L3-9.

Cosmology 3 - Cosmological Structure Formation

Nick Kaiser

March 29, 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

- We already considered the non-linear version of this model as a model for formation of clusters.
  - recall that we excise a sphere of the dust and replace it by a slightly smaller expanding sphere
- here we will look in more detail at how the density fluctuation evolves with time
  - for clusters we considered the case where the interior was on a bound orbit
  - but 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'
- 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

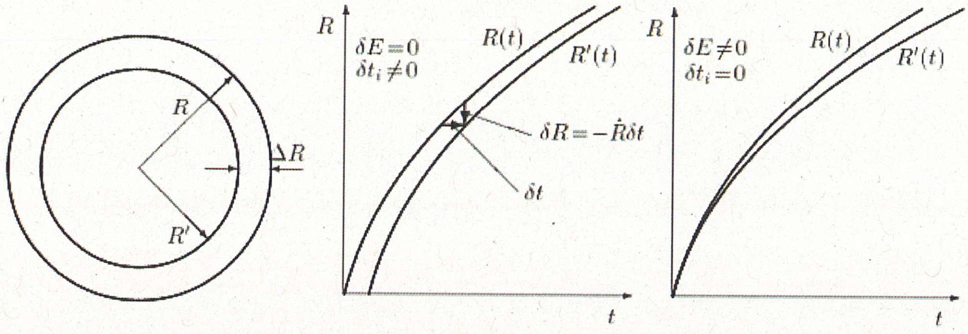

Figure 31.1: One can generate a perturbation of a dust-filled cosmology by excising a sphere of matter and replacing it with a smaller sphere of radius R'. The middle panel illustrates a decaying perturbation produced by a 'delayed bang'. The right panel shows the more interesting growing perturbation 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.

- it follows that
  - $v\delta v =$ constant so $\delta v = \delta\dot{R} \propto 1/v \propto \sqrt{R/GM} \propto t^{1/3}$
  - 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 $\delta R \sim t\delta\dot{R} \propto t^{4/3}$ 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 $\mathbf{v}_{\text{phys}}(\mathbf{r}, t)$ where $\mathbf{r}$ is the physical position
- Define *comoving spatial coordinates*
  - $\boxed{\mathbf{x} = \mathbf{r}/a(t)}$
- and define the density perturtbation
  - $\delta(\mathbf{x}, t) = \frac{\delta\rho(\mathbf{x},t)}{\overline{\rho}(t)} = \frac{\rho(a\mathbf{r},t)-\overline{\rho}(t)}{\overline{\rho}(t)}$
- and the *peculiar velocity*
  - $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_{\text{phys}}(a\mathbf{x}, t) - H\mathbf{r}$
- 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
- thus, given some initial density and (longitudinal) velocity perturbation we can calculate the amplitude of the growing mode $\delta^+(\mathbf{x}, t_i)$ which, at later times, will 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

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

1.1.3 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, absent any gravitational acceleration or pressure gradient, suffer a *cosmic drag*:

$$\dot{\mathbf{v}} = -H\mathbf{v}. \tag{1}$$

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

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

Adding the effect of the density perturbations, which 'source' a *peculiar gravity* $\delta\phi$ satisfying

$$\nabla_{\mathbf{r}}^2 \delta\phi = -4\pi G\delta\rho \tag{2}$$

(where the subscript on the Laplacian shows it is Laplacian with respect to physical coordinates) the equation of motion above becomes

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

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

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

where we have switched the spatial partial derivative to $\boldsymbol{\nabla}_{\mathbf{x}} = a\boldsymbol{\nabla}_{\mathbf{r}}$.

This is a linear equation so, if we decompose the potential and (longitudinal) velocity into comoving Fourier modes $\propto e^{i\mathbf{k}\cdot\mathbf{x}}$, applies to each mode individually. The spatial derivative operator applied to such a mode has $\boldsymbol{\nabla}_{\mathbf{x}} e^{i\mathbf{k}\cdot\mathbf{x}} = i\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{x}}$, so with $\mathbf{u}(\mathbf{x}, t) = \sum_{\mathbf{k}} u_{\mathbf{k}}(t)\hat{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}$ and $\delta\phi(\mathbf{x}, t) = \sum_{\mathbf{k}} \phi_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}}$ the equation of motion becomes

$$\dot{u}_{\mathbf{k}} = -2Hu_{\mathbf{k}} - ik\phi_{\mathbf{k}}/a^2 \tag{5}$$

where $\phi_{\mathbf{k}}$ satisfies the Fourier transform of Poisson's equation

$$|\mathbf{k}|^2\phi_{\mathbf{k}} = -4\pi G\overline{\rho}a^2\delta_{\mathbf{k}}. \tag{6}$$

The final step is to use the equation of continuity. This is $\partial_t\rho = -\boldsymbol{\nabla}_{\mathbf{r}}\cdot(\rho\mathbf{v}_{\text{phys}})$ or, in comoving coordinates $\partial_t(1+\delta) = -\boldsymbol{\nabla}_{\mathbf{x}}\cdot((1+\delta)\mathbf{u})$ and whose linearised version is

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

which again, being linear, applies to each Fourier model individually:

$$\dot{\delta}_{\mathbf{k}} = -i|\mathbf{k}|u_{\mathbf{k}} \tag{8}$$

using this and Poisson's equation to eliminate $u_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$ gives the single second order (in time) equation for the mode amplitude $\delta_{\mathbf{k}}$:

$$\boxed{\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\overline{\rho}\delta_{\mathbf{k}} = 0} \tag{9}$$

The subscript $\mathbf{k}$ here is somewhat redundant as none of the coefficients depends on $\mathbf{k}$. This is because we have not yet included pressure gradient forces: we have just obtained the equations for pressure free perturbations, but in a slightly more formal way. As a sanity check, if we give, as an ansätz $\delta_{\mathbf{k}} \propto t^\alpha$ this becomes the algebraic equation

$$3\alpha^2 + \alpha - 2 = 0 \tag{10}$$

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

All of this is valid in the matter dominated era. The generalisation to zero-pressure fluctuations at late times (in the presence of dark energy) will be considered later. 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.

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 (if we assume that the 'entropy per baryon' is constant – so called 'adiabatic' or 'isentropic' perturbations.

We will now analyse this, first in the matter dominated regime (in the relatively narrow period after $t_{\rm eq}$ but before recombination at $t_{\rm rec}$) and for wavelengths with are sub-horizon scale, so they can be treated using Newtonian mechanics.

1.2.1 The sound speed and the Jeans' length

We will assume, for now, that the wavelength is sufficiently large that photons cannot 'leak out' of the perturbations, so the fractional enhancement in the number of photons is the same as that of the baryons $\delta n_\gamma/n_\gamma = \delta\rho_{\rm b}/\rho_{\rm b}$. The pressure of the the radiation goes like the 4/3 power of the density of photons, so we have, for the radiation density,

$$\rho_{\rm r} = \frac{\overline{\rho}_{\rm r}}{\overline{\rho}_{\rm b}^{4/3}}\rho_{\rm b}^{4/3} \tag{11}$$

so

$$d\rho_{\rm r} = \frac{4}{3}\frac{\overline{\rho}_{\rm r}}{\overline{\rho}_{\rm b}}d\rho_{\rm b} \tag{12}$$

and therefore, since the radiation pressure is $P = \rho_{\rm r}c^2/3$, the sound speed is

$$c_s^2 = \frac{dP}{d\rho} = \frac{c^2}{3}\frac{d\rho_{\rm r}}{d\rho_{\rm b}} = \frac{4c^2}{9}\frac{\overline{\rho}_{\rm r}}{\overline{\rho}_{\rm b}} \tag{13}$$

so $c_s = (2/3)c$ at $t_{\rm eq}$ and decreases thereafter like $\sqrt{1+z}$.

The *Jeans length* is defined to be the distance that a sound wave can propagate in one expansion time: $\lambda_{\rm J} \sim c_s t$. It is often called the *'sound horizon'*. It is roughly equal to the (light) horizon size at (or before) $t_{\rm eq}$ but, since $t \sim 1/\sqrt{G\rho} \propto (1+z)^{-3/2}$, and $c_s \sim \sqrt{\overline{\rho}_{\rm r}/\overline{\rho}_{\rm b}} \propto (1+z)^{1/2}$ we have $\lambda_{\rm J} \propto (1+z)^{-1}$. That means that the *comoving Jeans length* (or comoving sound horizon) $\lambda_{\rm J}/a$ is independent of time between $t_{\rm eq}$ and $t_{\rm rec}$.

1.2.2 The equation of motion

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

$$\boxed{\ddot{\delta}_{\rm k} + 2H\dot{\delta}_{\rm k} - (4\pi G\overline{\rho} - c_s^2k^2/a^2)\delta_{\rm k} = 0} \tag{14}$$

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

$$k_{\rm J} \equiv \sqrt{4\pi G\overline{\rho}a^2/c_s^2} \tag{15}$$

the sign of the coefficient of $\delta_{\rm k}$ becomes positive and the result is that we get perturbations that oscillate with time.

They are however damped oscillations. A useful trick with such equations is to make a change of variables $\delta_{\rm k} = \Delta_{\rm k}t^\alpha$. In terms of $\Delta_{\rm k}$ the equation of motion becomes

$$\ddot{\Delta}_{\rm k} + 2\left(\frac{\alpha}{t} + H\right)\dot{\Delta}_{\rm k} + \left(\frac{c_s^2k^2}{a^2} - \frac{3}{2}H^2 + \frac{2H\alpha^2(\alpha-1)}{t^3}\right)\Delta_{\rm k} = 0 \tag{16}$$

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

$$\boxed{\ddot{\Delta}_{\mathbf{k}} + \left(\frac{c_s^2 k^2}{a^2} - H^2\right)\Delta_{\mathbf{k}} = 0} \tag{17}$$

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.

1.2.3 Adiabatic damping of sound waves

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 $\Omega_{\mathbf{k}} \simeq c_s k/a \propto a^{-3/2}$.

We can therefore apply the *principle of adiabatic invariance* which says that an oscillator with time varying frequency has energy $E \propto \Omega_{\mathbf{k}}$. Here the energy is proportional to the peculiar velocity squared: $E \propto v^2$ but $v \sim c_s \delta$, so $\delta \sim v/c_s \propto \Omega_{\mathbf{k}}^{1/2}/c_s \propto a^{-1/4}$.

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

$$\delta_{\mathbf{k}} \propto (1+z)^{1/4} \tag{18}$$

1.2.4 Silk damping

As the universal density decreases, the mean free path for photons increases. For sufficiently short wavelengths, the photons can diffuse a wavelength – this being most effective just before t_{rec}, at which time the photon mean free path rapidly increases and pressure switches off – and this can damp out the sound waves. This was first analysed by Joe Silk.

1.2.5 Sound waves in the radiation dominated era

Computing the evolution of sound waves prior to t_{eq} is more complicated. What is done is to use the continuity equation for the energy and momentum

$$T^{\mu\nu}{}_{;\mu} = 0 \tag{19}$$

and linearise this.

This is quite involved. Particularly so when one wants to consider the transition as modes enter the horizon.

The adiabatic damping in the radiation era can be easily determined. With $P = \rho c^2/3$ the sound speed is simply $c_s = c/\sqrt{3}$. The energy is now related to the velocity by $E \propto (1+z)v^2$ (since the mass per comoving volume is growing like $(1+z)$) while, as before $v \sim c_s \delta$. The frequency is $\Omega_{\mathbf{k}} \sim k c_s/a \propto (1+z)$ so the result is that the principle of adiabatic invariance ($E \propto \Omega_{\mathbf{k}}$) implies that the waves oscillate with constant amplitude:

$$\delta_{\mathbf{k}} \propto (1+z)^0. \tag{20}$$

1.2.6 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

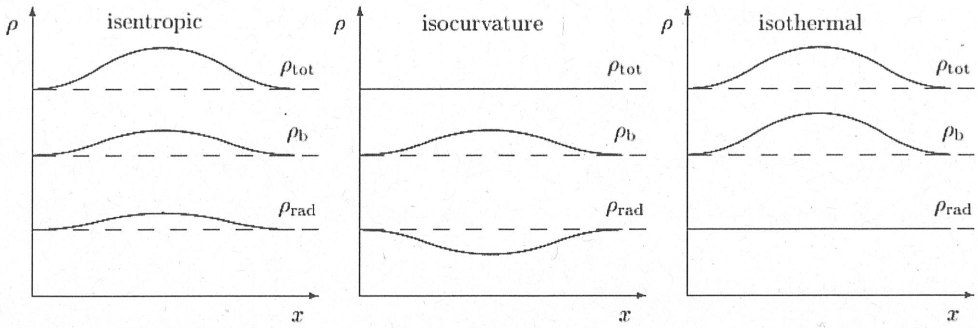

Figure 31.2: The left hand panel shows an 'adiabatic' or 'isentropic' perturbation of the kind we have been discussing above. In such a perturbation we crush the matter and radiation together. Such perturbations have a net density inhomogeneity and consequently have non-zero curvature or potential perturbations. 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.

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.

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

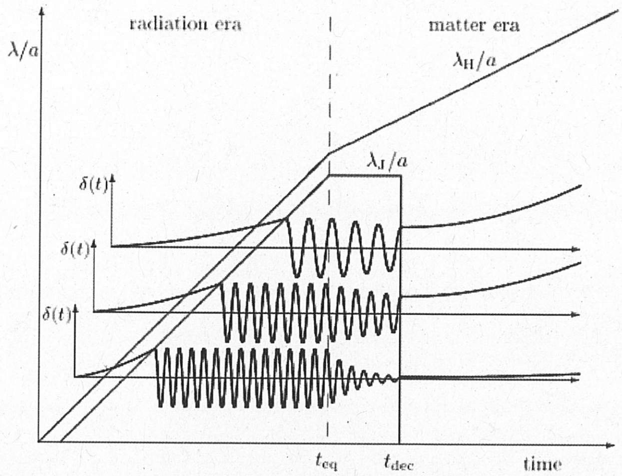

Figure 31.3: Evolution of initially adiabatic (or isentropic) perturbations is shown schematically for perturbations of three different wavelengths. The perturbation passes through three phases. First, when outside the horizon, the perturbation amplitude grows as $\delta \propto a^2$ in the radiation 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 (squared) 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 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_{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.

The evolution of perturbations in the HDM model (worked out by Bond and Efstathiou and others) is very different: perturbations entering the horizon when the neutrinos are still relativistic are washed out

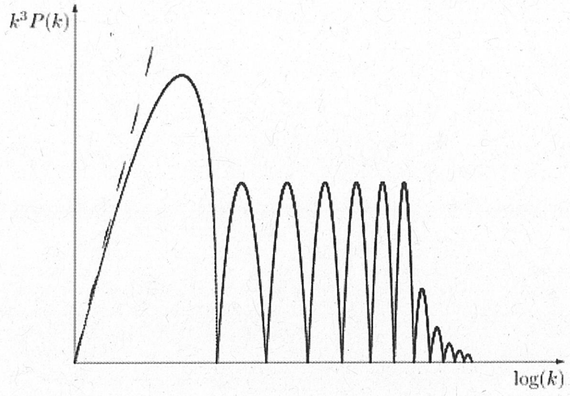

Figure 31.4: Power spectrum in the adiabatic-baryonic model (schematic). The dashed line indicates the initial power spectrum. The main peak is at a scale just larger than the maximum Jeans length, where the perturbations underwent continuous growth. Shorter waves entered the horizon before z_{eq} and 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

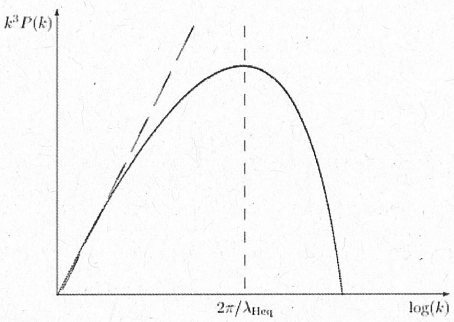

Figure 31.5: Power spectrum in the hot dark matter (HDM) model (schematic). The long-dashed line indicates the initial power spectrum. The vertical dashed line indicates the horizon size at the time the neutrinos become non-relativistic at $z \simeq z_{eq}$. In the HDM model the first structures to form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

3D numerical model of the Universe 903

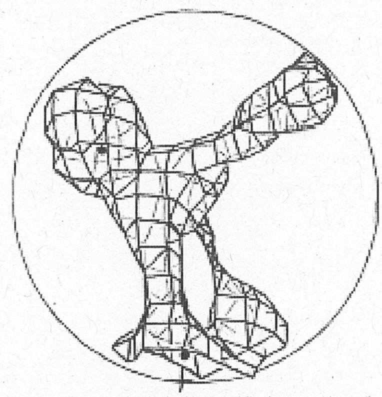

Figure 4. A surface of constant density level is plotted for the same region as that in Fig. 1.

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

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. Once the CDM comes to dominate over the radiation density, growth recommences. The resulting 'transfer function' is sketched in figure 6.

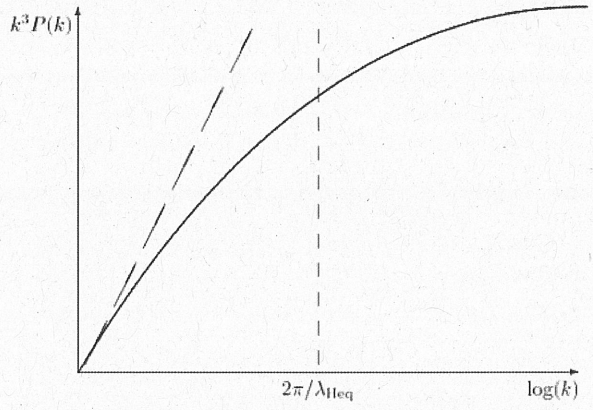

Figure 31.6: Power spectrum in the cold dark matter (CDM) model (schematic). The long-dashed line indicates the initial power spectrum. The vertical dashed line indicates the horizon size at the time the neutrinos become non-relativistic at $z \simeq z$eq.

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.

But the transition is rather gradual. As was already appreciated at the time – particularly by Rich 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 Bond and 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.

3 Seeds for structure formation

In the above we have implicitly assumed – following the compelling arguments of Harrison and of Zel'dovich – that if the primordial fluctuations originate in the extremely early universe they must have a 'scale-invariant' initial spectrum $n = 1$, for which the fluctuations in the gravitational potential are the same at different scales.

In addition, it is most commonly assumed that the initial 'seeds' for structure formation were *'gaussian'* random fields – as can be generated by summing Fourier modes with random phases – and this gives a complete description of the initial seeds. But, as we shall see, there are alternatives.

Below we will discuss three possibilities: The first is that the structure emerges 'spontaneously' by some astrophysical mechanism that rearranges matter from an initially uniform state. The second is a class of models, all involving scalar fields, in which structure forms via the *'Kibble mechanism'* and which includes *'domain walls'* and *'cosmic strings'*, both of which are examples of *'topological defects'*. The last is what is at the basis of the currently 'standard – or concordance – model' for structure formation in which the seeds were initially zero-point fluctuations of the modes of the 'inflaton field' driving inflation.

3.1 Spontaneous structure formation

The question here is what sort of large-scale structure might emerge if the universe was initially homogeneous, but developed small scale structure. One possibility might be the formation of stars that could explode and disturb the matter density on some scale ΔL.

Naively, one might imagine that there would be 'root-N' fluctuations with RMS density fluctuations on scale L given by $(\delta\rho/\rho)_L \sim 1/\sqrt{N}$ with $N \sim (L/\Delta L)^3$ the number of fluctuation regions.

This might look quite benign as $(\delta\rho/\rho)L \to 0$ for large L. But if we look at the potential fluctuations we get a very different story. These are $\delta\phi \sim G\delta M/L \sim G(\delta M/M) \times M/L$. But with $\delta M/M \sim \delta\rho/\rho \propto N^{-1/2} \propto M^{-1/2}$ and $L \propto M^{1/3}$ this gives $\delta\phi \propto M^{1/6}$, so these are actually growing with scale.

While this might look quite promising it is flawed. Subsequently, it was speculated that the leading order perturbations on large scales would be a 'surface effect'; with $\delta M \sim M_1\sqrt{N}$ with $N \sim (L/\Delta L)^2$, the number of fluctuation volumes within a distance $\sim \Delta L$ of the surface and $M_1 \sim \rho(\Delta L)^3$ the mass of one fluctuation volume. This would suggest $\delta M/M \propto L^{-2} \propto M^{-2/3}$ with $\delta\phi \propto L^0$, so scale invariant potential fluctuations.

Another line of argument was that while local spontaneous effects cannot generate a fluctuation in the net *mass* in a region – because mass is conserved, so a single fluctuation region has vanishing 'monopole' – it can *redistribute* the mass and thus, perhaps, generate a 'dipole' source for the large-scale gravitational field. A dipole produces a gravitational potential that falls off as $1/L^2$ on large scales (rather than as $1/L$ for a monopole source) and that results in $\delta M/M \propto M^{-5/6}$ and therefore $\delta\phi \propto M^{-1/6}$. So potential fluctuations weaker on large scales than scale-invariant, but only weakly so. But that doesn't work either; that is because *momentum* is conserved. In reality, the best one can do with local spontaneous effects is generate quadrupole sources – with $\delta\phi \propto 1/L^3$ for an individual source – and a random collection of these produces $\delta M/M \propto M^{-7/6}$ with $\delta\phi \propto M^{-1/2}$ with very weak effects on large scales. This is illustrated in figure 7.

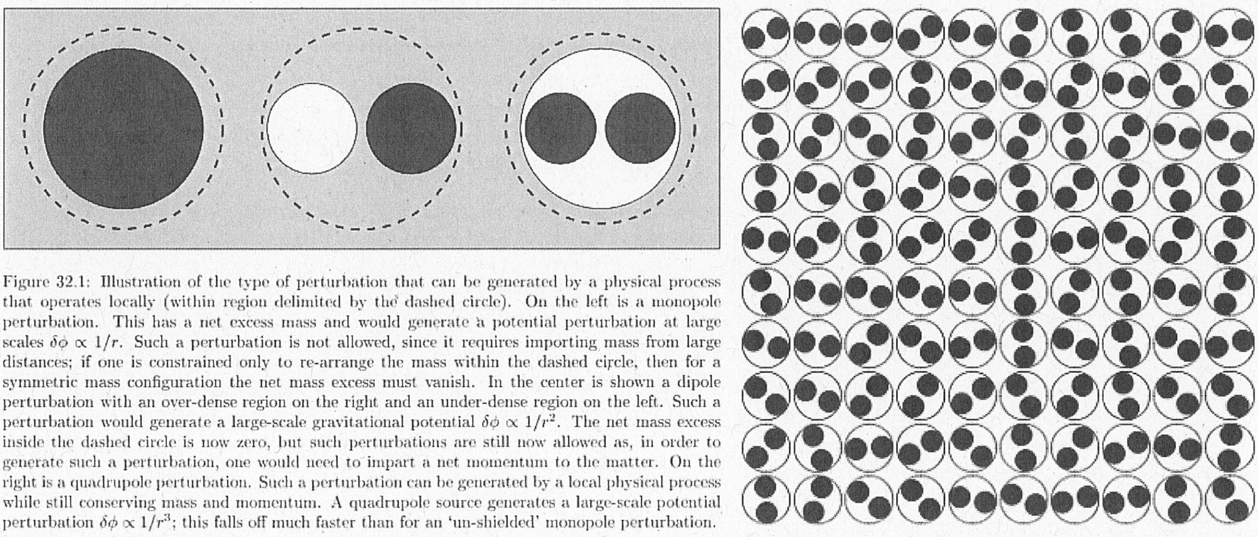

Figure 7: A locally generated disturbance of the matter cannot generate a 'monopole' density perturbation as mass is conserved. Nor can it generate a dipole as momentum is conserved. A random collection of quadrupole sources generates large scale perturbations with a $n = 4$ power spectrum ($P(k) \propto k^4$) with $\delta\phi \propto M^{-1/2}$. This is not a viable way to generate the observed large-scale structure.

3.2 Structure from topological defects

3.3 Primordial fluctuogenesis in inflation

As described in the previous lecture, the flatness and horizon problems can be solved if, prior to the 'big-bang' (here understood as the start of the radiation dominated universe), the universe was dominated by a nearly spatially homogeneous scalar field.

This was speculated to have been associated with 'grand unification' – thought to occur at an energy scale $E_{\rm GUT} \sim 10^{16}$GeV; a few orders of magnitude less than the Planck scale $E_{\rm P} \sim 10^{19}$GeV.

Many possibilities have been suggested. One is to have a field with Lagrangian density $\mathcal{L} = -\phi_{,\mu}\phi^{,\mu} - V(\phi)$ with potential $V(\phi) = \lambda\phi^4/\hbar c$. Provides $\phi \gg \sqrt{c^4/G}$, the potential term dominates in $\rho c^2 = \frac{1}{2}\dot{\phi}^2 + V$ and $P = \frac{1}{2}\dot{\phi}^2 - V$, giving $P \simeq -\rho$ with exponentially expanding solutions of the Friedmann equation $a \propto \exp(Ht)$.

The equation of motion (obtained by extremising $S = \int dt \int d^3x a^3(t)\mathcal{L}$ is

$$\ddot{\phi} + 3H\dot{\phi} + \frac{c^2}{a^2}\nabla^2\phi + 4\lambda\frac{c}{\hbar}\phi^3 = 0 \tag{21}$$

general idea in so-called 'chaotic' inflation is that the spatial derivative terms are negligible and the field is slowly 'rolling' down the potential at the 'terminal velocity' $\dot{\phi} = (4\lambda c/3H)\phi^3$. Ultimately the field reaches the bottom of the potential and would then, left to itself, oscillate about the minimum. But with coupling to other fields, the idea is the universe will 'reheat' to generate a hot thermal state with energy density $\rho \sim V/c^2$. If this is at the GUT scale, the horizon size is tiny compared to the present value: The energy density is $\rho \sim E^4$ with E being about 30 orders of magnitude greater than the current energy scale, so, with $t \sim 1/\sqrt{G\rho} \sim 10^{-60}t_0$ and the comoving horizon scale ct/a is about 30 orders of magnitude less than its current value. For this to work then requires roughly 70 e-foldings of expansion in order to drive the comoving horizon scale at reheating down to this value.

The remarkable thing that was worked out around 1980 is that, in this scenario, the fluctuations in the field cannot be arbitrarily small; there must be, at a minimum, zero point fluctuations in the Fourier modes, and these were shown to give density fluctuations at horizon crossing that are scale-invariant – this is a given – with amplitude at horizon crossing

$$\delta\rho/\rho \sim \sqrt{\frac{\hbar}{c}\frac{H^2}{\dot{\phi}}} \tag{22}$$

which, for suitable choice of the coupling constant λ, would be of the right size to explain the observed large-scale structure and CMB temperature anisotropies.

The way this is calculated is to decompose the field as $\phi = \phi_0 + \phi_1$, where ϕ_0 is the spatially uniform 'background' field that drives the accelerated expansion and the perturbation ϕ_1 describes the Fourier modes with non-vanishing $\mathbf{k}$.

The equation of motion for ϕ_1 is

$$\ddot{\phi}_1 + 3H\dot{\phi}_1 + \frac{c^2}{a^2}\nabla^2\phi_1 + 12\lambda\frac{\phi_0^2 c}{\hbar}\phi_1 = 0 \tag{23}$$

which is like that of a free massive field, and in fact, for the parameters that work to give inflation, the mass is effectively zero so one has, to a good approximation, a massless field.

When the modes are sub-horizon scale, the modes have zero point energy $E_{\mathbf{k}} = \frac{1}{2}\hbar\omega \sim \omega^2\phi_{\mathbf{k}}^2$ and evolve adiabatically. But as the modes become comparable in wavelength to the horizon scale, the adiabaticity breaks down and the mean occupation becomes non-zero (though still of order unity). This gives

$$\langle\phi_1^2\rangle \sim \hbar H^2/c. \tag{24}$$

What happens subsequently is that a region where the field had a positive ϕ_1 will inflate for slightly longer than one with a negative ϕ_1. The result is, after reheating, that there is more matter in that region (by an amount $\phi_1/H\dot{\phi}_0$).

How does the universe accommodate this excess of matter? After all, it reheats to the same temperature and density everywhere. It does so by 'herniating' slightly; i.e. it generates curvature fluctuations.

4 Non-linear structure via N-body simulations

4.1 The 'background-plus-perturbation' picture

One way to derive the equations for evolution of structure into the non-linear regime is to assume that there is a 'background' FLRW cosmology with metric $ds^2 = -dt^2 + a(t)^2dx^2$, whose scale factor $a(\tau)$ obeys the Friedmann equations

$$\ddot{a} + \frac{4\pi}{3}G\bar{\rho}a = 0 \tag{25}$$

where dot denotes time derivative, and where for a homogeneous background universe the mean density decreases as $\overline{\rho} \propto a^{-3}$. The density here may be augmented by additional terms for homogeneous dark energy or radiation backgrounds satisfying the appropriate continuity equations.

Onto this background are then added density perturbations, the potential of which appears as a perturbation to the metric $\delta g_{\alpha\beta} = -2\phi/c^2 \delta_{\alpha\beta}$ where

$$\nabla^2 \phi = 4\pi G(\rho - \overline{\rho})a^2 \tag{26}$$

which is the Newtonian Poisson's equation, but with only the density perturbation as the source.

For this to be valid does not require that the density perturbations be small; only that the velocities associated with structures are small compared to c.

The peculiar (i.e. non-Hubble) motions of non-relativistic particles such as dark matter or galaxies obey the 'structure evolution' equations

$$\dot{\mathbf{v}} + H\mathbf{v} = -\boldsymbol{\nabla}\phi/a \tag{27}$$

where $\mathbf{v} \equiv a\dot{\mathbf{x}}$, $H = \dot{a}/a$ and the spatial derivative is with respect to comoving coordinates $\mathbf{x}$.

This again looks very similar to the Newtonian equation of motion, but with the addition of a 'Hubble-dampling' term which acts like a frictional force.

This is the system of equations that is used in N-body simulations to obtain evolve structure into the non-linear regime. Particles are moved according to (27) and the potential is calculated either by binning the particle density onto a grid in comoving coordinates and solving (26) or by direct force summation (or a hybrid) while the scale factor of the background is evolved with (25).

Note that (25) for $a(\tau)$ is decoupled from the other equations; the scale factor appears in the equations of motion for the particles, but there is no 'back-reaction' of the emergence of structure on the dynamics 'background' expansion. This has sometimes been called into question.

As we will now show, there is an alternative way of deriving equations (27) and (26) that sheds some light on this. In this picture, the scale factor has no dynamical significance and is seen to be merely a 'book-keeping' device that suppresses the main part of the forces on the particles.

4.2 The Dmitriev & Zel'dovich equations

For N particles of mass m interacting under their mutual gravitational attraction there are $3N$ second order differential equations

$$\ddot{\mathbf{r}}_i = Gm \sum_{j \neq i} \frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3}. \tag{28}$$

These may be solved numerically provided initial positions $\mathbf{r}_i$ and velocities $\dot{\mathbf{r}}_i$ for the particles.

Writing this in terms of arbitrarily re-scaled coordinates $\mathbf{r} = a(t)\mathbf{x}$, so $\dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}}$ and $\ddot{\mathbf{r}} = \ddot{a}\mathbf{x} + 2\dot{a}\dot{\mathbf{x}} + a\ddot{\mathbf{x}}$, (28) becomes

$$\ddot{\mathbf{x}}_i + 2\frac{\dot{a}}{a}\dot{\mathbf{x}}_i = \frac{Gm}{a^3} \sum_{j \neq i} \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|^3} - \frac{\ddot{a}}{a}\mathbf{x}_i \tag{29}$$

where we have, somewhat arbitrarily, moved one of the terms in $\ddot{\mathbf{r}}$ over to the right hand side. On the left, the situation is similar to that in (27) where these is an apparent acceleration and a Hubble damping term. This looks a little different in that here is a factor 2. The reason for this is that the peculiar velocity is $\mathbf{v} \equiv a\dot{\mathbf{x}}$, so $\dot{\mathbf{v}} = a\ddot{\mathbf{x}} + \dot{a}\dot{\mathbf{x}}$ so the left hand side of (27) is $\dot{\mathbf{v}} + H\mathbf{v} = \dot{\mathbf{v}} + \dot{a}\dot{\mathbf{x}} = a \times (\ddot{\mathbf{x}} + 2H\dot{\mathbf{x}})$, so, aside from the factor a, the two left-hand sides are identical.

What we are interested in is the motion of particles with initial conditions that are close to being in uniform Hubble expansion with some initial expansion rate H (very close if we start at early times). So we might lay down particles on a regular grid in $\mathbf{r}$-space within some very large spherical boundary centred on the origin and give the particles small displacements $\delta\mathbf{r}$ and velocities $\dot{\mathbf{r}} = H\mathbf{r} + \delta\dot{\mathbf{r}}$ with 'peculiar' velocities $\delta\dot{\mathbf{r}}$ chosen to excite the growing mode. This is illustrated in figure 8. The corresponding initial conditions in terms of $\mathbf{x}$-coordinates are

$$\mathbf{x} = \mathbf{r}/a \quad \text{and} \quad \dot{\mathbf{x}} = ((H - \dot{a}/a)\mathbf{r} + \delta\dot{\mathbf{r}})/a. \tag{30}$$

The sum in (29) will have two components: A 'zeroth order' acceleration that, in the limit that the grid spacing becomes very small, is the same as the gravitational acceleration of a uniform density sphere, which