3. Structure Formation

Until now, we have discussed a universe which is perfectly homogeneous and isotropic. But that is not the universe we live in. Instead, our universe contains interesting objects which clump together, bound by the gravitational force, from planets and stars, to galaxies and clusters of galaxies. We would like to understand how these objects form.

The stakes become somewhat higher when we realise that the early universe was very much smoother than the one we live in today. Of course, there were no galaxies and planets, or even atoms, in the early fireball. But nor were there significant variations in the energy density. This can be clearly seen in the CMB, which has an almost uniform temperature T but exhibits tiny fluctuations on the scale

$$\frac{\delta T}{T} \approx 10^{-5}$$

We can compare this to the world we see around us today. As we learned in Section 1.4, the average energy density in the universe $\rho_{\rm crit,0}$, corresponds to about 1 hydrogen atom per cubic metre. But this hides the fact that most of this matter is contained in gravitationally bound objects. A measure analogous to $\delta T/T$ can be found by comparing the typical energy density contained in a galaxy to the average $\rho_{\rm crit,0}$: this turns out to be

$$\frac{\rho_{\rm galaxy}}{\rho_{\rm crit}} \approx 10^6$$

We see that the universe, like many of us, has become significantly more lumpy as it aged. The primary purpose of this section is to understand how this occurred: how did the small fluctuations seen in the CMB grow, ultimately resulting in the wondrous array clusters and galaxies that we see in the night sky. This process is known as *structure formation*.

There is also a second, more ambitious, purpose to this section, which is to trace the perturbations backwards in time. Ultimately, we would like to understand where the small fluctuations $\delta T/T$ seen in the CMB came from. Since these fluctuations grow to give rise to all the structure in the universe, this is really a rephrasing of one of the biggest questions of them all: where did we come from? We will see that when we evolve the fluctuations backwards in time, they take a very simple form, providing crucial information about what the universe looked like in its very earliest moments. Ultimately, in Section 3.5, we will offer an answer to this big question. We will argue that all the structure in the universe, including us, owes its existence to fluctuations of quantum fields, fluctuations that took place in the first few fractions of a second after the Big Bang.

3.1 Density Perturbations

In this section, we will assume that there are some small perturbations in the energy density of the universe. We will not (yet) ask where these perturbations came from. Instead, we will be interested in their fate. In particular, under what circumstances do they grow, and when do they fade away?

We start by considering non-relativistic matter. As we have seen, in our universe this is primarily dark matter. We know very little about the interactions of dark matter, but there is a wonderful universality in physics which tells us that, on suitably large distances, any substance can be described by the equations of fluid mechanics. This, then, will be our tool of choice: fluid mechanics applied on cosmological scales.

Our goal is to start with a homogeneous and isotropic fluid, and then see what happens when it is perturbed. First we need to specify our variables which, in contrast to earlier sections, now depend on both space and time. The standard variables of fluid mechanics are

- number density $n(\mathbf{x}, t)$. More precisely, we will be interested in the mass density. For now, we will consider a fluid made of a single type of particle of mass m, so the mass density is simply $mn(\mathbf{x}, t)$. For a non-relativistic fluid, the mass dominates the energy density which is given by $\rho(\mathbf{x}, t) = mn(\mathbf{x}, t)c^2$.
- Pressure $P(\mathbf{x}, t)$. As discussed in Section 1.2.1, non-relativistic fluids have $P \ll \rho$.
- Velocity $\mathbf{u}(\mathbf{x}, t)$.

Next, we need the relevant equations of fluid mechanics. These depend on the context. Ultimately, we want to understand fluids which gravitate in an expanding universe. However, we're going to build up slowly and introduce one ingredient at a time.

3.1.1 Sound Waves

First, we're going to consider fluids that don't experience gravity and live in a static spacetime. These fluids are described by three equations. The first is the *continuity equation* which captures the conservation of particles

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{u}) = 0 \tag{3.1}$$

This tells us that the particle density in some region can only change if it flows away, with the change due to its velocity $\mathbf{u}$.

The second is the *Euler equation*, which can be viewed as Newton's "F=ma" for a continuous system,

$$mn\left(\frac{\partial}{\partial t}+\mathbf{u}\cdot\nabla\right)\mathbf{u}=-\nabla P \tag{3.2}$$

The left-hand side is interpreted as mass $\times$ acceleration, while the pressure $-\nabla P$ on the right-hand side provides the force.

The last of our equations is the equation of state which, for now, we leave general as

$$P=P(n,T) \tag{3.3}$$

For much of this section, we will use ideal gas equation, $P=nk_BT$, which is the appropriate equation of state for a non-relativistic fluid. In time, we will also apply these ideas to other fluids.

The simplest solution to these equations describes a static fluid with $\mathbf{u}=0$ and constant density and pressure

$$n=\bar{n}\quad\text{and}\quad P=\bar{P}$$

This is a homogeneous and isotropic fluid. We take this to be our background and look at small perturbations. We take $\mathbf{u}$ to be small, and write

$$n(\mathbf{x},t)=\bar{n}+\delta n(\mathbf{x},t)\quad\text{and}\quad P(\mathbf{x},t)=\bar{P}+\delta P(\mathbf{x},t)$$

The equations (3.1) and (3.2) are linearised to give

$$\frac{\partial(\delta n)}{\partial t}=-\nabla\cdot(\bar{n}\mathbf{u})\quad\text{and}\quad m\bar{n}\frac{\partial\mathbf{u}}{\partial t}=-\nabla\delta P$$

We can combine these to find,

$$\frac{\partial^2(\delta n)}{\partial t^2}=-\bar{n}\nabla\cdot\frac{\partial\mathbf{u}}{\partial t}=\frac{1}{m}\nabla^2\delta P$$

At this point, we need to invoke the equation of state, relating P to n. It will be useful to give a new name to the quantity $\partial P/\partial n$: we write it as

$$\frac{\partial P}{\partial n}=mc_s^2 \tag{3.4}$$

We can then relate $\delta P=mc_s^2\delta n$ to find that the density perturbations obey

$$\left(\frac{\partial^2}{\partial t^2}-c_s^2\nabla^2\right)\delta n=0 \tag{3.5}$$

This is the *wave equation*. As its name suggests, its solutions are waves of the form

$$\delta n(\mathbf{x}, t) = A(\mathbf{k}) \cos(\omega t - \mathbf{k} \cdot \mathbf{x}) + B(\mathbf{k}) \sin(\omega t - \mathbf{k} \cdot \mathbf{x}) \tag{3.6}$$

We call these *sound waves*. The key property of the solution is the *wavevector* $\mathbf{k}$ which determines the direction of travel of the wave and the wavelength $\lambda = 2\pi/|\mathbf{k}|$. The frequency ω of the wave is given by

$$\omega = c_s |\mathbf{k}|$$

The proportionality constant c_s, defined in (3.4), has the interpretation of the speed of sound. (We'll compute examples below.) Finally, $A(\mathbf{k})$ and $B(\mathbf{k})$ are two, arbitrary integration constants. Because the wave equation is linear, we can add together as many solutions of the form (3.6) as we like, with different integration constants $A(\mathbf{k})$ and $B(\mathbf{k})$. In this way, we can build up wavepackets with different profiles.

In what follows, we will often write the solution (3.6) in complex form,

$$\delta n = C(\mathbf{k}) \exp(i(\omega t - \mathbf{k} \cdot \mathbf{x}))$$

for some complex C. This is standard, albeit inaccurate notation. Obviously the number density δn should be real. But because the wave equation is linear, we can always just take the real part of the right-hand-side to get a solution. This form of the solution is more useful simply because it's quicker to write exponentials rather than cos and sin.

Speed of Sound of a Non-Relativistic Fluid

Throughout Section 1, we treated the equation of state of a non-relativistic fluid as $P = 0$. What this really means is that $P \ll \rho$, where ρ is the energy density, mostly due to the rest mass of the fluid.

The equation for the sound speed (3.4) can alternatively be written in terms of the energy density $\rho = mnc^2$, as

$$\frac{\partial P}{\partial \rho} = \frac{c_s^2}{c^2} \tag{3.7}$$

Using $P = 0$ suggests that $c_s = 0$ for a non-relativistic fluid. But what this is really telling is simply that

$$c_s \ll c \tag{3.8}$$

This makes sense. The sound speed is related to the speed of the constituent particles in the fluid. In a non-relativistic fluid, this is necessarily much less than the speed of light.

In fact, we can do better and compute the sound speed as a function of temperature (which, itself, is related to the speed of the constituent particles). For the ideal gas, the equation of state is

$$P = nk_BT$$

It's tempting to simply differentiate $\partial P/\partial n$, with T fixed, to determine the speed of sound. But that's a little too hasty: as P and n vary, it is quite possibly that T varies as well.

To understand how this works, we need a little physical input. The energy of the ideal gas with some fixed total number of atoms $N = nV$ is (2.8)

$$E = \frac{3}{2}Nk_BT$$

If the volume changes, then the energy should change by the work done

$$\begin{aligned} dE = -P\,dV \quad &\Rightarrow \quad \frac{3}{2}Nk_B\,dT = -P\,dV \\ &\Rightarrow \quad \frac{3}{2}\frac{dT}{T} = -\frac{dV}{V} \end{aligned}$$

where, in the second line, we've used the equation of state. Integrating this expression tells us that $T^{3/2}V$ is constant. Alternatively, using $n = N/V$, we learn that $Tn^{-2/3}$ is constant. Such changes are referred to as *adiabatic*. (Underlying this is the statement that entropy is conserved for adiabatic changes; you can learn more about this in the lectures on Statistical Physics.) This means that

$$\left.\frac{\partial T}{\partial n}\right|_{\text{adiabatic}} = \frac{2}{3}\frac{T}{n}$$

The speed of sound (3.4) should be computed under the assumption of such an adiabatic change. We then have

$$\left.\frac{\partial P}{\partial n}\right|_{\text{adiabatic}} = \frac{5}{3}k_BT$$

From this, we compute the speed of sound in an ideal gas to be

$$c_s = \sqrt{\frac{5k_BT}{3m}}$$

Before we proceed, I will briefly mention another, more rigorous, approach to get this result. We could treat $T(\mathbf{x}, t)$ as a new dynamical field with its own equation of motion. This equation of motion turns out to be

$$\left(\frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla\right) T + \frac{2T}{3}\nabla\cdot\mathbf{u} = 0$$

A full derivation of this needs the Boltzmann equation, and can be found in the lectures on Kinetic Theory. It's straightforward to check that this equation combines with the continuity equation (3.1) to ensure that $Tn^{-2/3}$ is indeed constant along flow lines.

Sound Speed of a Relativistic Fluid

So far, our discussion of fluid dynamics was focussed on non-relativistic fluids. However, it should come as no surprise to learn that we will also be interested in relativistic fluids in the context of cosmology.

Much of the discussion above goes through for general fluids if the equations are phrased in terms of the energy density ρ instead of the number density n. In particular, the sound speed can be computed using (3.7). For a relativistic fluid, with $P = \rho/3$, we have

$$c_s^2 = \frac{1}{3}c^2 \tag{3.9}$$

This time we see that the speed of sound is tied to the speed of light. Again, this is to be expected: in a relativistic fluid, any constituent particles are flying around at close to the speed of light. The difference in the speed of sound between a non-relativistic fluid (3.8) and a relativistic fluid (3.9) will prove to be one of the important ingredients in the story of structure formation.

3.1.2 Jeans Instability

Our next step is to add in the effects of gravity. For now we will keep ourselves in a static spacetime. The continuity equation (3.1) remains unchanged. However, the Euler equation (3.2) picks up an extra term on the right-hand-side due to the gravitational field Φ experienced by the fluid,

$$mn\left(\frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla\right)\mathbf{u} = -\nabla P - mn\nabla\Phi \tag{3.10}$$

This gravitational field is determined by the matter in the fluid in the usual manner

$$\nabla^2\Phi = 4\pi G m n \tag{3.11}$$

We want to consider a homogeneous solution as before, with constant $n = \bar{n}$ and $P = \bar{P}$ and $\nabla\Phi = 0$. There's a small problem with this as a starting point: it doesn't obey the Poisson equation (3.11)! This is a famously fiddly aspect of the following derivation, one that stems from the fact that there is no infinite, static self-gravitating fluid. For now, we simply bury our head in the sand and ignore this issue, an approach which is sometimes known as the *Jeans' swindle*. But, for once, this approach will be rewarded: in the next section, we will consider perturbations in an expanding universe where this issue is resolved.

We now perturb the constant background. We require that the perturbed gravitational potential $\Phi + \delta\Phi$ obeys

$$\nabla^2\delta\Phi = 4\pi G m\,\delta n$$

The same linearisation that we saw previously now shows that the wave equation (3.5) is deformed to

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2\nabla^2 - 4\pi G m\bar{n}\right)\delta n = 0$$

This is again solved by the ansatz

$$\delta n = C(\mathbf{k})\,\exp\left(i(\omega t - \mathbf{k}\cdot\mathbf{x})\right)$$

but now with the frequency and wavevector related by

$$\begin{aligned}\omega^2 &= c_s^2k^2 - 4\pi G m\bar{n}\\ &= c_s^2(k^2 - k_J^2)\end{aligned}$$

Equations of this type, which relate the frequency to the wavenumber, are referred to as *dispersion relations*. In the second line we have defined the *Jeans wavenumber* k_J,

$$k_J = \sqrt{\frac{4\pi G m\bar{n}}{c_s^2}}$$

The qualitative properties of the solution now depend on the wavenumber $\mathbf{k}$. For small wavelengths, or large wavenumbers $k > k_J$, the solutions oscillate as before. These are sound waves. However, when the wavelengths are large, $k < k_J$ then the gravitational background becomes important. Here the frequency is imaginary, which has the interpretation that perturbations $\delta n \sim e^{i\omega t}$ grow or decay exponentially. For $k \ll k_J$ we have

$$\delta n \sim e^{\pm t/\tau} \quad \text{with} \quad \tau \approx \sqrt{\frac{1}{4\pi G m\bar{n}}} \tag{3.12}$$

We learn that long wavelength perturbations no longer oscillate as sound waves. Instead, any perturbation that has a size larger than the *Jeans length*,

$$\lambda_J = \frac{2\pi}{k_J} = c_s\sqrt{\frac{\pi}{G m\bar{n}}} \tag{3.13}$$

will typically grow exponentially quickly due to the effect of gravity. This is known as the *Jeans' instability*.

The derivation above also gives us a clue to the physical mechanism for the Jeans instability. It comes from attempting to balance the pressure and gravitational terms in the Euler equation (3.10). Consider an over-dense spherical region of radius R. In the absence of any pressure, this region would collapse with a time-scale τ given in (3.12). In a fluid, this collapse is opposed by the pressure. But the build-up of pressure is not instantaneous; it takes time given roughly by

$$t_{\rm pressure} \sim \frac{R}{c_s}$$

When R is small, $t_{\rm pressure} < \tau$ and the build-up of pressure stops the collapse and we get oscillating motion that we interpret as sound waves. In contrast, if R is large we have $\tau < t_{\rm pressure}$, there is no time for the pressure to build. In this case, the system suffers the Jeans instability and is susceptible to gravitational collapse.

3.1.3 Density Perturbations in an Expanding Space

Finally, we want to consider something more cosmological: the growth of density perturbations, interacting with gravity, in an expanding space with scale factor $a(t)$. Throughout we will work with flat space. (This means "$k = 0$" in the notation of Section 1. However, in this Section we use k to denote the wavenumber of perturbations.)

We need to revisit our equations once more. Consider a particle tracing out a trajectory $\mathbf{x}(t)$ in co-moving coordinates. The physical coordinates $\mathbf{r}(t)$ (called $\mathbf{x}_{\rm phys}$ in (1.14)) is given by

$$\mathbf{r}(t) = a(t)\mathbf{x}(t)$$

The physical velocity of a particle is

$$\mathbf{u} = \dot{\mathbf{r}} = H\mathbf{r} + \mathbf{v} \tag{3.14}$$

with $\mathbf{v} = a\dot{\mathbf{x}}$. In what follows, we will need to jump between physical and co-moving coordinates. The spatial derivatives are related simply by

$$\nabla_{\mathbf{r}} = \frac{1}{a}\nabla_{\mathbf{x}} \tag{3.15}$$

The temporal derivatives are a little more subtle since they differ depending on whether we keep $\mathbf{r}(t)$ fixed or $\mathbf{x}(t)$ fixed. In particular

$$\begin{aligned}\left.\frac{\partial}{\partial t}\right|_{\mathbf{r}} &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} + \left.\frac{\partial \mathbf{x}}{\partial t}\right|_{\mathbf{r}} \cdot \nabla_{\mathbf{x}} \\ &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} + \left.\frac{\partial (a^{-1}\mathbf{r})}{\partial t}\right|_{\mathbf{r}} \cdot \nabla_{\mathbf{x}} \\ &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} - H\mathbf{x} \cdot \nabla_{\mathbf{x}} \end{aligned} \tag{3.16}$$

Now we come to the equations describing the fluid. The equations that we dealt with previously should be viewed as given in terms of physical coordinates $\mathbf{r}$. However, it will turn out that subsequent calculations are somewhat easier if done in co-moving coordinates. We just have to translate from one to the other.

The Continuity Equation Revisited

The continuity equation (3.1) should be viewed in physical coordinates and so, in our new notation, reads

$$\left.\frac{\partial n}{\partial t}\right|_{\mathbf{r}} = -\nabla_{\mathbf{r}} \cdot (n\mathbf{u})$$

Changing to co-moving coordinates, it then becomes

$$\left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} - H\mathbf{x} \cdot \nabla_{\mathbf{x}} \right) n = -\frac{1}{a} \nabla_{\mathbf{x}} \cdot (n\mathbf{u})$$

In what follows, we drop the subscript $\mathbf{x}$ on everything; ∇ will always mean $\nabla_{\mathbf{x}}$ and $\frac{\partial}{\partial t}$ will always mean $\left.\frac{\partial}{\partial t}\right|_{\mathbf{x}}$.

We can make contact with the story of Section 1. Following (3.14), we write the velocity of the fluid as

$$\mathbf{u}(\mathbf{x}, t) = Ha\mathbf{x}(t) + \mathbf{v}(\mathbf{x}, t) \tag{3.17}$$

and the continuity equation becomes

$$\frac{\partial n}{\partial t} + 3Hn + \frac{1}{a} \nabla \cdot (n\mathbf{v}) = 0 \tag{3.18}$$

where we've used $\nabla \cdot \mathbf{x} = 3$. This form makes it clear that if we restrict to solutions in which $\mathbf{v} = 0$, so the velocity of the fluid simply follows the expansion of spacetime,

then we recover our earlier continuity equation (1.39), specialised to the case of non-relativistic matter,

$$\frac{\partial n}{\partial t} = -3Hn$$

(Recall that the energy density is given by $\rho = m\bar{n}c^2$.) This has the familiar solution

$$n(t) \sim \frac{1}{a^3} \tag{3.19}$$

which simply tells us that the number density dilutes as the universe expands.

Now we perturb the fluid,

$$\begin{aligned} n(\mathbf{x}, t) &= \bar{n}(t) + \delta n(\mathbf{x}, t) \\ &= \bar{n}(t)\Big[1 + \delta(\mathbf{x}, t)\Big] \end{aligned}$$

where $\bar{n}(t)$ is a spatially homogeneous density evolving as (3.19) and, in the second line, we've defined

$$\delta = \frac{\delta n}{\bar{n}} = \frac{\delta \rho}{\bar{\rho}}$$

The perturbation δ is referred to as the *density contrast*.

Let's now see what conditions the continuity equation (3.18) imposes on these perturbations. It reads

$$\begin{aligned} \frac{\partial}{\partial t}(\bar{n}\delta) + 3H\bar{n}\delta &= -\frac{1}{a}\nabla \cdot [\bar{n}(1+\delta)\mathbf{v}] \\ &= -\frac{\bar{n}}{a}\nabla \cdot \mathbf{v} + \mathcal{O}(\mathbf{v}\delta) \end{aligned}$$

We drop the second term on the grounds that is non-linear in the small quantities δ and $\mathbf{v}$. Using the fact that the background density $\bar{n}$ evolves as (3.19), this equation reduces to the simple requirement

$$\dot{\delta} = -\frac{1}{a}\nabla \cdot \mathbf{v} \tag{3.20}$$

This is the first of our perturbed equations.

The Euler and Poisson Equations Revisited

Next up we need to deal with the Euler equation (3.10) and Poisson equation (3.11). The Euler equation as written in (3.10) should again be viewed in physical coordinates,

$$mn\left(\left.\frac{\partial}{\partial t}\right|_{\mathbf{r}} + \mathbf{u}\cdot\nabla_{\mathbf{r}}\right)\mathbf{u} = -\nabla_{\mathbf{r}}P - mn\nabla_{\mathbf{r}}\Phi$$

After substituting in (3.17), (3.15) and (3.16), this becomes

$$mna\left(\frac{\partial}{\partial t} + \frac{\mathbf{v}}{a}\cdot\nabla\right)\mathbf{u} = -\nabla P - mn\nabla\Phi \tag{3.21}$$

where, as previously, the lack of any subscript on the derivatives means that they are taken holding $\mathbf{x}$ fixed. A similar, but simpler, story also holds for the Poisson equation. In physical coordinates, this is

$$\nabla^2_{\mathbf{r}}\Phi = 4\pi G mn$$

In co-moving coordinates, it becomes

$$\nabla^2\Phi = 4\pi G mna^2 \tag{3.22}$$

The background $\mathbf{u} = Ha\mathbf{x}$, with $\mathbf{v} = 0$, solves the Euler equation provided that we take $\nabla\bar{P} = 0$ and $\Phi = \bar{\Phi}$ such that

$$\nabla\bar{\Phi} = -\ddot{a}a\mathbf{x} \quad \Rightarrow \quad \nabla^2\bar{\Phi} = -3\ddot{a}a \tag{3.23}$$

This is now perfectly compatible with the Poisson equation; indeed, the two combine to give

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}m\bar{n}$$

But this is precisely the acceleration equation (1.52) that we met previously. Note that we didn't assume the Friedmann equation anywhere in this derivation. Nonetheless, we find the acceleration equation (which, recall, is the time derivative of the Friedmann equation) emerging as a consistency condition on our analysis! This isn't as miraculous as it may first appear. Our derivation of the Friedmann equation in Section 1.2.3 involved only Newtonian gravity, which is the same physics we have invoked here. However, in one particular sense, the current derivation using fluids is a considerable improvement on the derivation in Section 1.2.3, because we didn't have to make the misleading assumption that there is an origin of the universe from which all matter is expanding. Instead, the fluid treatment allows us to understand the expansion of a genuinely homogeneous and infinite universe.

For our immediate purposes, the thing that should make us most happy is that we no longer have to worry about the Jeans' swindle; our spatially homogeneous background satisfies the equations of motion as it should. At heart, the Jeans' swindle was telling us that a spatially homogeneous fluid is inconsistent in Newtonian gravity. But it is perfectly consistent if we allow for an expanding universe, with the gravitational potential $\bar{\Phi}$ for a homogeneous fluid driving the expansion of space, just as we learned in Section 1.

Next, we perturb around the background. We write $P = \bar{P} + \delta P$ and $\Phi = \bar{\Phi} + \delta\Phi$ and, as before, $\mathbf{u} = Ha\mathbf{x} + \mathbf{v}$. The linearised Euler equation reads

$$m\bar{n}a\left(\dot{\mathbf{v}} + H\mathbf{v}\right) = -\nabla\delta P - m\bar{n}\nabla\delta\Phi \tag{3.24}$$

where we've used the fact that $(\mathbf{v}\cdot\nabla)\mathbf{x} = \mathbf{v}$. If we drop the pressure and gravitational perturbation on the right-hand side, this equation tells us that $\dot{\mathbf{v}} = -H\mathbf{v}$, so the peculiar velocities redshift as $\mathbf{v} \sim 1/a$. This can be viewed as a consequence of Hubble friction, which slows the peculiar velocities as the universe expands.

Finally, the linearised Poisson equation is

$$\nabla^2\delta\Phi = 4\pi G m\bar{n}a^2\delta \tag{3.25}$$

Now we combine our three linearised equations (3.20), (3.24) and (3.25). Take the time derivative of (3.20) to get

$$\ddot{\delta} = \frac{H}{a}\nabla\cdot\mathbf{v} - \frac{1}{a}\nabla\cdot\dot{\mathbf{v}} = -H\dot{\delta} - \frac{1}{a}\nabla\cdot\dot{\mathbf{v}}$$

Take the gradient of (3.24) to get

$$\begin{aligned} m\bar{n}a\left(\nabla\cdot\dot{\mathbf{v}} - Ha\dot{\delta}\right) &= -\nabla^2\delta P - m\bar{n}\nabla^2\delta\Phi \\ &= -m\bar{n}\left(c_s^2\nabla^2\delta + 4\pi G m\bar{n}a^2\delta\right) \end{aligned}$$

where in the second line we've used $\delta P = mc_s^2\delta n = mc_s^2\bar{n}\delta$ and the Poisson equation (3.25). We now combine these two results to get a single equation telling us how the density perturbation δ evolves in an expanding spacetime

$$\ddot{\delta} + 2H\dot{\delta} - c_s^2\left(\frac{1}{a^2}\nabla^2 + k_J^2\right)\delta = 0 \tag{3.26}$$

where k_J is the physical Jeans wavenumber given, as before, by $k_J^2 = 4\pi G m\bar{n}/c_s^2$.

The most important addition from the expanding spac time is the friction-like term $2H\dot{\delta}$. This is referred to as *Hubble friction* or *Hubble drag*. (We saw an analogous term when discussing inflation in Section 1.5.)

To solve (3.26), it is simplest to find by working in Fourier space. We define

$$\delta(\mathbf{k},t) = \int d^3x\; e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{x},t)$$

where we are adopting the annoying but standard convention that the function $\delta(\mathbf{x})$ and its Fourier transform $\delta(\mathbf{k})$ are distinguished only by their argument. Since $\mathbf{x}$ is the co-moving coordinate, $\mathbf{k}$ is the co-moving wavevector.

The advantage of working in Fourier space is that the equation (3.26) decomposes into a separate equation for each value of $\mathbf{k}$,

$$\ddot{\delta}(\mathbf{k},t) + 2H\dot{\delta}(\mathbf{k},t) + c_s^2\left(\frac{k^2}{a^2} - k_J^2\right)\delta(\mathbf{k},t) = 0 \tag{3.27}$$

The slightly unusual factor of a in the final term arises because k is the co-moving wavenumber and so k/a is the physical wavenumber, but k_J refers to the physical Jeans wavenumber. Our challenge now is to solve this equation.

3.1.4 The Growth of Perturbations

Solutions to (3.27) have different behaviour depending on whether the perturbations have small or large wavelength compared to the Jeans' wavelength $\lambda_J = 2\pi/k_J$.

Small wavelength modes have $k/a \gg k_J$. Here, the equation (3.27) is essentially that of a damped harmonic oscillator, with the expanding universe providing the friction term $2H\dot{\delta}$. The solutions are oscillating sound-waves with the Hubble friction leading to an ever-decreasing amplitude.

If structure is to ultimately form in the universe, we need to find solutions that grow over time. These are supplied by the long-wavelength modes, with $k/a < k_J$, which suffer from the Jeans' instability. However, as we shall now see, the details of the Jeans' instability are altered in an expanding universe.

In what follows, we will see that there are two length scales at play for the growing modes. One is the Jeans' length scale $\lambda_J = 2\pi/k_J$,

$$\lambda_J = c_s\sqrt{\frac{\pi}{Gm\bar{n}}} = c_s c\sqrt{\frac{\pi}{G\bar{\rho}}} \tag{3.28}$$

Only modes with $\lambda > \lambda_J$ will grow. The other relevant physical length scale is set by the expansion of the universe,

$$d_H \approx cH^{-1} = c^2\sqrt{\frac{3}{8\pi G\bar{\rho}}} \tag{3.29}$$

This is called the *apparent horizon*. In the standard FRW cosmology (with ordinary matter or radiation) the apparent horizon coincides with the particle horizon, defined in (1.24). In such a situation, it would make little sense to talk about perturbations with wavelength $\lambda > d_H$. This is because the Fourier mode of a perturbation is a coherent wave and causality would appear to prohibit the formation of such perturbations on distances greater than d_H since there has been no time for light, or anything else, to cross this distance since the Big Bang.

This, however, is exactly the problem that is resolved by a period of inflation in the very early universe. The whole point of inflation is to stretch the particle horizon so that it sits way outside the apparent horizon. Indeed, we will see that perturbation modes with wavelengths $\lambda > d_H$ play an important role in the story of structure formation in our universe, strongly implying that a period of inflation is needed. In what follows, we will refer to the apparent horizon (3.29) simply as the "horizon".

Matter Perturbations in a Matter Dominated Universe

For non-relativistic fluids, the Jeans' length (3.28) always sits well within the horizon (3.29),

$$c_s \ll c \quad \Rightarrow \quad \lambda_J \ll cH^{-1}$$

This means that the perturbations which suffer the Jeans' instability include both sub-horizon and super-horizon wavelengths.

As the wavelength of the mode is sufficiently long, so $k/a \ll k_J$, then we can approximate (3.27) as

$$\ddot{\delta}(\mathbf{k}) + 2H\dot{\delta}(\mathbf{k}) - \frac{4\pi G\bar{\rho}}{c^2}\,\delta(\mathbf{k}) = 0 \tag{3.30}$$

Here we've left the t argument in $\delta(\mathbf{k}, t)$ implicit, but kept the $\mathbf{k}$ argument because it tells us that the mode is in Fourier space rather than real space.

In a matter dominated universe, $a \sim t^{2/3}$ so $H = 2/3t$. The third term is also related to the Hubble parameter through Friedmann equation $H^2 = 8\pi G\bar{\rho}/3c^2$. We then have

$$\ddot{\delta}(\mathbf{k}) + \frac{4}{3t}\dot{\delta}(\mathbf{k}) - \frac{2}{3t^2}\delta(\mathbf{k}) = 0$$

Substituting in the power-law ansatz $\delta(\mathbf{k}, t) \sim t^n$, we find two solutions, one decaying and one growing

$$\delta(\mathbf{k}, t) \sim \begin{cases} t^{2/3} \sim a \\ t^{-1} \sim a^{-3/2} \end{cases} \tag{3.31}$$

We see that the expansion of the universe slows down the rate at which objects undergo gravitational collapse, with the Hubble damping turning the exponential growth of the Jeans instability (3.12) into a power-law, one that scales linearly with the size of the universe.

Radiation Perturbations in a Radiation Dominated Universe

Although we have derived the perturbation equation (3.27) for non-relativistic fluids, it is not too difficult to modify them in a plausible way to give us an understanding of the perturbations in other fluid components. Here we will be interested in radiation, but things are clearer if we work with the general equation of state

$$P = w\rho$$

and only later restrict to $w = 1/3$.

We will work with the energy density $\rho(\mathbf{x}, t)$, rather than the number density $n(\mathbf{x}, t)$; for a non-relativistic fluid, they are related by $\rho = mnc^2$. We need to go through each of our original equations – continuity, Euler, and Poisson – and ask how they change for a general fluid. We will motivate each of these changes, but not derive them.

First, the continuity equation (3.18): this gets replaced by

$$\frac{\partial \rho}{\partial t} + 3(1+w)H\rho + \frac{1}{a}(1+w)\nabla \cdot (\rho \mathbf{v}) = 0$$

The equation of state parameter w appears twice. The first of these is unsurprising, since it guarantees that this equation reduces to our previous continuity equation (1.39) when $\mathbf{v} = 0$. We won't derive the presence of the $(1 + w)$ factor in the final term, but it arises in a similar way to the first.

The Euler equation (3.21) remains unchanged. Somewhat more subtle is the relativistic generalisation of the gravitational potential. In general relativity, both energy density and pressure gravitate. It turns out that the Poisson equation (3.22) should be replaced by

$$\nabla^2\Phi = \frac{4\pi G}{c^2}(1+3w)\rho a^2 \tag{3.32}$$

There is, in fact, a clue in the discussion above that strongly hints at this form. Recall that we avoided the Jeans' swindle in an expanding spacetime by relating the gravitational potential to the acceleration in (3.23). The Poisson equation then became equivalent to the acceleration equation (1.52) which, in general, reads

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(1+3w)\rho$$

We see the same distinctive factor of $(1+3w)$ appearing here.

Repeating the same steps as previously, the perturbation equation (3.27) is replaced by

$$\ddot{\delta}(\mathbf{k}) + 2H\dot{\delta}(\mathbf{k}) + c_s^2(1+w)\left(\frac{k^2}{a^2} - (1+3w)k_J^2\right)\delta(\mathbf{k}) = 0 \tag{3.33}$$

with $k_J = 2\pi/\lambda_J$ is the physical wavenumber, defined as in (3.28). It differs from the non-relativistic Jeans length only by the expression for the speed of sound c_s.

Let's now restrict to radiation with $w = 1/3$. We know from (3.9) that the speed of sound for a relativistic fluid is

$$c_s^2 = \frac{1}{3}c^2$$

This means that there is no parametric separation between the Jeans length (3.28) and the horizon (3.29). Instead, we have $\lambda_J \approx cH^{-1}$. Any perturbation that lies inside the horizon does not grow. Instead, the pressure of the radiation causes the perturbation to oscillate as a sound wave.

Outside the horizon it's a different story. In a radiation dominated universe, $a \sim t^{1/2}$ so $H = 1/2t$. For wavenumbers $k/a \ll k_J$, the equation (3.33) governing perturbations becomes

$$\ddot{\delta}_r(\mathbf{k}) + \frac{1}{t}\dot{\delta}_r(\mathbf{k}) - \frac{1}{t^2}\delta_r(\mathbf{k}) = 0$$

Once again, substituting in the power-law ansatz $\delta(\mathbf{k}, t) \sim t^n$, we find two solutions, one decaying and one growing

$$\delta_r(\mathbf{k}, t) \sim \begin{cases} t & \sim a^2 \\ t^{-1} & \sim a^{-2} \end{cases} \tag{3.34}$$

We learn that perturbations in the density of radiation grow outside the horizon. Indeed, they grow faster than the linear growth (3.31) seen in the matter dominated era.

Matter Perturbations in a Radiation Dominated Universe

We could also ask about density perturbations of matter in a radiation dominated universe. As we've seen, the Jeans length for matter is well within the horizon (because $c_s \ll c$). In a universe with multiple energy components ρ_i, matter perturbations δ_m with $k/a \ll k_J$ are described by a modified version of (3.30),

$$\ddot{\delta}_m(\mathbf{k}) + 2H\dot{\delta}_m(\mathbf{k}) - \frac{4\pi G}{c^2}\sum_i \bar{\rho}_i \delta_i(\mathbf{k}) = 0 \tag{3.35}$$

This final term can be traced to the gravitational potential $\delta\Phi$, which receives contributions from all energy sources. However, on sub-horizon scales, we have seen that the radiation perturbation does not grow, so we can set $\delta_r(\mathbf{k}) \approx 0$. Meanwhile, in the radiation dominated phase $\bar{\rho}_r \gg \bar{\rho}_m$ and so we can also ignore the $\bar{\rho}_m \delta_m(\mathbf{k})$ which will be sub-dominant to the $H\dot{\delta}_m(\mathbf{k})$ term. Using $H = 1/2t$, we have

$$\ddot{\delta}_m(\mathbf{k}) + \frac{1}{t}\dot{\delta}_m(\mathbf{k}) \approx 0 \quad \Rightarrow \quad \delta_m(\mathbf{k}, t) \sim \begin{cases} \log t \sim \log a \\ \text{constant} \end{cases} \tag{3.36}$$

We learn that, during the radiation dominated era, the matter perturbations inside the horizon grow only logarithmically. This slow growth occurs because the expansion of the universe is faster in the radiation dominated phase than in the matter dominated phase. A logarithmic increase is rather pathetic and it means that significant growth in sub-horizon scale perturbations gets going when we hit the matter dominated era at $z \approx 3400$

In contrast, the matter perturbations with wavelength larger than the horizon obey the same equation as the radiation perturbations in the radiation dominated era. (The term $\bar{\rho}_r \delta_r$ in (3.35) cannot now be neglected.) This means that those modes outside the horizon grow as $\delta \sim a^2$ as seen in (3.34).

The Cosmological Constant

We can also ask how matter perturbations grow in a universe dominated by a cosmological constant. We again use the perturbation equation (3.35). It is not possible to have perturbations δ_Λ. (This is what the "constant" in cosmological constant means!) Once again, $\bar{\rho}_m$ is negligible, so we have

$$\ddot{\delta}_m + 2H\dot{\delta}_m \approx 0$$

where, from Section 1.3.3, $H = \sqrt{\Lambda/3}$. The solutions are now

$$\delta_m \sim \begin{cases} \text{constant} \\ e^{-2Ht} \sim a^{-2} \end{cases}$$

We learn that in a universe dominated by a cosmological constant, there is no growth of perturbations. In other words, dark energy kills any opportunity to form galaxies. We will revisit this in Section 3.3.4.

3.1.5 Validity of the Newtonian Approximation

Everything we've done in this section relies on the perturbation equation (3.27), which was derived for non-relativistic matter using Newtonian gravity. However, as we stressed in Section 1.2, a proper description of expanding spacetime requires general relativity. So should we trust the Newtonian approximation?

We should be able to trust our equations on small length scales, for the simple reason that general relativity reduces to Newtonian gravity in this regime. However, when we get to perturbations whose length is comparable to the horizon d_H, we should be more nervous, since it seems plausible that the perturbations feel the curvature of spacetime in a way that our Newtonian approximation misses.

The only way to know if the Newtonian perturbation equation (3.27) is valid is to roll up our sleeves and perform the correct, general relativistic perturbation theory. This is a somewhat painful exercise that you will be given the opportunity to embrace in next year's Part III cosmology course. There you will learn that the question "is our equation (3.27) valid?" has a short answer and a long answer.

The short answer is: yes.

The long answer is substantially more subtle. It turns out that the matter perturbation in general relativity is not diffeomorphism invariant, which means that the answer you get depends on the coordinates you use. This is bad. Indeed, one of the main philosophical lessons of general relativity is that the coordinates you use should not matter one iota. Moreover, this issue is particularly problematic for super-horizon perturbations with $\lambda \gg d_H$, and an important part of the relativistic approach is to understand the right, diffeomorphism invariant quantity to focus on. For most choices of coordinates (so called "gauges") it turns out that the Poisson equation (3.25) is not valid on super-horizon scales. There is, however, a choice of coordinates – conformal Newtonian gauge – where the Poisson equation holds even on super-horizon scales and this is the one we are implicitly choosing[9]. All of this is to say that you can trust the physics that we've derived here but you should be careful when comparing to analogous results derived in a general relativistic setting where the answers may look different because, although the symbols are the same, they refer to subtly different objects.

3.1.6 The Transfer Function

There are a number of different questions that we could try to answer now. We posed one such question in the introduction to this section: can we compute the overall growth of the density perturbations to explain how we got from $\delta T/T \sim 10^{-5}$ in the CMB to the world we see around us. This, it turns out, requires some more discussion which we postpone to Section 3.3.2. Instead, we will ask about the relative growth of density perturbations of different wavenumber k.

We are interested in the perturbations of the matter, since this is what we're ultimately made of. If the density perturbations remain sufficiently small, so that the linearised analysis developed above holds then the linear analysis of this section remains valid which, in particular, means that each $\delta(\mathbf{k})$ evolves independently, as seen in, for example, (3.27). The evolution of a perturbation of a given wavevector $\mathbf{k}$ from an initial time t_i to the present can be distilled into a *transfer function* $T(k)$, defined as

$$\delta(\mathbf{k}, t_0) = T(k)\, \delta(\mathbf{k}, t_i) \tag{3.37}$$

The initial time t_i is usually taken to be early, typically just after the end of inflation.

Key to understanding the physics is the question of when perturbations enter the horizon. Recall that, in physical coordinates, the apparent horizon is $d_H \approx c/H$ as in

[9]More details can be found in the paper by Chisari and Zaldarriaga, "*Connection between Newtonian simulations and general relativity*, arXiv:1101.3555.

(3.29). It is simplest, however, to work in co-moving coordinates, where the apparent horizon is

$$\chi_H = \frac{c}{aH}$$

In the radiation dominated era, $a \sim t^{1/2}$ and so $H \sim 1/a^2$. In the matter dominated era, $a \sim t^{2/3}$ and so $H \sim 1/a^{3/2}$. In both cases, the co-moving horizon increases over time

$$\chi_H \sim \begin{cases} a & \text{radiation domination} \\ a^{1/2} & \text{matter domination} \end{cases} \tag{3.38}$$

The intuition behind this is that, as the universe expands, there is more that one can see and, correspondingly, the co-moving horizon grows.

The co-moving wavevector $\mathbf{k}$ remains unchanged over time. (This is the main advantage of working with the co-moving wavevector in the previous section. In contrast, the physical wavevector is $\mathbf{k}_{\rm phys} = \mathbf{k}/a$ shrinks over time as the physical wavelength $\lambda_{\rm phys} = 2\pi/k_{\rm phys} = 2\pi a/k$ is stretched by the expansion of the universe.) This means that, for each $\mathbf{k}$, there will be a time when the corresponding perturbation enters the horizon. It matters whether the time of entry occurs during the radiation or matter dominated eras.

At the time of matter-radiation equality (which occurred around $z = 3400$), modes with wavenumber $k_{\rm eq}$ have just entered the horizon, where

$$k_{\rm eq} = \frac{2\pi}{c}(aH)_{\rm eq}$$

Modes larger than this (i.e. with $k < k_{\rm eq}$) enter the horizon in the matter dominated era. Modes smaller than this (i.e. with $k > k_{\rm eq}$) enter the horizon during the radiation dominated era. Let's look at each of these in turn.

First the long wavelength modes with $k < k_{\rm eq}$. These were outside the horizon during the radiation era where δ grew as a^2, as seen in (3.34). As the universe entered the matter dominated era, the growth slows to $\delta \sim a$, as seen in (3.31). This means that, starting from an initial time t_i, they evolve to their present day value

$$\delta(\mathbf{k}, t_0) = \left(\frac{a_{\rm eq}}{a_i}\right)^2 \frac{a_0}{a_{\rm eq}}\, \delta(\mathbf{k}, t_i) \qquad \text{for } k < k_{\rm eq} \tag{3.39}$$

We learn that each mode grows by an amount independent of $\mathbf{k}$.

Things are more interesting for the short wavelength modes with $k > k_{\rm eq}$ which enter the horizon during the radiation era. Before entering the horizon, such modes grow as $\delta \sim a^2$ as in (3.34). However, when they enter the horizon, their growth slows to the logarithmic growth seen in (3.36). For our purposes, this is effectively constant. The growth only resumes when the universe enters the matter dominated era. This means that

$$\begin{aligned}\delta(\mathbf{k}, t_0) &= \left(\frac{a_{\rm enter}}{a_i}\right)^2 \frac{a_0}{a_{\rm eq}}\,\delta(\mathbf{k}, t_i) \\ &= \left(\frac{a_{\rm enter}}{a_{\rm eq}}\right)^2 \times \left[\left(\frac{a_{\rm eq}}{a_i}\right)^2 \frac{a_0}{a_{\rm eq}}\right] \delta(\mathbf{k}, t_i) \qquad \text{for } k > k_{\rm eq}\end{aligned} \tag{3.40}$$

The factor in square brackets is the same, constant amount (3.39) that the long wavelength modes grew by. However, the amplitude is suppressed by the factor of $a_{\rm enter}^2/a_{\rm eq}^2$, reflecting the fact that growth stalled during the radiation dominated era. For a given mode $\mathbf{k}$, the scale factor at horizon entry is given by

$$k = \frac{2\pi}{c}(aH)_{\rm enter}$$

Using $a \sim t^{1/2}$ in the radiation era, we have $H = 1/2t \sim 1/a^2$ so a given scale k enters the horizon at $k \sim (aH)_{\rm enter} \sim 1/a_{\rm enter}$. We can then write $(a_{\rm enter}/a_{\rm eq})^2 = k_{\rm eq}^2/k^2$.

Finally, all of this can be packaged into the transfer function (3.37). Assuming that the perturbations remain sufficiently small, so the linearised analysis is valid, the transfer function can be found in (3.39) and (3.40). It scales with the wavenumber as

$$T(k) \sim \begin{cases} 1 & k < k_{\rm eq} \\ k^{-2} & k > k_{\rm eq} \end{cases} \tag{3.41}$$

We will make use of this shortly.

3.2 The Power Spectrum

It should be obvious that we're not going to understand the density perturbations in the early universe to enough accuracy to predict the location of, say, my mum's house. Or even the location of the Milky Way galaxy. Instead, if we want to make progress then we must lower our ambitions. We will need to develop a statistical understanding of the distribution of galaxies in the universe.

To this end, we consider various averages of the density contrast,

$$\delta(\mathbf{x},t) = \frac{\delta\rho(\mathbf{x},t)}{\bar{\rho}(t)}$$

By construction, the spatial average of δ itself at a given time vanishes,

$$\langle\delta(\mathbf{x},t)\rangle = 0$$

The first non-trivial information lies in the correlation function, defined by the spatial average

$$\xi(|\mathbf{x}-\mathbf{y}|,t) = \langle\delta(\mathbf{x},t)\,\delta(\mathbf{y},t)\rangle \tag{3.42}$$

Our old friend, the cosmological principle, is implicit in the left-hand side where we have assumed that the universe is statistically homogeneous and isotropic, so that the function $\xi(\mathbf{x},\mathbf{y},t)$ depends only on $|\mathbf{x}-\mathbf{y}|$. The correlation function $\xi(r,t)$ tells us the likelihood that, at time t, two galaxies are separated by a distance r.

Further statistical information about $\delta(\mathbf{x})$ can be distilled into higher correlation functions, such as $\langle\delta\delta\delta\rangle$ However, in what follows we will limit ourselves to understanding the correlation function $\xi(r,t)$.

In Section 3.1, we learned that the evolution of the density perturbations is best described in momentum space,

$$\delta(\mathbf{k},t) = \int d^3x\; e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta(\mathbf{x},t) \tag{3.43}$$

The correlation function in momentum space is given by

$$\begin{aligned}
\langle\delta(\mathbf{k},t)\,\delta(\mathbf{k}',t)\rangle &= \int d^3x\,d^3y\; e^{i\mathbf{k}\cdot\mathbf{x}+i\mathbf{k}'\cdot\mathbf{y}}\,\langle\delta(\mathbf{x},t)\,\delta(\mathbf{y},t)\rangle \\
&= \int d^3x\,d^3y\; e^{i\mathbf{k}\cdot\mathbf{x}+i\mathbf{k}'\cdot\mathbf{y}}\,\xi(r,t) \\
&= \int d^3r\,d^3y\; e^{i\mathbf{k}\cdot\mathbf{r}+i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{y}}\,\xi(r,t) \\
&= (2\pi)^3\,\delta_D^3(\mathbf{k}+\mathbf{k}')\int d^3r\; e^{i\mathbf{k}\cdot\mathbf{r}}\,\xi(r,t)
\end{aligned} \tag{3.44}$$

where, in the second line, we've defined $\mathbf{r} = \mathbf{x}-\mathbf{y}$. The Dirac delta-function $\delta_D^3(\mathbf{k}+\mathbf{k}')$ reflects the underlying (statistical) translation invariance. (Note that I've added a

subscript D on the Dirac delta $\delta_D^3(\mathbf{k})$ to distinguish it from the density contrast $\delta(\mathbf{k}, t)$!) The remaining function is called the *power spectrum*,

$$P(k,t) = \int d^3r \; e^{i\mathbf{k}\cdot\mathbf{r}}\, \xi(r,t)$$

This is the three-dimensional Fourier transform of the correlation function. If we work in spherical polar coordinates, chosen so that $\mathbf{k}\cdot\mathbf{r} = kr\cos\theta$, then we have

$$\begin{aligned} P(k,t) &= \int_0^{2\pi} d\phi \int_{-1}^{+1} d(\cos\theta) \int_0^\infty dr \; r^2 e^{ikr\cos\theta} \xi(r,t) \\ &= 2\pi \int_0^\infty dr \; \frac{r^2}{ikr} \left[e^{ikr} - e^{-ikr}\right] \xi(r,t) \\ &= \frac{4\pi}{k} \int_0^\infty dr \; r\sin(kr)\, \xi(r,t) \end{aligned} \tag{3.45}$$

The spatial correlation function $\xi(r)$ can be measured by averaging over many galaxies in the sky. (We'll say more about this in Section 3.2.5.) Meanwhile, the power spectrum $P(k)$ is the most natural theoretical object to consider. The formula above relates the two.

3.2.1 Adiabatic, Gaussian Perturbations

To describe the structure of galaxies in our universe, we introduce a probability distribution for $\delta(k)$. The idea is that averages computed from the distribution will coincide with the spatial average which leads to $\xi(r)$ and, relatedly, $P(k)$.

There are two basic questions that we need to address:

- What is the initial probability distribution?
- How did this probability distribution subsequently evolve?

If we understand both of these well enough, we should be able to compare our results to the distribution of galaxies observed in the sky. We start by describing the initial probability distribution. We then see how it evolves in Section 3.2.3.

It may seem daunting to guess the form of the initial perturbations. However, the universe is kind to us and the observational evidence suggests that these perturbations take the simplest form possible. (We will offer an explanation for this in Section 3.5.)

First, the perturbations of each fluid component are correlated. In particular, the perturbation in any non-relativistic matter, such as baryons and cold dark matter, is the same: $\delta_B = \delta_{CDM}$. Furthermore, perturbations in matter and perturbations in radiation are related by

$$\delta_m = \frac{3}{4}\delta_r \tag{3.46}$$

Perturbations of this kind are called *adiabatic*.

It may seem like a minor miracle that the perturbations in all fluids are correlated in this way. What's really happening is that there is an initial perturbation in the gravitational potential (or, in the language of general relativity, in the metric) which, in turn, imprints itself on each of the fluids in the same way.

Logically, we could also have initial perturbations of the form $\delta\rho_m = -\delta\rho_r$. These are referred to as *isocurvature perturbations* because the net perturbation $\delta\rho = \delta\rho_m + \delta\rho_r = 0$ gives no change to the local curvature of spacetime. There is no hint of these isocurvature perturbations in our universe.

Since we have adiabatic perturbations, we need only specify a probability distribution for a single component, which we take to be $\delta \equiv \delta_m$. We take this distribution to be a simple Gaussian

$$\text{Prob}\Big[\delta(\mathbf{k})\Big] = \frac{1}{\sqrt{2\pi P(k)}} \exp\left(-\frac{\delta(\mathbf{k})^2}{2P(k)}\right) \tag{3.47}$$

This expression holds for each $\mathbf{k}$ independently. This means that there is no correlation between perturbations with different wavelengths. This is an assumption, and one that can be tested since it means that, at least initially, all higher point correlation functions are determined purely in terms of the one-point and two-point functions. For example, $\langle\delta\delta\delta\rangle = 0$.

Note that the power spectrum $P(k)$ arises in this distribution in the guise of the variance. This ensures that the two-point function is indeed given by

$$\langle\delta(\mathbf{k}, t_i)\,\delta(\mathbf{k}', t_i)\rangle = (2\pi)^3\,\delta_D^3(\mathbf{k} + \mathbf{k}')\,P(k) \tag{3.48}$$

It remains only to specify the form of the power spectrum $P(k)$ for these initial perturbations. These are usually taken to have the power-law form

$$P(k) = Ak^n \tag{3.49}$$

for constants A and n. The exponent n is called the *spectral index*.

A power-law $P \sim k^n$ gives rise to a real space correlation function $\xi(r) \sim 1/r^{n+3}$. (Actually, one must work a little harder to make sense of the inverse Fourier transform (3.45) at high k, or small r.) The choice $n = 0$ is what we would get if we sprinkle points at random in space; it is sometimes referred to as *white noise*. (We'll build more intuition for this in Section 3.2.2 below.) Meanwhile, any $n < -3$ means that $\xi(r) \to \infty$ as $r \to \infty$, so the universe gets more inhomogeneous at large scales, in contradiction to the cosmological principle. We'd like to ask: what choice of spectral index n describes our universe?

The Harrison-Zel'dovich Spectrum

A particularly special choice for the initial power spectrum is

$$n = 1$$

This is known as the *Harrison-Zel'dovich* power spectrum (named after Harrison, Zel'dovich, and Peebles and Yu). It is special for two reasons. First, and most importantly, it turns out to be almost (but not quite!) the initial spectrum of density perturbations in our universe. Second, it also has a special mathematical property.

To explain this mathematical property, we need some new definitions. We start by some simple dimensional analysis. The original perturbation $\delta(\mathbf{x}) = \delta\rho/\rho$ was dimensionless, so after a Fourier transform (3.43) the perturbation $\delta(\mathbf{k})$ has dimension [length]^3. The delta-function $\delta_D^3(\mathbf{k})$ also has dimension $[k]^{-3} = [\text{length}]^3$ which means that the power spectrum $P(k)$ also has dimension [length]^3. It is often useful to define the dimensionless power spectrum

$$\Delta(k) = \frac{4\pi k^3\, P(k)}{(2\pi)^3} \tag{3.50}$$

The factors of 2 and π are conventional. Because $\Delta(k)$ is dimensionless, it makes sense to say that, for example, $\Delta(k)$ is a constant. Unfortunately, as you can see, this does not give rise to the Harrison-Zel'dovich spectrum.

However, we can also look at fluctuations in other quantities. In particular, rather than talk about perturbations in the density ρ, we could instead talk about perturbations in the gravitational potential: $\Phi(\mathbf{x}) = \bar{\Phi}(\mathbf{x}) + \delta\Phi(\mathbf{x})$. The two are related by the Poisson equation (3.32)

$$\nabla^2\delta\Phi = \frac{4\pi G}{c^2}(1+3w)\bar{\rho}a^2\delta \quad \Rightarrow \quad -k^2\delta\Phi(\mathbf{k}) = \frac{4\pi G}{c^2}(1+3w)\bar{\rho}a^2\delta(\mathbf{k}) \tag{3.51}$$

We can then construct the power spectrum of gravitational perturbations

$$\langle \delta\Phi(\mathbf{k})\, \delta\Phi(\mathbf{k}')\rangle = (2\pi)^3\, \delta_D^3(\mathbf{k}+\mathbf{k}')\, P_\Phi(k) \tag{3.52}$$

and the corresponding dimensionless gravitational power spectrum

$$\Delta_\Phi = \frac{4\pi k^3\, P_\Phi(k)}{(2\pi)^3}$$

The Poisson equation (3.51) tells us that there's a simple relationship between $P_\Phi(k)$ and $P(k)$, namely

$$P_\Phi(k) \propto k^{-4} P(k) \tag{3.53}$$

where the proportionality factor hides the various constants arising from the Poisson equation. We can write this as

$$P(k) \propto k^4 P_\Phi(k) \propto k\Delta_\Phi$$

We see that the Harrison-Zel'dovich spectrum arises if the initial gravitational perturbations are independent of the wavelength, in the sense that Δ_Φ = constant. Such fluctuations are said to be *scale invariant*. We will see that such scale invariant perturbations in the gravitational potential are a good description of our universe, and hold an important clue to what was happening at the very earliest times. We will see what this clue is telling us in Section 3.5. First, however, it will be useful to pause to build some intuition for these different probability distributions.

3.2.2 Building Intuition For Gaussian Distributions

The discussion above can be bafflingly formal when you first meet it. At this stage, it's useful to build some intuition for what the different power spectra look like and, in particular, why $P_\Phi \sim 1/k^3$ corresponds to a scale invariant distribution.

To visualise what's going on, we'll ultimately show some pictures of distributions in $d = 2$ spatial dimensions. But, for now, let's keep the spatial dimension d arbitrary. We'll focus on the probability distribution of some scalar field $\Phi(\mathbf{x})$ which, in the cosmological context, you should think of as the gravitational perturbation $\delta\Phi$. However, for the purposes of our discussion, $\Phi(\mathbf{x})$ could be any scalar field. The Fourier transform is

$$\Phi(\mathbf{k}) = \int d^d x\; e^{i\mathbf{k}\cdot\mathbf{x}} \Phi(\mathbf{x})$$

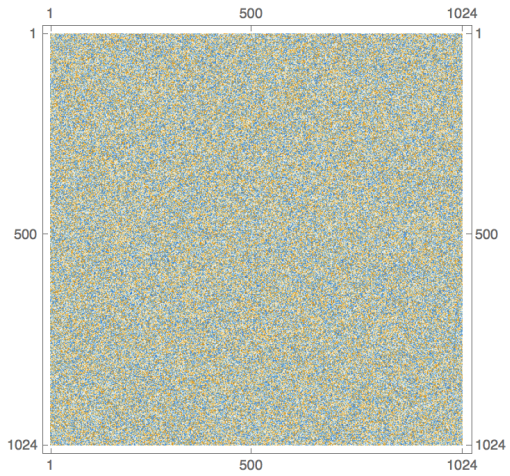

Figure 33: The white noise distribution in $d = 2$ dimensions with $n = 4$.

and we will ask that this takes values drawn from a Gaussian probability distribution of the form (3.52),

$$\langle \Phi(\mathbf{k})\Phi(\mathbf{k}')\rangle = (2\pi)^d\, \delta^d(\mathbf{k}+\mathbf{k}')\, P_\Phi(k)$$

where $\delta^d(\mathbf{k}+\mathbf{k}')$ is the usual d-dimensional delta function. The question that we'd like to ask is: what does such a distribution mean for $\Phi(\mathbf{x})$ and, in particular, how does the choice of power spectrum $P_\Phi(k)$ affect it?

In position space, the two-point correlation function is given by the Fourier transform of the power spectrum,

$$\begin{aligned}\langle \Phi(\mathbf{x})\Phi(\mathbf{y})\rangle &= \int \frac{d^dk}{(2\pi)^d}\int \frac{d^dk'}{(2\pi)^d} e^{-i\mathbf{k}\cdot\mathbf{x}-i\mathbf{k}'\cdot\mathbf{y}}\langle \Phi(\mathbf{k})\Phi(\mathbf{k}')\rangle \\ &= \int \frac{d^dk}{(2\pi)^d} e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} P_\Phi(k) \end{aligned} \tag{3.54}$$

We'll now look at what this means for a power spectrum of the form

$$P_\Phi(k) = k^{n-4} \tag{3.55}$$

for various choices of integer n. (The exponent here is chosen to match our previous conventions.)

Obviously, in cosmology we're interested in $d = 3$ spatial dimensions. However, below we'll plot distributions in $d = 2$ dimensions. The key physics is the same but, as we'll

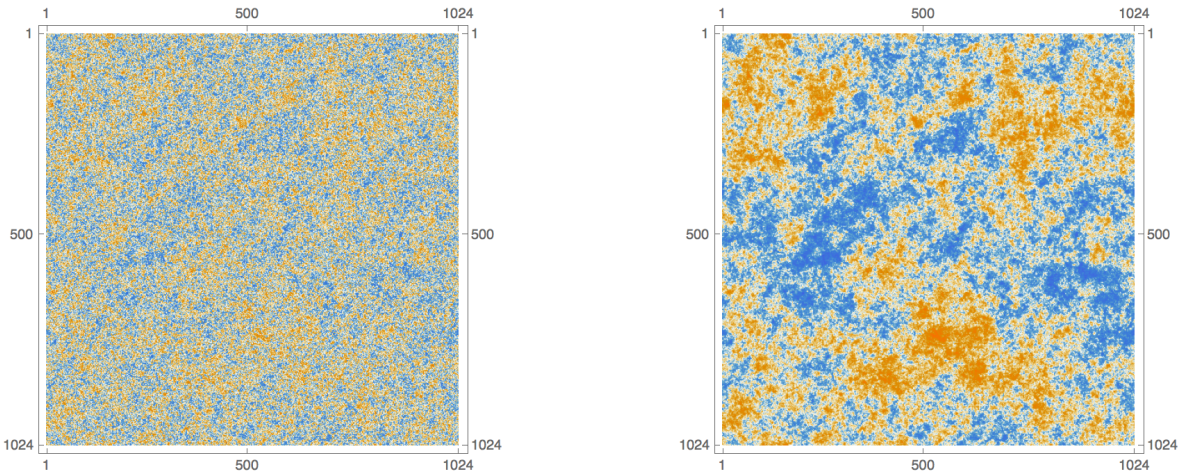

Figure 34: The distribution in $d = 2$ dimensions with $n = 3$, corresponding to $P_\Phi \sim 1/k$ (on the left) and $n = 2$, corresponding to $P_\Phi \sim 1/k^2$ (on the right). The latter is scale invariant in two dimensions.

see, occurs for different values of n. We start with constant power spectrum, or $n = 4$ in the convention of (3.55). Here we have

$$n = 4 \quad \Rightarrow \quad \langle \Phi(\mathbf{x})\Phi(\mathbf{y})\rangle \sim \delta^d(\mathbf{x}-\mathbf{y})$$

This means that there's no correlation between the value of Φ at different points. A typical configuration of $\Phi(\mathbf{x})$ is shown[10] in Figure 33. A distribution like this, with no correlation between neighbouring points, is known as *white noise*. (There's a perennial confusion here: white noise for Φ and white noise for the density perturbation occur for different values of n because the two distributions are related by a power of k^4.)

Now we look at what happens as we decrease n. For $n = 3$, corresponding to $P_\Phi(k) \sim k^{-1}$, the correlation between neighbouring points becomes stronger. A typical distribution is shown on the left in Figure 34. We see that if the field takes a particular value at some point $\mathbf{x}$, there is now an increased likelihood that it takes similar values at neighbouring points.

This likelihood increases further as we lower n. The distribution for $n = 2$ is shown on the right in Figure 34. This distribution is rather special since it gives $P_\Phi(k) \sim 1/k^2$

[10] All the images of distributions were created using Garrett Goon's publicly available mathematica script. Operationally, this starts with the white noise of Figure 33, Fourier transforms to momentum space, multiplies the resulting distribution by $P(k)$, and then Fourier transforms back. A clear and detailed account of this can be found on Garrett's webpage https://garrettgoon.com/gaussian-fields/.

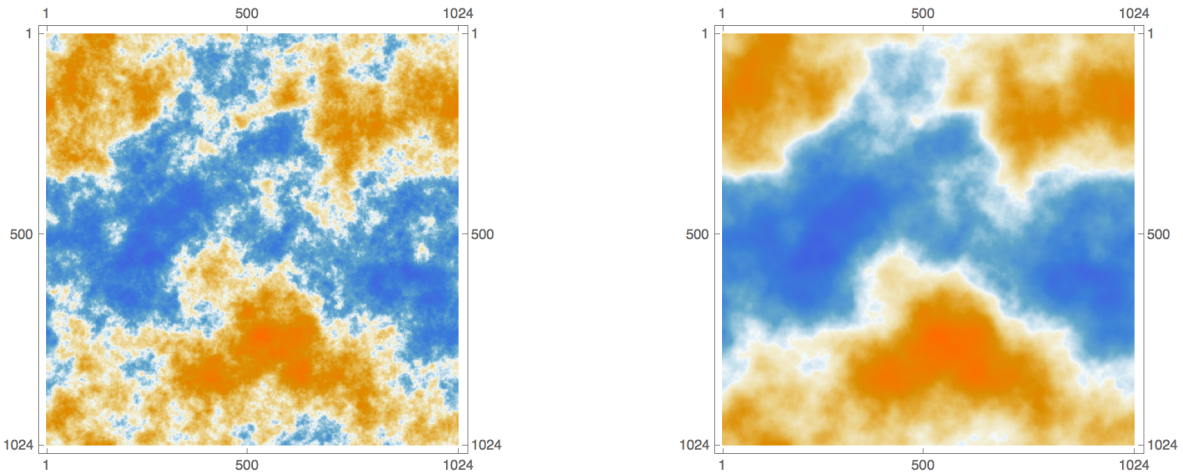

Figure 35: The distribution in two dimensions with $n = 1$ corresponding to $P_\Phi(k) = 1/k^3$ (on the left) and $n = 0$ corresponding to $P_\Phi(k) = 1/k^4$ (on the right).

and, in d spatial dimensions, the distribution $1/k^d$ is scale invariant. This means that the correlation between any two points is independent of the distance between those points! To see this, we simply need to rescale the correlation function (3.54) to find

$$\langle \Phi(\lambda \mathbf{x})\Phi(\lambda \mathbf{y})\rangle = \int \frac{d^d k}{(2\pi)^d} \frac{e^{-i\lambda \mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}}{k^d} = \langle \Phi(\mathbf{x})\Phi(\mathbf{y})\rangle$$

where the final equality holds by redefining $\mathbf{k}' = \lambda \mathbf{k}$ to remove λ from the exponent, and then noting that the factors of λ cancel between the measure factor $d^d k$ and the $1/k^d$ in the power spectrum.

We can decrease n still further, to find configurations in which the spatial correlation increases. Examples for $n = 1$ and $n = 0$ are shown in Figure 35.

3.2.3 The Power Spectrum Today

The Gaussian distribution (3.47) holds at some initial time t_i, which we take to be a very early time, typically just after inflation. As we have seen, the subsequent evolution of the density perturbations is described by the transfer function

$$\delta(\mathbf{k}, t_0) = T(k)\, \delta(\mathbf{k}, t_i)$$

We computed this for non-relativistic matter in (3.41); it is

$$T(k) \sim \text{constant} \times \begin{cases} 1 & k < k_{\rm eq} \\ k^{-2} & k > k_{\rm eq} \end{cases}$$

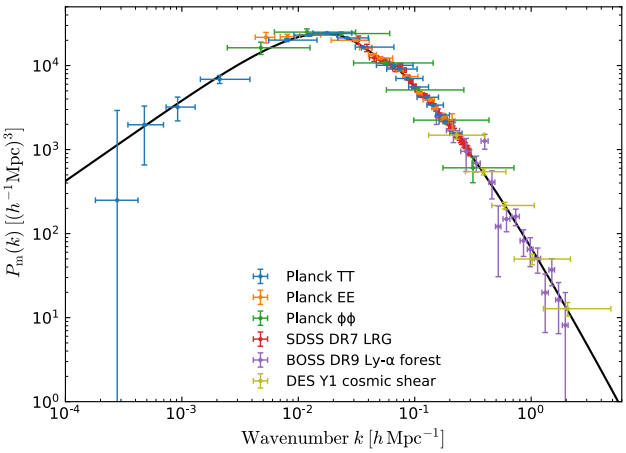

Figure 36: The observed matter power spectrum.

In general, each fluid component will have a separate transfer function, so that the adiabatic form of the initial perturbations (3.46) gets ruined as the universe evolves.

Provided that this linear analysis is valid, the distribution of fluctuations remains Gaussian, and only the power spectrum $P(k)$ changes. From the relation $P \sim \langle \delta\delta \rangle$, we have

$$P(k; t_0) = T^2(k)\, P(k; t_i)$$

As the density perturbations get large, linear perturbation theory breaks down and the evolution becomes non-linear. In this situation, perturbations with different wavevector $\mathbf{k}$ start to interact and the simple Gaussian distribution no longer holds. If we want to get a good handle on the late time universe, filled with galaxies and clusters, we must ultimately understand this non-linear behaviour. We'll start to explore this in Section 3.3 but, for now, we will content ourselves with the simple linear evolution.

If we start with the power-law spectrum $P \sim k^n$, then it subsequently evolves to

$$P(k) \sim \begin{cases} k^n & k < k_{\rm eq} \\ k^{n-4} & k > k_{\rm eq} \end{cases} \tag{3.56}$$

with the turnover near $ak \approx ak_{\rm eq} \sim 0.01$ Mpc^{-1}. A more careful analysis shows that the turnover at $k = k_{\rm eq}$ happens rather gradually.

We can now compare these expectations with the observed matter power spectrum. Data taken from a number of different sources, is shown[11] in Figure 36. At very large scales (small k) the data is taken from the CMB; we will discuss this further in Section 3.4. Longer wavelength structures are seen through various methods of measuring of structure in the universe today. One finds that the data fits very well with the initial Harrison-Zel'dovich power-law spectrum $n = 1$. More accurate observations reveal, a slight deviation from the perfect Harrison-Zel'dovich spectrum. Both large scale structure[12], and CMB measurements (which are discussed briefly in the next section) give

$$n \approx 0.97$$

The fact that perturbations in the early universe are almost, but not quite, described by the Harrison-Zel'dovich spectrum is an important clue for what was happening in the very early universe. A precise scale invariant Harrison-Zel'dovich spectrum is telling us that there must have been some symmetry in the early universe; the deviation is telling us that there was some dynamics taking place which breaks this symmetry. We will describe this more in Section 3.5.

3.2.4 Baryonic Acoustic Oscillations

There is a time in the early universe, bounded by redshifts

$$1100 \lesssim z \lesssim 3400$$

when the expansion was dominated by matter, but hydrogen had not yet formed. As we saw in Section 2, in this epoch protons, electrons and photons were in thermal equilibrium. In such a photon-baryon fluid, the speed of sound is determined by the photons rather than the matter, so $c_s \approx c/\sqrt{3}$. This means that the effective Jeans length for baryonic matter is much greater than the corresponding length for dark matter.

The consequence is that dark matter and baryonic matter behave differently in this epoch. Density perturbations in dark matter, which long ago decoupled from the photons, start to grow as $\delta \sim a$ as in (3.31). Meanwhile, density perturbations in

[11] This plot is taken from the Planck 2018 results, "*Overview and the cosmological legacy of Planck*, arXiv:1807.06205.

[12] For example, the paper *The one-dimensional Ly-alpha forest power spectrum from BOSS* by N. Palanque-Delabrouille et. al. arXiv:1306.5896 finds $n = 0.97 \pm 0.02$. Meanwhile, the Planck collaboration, in arXiv:1807.06211, quotes $n_s = 0.9649 \pm 0.0042$.

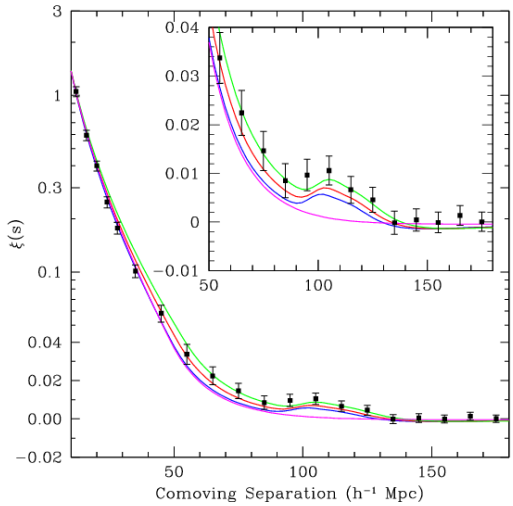

Figure 37: Baryonic acoustic oscillations seen in the distribution of galaxies.

baryonic matter are supported by the pressure from the photons and, at least on sub-horizon scale, oscillate. The resulting sound waves in the baryon-photon fluid are known as *baryonic acoustic oscillations*.

There are two important consequences of this. The immediate consequence is that dark matter has a head start in structure formation, with density perturbations starting to grow at $z \approx 3400$. By the time the baryons decouple at $z = 1100$, there are already well-established gravitational wells which act as seeds, expediting the formation of the baryonic structures that we call galaxies.

The second consequence is more subtle. At recombination, the photons stream away from the sound waves they have helped create. But the baryons are frozen in place, a remnant of this earlier time. The sound waves contain regions in which the baryons are more compressed, and regions in which they are more rarified, with the wavelength determined by the horizon at decoupling,

$$d_H \sim \frac{cH_0^{-1}}{(1+z)^{3/2}} \approx 0.1 \text{ Mpc}$$

using $cH_0^{-1} \approx 4 \times 10^3$ Mpc and $z \approx 1100$. In the subsequent evolution of the universe, these waves were stretched by a factor of $z \approx 1100$, leaving a faint imprint on the clustering of matter seen today, where there is an excess in galaxies separated by a distance

~ 150 Mpc. The effects of these baryonic acoustic oscillations in the distribution of galaxies was first observed in 2005; the correlation function is shown in Figure 37[13].

3.2.5 Window Functions and Mass Distribution

In this section, we've understood some of the mathematical properties of $\delta(\mathbf{x}, t)$. But, so far, we've not actually discussed how one might go about measuring such an object. And, as we now explain, there is a small subtlety.

Recall that $\delta(\mathbf{x})$ is a density contrast. But a density is, of course, energy per unit volume. Mathematically, there is no difficulty in defining the density at a point $\mathbf{x}$. But how do we construct $\delta(\mathbf{x})$ from observations? In particular, what volume do we divide by?!

At heart, this comes back to our initial discussion of the cosmological principle. If we observe many galaxies, each localised at some point $\mathbf{X}_i$, then the universe looks far from homogeneous. The same is true for any fluid if we look closely enough. But our interest is in a more coarse-grained description.

To this end, we introduce a *window function* which we denote as $W(\mathbf{x}; R)$. The purpose of this function is to provide a way to turn the observed density $\delta(\mathbf{x})$ into something that is smooth, and varies on length scales $\sim R$. We construct the smoothed density contrast as

$$\delta(\mathbf{x}; R) = \int d^3x' \; W(\mathbf{x} - \mathbf{x}'; R)\, \delta(\mathbf{x}') \tag{3.57}$$

In Fourier space, we have

$$\begin{aligned}
\delta(\mathbf{k}; R) &= \int d^3x \; e^{i\mathbf{k}\cdot\mathbf{x}}\, \delta(\mathbf{x}) \\
&= \int d^3x\, d^3x' \; e^{i\mathbf{k}\cdot\mathbf{x}}\, W(\mathbf{x} - \mathbf{x}'; R)\, \delta(\mathbf{x}') \\
&= \int d^3x\, d^3x' \; e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\, W(\mathbf{x} - \mathbf{x}'; R)\, e^{i\mathbf{k}\cdot\mathbf{x}'} \delta(\mathbf{x}') \\
&= \int d^3y\, d^3x' \; e^{i\mathbf{k}\cdot\mathbf{y}}\, W(\mathbf{y}; R)\, e^{i\mathbf{k}\cdot\mathbf{x}'} \delta(\mathbf{x}') \\
&= \tilde{W}(\mathbf{k}; R)\, \delta(\mathbf{k})
\end{aligned}$$

This is the statement that a convolution integral, like (3.57), in real space becomes a product in Fourier space.

[13]This data is taken from D. J. Eisenstein *et al.* [SDSS Collaboration], "*Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies,*" Astrophys. J. **633**, 560 (2005), astro-ph/0501171.

There is no canonical choice of window function. But there are sensible choices. These include:

- The Spherical Top Hat. This is a sharp cut-off in real space, given by

$$W(\mathbf{x};R) = \frac{1}{V} \times \begin{cases} 1 & |\mathbf{x}| \leq R \\ 0 & |\mathbf{x}| > R \end{cases} \quad \text{with } V = \frac{4\pi}{3}R^3$$

  In Fourier space, this becomes

$$\tilde{W}(\mathbf{k};R) = \frac{3}{(kR)^3}\Big[\sin kR - kR\cos kR\Big] \tag{3.58}$$

  Note that the Fourier transform $\tilde{W}(\mathbf{k};R) = \tilde{W}(kR)$; this will be true of all our window functions.

- The Gaussian: This provides a smooth cut-off in both position and momentum space,

$$W(\mathbf{x};R) = \frac{1}{(2\pi)^{3/2}R^3}\exp\left(\frac{-r^2}{2R^2}\right)$$

  which, in Fourier space, retains its Gaussian form

$$\tilde{W}(kR) = \exp\left(-\frac{k^2R^2}{2}\right)$$

- The Sharp k Filter: This is a sharp cut-off in momentum space

$$\tilde{W}(kR) = \begin{cases} 1 & kR \leq 1 \\ 0 & kR > 1 \end{cases} \tag{3.59}$$

  It looks more complicated in real space,

$$W(\mathbf{x};R) = \frac{1}{2\pi^2r^3}\left[\sin(r/R) - \frac{r}{R}\cos(r/R)\right]$$

  In contrast to the other two, this has window function has the property that it diverges logarithmically when integrated over all of space.

Note that, in each case, $\tilde{W}(kR=0) = 1$. Different window functions may be better suited to different measurements or calculations. We now provide an example.

The Mass Distribution

We now use the window function technology to address a simple question: what is the distribution of masses contained within a sphere of radius R?

For each of the window functions, we can define the average mass $M(R)$ inside a sphere of radius R. You might think that we could integrate the mass density multiplied by the window function over all space but this is problematic for the sharp k cut-off because it diverges. Instead we note that the window function has dimension 1/Volume and define

$$\bar{M}(R) = \frac{\bar{\rho}}{W(0;R)c^2} = \frac{\gamma V \bar{\rho}}{c^2} \tag{3.60}$$

where $V = 4\pi R^3/3$ is the usual volume inside a sphere and $\bar{\rho}(\mathbf{x})$ is the average energy density in the universe. Here γ is a constant that differs for each of the three window functions,

$$\gamma = \begin{cases} 1 & \text{Top Hat} \\ 3\sqrt{\pi/2} & \text{Gaussian} \\ 9\pi/2 & k \text{ Filter} \end{cases}$$

(You need to Taylor expand the sharp k filter to see that it is indeed finite at $r = 0$.)

Next, we want to look at deviations from the average. The smoothed mass distribution is related to the smoothed density contrast by

$$M(\mathbf{x};R) = \bar{M}(R)(1 + \delta(\mathbf{x};R))$$

So we can also interpret the smoothed density contrast as

$$\delta(\mathbf{x};R) = \frac{\delta M(\mathbf{x};R)}{\bar{M}(R)}$$

where $\delta M(\mathbf{x};R) = M(\mathbf{x};R) - \bar{M}(R)$. The variance in the mass distribution is then

$$\sigma^2(M) = \langle \delta^2(\mathbf{x};R) \rangle$$

This depends on both the choice of window function and, more importantly, on the scale R at which we do the smoothing. Using our definition (3.57), this is

$$\sigma^2(M) = \int d^3x'\, d^3x''\; W(\mathbf{x}-\mathbf{x}';R)W(\mathbf{x}-\mathbf{x}'';R)\, \langle \delta(\mathbf{x}')\delta(\mathbf{x}'') \rangle \tag{3.61}$$

We introduced the two-point correlation function in (3.42),

$$\xi(r) = \langle \delta(\mathbf{x}+\mathbf{r})\,\delta(\mathbf{x})\rangle = \int \frac{d^3k}{(2\pi)^3}\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, P(k)$$

where, following Section 3.2, we've written this in terms of the power spectrum $P(k)$. We then have

$$\sigma^2(M) = \int \frac{d^3k}{(2\pi)^3} \int d^3x'\, d^3x''\, W(\mathbf{x}-\mathbf{x}';R)W(\mathbf{x}-\mathbf{x}'';R)\, e^{-i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}'')}\, P(k)$$

But the integrations over spatial coordinates now conspire to turn the window functions into their Fourier transform. We're left with

$$\sigma^2(M) = \int \frac{d^3k}{(2\pi)^3}\, \tilde{W}^2(kR)\, P(k) = \frac{1}{2\pi^2}\int dk\; k^2\, \tilde{W}^2(kR)\, P(k)$$

Note that, as we smooth on smaller scales, so $kR \to 0$, we have $\tilde{W}(kR) \to 1$ and, correspondingly, $\sigma^2(R) \to \sigma^2$. This is what we would wish for a variance $\sigma^2(R)$ which is smoothed on scales R.

Now recall the power spectrum from (3.56),

$$P(k) \sim \begin{cases} k^n & k < k_{\rm eq} \\ k^{n-4} & k > k_{\rm eq} \end{cases}$$

where observations of galaxy distributions give $n \approx 0.97$. At this point, it is simplest to use the sharp k-filter window function (3.59). At the largest scales, where $P(k) \sim k^n$, we then have

$$\sigma^2(M) \sim \int_0^{1/R} dk\; k^{2+n} \sim \frac{1}{R^{3+n}} \sim \frac{1}{M^{(n+3)/3}}$$

where, in the final scaling, we've used (3.60). If we have $n < -3$, we would have increasingly large mass fluctuations on large scales. This would violate our initial assumption of the cosmological principle. Fortunately, we don't live in such a universe.

Meanwhile, on shorter scales we have $P(k) \sim k^{n-4}$. Here we have

$$\sigma^2(M) \sim \int_0^{1/R} dk\; k^{n-2} \sim \frac{1}{R^{n-1}} \sim \frac{1}{M^{(n-1)/3}}$$

For $n = 1$, this becomes logarithmic scaling.

What Cosmologists Measure

As a final aside: observational cosmologists quote the fundamental parameter

$$\sigma_8^2 := \frac{1}{2\pi^2}\int dk\ k^2\,\tilde{W}^2(kR)\,P(k) \tag{3.62}$$

Here $P(k)$ is the evolved linear power spectrum that we described in Section 3.2. Meanwhile, the window function $\tilde{W}(kR)$ is taken to be the top hat (3.58), evaluated at the scale $R = 8h^{-1}$ Mpc where galactic clusters are particularly rich. (Here $h \approx 0.7$ characterises the Hubble parameter, as defined in (1.16).) Until now, we've mostly focussed on the k-dependence of $P(k)$. The variable σ_8 characterises its overall magnitude. Larger values of σ_8 imply more fluctuations, and so structure formation started earlier. For what it's worth, the current measured value is $\sigma_8 \approx 0.8$.

3.3 Nonlinear Perturbations

So far, we have relied on perturbation theory to describe the growth of density fluctuations, working with the linearised equations. But this is only tenable when the fluctuations are small. As they grow to size $\delta\rho \approx \bar{\rho}$, or $\delta \approx 1$, perturbation theory breaks down. At this point, we must solve the full coupled equations in an expanding FRW universe. This is difficult.

There are a number of ways to proceed. At some point, we simply have to resort to difficult and challenging numerical simulations. However, there is a rather simple toy model which captures some of the relevant physics.

3.3.1 Spherical Collapse

For convenience, we will work with an the Einstein-de Sitter universe, filled only with dust, so $\Omega_m = 1$. This means that the average density is equal to the critical density, $\bar{\rho}(t) = \rho_{\rm crit}(t)$.

At some time t_i, when the average density is $\bar{\rho}_i$, we create a density perturbation. To do this, consider a spherical region of radius R_i, centred about some point which we take to be the origin. Take the matter within this region and compress it into a smaller spherical region of radius $r_i < R_i$, with constant density

$$\rho_i = \bar{\rho}_i(1 + \delta_i)$$

We will initially take δ_i to be small but, in contrast to previous sections, we won't assume that it remains small for all time. Instead, we will follow its evolution as it grows.

Between r_i and R_i, there is then a gap with no matter. The mass contained in the spherical region $r < r_i$ is

$$M_i c^2 = \frac{4\pi}{3} R_i^3 \bar{\rho}_i = \frac{4\pi}{3} r_i^3 \rho_i = \frac{4\pi}{3} r_i^3 \bar{\rho}_i (1 + \delta_i)$$

Furthermore, the total mass in the perturbation remains constant at M_i, even as all the other variables, $\bar{\rho}$, δ and the edge of the over-dense region r evolve in time.

We would like to understand how this density perturbation evolves. To do this, we can revert to the simple Newtonian argument that we used in Section 1.2.3 when first deriving the Friedmann equation. Recall that, for a spherically symmetric distribution of masses, the gravitational potential at some point r depends only on the mass contained inside r and does not depend at all on the mass outside. Consider a particle at some radius r, either inside or outside the over-dense region. The conservation of energy for this particle reads

$$\frac{1}{2}\dot{r}^2 - \frac{GM(r)}{r} = E \tag{3.63}$$

where $M(r)$ is the mass contained within the radius r and is constant: by mass conservation $M(r)$ doesn't change as r evolves. Meanwhile E is also a constant (and is identified with energy divided by the mass of a single particle).

We can now apply this formula to particles both inside and outside the over-dense region. First we look at the particles outside, with $r(t_i) \geq R_i$. For these particles, the mass $M(r)$ is the same as it was before we perturbed the distribution, so they carry on as before. But our starting point was an Einstein-de Sitter universe with critical energy density, which corresponds to $E = 0$. Integrating (3.63) gives

$$r(t) = \left(\frac{9GM(r)}{2}\right)^{1/3} t^{2/3} \qquad \text{if } r(t_i) > R_i \tag{3.64}$$

with $M(r)$ constant. This is the usual expansion of a flat, matter dominated universe. The average energy density is

$$\bar{\rho}(t) = \frac{M(r)c^2}{(4\pi/3)r^3(t)} = \frac{c^2}{6\pi G}\frac{1}{t^2} \tag{3.65}$$

which reproduces the usual time evolution of the critical energy density (1.51).

In contrast, inside the over-dense region (i.e when $r(t_i) \leq r_i$), we have $E < 0$. This means that the over-dense region acts like a universe with positive curvature (i.e. $k = +1$). The inner sphere will then behave like the closed universe we met in Section 1.3.2: it first continues to expand, before slowing and subsequently collapsing back in on itself.

We presented the solution for a closed universe in parametric form in (1.57) and (1.58); you can check that the following expressions satisfy (3.63)

$$\begin{aligned} r(\tilde{\tau}) &= A\,(1-\cos\tilde{\tau}) \\ t(\tilde{\tau}) &= B\,(\tilde{\tau}-\sin\tilde{\tau}) \end{aligned} \tag{3.66}$$

where the constants are

$$A = \frac{GM}{2|E|} \quad \text{and} \quad B = \frac{GM}{(2|E|)^{3/2}} \quad \Rightarrow \quad A^3 = GMB^2 \tag{3.67}$$

We can apply the solution (3.66) to the edge of the over-dense region, i.e. the point with $r(t_i) = r_i$. We see that the spatial extent of the perturbation continues to grow for some time, swept along by the expansion of the universe. At early times $\tilde{\tau} \ll 1$, we can linearise the solution to find

$$r(\tilde{\tau}) \approx \frac{1}{2}A\tilde{\tau}^2 \quad \text{and} \quad t(\tilde{\tau}) \approx \frac{1}{6}B\tilde{\tau}^3 \quad \Rightarrow \quad r(t) \approx \frac{A}{2}\left(\frac{6}{B}\right)^{2/3} t^{2/3} \tag{3.68}$$

Thus, initially, the growth of the over-dense region has the same time dependence as the region outside the shell (3.64).

However, the excess mass in the over-dense region causes the expansion to slow. From (3.66), we see that the expansion halts and then starts to collapse again at time $\tilde{\tau}_{\rm turn} = \pi$. This is the turn-around time.

Taken at face value, the solution (3.66) then collapses back to a point at the time $\tilde{\tau}_{\rm col} = 2\pi$. We will discuss what really happens here in Section 3.3.2.

The Density in Spherical Collapse

From the solution (3.66), it is straightforward to figure out how the density evolves. At a given time, the density of the over-dense region is

$$\rho(\tilde{\tau}) = \frac{M_i c^2}{(4\pi/3)r^3} = \frac{3M_i c^2}{4\pi A^3}\frac{1}{(1-\cos\tilde{\tau})^3}$$

Meanwhile, the critical density evolves as (3.65)

$$\bar{\rho}(\tilde{\tau}) = \frac{c^2}{6\pi G}\frac{1}{t^2} = \frac{c^2}{6\pi GB^2}\frac{1}{(\tilde{\tau}-\sin\tilde{\tau})^2}$$

The density contrast $\delta = \delta\rho/\bar{\rho}$ can be computed from the ratio of the two,

$$(1+\delta) = \frac{\rho}{\bar{\rho}} = \frac{9}{2}\frac{(\tilde{\tau}-\sin\tilde{\tau})^2}{(1-\cos\tilde{\tau})^3} \tag{3.69}$$

where we've used the fact that $A^3 = GMB^2$.

Again, we can see what happens at early times. We Taylor expand each of the terms, but this time we need to go to second order: $\tilde{\tau} - \sin\tau \approx \frac{1}{3!}\tilde{\tau}^3 - \frac{1}{5!}\tilde{\tau}^5$ and $1 - \cos\tilde{\tau} \approx \frac{1}{2}\tilde{\tau}^2 - \frac{1}{4!}\tilde{\tau}^4$. This gives

$$1 + \delta_{\rm lin}(\tilde{\tau}) \approx \frac{(1 - \frac{1}{20}\tilde{\tau}^2)^2}{(1 - \frac{1}{12}\tilde{\tau}^2)^3} \approx 1 + \frac{3}{20}\tilde{\tau}^2 \tag{3.70}$$

But, from (3.68), we can write this as

$$\delta_{\rm lin}(t) = \frac{3}{20}\left(\frac{6}{B}\right)^{2/3} t^{2/3} \tag{3.71}$$

Happily, this coincides with the $t^{2/3}$ time dependence that we found in (3.31) when discussing linear perturbation theory.

When we reach turn-around, at $\tilde{\tau} = \pi$, the density is

$$\delta(\tilde{\tau}_{\rm turn}) = \frac{9\pi^2}{16} - 1 \approx 4.55$$

For what follows, it will prove useful to ask the following, slightly artificial question: what would the density contrast be at turn-around if we were to extrapolate the linear solution? From (3.66), we have $t_{\rm turn} = B\pi$, so we can write the linear solution (3.71) as

$$\delta_{\rm lin}(t) = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_{\rm turn}}\right)^{2/3} \quad \Rightarrow \quad \delta_{\rm lin}(t_{\rm turn}) = \frac{3}{20}(6\pi)^{2/3} \approx 1.06$$

Meanwhile, when the perturbation has completely collapsed at $\tilde{\tau}_{\rm col} = 2\pi$, the true density is

$$\delta(\tilde{\tau}_{\rm col}) = \infty$$

and we'll see how to interpret this shortly. We can again ask the artificial question: what would the density contrast be at collapse if we were to extrapolate the linear solution. This time, from (3.66), we have $t_{\rm col} = 2B\pi = 2t_{\rm turn}$, so

$$\delta_{\rm lin}(t_{\rm col}) = \frac{3}{20}(12\pi)^{2/3} \approx 1.69$$

A simplistic interpretation of this result is as follows: if we work within linear perturbation theory, and the density contrast reaches $\delta_{\rm lin} \approx 1.69$, then we should interpret this as a complete collapse.

3.3.2 Virialisation and Dark Matter Halos

As we have seen, the simple spherical collapse model predicts that an initial over-density will ultimately collapse down to a point with infinite density. The interpretation of such a singularity is a black hole.

Yet our universe is not dominated by black holes. This is because the assumption of spherical collapse is not particularly realistic, and while this is not too much of a problem for much of the discussion, it becomes important as the end point nears. Here, the random motion of the matter, together with interactions, means that the matter will ultimately settle down into an equilibrium configuration with the kinetic energy balanced by the potential energy. The end result is a *dark matter halo*, an extended region of dark matter in which galaxies are embedded.

This process in which equilibrium is reached is known, rather wonderfully, as *violent relaxation*. Or, less evocatively, as *virialisation*. This latter name reflects the fact that by the time the system has settled down, it obeys the virial theorem, with the average kinetic energy T related to the average potential energy V by

$$\bar{T} = -\frac{1}{2}\bar{V}$$

We proved this theorem in Section 1.4.3.

Let's now apply this to our collapse model. Our original formula (3.63) is conveniently written in terms of the kinetic energy $T = \frac{1}{2}\dot{r}^2$ and the potential energy $V = -GM/r$. We can start by considering the turn-around point, where the kinetic energy vanishes, $T = 0$, and

$$V_{\rm turn} = -\frac{GM}{r_{\rm turn}}$$

The total energy $E = T + V$ is conserved. This means that after virialisation, when $T = -\frac{1}{2}V$, we must have

$$T_{\rm vir} + V_{\rm vir} = \frac{1}{2}V_{\rm vir} = V_{\rm turn} \quad \Rightarrow \quad \begin{cases} r_{\rm vir} = \frac{1}{2}r_{\rm turn} \\ \rho_{\rm vir} = 8\rho_{\rm turn} \end{cases}$$

Our real interest is in the density contrast, $1 + \delta_{\rm vir} = \rho_{\rm vir}/\bar{\rho}_{\rm vir}$. We take the virialisation time to coincide with the collapse time, $t_{\rm vir} = t_{\rm col} = 2t_{\rm turn}$. Since the universe scales as $a \sim t^{2/3}$, the critical energy has diluted by a factor of 4 between turn-around and virialisation, so $\bar{\rho}_{\rm vir} = \bar{\rho}_{\rm turn}/4$. Putting this together, we have

$$\delta_{\rm vir} = \frac{\rho_{\rm vir}}{\bar{\rho}_{\rm vir}} - 1 = 32\,\frac{\rho_{\rm turn}}{\bar{\rho}_{\rm turn}} - 1$$

But from (3.69), using $\tau_{\rm turn} = \pi$, we have $\rho_{\rm turn}/\bar{\rho}_{\rm turn} = 9\pi^2/16$. The upshot is that the density contrast in a dark matter halo is expected to be

$$\delta_{\rm vir} = 18\pi^2 - 1 \approx 177$$

Once again referring to our linear model, we learn that whenever $\delta_{\rm lin} \gtrsim 1.69$, we may expect to form a dark matter halo whose density ρ is roughly 200 times greater than the background density $\bar{\rho}$.

3.3.3 Why the Universe Wouldn't be Home Without Dark Matter

We can try to put together some of the statements that we have seen so far to get a sense for when structures form.

The right way to do this is to use the window function that we introduced in Section 3.2.5, to define spatial variations smoothed on different scales R. The spatial variations are computed by integrating the power spectrum against the window function, as in (3.61). We can then trace the evolution of these spatial perturbations to see how they evolve.

Here, instead, we're going to do a quick and dirty calculation to get some sense of the time scale. Indeed, taken at face value, there seems to be a problem. The CMB tells us that $\delta T/T \sim 10^{-5}$ at redshift $z \approx 1000$. Yet we know that, in the matter dominated era, perturbations grow linearly with scale (3.31). This would naively suggest that, even today, we have only $\delta \sim 10^{-2}$ which, given our discussion above, is not enough for structures to form. What's going on?

In large part, this issue arises because we need to do a better job of defining the spatial variations. But there is also some important physics buried in this simple observation which we mentioned briefly before, but is worth highlighting. The CMB figure of $\delta T/T \sim 10^{-5}$ is telling us about the fluctuations in radiation and, through this, fluctuations in baryonic matter at recombination. This is not sufficient for galaxies to form. To get the universe we see today, it's necessary to have dark matter. Between $z \approx 3000$ and $z \approx 1000$, when the universe was matter dominated, perturbations in dark matter were growing while the baryon-photon fluid was sloshing back and forth. This can be further enhanced by the logarithmic growth (3.36) of dark matter perturbations during the radiation dominated era.

Even accounting for dark matter, it's not obvious, using our results above, that there is enough time for structures to form. Fortunately, there are a bunch of scrappy factors floating around which get us close to the right ballpark. For example, the fluctuations

in matter density are related to those in temperature by $\delta_m \approx 3 \times \delta T/T$. (We will see this in (3.73).) Furthermore, we should focus on the peaks of the fluctuations rather than the average: these come in around $\delta T/T \approx 6 \times 10^{-5}$. The Sachs-Wolfe effect (which we will describe in Section 3.4 provides another small boost. All told, these factors conspire to give $\delta_m \approx 10^{-3}$ at $z \approx 1000$. This tells us that we expect dark matter halos to form at redshift $z \approx 1$ which is roughly right.

However, an important take-home message is that the existence of dark matter, which is decoupled from the photon fluid and so starts to grow as soon as the universe is matter dominated, is crucial for structure to form on a viable time scale.

3.3.4 The Cosmological Constant Revisited

We can repeat the argument above in the presence of a cosmological constant. We saw in (1.60) that the cosmological constant changes the equation (3.63), describing the radial motion of a particle, to include a term that looks like an inverted harmonic oscillator

$$\frac{1}{2}\dot{r}^2 - \frac{GM(r)}{r} - \frac{1}{6}\Lambda r^2 = E \tag{3.72}$$

Let's now play our earlier game. We start with a universe comprising of both matter and a cosmological constant with critical density, so that $E = 0$.

Now we create an over-density by squeezing the sphere at $r = R_i$ to a smaller radius, $r = r_i$. Particles with $r(t_i) < r_i$ have negative energy $E < 0$. If, as previously, this over-dense region is to turn around and subsequently collapse then there must be a time when $\dot{r} = 0$ and $r(t)$ solves the cubic equation

$$\frac{1}{6}\Lambda r^3(t) - |E|r(t) + GM = 0$$

with M the constant mass contained in the over-dense region. We want to know if this equation has a solution with $r(t) > 0$?

To answer this, first note that the cubic has stationary points at $r = \pm\sqrt{2|E|/\Lambda}$. The cubic only has a root with $r > 0$ if the positive stationary point lies below the real axis, or

$$\frac{1}{6}\Lambda\left(\frac{2|E|}{\Lambda}\right)^{3/2} - |E|\left(\frac{2|E|}{\Lambda}\right)^{1/2} + GM < 0 \quad \Rightarrow \quad \Lambda^{1/2} < \frac{(2|E|)^{3/2}}{3GM}$$

We write this upper bound on Λ as

$$\Lambda^{1/2} < \frac{1}{3B}$$

where $B = GM/(2|E|)^{3/2}$ was defined previously in (3.67). We need to relate this constant B to the initial density perturbation. For this, note that if we make the density perturbation at early times, then the cosmological constant is negligible and the universe evolves as if it is matter dominated. In this case, we can use our earlier result (3.71)

$$\delta(t) = \frac{3}{20}\left(\frac{6}{B}\right)^{2/3} t^{2/3}$$

Using this to eliminate B, and evaluating the various constants, we have an upper bound on Λ

$$\Lambda^{1/2} \lesssim 0.1\,\frac{\delta^{3/2}}{t}$$

Note that $\delta^{3/2}/t$ is the combination which, in linear perturbation theory, stays constant in the matter dominated era as seen in (3.31). We see that if we want gravitational collapse to occur and galaxies to form (which, let's face it, would be nice) then there is an upper bound on the cosmological constant Λ, which depends on the strength of the initial perturbations.

What is this bound for our universe? It's a bit tricky to get an accurate statement using the information that we have gathered so far in this course, but we can get a ball-park figure. We argued in Section 3.3.3 that it is sensible to take $\delta_m \sim 10^{-3}$ at $z \approx 1000$, which is roughly the time of last scattering $t_{\rm last} \approx 350,000$ years $\approx 10^{13}$ s. This gives an upper bound on the cosmological constant of

$$\Lambda \lesssim 10^{-37}\ {\rm s}^{-2}$$

and a corresponding bound on the vacuum energy of

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} = \frac{M_{\rm pl}^2 c^4 \Lambda}{\hbar c^3} \approx (10^{47}\Lambda)\ {\rm eV\,m^{-3}s^2} \lesssim 10^{10}\ {\rm eV\,m^{-3}}$$

This is only a factor of 10 higher than the observed value of $\rho_\Lambda \approx 10^9$ eV m^{-3}! Although the calculation above involved quite a lot of hand-waving and order-of-magnitude estimates, the conclusion is the right one[14]: if the cosmological constant were much larger than we observe today, then galaxies would not have formed. We are, it appears, living on the edge.

[14]A better version of this calculation models the size of density perturbations using the σ_8 variable defined in (3.62), and takes into account the non-vanishing radiation contribution to the energy density in the early universe. Some of this discussion can be found in the original paper by Weinberg "*Anthropic Bound on the Cosmological Constant*" in Physical Review Letters vol 59 (1987).

3.4 The Cosmic Microwave Background

The cosmic microwave background (CMB) provides the snapshot of the early universe. In section 2.2, we described the how the CMB is an almost perfect blackbody. At temperature $T \approx 2.73$ K. However, there are small fluctuations in the CMB, with magnitude

$$\frac{\delta T}{T} \approx 10^{-5}$$

We already mentioned this at the very start of these lectures as evidence that the early universe was homogeneous and isotropic. As we now explain, these temperature fluctuations contain a near-perfect imprint of the anisotropies at the time of recombination. Moreover, we can trace the fate of these perturbations back in time to get another handle on the primordial power spectrum.

In Section 3.2.1, we stated that the perturbations in the early universe were adiabatic, meaning that perturbations in all fluids are proportional. In particular, the density perturbations in matter and radiation are related by

$$\delta_r = \frac{4}{3}\delta_m$$

It is more convenient to express this in terms of the temperature of the CMB. From our discussion of blackbody radiation, we know that $\rho_r \sim T^4$, so

$$\delta_r = \frac{\delta\rho_r}{\rho_r} = 4\frac{\delta T}{T} \quad \Rightarrow \quad \frac{\delta T}{T} = \frac{1}{3}\delta_m \tag{3.73}$$

We might, therefore expect that temperature fluctuations of the CMB contain a direct imprint of the matter fluctuations in the early universe. In fact, there is a subtlety which means that this is not quite true.

3.4.1 Gravitational Red-Shift

The new physics is gravitational redshift. This is an effect that arises from general relativity. Here we just give a heuristic sketch of the basic idea.

As a warm-up, first consider throwing a particle from the Earth upwards into space. We know that it must lose kinetic energy to escape the Earth's gravitational potential $\Phi = -GM/R$.

What happens if we do the same for light? Clearly light can't slow down, but it does lose energy. This manifests itself in a reduction in the frequency of the light, or a stretching of the wavelength. In other words, the light is redshifted. In the Newtonian limit, this redshift is

$$\frac{\delta\lambda}{\lambda} = -\frac{\Phi}{c^2} \tag{3.74}$$

Now consider a spatially varying gravitational potential $\delta\Phi(\mathbf{x})$ of the kind that permeates the early universe. To reach us, the photons from any point in space $\mathbf{x}$ will have to climb out of the gravitational potential and will be redshifted. This, in turn, shifts the temperature of the CMB. A straightforward generalisation of (3.74) suggests

$$\frac{\delta T(\hat{\mathbf{n}})}{T} = \frac{\delta\Phi(\mathbf{x}_{\rm last})}{c^2}$$

where $\mathbf{x}_{\rm last} = |\mathbf{x}_{\rm last}|\hat{\mathbf{n}}$ sits on the surface of last scattering, where the CMB was formed. In fact, this too misses an important piece of physics. The slight increase in $\delta\Phi$ results in a slight change in the local expansion rate of the universe which, since the CMB forms in the matter dominated era, scales as $a(t) \sim t^{2/3}$. This is known as the *Sachs-Wolfe effect*. It turns out that this gives an extra contribution of $-\frac{2}{3}\Phi/c^2$. This means that the temperature fluctuation in the CMB is related to the gravitational perturbation by

$$\frac{\delta T(\hat{\mathbf{n}})}{T} = \frac{\delta\Phi(\mathbf{x}_{\rm last})}{3c^2} \tag{3.75}$$

We learn that there are two, competing contributions to the temperature fluctuations in the CMB: the initial adiabatic perturbation (3.73) and the gravitational perturbation leading to the redshift (3.75). The question is: which is bigger?

The two contributions are not independent. They are related by the Poisson equation (3.51),

$$\delta\Phi(\mathbf{k}) = -\frac{4\pi G}{c^2k^2}\bar{\rho}a^2\delta_m(\mathbf{k}) \tag{3.76}$$

We see that the redshift contribution dominates for large wavelengths (k small) while the adiabatic contribution dominates for small wavelengths (k large). The cross-over happens at the critical value of k

$$k_{\rm crit}^2 \sim \frac{4\pi G}{c^4}\bar{\rho}a^2 \quad \Rightarrow \quad k_{\rm crit} \sim \frac{aH}{c}$$

But we recognise this as the size of the co-moving horizon. This means that modes that are were outside the horizon at last scattering will be dominated by the redshift and the Sachs-Wolfe effect; those which were inside the horizon at last scattering will exhibit the matter power spectrum.

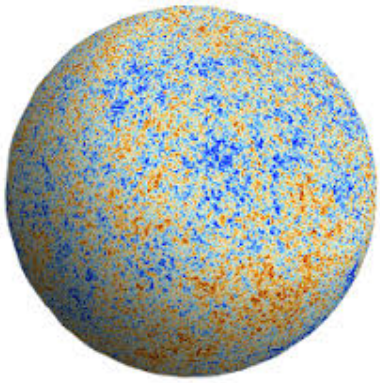

Figure 38: The CMB in its natural setting.

3.4.2 The CMB Power Spectrum

We don't have a three-dimensional map of the microwave background. Instead, the famous picture of the CMB lives on a sphere which surrounds us, as shown in the figure. This is clear in (3.75), where the temperature fluctuations depends only on the direction $\hat{\mathbf{n}}$.

We introduce spherical polar coordinates, and label the direction $\hat{\mathbf{n}}$ by the usual angles θ and ϕ. We then expand the temperature fluctuation in spherical polar coordinates as

$$\frac{\delta T(\hat{\mathbf{n}})}{T} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{l,m}\, Y_{l,m}(\theta, \phi)$$

Here $Y_{l,m}(\theta, \phi)$ are spherical harmonics, given by

$$Y_{l,m}(\theta, \phi) = N_{l,m} e^{im\phi} P_l^m(\cos\theta)$$

with $P_l^m(\cos\theta)$ the associated Legendre polynomial and $N_{l,m}$ an appropriate normalisation. Shortly, we will need $N_{l,0} = (2l+1)/4\pi$.

The measured coefficients $a_{l,m}$ the temperature anisotropies at different angular separation. Small l corresponds to large angles on the sky. We will now relate these to the primordial power spectrum $P(k)$.

As in the previous section, we are interested in correlations in the temperature fluctuations. The temperature two-point correlation function boils down to understanding the spatial average of

$$\langle a_{l,m}\, a^{\star}_{l',m'} \rangle = C_l\, \delta_{l,l'} \delta_{m,m'}$$

where statistical rotational invariance ensures that the average depends only on the angular momentum label l, and not on m. The coefficients C_l are called *multipole moments*.

The temperature correlation function can be written in terms of C_l. We pick spherical polar coordinates such that $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos\theta$. Using θ and ϕ. Using $P_l^0(1) = 1$ and $P_l^m(1) = 0$ for $m \neq 0$, we then have

$$\begin{aligned}\frac{\langle \delta T(\hat{\mathbf{n}}) \delta T(\hat{\mathbf{n}}')\rangle}{T^2} &= \sum_{l,m}\sum_{l'm'} \langle a_{l,m} a_{l'm'}\rangle \; Y_{l,0}(\theta,\phi) \\ &= \sum_l \frac{2l+1}{4\pi} C_l \, P_l(\cos\theta)\end{aligned}$$

We would like to relate these coefficients C_l to the power spectrum. We will focus on large scales, with small l, where, as discussed above, we expect the temperature fluctuations to be dominated by the Sachs-Wolfe effect (3.75). In practice, this holds for $l \lesssim 50$.

It is a straightforward, if somewhat fiddly, exercise to write C_l in terms of the gravitational power spectrum (3.52).

$$\langle \delta\Phi(\mathbf{k}) \, \delta\Phi(\mathbf{k}')\rangle = (2\pi)^3 \, \delta_D^3(\mathbf{k}+\mathbf{k}') \, P_\Phi(k)$$

We do not give all the details here. (See, for example, the book by Weinberg.) After decomposing the Fourier mode $\delta\Phi(\mathbf{k})$ in spherical harmonics, one finds that the coefficients of the two-point function can be written as

$$C_l = \frac{16\pi T^2}{9} \int dk \; k^2 P_\Phi(k) j_l^2(kr)$$

with $j_l(kr)$ a spherical Bessel function. The primordial gravitational power spectrum takes the form (3.53)

$$P_\Phi(k) \sim k^{n-4}$$

which differs by a power of k^{-4} compared to the matter power spectrum, a fact which follows from the relation (3.76). For the Harrison-Zel'dovich spectrum, $n = 1$, one then finds

$$C_l \sim \frac{1}{l(l+1)}$$

It remains to compare this to the observed CMB power spectrum.

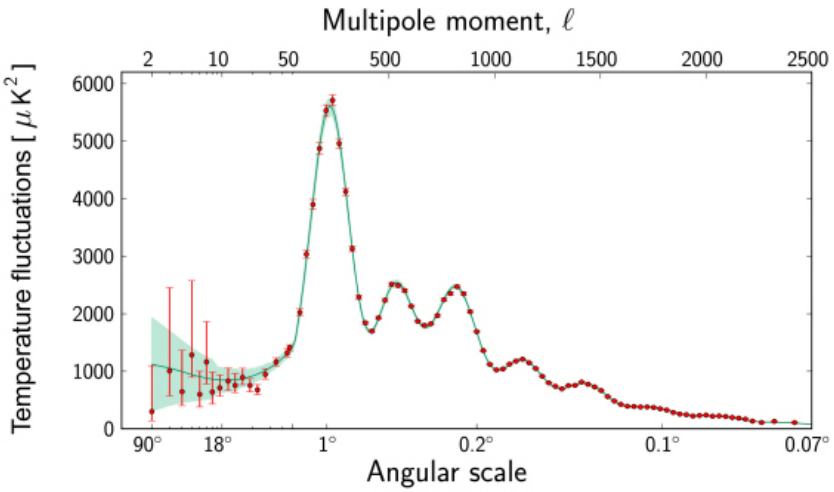

Figure 39: The CMB power spectrum measured by Planck. The combination $l(l+1)C_l$ is plotted on the vertical axis.

3.4.3 A Very Brief Introduction to CMB Physics

There has been an enormous effort, over many decades, to accurately measure the fluctuation coefficients C_l. The results from the Planck satellite are shown in Figure 39, with the combination $l(l+1)C_l$ plotted on the vertical axis; the red dots are data, shown with error bars, while the green line is the best theoretical fit.

The power spectrum exhibits a distinctive pattern of peaks and troughs. These are again a remnant of the acoustic oscillations in the early universe. A quantitative understanding of how these arise is somewhat beyond what this course. (You can learn more next year in Part III.) Here we give just a taster:

- At low l, the temperature fluctuations have the advertised scale $\delta T/T \approx 10^{-5}$. Here the plot is roughly constant. This confirms that the CMB is close to the Harrison-Zel'dovich spectrum, with $C_l \sim 1/l(l+1)$, as expected. In fact, a detailed analysis gives

  $$n \approx 0.97$$

  in good agreement with the measurements from galaxy distributions.

- The first peak sits at $l \approx 200$ and sets the characteristic angular scales of fluctuations that one can see by eye in the CMB maps. At this point, the fluctuations have risen to $\delta T/T \approx 6 \times 10^{-5}$.

This peak arises from an acoustic wave that had time to undergo just a single compression before decoupling. This is the same physics that led to the baryon acoustic peak shown in Figure 37. The angular size in the sky is determined both by the horizon at decoupling (usually referred to as the *sound horizon*) and the subsequent expansion history of the universe. In particular, its angular value is very sensitive to the curvature of the universe. The location of this first peak is our best evidence that the universe is very close to flat (or $k = 0$ in the language of Section 1.)

Given the observed fact that the matter and radiation in the universe sits well below the critical value, the position of the first peak also provides corroborating evidence for dark energy.

- The second and third peaks contain information about the amount of baryonic and dark matter in the early universe. This is because the amplitudes of successive oscillations depends on both the baryon-to-photon ratio in the plasma, and the gravitational potentials created by dark matter.

- The microwave background doesn't just contain information from the temperature anisotropies. One can also extract information from the polarisation of the photons. These are two kinds of polarisation pattern, known as *E-modes* and *B-modes*.

  The E-mode polarisation has been measured and is found to be correlated with the temperature anisotropies. Interestingly, these correlations (really anti-correlations) extend down below $l < 200$. This is important because modes of this size were outside the horizon at the time the CMB was formed. Such correlations could only arise if there was some causal interaction between the modes, pointing clearly to the need for a period of inflation in the very early universe.

  B-modes in the CMB have been found but, somewhat disappointingly, arise because of contamination due to interstellar dust. A discovery of primordial B-modes would be *extremely* exciting since they are thought to be generated by gravitational waves, created by quantum effects at play during in inflation. The observation of primordial B-modes imprinted in the CMB would provide our first experimental window into quantum gravity!

- For very low $l \lesssim 10$, there are both large error bars and poor agreement with the theoretical expectations. The large error bars arise because we only have one sky to observe and only a handful of independent observables, with $-l \leq m \leq l$. This

issue is known as *cosmic variance*. It makes it difficult to know if the disagreement with theory is telling us something deep, or is just random chance.

3.5 Inflation Revisited

> "With the new cosmology the universe must have started off in some very simple way. What, then, becomes of the initial conditions required by dynamical theory? Plainly there cannot be any, or they must be trivial. We are left in a situation which would be untenable with the old mechanics. If the universe were simply the motion which follow from a given scheme of equations of motion with trivial initial conditions, it could not contain the complexity we observe. Quantum mechanics provides an escape from the difficulty. It enables us to ascribe the complexity to the quantum jumps, lying outside the scheme of equations of motion."
>
> A very prescient Paul Dirac, in 1939

Until now, we have only focussed only on the evolution of some initial density perturbations that were mysteriously laid down in the very early universe. The obvious question is: where did these perturbations come from in the first place?

There is an astonishing answer to this question. The density perturbations are quantum fluctuations from the very first moment after the Big Bang, fluctuations which were caught in the act and subsequently stretched to cosmological scales by the rapid expansion of the universe during inflation, where they laid the seeds for the formation of galaxies and other structures that we see around us.

This idea that the origin of the largest objects in the universe can be traced back to quantum fluctuations taking place at the very earliest times is nothing short of awe-inspiring. Yet, as we will see, the process of inflation generates perturbations on a super-horizon scale. These perturbations are adiabatic, Gaussian and with a power spectrum $P(k) \sim k^n$ with $n \approx 1$. In other words, the perturbations are exactly of the form required to describe our universe.

3.5.1 Superhorizon Perturbations

Before we get to the nitty gritty, let's first understand why inflation provides a very natural environment in which to create perturbations which, subsequently, have wavelength greater than the apparent horizon. During inflation, the universe undergoes an accelerated expansion (1.90) which, for simplicity, we approximate as an exponential de Sitter phase,

$$a(t) = a(0) \exp\left(H_{\rm inf} t\right)$$

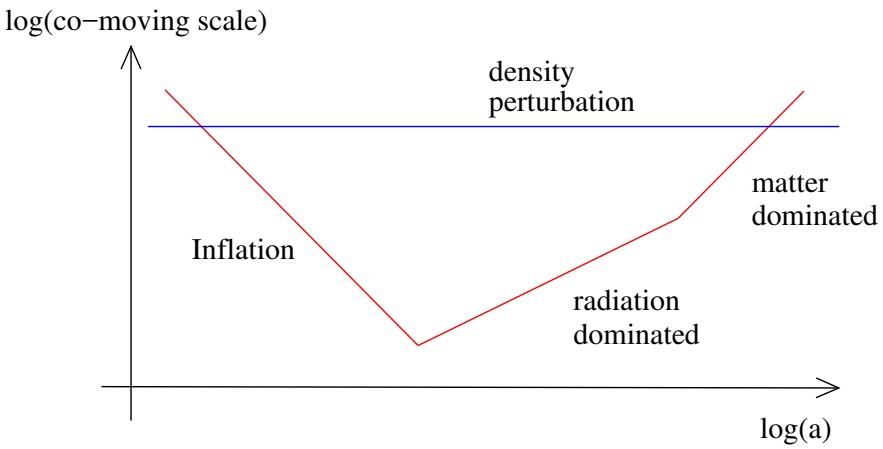

Figure 40: The density perturbations are created during inflation and exit the co-moving horizon, shown in red. Then they wait. Later, during the hot Big Bang phase of radiation or matter domination, the co-moving horizon expands and the density perturbations re-enter where we see them today.

The key observation is that, in an accelerating phase of this type, the co-moving horizon is shrinking,

$$\chi_H = \frac{c}{aH_{\rm inf}} \tag{3.77}$$

Focussing on the co-moving horizon (rather than the physical horizon) gives us a view of inflation in which we zoom into some small patch of space, which subsequently becomes our entire universe.

Any perturbation created during inflation with co-moving wavevector $\mathbf{k}$ will rapidly move outside the horizon, where they linger until the expansion of the universe slows to a more sedentary pace, after which the co-moving horizon expands, as in (3.38), and the perturbations created during inflation can now re-enter. This is shown in Figure 40. In this way, inflation can naturally generate superhorizon perturbations that seem to be needed to explain the universe we see around us. This picture also makes it clear that the longer wavelength perturbations must have been created earlier in the universe's past.

3.5.2 Classical Inflationary Perturbations

It remains for us to explain how these density perturbations arose in the first place. A full discussion requires both quantum field theory and general relativity. Here we give the essence of the idea.

Recall that inflation requires the introduction of a new degree of freedom, the *inflaton* scalar field with action (1.82),

$$S = \int d^3x\, dt\; a^3(t) \left[\frac{1}{2}\dot{\phi}^2 - \frac{c^2}{2a^2(t)} \nabla\phi \cdot \nabla\phi - V(\phi) \right]$$

The scalar field ϕ rolls from some initial starting point, high up on the potential, and in doing so, drives inflation. In this process, ϕ also undergoes quantum fluctuations; these will be the seeds for density perturbations.

We start by looking at a stripped down version of this story. We will take the potential $V(\phi) =$ constant, which is the same thing as a cosmological constant. This ensures that the universe sits in a de Sitter phase with $a(t) \sim e^{H_{\rm inf} t}$. We then look at the dynamics of ϕ in this background. The classical equation of motion is

$$\frac{d^2\phi}{dt^2} + 3H_{\rm inf}\frac{d\phi}{dt} - \frac{c^2}{a^2}\nabla^2\phi = 0 \tag{3.78}$$

Ultimately, we want to treat $\phi(\mathbf{x}, t)$ as a quantum variable. To do this, we will massage the equation of motion in various ways until it looks like something more familiar. First, we decompose the spatial variation of $\phi(\mathbf{x}, t)$ in Fourier modes,

$$\phi(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3}\; e^{-i\mathbf{k}\cdot\mathbf{x}}\, \phi_{\mathbf{k}}(t)$$

The reality of $\phi(\mathbf{x}, t)$ means that we must have $\phi^\star_{\mathbf{k}} = \phi_{-\mathbf{k}}$. The equation of motion (3.78) then becomes decoupled equations for each $\phi_{\mathbf{k}}$,

$$\frac{d^2\phi_{\mathbf{k}}}{dt^2} + 3H_{\rm inf}\frac{d\phi_{\mathbf{k}}}{dt} + \frac{c^2k^2}{a^2}\phi_{\mathbf{k}} = 0 \tag{3.79}$$

This equation takes the form of a damped harmonic oscillator, with some time dependence hiding in the $1/a^2$ part of the final term. A time dependent frequency is something we can deal with in quantum mechanics, but friction is not. For this reason, we want to make a further change of variables that gets rid of the damping term proportional to $\dot{\phi}_{\mathbf{k}}$. To achieve this, we work in conformal time (1.26)

$$\tau = \int^t \frac{dt'}{a(t')} = -\frac{1}{aH_{\rm inf}}$$

Note that, for a de Sitter universe, conformal time sits in the range $\tau \in (-\infty, 0)$ so $\tau \to 0^-$ is the far future. We then have

$$\frac{d^2\phi}{dt^2} = \frac{1}{a^2}\frac{d^2\phi}{d\tau^2} - \frac{H}{a}\frac{d\phi}{d\tau} \quad \text{and} \quad \frac{d\phi}{dt} = \frac{1}{a}\frac{d\phi}{d\tau}$$

and the equation of motion (3.79) becomes an equation for $\phi_{\mathbf{k}}(\tau)$,

$$\frac{d^2\phi_{\mathbf{k}}}{d\tau^2} - \frac{2}{\tau}\frac{d\phi_{\mathbf{k}}}{d\tau} + c^2k^2\phi_{\mathbf{k}} = 0$$

This doesn't seem to have done much good, simply changing the coefficient of the damping term. But things start looking rosier if we define

$$\tilde{\phi}_{\mathbf{k}} = -\frac{1}{H_{\rm inf}\tau}\,\phi_{\mathbf{k}} \tag{3.80}$$

Using $\dot{a} = H_{\rm inf}a$, the equation becomes

$$\frac{d^2\tilde{\phi}_{\mathbf{k}}}{d\tau^2} + \left(c^2k^2 - \frac{2}{\tau^2}\right)\tilde{\phi}_{\mathbf{k}} = 0 \tag{3.81}$$

This is the final form that we want. Each $\tilde{\phi}_{\mathbf{k}}$ obeys the equation of a harmonic oscillator, with a frequency

$$\omega_k^2 = c^2k^2 - \frac{2}{\tau^2} \tag{3.82}$$

that depends on both k and on conformal time τ. In the far past, $\tau \to -\infty$, the time-dependent $1/\tau^2$ term is negligible. However, as we move forward in time, ω^2 first goes to zero and then becomes negative, corresponding to a harmonic oscillator with an upside-down potential. The co-moving horizon (3.77) is $\chi_H = c/aH_{\rm inf} = -c\tau$. This means that, for a given perturbation $\mathbf{k}$, the wavelength $\lambda = 2\pi/k$ exits the horizon at more or less the time that the frequency of the associated harmonic oscillator is $\omega_k^2 = 0$.

It is not too difficult to write down a solution to the time-dependent harmonic oscillator (3.81). It is a second order differential equation, so we expect two linearly independent solutions. You can check that the general form is given by

$$\tilde{\phi}_{\mathbf{k}} = \alpha e^{-ick\tau}\left(1 - \frac{i}{ck\tau}\right) + \beta e^{+ick\tau}\left(1 + \frac{i}{ck\tau}\right) \tag{3.83}$$

where α and β are integration constants. In the far past, $ck\tau \to -\infty$, these modes oscillate just like a normal harmonic oscillator. But as inflation proceeds, and $ck\tau \to 0^-$, the oscillations stop. Expanding out the $e^{\pm ick\tau}$ in this limit, we find that the modes grow as $\tilde{\phi}_{\mathbf{k}} \approx (\beta - \alpha)/ck\tau$. If we then translate back to the original field $\phi_{\mathbf{k}}$ using (3.80), we find that the Fourier modes obey

$$\phi_{\mathbf{k}} = -\frac{\alpha H_{\rm inf}}{ck}e^{-ick\tau}\,(ck\tau - i) - \frac{\beta H_{\rm inf}}{ck}e^{+ick\tau}\,(ck\tau + i)$$

These modes now oscillate wildly at the beginning of inflation, $ck\tau \to -\infty$, but settle down to become constant after the mode has exited the horizon and $ck\tau \to 0^-$.

3.5.3 The Quantum Harmonic Oscillator

Our ultimate goal is to understand the quantum fluctuations of the inflaton field $\phi(\mathbf{x}, t)$. At first glance, this sounds like a daunting problem. But the analysis above shows the way forward, because each (rescaled) Fourier mode $\tilde{\phi}_{\mathbf{k}}$ obeys the equation for a simple harmonic oscillator (3.81). And we know how to quantise the harmonic oscillator. The only subtlety is that the frequency ω_k is time dependent. But this too is a problem that we can address purely within quantum mechanics.

A Review of the Harmonic Oscillator

Let's first review the solution to the familiar harmonic oscillator in which the frequency ω does not vary with time. The Hamiltonian is

$$\hat{H} = \frac{1}{2}\hat{p}^2 + \frac{1}{2}\omega^2\hat{q}^2$$

where we've set the usual mass $m = 1$. The position and momentum obey the canonical commutation relation

$$[\hat{q}, \hat{p}] = i\hbar$$

The slick way to solve this is to introduce annihilation and creation operators. These are defined by

$$\hat{a} = \sqrt{\frac{\omega}{2\hbar}}\hat{q} + i\sqrt{\frac{1}{2\hbar\omega}}\hat{p} \quad \text{and} \quad \hat{a}^\dagger = \sqrt{\frac{\omega}{2\hbar}}\hat{q} - i\sqrt{\frac{1}{2\hbar\omega}}\hat{p}$$

and the inverse is

$$\hat{q} = \sqrt{\frac{\hbar}{2\omega}}(\hat{a} + \hat{a}^\dagger) \quad \text{and} \quad \hat{p} = -i\sqrt{\frac{\hbar\omega}{2}}(\hat{a} - \hat{a}^\dagger) \tag{3.84}$$

You can check that these obey the commutation relations

$$[\hat{a}, \hat{a}^\dagger] = 1$$

When written in terms of annihilation and creation operators, the Hamiltonian takes the simple form

$$\hat{H} = \frac{1}{2}\hbar\omega(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}) = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right)$$

Now it is straightforward to build the energy eigenstates of the system. The ground state is written as $|0\rangle$ and obeys

$$\hat{a}|0\rangle = 0$$

Excited states then constructed by acting with $\hat{a}^\dagger$, giving

$$|n\rangle = \frac{1}{\sqrt{n!}}\hat{a}^{\dagger n}|0\rangle \quad \Rightarrow \quad \hat{H}|n\rangle = \hbar\omega\left(n+\frac{1}{2}\right)|n\rangle$$

In what follows, we will be particularly interested in the variance in the ground state $|0\rangle$. First, recall that the expectation value of $\hat{q}$ vanishes in the ground state (or, indeed, in any energy eigenstate),

$$\langle 0|\hat{q}|0\rangle = \sqrt{\frac{\hbar}{2\omega}}\langle 0|(\hat{a}+\hat{a}^\dagger)|0\rangle = 0$$

where we use the property of the ground state $\hat{a}|0\rangle = 0$ or, equivalently, $\langle 0|\hat{a}^\dagger = 0$. However, the variance is non-vanishing, and given by

$$\langle 0|\hat{q}^2|0\rangle = \frac{\hbar}{2\omega}\langle 0|(\hat{a}+\hat{a}^\dagger)^2|0\rangle = \frac{\hbar}{2\omega}\langle 0|\hat{a}^\dagger\hat{a}|0\rangle = \frac{\hbar}{2\omega}$$

We write this as

$$\langle \hat{q}^2\rangle = \frac{\hbar}{2\omega} \tag{3.85}$$

These will be the fluctuations which we will apply to the inflaton field. But first we need to see the effects of a time dependent frequency.

A Review of the Heisenberg Picture

There are two ways to think about time evolution in quantum mechanics. In the first, known as the *Schrödinger picture*, the states evolve in time while the operators are fixed. In the second, known as the *Heisenberg picture*, the states are fixed while the operators evolve in time. Both give the same answers for any physical observable (i.e. expectation functions) but one approach may be more convenient for any given problem. It will turn out that the Heisenberg picture is best suited for cosmological purposes, so we pause to review it here.

The Schrödinger picture is perhaps the most intuitive. Here the evolution of states is determined by the time-dependent Schrödinger equation

$$i\hbar\frac{d|\psi\rangle}{dt} = \hat{H}|\psi\rangle$$

Alternatively, we can introduce a unitary evolution operator $U(t)$ which dictates how the states evolve,

$$|\psi(t)\rangle = \hat{U}(t)|\psi(0)\rangle$$

The Schrödinger equation tells us that this operator must obey

$$i\hbar\frac{d\hat{U}}{dt} = \hat{H}\hat{U} \tag{3.86}$$

If $\hat{H}$ is time-independent then this is solved by $\hat{U} = \exp\left(-i\hat{H}t/\hbar\right)$. However, if $\hat{H}$ is time-dependent (as it will be for us) we must be more careful.

In the Heisenberg picture, this time dependence is moved onto the operators. We consider the state to be fixed, while operators evolve as

$$\hat{\mathcal{O}}(t) = U^\dagger(t)\,\hat{\mathcal{O}}\,\hat{U}(t)$$

From (3.86), we find that these time-dependent operators obey

$$\frac{d\hat{\mathcal{O}}}{dt} = \frac{i}{\hbar}[\hat{H}, \hat{\mathcal{O}}] \tag{3.87}$$

We can look at how this works for the harmonic oscillator with a fixed frequency ω. The creation and annihilation operators $\hat{a}$ and $\hat{a}^\dagger$ have a particularly simple time evolution,

$$\begin{aligned} [\hat{H}, \hat{a}] = -\hbar\omega\hat{a} \quad &\Rightarrow \quad \hat{a}(t) = e^{-i\omega(t-t_0)}\,\hat{a}(t_0) \\ [\hat{H}, \hat{a}^\dagger] = +\hbar\omega\hat{a}^\dagger \quad &\Rightarrow \quad \hat{a}(t) = e^{+i\omega(t-t_0)}\,\hat{a}^\dagger(t_0) \end{aligned}$$

We can then simply substitute this into (3.84) to see how $\hat{q}(t)$ and $\hat{p}(t)$ evolve in time. We have

$$\begin{aligned} \hat{q}(t) &= \sqrt{\frac{\hbar}{2\omega}}\left(e^{-i\omega(t-t_0)}\,\hat{a}(t_0) + e^{+i\omega(t-t-0)}\,\hat{a}^\dagger(t_0)\right) \\ \hat{p}(t) &= -i\sqrt{\frac{\hbar\omega}{2}}\left(e^{-i\omega(t-t_0)}\,\hat{a}(t_0) - e^{+i\omega(t-t-0)}\,\hat{a}^\dagger(t_0)\right) \end{aligned} \tag{3.88}$$

Note that these obey the operator equation of motion (3.87), with

$$\frac{d\hat{q}}{dt} = \frac{i}{\hbar}[\hat{H}, \hat{q}] = \hat{p} \quad \text{and} \quad \frac{d\hat{p}}{dt} = \frac{i}{\hbar}[\hat{H}, \hat{p}] = -\omega^2\hat{q}$$

The Time-Dependent Harmonic Oscillator

For our cosmological application, we need to understand the physics of a harmonic oscillator with a time-dependent frequency,

$$\hat{H}(t) = \frac{1}{2}\hat{p}^2 + \frac{1}{2}\omega^2(t)\hat{q}^2$$

Our real interest is in the specific time-dependence (3.82) but, for now, we will keep $\omega(t)$ arbitrary.

A time-dependent Hamiltonian opens up different kinds of questions. We could, for example, pick some fixed moment in time t_0 at which we diagonalise the Hamiltonian. We do this by introducing the usual annihilation and creation operators, and place the system in the instantaneous ground state

$$\hat{a}(t_0)|0\rangle = 0$$

Now the system subsequently evolves. But, with a time-dependent Hamiltonian it will no longer sit in the ground state (in the Schrödinger picture). This is related to the fact that energy is no longer conserved when the Hamiltonian is time-dependent. We want to understand how the variance (3.85) evolves in this situation.

We will work in the Heisenberg picture. In analogy with (3.88), we expand the position operator in terms of $\hat{a}(t_0)$ and $\hat{a}^\dagger(t_0)$, with some time-dependent coefficients

$$\hat{q}(t) = v(t)\,\hat{a}(t_0) + v^\star(t)\,\hat{a}^\dagger(t_0) \tag{3.89}$$

The momentum is then

$$\hat{p}(t) = \frac{d\hat{q}}{dt} = \dot{v}(t)\,\hat{a}(t_0) + \dot{v}^\star(t)\,\hat{a}^\dagger(t_0)$$

Taking a second time derivative, we have

$$\frac{d\hat{p}}{dt} = \ddot{v}(t)\,\hat{a}(t_0) + \ddot{v}^\star(t)\,\hat{a}^\dagger(t_0) = -\omega^2(t)\hat{q}(t)$$

where the second equality comes from the operator equation of motion (3.87). Comparing coefficients of $\hat{a}(t_0)$ and $\hat{a}^\dagger(t_0)$, we see that the coefficient $v(t)$ must obey the original equation of motion

$$\ddot{v} + \omega^2(t)v = 0 \tag{3.90}$$

Meanwhile, we can normalise $v(t)$ by insisting that $[\hat{q}(t), \hat{p}(t)] = i\hbar$ and $[\hat{a}(t_0), \hat{a}^\dagger(t_0)] = 1$. These are compatible provided

$$v\dot{v}^\star - v^\star\dot{v} = i\hbar \tag{3.91}$$

When ω is constant, this agrees with what we saw before: we had $v = \sqrt{\hbar/2\omega}e^{-i\omega(t-t_0)}$, which is a solution to the harmonic oscillator (3.90), with the normalisation fixed by (3.91).

Finally, we can answer the main question: if we place the time-dependent harmonic oscillator in the ground state $|0\rangle$ at some time t_0, how does the variance of $\hat{q}(t)$ subsequently evolve? Using (3.89), we have

$$\langle \hat{q}^2(t) \rangle = |v(t)|^2 \tag{3.92}$$

This is the result we need to evaluate the size of quantum fluctuations during inflation.

3.5.4 Quantum Inflationary Perturbations

We can now import the quantum mechanical story above directly to the inflaton field. Recall that each (rescaled) Fourier mode of the inflaton acts like a harmonic oscillator with a time-dependent frequency,

$$\frac{d^2\tilde{\phi}_{\mathbf{k}}}{d\tau^2} + \omega_{\mathbf{k}}^2(\tau)\tilde{\phi}_{\mathbf{k}} = 0 \quad \text{with} \quad \omega_k^2(\tau) = c^2k^2 - \frac{2}{\tau^2}$$

We treat each Fourier component as an independent quantum operator which, piling hat on hat, we write as $\hat{\tilde{\phi}}_{\mathbf{k}}$. This is analogous to $\hat{q}$ in the harmonic oscillator that we described above. Following (3.89), we write

$$\hat{\tilde{\phi}}_{\mathbf{k}}(\tau) = v_{\mathbf{k}}(\tau)\,\hat{a}_{\mathbf{k}}(\tau_0) + v_{\mathbf{k}}^\star(\tau)\,\hat{a_{\mathbf{k}}}^\dagger(\tau_0) \tag{3.93}$$

where, as we've seen, $v(\tau)$ must obey the original harmonic oscillator equation (3.90), together with the normalisation condition (3.91) (with $\dot{v} = dv/d\tau$ in these equations).

First, we must decide when we're going to place the system in its ground state. The only sensible option is to do this right at the beginning of inflation, with $\tau_0 \to -\infty$. At this point, the frequency is simply $\omega_k^2 = c^2k^2$ and we get the normal harmonic oscillator. In the context of inflation, this choice is referred to as the *Bunch-Davies vacuum*. As we will see, this simple choice for the initial conditions at the very beginning of the universe is the one that ultimately agrees with what we see around us today.

Next, we must determine the coefficient $v_{\mathbf{k}}(\tau)$. We know that the general solution to (3.90) is (3.83)

$$v_{\mathbf{k}}(\tau) = \alpha e^{-ick\tau}\left(1 - \frac{i}{ck\tau}\right) + \beta e^{+ick\tau}\left(1 + \frac{i}{ck\tau}\right)$$

We need only to fix the integration constants α and β. We set $\beta = 0$ to ensure that, as $\tau \to -\infty$, the operator expansion (3.93) agrees with that of the normal harmonic oscillator. The normalisation of α is then fixed by (3.91)

$$v_{\mathbf{k}}\dot{v}_{\mathbf{k}}^\star - v_{\mathbf{k}}^\star\dot{v}_{\mathbf{k}} = 2\alpha^2 ick = i\hbar \quad \Rightarrow \quad \alpha^2 = \frac{\hbar}{2ck}$$

Now we're home and dry. The time-dependent coefficient in the expansion of the Fourier mode $\hat{\tilde{\phi}}_{\mathbf{k}}$ is

$$v_{\mathbf{k}}(\tau) = \sqrt{\frac{\hbar}{2ck}} e^{-ick\tau}\left(1 - \frac{i}{ck\tau}\right)$$

So the quantum fluctuations in the field $\tilde{\phi}_{\mathbf{k}}$ can be read off from (3.92),

$$\langle \hat{\tilde{\phi}}_{\mathbf{k}} \hat{\tilde{\phi}}^{\dagger}_{\mathbf{k}} \rangle = \frac{\hbar}{2ck}\left(1 + \frac{1}{c^2k^2\tau^2}\right)$$

where we have to take $\hat{\phi}\hat{\phi}^{\dagger}$ because, in contrast to $\hat{q}$, the Fourier mode $\hat{\phi}_{\mathbf{k}}$ is complex. Our interest is in the original field $\phi_{\mathbf{k}} = -H_{\rm inf}\tau\tilde{\phi}_{\mathbf{k}}$. (This rescaling was introduced back in (3.80).) The fluctuations of this field are given by

$$\langle \hat{\phi}_{\mathbf{k}} \hat{\phi}^{\dagger}_{\mathbf{k}} \rangle = \frac{\hbar H^2_{\rm inf}}{2ck}\left(\frac{1}{c^2k^2} + \tau^2\right)$$

At early times, the fluctuations are large. However, at late times, $ck\tau \to 0^-$, the fluctuations become constant in time. The cross-over happens at $ck\tau \approx 1$, which is when the fluctuations exit the horizon. At later times, the k dependence of the fluctuations is given by

$$\lim_{ck\tau\to 0^-} \langle \hat{\phi}_{\mathbf{k}} \hat{\phi}^{\dagger}_{\mathbf{k}} \rangle = \frac{\hbar H^2_{\rm inf}}{2c^3k^3} \tag{3.94}$$

This is the famous inflationary power spectrum. It takes the Harrison-Zel'dovich "scale invariant" form, a statement which, as we explained in Section 3.2.1, is manifest only when written in terms of the power spectrum introduced in (3.50),

$$\Delta_{\phi}(k) = \frac{4\pi k^3 \langle \hat{\phi}_{\mathbf{k}} \hat{\phi}^{\dagger}_{\mathbf{k}} \rangle}{(2\pi)^3} = \frac{\hbar H^2_{\rm inf}}{4\pi^2 c^3}$$

This is indeed independent of k. These fluctuations remain frozen outside the horizon, until they subsequently re-enter during the radiation dominated era or, for very long wavelength, matter dominated era.

The fact that the power spectrum $\Delta(k)$ does not depend on the wavelength can be traced to an underlying, scale invariance symmetry of de Sitter space.

A Rolling Inflation

The calculation above holds for a scalar field ϕ with $V(\phi) = \text{constant}$. This, of course, is not the realistic situation for inflation, but it's a good approximation when the scalar field rolls down a rather flat potential. In this case, the shorter wavelength modes (larger k) which exit the horizon later will have a slightly smaller H and, correspondingly, slightly smaller fluctuations. This means that the power spectrum is almost, but not quite, scale invariant.

We will not present this longer calculation here; we quote only the answer which we write as

$$\Delta_\phi(k) \sim k^{n_S - 1}$$

Here *scalar spectral index* n_s is close to 1. It turns out that, to leading order,

$$n_S = 1 - 2\epsilon \tag{3.95}$$

where ϵ is a dimensionless number known as a *slow-roll parameter*. It is one of two such parameters which are commonly used to characterise the shape of the inflaton potential,

$$\epsilon = \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 \quad \text{and} \quad \eta = M_{\rm pl}^2 \frac{V''}{V}$$

with the Planck mass given by $M_{\rm pl}^2 = \hbar c/8\pi G$.

The Gravitational Power Spectrum

To compare to observations, we must turn the fluctuations of the inflaton field ϕ into fluctuations in the energy density or, as explained in (3.2.1), the gravitational potential Φ. As with many details, a full treatment needs a relativistic analysis. It turns out that the inflationary perturbations imprint themselves directly as fluctuations of the gravitational potential,

$$\Delta_\phi(k) \mapsto \Delta_\Phi(k)$$

But this is exactly what we need! The almost scale-invariant power spectrum of the inflaton gives rise to the almost scale-invariant power spectrum needed to explain the structure of galaxies in our universe. Moreover, the observed spectral index $n \approx 0.97$ can be used to infer something about the dynamics of the inflaton in the early universe.

There are many remarkable things about the inflationary origin of density perturbations. Here is another: the fluctuations that we computed in (3.94) are quantum. They measure the spread in the wavefunction. Yet these must turn into classical probabilities which, subsequently, correspond to the random distribution of galaxies in the universe. This is, at heart, no different from the quantum measurement problem in any other setting, now writ large across the sky. But one may worry that, in the absence of any observers, the problem is more acute. Closer analysis suggests that the modes decohere, and evolve from quantum to classical, as they exit the horizon.

3.5.5 Things We Haven't (Yet?) Seen

There is much more to tell about inflation, both things that work and things that don't. Here, as a taster, is a brief description of two putative features of inflation which might, with some luck, be detected in the future.

Gravitational Waves

It's not just the inflaton that suffers quantum fluctuations during inflation. There are also quantum fluctuations of spacetime itself.

It's a common misconception that we don't understand quantum gravity. There is, of course, some truth to this: there are lots of things that we don't understand about quantum gravity, such as what happens inside the singularity of a black hole. But provided that the curvature of spacetime is not too large, we can do trustworthy quantum gravity calculations, and inflation provides just such an opportunity.

These quantum gravity fluctuations leave an imprint on spacetime and, subsequently, on the CMB. This can be traced back to the fact that the graviton is a particle with spin 2. Correspondingly, these fluctuations have a distinctive swirly pattern, known as B-mode polarisation.

We have not yet observed such B-modes in the CMB, although it's not for the want of trying. Finding them would be a very big deal: not only would it be our first observational evidence of quantum gravity, but they would tell us directly the scale at which inflation occurs, meaning that we can determine $H_{\rm inf}$, or equivalently, the magnitude of the potential $V(\phi)$. (In contrast, the density perturbations that we have observed depend on both $V(\phi)$ and the slow-roll parameter ϵ as we can see in (3.95).)

The power spectrum of tensor modes is denoted Δ_T (with T for tensor). It also predicted to take (almost) Harrison-Zel'dovich form, but with a slightly different spectral index from the scalar modes. Cosmologists place limits on the strength of these tensor

perturbations relative to the scalar modes Δ_ϕ formed by the inflaton. The ratio is defined to be

$$r = \frac{\Delta_T}{\Delta_\phi}$$

Currently, the lack of observation only allows us to place an upper limit of $r \leq 0.07$, although it's possible to relax this if we allow some flexibility with other parameters. Roughly speaking, if inflation is driven by physics close to the Planck scale or GUT scale then we have a hope of detecting $r \neq 0$. If, however, the scale of inflation is closer to the TeV scale (the current limit of our knowledge in particle physics) then it seems unlikely we will find tensor modes in our lifetime.

Non-Gaussianity

We saw in Section 3.2.1 that the observed spectrum of density perturbations is well described by a Gaussian probability distribution. This too is a success of inflation: one can show that in slow-roll inflation three point functions $\langle \hat{\phi}_{\mathbf{k}_1} \hat{\phi}_{\mathbf{k}_2} \hat{\phi}_{\mathbf{k}_3} \rangle$ are suppressed by the slow-roll parameters ϵ^2 and η^2.

Nonetheless, this hasn't stopped people hoping. The discovery of non-Gaussian primordial density fluctuations would provide us with a wealth of precious information about the detailed dynamics of the inflation in the early universe. While the two-point function tells us just two numbers — n_S and the overall scale of the power spectrum — the three-point correlator $\langle \hat{\phi}_{\mathbf{k}_1} \hat{\phi}_{\mathbf{k}_2} \hat{\phi}_{\mathbf{k}_3} \rangle \sim f_{NL}\, \delta_D^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)$ is a function of every triangle you can draw on the (Fourier transformed) sky. For this reason, there has been a big push to try to detect a primordial non-Gaussian signal in the CMB or large scale structure. Alas, so far, to no avail. Meanwhile, ever optimistic theorists have proposed more creative versions of inflation which give rise to non-Gaussianity at a detectable level[15]. Sadly, there is little evidence that these theorists are going to be validated any time soon.

[15] See, for example, M. Alishahiha, E. Silverstein and D. Tong, "*DBI in the sky*", Phys. Rev. D **70**, 123505 (2004) [hep-th/0404084].