gives infinitesimally thin walls. This is also obviously unrealistic, but actually receives some justification if we think of the Universal expansion adiabatically stretching a self-gravitating sheet. If the thickness of the sheet is T and the surface density Σ, then the acceleration of a particle at the surface is $\ddot{r} \sim G\Sigma$, and the frequency of oscillation of particles through the sheet is

$$\omega \sim \sqrt{\frac{\ddot{r}}{T}} \sim \sqrt{\frac{G\Sigma}{T}}. \tag{49}$$

Now the surface density decreases as $\Sigma \propto 1/a^2$ for a sheet expanding in the transverse direction at the Hubble rate, so applying the law of adiabatic invariance $r = A\cos(\omega t)$ with amplitude $A \propto 1/\sqrt{\omega}$ and requiring $A \simeq T$ we find that the (physical) thickness must evolve as

$$T \propto a^{2/3}. \tag{50}$$

This increases with time, but not as fast as the scale factor $a(t)$, so in comoving coordinates the sheet should indeed become thin.

Another nice feature of the Zel'dovich approximation is that one can compute the non-linear power spectrum analytically in terms of the initial power spectrum, as described in detail in chapter **??**.

3.4 Press-Schechter Mass Function

The *Press-Schechter approximation* is designed for hierarchical type initial fluctuation fields. It provides one with a useful approximation for the *differential mass function* $n(M) = dN(> M)/dM$, where the *cumulative mass function* $N(> M)$ is the comoving number density of bound structures with mass $> M$.

The idea is that one identify two quantities: The first is the fraction of space where the initial density contrast field $\delta(\mathbf{r})$, when filtered with a kernel of mass M, lies above the threshold $\delta_{\rm crit}$ for formation of non-linear condensations.

$$f(\delta > \delta_{\rm crit}; M) = \int_{\delta_{\rm crit}/\sigma(M)}^{\infty} \frac{d\nu}{\sqrt{2\pi}} \exp\bigl(-\nu^2/2\bigr). \tag{51}$$

The second is the fraction of mass in objects more massive than M

$$f(> M) = \int_M^\infty dM M n(M). \tag{52}$$

Differentiating (51) and (52) with respect to M and equating we get

$$n(M) = \frac{\delta_{\rm crit} d\sigma(M)/dM}{\sqrt{2\pi} M \sigma^2(M)} \exp\bigl(-\delta_{\rm crit}^2 / 2\sigma^2(M)\bigr) \tag{53}$$

While hard to justify rigorously, the idea obviously contains an element of truth, and moreover seems to give predictions which agree with the results of N-body experiments.

If one assumes a power-law spectrum $P(k) \propto k^n$ then the variance as a function of smoothing mass M is also a power law, $\sigma^2(M) \propto M^{-(n+3)/3}$. In 'hierarchical models' (those with $n > -3$) the mass variance increases with decreasing mass. At sufficiently low masses we must have $\sigma \gg \delta_{\rm crit}$ and the exponential factor becomes close to unity and the theory predicts a power-law differential mass function. For $n = -2$, for instance, which is the slope of the CDM spectrum around the mass scale of galaxies the theory predicts $n(M) \propto M^{-5/6}$. At high masses, when $\delta_{\rm crit}/\sigma(M)$ starts to exceed unity the exponential factor becomes very small. The general prediction is for a power-law mass function which becomes exponentially cut off above some characteristic mass scale; the mass $M_\star$ where $\sigma(M_\star) \simeq 1$. This is just the kind of behavior seen in the *galaxy luminosity function* and also in the *cluster mass function*.

3.5 Biased Clustering

In the Press-Schechter theory, collapsed objects are associated with regions where the initial over-density, smoothed on an appropriate mass-scale, is sufficiently large. A consequence of this is that objects on the high end of the mass function — those with $M \gtrsim M_\star$ that is — will tend to have amplified large large-scale clustering properties. Their clustering is said to be positively biased. The effect is illustrated in figure 9 which shows how the density of over-dense regions is modulated by long wavelength modes of the density field. This effect is fairly obvious, but what is less obvious is how the strength of the modulation increases as one raises the threshold. This is consequence of the peculiar property of a Gaussian distribution. The Gaussian distribution is $P(\nu) \propto \exp\big(-\nu^2/2\big)$. In the vicinity of some value $\nu = \nu_0$ the distribution for $\Delta\nu = \nu - \nu_0$ is $P(\Delta\nu) \propto \exp\big(-(\nu_0 + \Delta\nu)^2/2\big)$. Expanding the quadratic factor in the exponential, and assuming $\nu_0 \gg \Delta\nu$ gives

$$P(\Delta\nu) \sim \exp(-\nu_0 \Delta\nu). \tag{54}$$

Thus a Gaussian looks locally exponential: $P(\Delta\nu) \sim \exp(-\Delta\nu/\sigma_{\Delta\nu})$ with exponential scale length $\sigma_{\Delta\nu} = 1/\nu_0$ which decreases with increasing ν_0. Thus the further out we go on the tail of a Gaussian the steeper the distribution becomes.

If we add a positive background field δ_b, the fractional change in the probability to exceed the threshold is then $\Delta P/P \simeq \nu_0 \delta_b/\sigma$. The fluctuation in the number density of upward excursions is then

$$1 + \frac{\delta n}{n} = 1 + b\delta_b \tag{55}$$

where the *bias factor* is

$$b = \frac{\nu_0}{\sigma} = \frac{\delta_{\rm crit}}{\sigma^2}. \tag{56}$$

Since $\delta_{\rm crit}$ is constant here, the bias factor rapidly increases with the mass of the objects (because $\sigma^2(M)$ decreases with increasing mass). This is the linearized bias; valid for very small δ_b, such that $b\delta_b \ll 1$. It is not difficult to show that for $\delta_b \lesssim 1$, the density of upward fluctuations is proportional to $\exp(b\delta_b)$. Thus the density of objects is the exponential of the background field.

One solid application of this theory is to clusters of galaxies; these are the most massive gravitationally collapsed objects, and so are naturally identified with particularly high peaks. For a long time, the very strong clustering of such objects was a puzzle; they have a correlation length of about 20Mpc as compared to about 5Mpc for galaxies. Now we understand that this is just about what one would expect given Gaussian initial density fluctuations. It is tempting to apply this theory also to galaxies, but there the connection between theory and observation is more tenuous. However, at high redshift one would expect the rare, most massive galaxies to be the analog of very massive clusters today, and this theory then provides a natural explanation for the rather strong clustering of 'Lyman-break' galaxies at $z \sim 3$.

3.6 Evolution of the cluster mass function

- the cluster mass- or X-ray luminosity-function has a form rather similar to the galaxy luminosity function with a 'knee', above which the number of clusters drops exponentially
- the evolution of the cluster mass function $n(M)dM$ depends on
  1. the initial seeds for structure
  2. how the structure grows (which depends on expansion history)

3.6.1 Self-similar evolution

- during the matter dominated era
  - which, it used to be thought continued up to the present day,

  the 'background' cosmology is 'scale invariant': density, scale factor etc. just vary as power laws with time.
- and the primordial fluctuations are also approximately scale invariant

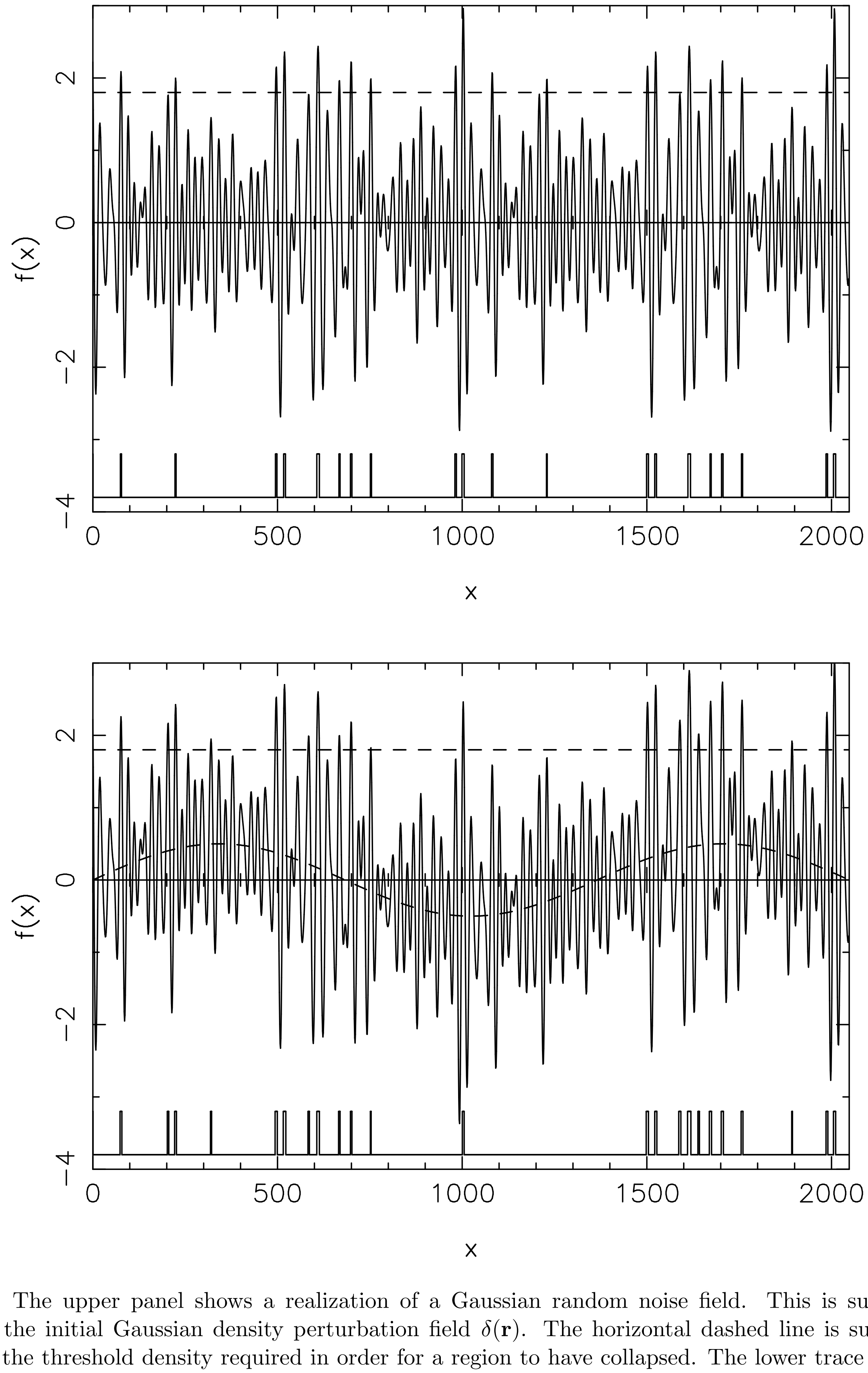

Figure 9: The upper panel shows a realization of a Gaussian random noise field. This is supposed to represent the initial Gaussian density perturbation field $\delta(\mathbf{r})$. The horizontal dashed line is supposed to represent the threshold density required in order for a region to have collapsed. The lower trace shows the 'excursion set' for this threshold (here taken to be 1.8 times the root mean squared fluctuation. This function is one or zero depending on whether $f(x)$ exceeds the threshold. The positive parts of the excursion set are randomly distributed with position. The lower panel shows the same thing, but where we have added a long-wavelength sinusoidal 'background' field. Clearly, and not surprisingly, the background field has modulated the density of the regions exceeding the critical threshold.

- e.g. Gaussian random field with power-law power spectrum $\Delta^2_\rho(k) \equiv k^3 P_\rho(k) \propto k^{n+3}$
- where k is a 'co-moving' wavenumber $\mathbf{k} = a(t)\mathbf{k}_{\rm phys}$
- and $n \simeq -1$ on relevant scales (spectral index varies slowly with wavelength

the fact that (as we will see later) the density fluctuations grow with time as $\Delta\rho/\rho \propto a(t) \propto 1/(1+z)$ means that the wave-number (inverse comoving scale) of non-linearity grows like

- $\boxed{k_\star \propto (1+z)^{2/(n+3)}}$
- and hence the characteristic mass (knee of the mass function) varies as
- $\boxed{M_\star \propto k_\star^{-3} \propto (1+z)^{-6/(n+3)}}$
- which is quite a strong rate of evolution

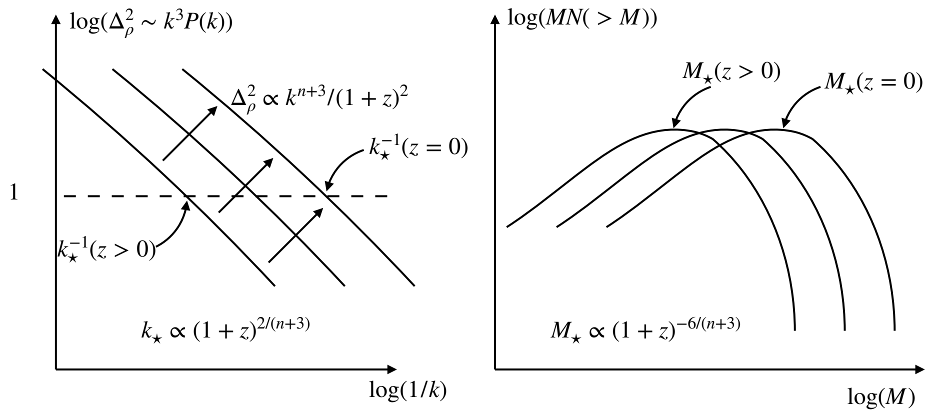

Figure 10: Self-similar model for evolution of the cluster mass function. Left panel show the evolution of the power spectrum of density perturbations for (nearly) scale invariant initial conditions. In a Einstein-de Sitter model this evolves with time as $(1+z)^2$ so the 'characteristic scale of non-linearity $k_\star^{-1}$ increases with time as a power-law also. Right hand panel shows mass function.

- the radius scales like the cube root of the mass divided by the density
- the latter scales like $\rho \propto (1+3)^3$ so the characteristic radius scales as $R_\star \propto a/k_\star \propto (1+z)^{-1}(1+z)^{-2/n+3}$ or
- $\boxed{R_\star \propto (1+z)^{-(n+5)/(n+3)}}$
- and the characteristic temperature (from hydrostatic equilibrium) goes like $T_\star \propto M_\star/R_\star$ or
- $\boxed{T_\star \propto (1+z)^{(n-1)/(n+3)}}$
  - the idea here is that the clustering is growing 'hierarchically' with small halos merging into larger ones as the universe ages
  - this is a highly complex process – at the time these models were developed it was not possible to model this using hydrodynamical simulations. Even today this is challenging.
  - the beauty of the model is that the scale invariance of the initial fluctuations and the background within which these are evolving means that one can predict the population at one time from observations at another simply by scaling the physical quantities appropriately
- this model worked quite well, but not perfectly
  - understandable since lower-mass clusters were quite likely to have been affected by early energy ejection
  - gives the gas an entropy larger than would have arisen from shocking in the self-similar evolution

but even allowing for this (e.g. focussing on high-mass end of distribution function) there was a problem

- models predicted too rapid evolution
- this was an early indications of the need for dark-energy or cosmological constant (ΛCDM) (e.g. Pat Henry)
  - along with problems with the age of the stars vs. age of the universe

3.7 Davis and Peebles Scaling Solution

The goes beyond the self-similar scaling and attempts to determine the slope of the two-point correlation function in the non-linear regime from the slope n of the initial power spectrum, assumed to be power-law like with $P(k) \propto k^n$.

The original discussion was couched in terms of the BBGKY hierarchy, but the essential result can be easily obtained from conservation of energy considerations, much as we did for the accretion onto a point mass.

With the initial spectrum for the density fluctuations δ and with $\nabla^2 \phi = 4\pi G \rho \delta$, so $\delta\phi_k = 4\pi G \rho \delta_k / k^2$ the root mean square potential fluctuations on scale r are

$$\langle \delta\phi^2 \rangle_r^{1/2} \sim \left[\int d^3k k^{-4} k^n \tilde{W}_r(k) \right]^{1/2} \tag{57}$$

where $\tilde{W}_r(k)$ is the transform of the smoothing kernel, which falls rapidly for $k \gg 1/r$. This gives

$$\langle \delta\phi^2 \rangle_r^{1/2} \propto r^{(1-n)/2}. \tag{58}$$

In terms of mass scale, $M \propto r^{1/3}$ this is

$$\langle \delta\phi^2 \rangle_r^{1/2} \propto M^{(1-n)/6}. \tag{59}$$

One the other hand, in the non-linear regime, we have a power-law mass auto-correlation function $\xi(r) \propto r^{-\gamma}$. Now imagine the mass distribution to be a set of randomly distributed clumps of size r and over-density $\delta_* \gg 1$. The fraction of space occupied by the clumps is $f \sim 1/\delta_*$, so the density fluctuation variance is $\xi(r) \simeq \langle \delta^2 \rangle \sim f \delta_*^2 \sim \delta_*$. The mass of a lump is $M \sim \overline{\rho} \delta_* r^3$, so the characteristic mass of clumps of size r is $M \propto r^{3-\gamma}$. The binding energy of clumps then scales with their radius and mass as

$$\delta\phi \sim M/r \propto r^{2-\gamma} \propto M^{(2-\gamma)/(3-\gamma)}. \tag{60}$$

Equating (59) and (60), we obtain the relation

$$\gamma = \frac{9+3n}{5+n} \tag{61}$$

which would fit with the empirically observed slope $\gamma \simeq 1.8$ for a white noise spectrum $n = 0$.

While the derivation here is similar to that for spherical accretion, the result is much less robust. While it makes perfect sense to say that the binding energy of structures when they first form is given, within a geometrical factor of order unity, by the initial binding energy, the calculation here assumes that even when much larger mass objects have collapsed, the small clumps still preserve the binding energy with which they are born. This is not likely to be the case, as there will be transfer of energy between the different scales of the hierarchy. As we have argued above, entropy considerations suggest that such interactions will tend to erase sub-structure. Numerical simulations do not provide much support for this theory.

3.8 Cosmic Virial Theorem

The *cosmic virial theorem* (Davis and Peebles again) attempts to relate the low order correlation functions for galaxies to the relative motions of galaxies and thereby obtain an estimate of the mass-to-light ratio of mass clustered along with galaxies.

In essence, their argument is as follows: Assume that galaxies cluster like the mass — this means that the excess mass within distance r of a galaxy grows like $M \propto \int^r d^3r\ \xi(r) \propto r^{3-\gamma}$. The potential well depth is then $\delta\phi \sim GM/r \propto r^{2-\gamma}$. One would expect the relative velocity of galaxies at separation r to scale as

$$\sigma^2(r) \propto r^{2-\gamma} \simeq r^{0.2} \tag{62}$$

This prediction seems to be remarkably well obeyed on scales from a few tens of kpc out to about 1 Mpc (and one would not expect the result to hold at larger separations where things have yet to stabilize anyway).

From the size of the peculiar motions, one infers that the mass-to-light ratio of material clustered around galaxies on scale ~ 1 Mpc or less is $M/L \simeq 300h$ in solar units. If representative of the universal value, this would imply $\Omega \simeq 0.2$. This is similar to the mass-to-light ratio from virial analysis of individual clusters of galaxies, and provides strong supporting evidence for copious amounts of dark matter. It also supports the hypothesis that the galaxies cluster like the mass, and therefore that the universal density parameter is $\Omega \simeq 0.2$ rather than the aesthetically pleasing $\Omega = 1$.