LAGRANGIANS AND FIELDS To understand what scalar fields can do for cosmology, it is necessary to use some elements of the more powerful Lagrangian description of the dynamics. We will try to keep this fairly informal. Consider first a classical system of particles: the **Lagrangian** L is defined as the difference of the kinetic and potential energies, $L = T - V$, for some set of particles with coordinates $q_i(t)$. **Euler's equation** gives an equation of motion for each particle

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i}. \tag{99}$$

As a sanity check, consider a single particle in a potential in 1D: $L = m\dot{x}^2/2 - V(x)$. $\partial L/\partial \dot{x} = m\dot{x}$, so we get $m\ddot{x} = -\partial V/\partial x$, as desired. The advantage of the Lagrangian formalism, of course, is that it is not necessary to use Cartesian coordinates. In passing, we note that the formalism also supplies a general definition of momentum:

$$p_i \equiv \frac{\partial L}{\partial \dot{q}_i}, \tag{100}$$

which again is clearly sensible for Cartesian coordinates.

A field may be regarded as a dynamical system, but with an infinite number of degrees of freedom. How do we handle this? A hint is provided by electromagnetism, where we are familiar with writing the total energy in terms of a density which, as we are dealing with generalized mechanics, we may formally call the Hamiltonian density:

$$H = \int \mathcal{H}\, dV = \int \left(\frac{\epsilon_0 E^2}{2} + \frac{B^2}{2\mu_0}\right)\, dV. \tag{101}$$

This suggests that we write the Lagrangian in terms of a **Lagrangian density** $\mathcal{L}$: $L = \int \mathcal{L}\, dV$. This quantity is of such central importance in quantum field theory that it is usually referred to (incorrectly) simply as 'the Lagrangian'. The equation of motion that corresponds to Euler's equation is now the **Euler–Lagrange equation**

$$\frac{\partial}{\partial x^\mu}\left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right] = \frac{\partial \mathcal{L}}{\partial \phi}, \tag{102}$$

where we use the shorthand $\partial_\mu \phi \equiv \partial \phi/\partial x^\mu$. Note the downstairs index for consistency: in special relativity, $x^\mu = (ct, \mathbf{x})$, $x^\mu = (ct, -\mathbf{x}) = g_{\mu\nu}x^\nu$. The Lagrangian $\mathcal{L}$ and the field equations are therefore generally equivalent, although the Lagrangian arguably seems more fundamental: we can obtain the field equations given the Lagrangian, but inverting the process is less straightforward.

For quantum mechanics, we want a Lagrangian that will yield the Klein–Gordon equation. If ϕ is a single real scalar field, then the required Lagrangian is

$$\mathcal{L} = \tfrac{1}{2}\partial^\mu\phi\partial_\mu\phi - V(\phi); \quad V(\phi) = \tfrac{1}{2}\mu^2\phi^2. \tag{103}$$

Again, we will be content with checking that this does the right thing in a simple case: the homogeneous model, where $\mathcal{L} = \dot{\phi}^2/2 - V(\phi)$. This is now just like the earlier example, and gives $\ddot{\phi} = -\partial V/\partial\phi$, as required.

NOETHER'S THEOREM The final ingredient we need before applying scalar fields to cosmology is to understand that they can be treated as a fluid with thermodynamic properties like pressure. these properties are derived by a profoundly important general argument that relates the existence of global symmetries to conservation laws in physics. In classical mechanics, conservation of energy and momentum arise by considering Euler's equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \tag{104}$$

where $L = \sum_i T_i - V_i$ is a sum over the difference in kinetic and potential energies for the particles in a system. If L is independent of all the position coordinates q_i, then we obtain conservation of momentum (or angular momentum, if q is an angular coordinate): $p_i \equiv \partial L/\partial \dot{q}_i =$ constant for each particle. More realistically, the potential will depend on the q_i, but homogeneity of space says that the Lagrangian as a whole will be unchanged by a **global translation**: $q_i \to q_i + dq$, where dq is some constant. Using Euler's equation, this gives conservation of total momentum:

$$dL = \sum_i \frac{\partial L}{\partial q_i}\, dq \quad \Rightarrow \quad \frac{d}{dt}\sum_i p_i = 0. \tag{105}$$

If L has no explicit dependence on t, then

$$\frac{dL}{dt} = \sum_i \left(\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i\right) = \sum_i (\dot{p}_i\dot{q}_i + p_i\ddot{q}_i), \tag{106}$$

which leads us to define the **Hamiltonian** as a further constant of the motion

$$H \equiv \sum_i p_i \dot{q}_i - L = \text{constant}. \tag{107}$$

Something rather similar happens in the case of quantum (or classical) field theory: the existence of a global symmetry leads directly to a conservation law. The difference between discrete dynamics and field dynamics, where the Lagrangian is a *density*, is that the result is expressed as a **conserved current** rather than a simple constant of the motion. Suppose the Lagrangian has no explicit dependence on spacetime (*i.e.* it depends on x^μ only implicitly through the fields and their 4-derivatives). As above, we write

$$\frac{d\mathcal{L}}{dx^\mu} = \frac{\partial \mathcal{L}}{\partial \phi}\frac{\partial \phi}{\partial x^\mu} + \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi)}\frac{\partial(\partial_\nu \phi)}{\partial x^\mu}, \tag{108}$$

Using the Euler–Lagrange equation to replace $\partial\mathcal{L}/\partial\phi$ and collecting terms results in

$$\frac{d}{dx^\nu}\left[\frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi)}\frac{\partial \phi}{\partial x^\mu} - \mathcal{L}g^{\mu\nu}\right] \equiv \frac{d}{dx^\nu}T^{\mu\nu} = 0. \tag{109}$$

This is a conservation law, as we can see by analogy with a simple case like the conservation of charge. There, we would write

$$\partial_\mu J^\mu = \dot{\rho} + \boldsymbol{\nabla}\cdot\mathbf{j} = 0, \tag{110}$$

where ρ is the charge density, $\mathbf{j}$ is the current density, and J^μ is the 4-current. We have effectively four such equations (one for each value of ν) so there must be four conserved quantities: clearly energy and the four components of momentum. Conservation of 4-momentum is expressed by $T^{\mu\nu}$, which is the 4-current of 4-momentum. For a simple fluid, it is just

$$T^{\mu\nu} = \text{diag}(\rho c^2, p, p, p), \tag{111}$$

so now we can read off the density and pressure generated by a scalar field. Note immediately the important consequence for cosmology: a potential term $-V(\phi)$ in the Lagrangian produces $T^{\mu\nu} = V(\phi)g^{\mu\nu}$. This is the $p = -\rho$ equation of state characteristic of the cosmological constant. If we now follow the evolution of ϕ, the cosmological 'constant' changes and we have the basis for models of inflationary cosmology.

4 Inflation – II

Topics to be covered:

- Models for inflation
- Slow roll dynamics
- Ending inflation

4.1 Equation of motion

Most of the main features of inflation can be illustrated using the simplest case of a single real scalar field, with Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - V(\phi) = \tfrac{1}{2}(\dot\phi^2 - \nabla^2\phi) - V(\phi). \tag{112}$$

It turns out that we can get inflation with even the simple mass potential $V(\phi) = m^2\,\phi^2/2$, but it is easy to keep things general. Noether's theorem gives the energy–momentum tensor for the field as

$$T^{\mu\nu} = \partial^\mu\phi\partial^\nu\phi - g^{\mu\nu}\mathcal{L}. \tag{113}$$

From this, we can read off the energy density and pressure:

$$\begin{aligned} \rho = T^{00} &= \tfrac{1}{2}\dot\phi^2 + V(\phi) + \tfrac{1}{2}(\nabla\phi)^2 \\ p = T^{11} &= \tfrac{1}{2}\dot\phi^2 - V(\phi) - \tfrac{1}{6}(\nabla\phi)^2. \end{aligned} \tag{114}$$

If the field is constant both spatially and temporally, the equation of state is then $p = -\rho$, as required if the scalar field is to act as a cosmological constant; note that derivatives of the field spoil this identification.

We now want to revisit the equation of motion for the scalar field, but with the critical difference that we place the field in the expanding universe. Everything so far has been special relativity, so we don't have quite enough formalism to derive the full equation of motion, but it is

$$\ddot\phi + 3H\dot\phi - \nabla^2\phi + dV/d\phi = 0. \tag{115}$$

This is a wave equation similar to the one in flat space. The **Hubble drag** term $3H\dot{\phi}$ is the main new feature: loosely, it reflects the fact that the redshifting effects of expansion will drain energy from the field oscillations.

This is not hard to prove in the homogeneous case, which is the main one of interest for inflationary applications. This is because $\nabla\phi = \nabla_{\rm comoving}\,\phi/R$. Since R increases exponentially, these perturbations are damped away: assuming V is large enough for inflation to start in the first place, inhomogeneities rapidly become negligible. In the homogeneous limit, we can simply appeal to energy conservation:

$$\frac{d\ln\rho}{d\ln a} = -3(1+w) = -3\dot{\phi}^2/(\dot{\phi}^2/2+V), \tag{116}$$

following which the relations $H = d\ln a/dt$ and $\dot{V} = \dot{\phi}V'$ can be used to change variables to t, and the damped oscillator equation for ϕ follows.

4.2 The slow-roll approximation

The solution of the equation of motion becomes tractable if we both ignore spatial inhomogeneities in ϕ and make the **slow-rolling approximation** that the $\ddot{\phi}$ term is negligible. The physical motivation here is to say that we are most interested in behaviour close to de Sitter space, so that the potential dominates the energy density. This requires

$$\dot{\phi}^2/2 \ll |V(\phi)|; \tag{117}$$

differentiating this gives $\ddot{\phi} \ll |dV/d\phi|$, as required. We therefore have a simple slow-rolling equation for homogeneous fields:

$$3H\dot{\phi} = -dV/d\phi. \tag{118}$$

In combination with Friedmann's equation in the natural-unit form

$$H^2 = \frac{8\pi}{3m_{\rm P}^2}(\dot{\phi}^2/2+V) \simeq \frac{8\pi}{3m_{\rm P}^2}V, \tag{119}$$

This gives a powerful but simple apparatus for deducing the expansion history of any inflationary model.

The conditions for inflation can be cast into useful dimensionless forms. The basic condition $V \gg \dot{\phi}^2$ can now be rewritten using the slow-roll relation as

$$\epsilon \equiv \frac{m_{\rm P}^2}{16\pi} \, (V'/V)^2 \ll 1. \tag{120}$$

Also, we can differentiate this expression to obtain the criterion $V'' \ll V'/m_{\rm P}$, or $m_{\rm P}V''/V \ll V'/V \sim \sqrt{\epsilon}/m_{\rm P}$. This gives a requirement for the second derivative of V to be small, which we can write as

$$\eta \equiv \frac{m_{\rm P}^2}{8\pi} \, (V''/V) \ll 1 \tag{121}$$

These two criteria make perfect intuitive sense: the potential must be flat in the sense of having small derivatives if the field is to roll slowly enough for inflation to be possible.

4.3 Inflationary models

The curse and joy of inflationary modelling is that nothing is known about the **inflaton** field ϕ, nor about its potential. We therefore consider simple classes of possible example models, with varying degrees of physical motivation.

If we think about a single field, models can be divided into two basic classes, as illustrated in figure 10. The simplest are **large-field inflation** models, in which the field is strongly displaced from the origin. There is nothing to prevent the scalar field from reaching the minimum of the potential – but it can take a long time to do so, and the universe meanwhile inflates by a large factor. In this case, inflation is realized by means of 'inertial confinement'. The opposite is when the potential is something like the Higgs potential, where the gradient vanishes at the origin: this is a model of **small-field inflation**. In principle, the field can stay at $\phi = 0$ forever if it is placed exactly there. One would say that the universe then inhabited a state of **false vacuum**, as opposed to the true vacuum at $V = 0$ (but it is important to be clear that there is no fundamental reason why the minimum should be at zero density exactly; we will return to this point).

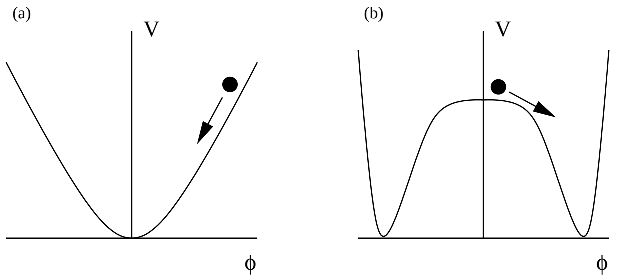

Figure 10. The two main classes of single-field inflation models: (a) large-field inflation; (b) small-field inflation. The former is motivated by a mass-like potential, the latter by something more like the Higgs potential.

The first inflation model (Guth 1981) was of the small-field type, but large-field models have tended to be considered more plausible, for two reasons. The first is to do with initial conditions. If inflation starts from anywhere near to thermal equilibrium at a temperature $T_{\rm GUT}$, we expect thermal fluctuations in ϕ; the potential should generally differ from its minimum by an amount $V \sim T_{\rm GUT}^4$. How then is the special case needed to trap the potential near $\phi = 0$ to arise? We have returned to the sort of fine-tuned initial conditions from which inflation was designed to save us. The other issue with simple small-field models relates to the issue of how inflation ends. This can be viewed as a form of phase transition, which is continuous or second order in the case of large-field models. For small-field models, however, the transition to the true vacuum can come about by quantum tunnelling, so that the transition is effectively discontinuous and first order. As we will discuss further below, this can lead to a universe that is insufficiently homogeneous to be consistent with observations.

CHAOTIC INFLATION MODELS Most attention is therefore currently paid to the large-field models where the field finds itself some way from its potential minimum. This idea is also termed **chaotic inflation**: there could be primordial chaos, within which conditions might vary. Some parts may attain the vacuum-dominated conditions needed for inflation, in which case they will expand hugely, leaving a universe inside a single bubble – which could be the one we inhabit. In principle this bubble has an edge, but if inflation persists for sufficiently long, the distance to this nastiness is so much greater than the current particle horizon that its existence has no testable consequences.

A wide range of inflation models of this kind is possible, but it will suffice here to discuss two simple special cases:

(1) **Polynomial inflation.** If the potential is taken to be $V \propto \phi^\alpha$, then the scale-factor behaviour can be very close to exponential. This becomes less true as α increases, but investigations are usually limited to ϕ^2 and ϕ^4 potentials on the grounds that higher powers are nonrenormalizable.

(2) **Power-law inflation.** On the other hand, $a(t) \propto t^p$ would suffice, provided $p > 1$. The potential required to produce this behaviour is

$$V(\phi) \propto \exp\left(\sqrt{\frac{16\pi}{p\, m_{\rm P}^2}}\,\phi\right). \tag{122}$$

This is an exact solution, not a slow-roll approximation.

HYBRID INFLATION One way in which the symmetric nature of the initial condition for small-field inflation can be made more plausible is to go beyond the space of single-field inflation. The most popular model in this generalized class is **hybrid inflation**, in which there are two fields, with potential

$$V(\phi, \psi) = \frac{1}{4\lambda}(M^2 - \lambda\psi^2)^2 + U(\phi) + \tfrac{1}{2}g^2\psi^2\phi^2. \tag{123}$$

We can think of this as being primarily $V(\psi)$, but with the form of V controlled by the second field, ϕ. For $\phi = 0$, we have the standard symmetry-breaking potential; but for large ϕ, $\phi > M/g$, the dependence on ψ becomes parabolic. Evolution in this parabolic trough at large ϕ can thus naturally lower ψ close to $\psi = 0$. If this happens, we have inflation driven by ϕ as the inflaton, with $V(\phi) = U(\phi) + \lambda M^2/4$. This extra constant in the potential raises H, so the Hubble damping term is particularly high, keeping the field from rolling away from $\psi = 0$ until near to $\phi = 0$.

Hybrid inflation therefore has the ability to make some of the features of the simplest inflation models seem more plausible, while introducing sufficient extra complexity that one can try to test

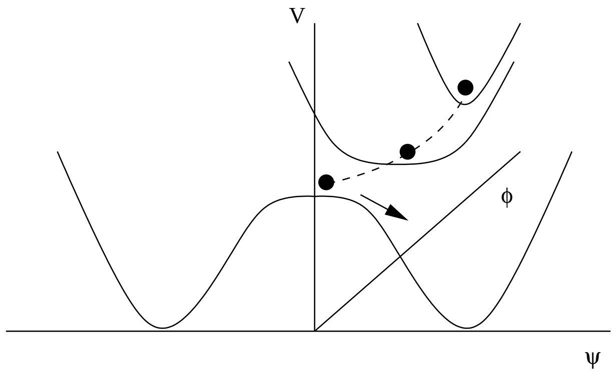

Figure 11. A sketch of the potential in hybrid inflation. For $\phi = 0$, $V(\psi)$ has the symmetry-breaking form of the potential for small-field inflation, but for large ϕ there is a simple quadratic minimum in $V(\psi)$. Evolution in this potential can drive conditions towards $\psi = 0$ while ϕ is large, preparing the way for something similar to small-field inflation.

the robustness of the predictions of the simple models. The form of the Lagrangian is also claimed to have some fundamental motivation (although this has been said of many Lagrangians). As a result, hybrid inflation is rather popular with inflationary theorists.

CRITERIA FOR INFLATION Successful inflation in any of these models requires > 60 e-foldings of the expansion. The implications of this are easily calculated using the slow-roll equation, which gives the number of e-foldings between ϕ_1 and ϕ_2 as

$$N = \int H \, dt = -\frac{8\pi}{m_{\rm P}^2} \int_{\phi_1}^{\phi_2} \frac{V}{V'} \, d\phi \tag{124}$$

For a potential that resembles a smooth polynomial, $V' \sim V/\phi$, and so we typically get $N \sim (\phi_{\rm start}/m_{\rm P})^2$, assuming that inflation terminates at a value of ϕ rather smaller than at the start. The criterion for successful inflation is thus that the initial value of the field exceeds the Planck scale:

$$\phi_{\rm start} \gg m_{\rm P}. \tag{125}$$

This is the real origin of the term 'large-field': it means that ϕ has to be large in comparison to the Planck scale. By the same argument, it is easily seen that this is also the criterion needed to make the slow-roll parameters ϵ and $\eta \ll 1$. To summarize, any model in which the potential is sufficiently flat that slow-roll inflation can commence will probably achieve the critical 60 e-foldings.

It is interesting to review this conclusion for some of the specific inflation models listed above. Consider a mass-like potential $V = m^2\phi^2$. If inflation starts near the Planck scale, the fluctuations in V are presumably $\sim m_{\rm P}^4$ and these will drive $\phi_{\rm start}$ to $\phi_{\rm start} \gg m_{\rm P}$ provided $m \ll m_{\rm P}$; similarly, for $V = \lambda\phi^4$, the condition is weak coupling: $\lambda \ll 1$. Any field with a rather flat potential will thus tend to inflate, just because typical fluctuations leave it a long way from home in the form of the potential minimum.

This requirement for weak coupling and/or small mass scales near the Planck epoch is suspicious, since quantum corrections will tend to re-introduce the Planck scale. In this sense, especially with the appearance of the Planck scale as the minimum required field value, it is not clear that the aim of realizing inflation in a classical way distinct from quantum gravity has been fulfilled.

4.4 Ending inflation

BUBBLE NUCLEATION AND THE GRACEFUL EXIT In small-field inflation, as in with Guth's initial idea, the potential is trapped at $\phi = 0$, and eventually undergoes a first-order phase transition. This model suffers from the problem that it predicts residual inhomogeneities after inflation is over that are far too large. This is easily seen: because the transition is first-order, it proceeds by **bubble nucleation**, where the vacuum tunnels between false and true vacua. However, the region occupied by these bubbles will grow as a causal process, whereas outside the bubbles the exponential expansion of inflation continues. This means that it is very difficult for the bubbles to percolate and eliminate the false vacuum everywhere, as is needed for an end to inflation. Instead, inflation continues indefinitely, with the bubbles of true vacuum having only a small filling factor at any time. This **graceful exit problem** motivated variants in which the phase transition is second order, and proceeds continuously by the field rolling slowly but freely down the potential.

REHEATING As we have seen, slow-rolling behaviour requires the field derivatives to be negligible; but the relative importance of time derivatives increases as V approaches zero (if the minimum is indeed at zero energy). Even if the potential does not steepen, sooner or later we will have $\epsilon \simeq 1$ or $|\eta| \simeq 1$ and the inflationary phase will cease. Instead of rolling slowly 'downhill', the field will oscillate about the bottom of the potential, with the oscillations becoming damped by the $3H\dot{\phi}$ friction term. Eventually, we will be left with a stationary field that either continues to inflate without end, if $V(\phi = 0) > 0$, or which simply has zero density.

However, this conclusion is incomplete, because we have so far neglected the couplings of the scalar field to matter fields. Such couplings will cause the rapid oscillatory phase to produce particles, leading to **reheating**. Thus, even if the minimum of $V(\phi)$ is at $V = 0$, the universe is left containing roughly the same energy density as it started with, but now in the form of normal matter and radiation – which starts the usual FRW phase, albeit with the desired special 'initial' conditions.

As well as being of interest for completing the picture of inflation, it is essential to realize that these closing stages of inflation are the *only* ones of observational relevance. Inflation might well continue for a huge number of e-foldings, all but the last few satisfying $\epsilon, \eta \ll 1$. However, the scales that left the de Sitter horizon at these early times are now vastly greater than our observable horizon, c/H_0, which exceeds the de Sitter horizon by only a finite factor – about e^{60} for GUT-scale inflation, as we saw earlier. Realizing that the observational regime corresponds only to the terminal phases of inflation is both depressing and stimulating: depressing, because ϕ may well not move very much during the last phases – our observations relate only to a small piece of the potential, and we cannot hope to recover its form without substantial *a priori* knowledge; stimulating, because observations even on very large scales must relate to a period where the simple concepts of exponential inflation and scale-invariant density fluctuations were coming close to breaking down. This opens the possibility of testing inflation theories in a way that would not be possible with data relating to only the simpler early phases. These tests take the form of tilt and gravitational waves in the final perturbation spectrum, to be discussed further below.

5 Fluctuations from inflation

Topics to be covered:

- Description of inhomogeneity
- Mechanisms for fluctuation generation
- Tilt and tensor modes
- Eternal inflation

5.1 The perturbed universe

We now need to consider the greatest achievement of inflation, which was not anticipated when the theory was first put forward: it provides a concrete mechanism for generating the seeds of structure in the universe. In essence, the idea is that the inevitable small quantum fluctuations in the inflaton field ϕ are transformed into residual classical fluctuations in density when inflation is over. The details of this process can be technical, and could easily fill a lecture course. The following treatment is therefore simplified as far as possible, while still making contact with the full results.

QUANTIFYING INHOMOGENEITY The first issue we have to deal with is how to quantify departures from uniform density. Frequently, an intuitive Newtonian approach can be used, and we will adopt this wherever possible. But we should begin with a quick overview of the relativistic approach to this problem, to emphasise some of the big issues that are ignored in the Newtonian method.

Because relativistic physics equations are written in a covariant form in which all quantities are independent of coordinates, relativity does not distinguish between *active* changes of coordinate (e.g. a Lorentz boost) or *passive* changes (a mathematical change of variable, normally termed a gauge transformation). This generality is a problem, as we can see by asking how some scalar quantity S (which might be density, temperature etc.) changes under a gauge transformation $x^\mu \to x'^\mu = x^\mu + \epsilon^\mu$. A gauge transformation induces the usual Lorentz transformation coefficients dx'^μ/dx^ν (which have no effect on a scalar), but also involves a translation that relabels spacetime points. We therefore have $S'(x^\mu + \epsilon^\mu) = S(x^\mu)$, or

$$S'(x^\mu) = S(x^\mu) - \epsilon^\alpha \partial S/\partial x^\alpha. \tag{126}$$

Consider applying this to the case of a uniform universe; here ρ only depends on time, so that

$$\rho' = \rho - \epsilon^0 \dot{\rho}. \tag{127}$$

An effective density perturbation is thus produced by a local alteration in the time coordinate: when we look at a universe with a fluctuating density, should we really think of a uniform model in which time is wrinkled? This ambiguity may seem absurd, and in the laboratory it could be resolved empirically by constructing the coordinate system directly – in principle by using light signals. This shows that the cosmological horizon plays an important role in this topic: perturbations with wavelength $\lambda \lesssim ct$ inhabit a regime in which gauge ambiguities can be resolved directly via common sense. The real difficulties lie in the super-horizon modes with $\lambda \gtrsim ct$. Within inflationary models, however, these difficulties can be overcome, since the true horizon is $\gg ct$.

The most direct general way of solving these difficulties is to construct perturbation variables that are explicitly independent of gauge. A comprehensive technical discussion of this method is given in chapter 7 of Mukhanov's book, and we summarize the essential elements here, largely without proof.

Firstly, metric perturbations can be split into three classes: **scalar perturbations**, which are described by scalar functions of spacetime coordinate, and which correspond to growing density perturbations; **vector perturbations**, which correspond to vorticity perturbations, and **tensor perturbations**, which correspond to gravitational waves. Here, we shall concentrate mainly on scalar perturbations.

A key result is that scalar perturbations can be described by just two gauge-invariant 'potentials' (functions of spacetime coordinates). Since these are gauge-invariant, we may as well write the perturbed metric in a particular gauge that makes things look as simple as possible. This is the **longitudinal gauge** in which the time and space parts of the RW metric are perturbed separately:

$$d\tau^2 = (1 + 2\Psi)dt^2 - (1 - 2\Phi)\gamma_{ij}\, dx^i\, dx^j. \tag{128}$$

Health warning: there are different conventions, and the symbols for the potentials are sometimes swapped, or signs flipped.

A second key result is that inserting the longitudinal metric into the Einstein equations shows that Ψ and Φ are identical in the case of fluid-like perturbations where off-diagonal elements of the energy–momentum tensor vanish. In this case, the longitudinal gauge becomes identical to the **Newtonian gauge**, in which perturbations are described by a single scalar field, which is the gravitational potential:

$$d\tau^2 = (1 + 2\Phi)dt^2 - (1 - 2\Phi)\gamma_{ij}\, dx^i\, dx^j, \tag{129}$$

and this should be quite familiar. If we consider small scales, so that the spatial metric γ_{ij} becomes that of flat space, then this form matches, for example, the Schwarzschild metric with $\Phi = -GM/r$, in the limit $\Phi/c^2 \ll 1$.

The conclusion is thus that the gravitational potential can for many purposes give an effectively gauge-invariant measure of cosmological perturbations. The advantage of this fact is that the gravitational potential is a familiar object, which we can manipulate and use our Newtonian intuition. This is still not guaranteed to give correct results on scales greater than the horizon, however, so a fully relativistic approach is to be preferred. But with the length restrictions of this course, it is hard to go beyond the Newtonian approach. The main results of the full theory can at least be understood and made plausible in this way.

FLUCTUATION POWER SPECTRA From the Newtonian point of view, potential fluctuations are directly related to those in density via Poisson's equation:

$$\nabla^2\Phi/a^2 = 4\pi G(1+3w)\,\bar\rho\,\delta, \tag{130}$$

where we have defined a dimensionless fluctuation amplitude

$$\delta \equiv \frac{\rho-\bar\rho}{\bar\rho}. \tag{131}$$

the factor of a^2 is there so we can use comoving length units in ∇^2 and the factor $(1+3w)$ accounts for the relativistic active mass density $\rho+3p$.

We are very often interested in asking how these fluctuations depend on scale, which amounts to making a Fourier expansion:

$$\delta(\mathbf{x}) = \sum \delta_k e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{132}$$

where $\mathbf{k}$ is the comoving wavevector. What are the allowed modes? If the field were periodic within some box of side L, we would have the usual harmonic boundary conditions

$$k_x = n\,\frac{2\pi}{L}, \qquad n = 1, 2\cdots, \tag{133}$$

and the inverse Fourier relation would be

$$\delta_k(\mathbf{k}) = \left(\frac{1}{L}\right)^3 \int \delta(\mathbf{x})\exp\big(i\mathbf{k}\cdot\mathbf{x}\big)\; d^3x. \tag{134}$$

Working in Fourier space in this way is powerful because it immediately gives a way of solving Poisson's equation and relating fluctuations in density and potential. For a single mode, $\nabla^2 \to -k^2$, and so

$$\Phi_k = -4\pi G(1+3w)a^2\,\bar{\rho}\,\delta_k/k^2. \tag{135}$$

The fluctuating density field can be described by its statistical properties. The mean is zero by construction; the variance is obtained by taking the volume average of δ^2:

$$\langle \delta^2 \rangle = \sum |\delta_k|^2. \tag{136}$$

To see this result, write the lhs instead as $\langle \delta\delta^* \rangle$ (makes no difference for a real field), and appreciate that all cross terms integrate to zero via the boundary conditions. For obvious reasons, the quantity

$$P(k) \equiv |\delta_k|^2 \tag{137}$$

is called the **power spectrum**. Note that, in an isotropic universe, we assume that P will be independent of direction of the wavevector in the limit of a large box: the fluctuating density field is statistically **isotropic**. In applying this apparatus, we would not want the (arbitrary) box size to appear. This happens naturally: as the box becomes big, the modes are finely spaced and a sum over modes is replaced by an integral over k space times the usual density of states, $(L/2\pi)^3$:

$$\langle \delta^2 \rangle = \sum |\delta_k|^2 \to \frac{L^3}{(2\pi)^3} \int P(k)\; d^3k = \int \Delta^2(k)\; d\ln k. \tag{138}$$

In the last step, we have defined the combination

$$\Delta^2(k) \equiv \frac{L^3}{(2\pi)^3}\, 4\pi k^3\, P(k), \tag{139}$$

which absorbs the box size into the definition of a dimensionless power spectrum, which gives the contribution to the variance from each logarithmic range of wavenumber (or wavelength). Despite the attraction of a dimensionless quantity, one still frequently sees plots of $P(k)$ – and often in a dimensionally fudged form in which $L = 1$ is assumed, and P given units of volume.

5.2 Relic fluctuations from inflation

OVERVIEW It was realized very quickly after the invention of inflation that the theory might also solve the other big puzzle with the initial condition of the universe. When we study gravitational instability, we will see that the present-day structure requires that the universe at even the Planck era would have had to possess a finite degree of inhomogeneity. Inflation suggests an audacious explanation for this structure, which is that it is an amplified form of the quantum fluctuations that are inevitable when the universe is sufficiently small. The present standard theory of this process was worked out by a number of researchers and generally agreed at a historic 1982 Nuffield conference in Cambridge.

The essence of the idea can be seen in figure 12. This reminds us that de Sitter space contains an **event horizon**, in that the comoving distance that particles can travel between a time t_0 and $t = \infty$ is finite,

$$r_{\rm EH} = \int_{t_0}^{\infty} \frac{c\, dt}{R(t)}; \tag{140}$$

this is not to be confused with the particle horizon, where the integral would be between 0 and t_0. With $R \propto \exp(Ht)$, the proper radius of the horizon is given by $R_0 r_{\rm EH} = c/H$. The exponential expansion literally makes distant regions of space move faster than light, so that points separated by $> c/H$ can never communicate with each other. If we imagine expanding the inflaton, ϕ, using comoving Fourier modes, then there are two interesting limits for the mode wavelength:

(1) 'Inside the horizon': $a/k \ll c/H$. Here the de Sitter expansion is negligible, just as we neglect the modern vacuum energy in the Solar system. The fluctuations in ϕ can be calculated exactly as in flat-space quantum field theory.

(2) 'Outside the horizon': $a/k \gg c/H$. Now the mode has a wavelength that exceeds the scale over which causal influences can operate. Therefore, it must now act as a 'frozen' quantity, which has the character of a classical disturbance. This field fluctuation can act as the seed for subsequent density fluctuations.

Before going any further, we can immediately note that a natural prediction will be a spectrum of perturbations that are nearly *scale invariant*. This means that the metric fluctuations of spacetime receive equal levels of distortion from each decade of perturbation wavelength, and may be quantified

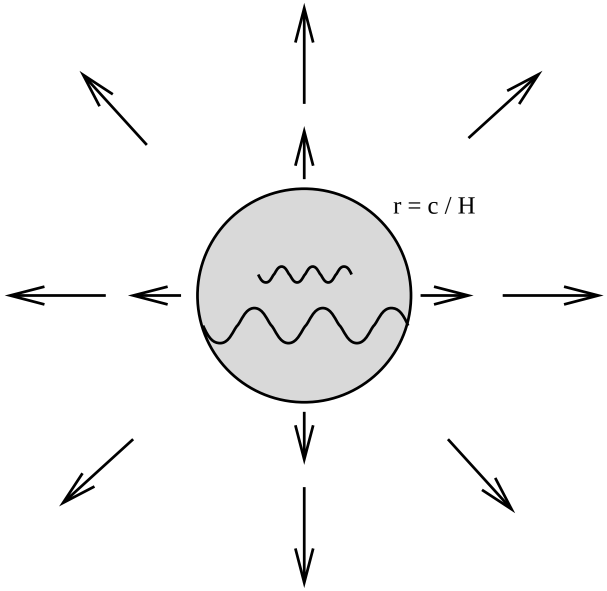

Figure 12. The event horizon in de Sitter space. Particles outside the sphere at $r = c/H$ can never receive light signals from the origin, nor can an observer at the origin receive information from outside the sphere. The exponential expansion quickly accelerates any freely falling observers to the point where their recession from the origin is effectively superluminal. The wave trains represent the generation of fluctuations in this spacetime. Waves with $\lambda \ll c/H$ effectively occupy flat space, and so undergo the normal quantum fluctuations for a vacuum state. As these modes (of fixed comoving wavelength) are expanded to sizes $\gg c/H$, causality forces the quantum fluctuation to become frozen as a classical amplitude that can seed large-scale structure.

in terms of the dimensionless power spectrum, Δ^2_Φ, of the Newtonian gravitational potential, Φ ($c = 1$):

$$\Delta^2_\Phi \equiv \frac{d\,\sigma^2(\Phi)}{d\,\ln k} = \text{constant} \equiv \delta^2_{\rm H}. \tag{141}$$

The origin of the term 'scale-invariant' is clear: since potential fluctuations modify spacetime, this is equivalent to saying that spacetime must be a fractal: it has the same level of deviation from the exact RW form on each level of resolution. It is common to denote the level of metric fluctuations by $\delta_{\rm H}$ – the **horizon-scale amplitude** (which we know to be about 10^{-5}). The justification for this name is that the potential perturbation is of the same order as the density fluctuation on the scale of the horizon at any given time. We can see this from Poisson's equation in Fourier space:

$$\Phi_k = -\frac{a^2}{k^2}\,4\pi G\,\bar{\rho}\,\delta_k = -(3/2)\,\frac{a^2}{k^2}\,\Omega_m H^2\,\delta_k \tag{142}$$

(where we have taken a $w = 0$ pressureless equation of state). This says that $\Phi_k/c^2 \sim \delta_k$ when the reciprocal of the physical wavenumber is c/H, i.e. is of order the horizon size.

The intuitive argument for scale invariance is that de Sitter space is invariant under time translation: there is no natural origin of time under exponential expansion. At a given time, the only length scale in the model is the horizon size c/H, so it is inevitable that the fluctuations that exist on this scale are the same at all times. By our causality argument, these metric fluctuations must be copied unchanged to larger scales as the universe exponentiates, so that the appearance of the universe is independent of the scale at which it is viewed.

If we accept this rough argument, then the implied density power spectrum is interesting. because of the relation between potential and density, it must be

$$\Delta^2(k) \propto k^4, \tag{143}$$

So the density field is very strongly inhomogeneous on small scales. Another way of putting this is in terms of a standard power-law notation for the non-dimensionless spectrum:

$$P(k) \propto k^n; \quad n = 1. \tag{144}$$

To get a feeling for what this means, consider the case of a matter distribution built up by the random placement of particles. It is not hard to show that this corresponds to **white noise**: a power spectrum that is independent of scale – i.e. $n = 0$. Recall the inverse Fourier relation:

$$\delta_k(\mathbf{k}) = \left(\frac{1}{L}\right)^3 \int \delta(\mathbf{x})\exp\big(i\mathbf{k}\cdot\mathbf{x}\big)\; d^3x. \tag{145}$$

Here, the density field is a sum of spikes at the locations of particles. Because the placement is random, the contribution of each spike is a complex number of phase uniformly distributed between 0 and 2π, independent of k. Conversely, the $n = 1$ 'scale-invariant' spectrum thus represents a density field that is super-uniform on large scales, but with enhanced small-scale fluctuations.

This $n = 1$ spectrum was considered a generic possibility long before inflation, and is also known as the **Zeldovich spectrum**. It is possible to alter this prediction of scale invariance only if the expansion is non-exponential; but we have seen that such deviations must exist towards the end of inflation. As we will see, it is natural for n to deviate from unity by a few %, and this is one of the predictions of inflation.

A MORE DETAILED TREATMENT We now need to give an outline of the exact treatment of inflationary fluctuations, which will allow us to calculate both the scale dependence of the spectrum and the absolute level of fluctuations. This can be a pretty technical subject, but it is possible to take a simple approach and still give a flavour of the main results and how they arise.

To anticipate the final answer, the inflationary prediction is of a horizon-scale amplitude

$$\delta_{\rm H} = \frac{H^2}{2\pi\,\dot\phi} \tag{146}$$

which can be understood as follows. Imagine that the main effect of fluctuations is to make different parts of the universe have fields that are perturbed by an amount $\delta\phi$. In other words, we are dealing with various copies of the same rolling behaviour $\phi(t)$, but viewed at different times

$$\delta t = \frac{\delta\phi}{\dot\phi}. \tag{147}$$

These universes will then finish inflation at different times, leading to a spread in energy densities (figure 13). The horizon-scale density amplitude is given by the different amounts that the universes have expanded following the end of inflation:

$$\delta_{\rm H} \simeq H\ \delta t = \frac{H}{\dot\phi}\,\delta\phi = \frac{H}{\dot\phi} \times \frac{H}{2\pi} = \frac{H^2}{2\pi\,\dot\phi}. \tag{148}$$

The $\delta_{\rm H} \simeq H\ \delta t$ argument relies on $R(t) \propto \exp(Ht)$ and that $\delta_{\rm H}$ is of order the fractional change in R. We will not attempt here to do better than justify the order of magnitude.

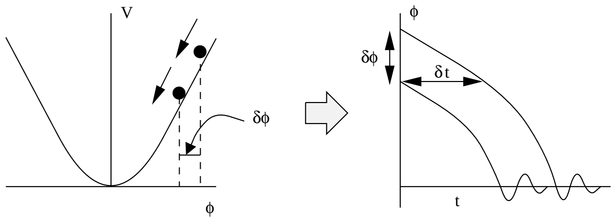

Figure 13. This plot shows how fluctuations in the scalar field transform themselves into density fluctuations at the end of inflation. Different points of the universe inflate from points on the potential perturbed by a fluctuation $\delta\phi$, like two balls rolling from different starting points. Inflation finishes at times separated by δt in time for these two points, inducing a density fluctuation $\delta = H\delta t$.

The last step uses the crucial input of quantum field theory, which says that the rms $\delta\phi$ is given by $H/2\pi$, and we now sketch the derivation of this result, What we need to do is consider the equation of motion obeyed by perturbations in the inflaton field. The basic equation of motion is

$$\ddot{\phi} + 3H\dot{\phi} - \nabla^2\phi + V'(\phi) = 0, \tag{149}$$

and we seek the corresponding equation for the perturbation $\delta\phi$ obtained by starting inflation with slightly different values of ϕ in different places. Suppose this perturbation takes the form of a comoving plane-wave perturbation of comoving wavenumber k and amplitude A: $\delta\phi = A\exp(i\mathbf{k}\cdot\mathbf{x} - ikt/a)$. If the slow-roll conditions are also assumed, so that V' may be treated as a constant, then the perturbed field $\delta\phi$ obeys the first-order perturbation of the equation of motion for the main field:

$$[\ddot{\delta\phi}] + 3H[\dot{\delta\phi}] + (k/a)^2[\delta\phi] = 0, \tag{150}$$

which is a standard wave equation for a massless field evolving in an expanding universe.

Having seen that the inflaton perturbation behaves in this way, it is not much work to obtain the quantum fluctuations that result in the field at late times (*i.e.* on scales much larger than the de Sitter horizon). First consider the fluctuations in flat space on scales well inside the horizon. In principle, this requires quantum field theory, but the vacuum fluctuations in ϕ can be derived by a simple argument using the uncertainty principle. First of all, note that the sub-horizon equation of motion is just that for a simple harmonic oscillator: $[\ddot{\delta\phi}] + \omega^2[\delta\phi] = 0$, where $\omega = k/a$. For an oscillator of mass m and position coordinate q, the rms uncertainty in q in the ground state is

$$q_{\rm rms} = \left(\frac{\hbar}{2m\omega}\right)^{1/2}. \tag{151}$$

This can be derived immediately from the uncertainty principle, which says that the minimum uncertainty is

$$\langle(\delta p)^2\rangle\langle(\delta q)^2\rangle = \hbar^2/4. \tag{152}$$

For a classical oscillation with $q(t) \propto e^{i\omega t}$, the momentum is $p(t) = m\dot{q} = i\omega m q(t)$. Quantum uncertainty can be thought of as saying that we lack a knowledge of the amplitude of the oscillator, but in any case the amplitudes in momentum and coordinate must be related by $p_{\rm rms} = m\omega q_{\rm rms}$. The uncertainty principle therefore says

$$m^2\omega^2 q_{\rm rms}^4 = \hbar^2/4, \tag{153}$$

which yields the required result.

For scalar field fluctuations, our 'coordinate' q is just the field $\delta\phi$, the oscillator frequency is $\omega = k/a$, and we now revert to $\hbar = 1$. What is the analogue of the mass of the oscillator in this case? Recall that a Lagrangian, L, has a momentum $p = \partial L/\partial\dot{q}$ corresponding to each coordinate. For the present application, the kinetic part of the Lagrangian density is

$$\mathcal{L}_{\rm kinetic} = a^3\dot{\phi}^2/2, \tag{154}$$

and the 'momentum' conjugate to ϕ is $p = a^3\dot{\phi}$. In the current case, p is a momentum *density*, since $\mathcal{L}$ is a Lagrangian density; we should therefore multiply p by a comoving volume V, so the analogue of the SHO mass is $m = a^3V$.

The uncertainty principle therefore gives us the variance of the zero-point fluctuations in $\delta\phi$ as

$$\langle(\delta\phi)^2\rangle = \big(2\,(a^3V)\,(k/a)\big)^{-1}, \tag{155}$$

so we adopt an rms field amplitude from quantum fluctuations of

$$\delta\phi = (a^3 V)^{-1/2}\, (2k/a)^{-1/2}\, e^{-ikt/a}. \tag{156}$$

This is the correct expression that results from a full treatment in quantum field theory.

With this boundary condition, it straightforward to check by substitution that the following expression satisfies the evolution equation:

$$\delta\phi = (a^3 V)^{-1/2}\, (2k/a)^{-1/2}\, e^{ik/aH}\, (1 + iaH/k) \tag{157}$$

(remember that H is a constant, so that $(d/dt)[aH] = H\dot{a} = aH^2$ *etc.*). At early times, when the horizon is much larger than the wavelength, $aH/k \ll 1$, and so this expression is the flat-space result, except that the time dependence looks a little odd, being $\exp(ik/aH)$. However, since $(d/dt)[k/aH] = -k/a$, we see that the oscillatory term has a leading dependence on t of the desired $-kt/a$ form.

At the opposite extreme, $aH/k \gg 1$, the squared fluctuation amplitude becomes frozen out at the value

$$\langle 0|\, |\phi_k|^2\, |0\rangle = \frac{H^2}{2k^3 V}, \tag{158}$$

where we have emphasised that this is the vacuum expectation value. The fluctuations in ϕ depend on k in such a way that the fluctuations per decade are constant:

$$\frac{d\,(\delta\phi)^2}{d\,\ln k} = \frac{4\pi k^3 V}{(2\pi)^3}\, \langle 0|\, |\phi_k|^2\, |0\rangle = \left(\frac{H}{2\pi}\right)^2 \tag{159}$$

(the factor $V/(2\pi)^3$ comes from the density of states in the Fourier transform, and cancels the $1/V$ in the field variance; $4\pi k^2\, dk = 4\pi k^3\, d\ln k$ comes from the k-space volume element).

This completes the argument. The initial quantum zero-point fluctuations in the field have been transcribed to a constant classical fluctuation that can eventually manifest itself as large-scale structure. The rms value of fluctuations in ϕ can be used as above to deduce the power spectrum of mass fluctuations well after inflation is over. In terms of the variance per $\ln k$ in potential perturbations, the answer is

$$\begin{aligned}
\delta_{\rm H}^2 \equiv \Delta_\Phi^2(k) &= \frac{H^4}{(2\pi\dot{\phi})^2} \\
H^2 &= \frac{8\pi}{3}\frac{V}{m_{\rm P}^2} \\
3H\dot{\phi} &= -V',
\end{aligned} \tag{160}$$

where we have also written once again the slow-roll condition and the corresponding relation between H and V, since manipulation of these three equations is often required in derivations.

TENSOR PERTURBATIONS Later in the course, we will compare the predictions of this inflationary apparatus with observations of the fluctuating density field of the contemporary universe. It should be emphasised again just what an audacious idea this is: that all the structure around us was seeded by quantum fluctuations while the universe was of subnuclear scale. It would be nice if we could verify this radical assumption, and there is one basic test: if the idea of quantum fluctuations is correct, it should apply to every field that was present in the early universe. In particular, it should apply to the gravitational field. This corresponds to metric perturbations in the form of a tensor $h^{\mu\nu}$, whose coefficients have some typical amplitude h (not the Hubble parameter). This **spatial strain** is what is measured by gravity-wave telescopes such as LIGO: the separation between a pair of freely-suspended masses changes by a fractional amount of order h as the wave passes. These experiments can be fabulously precise, with a current sensitivity of around $h = 10^{-21}$.

What value of h does inflation predict? For scalar perturbations, small-scale quantum fluctuations lead to an amplitude $\delta\phi = H/2\pi$ on horizon exit, which transforms to a metric fluctuation $\delta_{\rm H} = H\delta\phi/\dot{\phi}$. Tensor modes behave similarly – except that h must be dimensionless, whereas ϕ has dimensions of mass. On dimensional grounds, then, the formula for the tensor fluctuations is plausible:

$$h_{\rm rms} \sim H/m_{\rm P}. \tag{161}$$

But unlike fluctuations in the inflaton, the tensor fluctuation do not affect the progress of inflation: once generated, they play no further part in events and survive to the present day. Detection of these primordial tensor perturbations would not only give confidence in the basic inflationary picture, but would measure rather directly the energy scale of inflation.

INFLATON COUPLING The calculation of density inhomogeneities sets an important limit on the inflation potential. From the slow-rolling equation, we know that the number of e-foldings of inflation is

$$N = \int H \, dt = \int H \, d\phi / \dot{\phi} = \int 3H^2 \, d\phi / V'. \tag{162}$$

Suppose $V(\phi)$ takes the form $V = \lambda\phi^4$, so that $N = H^2/(2\lambda\phi^2)$. The density perturbations can then be expressed as

$$\delta_{\rm H} \sim \frac{H^2}{\dot{\phi}} = \frac{3H^3}{V'} \sim \lambda^{1/2} N^{3/2}. \tag{163}$$

Since $N \gtrsim 60$, the observed $\delta_{\rm H} \sim 10^{-5}$ requires

$$\lambda \lesssim 10^{-15}. \tag{164}$$

Alternatively, in the case of $V = m^2\phi^2$, $\delta_{\rm H} = 3H^3/(2m^2\phi)$. Since $H \sim \sqrt{V}/m_{\rm P}$, this gives $\delta_{\rm H} \sim m\phi^2/m_{\rm P}^3 \sim 10^{-5}$. Since we have already seen that $\phi \gtrsim m_{\rm P}$ is needed for inflation, this gives

$$m \lesssim 10^{-5} m_{\rm P}. \tag{165}$$

These constraints appear to suggest a defect in inflation, in that we should be able to use the theory to *explain* why $\delta_{\rm H} \sim 10^{-5}$, rather than using this observed fact to constrain the theory. The amplitude of $\delta_{\rm H}$ is one of the most important numbers in cosmology, and it is vital to know whether there is a simple explanation for its magnitude. One way to view this is to express the horizon-scale amplitude as

$$\delta_{\rm H} \sim \frac{V^{1/2}}{m_{\rm P}^2 \, \epsilon^{1/2}}. \tag{166}$$

We have argued that inflation will end with ϵ of order unity; if the potential were to have the characteristic value $V \sim E_{\rm GUT}^4$ then this would give the simple result

$$\delta_{\rm H} \sim \left(\frac{m_{\rm GUT}}{m_{\rm P}}\right)^2. \tag{167}$$

TILT Finally, deviations from exact exponential expansion must exist at the end of inflation, and the corresponding change in the fluctuation power spectrum is a potential test of inflation. Define the **tilt** of the fluctuation spectrum as follows:

$$\mathrm{tilt} \equiv 1 - n \equiv -\frac{d \ln \delta_{\rm H}^2}{d \ln k}. \tag{168}$$

We then want to express the tilt in terms of parameters of the inflationary potential, ϵ and η. These are of order unity when inflation terminates; ϵ and η must therefore be evaluated when the observed universe left the horizon, recalling that we only observe the last 60-odd e-foldings of inflation. The way to introduce scale dependence is to write the condition for a mode of given comoving wavenumber to cross the de Sitter horizon,

$$a/k = H^{-1}. \tag{169}$$

Since H is nearly constant during the inflationary evolution, we can replace $d/d \ln k$ by $d \ln a$, and use the slow-roll condition to obtain

$$\frac{d}{d \ln k} = a \, \frac{d}{da} = \frac{\dot\phi}{H} \, \frac{d}{d\phi} = -\frac{m_{\rm P}^2}{8\pi} \, \frac{V'}{V} \, \frac{d}{d\phi}. \tag{170}$$

We can now work out the tilt, since the horizon-scale amplitude is

$$\delta_{\rm H}^2 = \frac{H^4}{(2\pi\dot\phi)^2} = \frac{128\pi}{3} \left(\frac{V^3}{m_{\rm P}^6 \, V'^2} \right), \tag{171}$$

and derivatives of V can be expressed in terms of the dimensionless parameters ϵ and η. The tilt of the density perturbation spectrum is thus predicted to be

$$1 - n = 6\epsilon - 2\eta \tag{172}$$

For most models in which the potential is a smooth polynomial-like function, $|\eta| \simeq |\epsilon|$. Since ϵ has the larger coefficient and is positive by definition, the simplest inflation models tend to predict that the spectrum of scalar perturbations should be slightly tilted, in the sense that n is slightly less than unity.

It is interesting to put flesh on the bones of this general expression and evaluate the tilt for some specific inflationary models. This is easy in the case of power-law inflation with $a \propto t^p$ because the inflation parameters are constant: $\epsilon = \eta/2 = 1/p$, so that the tilt here is always

$$1 - n = 2/p \tag{173}$$

In general, however, the inflation derivatives have to be evaluated explicitly on the largest scales, 60 e-foldings prior to the end of inflation, so that we need to solve

$$60 = \int H\, dt = \frac{8\pi}{m_{\rm P}^2} \int_{\phi_{\rm end}}^{\phi} \frac{V}{V'}\, d\phi. \tag{174}$$

A better motivated choice than power-law inflation would be a power-law potential $V(\phi) \propto \phi^\alpha$; many chaotic inflation models concentrate on $\alpha = 2$ (mass-like term) or $\alpha = 4$ (highest renormalizable power). Here, $\epsilon = m_{\rm P}^2 \alpha^2/(16\pi\phi^2)$, $\eta = \epsilon \times 2(\alpha - 1)/\alpha$, and

$$60 = \frac{8\pi}{m_{\rm P}^2} \int_{\phi_{\rm end}}^{\phi} \frac{\phi}{\alpha}\, d\phi = \frac{4\pi}{m_{\rm P}^2 \alpha}(\phi^2 - \phi_{\rm end}^2). \tag{175}$$

It is easy to see that $\phi_{\rm end} \ll \phi$ and that $\epsilon = \alpha/240$, leading finally to

$$1 - n = (2 + \alpha)/120. \tag{176}$$

The predictions of simple chaotic inflation are thus very close to scale invariance in practice: $n = 0.97$ for $\alpha = 2$ and $n = 0.95$ for $\alpha = 4$. However, such a tilt has a significant effect over the several decades in k from CMB anisotropy measurements to small-scale galaxy clustering. These results are in some sense the default inflationary predictions: exact scale invariance would be surprising, as would large amounts of tilt. Either observation would indicate that the potential must have a more complicated structure, or that the inflationary framework is not correct.

5.3 Stochastic eternal inflation

These fluctuations in the scalar field can affect the progress of inflation itself. They can be thought of as adding a random-walk element to the classical rolling of the scalar field down the trough defined by $V(\phi)$. In cases where ϕ is too close to the origin for inflation to persist for sufficiently long, it is possible for the quantum fluctuations to push ϕ further out – creating further inflation in a self-sustaining process. This is the concept of **stochastic inflation**.

Consider the scalar field at a given point in the inflationary universe. Each e-folding of the expansion produces new classical fluctuations, which add incoherently to those previously present. If the field is sufficiently far from the origin in a polynomial potential, these fluctuations produce a random walk of $\phi(t)$ that overwhelms the classical trajectory in which ϕ tries to roll down the potential, as follows. The classical amplitude from quantum fluctuations is $\delta\phi = H/2\pi$, and a new disturbance of the same rms will be added for every $\Delta t = 1/H$. The slow-rolling equation says that the trajectory is $\dot{\phi} = -V'/3H$; we also have $H^2 = 8\pi V/3m_{\rm P}^2$, so that the classical change in ϕ is $\Delta\phi = -m_{\rm P}^2 V'/8\pi V$ in a time $\Delta t = 1/H$. Consider $V = \lambda|\phi|^n/(n m_{\rm P}^{n-4})$, for which these two changes in ϕ will be equal at $\phi \sim \phi^* = m_{\rm P}/\lambda^{1/(n+2)}$. For smaller ϕ, the quantum fluctuations will have a negligible effect on the classical trajectory; for larger ϕ, the equation of motion will become stochastic. The resulting random walk will send some parts of the universe to ever larger values of ϕ, so inflation never entirely ends. This **eternal inflation** is the basis for the concept of the **inflationary multiverse**: different widely-separated parts of the universe will inflate by different amounts, producing in effect separate universes with distinct formation histories.

6 Structure formation – I

6.1 Newtonian equations of motion

We have decided that perturbations will in most cases effectively be described by the Newtonian potential, Φ. Now we need to develop an equation of motion for Φ, or equivalently for the density fluctuation $\rho \equiv (1+\delta)\bar{\rho}$. In the Newtonian approach, we treat dynamics of cosmological matter exactly as we would in the laboratory, by finding the equations of motion induced by either pressure or gravity. We begin by casting the problem in comoving units:

$$\begin{aligned} \mathbf{x}(t) &= a(t)\mathbf{r}(t) \\ \delta\mathbf{v}(t) &= a(t)\mathbf{u}(t), \end{aligned} \tag{177}$$

so that $\mathbf{x}$ has units of proper length, i.e. it is an **Eulerian coordinate**. First note that the comoving peculiar velocity $\mathbf{u}$ is just the time derivative of the comoving coordinate $\mathbf{r}$:

$$\dot{\mathbf{x}} = \dot{a}\mathbf{r} + a\dot{\mathbf{r}} = H\mathbf{x} + a\dot{\mathbf{r}}, \tag{178}$$

where the rhs must be equal to the Hubble flow $H\mathbf{x}$, plus the peculiar velocity $\delta\mathbf{v} = a\mathbf{u}$.

The equation of motion follows from writing the Eulerian equation of motion as $\ddot{\mathbf{x}} = \mathbf{g}_0 + \mathbf{g}$, where $\mathbf{g} = -\boldsymbol{\nabla}\Phi/a$ is the peculiar acceleration, and $\mathbf{g}_0$ is the acceleration that acts on a particle in a homogeneous universe (neglecting pressure forces to start with, for simplicity). Differentiating $\mathbf{x} = a\mathbf{r}$ twice gives

$$\ddot{\mathbf{x}} = a\dot{\mathbf{u}} + 2\dot{a}\mathbf{u} + \frac{\ddot{a}}{a}\mathbf{x} = \mathbf{g}_0 + \mathbf{g}. \tag{179}$$

The unperturbed equation corresponds to zero peculiar velocity and zero peculiar acceleration: $(\ddot{a}/a)\,\mathbf{x} = \mathbf{g}_0$; subtracting this gives the perturbed equation of motion

$$\dot{\mathbf{u}} + 2(\dot{a}/a)\mathbf{u} = \mathbf{g}/a = -\boldsymbol{\nabla}\Phi/a^2. \tag{180}$$

This equation of motion for the peculiar velocity shows that $\mathbf{u}$ is affected by gravitational acceleration and by the **Hubble drag** term, $2(\dot{a}/a)\mathbf{u}$. This arises because the peculiar velocity falls with time as a particle attempts to catch up with successively more distant (and therefore more rapidly receding) neighbours. In the absence of gravity, we get $\delta v \propto 1/a$: momentum redshifts away, just as with photon energy.

The peculiar velocity is directly related to the evolution of the density field, through conservation of mass. This is described by the usual continuity equation $\dot{\rho} = -\boldsymbol{\nabla}\cdot(\rho\mathbf{v}$, where $\rho = \bar{\rho}(1+\delta)$ and proper length units are assumed. If we use comoving length units, the mean density is constant and this is easily transformed to

$$\dot{\delta} = -\boldsymbol{\nabla}\cdot[(\mathbf{1}+\delta)\mathbf{u}]. \tag{181}$$

This simplifies further if we restrict ourselves to a the linear approximation where $\delta \ll 1$, and neglect terms that are second order in the perturbation, yielding the linearized continuity equation

$$\dot{\delta} = -\boldsymbol{\nabla}\cdot\mathbf{u}. \tag{182}$$

The solutions of these equations can be decomposed into modes either parallel to $\mathbf{g}$ or independent of $\mathbf{g}$ (these are the homogeneous and inhomogeneous solutions to the equation of motion). The homogeneous case corresponds to no peculiar gravity – i.e. zero density perturbation. This is consistent with the linearized continuity equation, $\boldsymbol{\nabla} \cdot \mathbf{u} = -\dot{\delta}$, which says that it is possible to have **vorticity modes** with $\boldsymbol{\nabla} \cdot \mathbf{u} = 0$ for which $\dot{\delta}$ vanishes, so there is no growth of structure in this case. The proper velocities of these vorticity modes decay as $v = au \propto a^{-1}$, as with the kinematic analysis for a single particle.

GROWING MODE For the growing mode, it is most convenient to eliminate $\mathbf{u}$ by taking the divergence of the equation of motion for $\mathbf{u}$, and the time derivative of the continuity equation. This requires a knowledge of $\boldsymbol{\nabla} \cdot \mathbf{g}$, which comes via Poisson's equation: $\boldsymbol{\nabla} \cdot \mathbf{g} = 4\pi G a \rho_0 \delta$. The resulting 2nd-order equation for δ is

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G \rho_0 \, \delta. \tag{183}$$

This is easily solved for the $\Omega_m = 1$ case, where $4\pi G\rho_0 = 3H^2/2 = 2/3t^2$, and a power-law solution works:

$$\delta(t) \propto t^{2/3} \quad \text{or} \quad t^{-1}. \tag{184}$$

The first solution, with $\delta(t) \propto a(t)$ is the growing mode, corresponding to the gravitational instability of density perturbations. Given some small initial seed fluctuations, this is the simplest way of creating a universe with any desired degree of inhomogeneity.

RADIATION-DOMINATED UNIVERSE The analysis so far does not apply when the universe was radiation dominated ($c_s = c/\sqrt{3}$). For this period of the early Universe it is therefore common to resort to general relativity perturbation theory or use special relativity fluid mechanics and Newtonian gravity with a relativistic source term (see e.g. Section 15.2 of Peacock 1999). In the interests of brevity and completeness we simply quote the result of this analysis, which is

$$\delta(t) \propto t \quad \text{or} \quad t^{-1}; \tag{185}$$

thus the growing mode during radiation domination ($a \propto t^{1/2}$) has $\delta(t) \propto a(t)^2$.

The results for matter domination and radiation domination can be combined to say that gravitational potential perturbations are independent of time (at least while $\Omega = 1$). Poisson's equation tells us that $-k^2\Phi/a^2 \propto \rho\,\delta$; since $\rho \propto a^{-3}$ for matter domination or a^{-4} for radiation, that gives $\Phi \propto \delta/a$ or δ/a^2 respectively, so that

$$\Phi = \text{constant} \tag{186}$$

in either case. In other words, the metric fluctuations resulting from potential perturbations are frozen, at least for perturbations with wavelengths greater than the horizon size. This simple result at least has an intuitive plausibility for the radiation-dominated case, despite the lack of a derivation.

MODELS WITH NON-CRITICAL DENSITY We have solved the growth equation for the matter-dominated $\Omega = 1$ case. It is possible to cope with other special cases (e.g. matter + curvature) with some effort. In the general case (especially with a general vacuum having $w \neq -1$), it is necessary to integrate the differential equation numerically. At high z, we always have the matter-dominated $\delta \propto a$, and this serves as an initial condition. In general, we can write

$$\delta(a) \propto a\, f[\Omega_m(a)], \tag{187}$$

where the factor f expresses a deviation from the simple growth law. For flat models with $\Omega_m + \Omega_v = 1$, a useful approximation is $f \simeq \Omega_m^{0.23}$, which is less marked than $f \simeq \Omega_m^{0.65}$ in the $\Lambda = 0$ case. This reflects the more rapid variation of Ω_v with redshift; if the cosmological constant is important dynamically, this only became so very recently, and the universe spent more of its history in a nearly Einstein–de Sitter state by comparison with an open universe of the same Ω_m. Interestingly, this difference is erased if we look at the growth *rate* , in which case we have the almost universal formula

$$f_g \equiv \frac{d\ln\delta}{d\ln a} = \Omega_m(a)^\gamma, \tag{188}$$

where γ is close to 0.6, independently of whether there is significant vacuum energy.

6.2 Pressure and the shape of the matter power spectrum

So far, we have mainly considered the collisionless component. For the photon-baryon gas, all that changes is that the peculiar acceleration gains a term from the pressure gradients:

$$\mathbf{g} = -\boldsymbol{\nabla}\Phi/a - \boldsymbol{\nabla}p/(a\rho). \tag{189}$$

The pressure fluctuations are related to the density perturbations via the sound speed

$$c_s^2 \equiv \frac{\partial p}{\partial \rho}. \tag{190}$$

Now think of a plane-wave disturbance $\delta \propto e^{-i\mathbf{k}\cdot\mathbf{r}}$, where $\mathbf{k}$ and $\mathbf{r}$ are in comoving units. All time dependence is carried by the amplitude of the wave. The linearized equation of motion for δ then gains an extra term on the rhs from the pressure gradient:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \delta\big(4\pi G\rho_0 - c_s^2 k^2/a^2\big). \tag{191}$$

This shows that there is a critical proper wavelength, the **Jeans length**, at which we switch from the possibility of gravity-driven growth for long-wavelength modes to standing sound waves at short wavelengths. This critical length is

$$\lambda_{\rm J}^{\rm proper} = \frac{2\pi}{k_{\rm J}^{\rm proper}} = c_s\sqrt{\frac{\pi}{G\rho}}. \tag{192}$$

Prior to matter-radiation equality, the speed of sound for a radiation-dominated fluid with $p = u/3$ is $c_s = c/\sqrt{3}$, so this Jeans length is close to the horizon size.

A perturbation of given comoving wavelength will start out larger than the horizon, but the horizon grows with time, and so the perturbation 'enters the horizon' (not very good terminology, but standard). After that time, pressure forces dominate over gravity, and the perturbation oscillates as a standing sound wave, which turns out to have a constant amplitude in δ during the radiation era; but zero-pressure growth would result in δ increasing with time, so the amplitude of small-scale fluctuations falls relative to the growing mode. Since the growing mode corresponds to constant potential, this means that the potential decays as $1/\delta_{\rm grow} \propto 1/a^2$. All this can be solved exactly for the radiation-dominated era, and the effective damping of the initial potential fluctuation is

$$\frac{\Phi}{\Phi_i} = 3(\sin x - x\cos x)/x^3; \quad x \equiv kc_s\eta, \tag{193}$$

where η is conformal time, $d\eta \equiv dt/a(t)$, which is thus equal to the comoving particle horizon size. We will see later that the imprint of these **acoustic oscillations** is visible in the microwave background.

At these early times, the dark matter is a minority constituent of the universe, but it suffers an interesting and critical effect from the above behaviour of the coupled baryon-photon fluid (which is glued together by Thomson scattering). The small-scale damping of the waves results in the radiation becoming smooth, which breaks the usual adiabatic relation in which the matter density and the photon number density have equal perturbations. Now recall the growth equation for matter perturbations, neglecting pressure:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G\rho_0\,\delta. \tag{194}$$

The rhs contains the combination $\rho_0\delta$; this is just the fluctuation in density, which drives the gravitational growth. But the growing mode with Φ constant only arises if ρ_0 is the total density, which also sets the timescale for expansion. If the majority constituent of the universe (the radiation, at early times) is uniform, then the rhs becomes $\propto \rho_m\delta$, which is $\ll \rho_0\delta$. Thus the growth switches off.

Figure 14 shows a schematic of the resulting growth history for matter density fluctuations. For scales greater than the horizon, perturbations in matter and radiation can grow together, so fluctuations at early times grow at the same rate, independent of wavenumber. But this growth ceases once the perturbations 'enter the horizon' – i.e. when the horizon grows sufficiently to exceed the perturbation wavelength. At this point, growth ceases. For fluids (baryons) it is the radiation pressure that prevents the perturbations from collapsing further. For collisionless matter the rapid radiation driven expansion prevents the perturbation from growing again until matter radiation equality.

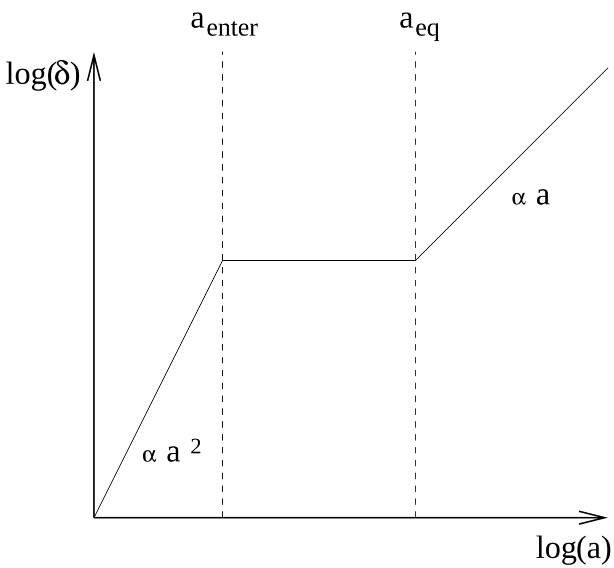

Figure 14. A schematic of the suppression of fluctuation growth during the radiation dominated phase when the density perturbation enters the horizon at $a_{\rm enter} < a_{\rm eq}$.

This effect (called the Mészáros effect) is critical in shaping the late-time power spectrum (as we will show) as the universe preserves a 'snapshot' of the amplitude of the mode at horizon crossing. Before this process operates, inflation predicts an approximately scale invariant initial Zeldovich spectrum where $P_i(k) \propto k$. How does the Mészáros effect modify the shape of this initial power spectrum?

Figure 15 shows that the smallest physical scales (largest k scales) will be affected first and experience the strongest suppression to their amplitude. The largest physical scale fluctuations (smallest k scales) will be unaffected as they will enter the horizon after matter-radiation equality. We can therefore see that there will be a turnover in the power spectrum at a characteristic scale given by the horizon size at matter-radiation equality.

From Figure 14 we can see that when a fluctuation enters the horizon before matter-radiation equality its growth is suppressed by $f = (a_{\rm enter}/a_{\rm eq})^2$. A fluctuation k enters the horizon when

$D_{\rm H} \simeq 1/k$. As $D_{\rm H} = c/aH(a)$ and $H(a) \propto a^{-2}$ during radiation domination we see that the fluctuations are suppressed by a factor $f \propto k^{-2}$ and that the power spectrum on large k scales follows a k^{-3} power law.

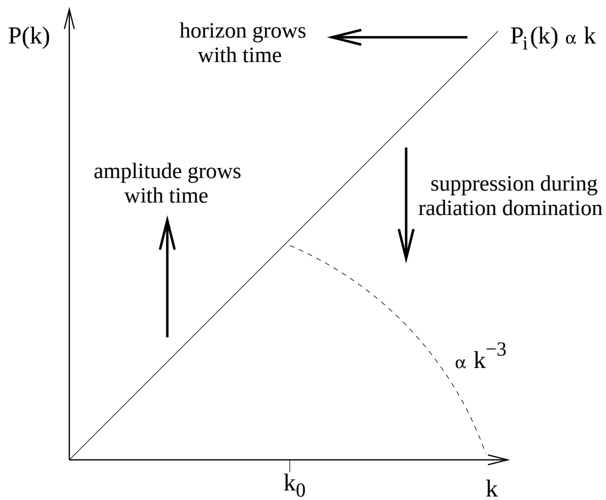

Figure 15. Schematic of the how the Mészáros effect modifies the initial power spectrum. Note log scale.

6.3 Transfer functions and characteristic scales

The above discussion can be summed up in the from of the linear **transfer function** for density perturbations, where we factor out the long-wavelength growth law from a term that expresses how growth is modulated as a function of wavenumber:

$$\delta(a) \propto g(a) T_k. \tag{195}$$

In principle, there is a transfer function for each constituent of the universe, and these evolve with time. As we have discussed, however, the different matter ingredients tend to come together at late times, and the overall transfer function tends to something that is the same for all matter components and which does not change with time for low redshifts. This late-time transfer function is therefore an important tool for cosmologists who want to predict observed properties of density fields in the current universe.

We have discussed the main effects that contribute to the form of the transfer function, but a full calculation is a technical challenge. In detail, we have a mixture of matter (both collisionless dark particles and baryonic plasma) and relativistic particles (collisionless neutrinos and collisional photons), which does not behave as a simple fluid. Particular problems are caused by the change in the photon component from being a fluid tightly coupled to the baryons by Thomson scattering, to being collisionless after recombination. Accurate results require a solution of the Boltzmann equation to follow the evolution of the full phase-space distribution. This was first computed accurately by Bond & Szalay (1983), and is today routinely available via public-domain codes such as CMBFAST.

Some illustrative results are shown in figure 16. Leaving aside the isocurvature models, all adiabatic cases have $T \rightarrow 1$ on large scales – i.e. there is growth at the universal rate (which is such that the amplitude of potential perturbations is constant until the vacuum starts to be important at $z \lesssim 1$). The different shapes of the functions can be understood intuitively in terms of a few special length scales, as follows:

(1) Horizon length at matter-radiation equality. The main bend visible in all transfer functions is due to the Mészáros effect (discussed above), which arises because the universe is radiation dominated at early times.

$$T_k \simeq \begin{cases} 1 & kD_{\rm H}(z_{\rm eq}) \ll 1 \\ [kD_{\rm H}(z_{\rm eq})]^{-2} & kD_{\rm H}(z_{\rm eq}) \gg 1. \end{cases} \tag{196}$$

This process continues until the universe becomes matter dominated. We therefore expect a characteristic 'break' in the fluctuation spectrum around the comoving horizon length at this time, which we have seen is $D_{\rm H}(z_{\rm eq}) = 16\,(\Omega_m h^2)^{-1}{\rm Mpc}$. Since distances in cosmology always scale as h^{-1}, this means that $\Omega_m h$ should be observable.

(2) Free-streaming length. This relatively gentle filtering away of the initial fluctuations is all that applies to a universe dominated by Cold Dark Matter, in which random velocities are negligible. A CDM universe thus contains fluctuations in the dark matter on all scales, and structure formation proceeds via hierarchical process in which nonlinear structures grow via mergers. Examples of CDM would be thermal relic WIMPs with masses of order 100 GeV, but a more interesting case arises when thermal relics have lower masses. For collisionless dark matter,

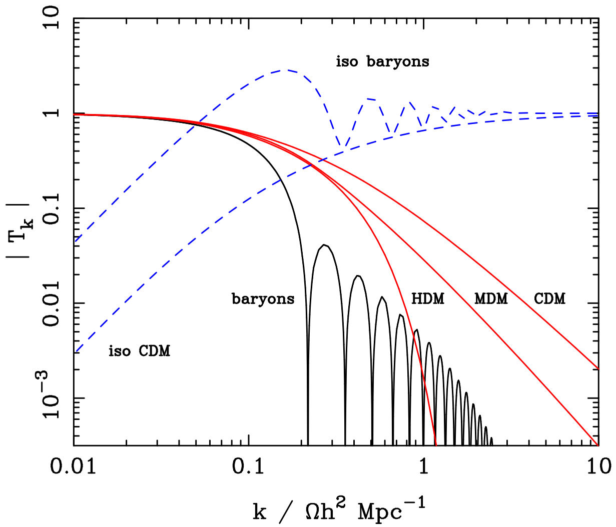

Figure 16. A plot of transfer functions for various adiabatic models, in which $T_k \to 1$ at small k. A number of possible matter contents are illustrated: pure baryons; pure CDM; pure HDM. For dark-matter models, the characteristic wavenumber scales proportional to $\Omega_m h^2$, marking the break scale corresponding to the horizon length at matter-radiation equality. The scaling for baryonic models does not obey this exactly; the plotted case corresponds to $\Omega_m = 1$, $h = 0.5$.

perturbations can be erased simply by free streaming: random particle velocities cause blobs to disperse. At early times ($kT > mc^2$), the particles will travel at c, and so any perturbation that has entered the horizon will be damped. This process switches off when the particles become non-relativistic, so that perturbations are erased up to proper lengthscales of $\simeq ct(kT = mc^2)$. This translates to a comoving horizon scale ($2ct/a$ during the radiation era) at $kT = mc^2$ of

$$L_{\rm free-stream} = 112\,(m/{\rm eV})^{-1}\,{\rm Mpc} \tag{197}$$

(in detail, the appropriate figure for neutrinos will be smaller by $(4/11)^{1/3}$ since they have a smaller temperature than the photons). A light neutrino-like relic that decouples while it is relativistic satisfies

$$\Omega_\nu h^2 = m/94.1\,{\rm eV} \tag{198}$$

Thus, the damping scale for HDM (Hot Dark Matter) is of order the bend scale. The existence of galaxies at $z \simeq 6$ tells us that the coherence scale must have been below about 100 kpc, so the DM mass must exceed about 1 keV.

A more interesting (and probably practically relevant) case is when the dark matter is a mixture of hot and cold components. The free-streaming length for the hot component can therefore be very large, but within range of observations. The dispersal of HDM fluctuations reduces the CDM growth rate on all scales below $L_{\rm free-stream}$ – or, relative to small scales, there is an enhancement in large-scale power.

(3) Acoustic horizon length. The horizon at matter-radiation equality also enters in the properties of the baryon component. Since the sound speed is of order c, the largest scales that can undergo a single acoustic oscillation are of order the horizon at this time. The transfer function for a pure baryon universe shows large modulations, reflecting the number of oscillations that have been completed before the universe becomes matter dominated and the pressure support drops. The lack of such large modulations in real data is one of the most generic reasons for believing in collisionless dark matter. Acoustic oscillations persist even when baryons are subdominant, however, and can be detectable as lower-level modulations in the transfer function. At matter-radiation equality, the dark matter has a smoothly declining transfer function – but the baryons have an oscillating transfer function, so the spatial distribution of these two components is different. Once the sound speed drops, gravity will pull these components together, and their transfer functions will tend to become identical. But since baryons are about 5% of the total matter content, the resulting final 'compromise' transfer function has acoustic oscillations imprinted into it at the few per cent level. We will say more about this later.

SPECTRUM NORMALIZATION We now have a full recipe for specifying the matter power spectrum: Historically, this is done in a slightly awkward way. First suppose we wanted to consider smoothing the density field by convolution with some **window**. One simple case is to imagine averaging within a sphere of radius R. For the effect on the power spectrum, we need the Fourier transform of this filter:

$$\sigma^2(R) = \int \Delta^2(k)\, |W_k|^2\, d\ln k; \quad W_k = \frac{3}{(kR)^3}(\sin kR - kR\cos kR). \tag{199}$$

Unlike the power spectrum, $\sigma(R)$ is monotonic, and the value at any scale is sufficient to fix the normalization. The traditional choice is to specify σ_8, corresponding to $R = 8\,h^{-1}\,\mathrm{Mpc}$. As a final complication, this measure is normally taken to apply to the rms in the filtered *linear-theory* density field. The best current estimate is $\sigma_8 \simeq 0.8$, so clearly nonlinear corrections matter in interpreting this number. The virtue of this convention is that it is then easy to calculate the spectrum normalization at any early time.

7 Structure formation – II

The equations of motion are nonlinear, and we have only solved them in the limit of linear perturbations. We now discuss evolution beyond the linear regime, first considering the full numerical solution of the equations of motion, and then a key analytic approximation by which the 'exact' results can be understood.

N-BODY MODELS The exact evolution of the density field is usually performed by means of an **N-body simulation**, in which the density field is represented by the sum of a set of fictitious discrete particles. We need to solve the equations of motion for each particle, as it moves in the gravitational field due to all the other particles. Using comoving units for length and velocity ($\mathbf{v} = a\mathbf{u}$), we have previously seen the equation of motion

$$\frac{d}{dt}\mathbf{u} = -2\frac{\dot{a}}{a}\mathbf{u} - \frac{1}{a^2}\boldsymbol{\nabla}\Phi, \tag{200}$$

where Φ is the Newtonian gravitational potential due to density perturbations. The time derivative is already in the required form of the convective time derivative observed by a particle, rather than the partial $\partial/\partial t$.

In outline, this is straightforward to solve, given some initial positions and velocities. Defining some timestep dt, particles are moved according to $d\mathbf{x} = \mathbf{u}\, dt$, and their velocities updated according to $d\mathbf{u} = \dot{\mathbf{u}}\, dt$, with $\dot{\mathbf{u}}$ given by the equation of motion (in practice, more sophisticated time integration schemes are used). The hard part is finding the gravitational force, since this involves summation over $(N-1)$ other particles each time we need a force for one particle. All the craft in the field involves finding clever ways in which all the forces can be evaluated in less than the raw $O(N^2)$ computations per timestep. We will have to omit the details of this, unfortunately, but one obvious way of proceeding is to solve Poisson's equation on a mesh using a Fast Fourier Transform. This can convert the $O(N^2)$ time scaling to $O(N \ln N)$, which is a qualitative difference given that N can be as large as 10^{10}.

These non-linear effects boost the amplitude of the power spectrum at small physical scales (large k scales) as can be seen in Figure 17 . For cosmological observations we need to understand

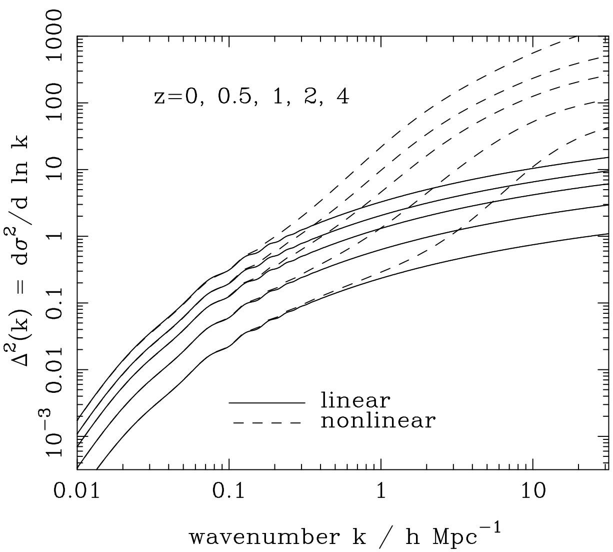

Figure 17. ΛCDM power spectrum normalised by $\sigma_8 = 0.9$. The linear power spectrum is show solid and the non-linear power spectrum is shown dashed using the fitting formula from Smith et al 2003.

these non-linear effects to high precision. This is one of the issues facing modern day cosmology and non-linear effects can only be calculated through large scale suites of HPC N-body simulations.

THE SPHERICAL MODEL *N*-body models can yield evolved density fields that are nearly exact solutions to the equations of motion, but working out what the results mean is then more a question of data analysis than of deep insight. Where possible, it is important to have analytic models that guide the interpretation of the numerical results. The most important model of this sort is the spherical density perturbation, which can be analysed immediately using the tools developed for the Friedmann models, since Birkhoff's theorem tells us that such a perturbation behaves in exactly the same way as part of a closed universe. The equations of motion are the same as for the scale factor, and we can therefore write down the **cycloid solution** immediately. For a matter-dominated universe, the relation between the proper radius of the sphere and time is

$$\begin{aligned} r &= A(1-\cos\theta) \\ t &= B(\theta-\sin\theta). \end{aligned} \tag{201}$$

It is easy to eliminate θ to obtain $\ddot{r} = -GM/r^2$, and the relation $A^3 = GMB^2$ (use e.g. $\dot{r} = (dr/d\theta)/((dt/d\theta)$, which gives $\dot{r} = [A/B]\sin\theta/[1-\cos\theta]$). Expanding these relations up to order θ^5 gives $r(t)$ for small t:

$$r \simeq \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left[1-\frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right], \tag{202}$$

and we can identify the density perturbation within the sphere:

$$\delta \simeq \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{203}$$

This all agrees with what we knew already: at early times the sphere expands with the $a \propto t^{2/3}$ Hubble flow and density perturbations grow proportional to a.

We can now see how linear theory breaks down as the perturbation evolves. There are three interesting epochs in the final stages of its development, which we can read directly from the above solutions. Here, to keep things simple, we compare only with linear theory for an $\Omega = 1$ background.

(1) **Turnround**. The sphere breaks away from the general expansion and reaches a maximum radius at $\theta = \pi$, $t = \pi B$. At this point, the true density enhancement with respect to the background is just $[A(6t/B)^{2/3}/2]^3/r^3 = 9\pi^2/16 \simeq 5.55$.

(2) **Collapse**. If only gravity operates, then the sphere will collapse to a singularity at $\theta = 2\pi$.

(3) **Virialization**. Clearly, collapse to a point is highly idealized. Consider the time at which the sphere has collapsed by a factor 2 from maximum expansion ($\theta = 3\pi/2$). At this point, it has kinetic energy K related to potential energy V by $V = -2K$. This is the condition for equilibrium, according to the **virial theorem**. Conventionally, it is assumed that this