Dr. Jean-Luc Lehners

Early Universe Cosmology
Autumn Blockkurs 2014, Berlin

Contents

1 Lecture I

1.1 Big Bang Cosmology

1.1.1 The Cosmological Principle

Cosmology is the science of the universe treated as a whole. It is the science that tries to understand the origin (if there was one!), the shape and the evolution of the universe, including predictions for its future. As such, it also tries to determine what the place of humans is in the universe. It is the only science where the object under study cannot be separated from "the rest", and this feature leads to fundamentally different challenges that cosmology must face.

The universe cannot be reproduced in the laboratory, and moreover we are severely limited in our ability to travel across the universe. In fact, all our knowledge about the universe must be inferred from the information that arrives here on Earth now! This, however, is not as limiting as it may sound, because of the finiteness of the speed of light. Light originating from distant places, but which we are observing now, shows us these distant places not as they are now, but as they were when the light was emitted. In this way, when looking further and further into the universe, we see further and further into the past of the universe, and this allows us to reconstruct (up to a point) the history of our universe. Of course, we do not directly observe our past, but rather the past of increasingly distant regions. This strategy then allows us to reconstruct our history to the extent that the universe is evolving in the same way in different places, *i.e.* to the extent that the universe is spatially *homogeneous*.

Observations of the distribution of galaxies in the universe show that these are distributed along filaments, with huge voids of over 100 million light-years across in between; see Fig. 1. These observations also show that on even bigger scales, namely bigger than about 300 million light years, the universe becomes very similar in all directions. In other words, on these scales the universe becomes *isotropic*.

Thus, in part motivated by observations, and in part by wishful thinking, we are led to postulate the *Cosmological Principle*, namely that the universe is spatially *homogeneous* and *isotropic* on the largest scales. Some comments:

- If the cosmological principle was not at least approximately true, we would know very little about the distant universe. Only the observation that the universe appears to obey the same laws, and have the same rough properties, over large distances has allowed us to gain knowledge about the behavior of the universe on the largest observable scales.

- The cosmological principle certainly breaks down on small scales (we are living proof!), but it may very well also break down on scales much larger than the currently observable part of the universe - later on, when discussing different models of the universe, we will return to this topic.

- The cosmological principle treats space and time on different footings! This is because the universe is only homogeneous and isotropic in a frame which is *co-moving* with typical galaxies. In other words, in the universe there is a preferred frame, and the universe only appears isotropic to the extent that one is at rest with respect to that frame. A further subtlety is that the existence of a preferred frame does not imply that

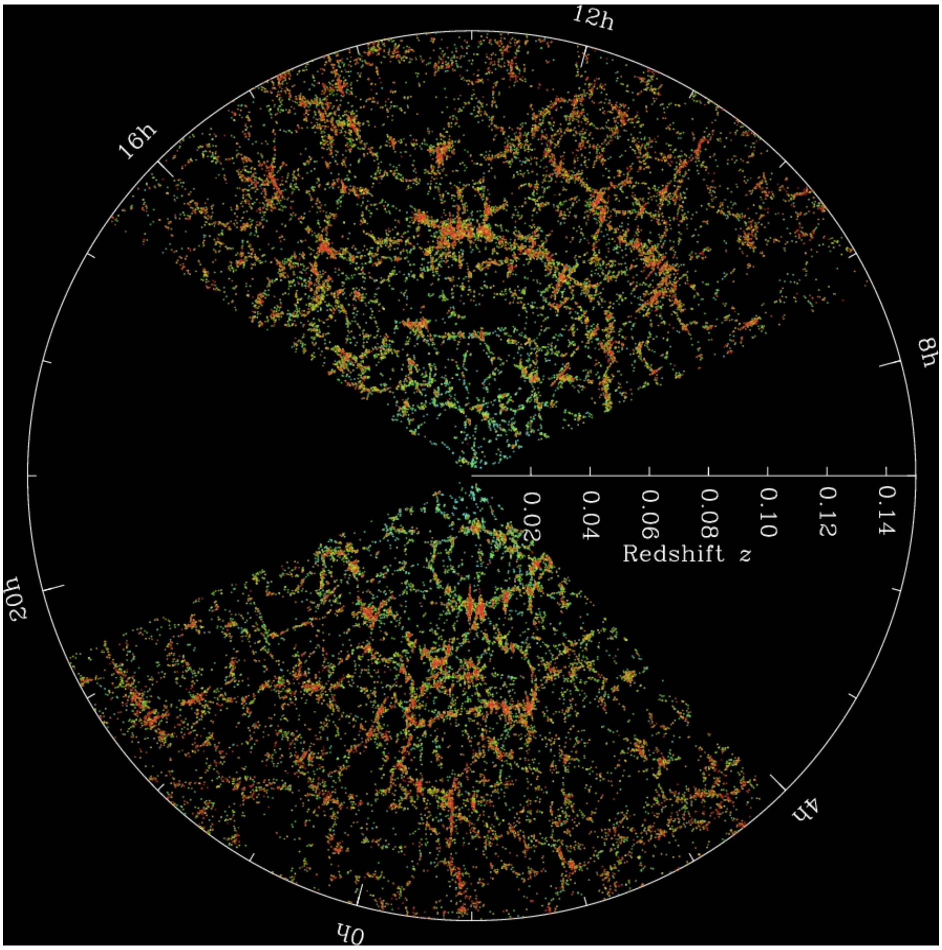

Figure 1: The distribution of galaxies in two wedges centered on Earth, according to the Sloan Digital Sky Survey. Each dot on the figure represents the location of a galaxy, and redshift 0.01 corresponds to a distance of about 140 million light years. One of the goals of cosmology is to understand the statistical properties of the distribution of galaxies in the universe. *Credit: M. Blanton and the Sloan Digital Sky Survey*

far-away galaxies must be approximately at rest with respect to each other - on the contrary, there is overwhelming evidence that the universe is expanding. The preferred frame can then be imagined to follow (and largely to be defined by) this expansion.

- Despite its name, the cosmological principle is simply an observed feature of our universe, and does not stem from fundamental physical principles. It is one of the main aims of early universe cosmology to find an *explanation* for why the universe is so homogeneous and isotropic.

1.1.2 Robertson-Walker Universes

We can use general relativity to describe the evolution of the universe, as general relativity is the (classical) dynamical theory of space, time and matter. We will see later how quantum theory has likely played an important role too in shaping the universe, but quantum effects only need to be taken into account at special times in the history of the universe.

At first, we will assume exact spatial homogeneity and isotropy. This is of course an idealization, but the cosmological principle implies that this should be a good approximation in describing the universe on large scales. This assumption is very restrictive, and in fact one can prove that it allows for only three distinct spatial geometries, namely flat 3-dimensional space, a 3-sphere (with constant positive curvature) or a hyperbolic 3-sphere (with constant negative curvature). We denote the 3-dimensional line element by $\mathrm{d}l^2$. Then, using cartesian coordinates x, y, z, the line element of flat space is simply given by

$$\mathrm{d}l^2_{\text{flat}} = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{1.1}$$

In polar coordinates defined via $x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi, z = r\cos\theta$, the line element becomes

$$\mathrm{d}l^2_{\text{flat}} = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2). \tag{1.2}$$

We can obtain the metric for the 3-sphere by using the fact that it can be embedded in 4-dimensional flat space. If we denote the fourth spatial coordinate by w, then the line element for 4-dimensional euclidean space is $\mathrm{d}l_4^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \mathrm{d}w^2$, and within this space the unit 3-sphere is defined via the equation

$$x^2 + y^2 + z^2 + w^2 = 1. \tag{1.3}$$

Differentiating, we learn that $x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z + w\mathrm{d}w = 0$, or

$$\mathrm{d}w = \pm\frac{x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z}{\sqrt{1 - x^2 - y^2 - z^2}}. \tag{1.4}$$

We can now obtain the metric for the 3-sphere by plugging this expression for $\mathrm{d}w$ into the 4-dimensional line element, thus restricting the line element to the 3-sphere. The result is

$$\mathrm{d}l^2_{3-\text{sphere}} = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \frac{(x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z)^2}{1 - x^2 - y^2 - z^2}. \tag{1.5}$$

In a similar way one can obtain the metric of the 3-dimensional hypersphere by embedding the hyperboloid $x^2 + y^2 + z^2 - w^2 = -1$ into 4-dimensional Minkowski space $\mathrm{d}l_4^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - \mathrm{d}w^2$. All three metrics can be conveniently summarized by

$$\mathrm{d}l_3^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \frac{k(x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z)^2}{1 - k(x^2 + y^2 + z^2)} \equiv g_{(3)ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{1.6}$$

or, in polar coordinates,

$$\mathrm{d}l_3^2 = \frac{1}{1 - kr^2}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2), \tag{1.7}$$

where $k = 0, +1, -1$ for the flat, spherical and hyperspherical spaces respectively. Above, we derived the metric on the 3-sphere only for the case where it has unit radius. We can allow for an arbitrary, and in general time-dependent, size by multiplying the line element by a factor

$a^2(t)$, and likewise for the hypersphere. In the case of flat space, the multiplicative factor $a^2(t)$ has no absolute meaning, and only relative values at different times are meaningful. The full 4-dimensional *Robertson-Walker* metric can then be obtained by adding a time direction, yielding

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)[\frac{1}{1-kr^2}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)]. \tag{1.8}$$

The function $a(t)$ is known as the *scale factor* of the universe. Its meaning can be clarified by using the metric above to calculate the spatial distance between an observer at the origin $r = 0$ and one at $r = R$ (at fixed time t, and with the angular coordinates also being fixed), which is given by

$$\int_{r=0}^{r=R} \mathrm{d}s = \int_{r=0}^{r=R} a(t)\frac{1}{\sqrt{1-r^2}}\mathrm{d}r = a(t)\begin{cases} \text{arcsinh}\, R & (k=-1) \\ R & (k=0) \\ \arcsin R & (k=+1) \end{cases} \tag{1.9}$$

thus observers at fixed r, θ, ϕ only change their relative distance via the global, space-independent scale factor. In other words, the scale factor describes the overall expansion or contraction of the universe. The coordinates r, θ, ϕ are so-called *comoving* coordinates, and the result above implies that the proper distance between two comoving observers evolves as $a(t)$. Moreover, for a comoving observer, proper time is measured according to $\int \sqrt{-\mathrm{d}s^2} = \int \mathrm{d}t = \Delta t$, which shows that the time coordinate t is simply the *proper time* measured by comoving observers. Sometimes, it is convenient to define a *conformal time* τ via $\mathrm{d}t = a\mathrm{d}\tau$, so that the metric has an overall factor of a^2:

$$\mathrm{d}s^2 = a^2(\tau)[-\mathrm{d}\tau^2 + \frac{1}{1-kr^2}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)]. \tag{1.10}$$

The general form of the Robertson-Walker metric allows us to understand a crucial effect in cosmology, namely the shifting of spectral lines observed in the light coming from distant galaxies. When the scale factor is non-constant, the frequency of light is in general different at the time of emission than at the time of observation. This can be seen as follows: we assume that we are observing (at time t_0) the light originating (at time t_R) from a galaxy located at $r = R$. Our own position can be chosen to be at $r = 0$. Since light rays propagate according to $\mathrm{d}s^2 = 0$, we obtain from Eq. (1.8) that

$$\mathrm{d}t = -a(t)\frac{\mathrm{d}r}{\sqrt{1-kr^2}}, \tag{1.11}$$

where the minus sign was chosen because the light is coming towards us along decreasing values of r. Hence the times of emission and observation are related to the coordinate separation R by

$$\int_{t_0}^{t_R} \frac{\mathrm{d}t}{a(t)} = \int_0^R \frac{\mathrm{d}r}{\sqrt{1-kr^2}}. \tag{1.12}$$

Subsequent wave crests of the light wave are emitted at time intervals Δt_R, if the frequency of the light is $\nu_R = 1/\Delta t_R$. Then, since the right-hand side of the equation above is independent of time, the differential of the equation implies that

$$\frac{\Delta t_R}{a(t_R)} = \frac{\Delta t_0}{a(t_0)}. \tag{1.13}$$

In other words, the frequency at observation is related to the frequency at emission via

$$\nu_0 = \nu_R \frac{a(t_R)}{a(t_0)}. \tag{1.14}$$

Equivalently, the wavelength at observation λ_0 is shifted from the wavelength at emission λ_R via

$$\lambda_0 = \lambda_R \frac{a(t_0)}{a(t_R)} \equiv \lambda_R(1+z). \tag{1.15}$$

In an expanding universe, the wavelength at observation is longer than at absorption, $z > 0$, and the light is *redshifted.* In a contracting universe, light decreases in wavelength, $z < 0$, and in that case light is *blueshifted.* Astronomical observations, first carried out by Vesto Slipher in the 1920s and to the required accuracy by Edwin Hubble in the 1930s, show that the light from almost all galaxies is redshifted, thus providing evidence that our universe is expanding. Only the light from a few nearby galaxies (such as the Andromeda galaxy) is blueshifted, which can be understood by the strong gravitational interactions between these galaxies and the Milky Way, causing them to have significant *peculiar velocities* superimposed on the comoving *Hubble flow.* For nearby objects, it is straightforward to show that the redshift increases linearly with proper distance d (for nearby objects, $d \approx t_R - t_0$)

$$z \approx H_0 d, \tag{1.16}$$

where the constant of proportionality is the *Hubble parameter*

$$H \equiv \frac{\dot{a}}{a}, \tag{1.17}$$

and the subscript 0 customarily refers to a quantity evaluated at the present time.

We should stress that the Robertson-Walker metric is only intended as a zeroth approximation to the coarse-grained structure of the universe. In thinking about the expansion/contraction of space, it is important to keep in mind that near localized sources (such as stars) the metric is locally not of Robertson-Walker form, but rather of Schwarzschild form. In the Schwarzschild metric, space is not expanding, and, as a consequence, in our solar system space is neither expanding nor contracting. As an improved approximation, one could think of the entire universe as a kind of “Swiss cheese”, with the cheese being well described by the Robertson-Walker metric, and the (small) holes by the Schwarzschild metric. This avoids some standard misconceptions about the expansion of space.

1.1.3 The Friedmann Equations and the Big Bang

The cosmological principle led to the Robertson-Walker form of the metric, Eq. (1.8). In euclidean coordinates this metric can also be written as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t) g_{(3)ij} \mathrm{d}x^i \mathrm{d}x^j, \tag{1.18}$$

where the spatial 3-metric $g_{(3)}$ was defined in Eq. (1.6). It is a straightforward exercise to derive the associated connections and curvature tensors:

$$\Gamma^0_{ij} = \tfrac{\dot{a}}{a} g_{ij} = a\dot{a} g_{(3)ij} \tag{1.19}$$
$$\Gamma^k_{0j} = \tfrac{\dot{a}}{a} \delta^k_j \tag{1.20}$$
$$\Gamma^k_{ij} = \Gamma^k_{(3)ij} = k g_{(3)ij} x^k \tag{1.21}$$
$$\Gamma^0_{00} = \Gamma^0_{0j} = \Gamma^k_{00} = 0 \tag{1.22}$$
$$R_{00} = -3\tfrac{\ddot{a}}{a} \tag{1.23}$$
$$R_{ij} = (\tfrac{\ddot{a}}{a} + 2\tfrac{\dot{a}^2}{a^2} + 2\tfrac{k}{a^2}) g_{ij} = (a\ddot{a} + 2\dot{a}^2 + 2k) g_{(3)ij} \tag{1.24}$$
$$R_{0j} = 0 \tag{1.25}$$
$$R = 6(\tfrac{\ddot{a}}{a} + \tfrac{\dot{a}^2}{a^2} + \tfrac{k}{a^2}). \tag{1.26}$$

The assumptions of spatial isotropy and homogeneity also lead to strong restrictions on the form of the energy-momentum tensor on large scales. Isotropy, *i.e.* invariance under spatial rotations, implies that (at a co-moving point) the spatial part T_{ij} of the energy-momentum tensor should be proportional to the metric g_{ij}, and that the mixed, vectorial component t_{0i} should vanish. Homogeneity means that we must impose these conditions everywhere in space, implying that the function of proportionality in T_{ij} can depend on time alone. Likewise, T_{00} can depend on time alone. The conventional definitions are

$$T_{00} \equiv \rho(t) \quad T_{0i} = 0 \quad T_{ij} \equiv p(t) g_{ij}, \tag{1.27}$$

where $\rho(t)$ is the proper *energy density* and $p(t)$ the *pressure*. This particular form of the energy-momentum tensor corresponds to that of an ideal fluid. The energy density and the pressure are related to each other via the *equation of state*

$$p = w\rho. \tag{1.28}$$

In many cases of interest, w is constant, and we will assume this to be the case unless otherwise noted. In general relativity, the Bianchi identity for the Einstein tensor ($G^{\mu\nu}{}_{;\nu} = 0$) requires that the energy-momentum tensor be (covariantly) conserved, *i.e.* $T^{\mu\nu}{}_{;\nu} = 0$. The $\mu = i$ component of this equation corresponds to momentum conservation, and it is automatically satisfied in a Robertson-Walker background. The $\mu = 0$ component leads to

$$\dot{\rho} + 3H(\rho + p) = 0. \tag{1.29}$$

This equation is known as the *equation of continuity*. Note that it shows that in general, in a dynamical universe energy is *not* conserved. For a constant equation of state, the equation of continuity can be integrated to yield

$$\rho \propto \frac{1}{a^{3(1+w)}}. \tag{1.30}$$

Thus, if we know the equation of state of a certain matter component, then we know how its energy density (and thus its relative importance) scales as the universe contracts or expands. The most commonly considered ideal fluids are

- Pressure free matter/dust, $w = 0$. Ordinary baryonic matter as well as dark matter fall into this category. Their energy density simply scales inversely to the volume of a given region of space, $\rho \propto a^{-3}$, as expected.

- Relativistic particles/radiation, $w = \frac{1}{3}$. The energy density of radiation or a gas of relativistic particles scales as a^{-4}. This means that in an expanding universe, the energy density of radiation falls off faster than the volume of space increases. Heuristically one can understand the exponent -4 as follows: the number of photons in a given co-moving volume scales with the volume, as a^{-3}. However, the wavelength of a photon also scales with the scale factor, and thus its frequency/energy scale as a^{-1}. This additional factor of a^{-1} then leads to the overall scaling as a^{-4}.

- Cosmological constant, $w = -1$. The energy density of a cosmological constant is, as the name suggests, constant over time and unaffected by cosmic expansion/contraction. For this reason it is also often described as vacuum energy. Note that $p = -\rho$ implies that the energy-momentum tensor is proportional to the metric, $T_{\mu\nu} \propto g_{\mu\nu}$, which is consistent with the Lorentz invariance of the vacuum. Another name that is used to describe a cosmological constant, or in fact any type of matter with an equation of state that is close to -1 is *dark energy*. While observations suggest that dark energy currently provides the dominant contribution to the overall energy density of the universe, this type of matter remains by far the least understood.

There is one important omission in this list: scalar fields, which are often used in modelling the dynamics of the universe. We will treat these in detail later.

We are finally ready to derive the *Friedmann equations*. These equations arise from the Einstein field equations by assuming that the metric takes on a Robertson-Walker form. The Einstein equations state that the Einstein tensor $G_{\mu\nu}$, which is directly related to the curvature tensor (Ricci tensor) $R_{\mu\nu}$ and which describes the curvature of spacetime, is equal to the stress-energy tensor $T_{\mu\nu}$, which describes the energy, momentum and pressure of matter,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \equiv G_{\mu\nu} = T_{\mu\nu} \quad (\text{in reduced Planck units}, 8\pi G = 1) \tag{1.31}$$

– in the words of J.A. Wheeler "spacetime tells matter how to move; matter tells spacetime how to curve". With an energy-momentum tensor of perfect fluid form, the Einstein equations then reduce to

$$H^2 + \frac{k}{a^2} = \frac{1}{3}\rho, \tag{1.32}$$

$$\frac{\ddot{a}}{a} = -\frac{1}{6}(\rho + 3p) = -\frac{1}{6}\rho(1 + 3w). \tag{1.33}$$

The first of these equations is sometimes referred to as simply the *Friedmann equation*, while the second one is often called the *acceleration equation*. These equations determine the evolution of the observable universe on the largest scales. One should think of ρ and p as denoting the total energy density and pressure – these then consist of a sum over all the individual matter types that are present in the universe. First, though, it is useful to look at a few solutions to these equations when only a single matter component is present. We will first do this for the examples of matter types enumerated above, for the case of a spatially flat universe ($k = 0$):

- Pressure free matter/dust, $w = 0$. The Friedmann equation then implies that the scale factor evolves as $a(t) \propto t^{\frac{2}{3}}$, while the acceleration equation shows that the universe is decelerating. (Including integration constants, the solution can be written as $a(t) = \frac{(t-t_\star)^{2/3}}{(t_0-t_\star)^{2/3}}$, where t_0 is the time today, and $t_\star$ the time of the big bang, such that $a_0 = 1$.)

- Relativistic particles/radiation, $w = \frac{1}{3}$. In this case the scale factor evolves as $a(t) \propto t^{\frac{1}{2}}$, and this also corresponds to decelerated expansion.

- Cosmological constant, $w = -1$. The Friedmann Eq. (1.32) implies that $\rho > 0$ since $k = 0$. In this case the scale factor evolves exponentially, $a(t) \propto e^{Ht}$, and the Hubble parameter is constant. This solution is known as the *de Sitter* universe. According to Eq. (1.33), in the de Sitter solution the expansion is accelerating. Above, we mentioned that matter with equation of state close to -1 is also sometimes called dark energy. Here, we can refine that definition. The acceleration equation implies that, for a positive energy density, the universe is accelerating whenever $w < -\frac{1}{3}$. We have presently entered such an era of accelerated expansion.

The solutions for $k = 0$ are those that are used most often in cosmology, as current observations indicate that our universe is very close to spatially flat. However, merely the fact that the matter in the universe is distributed inhomogeneously on small scales indicates that, at least when one looks at smaller regions of the universe, the average curvature must necessarily be positive or negative. Hence it is certainly also important to look at solutions with $k = \pm 1$. The corresponding solutions are more easily found using conformal time τ. In terms of conformal time, the Friedmann equations read

$$a'^2 + ka^2 = \frac{1}{3}\rho a^4 \quad (1.34)$$

$$a'' + ka = \frac{1}{6}(\rho - 3p)a^3, \quad (1.35)$$

where a prime denotes a derivative with respect to conformal time.

For $k = +1$, the spatial sections of the universe are spheres, and this is known as the *closed* model of the universe. The Friedmann Eq. (1.34) shows that the expansion can come to a halt, even when the energy density is positive. After such a halt, the universe re-collapses, as can be seen in the solutions obtained in the presence of dark matter or radiation

$$a(\tau) = \begin{cases} 1 - \cos\tau & w = 0 \quad 0 < \tau < 2\pi \\ \sin\tau & w = \frac{1}{3} \quad 0 < \tau < \pi. \end{cases} \quad (1.36)$$

For $k = -1$, the universe is called *open*, and the corresponding solutions are

$$a(\tau) = \begin{cases} \cosh\tau - 1 & w = 0 \quad 0 < \tau < \infty \\ \sinh\tau & w = \frac{1}{3} \quad 0 < \tau < \infty. \end{cases} \tag{1.37}$$

For those solutions, the universe expands forever.

Note that for a negative cosmological constant (ρ constant and negative), the Friedmann equation implies that we must have $k = -1$ to obtain a solution at all. Moreover, if the universe expands, the curvature term $\frac{k}{a^2}$ becomes increasingly subdominant compared to the cosmological constant term, so that the expansion must necessarily come to a halt, after which the universe collapses.

After this brief survey of solutions to the Friedmann equations, a few general remarks about these equations are in order: the Friedmann equations describe the evolution of the curvature of the universe over time. Even spatially flat universes ($k = 0$) are curved in a 4-dimensional sense – their curvature arises from the evolution of the scale factor. Whether the universe has positive or negative spatial curvature can be deduced by comparing its density to the *critical density*

$$\rho_{\text{crit}} = 3H^2, \tag{1.38}$$

which at any given time represents the energy density of a spatially flat universe with the identical Hubble rate at that instant. Currently, the critical density corresponds to approximately

$$\rho_{\text{crit},0} = 3H_0^2 \approx \frac{3}{8\pi G}(70\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1})^2 \approx 10^{-26}\,\text{kg/m}^3. \tag{1.39}$$

The Friedmann equations describe both expanding and contracting universes. As discussed in section 1.1.2, measurements of the spectra of distant galaxies indicate that all but a few close galaxies are receding from us (with increasing velocity as their distance to us increases), and this is a clear indication that the universe is currently expanding. However, one should keep an open mind regarding the possibility that at other cosmic epochs, the universe might have been contracting. We will study cosmological models involving contracting phases later on.

The velocity of an object flying through the universe is well-defined with respect to the local Hubble flow, *i.e.* with respect to a local inertial frame. However, the relative velocity between two distant objects is not very meaningful, because their individual velocities can only be measured in the respective local frames. In an expanding universe, one is tempted to say that increasingly distant regions recede at increasing speeds, and, given a sufficient separation, that speed exceeds the speed of light, thus apparently violating a basic law of nature. However, by the argument just given, such a description is misleading. Our discussion of horizons in section 1.1.6 will treat this topic, and the associated issues of causality, in more detail.

The Friedmann equations have two huge consequences for the picture that we have of the universe. The first was already discussed, namely that the universe is not static, but it is an evolving entity. Space and time are evolving quantities, even on the largest observable scales. This marks a major shift in our thinking about the universe: up until 100 years ago, people were mostly convinced (on philosophical grounds) that the universe had to be static. The discovery of the expansion of the universe has brought with it a major conceptual paradigm shift, as the universe is now seen as en evolving object in itself. This shows in a very clear way how space, time and matter in the universe are tightly bound to each other. The second consequence becomes apparent when looking at the cosmological solutions described above: in all of them except for the de Sitter solution the universe reaches a point where the scale factor vanishes. That this point does not in general correspond to a coordinate singularity is apparent from the expression for the Ricci scalar (1.26), which is a curvature invariant. The Ricci scalar generically blows up as $a \to 0$, thus indicating that this moment represents a singularity. This singularity is called the *Big Bang*, and in the standard Big Bang theory it is interpreted as the beginning of the universe (and often in fact also as the beginning of space and time). As we will discuss shortly, such an interpretation appears to be fallacious, but nevertheless, the Big Bang represents a special moment in the history of the universe which every self-respecting cosmological model should aim to explain! The singular nature of the Big Bang is an indication that general relativity breaks down around that time, and a quantum theory of gravity is likely needed in order to resolve this singularity.

1.1.4 The Thermal History of the Universe

It is instructive to study the Friedmann equations in the presence of several matter components:

$$3H^2 = \frac{\rho_{\mathrm{r},0}}{a^4} + \frac{\rho_{\mathrm{m},0}}{a^3} - \frac{3k}{a^2} + \Lambda. \tag{1.40}$$

If we choose units in which the current scale factor is $a_0 = 1$, then $\rho_{\mathrm{r},0}$ and $\rho_{\mathrm{m},0}$ denote the current energy densities in radiation and pressure free matter, while Λ denotes the contribution from a cosmological constant. As the universe expands, it is clear that matter types that scale with a less negative power of the scale factor eventually dominate over those which scale with more negative powers of a. Thus, if Λ is non-zero, it will eventually dominate the energy density in the universe, and determine its evolution. A useful way to re-write Eq. (1.40) can be obtained by dividing through by the current critical density $3H_0^2$, to obtain

$$\left(\frac{H}{H_0}\right)^2 = \frac{\Omega_\mathrm{r}}{a^4} + \frac{\Omega_\mathrm{m}}{a^3} + \frac{\Omega_\mathrm{k}}{a^2} + \Omega_\Lambda, \tag{1.41}$$

where the current fractional energy densities are conventionally defined as

$$\Omega_\mathrm{r} = \frac{\rho_{r,0}}{3H_0^2}, \quad \Omega_\mathrm{m} = \frac{\rho_{m,0}}{3H_0^2}, \quad \Omega_\mathrm{k} = \frac{-k}{H_0^2}, \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}. \tag{1.42}$$

In fact, observational evidence indicates that about 68.5% of the *current* energy density must be due to a cosmological constant (or something similar). This realisation stems foremost from the observation of distant supernovae, which allow one to infer the expansion history of the universe. Such measurements indeed indicate that our universe has entered a phase of accelerated expansion about 5 billion years ago (which, coincidentally, was around the time that the solar system formed). Furthermore, observations of the cosmic microwave background indicate that the total energy density is equal to the critical density with an accuracy better than one part in 100, thus indicating that to a good approximation we can take $k = 0$ to describe the recent expansion history of the universe. Dark matter currently comprises about 26.6% of the total energy density, and baryonic matter 4.9%. Radiation adds only about a fraction 6×10^{-5} to the total. The Planck satellite has measured the current Hubble rate to be about $H_0 \approx 67.3 kms^{-1}Mpc^{-1}$ (with $1pc = 3.26$ light-years).

We are now in a position to calculate the time that has elapsed since the Big Bang. Defining $u \equiv \frac{a}{a_0}$ and using the new notation, the Friedmann Eq. (1.32) can be recast as the differential

$$\mathrm{d}t = \frac{\mathrm{d}u}{H_0\sqrt{\Omega_\mathrm{r}u^{-2} + \Omega_\mathrm{m}u^{-1} + \Omega_\mathrm{k} + \Omega_\Lambda u^2}}. \tag{1.43}$$

Integrating from $u = 0$ to $u = 1$ gives us the time since the Big Bang, namely

$$t_{BB} \approx 0.95\frac{1}{H_0} \approx 13.8 \times 10^9 \,\text{years}. \tag{1.44}$$

Note that, for quick order-of-magnitude calculations, the time since the Big Bang and the size of the observable universe (on the order of 10 billion years and 10 billion light years) are approximately given by 10^{60} Planck times and 10^{60} Planck lengths respectively!

We can also easily estimate the redshift at the time of radiation-matter equality: equating $\Omega_\mathrm{r}/a_\mathrm{eq}^4 = \Omega_\mathrm{m}/a_\mathrm{eq}^3$ we find

$$a_\mathrm{eq} = \frac{1}{1+z_\mathrm{eq}} = \frac{\Omega_\mathrm{r}}{\Omega_\mathrm{m}} \approx \frac{1}{5000}. \tag{1.45}$$

(To find t_eq one would simply integrate Eq. (1.43) from $u = 0$ to $u = a_\mathrm{eq} = 1/5000$.)

Photons have energy $E = h\nu = hc/\lambda = k_BT$, where ν is the frequency, λ the wavelength, k_B Boltzmann's constant (introduced by Planck) and T the absolute temperature. Thus the wavelength can be re-expressed as $\lambda = hc/(k_BT)$. As the universe expands or contracts, the wavelength changes with the scale factor $\lambda \propto a(t)$, and thus the temperature evolves as

$$T(t) = T_0\frac{a_0}{a(t)}. \tag{1.46}$$

We already saw this relation in disguise when calculating the scaling of the energy density of radiation with the scale factor ($\rho_\mathrm{r} \propto a^{-4}$), and we will see another derivation below. This relation implies that the universe was *hotter* in the past, hence the occasionally used name *Hot Big Bang model*. In the very early universe, photons and other matter particles were tightly coupled, implying that in fact all matter constituents were hot. Here we will briefly sketch the thermal history of the universe. In the early universe, the dynamics was principally determined by the energy density of radiation. For this epoch, there exists a useful relation between temperature and time; the Friedmann equation implies that

$$H^2 = \frac{1}{4t^2} = \frac{1}{3}\rho \approx T^4, \tag{1.47}$$

where we have neglected a factor of order unity relating the energy density in radiation to the fourth power of temperature. The relation above, namely $T \approx 1/\sqrt{t}$ holds in Planck units, and if we restore units, we get the formula

$$T_\mathrm{MeV} \approx \frac{1}{\sqrt{t_\mathrm{sec}}}, \tag{1.48}$$

where the temperature is expressed in MeV and the time in seconds (counting from a putative Big Bang). Major events in the history of the universe were:

$t < 10^{-14}\,$s, $T >$ TeV

We still know very little about this epoch, as the corresponding particle physics currently cannot be reproduced in accelerator experiments. From our current laws of physics, we can assume that the electroweak symmetry was unbroken during this early phase, and in fact the gauge group describing all matter interactions could have been much bigger (such as $SO(10)$, allowing a grand unified description of particle physics.) Also, supersymmetry might have been a good symmetry of nature at those early times, but it remains too early to tell.

$10^{-14} < t < 10^{-10}\,$s, 100GeV $< T <$ 10TeV

During this time, the electroweak symmetry $SU(2)\times U(1)$ was broken, and the W, Z bosons became massive. The particle physics that takes place at these energies is currently being probed by the Large Hadron Collider at CERN, and is, to a large extent, well understood.

$t \sim 10^{-5}\,$s, $T \sim$ 200MeV

As the universe cools, free quarks and gluons bind into baryons (particles with 3 quarks, such as protons and neutrons) and mesons (particles with a quark and an anti-quark).

$t \sim 0.2\,$s, $T \sim 1 - 2$MeV

The cross-sections of the weak interactions become very small as the temperature drops to these scales, and thus the neutrinos decouple from the rest of the matter in the universe. This neutrino background evolves independently from that time on. In the long term future, one may hope that it will be possible to build a neutrino telescope (perhaps on the Moon) that can measure these primordial neutrinos directly. This would give us direct information about the physical conditions at these times close to the Big Bang.

As a byproduct, the ratio of neutrons to protons also *freezes out* around this time. Up until this time, protons could easily convert into neutrons, and vice versa, via the exchange of neutrinos, electrons and positrons,

$$p + e^- \leftrightarrow n + \nu, \qquad p + \bar{\nu} \leftrightarrow n + e^+. \tag{1.49}$$

The neutron is heavier than the proton by an amount $Q = m_n - m_p \approx 1.3$MeV, where $m_{n,p}$ denote the masses of a neutron and a proton, respectively. Standard thermodynamics then implies that, as long as the two species are still in equilibrium, *i.e.* as long as the reactions above are efficient, the number densities of neutrons (n_n) and protons (n_p) are related by

$$\frac{n_n}{n_p} \approx e^{-Q/T}. \tag{1.50}$$

The reactions above require an energy $m_n - m_p - m_e \approx 1.3 - 0.5 \approx 0.8$MeV to be efficient, where m_e denotes the mass of an electron. Below this temperature, the ratio of neutrons to protons gets frozen in, with $n_n/n_p \approx e^{-1.3/0.8} \approx 1/5$. This has important implications for big bang *nucleosynthesis*, as we will see shortly.

$t \sim 1\,\mathrm{s}$, $T \sim 0.5\mathrm{MeV}$

When the temperature drops to the value of the rest mass of electrons and positrons, these start annihilating each other into photons, while the converse process, spontaneous creation of electron-positron pairs out of photons, becomes rare. The photons that are produced in this way are in thermal equilibrium, and their temperature is slightly higher than that of the neutrinos, which already decoupled earlier. The end result of electron-positron annihilation is that about one electron was left over for each billion photons. Thus there must have been a very slight excess of matter over anti-matter in the early universe. This asymmetry remains largely unexplained.

$t < 5\,\mathrm{min}$, $T \sim 0.05\mathrm{MeV}$

Above, we saw that the ratio of neutrons to protons freezes out at a value of about 1/5. However, neutrons are only stable when they are locked inside an atomic nucleus. Free neutrons decay into a proton, an electron and a neutrino with a half life of about 15min. Over the next few minutes, these decays reduce the neutron to proton number density fraction to

$$\frac{n_n}{n_p} \approx \frac{1}{7}. \tag{1.51}$$

Meanwhile, neutrons can combine with protons to form deuterons

$$p + n \to d + \gamma, \tag{1.52}$$

while, in turn, deuterons can combine via the following chain to end up as helium nuclei

$$d + d \to {}^3He + n, \qquad {}^3He + d \to {}^4He + p. \tag{1.53}$$

Assuming that all neutrons end up in helium atoms (which is in fact a good approximation), we would deduce that the total mass fraction in helium should be

$$Y = \frac{n_{He} m_{He}}{n_n m_n + n_p m_p} \approx \frac{(\frac{1}{2} n_n)(4 m_n)}{(n_n + n_p) m_n} = \frac{2 n_n / n_p}{1 + n_n / n_p} \approx \frac{1}{4}. \tag{1.54}$$

Helium is also produced via nuclear fusion in stars. However, looking at stars that are further and further away (and thus looking further into the past) observations show that helium levels decrease and level off around a value of 1/4. Measurements of other light elements, such as deuterium and lithium, are also in good agreement with theoretical estimates of their production via primordial *nucleosynthesis*. This is striking evidence that the universe was once hot and dense, and provides strong support of the hot big bang model!

$T \sim \mathrm{eV}$

After a time on the order of 10000 years, matter comes to dominate over the radiation energy density, and the universe starts expanding according to $a \propto t^{2/3}$. By the time the universe grows by a further order of magnitude in each direction, the helium nuclei and protons combine with the electrons to form helium and hydrogen atoms. This event is known as *recombination* and will be discussed in more detail in the next section. The name recombination is slightly inappropriate, as this is the first time after the big bang that atoms form.

$T \sim$ meV
As we will discuss, small fluctuations in the distribution of matter and radiation are present already at early times in the history of the universe. Under the influence of gravity, small initial overdensities grow and eventually collapse into stars and galaxies.

About 9 billion years after the big bang, the dark energy comes to dominate the energy density in the universe, and the expansion starts accelerating. The future dynamics of the universe depends crucially on the nature of dark energy. We will speculate about possible futures at various stages in this book.

1.1.5 A Relic from the Big Bang: The Cosmic Microwave Background

At the time of radiation-matter equality, the universe can still be described as a primordial soup of mostly photons, electrons, protons and helium ions. Via Thomson scattering, the photons constantly bounce off the free electrons, and hence the universe is not transparent at that time. As the universe cools, at some point it will become energetically favorable for the electrons to bind with the free helium ions and protons to form helium and hydrogen atoms respectively. In fact, because helium has a larger ionization potential, helium atoms form first. Afterwards, there are still many free electrons left over, and these can combine with protons to form neutral hydrogen atoms. Since the binding energy of hydrogen is about 13.6eV, a first guess is that hydrogen recombination will occur when the universe has reached that temperature. However, it actually occurs a little later, when the universe has cooled off a little more. This is because there are so many more photons than electrons (about a billion times as many), so that a small fraction of atypically energetic photons is enough to keep hydrogen ionized. In the tail of the Planck distribution of the photons, there are about $e^{-B/T}$ photons with energy larger than B when the temperature is T. Hence, for a binding energy $B \sim 13.6$eV, there are enough energetic photons as long as the temperature is above $T \approx -B/(\ln 10^{-9}) \approx 3000\,\mathrm{K}$. This corresponds to a redshift of $z \approx 1090$.

At that time, electrons and protons combine into neutral hydrogen atoms. Photons do not interact much with hydrogen atoms, because as a bound system hydrogen is neutral (of course, there is some scattering, called Rayleigh scattering, due to the fact that on some level photons will see the constituent proton and electron). But this means that, at this particular moment, photons can travel unhindered over large distances for the first time and the universe becomes transparent! The newly freed photons pervade the universe, flying in all directions. Since the photons, electrons and protons were in thermal equilibrium up to this time, the newly released radiation obeys a black body distribution with temperature $T \approx 3000K$. As the universe expands, this radiation cools, but the black body spectrum is preserved, as we will now show. We will assume here that recombination occurred instantaneously, and that, consequently, the radiation was in equilibrium with matter up until recombination, afar which time the radiation evolved independently and freely. In equilibrium, the energy density of the radiation is given by

$$\rho = 2\int \frac{\mathrm{d}^3 p}{(2\pi)^3} f_{\mathrm{B-E}} E \tag{1.55}$$

where p denotes the momentum of a photon, $f_{\mathrm{B-E}} = 1/(e^{E/T} - 1)$ denotes the Bose-Einstein distribution appropriate for bosons, E is the energy of a photon and is simply $E = p$, and the overall factor of 2 arises because the photon has two helicities. This gives

$$\rho = 2\int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{p}{e^{p/T} - 1} = \frac{\pi^3}{15} T^4 \tag{1.56}$$

But radiation scales as $\rho \propto a^{-4}$ (because it has an equation of state $w = \frac{1}{3}$), and hence we find once more that the temperature of the cosmic microwave background (and of all matter types that it is in equilibrium with during the early universe) scales as

$$T(t) \propto \frac{1}{a(t)}. \tag{1.57}$$

Inside the integral above, the momentum p and the temperature T scale in such a way with the expansion of the universe that the functional form remains preserved, in particular p/T remains constant. Thus, the spectrum remains that of a black body, with a temperature decreasing inversely to the expansion of the universe.

When released, the cosmic background radiation was at a temperature of $3000K$. Such radiation would have been easily visible to the human eye (*cf.* the Sun's surface temperature of about $6000K$). By today, however, the wavelength of the emitted background radiation is bigger by a factor 1090. When emitted, the typical wavelength was on the order of a μm. Now it is on the order of a mm, and thus the radiation is in the microwave frequency region. This is the origin of the name *Cosmic Microwave Background* radiation (CMB). The CMB can be measured via radio telescopes, with the most precise measurements to date being carried out by specifically dedicated satellite missions, such as WMAP and PLANCK.

The radiation that we measure today has flown at the speed of light for 13.8 billion years. It is of interest to calculate the current distance to the regions from where it was emitted: using conformal time, and a flat model of the universe, $k = 0$, light propagates along

$$\mathrm{d}\tau = \pm \mathrm{d}r. \tag{1.58}$$

Hence, using units where $a_0 = 1$, the physical (co-moving) distance to the surface of last scattering (LS) is (neglecting radiation, which is a good approximation)

$$\begin{aligned} d_{LS} &= a_0(r_{LS} - r_0) \\ &= r_{LS} \\ &= \int \mathrm{d}\tau \\ &= \int_{t_{\mathrm{LS}}}^{t_{BB}} \frac{\mathrm{d}t}{a(t)} \end{aligned}$$

$$
\begin{aligned}
&= \int_{a=1/1090}^{a=1} \frac{\mathrm{d}a}{H_0\sqrt{\Omega_{\mathrm{m}} a + \Omega_\Lambda a^4}} \\
&\approx \frac{3.13}{H_0}
\end{aligned}
\tag{1.59}
$$

Since $1/H_0 \approx 14.5$ billion (light) years, we get a present distance to those regions from which the currently observed CMB was emitted of about 45 billion light years. This is the current radius of the *cosmic sphere*, and it represents the current distance to the furthest places that can be observed with light. This distance is larger than the naive estimate of 13.8 billion light-years because of the expansion of the universe that has taken place between the time of emission and detection of the CMB radiation. At the time that the CMB was emitted, the corresponding radius was 1090 times smaller, *i.e.* only about 42 million light years, which is smaller than the scale on which the universe is homogeneous today. As an aside, we can relate this size to the scale at which the universe becomes approximately homogeneous, namely 300 million light years. Thus, the observable universe is bigger than the scale of homogeneity by a factor of about 100, or, in volume, the observable universe consists of about 10^6 regions of the scale of homogeneity. This gives us an indication to what extent the cosmological principle applies, *i.e.* to what extent spatial isotropy and homogeneity are good approximations.

The CMB provides the best evidence that we have for the isotropy of the universe, but only in a special frame. The microwave background radiation itself defines a preferred frame in the universe, which can be considered to be the (co-moving) "rest frame." To someone traveling at a considerable fraction of the speed of light with respect to that frame, the universe would not look isotropic at all. And in fact measurements of the CMB show a dipole with maximal intensity of about $\frac{\Delta T}{T} \sim 1/500$, which can be interpreted as indicating that we are moving with respect to the rest frame of the universe with a speed of about $600\,\mathrm{km\,s^{-1}}$, in the direction of the Hydra-Centaurus galaxy supercluster. When this dipole is subtracted from the data, the universe appears isotropic to one part in 10^5 in all directions; see Fig. 2.

Along with big bang nucleosynthesis, the cosmic background radiation is also the best evidence that we have for the hot big bang model. However, just as for nucleosynthesis, it is important to realise that it is not evidence for the big bang itself. Rather, it provides clear evidence that the universe was once very dense and at least a thousand times hotter than presently.

1.1.6 Puzzles of the Big Bang Model

Even though the hot big bang model explains the abundances of light elements in the early universe, as well as the existence of the cosmic background radiation, it leaves many questions unanswered. Here we will present some of the main puzzles:

Singularity Problem

When $a = 0$, the Robertson-Walker metric becomes singular, as curvature invariants like the Ricci scalar R become infinite. Thus the model breaks down at the big bang. To explain the origin of the universe, we need something that goes beyond the big bang model.

Flatness Puzzle

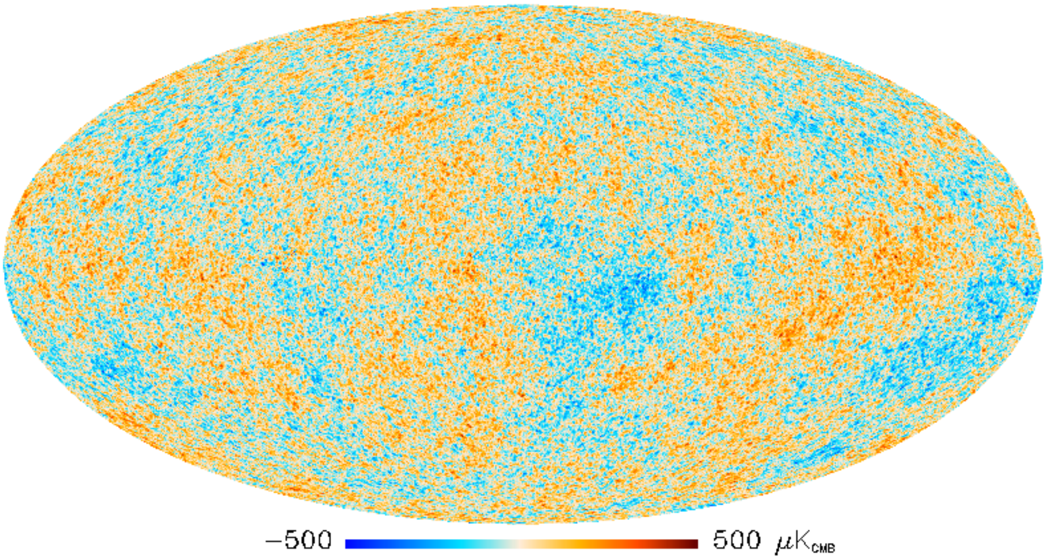

Figure 2: All-sky map of the cosmic microwave background radiation, as observed by the PLANCK satellite. In this picture, the dipole has been subtracted off, as have galaxy foregrounds, leaving only the small temperature fluctuations of a size of tens of μK. Red coloured regions are hotter than average, and blue ones colder. *Credit: ESA / PLANCK*

If we consider the Friedmann equation again, e.g. in the form of Eq. (1.41), we can see that the term representing the homogeneous curvature of the universe contributes a fractional amount $|\Omega_{\rm k}| = |k|/(aH)^2$ to the total energy content of the universe. The dependence on the scale factor and Hubble rate implies that, as the universe expands during the radiation and matter-dominated epochs, this fractional contribution becomes larger and larger. Present-day observations put an upper bound of

$$|\Omega_{\rm k,0}| \lesssim 10^{-2}. \tag{1.60}$$

But this means then that at early times, the average curvature of the universe must have been extremely tiny. Assuming pure radiation-domination as a first approximation, we find that at the Planck time, we must have had

$$|\Omega_{{\rm k},Pl}| \lesssim 10^{-2}\frac{(aH)_0^2}{(aH)_{Pl}^2} = 10^{-62}, \tag{1.61}$$

an incredibly small quantity. If we do not go back in time quite as far, the argument becomes slightly weaker but not really less troubling. How could the universe have been so incredibly flat at very early times?

Horizon Puzzle

The following argument can be made more precise, but in order to understand the essence it is already sufficient to assume once again that the universe was dominated by radiation throughout its history. The horizon problem stems from the fact that the universe, when evolved back in time, does not reduce to a small volume compared to its age. For the sake of argument, consider the presently known universe, with a (linear) size of 10^{60} Planck lengths (of course, the universe might very well be much larger, which would only strengthen the argument). If the universe is dominated by radiation, then the scale factor evolves as

$$a(t) = a(t_0)\left(\frac{t}{t_0}\right)^{1/2}, \tag{1.62}$$

where t_0 is a reference time. Setting $a(t_0) = 10^60$ and evolving our part of the universe back to the Planck time, we would obtain

$$a(t_{Pl}) = a(t_0)10^{-30} = 10^{30}, \tag{1.63}$$

i.e. we get a universe that is one Planck time old, yet 10^{30} Planck lengths in extension. Yet all these regions have undergone approximately the same expansion history, as evidenced by observations of the cosmic background radiation. How could the "initial conditions" have been set in all these regions in a causal manner? The particle horizon was evidently much too small to envisage a causal explanation. Another way to phrase the question is to ask how the big bang could have been synchronised over such a large region, assuming that the big bang was the beginning of space and time? From this point of view, one would have needed 10^{90} simultaneous big bangs, yet there was no time to synchronise them (synchronisation requires the causal transfer of information) if this was the beginning of time, and there was no place to put them if this was the beginning of space. The "beginning of time" proposition is clearly nonsensical! Note that the argument would be slightly weaker if we did not go back on time to the Planck time, but unchanged in essence. Thus we may also reformulate the question: what could have synchronised the big bang?

We can make the intuitive arguments above more precise by calculating the behaviour of the particle horizon, which represents the distance over which events moving with the Hubble flow could have been in causal contact with each other. For this, it is easier to use conformal time τ, with metric (expressed in spherical coordinates)

$$ds^2 = a(\tau)^2\left[-d\tau^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right]. \tag{1.64}$$

We take the radial direction to be that of a light ray, with spacetime path specified by

$$ds^2 = 0 \quad \rightarrow \quad d\tau = \pm dr \tag{1.65}$$

Starting from time τ_i at position r_i the particle horizon at time τ is thus given by

$$\begin{aligned} d_{horizon} &= a(r - r_i) && (1.66)\\ &= a(\tau - \tau_i) && (1.67)\\ &= a(t)\int^t \frac{dt'}{a} = a(t)\int \frac{da}{Ha^2} = a(t)\int \frac{d\ln(a)}{aH}. && (1.68) \end{aligned}$$

We saw above that during radiation and matter domination, the “comoving Hubble radius” $1/(aH)$ has been growing, so that the particle horizon used to be smaller and smaller in the past. In other words, more and more regions in causal contact now could not have been in causal contact at increasingly early times. Thus, mathematically speaking, the horizon puzzle has the same origin as the flatness puzzle.

Inhomogeneity/Anisotropy Puzzle

The CMB is isotropic to a high degree, but there are temperature fluctuations of order $\Delta T/T \sim \mathcal{O}(10^{-4})$, which are in fact responsible for al the gravitationally bound and collapsed structure we see in the universe today. How did these primordial temperature perturbations arise?

Brief List of Additional Puzzles

- The early universe was in a very peculiar state: the matter was very hot (high entropy), but the geometry very flat (low entropy). Overall the entropy of the universe was very low, which, from the point of view of thermodynamics, is synonymous with the state of the early universe having been very unlikely. Can we understand this special state?
- If particle physics is described by a grand unified theory at very high energies, then one would expect topological defects to have formed when the symmetry broke as the universe cooled. Why are none seen?
- Why is there so much more matter than antimatter in the universe?
- What is the role of dark energy? In the big bang model, it plays no role whatsoever. Is there a deep reason why dark energy exists?

The fact that all of these puzzles cannot be explained by the hot big bang model means that we must go beyond in order to understand the early universe. In other words, we need at least one new ingredient! As we will see, a scalar field, such as the Higgs field that was recently discovered, could be an important missing ingredient as it can change the dynamics of the early universe rather drastically by inducing either an inflationary or an ekpyrotic phase.

1.2 The Inflationary Universe

1.2.1 Basic Idea

The flatness and horizon problems would be solved if there was an earlier phase of evolution (i.e. a phase taking place before the radiation-dominated expansion period) during which $1/(aH)$ was shrinking by a large amount (extrapolating the earlier arguments back to the grand unified scale, we would need the shrinking in $1/(aH)$ to be by at least a factor 10^{-25}). Earlier, we saw that for constant w we have $\rho \propto a^{-3(1+w)}$. Using the Friedmann equation, this leads to

$$\dot{a}^2 \propto a^{-1-3w} \tag{1.69}$$

$$a \propto t^{\frac{2}{3(1+w)}} \qquad \text{assuming } w > -1 \tag{1.70}$$

$$\frac{1}{aH} \propto t^{\frac{1+3w}{3(1+w)}} \propto a^{(1+3w)/2} \tag{1.71}$$

Thus, in an expanding universe, the comoving Hubble radius $1/(aH)$ shrinks if $w < -1/3$. Recall that the acceleration equation reads $\ddot{a}/a = -(1 + 3w)\rho/6$, we can see that this condition is equivalent to having a period of *accelerated expansion*. This is called *inflation*.

Above we restricted to the case $w > -1$, however the case $w = -1$ is in fact of special interest. Then $\rho + p = 0$ and consequently $\dot{\rho} = 0$, i.e. we have a constant energy density, in other words a cosmological constant. In that case the solution to the equations of motion is given by the de Sitter solution

$$a = a_0 e^{Ht} \tag{1.72}$$

for some constant a_0 and where the Hubble rate is now constant, $H = \sqrt{\rho/3}$. Evidently, $1/(aH) \propto e^{-Ht}$ also shrinks in that case.

How much accelerated expansion do we need in order to solve the flatness and horizon problems? We require

$$\frac{(aH)_{beg}}{(aH)_{end}} < 10^{-25} \rightarrow \frac{a_{end}}{a_{beg}} > 10^{-25\frac{2}{(1+3w)}}, \tag{1.73}$$

where *beg* and *end* refer to the beginning and end of the inflationary phase respectively. For $w \approx -1$ this leads to the requirement

$$\frac{a_{end}}{a_{beg}} > 10^{25} \approx e^{60} \tag{1.74}$$

Thus often it is said that we require at least 60 "e-folds" of inflation in order to solve the flatness and horizon problems.

1.2.2 Scalar Field Implementation

The most popular way to model an inflationary phase is by considering the dynamics of a scalar field ϕ coupled to gravity and moving in an appropriate potential $V(\phi)$. The action is then given by

$$S = \int d^4x\sqrt{-g}\left[\frac{R}{2} - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right], \tag{1.75}$$

where we are working in reduced Planck units $8\pi G = 1$. Here the kinetic term is abbreviated by $(\partial\phi)^2 \equiv g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$. The equations of motion are obtained by varying the action with respect to the metric $g^{\mu\nu}$ and the scalar field ϕ, yielding respectively

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \frac{1}{2}g_{\mu\nu}(\partial\phi)^2 - g_{\mu\nu}V\,, \tag{1.76}$$

$$\Box\phi = V_{,\phi}\,. \tag{1.77}$$

In a flat ($k = 0$) FLRW background spacetime, these equations reduce to

$$3H^2 = \frac{1}{2}\dot{\phi}^2 + V\,, \tag{1.78}$$

$$\dot{H} = -\frac{1}{2}\dot{\phi}^2 , \tag{1.79}$$
$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0 . \tag{1.80}$$

In this background the stress-energy tensor $T_{\mu\nu}$ has the form of a perfect fluid, with energy density and pressure given by

$$\rho = \frac{1}{2}\dot{\phi}^2 + V , \tag{1.81}$$
$$p = \frac{1}{2}\dot{\phi}^2 - V . \tag{1.82}$$

Thus the equation of state is given by

$$w = \frac{p}{\rho} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V} . \tag{1.83}$$

The acceleration equation can now be rewritten as $\ddot{a}/a = -(\rho + 3p)/6 = -(\dot{\phi}^2 - V)/3$, so in order to obtain inflation we see that we need the potential energy to dominate over (twice) the kinetic energy, $\dot{\phi}^2 < V$.

We can also define the often-used parameter ϵ via

$$\epsilon \equiv \frac{3}{2}(1 + w) = \frac{1}{2}\frac{\dot{\phi}^2}{H^2} \tag{1.84}$$

and we can immediately see that inflation is obtained for

$$\epsilon < 1, \qquad \text{condition for inflation.} \tag{1.85}$$

A useful example is provided by the exponential potential

$$V = V_0 e^{-c\phi}, \tag{1.86}$$

where V_0, c are positive constants, since the equations of motion can be solved exactly in this case. The solution is given by

$$a = a_0 t^{1/\epsilon}, \quad H = \frac{1}{\epsilon t}, \quad \phi = \frac{1}{\sqrt{2\epsilon}} \ln\left(\frac{V_0 \epsilon^2}{3 - \epsilon} t^2\right), \quad V = \frac{3 - \epsilon}{\epsilon^2 t^2} \tag{1.87}$$

where a_0 is a constant and

$$\epsilon = \frac{c^2}{2} \left(= \frac{1}{2}\frac{\dot{\phi}^2}{H^2} = \frac{1}{2}\frac{V_{,\phi}^2}{V^2} \right) . \tag{1.88}$$

We can see that we need $c^2 < 2$ in order to obtain inflation. Note that ϵ above is constant and agrees with the earlier definition. The solution for the scale factor a clearly shows again that accelerated expansion corresponds to $\epsilon < 1$. Note also that the solution above is a scaling solution: in the equations of motion, each term has the same time-dependence ($\propto 1/t^2$), so that the relative contributions of the various terms remains unchanged over time (this being related to the fact that a shift in ϕ can be absorbed by a re-definition of V_0). In this model, enough inflation is obtained if

$$\frac{a_{end}}{a_{beg}} = \left(\frac{t_{end}}{t_{beg}}\right)^{1/\epsilon} = e^{\frac{\phi_{end}-\phi_{beg}}{\sqrt{2\epsilon}}} > e^{60} \quad \rightarrow \quad \phi_{end} - \phi_{beg} > 60\sqrt{2\epsilon}. \tag{1.89}$$

An important simplification arises in general when we are in the "slow-roll" regime, where the kinetic energy is vastly subdominant to the potential energy, $\dot{\phi}^2 \ll V$. Further assuming that $|\ddot{\phi}| \ll H|\dot{\phi}|$, so that the scalar field kinetic energy remains small over an extended period, implies that the equations of motion can be approximated by

$$3H^2 \approx V \tag{1.90}$$

$$3H\dot{\phi} \approx -V_{,\phi} \tag{1.91}$$

These slow-roll equations are used all the time in inflationary cosmology. Using these relations, ϵ can be rewritten

$$\epsilon \approx \frac{1}{2}\frac{V_{,\phi}^2}{V^2} \quad \leftrightarrow \quad \epsilon \ll 1, \tag{1.92}$$

and under these circumstances ϵ is usually called the (first) slow-roll parameter. In line with the above approximation, a second slow-roll parameter is often also defined via

$$\eta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}}, \tag{1.93}$$

and the second slow-roll condition is that

$$|\eta| \ll 1. \tag{1.94}$$

Note that using the slow-roll equations, η can be rewritten as

$$\eta \approx \frac{V_{,\phi\phi}}{V} - \epsilon\,. \tag{1.95}$$

Thus the smallness of the slow-roll parameters is guaranteed when the potential is very flat (compared to its magnitude) over an extended field range.

A useful estimate for the number of e-folds N is given by

$$N = \ln\left(\frac{a_{end}}{a_{beg}}\right) = \int H dt = \int \frac{1}{\sqrt{2\epsilon}} d\phi \approx \int \frac{V_{,\phi}}{V} d\phi \approx \frac{V}{V_{,\phi\phi}} \tag{1.96}$$

Thus $1/|\eta|$ provides a useful estimate for the number of e-folds, and one typically needs $|\eta| \sim \mathcal{O}(10^{-2})$ in successful models.

A second important example is provided by a quadratic potential, i.e. by a simple mass term

$$V(\phi) = \frac{1}{2}m^2\phi^2\,. \tag{1.97}$$

In this potential $\epsilon \approx V_{,\phi}^2/(2V^2) = 2/\phi^2$ and hence inflation occurs as long as $|\phi| > \sqrt{2}$ (in reduced Planck units). One may not think of a ϕ^2 potential as being very flat, but as we can see here, this potential's flatness relative to its magnitude becomes small as the field value is increased. Note that η is automatically very small in the slow-roll regime, as $\eta \approx V_{,\phi\phi}/V - \epsilon \approx 0$. The number of slow-roll e-foldings is given by the integral

$$N = \int \frac{1}{\sqrt{2\epsilon}} d\phi = \int \frac{1}{2}\phi\phi = \frac{1}{4}(\phi_{beg}^2 - \phi_{end}^2)\,. \tag{1.98}$$

As we just saw, inflation ends when ϕ reaches the value $\sqrt{2}$ and thus we require $|\phi_{beg}| \geq 15$. In this potential, and more generally in potentials given by a simple monomial, the field range must thus be substantially larger than one Planck mass. In order for classical general relativity to still be a meaningful approximation, we must make sure that the energy scale remains below the Planck scale, i.e. we need $V_{beg} = \frac{1}{2}m^2\phi_{beg}^2 \ll 1$, implying $m \ll 1/10$. Thus, the inflaton field must be very light (compared to the Planck scale).

Inflation ends when ϵ becomes larger than 1. At that time, the inflaton fields undergoes damped oscillations around the potential minimum. The theory of *reheating* assumes that at that point the field quanta start to decay into the particles of the standard model. This of course requires that the inflaton field couples to the standard model particles - an assumption that a realistic inflationary model must fulfil. As the inflaton quanta decay into relativistic particles, the universe heats up and the radiation dominated expansion period can begin. Note that the term *re*-heating is a little misleading, as there is currently no reason to assume the universe was hot at any previous time. Also, it should be pointed out that reheating remains only insufficiently understood, and that a detailed understanding depends on the precise way in which the hypothetical inflaton field fits together with the standard model of particle physics. There is one important constraint on reheating, which is that the temperature that is reached should remain below the grand unified scale - otherwise at symmetry breaking many topological defects would be formed via the Kibble mechanism, and none have been observed in our universe.

2 Lecture II

2.1 Perturbations

Inflation has the potential to explain the small temperature fluctuations observed in the CMB. For a heuristic picture, consider fluctuations in the scalar field

$$\phi(t, \vec{x}) = \bar{\phi}(t) + \delta\phi(t, \vec{x}) \tag{2.1}$$

Since the scalar field drives inflation, small fluctuations lead to slightly different expansion histories in different regions (i.e. reheating will happen at different times in different regions), thus some regions expand a little more and are a little colder, while other regions expand a little less and are still hotter. Hence such fluctuations in the scalar field can lead to the primordial temperature fluctuations. But why are there fluctuations in the scalar field in the first place? As we will see, the origin for these fluctuations comes from quantum fluctuations (cf. uncertainty principle) that get amplified during inflation. The calculation is rather technical, especially because in GR we have the freedom to change coordinates, so that it is a priori not clear what constitutes a physical fluctuation and what represents just a different choice of coordinates. We will now present the calculation in some detail.

We start by looking at perturbations of the metric $g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}$. The perturbed metric can be written as

$$ds^2 = -(1+2A)dt^2 + 2a(t)(B_{,i} + G_i)dx^i dt + a^2(t)[\delta_{ij} + h_{ij}]dx^i dx^j \tag{2.2}$$

with

$$h_{ij} = 2\psi\delta_{ij} + 2\partial_i\partial_j E + 2_{(}\partial_i E_{j)} + \gamma_{ij}, \tag{2.3}$$

with the additional conditions

$$\partial^i G_i = 0, \quad \partial^i E_i = 0, \quad \gamma^i_i = 0, \quad \partial^i \gamma_{ij} = 0. \tag{2.4}$$

The perturbations can be separated into three categories, according to their transformation properties with respect to three-dimensional rotations (or, expressed more prosaically, according to their spatial $i, j, ...$ index structure):

- scalar – no index : A, ψ, B, E
- vector – one i index, divergence free: G_i, E_i
- tensor – symmetric ij indices, transverse and traceless: γ_{ij}

Vector perturbations decay in all the models we will consider in these lectures, hence we will ignore them and focus only on scalar and tensor modes.

2.1.1 Gauge Transformations

We consider a small local coordinate change

$$x \to x'^\mu = x^\mu + \xi^\mu, \tag{2.5}$$

where $\xi^\mu = (\xi^0, \xi^i)$ with $\xi^i = \xi^i_T + \partial^i\xi$. Here ξ is a scalar and $\partial_i\xi^i_T = 0$ is a divergence free 3-vector. Thus ξ^0 and ξ are the two scalar transformation parameters. In general, under such a coordinate transformation the fields and their perturbations change. It turns out to be convenient to look at *gauge transformations*, where the background fields remain unperturbed and the entire change is accounted for by the field perturbations. Thus, for scalar and tensor quantities we write

$$s(x^\mu) \;\to\; s'(x^\mu) = s(x^\mu) + \Delta s(x^\mu) \tag{2.6}$$

$$t_{\rho\sigma}(x^\mu) \;\to\; t'_{\rho\sigma}(x^\mu) = t_{\rho\sigma}(x^\mu) + \Delta t_{\rho\sigma}(x^\mu) \tag{2.7}$$

Then, using the definition of a tensor quantity $t'_{\rho\sigma}(x'^\mu) = t_{\lambda\kappa}(x^\mu)\frac{\partial x^\lambda}{\partial x'^\rho}\frac{\partial x^\kappa}{\partial x'^\sigma}$ we find that scalars and tensors transform as

$$\Delta s(x) = s'(x) - s(x) = s'(x') - \xi^\mu\partial_\mu s(x) - s(x) = -\xi^\mu\partial_\mu s(x) \tag{2.8}$$

$$\Delta t_{\rho\sigma}(x) = t'_{\rho\sigma}(x') - t_{\rho\sigma}(x) = -\xi^\mu(x)\partial_\mu t_{\rho\sigma}(x) - t_{\lambda\sigma}(x)\partial_\rho\left(\xi^\lambda(x)\right) - t_{\lambda\rho}(x)\partial_\sigma\left(\xi^\lambda(x)\right) \tag{2.9}$$

under gauge transformations. Thus we can calculate how the metric (2.2) transforms. The scalars transform as

$$
\begin{aligned}
A &\rightarrow A + \dot{\xi}_0 & (2.10)\\
B &\rightarrow B + \frac{1}{a}(-\xi_0 - \dot{\xi} + 2H\xi) & (2.11)\\
\psi &\rightarrow \psi + H\xi_0 & (2.12)\\
E &\rightarrow E - \frac{1}{a^2}\xi, & (2.13)
\end{aligned}
$$

while the tensors γ_{ij} are left unchanged. Even though the scalar metric perturbations transform under gauge transformations, one can find *gauge-invariant* quantities that do not. The best known of these are the Bardeen potentials

$$
\begin{aligned}
\Phi &= A + \frac{d}{dt}[a(B - a\dot{E})] & (2.14)\\
\Psi &= \psi - aH(B - a\dot{E}), & (2.15)
\end{aligned}
$$

which do not transform. So if we want to make sure that we are talking about physical quantities only, there are in general two options:

1. Work with gauge-invariant quantities [1]
2. Fix a gauge, i.e. fix the a priori arbitrary ξ^0, ξ functions such that the perturbations become unambiguous

A useful example is *comoving gauge*, where one sets the off-diagonal part of the stress-energy tensor to zero, $\delta T_{0i} = 0$. In this gauge the perturbed spacetime is still comoving with the background matter, e.g. for a scalar field the constant ϕ surfaces remain constant time slices even in the perturbed spacetime. Note that Eq. (2.8) implies that a scalar field perturbation transforms under a change of time slicing ($t \rightarrow t + \xi^0 = t + \delta t$) as

$$\delta\phi \rightarrow \delta\phi - \dot{\phi}\delta t. \tag{2.16}$$

Given that

$$T_{0i} \sim \partial_i \delta\phi(t, x^i)\dot{\phi}(t), \tag{2.17}$$

we can see that we need

$$\delta\phi_{com} = 0 \tag{2.18}$$

to stay in comoving gauge, i.e. we need

$$\delta t = \frac{\delta\phi}{\dot{\phi}}. \tag{2.19}$$

[1]There is one caveat here: the definitions of the gauge-invariant quantities must remain well-defined throughout, i.e. the gauge transformations themselves cannot become degenerate or singular.

Keeping in mind that

$$\psi \to \psi + H\delta t \tag{2.20}$$

we can see that

$$\mathcal{R} = \psi - H\frac{\delta\phi}{\dot{\phi}} \tag{2.21}$$

has an invariant meaning. It is the *comoving curvature perturbation*. In general it is defined via

$$\mathcal{R} = \psi + \frac{H}{\rho + p}\delta q \tag{2.22}$$

where $\delta T^0_i = \partial_i \delta q$. In comoving gauge we simply have $\mathcal{R}_{com} = \psi$ and thus in that gauge the comoving curvature perturbation specifies the fluctuations in the scale factor (on constant-ϕ surfaces). Hence its name. The comoving curvature perturbation is important because under certain conditions it is conserved on very large scales, and moreover it is directly related to what we observe in the CMB.

2.1.2 Perturbed Action

The cosmological background solutions that we are interested in effectively split spacetime into space and time, as the background fields (the metric and the scalar field) depend on time alone. Thus surfaces of constant scalar field correspond to equal-time slices through the spacetime. In calculating the perturbations around these cosmological spacetimes, it is useful to split the metric in a similar fashion, as first described by Arnowitt-Deser-Misner. This ADM formalism thus starts by considering a metric of the form

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt), \tag{2.23}$$

where N is the *lapse* function (changing N corresponds to changing the time coordinate) and N^i is the *shift*. We will plug this metric into the action:

$$S = \int dx^4 \sqrt{-g}[\frac{1}{2}R - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)]. \tag{2.24}$$

Note that $\sqrt{-g} = N\sqrt{h}$. We relate the 4-dimensional Ricci scalar R to the 3-dimensional Ricci scalar $R^{(3)}$ via the contracted Gauss-Codazzi equation:

$$R = R^{(3)} + K^{ij}K_{ij} - K^2, \tag{2.25}$$

where K_{ij} is the extrinsic curvature. It is useful to define

$$E_{ij} = NK_{ij}, \tag{2.26}$$

with

$$E_{ij} \;=\; \frac{1}{2}(\dot{h}_{ij}) - \nabla_i N_j - \nabla_j N_i = \frac{1}{2}(\dot{h}_{ij}) - 2\nabla_{(i}N_{j)}, \tag{2.27}$$

$$E = E^i_i = h^{ij} E_{ij}. \tag{2.28}$$

The action then becomes

$$S = \frac{1}{2}\int d^4x\sqrt{h}N\left(R^{(3)} - h^{ij}\partial_i\phi\partial_j\phi - 2V\right) \tag{2.29}$$

$$+\frac{1}{2}\int d^4x\sqrt{h}N^{-1}\left(E_{ij}E^{ij} - E^2 + (\dot{\phi} - N^i\partial_i\phi)^2\right) \tag{2.30}$$

We will choose comoving gauge, in which the inflaton fluctuation vanishes

$$\delta\phi = 0 \tag{2.31}$$

and the scalar degree of freedom is represented by the comoving curvature perturbation $\mathcal{R}$. The metric for the spatial hypersurfaces is given by:

$$h_{ij} = a^2[(1 + 2\mathcal{R})\delta_{ij} + \gamma_{ij}] \tag{2.32}$$

Note that this gauge is fully fixed: ξ^0 is chosen to remove $\delta\phi$ and ξ to remove E, while A is absorbed into the lapse function and B into the shift. First, we drop the tensor fluctuations γ_{ij} – at linear order, they evolve independently, and thus we can deal with them separately later on.

In the ADM formalism, the lapse and shift functions appear in the action without derivatives, i.e. they are Lagrange multipliers and their equations of motion are constraint equations:

$$\frac{\delta\mathcal{L}}{\delta N} = R^{(3)} - N^{-2}(E_{ij}E^{ij} - E^2) - N^{-2}\dot{\phi}^2 - 2V = 0 \tag{2.33}$$

$$\frac{\delta\mathcal{L}}{\delta N_i} = \nabla_i[N^{-1}(E^i_j - \delta^i_j E)] = 0 \tag{2.34}$$

We chose the following ansatz to solve the equations (2.33) and (2.34),

$$N_i = \partial_i\psi + \tilde{N}_i \tag{2.35}$$

$$\partial^i\tilde{N}_i = 0 \tag{2.36}$$

$$N = 1 + \alpha, \tag{2.37}$$

where one then expands $\alpha = \alpha^{(1)} + \alpha^{(2)} + ...$, $\psi = \psi^{(1)} + \psi^{(2)} + ...$, $\tilde{N}_i = \tilde{N}_i^{(1)} + \tilde{N}_i^{(2)} + ...$ with the subscripts $(1), (2), ...$ denoting the order of $\mathcal{R}$. One then finds the following solutions to first order:

$$N = 1 + \frac{\dot{\mathcal{R}}}{H} \tag{2.38}$$

$$N_i = \partial_i\psi, \qquad \psi = -\frac{\mathcal{R}}{H} + a^2\frac{\dot{\phi}^2}{2H^2}\partial^{-2}\dot{\mathcal{R}} \tag{2.39}$$

We can now write out explicitly the second order action. After using the background equations of motion and integrating by parts, one obtains

$$S_2 = \int d^3x dt\, \epsilon \left[a^3 \dot{\mathcal{R}}^2 - a(\partial_i \mathcal{R})^2\right] \tag{2.40}$$

Note that in the quadratic action, the potential of the theory does not appear explicitly. However, the evolution of the scale factor a is of course determined via the potential. Note also that the comoving curvature perturbation is massless. This is important, as it implies that on large scales, where spatial gradients may be neglected, $\mathcal{R}$ is conserved. Incidentally, the conservation of $\mathcal{R}$ shows that inflation is an attractor (if it were a repeller, perturbations would blow up over time).

In general, given that the field $\mathcal{R}$ is small we can expand the Lagrangian in a power series of the scalar field.

$$S = \int d^4x \mathcal{L}[\mathcal{R}(x)] = \int d^4x \mathcal{L}^{(2)}[\mathcal{R}(x)] + \mathcal{L}^{(3)}[\mathcal{R}(x)] + ... \tag{2.41}$$

We just derived $\mathcal{L}^{(2)}$, which describes the free propagation of the field. At next order, $\mathcal{L}^{(3)}$ describes the self-interaction of the field. Since the perturbations are small, in many models $\mathcal{L}^{(3)}$ can be neglected. The path integral then becomes a Gaussian integral (after Wick rotation), and for this reason one says that the perturbations are highly Gaussian. The term $L^{(3)}$ then typically represents the leading *non-Gaussian* corrections - the PLANCK satellite has put rather stringent upper bounds on their size. However, as the observations are continually improving, corrections arising from $\mathcal{L}^{(3)}$ and perhaps even $\mathcal{L}^{(4)}$ may be detected in the near future.

2.1.3 Quantisation

We would like to teat the small curvature perturbations quantum-mechanically, assuming a classical curved background (justifying this assumption is one of the central aims of "quantum cosmology"). For the quantisation of the perturbations, it is useful to define the canonically normalised Mukhanov-Sasaki variable

$$v = z\mathcal{R} \quad , \quad z^2 = a^2 \frac{\dot{\phi}^2}{H^2} = 2a^2\epsilon \tag{2.42}$$

Switching to conformal time, the action now becomes

$$S_2 = \int d^3x d\tau [(v')^2 - (\partial_i v)^2 + \frac{z''}{z} v^2] \tag{2.43}$$

The action is quadratic, and will thus lead to linear equations of motion. This implies that it will be useful to expand the perturbations in their Fourier modes

$$v(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} v_k(\tau) e^{i\mathbf{kx}} \tag{2.44}$$

Spatial isotropy implies that the Fourier modes v_k only depend on the modulus $k = |\mathbf{k}|$. Then the equation of motion for each Fourier mode is

$$v_k'' + (k^2 - \frac{z''}{z})v_k = 0 \tag{2.45}$$

The linearity of the equation implies that each mode evolves independently, and there is no mode mixing.

We can quantise the system by promoting the fields to operators, and writing these new operators as a linear combination of annihilation and creation operators

$$\hat{v}_k = f\hat{a}_{\mathbf{k}} + f^{\star}\hat{a}^{\dagger}_{-\mathbf{k}}. \tag{2.46}$$

We then require the annihilation/creation operators to satisfy the canonical quantisation condition

$$[\hat{a}_{\mathbf{k}}, \hat{a}^{\dagger}_{-\mathbf{k}'}] = (2\pi)^2\delta(\mathbf{k} - \mathbf{k}') \tag{2.47}$$

Here $f, f^{\star}$ are time-dependent (complex) solutions of the equations of motion (2.45). The following quantity (called the Wronskian) is a constant of motion,

$$ff^{\star\prime} - f^{\star}f' = i, \tag{2.48}$$

where we have fixed the right hand side in such a way as to ensure the canonical normalisation of the mode functions.

De Sitter Limit

An important limit, in which the calculation simplifies considerably, is the de Sitter limit where the Hubble rate is constant

$$\epsilon \to 0 \quad , \quad H = const. \tag{2.49}$$

In this case we have that $z = a\sqrt{2\epsilon}$ and

$$a = -\frac{1}{H\tau}\,, \qquad -\infty < \tau < 0 \tag{2.50}$$

Note the range of the conformal time coordinate! It then follows that

$$\frac{z''}{z} = \frac{a''}{a} = \frac{2}{\tau^2} \tag{2.51}$$

so that in the de Sitter limit the mode equation becomes

$$v_k'' + (k^2 - \frac{2}{\tau^2})v_k = 0. \tag{2.52}$$

It is important to understand this equation qualitatively at first: at early times and/or on small scales ($|k\tau| \gg 1$) the $2/\tau^2$ term may be dropped, so that we simply find the mode equation in Minkowski space

$$v_k'' + k^2 v_k = 0 \tag{2.53}$$

That is the equation of motion for a harmonic oscillator and it has the positive-frequency solution

$$f_k = \frac{1}{\sqrt{2k}} e^{-ik\tau} \tag{2.54}$$

In the present context, this early-time solution is often referred to as the Bunch-Davies initial condition. At late times and/or on large scales ($|k\tau| \ll 1$), the equation changes to a harmonic oscillator with time-dependent tachyonic mass

$$v_k'' - \frac{2}{\tau^2} v_k = 0 \tag{2.55}$$

in such a way that the solution now grows, $f \propto 1/\tau$. Thus we learn that the fluctuations get amplified as soon as they *exit the horizon*, i.e. from the moment when $|k\tau| \sim 1$. These general features remain true in all cases of interest.

One can in fact solve the de Sitter mode equation exactly, and the solution is given by

$$f_k = \frac{\alpha}{\sqrt{2k}} e^{-ik\tau}\left(1 - \frac{i}{k\tau}\right) + \frac{\beta}{\sqrt{2k}} e^{ik\tau}\left(1 + \frac{i}{k\tau}\right) \tag{2.56}$$

where α, β are constants. If we want the modes to be in the Bunch-Davies vacuum at early times, then we must choose $\alpha = 1, \beta = 0$.

Statistics

We can now calculate the 2-point correlation function of the comoving curvature perturbation

$$\langle \mathcal{R}_{\mathbf{k}}(\tau) \mathcal{R}_{\mathbf{k}'}(\tau) \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') \frac{H^2}{\epsilon} \frac{1}{k^3} (1 + k^2 \tau^2), \tag{2.57}$$

which at late times tends to a constant (note that here we have assumed that we are close to de Sitter, but not exactly in the de Sitter limit where $\epsilon = 0$). In the formula above, if there is a slight time dependence in the Hubble rate and the slow-roll parameter ϵ, then these quantities are to be evaluated at horizon crossing, i.e. when $a(t_\star)H(t_\star) = k$.

Often one defines the power spectrum $P_{\mathcal{R}}$ and the variance Δ_s^2 of the scalar perturbations via

$$\langle \mathcal{R}_{\mathbf{k}}(\tau) \mathcal{R}_{\mathbf{k}'}(\tau) \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}')\, P_{\mathcal{R}}, \quad \Delta_s^2 = \frac{k^3}{2\pi^2} P_{\mathcal{R}} \tag{2.58}$$

With these definitions the real space variance of the comoving curvature perturbation is then given by

$$\langle \mathcal{R}\mathcal{R} \rangle \sim \int_0^\infty d\mathrm{ln}k\, \Delta_s^2\,. \tag{2.59}$$

Also, one conventionally writes the variance in power-law form

$$\Delta_s^2 = A(k_0)\left(\frac{k}{k_0}\right)^{n_s-1} \tag{2.60}$$

where k_0 denotes a reference scale and A the amplitude at that scale. Here n_s is called the spectral index.

Thus, for approximately de Sitter space, the variance is given by

$$\Delta_s^2 = \frac{H_\star^2}{8\pi^2\epsilon_\star} \tag{2.61}$$

The fact that this is independent of k means that we have a *scale-invariant* spectrum $n_s = 1$. A more precise treatment leads to small but important corrections, as we will now see.

General Slow-Roll Case

We can also solve for the mode functions more generally, making the slow-roll approximation $\epsilon \ll 1, \eta \ll 1$. We can then expand the z''/z term in the mode equation to first order in slow roll parameters,

$$\frac{z''}{z} = a^2H^2[2 + 2\epsilon - 3\eta + \mathcal{O}(\varepsilon^2)] \tag{2.62}$$

Defining

$$\nu = \frac{3}{2} + 2\epsilon - \eta + \mathcal{O}(\varepsilon^2) \tag{2.63}$$

we can rewrite the mode equation as

$$v_k'' + (k^2 - \frac{\nu^2 - \frac{1}{4}}{\tau^2})v_k = 0 \tag{2.64}$$

This is a Bessel equation with the general solution

$$f_k(\tau) = \sqrt{-\tau}[\alpha_k H_\nu^{(1)}(-k\tau) + \beta_k H_\nu^{(2)}(-k\tau)] \tag{2.65}$$

where $H^{(1,2)}$ are Hankel functions of the first and second kind, and α_k, β_k are constants. We can fix these constants by imposing once again Bunch-Davies boundary conditions at early times and on small scales. For this we need the limit

$$\lim_{x\to\infty} H_\nu^{(1)}(x) = \sqrt{\frac{2}{\pi x}}e^{[i(x-\frac{\nu\pi}{2}-\frac{\pi}{4})]}, \quad \lim_{x\to\infty} H_\nu^{(2)}(x) = \sqrt{\frac{2}{\pi x}}e^{[-i(x-\frac{\nu\pi}{2}-\frac{\pi}{4})]} \tag{2.66}$$

Thus we must choose

$$\alpha_k = \sqrt{\frac{\pi}{4}} \quad , \quad \beta_k = 0\,. \tag{2.67}$$

Having found the appropriate mode functions, we can study their late time limit as well.

For this we need the relation

$$\lim_{x\to 0} H_\nu^{(1)}(x) = \frac{i}{\pi}\Gamma(\nu)(\frac{x}{2})^{-\nu} \tag{2.68}$$

leading to the late-time expression

$$f_k(\tau) = \frac{2^\nu i}{\sqrt{4\pi}}\Gamma(\nu)\frac{1}{k^\nu(-\tau)^{\nu-1/2}} \quad (|k\tau| \ll 1) \tag{2.69}$$

$$\approx \frac{i}{\sqrt{2}}\frac{1}{k^{3/2+2\epsilon-\eta}(-\tau)^{1+2\epsilon-\eta}}. \tag{2.70}$$

Here Γ is the factorial function, and we have used $\Gamma(\nu) \approx \Gamma(3/2) = \sqrt{\pi}/2$. Using

$$\tau = -\frac{1}{aH(1-\epsilon)} \tag{2.71}$$

we can see that the power spectrum is (approximately) as above in the de Sitter limit,

$$\Delta_s^2 = \frac{H_\star^2}{8\pi^2\epsilon_\star}, \tag{2.72}$$

while the k-dependence of the mode function translates into a spectral index

$$n_s = 1 - 4\epsilon + 2\eta. \tag{2.73}$$

Thus, we may typically expect slow-roll models of inflation to yield a slight red tilt, i.e. a power spectrum slightly smaller than 1. This tilt is referred to as red since it corresponds to having more power on larger scales compared to a pure scale-invariant spectrum. A spectral index above 1 is referred to as a blue tilt.

Current observations by the PLANCK satellite indicate that the variance has a value of $\Delta_s^2 = (2.20 \pm 0.12) \times 10^{-9}$, while the spectral index has been measured to be $n_s = 0.9600.015$ at 95% confidence level. Note that inflation does not predict the variance - it has to be fitted after the measurement. Thus inflation, in its current form, does not explain the amplitude of the perturbations. The red tilt of the spectral index however fits well with general slow-roll expectations.

2.1.4 Tensor modes

We can now repeat this calculation for tensor fluctuations. In fact the calculation proceeds in an entirely analogous fashion, hence we will only sketch it here. As stated earlier, at linear order the tensor modes are evolving independently of scalar and vector modes, and so in the present section it is sufficient to consider the perturbed metric

$$ds^2 = -dt^2 + a^2(t)[\delta_{ij} + \gamma_{ij}(t, \underline{x})]dx^i dx^j \tag{2.74}$$

with

$$\gamma^i{}_i = \partial^i\gamma_{ij} = 0\,. \tag{2.75}$$

We can transform to Fourier space and decompose the tensor perturbation into two polarisation states

$$\gamma_{ij} = \int \frac{d^3k}{(2\pi)^3} \sum_{s=+,\times} \epsilon^s_{ij} \frac{2}{a} v^s_{\underline{k}} e^{i\underline{k}\cdot\underline{x}} \tag{2.76}$$

with

$$\epsilon_{ii} = k^i \epsilon_{ij} = 0 \tag{2.77}$$

where $v^s_{\underline{k}}$ is the canonically normalised variable. This variable satisfies the equation of motion

$$v''_k + \left(k^2 - \frac{a''}{a}\right) v_k = 0\,, \tag{2.78}$$

where a prime denotes a derivative with respect to conformal time τ, with $dt \equiv a d\tau$ as usual. During inflation

$$\tau = -\frac{1}{aH} \quad (-\infty < \tau < 0), \qquad \frac{a''}{a} = \frac{2}{\tau^2} \tag{2.79}$$

so that we obtain

$$v''_k + \left(k^2 - \frac{2}{\tau^2}\right) v_k = 0\,. \tag{2.80}$$

Quantising as before, and imposing the same boundary conditions, we find that at early times $|k\tau| \gg 1$ the solution is oscillatory

$$v_k = \frac{\sqrt{\hbar}}{\sqrt{2k}} e^{-ik\tau} \tag{2.81}$$

at late times $|k\tau| \leq 1$ the solution is growing

$$v_k \approx \frac{\sqrt{\hbar}}{\sqrt{2} k^{3/2} \tau} e^{-ik\tau} \tag{2.82}$$

and thus gravitational waves get *amplified* during inflation. The variance is given by

$$\Delta^2_t = \frac{4}{a^2 \pi^2} k^3 |v_k|^2 = \frac{2}{\pi^2} \frac{H^2_\star}{M^2_{Pl}} \tag{2.83}$$

where $H_\star$ denotes the value of H at horizon exit, i.e. when $|k\tau| = 1$. Note the important result that the amplitude of the gravitational waves thus produced depends only on the Hubble rate during inflation. Here one also defines a spectral index n_t via an assumed power law form

$$\Delta^2_t(k) = A_t(k_0) \left(\frac{k}{k_0}\right)^{n_t}. \tag{2.84}$$

where A_t is the amplitude at a reference scale k_0. Note that according to this conventional definition $n_t = 0$ corresponds to scale invariance. More precise calculations involving Hankel functions, as performed above for the scalar case, lead to the slow-roll result

$$n_t = -2\epsilon. \tag{2.85}$$

The tensor-to-scalar ratio r is defined by

$$r \equiv \frac{\Delta_t^2}{\Delta_s^2} = 16\epsilon \tag{2.86}$$

We can infer Δ_s^2 from observations of the CMB

$$\Delta_s^2 = \frac{1}{8\pi^2}\frac{H_\star^2}{M_{Pl}^2}\frac{1}{\epsilon_\star} = 2 \times 10^{-9} \tag{2.87}$$

Since $H^2 \approx V$ we immediately obtain the energy scale during inflation

$$V^{1/4} = \left(\frac{r}{0.01}\right)^{1/4} \times 10^{16} GeV. \tag{2.88}$$

Thus, assuming an inflationary origin, a measurement of r allows one to immediately obtain the energy scale during inflation, and thus, in conjunction with a measurement of the scalar amplitude Δ_s^2, a value for the slow-roll parameter ϵ.

We can also rewrite

$$r = 16\epsilon = 8\left(\frac{d\phi}{dN}\right)^2 \tag{2.89}$$

Integrating and assuming r constant we get the *Lyth bound*

$$\frac{\Delta\phi}{M_{Pl}} \approx 2 \times \sqrt{\frac{r}{0.01}} \tag{2.90}$$

for $N = 60$ e-folds of inflation. This equation shows that in order to obtain a gravitational signal large enough to be observable, the scalar field must typically travel a distance of at least one Planck unit in field space. Models in which this is the case are often referred to as *large-field models.*

2.1.5 Example: $m^2\phi^2$ Inflation

For this potential we calculated the slow-roll parameters earlier, finding

$$\epsilon = 2/\phi^2 \quad \eta \sim \mathcal{O}(\epsilon^2) \tag{2.91}$$

Also, we found that the number of e-folds of inflation remaining is roughly given by $N = \frac{1}{4}\phi^2$. Given that the modes that we observe on the CMB now would have been produced about 60 e-folds before the end of inflation, we have $N_{CMB} = 60$. The amplitude of the scalar perturbations is given by

$$\frac{H^2}{8\pi^2\epsilon} \approx \frac{V}{8\pi^2\epsilon} \approx V = \frac{1}{2}m^2\phi_{CMB}^2 = 2N_{CMB}m^2 \tag{2.92}$$

Thus, to obtain the observed variance of 10^{-9} we have to assume an inflaton mass of about $m \approx 10^{-6}$ in natural units. Having fixed these relations by hand, we then obtain an expected value for the spectral index of

$$n_s \approx 0.96 \tag{2.93}$$

and a tensor-to-scalar ratio

$$r \approx 0.13\,. \tag{2.94}$$

The observational bounds are so good that we will soon know whether this model (which is surely the most widely used model of inflation, and arguably the simplest) remains viable or is ruled out.

2.1.6 Quantum To Classical

The fact that the quantised curvature perturbations start behaving increasingly classically can best be seen in the Schrödinger picture. We start again from the Lagrangian for each Fourier mode,

$$L = \frac{1}{2}v'^2 - \frac{1}{2}\left(k^2 - \frac{z''}{z^2}\right)v^2 \tag{2.95}$$

The canonical momentum is then given by

$$\pi = \frac{\partial L}{\partial v'} = v' \tag{2.96}$$

Note that, making use of the Wronskian condition, the annihilation operator can now be rewritten in terms of the field and its momentum

$$i\hat{a} = f^{\star\prime}\hat{v} - f^{\star}\hat{\pi} \tag{2.97}$$

The vacuum state is defined by

$$\hat{a}|0>= 0 \tag{2.98}$$

Then using the expression above and the canonical replacement $\pi \to -i\frac{\partial}{\partial v}$, we find that the ground state Schrödinger wave function is given by

$$\Psi(v) = N \exp\left(-\frac{1}{2}Cv^2\right) \tag{2.99}$$

where N is a normalisation factor. Here C is the correlator and it is given by

$$C = -i\frac{f^{\star\prime}}{f^{\star}} \tag{2.100}$$

At early times, we know that the mode functions are approximately in the Bunch-Davies state, such that $f^{\star} \approx \frac{1}{\sqrt{2k}}e^{-ik\tau}$. Then we have

$$C \approx k \qquad (|k\tau| \gg 1) \tag{2.101}$$

On the other hand, at late times, we have the solution (2.69). Using the asymptotic forms of the Hankel functions, one finds that

$$C \approx -\frac{2\pi}{\Gamma(\nu)^2\tau}(\frac{-k\tau}{2})^\nu - i\left(\frac{1}{2}-\nu\right)\frac{1}{\tau} \qquad (|k\tau| \ll 1) \tag{2.102}$$

$$\approx k^3\tau^2 + \frac{i}{\tau} \tag{2.103}$$

where we have kept the leading real and imaginary contributions. It is now useful to recall the WKB criterion for classicality: a wavefunction behaves approximately classically when its amplitude varies slowly compared to the variation of its phase. Clearly, at early times the wavefunction is very quantum, while at late times the variation of its amplitude goes to zero while its phase is continually speeding up. Thus, as time goes on, the perturbations behave increasingly classically. Note also that the real part of C becomes small at late times. This means that the dispersion of the fluctuations modes becomes large – in other words, these modes get copiously produced.

2.1.7 Relation To Observations - Angular Power Spectrum and Polarisation

The small anisotropies in the CMB, which manifest themselves as orientation-dependent fluctuations in its temperature, are usefully characterised by expanding them in spherical harmonics Y_l^m,

$$\delta T(\hat{n}) = \sum_{lm} a_{lm} Y_l^m(\hat{n}), \tag{2.104}$$

where $\hat{n}$ denotes the direction in the sky. One can form the rotationally invariant quantities

$$C_l^{\mathrm{TT}} = \frac{1}{2l+1}\sum_m \langle a_{lm}^* a_{lm}\rangle. \tag{2.105}$$

It is these C_l^{TT} coefficients that are often plotted in graphical representations of the statistical properties of the cosmic background radiation; see Fig. 3. The characteristic pattern of peaks and troughs is well understood in terms of the physics of the radiation-matter plasma that dominated the universe prior to recombination, *if* one assumes the pre-existence of a nearly scale-invariant fluctuation spectrum at even earlier times.

We will give a heuristic explanation of the pattern seen in Fig. 3. Detailed calculations are usually performed numerically, and involve solving the full Boltzmann equations describing the various matter components in the early universe. For our purposes, it is sufficient to have a qualitative understanding of the results of these calculations. Our strategy will be to assume the existence of temperature fluctuations with a scale-invariant spectrum at early times, and then explain how this reproduces the pattern seen in the figure above. On large scales (small l), the perturbation modes do not evolve much, and hence we may expect a flat plateau for $l < 20$. On these scales we get a direct glimpse of the primordial power spectrum, giving us an estimate of the amplitude of the fluctuations. However, observations on these scales are severely limited by cosmic variance, i.e. by the fact that for such long-wavelength modes only a few fit on the sky and thus random unknowable fluctuations play an important role. On intermediate scales, there are acoustic oscillations leading to the characteristic peaks

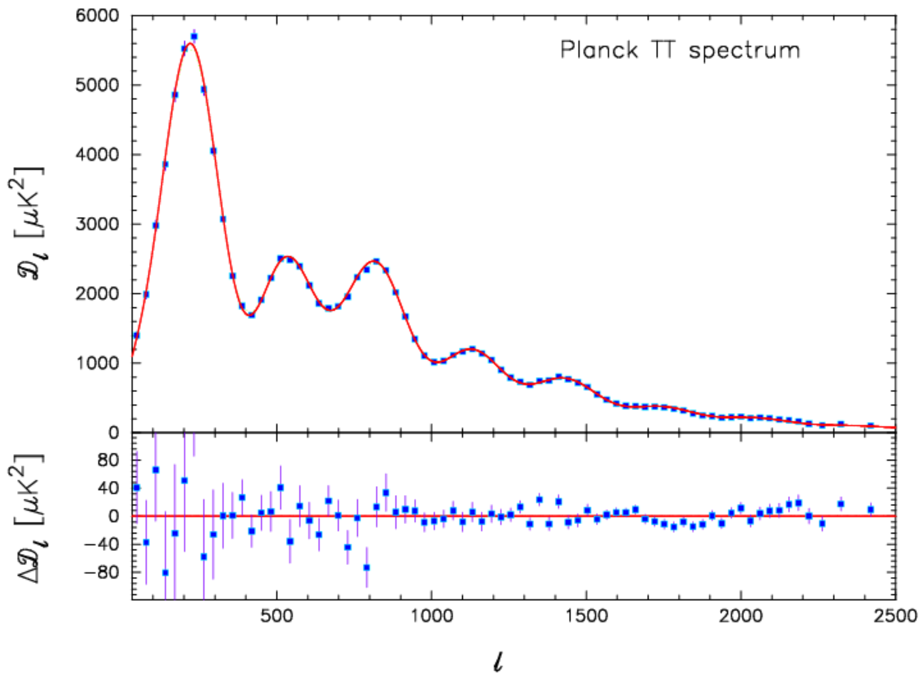

Figure 3: This graph shows the angular power spectrum of the CMB radiation, *i.e.* it shows the amplitude of the temperature fluctuations (to be precise: what is depicted on the vertical axis is the combination $l(l+1)C_l^{\rm TT}/2\pi$) as a function of the multipole moment l. *Credit: ESA / PLANCK*

and troughs in the spectrum. Indeed, in the early radiation-matter plasma, the fluctuations oscillate as acoustic waves, and this can lead to constructive/destructive interference. Here it is important to realise that, since the super-Hubble curvature perturbations are frozen in, they all start their evolution *with the same phase*: when they enter the horizon, they have a vanishing time derivative and then they start oscillating. If this were not the case, but rather if all temporal phases were random, we would not see any such peaks and troughs in the angular power spectrum. Moreover, using known plasma physics, one can calculate the size of the oscillations that are at a peak, i.e. this gives us a standard ruler, and by measuring the angular size on the sky, we can find out (since we know the redshift) what the geometry of the universe is in between last scattering and us. Such measurements allow us to conclude that the spatial geometry of the universe is flat.

At later stages, these density fluctuations collapse to stars and galaxies under the influence of gravity. Thus we arrive at the astonishing picture that all the structure in the universe may have originated from primordial quantum fluctuations!

We should also discuss another aspect of the CMB radiation, namely its polarisation. The light from the CMB, like every light source, has a polarisation that can be described via an intensity matrix I_{ij} in the plane perpendicular to propagation, with

$$I_{ij} = \begin{pmatrix} T+Q & U \\ U & T-Q \end{pmatrix} \tag{2.106}$$

T denotes the temperature anisotropy, while Q and U are the Stokes parameters, and they can be decomposed into spin-2 spherical harmonics

$$(Q \pm iU)(\hat{n}) = \sum_{l,m} a_{\pm 2,lm} \, {}_{\pm 2}Y_{lm}(\hat{n}) \tag{2.107}$$

Particular linear combinations of these coefficients are then called E-mode (symmetric under parity transformations) and B-mode polarisations (anti-symmetric under parity transformations),

$$a_{E,lm} = -\frac{1}{2}(a_{2,lm} + a_{-2,lm}) \qquad a_{B,lm} = -\frac{1}{2i}(a_{2,lm} - a_{-2,lm}) \tag{2.108}$$

This then implies that on top of the temperature-temperature cross-correlation described earlier in this section, we have more possibilities, as we can consider all of the following angular 2-point functions

$$C_l^{\mathrm{XY}} = \frac{1}{2l+1} \sum_m \langle a^*_{\mathrm{X},lm} a_{\mathrm{Y},lm} \rangle \tag{2.109}$$

where now $\mathrm{X}, \mathrm{Y} = T, E, B$. These spectra are useful, because scalar perturbation modes produce only E-modes, while tensor perturbations produce both E-modes and B-modes. These B-modes arise because tensor fluctuations lead to a quadrupole in the radiation field, and then via Thomson scattering in such a quadrupole, B-mode polarisation is generated. Because of this, a detection of *primordial* B-modes in the CMB would be indirect evidence for the presence of gravitational waves in the early universe. In fact, in March 2014 the BICEP2 team has announced the detection of a primordial gravitational wave signal of $r \approx 0.2$ precisely due to the indirect effect of such gravitational waves on the B-mode polarisation of the CMB. Currently, it is however not clear if the B-mode signal really is primordial or if it may rather be caused by dust foregrounds in our galaxy (which also induce B-modes), and therefore it is prudent to wait for better foreground maps, which will hopefully be provided by the PLANCK satellite later this year.

3 Lecture III

3.1 Puzzles of Inflation

Particle physics problems:

- What is the nature of the inflaton?
- How does the inflaton couple to all other particles, in particular to the standard model particles? Once this is known - does reheating really work?
- How does the inflationary potential arise? How can it be so flat over such an extended field range? What determines its magnitude?

Cosmological problems:

- How were the initial conditions for inflation set?
  For inflation to start, you need a super-Hubble sized patch that is uniform to a high

degree (initial anisotropies must already be subdominant in the Friedmann equation in order for inflation to be able to begin) and with the scalar field being up on the potential with very small velocity
Such initial conditions are far from typical - so how do you get them?

- Past-incompleteness
Inflation is necessarily preceded by a singularity - how is this resolved, and what are the implications for inflation? This is another facet of the initial conditions problem.

- The unlikeliness problem
Inflation occurring at a higher Hubble rate would produce a larger universe with more galaxies (because the perturbation amplitude would be larger)
Then why does our universe emerge from a lower part of the potential?

- The entropy problem
In de-Sitter space with expansion rate H the entropy is given by $S = e^{1/H^2}$ - hence the entropy of the inflationary phase was smaller by an amount $e^{10^{100}}$ than the maximal possible entropy (calculated from assuming a cosmological constant at the currently measured value of dark energy)
Also, the entropy of the present universe is about $S_0 \sim 10^{90}$ - again, from a thermodynamic point of view, inflation is highly unlikely. So why did the universe start in such a highly unlikely state? (Note that the unlikeliness and entropy problems lead to tension in roughly opposite directions)

- The multiverse problem - is inflation actually predictive?
Consider one Hubble patch (size $1/H^3$) over one Hubble time $\Delta t = 1/H$ and assume slow-roll inflation. Then the classical motion of the field is $\Delta\phi_{cl} = \dot{\phi}\Delta t = \frac{\dot{\phi}H}{H^2} \sim \frac{V_{,\phi}}{V}$. Meanwhile the quantum evolution is of order $|\Delta\phi_{qu}| \sim \frac{H}{2\pi} \sim V^{1/2}$. Taking as an example $m^2\phi^2$ inflation, we can see that the quantum evolution is typically larger than the classical evolution when $|\phi| > m^{-1/2}$. Over one Hubble time, the universe grows by a factor $e^3 \approx 20$. Thus, for large field values, after one Hubble time there will be 10 Hubble patches where the field is kicked *up* the potential, implying that regions where inflation occurs grow exponentially over time and thus inflation never comes to an end, globally speaking (of course, at any given place inflation eventually does come to an end, but the point is that continually new inflating regions are created). As the field rolls down the potential, quantum jumps also occur, and these change the predictions that we calculated earlier, as they change the slow-roll background that we assumed in our calculations. Because all possible quantum jumps will eventually occur, and because the number of created inflationary regions is infinite, this implies that *all* possible values for the amplitude, the spectral index, non-Gaussianities,..., will be created. Without a measure, inflation is thus not predictive, and thus, strictly speaking, not a scientific theory (as it is unfalsifiable). Is this simple picture of eternal inflation correct? If it is, then what is the correct measure? No good answer to this question has been found to date.

Interestingly, there exists an entirely different approach to early universe cosmology, based on the assumption that instead of an inflationary phase there was a phase of ultra-slow contraction before the hot big bang.

3.2 The Cyclic Universe

3.2.1 The Ekpyrotic Phase

Let us go back to the Friedmann equation in the presence of different matter types, represented here by their energy densities ρ,

$$3H^2 = \left(\frac{-3\kappa}{a^2} + \frac{\rho_m}{a^3} + \frac{\rho_r}{a^4} + \frac{\sigma^2}{a^6} + \ldots + \frac{\rho_\phi}{a^{3(1+w_\phi)}} \right) \tag{3.1}$$

The subscript m refers to non-relativistic matter and includes dark matter, r refers to radiation and σ denotes the energy density of anisotropies in the curvature of the universe.The associated scaling with a can be calculated as follows: consider a metric of the form

$$ds^2 = -dt^2 + a(t)^2 \sum_i e^{2\beta_i} dx^{i2} \tag{3.2}$$

with $\sum_i \beta_i = 0$ so that a denotes the average scale factor, while the β_i parameterise anisotropies/gradients in the x, y, z spatial directions. Then the Friedmann equation picks up a new term

$$3H^2 = \cdots + \frac{1}{2} \sum_i \dot{\beta}_i^2 \tag{3.3}$$

while the ij Einstein equations give

$$\ddot{\beta} + 3H\dot{\beta} = 0 \quad \rightarrow \quad \dot{\beta} \propto \frac{1}{a^3} \tag{3.4}$$

so that the new term in the Friedmann equation scales as $\sum_i \dot{\beta}_i^2 \propto 1/a^6$.

In an expanding universe, as the scale factor a grows, matter components with a slower fall-off of their energy density come to dominate. If there is an inflationary scalar field, the eventually the inflaton, whose energy density is roughly constant, dominates the cosmic evolution and determines the (roughly constant) Hubble parameter while causing the scale factor to grow exponentially, $a \propto e^{Ht}$. We can define the relative energy density in the curvature as $\Omega_\kappa \equiv -\kappa/(a^2 H^2)$ and in the anisotropies as $\Omega_\sigma \equiv \sigma^2/(3a^6 H^2)$. During inflation, these relative densities fall off quickly, and the universe is rendered exponentially flat; as we calculated earlier, the flatness puzzle is then resolved as long as the scale factor grows by at least 60 e-folds.

Now we will show how the same problem can be solved by having a contracting phase before the standard expanding phase of the universe. The Friedmann equation relates the Hubble parameter to the total energy density in the universe, which is the sum of kinetic and potential energy. Now suppose that, instead of a flat potential, the scalar ϕ has a very steep, negative potential, as shown in Fig. 4. As a concrete example, one can model the potential with a negative exponential

$$V(\phi) = -V_0 e^{-c\phi}, \tag{3.5}$$

where $V_)$ and c are constants. In a contracting universe, the argument presented in the previous paragraph is reversed, and one would initially expect the anisotropy term (proportional

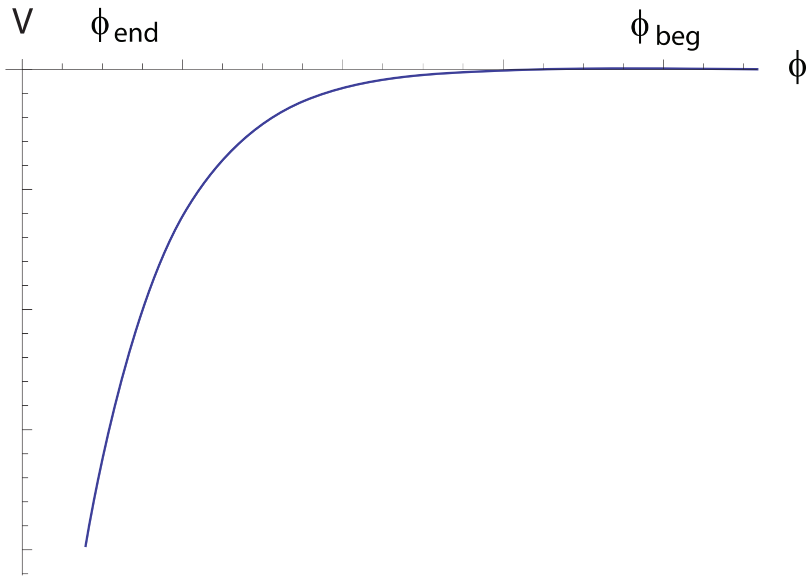

Figure 4: The potential during ekpyrosis is negative and steeply falling; it can be modelled by the exponential form $V(\phi) = -V_0 e^{-c\phi}$.

to a^{-6}) to come to dominate the cosmic evolution. However, if there is a matter component with $w > 1$, then this component will scale with an even larger negative power of a, and hence will come to dominate the cosmic evolution in a contracting universe in the same way as the inflaton comes to dominate in an expanding universe. In fact, for a steep negative exponential potential, neglecting all other matter fields for the moment, the equations of motion are solved by the scaling solution

$$a \propto (-t)^{1/\epsilon}, \quad \phi = \sqrt{\frac{2}{\epsilon}} \ln\left[-\left(\frac{V_0 \epsilon^2}{(\epsilon - 3)}\right)^{\frac{1}{2}} t\right], \quad \epsilon = \frac{c^2}{2}, \tag{3.6}$$

This is called a scaling solution because in the equations of motion each term has the same time dependence, as you should verify. Note that ϵ here has the same definition as in inflation, and it is related to the equation of state $w = p/\rho$ via the standard relation

$$\epsilon = \frac{3}{2}(1 + w). \tag{3.7}$$

For $w > 1$ this implies $\epsilon > 3$ or equivalently $c^2 > 6$. Thus here we are in the presence of *fast-roll*, and although the scalar field is quickly rolling down its potential, the universe contracts very slowly. Note also the useful relation

$$V(\phi) = -\frac{1}{\epsilon t^2}. \tag{3.8}$$

We are using a coordinate system in which the big crunch occurs at $t = 0$; in other words, the time coordinate is negative during the ekpyrotic phase. The steeply falling scalar fields act as a very stiff fluid, and, in fact, one can take the condition $w > 1$ to be the defining feature of ekpyrosis. The matter content does not necessarily have to be composed of scalar fields, but it is easy to model the ekpyrotic phase that way, and scalar fields commonly appear in effective theories arising from higher dimensions. The main consequence is that the extra term in the Friedmann equation (3.1) with $w > 1$ comes to dominate the cosmic evolution, and once more, the fractional energy densities $\Omega_\kappa \propto a^{-2}H^{-2}$ and $\Omega_\sigma \propto a^{-6}H^{-2}$ quickly decay. Thus, neglecting quantum effects, the universe is left exponentially flat and isotropic as it approaches the big crunch. The flatness problem is thus solved if the ekpyrotic phase lasts long enough.

It is interesting to calculate the length of the ekpyrotic phase, assuming the ekpyrotic potential, at its minimum, reaches the GUT scale (2 orders of magnitude below the Planck energy scale) and that the coefficient in the potential is, for example, $c = 15$. As shown earlier, aH needs to grow by at least a factor e^{60} in order to solve both the flatness and the horizon problems. Since a is roughly constant, and $H \propto 1/t$ during ekpyrosis, we have $t_{ek-beg}/t_{ek-end} = e^{60}$ at least. The scaling solution implies that $V = -\frac{2}{c^2t^2}$, and hence if we assume $|V_{min}| \sim (10^{-2}M_{Planck})^4$, we get $t_{ek-end} \sim 10^3 t_{Planck}$. This implies

$$t_{ek-beg} \gtrsim 10^{30} t_{Planck} \approx 10^{-13} s, \tag{3.9}$$

a short time in cosmology, although long by microphysical standards. In the cyclic picture of the universe, the potential V interpolates between the GUT scale and the dark energy scale. From the scaling solution (3.6), we have that $V \propto t^{-2}$, and this leads to

$$\begin{aligned} |t_{ek-beg}| &= \sqrt{\frac{V_{ek-end}}{V_{ek-beg}}}|t_{ek-end}| \approx \sqrt{10^{112}}10^3 t_{Planck} \\ &= 10^{16} s. \end{aligned} \tag{3.10}$$

In this case, the ekpyrotic phase lasts on the order of a billion years.

In the same way, anisotropies are suppressed, and chaotic mixmaster behavior does not arise. As we will see later, the inclusion of quantum effects superposes small fluctuations on this classical background, and these will address the inhomogeneity problem.

The relic problem is solved if in the transition from the contracting phase to the expanding phase the temperature never went above the GUT temperature.

3.2.2 Cyclic Scenario

The cyclic model is an ambitious attempt at providing a complete history of the universe, incorporating both the ekpyrotic mechanism and dark energy in an essential way. Here we will just present a brief overview. The cyclic model relies on having a scalar field potential of the general shape depicted in figure 5. The idea is that the ekpyrotic phase, which precedes the big crunch/big bang transition, is itself preceded by a phase where the potential is positive and flat. During this phase, the scalar field rolls down the potential very slowly, and this in fact provides a period of dark energy domination. But this could be our current universe! At some point in the future, the potential energy becomes negative, and the universe reverts from expansion to contraction. We then enter an ekpyrotic phase, which locally flattens and homogenises the universe. After a brief phase of kinetic energy domination, we reach the big crunch/big bang, and matter and radiation are produced. This "bounce" is discussed

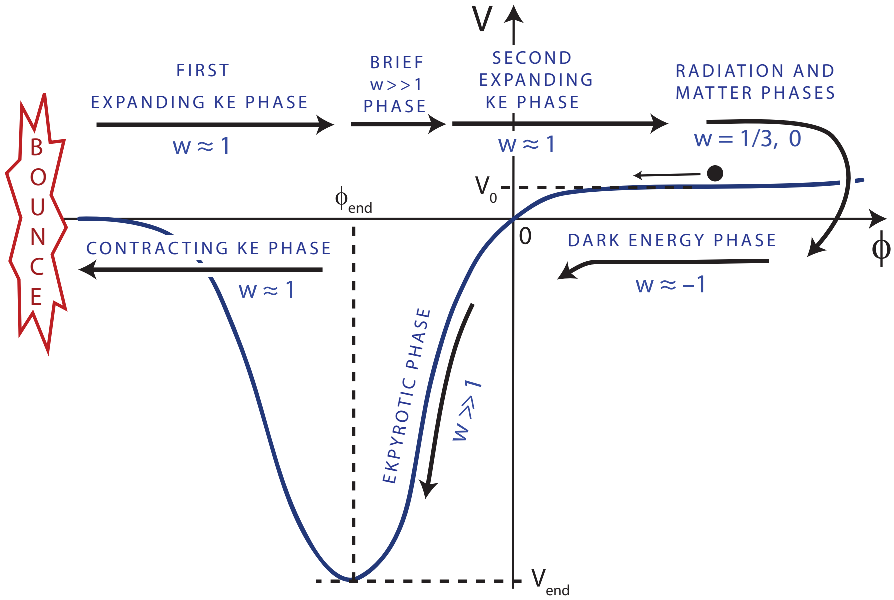

Figure 5: The potential for the cyclic universe.

further in Lecture IV. After the bang, the scalar field acquires a small boost and quickly rolls back across the potential well and onto the positive energy plateau, where the scalar gets quasi-stabilised. Now the universe undergoes the usual periods of radiation and matter domination, while the scalar field starts rolling slowly back. Eventually, the resulting dark energy comes to dominate the energy density of the universe, and the whole cycle starts again. In this way, physical processes in the universe today provide the initial conditions for the next cycle. What should be noted is the economy of ingredients. Also, there exists a nice interpretation of this evolution in terms of branes and higher dimensions.

An important aspect should be noted: even though the sequence of cosmological phases is cyclic, the scale factor for example does not evolve cyclically: the universe expands by large amounts during radiation, matter and dark energy domination, while it only contracts a little during the ekpyrotic phase. Thus, over the course of each cycle, the universe grows by a huge factor. This is a beneficial feature, as it prevents the build-up of entropy and a possible conflict with the 2nd law of thermodynamics.

A second aspect to note is that in this scenario, dark energy plays a crucial role in stabilising the cycles, and in contributing to the flattening of the universe. Thus, contrary to hot big bang and inflationary cosmology, dark energy actually has a “raison d’être”.

3.2.3 Perturbations

Single Field

In order to calculate the fluctuations in this model, we can take over the formalism developed earlier in the context of inflation, as here also we have gravity coupled to a scalar field with a potential. There, we saw that the mode functions of the comoving curvature perturbation $\mathcal{R}$ obey the equation of motion

$$v'' + \left(k^2 - \frac{z''}{z}\right) v = 0 \tag{3.11}$$

where $v = z\mathcal{R}$ with $z = a\sqrt{2\epsilon}$. For an exponential potential, ϵ is constant, and thus $z''/z = a''/a$. Using $a \propto (-\tau)^{1/(\epsilon-1)}$, we find

$$\frac{a''}{a} = -\frac{\epsilon - 2}{(\epsilon-1)^2\tau^2} \rightarrow \nu = \frac{\epsilon - 3}{2(\epsilon - 1)} \tag{3.12}$$

Given that $\epsilon > 3$, we have $0 < \nu < 1/2$. If we now inspect the late time behaviour of the mode functions,

$$f \sim \frac{(-\tau)^{\frac{1}{2}-\nu}}{k^\nu}, \quad (|k\tau| \ll 1) \tag{3.13}$$

we can see that these modes are not amplified. Thus, as the model stands, it renders the universe completely flat, both classically and quantum-mechanically. This is great news regarding the flatness problem, but it means that so far we have not explained the origin of the small temperature fluctuations in the CMB. It turns out that this can be done by adding one more scalar field, as we will now show.

Two Fields - Entropic Mechanism

Motivated by the previous discussion, we will add a second scalar field χ to the ekpyrotic model, by considering the Lagrangian (in natural units $8\pi G = M_{Pl}^{-2} = 1$)

$$\mathcal{L} = \sqrt{-g}\left[\frac{R}{2} - \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}e^{-b\phi}\partial_\mu\chi\partial^\mu\chi + V_0 e^{-c\phi}\right], \tag{3.14}$$

where we assume b, c to be constants. The reason for considering a coupling term $e^{-b\phi}$ in the kinetic term for χ will become clear below. Note that we have left the potential unchanged, and in particular it does not depend on the second scalar. In a flat Friedmann–Lemaitre-Robertson-Walker (FLRW) universe, the equations of motion are given by

$$\ddot{\phi} + 3H\dot{\phi} + cV_0 e^{-c\phi} = -\frac{1}{2}be^{-b\phi}\dot{\chi}^2, \tag{3.15}$$

$$\ddot{\chi} + \left(3H - b\dot{\phi}\right)\dot{\chi} + e^{b\phi}V_{,\chi} = 0, \tag{3.16}$$

$$H^2 = \frac{1}{6}\left(\dot{\phi}^2 + e^{-b\phi}\dot{\chi}^2 - 2V_0 e^{-c\phi}\right). \tag{3.17}$$

Since the potential $V(\phi)$ does not depend on χ, it is clear that $\chi =$ *constant* is a solution. The remaining equations then reduce to those for a single scalar in an ekpyrotic potential, and they admit the scaling solution(3.6).

Having discussed the background, we now turn our attention to the fluctuations. The main interest of course lies in the fluctuations of χ. Given that the background trajectory is along a $\chi =$ *constant* line in field space, it is clear that fluctuations in χ are transverse to the background evolution. Thus, in contrast to fluctuations of ϕ, which change the expansion history locally and correspond to curvature fluctuations, the $\delta\chi$ fluctuations are isocurvature (or entropy) perturbations. We will come back to this point later on. Note that $\delta\chi$ is immediately gauge-invarient, as $\Delta\delta\chi = -\dot{\chi}\xi^0 = 0$ (cf. Eq. (2.8)). A useful definition for the entropy perturbation is given by the gauge-invariant quantity $\delta s = e^{-\frac{b}{2}\phi}\delta\chi$, whose linearized equation of motion in Fourier space is given by

$$\ddot{\delta s} + 3H\dot{\delta s} + \left[\frac{k^2}{a^2} - \frac{b^2}{4}\dot{\phi}^2 - \frac{b}{2}V_{,\phi}\right]\delta s = 0, \tag{3.18}$$

where k denotes the wavenumber of the fluctuation mode. Switching to conformal time and defining the canonically normalized entropy perturbation $v_s \equiv a\delta s$, we obtain

$$v_s'' + \left[k^2 - \frac{a''}{a} - \frac{b^2}{4}\phi'^2 - \frac{b}{2}a^2V_{,\phi}\right]v_s = 0. \tag{3.19}$$

Imposing the usual boundary condition that in the far past/on small scales the mode function is that of a fluctuation in Minkowski space, $\lim_{k\tau\to-\infty} v_s = \frac{1}{\sqrt{2k}}e^{-ik\tau}$, up to an irrelevant phase the solution is

$$v_s = \sqrt{\frac{\pi}{4}}\sqrt{-\tau}H_\nu^{(1)}(-k\tau), \tag{3.20}$$

where $H_\nu^{(1)}$ denotes a Hankel function of the first kind. At late times/on large scales, the entropy perturbations then scale as

$$v_s \propto k^{-\nu}(-\tau)^{1/2-\nu} \qquad (|k\tau| \ll 1). \tag{3.21}$$

Defining a parameter $\Delta \equiv \frac{b}{c} - 1$, the spectral index comes out as

$$n_s = 4 - 2\nu = 1 - 2\Delta\frac{\epsilon}{(\epsilon - 1)}, \tag{3.22}$$

where we did not have to make any approximations. When the two exponents b and c in the original Lagrangian (3.14) are equal, we obtain an exactly scale-invariant spectrum, $n_s = 1$. However, when b and c differ sightly, we obtain deviations from scale-invariance. Since we have $\epsilon > 3$, the deviation from scale-invariance is always between -3Δ and -2Δ. Thus, if b is larger than c by about two percent, we obtain the central value $n_s = 0.96$ reported by the PLANCK team. Note that these modes also satisfy the WKB classicality conditions increasingly well, in direct analogy with the inflationary calculation performed earlier.

Using the large-scale expression for the mode functions (3.21), with ν given in (3.22), we can also find the time dependence of the original scalar field fluctuation $\delta\chi$:

$$\delta\chi = e^{\frac{b}{2}\phi}\frac{v_s}{a} \propto (-a\tau)^{\frac{b}{c}}(-\tau)^{\frac{1}{2}-\nu}\frac{1}{a} = constant. \tag{3.23}$$

Thus $\delta\chi$ tends to a constant on large scales, irrespective of the values of b and c, implying that this solution is stable for any values of the spectrum.

What we observe in the cosmic background radiation are not entropy perturbations, but rather the observed temperature fluctuations stem directly from curvature perturbations. Thus, for the model to be viable, we must ensure that the entropy perturbations can get converted into curvature fluctuations. One possibility is that after the ekpyrotic phase, there could be a turn in the background field space trajectory. Indeed, when two fields are present, the evolution equation of the curvature perturbation on large scales becomes more involved,

$$\dot{\mathcal{R}} = \frac{2H}{\dot{\phi}}\dot{\theta}\delta s = \sqrt{\frac{2}{\epsilon_c}}\dot{\theta}\delta s, \tag{3.24}$$

where ϵ_c denotes the value of ϵ during the conversion process and $\dot{\theta} = V_{,s}/\dot{\phi}$.. This equation illustrates that whenever the background trajectory bends, curvature perturbations are generated. Since there is no k-dependence in Eq. (3.24), their spectrum will be identical to that of the entropy perturbations that source them, and thus will be given by Eq. (3.22). A bending could either occur at the end of the ekpyrotic phase, or during the subsequent kinetic phase where the ekpyrotic potential has become unimportant. For these two possibilities, we can estimate the amplitude of the curvature perturbation by approximating ϵ_c and δs as constants over the time of the conversion, and assuming a total bending angle of about 1 radian, $\Delta\theta \approx 1$, giving $\mathcal{R}_{final} \approx \frac{1}{\sqrt{\epsilon_c}}\delta s_{ek-end}$ and leading to a variance

$$\Delta_s^2 = \frac{k^3}{(2\pi)^2}\langle\mathcal{R}_{final}^2\rangle \approx \frac{(\epsilon-1)^2}{\epsilon_c(\epsilon-3)}\frac{V_{ek-end}}{(2\pi)^2}, \tag{3.25}$$

where V_{ek-end} corresponds to the energy scale of the deepest point in the potential. Unless the fast-roll parameter ϵ during the ekpyrotic phase is very close to 3, this implies that the potential has to reach approximately the grand unified scale $V_{ek-end} \approx (10^{-2}M_{Pl})^4$ in order for the curvature perturbations to have an amplitude in agreement with the observed value of about 2×10^{-9}.

Without having the time to go into detail, we also note that in this model non-Gaussian corrections are small and in agreement with observations. (This is essentially due to the fact that in th original Lagrangian there are no terms in χ of higher order than quadratic.) Thus, this example of the "non-minimal entropic mechanism" provides an alternative model in agreement with observations. In saying this, we have assumed that the perturbations can get through the bounce phase unchanged. Bounce models in which this is the case can be constructed, and we will discuss this issue in a bit more detail in lecture IV.

Gravitational Waves

In order to calculate the tensor fluctuations in this model, we can again take over the formalism developed earlier in the context of inflation. The tensor mode functions obey the equation of motion

$$v'' + \left(k^2 - \frac{a''}{a}\right) v = 0 \tag{3.26}$$

where v corresponds to a polarisation mode of the off-diagonal spatial metric perturbations. In analogy with the single scalar field calculation above, we have that

$$\frac{a''}{a} = -\frac{\epsilon - 2}{(\epsilon - 1)^2 \tau^2} \rightarrow \nu = \frac{\epsilon - 3}{2(\epsilon - 1)} \tag{3.27}$$

Given that $\epsilon > 3$, we have $0 < \nu < 1/2$. Hence, inspecting once more the late time behaviour of the mode functions,

$$f \sim \frac{(-\tau)^{\frac{1}{2}-\nu}}{k^\nu}, \quad (|k\tau| \ll 1) \tag{3.28}$$

we can see that these modes are not amplified. Thus, during an ekpyrotic phase, tensor modes are not produced and gravitational waves are not generated. This is in marked contrast with inflation, and serves as a possible observational characteristic to distinguish between the two kinds of models. However, two things complicate this simple diagnostic:

- Not all inflationary models produce gravity waves at a level that is detectable in the foreseeable future. This depends entirely on the energy scale during inflation.
- In ekpyrotic and cyclic models, gravitational waves might get produced at a different stage, in particular during the bounce phase. This possibility has not been explored much yet, and remains incompletely understood.

3.2.4 Puzzles

Ekpyrotic and cyclic models present similar open problems/challenges than inflationary ones, with a few interesting differences.

- What is the nature of the ekpyrotic scalar field?
- How does it couple to all other particles, in particular to the standard model particles? How does reheating work in cyclic models?
- How does the cyclic potential arise? How can it be so steep over such an extended field range?
- Can cycles last indefinitely, or is there a limit to how many cycles are viable?
- How did the bounce occur? And can we follow cosmological perturbations unambiguously from the contracting into the expanding phase? Are classically singular bounces resolved by quantum gravity? In the meantime, can we construct sensible non-singular bounces? Some of these aspects will be discussed a little further below.
- Evidently, there is less of an initial conditions problem, as each cycle sets up the initial conditions for the next one in a dynamical fashion. However, one may ask how the cycling started? How come a classical spacetime came into existence in the first place?

One point should be noted, which is in contrast with inflation: during the ekpyrotic phase, there is no runaway as in eternal inflation. Large quantum fluctuations during ekpyrosis simply lead to a local time-delay in the cycle, but they do not get amplified. Thus they do not disrupt the cyclic evolution in any dangerous way, and in particular they do not lead to an infinity of different universes.

3.3 Modelling Bounces

In classical GR, the big bang corresponds to a singularity, where certain curvature invariants become infinite. A description beyond such a singularity is then impossible. Likewise, there exist models in string theory, e.g. models where two branes collide, in which the big bang corresponds to a classically singular moment. These models are interesting, amongst others because they provide support for a number of assumptions that we have made earlier: in these models, shortly before the brane collision, a turn in the scalar field trajectory automatically happens (so that entropy perturbations get converted into curvature perturbations). Moreover, at the brane collision, some of the brane kinetic energy can be converted into particles and radiation, thus heating the universe. In these models, the two branes are of opposite tension. If more matter/radiation gets produced on the negative-tension brane than on the positive-tension one, the effective 4d scalar field also gets a little extra kinetic energy, as is necessary for a cyclic implementation. For a comprehensive overview, see section 6 of my review arXiv:0806.1245.The draw-back of such models is that the brane collision is classically singular. One may expect that quantum gravitational effects resolve such singularities, i.e. that in quantum gravity the quantum effects near such singularities play an important role and avoid the development of infinities. However, currently no (fully convincing) quantum gravitational resolutions of cosmological singularities are known.

Because of this, it is of interest to try to find descriptions of classically non-singular bounces, where the universe contracts to a minimum scale factor and then smoothly re-expands. Such non-singular descriptions may serve two purposes: the first is that it would be interesting to know whether non-singular bouncing universes are allowed in fundamental theory (the answer to this question is currently unknown), and even if not, classically non-singular descriptions may provide some approximate description of the physics of quantum-resolved bounces. It is not easy to find classically non-singular bounce solutions, because of the following fact: in a flat FLRW universe, the acceleration equation can be written as

$$\dot{H} = -\frac{1}{2}(\rho + p). \tag{3.29}$$

Thus, if we want to go from contraction to expansion, i.e. if we want $\dot{H} > 0$, we require that the sum of energy density and pressure be negative,

$$\rho + p < 0 \quad \text{at} \quad \text{a} \quad \text{bounce} \tag{3.30}$$

This condition corresponds to a violation of the null energy condition. Typically, such a violation is associated with ghost instabilities, which are excitations with negative kinetic energy, and other pathologies. Note that negative energy fluctuations are catastrophic as ever wilder fluctuations decrease the energy more and more, and are thus favourable. Thus, the system immediately becomes catastrophically unstable. From a quantum-mechanical point of view, the situation is even worse, as a theory with ghosts does not even have a vacuum state. However, in recent years models have been found that can evade many such potential pathologies, and better models are developed all the time.

Here we will describe one simple model of a classically non-singular bounce, namely a *ghost condensate* bounce. The model is not perfect, but it is easy to describe and it highlights the main advantages and issues arising in the description of non-singular bounces. To this end, consider the Lagrangian

$$\mathcal{L} = \sqrt{-g}\left[\frac{R}{2} + K(\phi)X + T(\phi)X^2 - V(\phi)\right] \tag{3.31}$$

where we have defined $X \equiv -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$ as shorthand for the ordinary kinetic term. The term proportional to $T(\phi)$ is the leading higher-derivative correction to the kinetic term, and it will play an important role here. During an ekpyrotic contracting phase, we have $K(\phi) = 1, T(\phi) = 0, V(\phi) = -V_0 e^{-c\phi}$. If this were all there was, then the universe would eventually reach a big crunch singularity. However, here we imagine that near $\phi = 0$, the sign of K switches to negative. This would ordinarily lead to a ghost, but here we imagine that simultaneously the higher-derivative term TX^2 turns on and reaches a value T_b at the bounce. An example is provided by the following choice

$$K = 1 - \frac{2}{(1+\phi^2)}, \quad T = \frac{T_b}{(1+\phi^2)}\,. \tag{3.32}$$

In this theory the energy density and pressure are given by

$$\rho = \frac{1}{2}K\dot{\phi}^2 + \frac{3}{4}T\dot{\phi}^4 + V, \tag{3.33}$$

$$p = \frac{1}{2}K\dot{\phi}^2 + \frac{1}{4}T\dot{\phi}^4 - V, \tag{3.34}$$

so that the sum of energy density and pressure is $\rho + p = K\dot{\phi}^2 + T\dot{\phi}^4$. This sum can be negative at the bounce, where $K = -1, T = T_b$, as long as

$$\rho_b + p_b < 0 \quad \rightarrow \quad \dot{\phi}_b^2 < \frac{1}{T_b} \quad \text{condition for bounce} \tag{3.35}$$

Thus, if the kinetic energy is not too large, a bounce can occur. We can take a look at the Friedmann equation at the moment of the bounce to see whether it is possible to achieve this (note that by definition the bounce occurs at $H = 0$),

$$3H^2 = 0 = -\frac{1}{2}\dot{\phi}_b^2 + \frac{3}{4}T_b\dot{\phi}_b^4 + V \tag{3.36}$$

The solution to this equation is given by

$$\dot{\phi}_b^2 = \frac{1}{3T_b}(1 \pm \sqrt{1 - 12T_bV}) \tag{3.37}$$

Here the plus sign corresponds to the ghost condensate branch that we are interested in. We see that solutions with $\dot{\phi}_b^2 < 1/T_b$ are easily possible, as long as $-1/(4T_b) < V < 1/(12T_b)$. Given the general arguments mentioned above, it is important to check the stability of the model. We are principally interested in the stability of the scalar field, since it has a wrong-sign kinetic term at the bounce. For this calculation, we may momentarily neglect the coupling to gravity and the associated issues of gauge-dependence, and just look at the (derivative) fluctuations of the scalar, which we write as

$$\phi = \bar{\phi}(t) + \delta\phi(t, x^i) \tag{3.38}$$

Here $\bar{\phi}$ denotes the background value of the scalar field. Collecting quadratic time and space derivative fluctuation terms, we obtain at the moment of the bounce (which is the most dangerous time)

$$\delta\mathcal{L} \supset a^3(\dot{\delta\phi})^2\left[-\frac{1}{2}+\frac{3}{2}T_b\dot{\bar{\phi}}_b^2\right] \tag{3.39}$$

$$-a^3\delta\phi^{,i}\delta\phi_{,i}\left[-\frac{1}{2}+\frac{1}{2}T_b\dot{\bar{\phi}}_b^2\right] \tag{3.40}$$

Absence of ghosts and of gradient instabilities means that the two respective terms in square brackets should be positive. (If these terms are positive, solutions to the associated equation of motion for $\delta\phi$ are oscillatory in nature - however, if either of these terms is negative, the solutions become exponentials, and thus will contain an unstable growing mode.) Absence of ghosts requires

$$\dot{\bar{\phi}}_b^2 > \frac{1}{3T_b} \quad \text{no} \quad \text{ghosts}, \tag{3.41}$$

while absence pif gradient instabilities requires

$$\dot{\bar{\phi}}_b^2 > \frac{1}{T_b} \quad \text{no} \quad \text{gradient} \quad \text{instabilities}, \tag{3.42}$$

These conditions are not quite compatible with the bounce condition. The best we can do is to have a potential that is close to, but slightly above, $-1/4T_b$ at the bounce. Then $\dot{\bar{\phi}}_b^2$ is slightly less than $1/T_b$. In that case, a bounce is obtained and ghosts are avoided. We can then also avoid gradient instabilities if we include additional higher-derivative terms in the theory. Normally, such terms would not play a role, but here, because the coefficient of $\delta\phi^{,i}\delta\phi_{,i}$ is so small, additional terms can contribute – e.g. a term $-(\Box\phi)^2$ would then stabilise the bounce for high momentum fluctuations. Note that there is one problem with this bounce model: at the moment of the bounce, the higher-derivative X^2 term is just as large as the X term - thus one must assume that all higher-derivative terms of the form X^n with $n > 2$ appear with small coefficients, which may not be easy to achieve. There exist models (e.g. models based on Galileon terms) that are better-behaved in this regard, though they are also more complicated to analyse.

An important question in bouncing models is what happens to the cosmological perturbations as they pass through the bounce. Naively, there are two opposing viewpoints: on the one hand, the bounce occurs on small scales so that one may guess that the long-wavelength modes of cosmological interest are unaffected by the bounce. On the other hand, at the bounce the Hubble rate passes through zero, so that the horizon becomes infinite. Thus all modes re-enter the Hubble radius and conservation is no longer guaranteed. We can analyse this situation by looking in more detail at the action and evolution equation for the co-moving curvature perturbation.

For an action of the form

$$S = \int d^4x\sqrt{-g}\left[\frac{R}{2} + P(X,\phi)\right] \tag{3.43}$$

where P is an arbitrary function of the scalar field and its kinetic term X, the co-moving curvature perturbation $\mathcal{R}$ has the quadratic action (in conformal time)

$$S_{(2)} = \int d\tau d^3x z^2 \left[\mathcal{R}'^2 - c_s^2 k^2 \mathcal{R}^2\right] \,, \tag{3.44}$$

with

$$z^2 = \frac{a^2\phi'^2}{2(H)^2}(P_{,X} + 2XP_{,XX}), \qquad c_s^2 = \frac{P_{,X}}{P_{,X} + 2XP_{,XX}} \,. \tag{3.45}$$

This leads to the equation of motion

$$\mathcal{R}'' + 2\frac{z'}{z}\mathcal{R} + c_s^2 k^2 \mathcal{R} = 0 \,. \tag{3.46}$$

In our case we have $P = KX + TX^2 - V$, and it is straightforward to see that near the bounce one obtains

$$c_s^2 \mid_{bounce} \approx \text{constant} \qquad z^2 \mid_{bounce} \approx \frac{\text{constant}}{\mathcal{H}^2} \,. \tag{3.47}$$

Near the bounce, the Hubble rate goes from negative to positive, and a good approximation close to the bounce is simply to take $\mathcal{H} = \tau$, so that $\tau = 0$ marks the time of the bounce. Note that then

$$\frac{z'}{z} \approx -\frac{1}{\tau} \tag{3.48}$$

and thus the equation for $\mathcal{R}$ is singular at the bounce. Nevertheless, we can go ahead and try to solve it, by using an Ansatz $\mathcal{R} = \tau^\alpha + c\tau^\beta + \cdots$. This leads to the solution

$$\mathcal{R} = c_1\left(1 + \frac{1}{2}c_s^2k^2\tau^2 + \cdots\right) + c_2\left(\tau^3 - \frac{1}{10}c_s^2k^2\tau^5 + \cdots\right) \,. \tag{3.49}$$

Thus, even though the equation is singular, the solution is not! As we can see, all perturbation modes become near-constant near the bounce, and longer wavelength modes (small k) evolve less than short modes. In fact, long-wavelength modes are conserved through the bounce with high precision, as a full numerical analysis shows. (The reason why the perturbation is singular is that co-moving gauge becomes ill-defines at the bounce. However, there exist other gauges, for instance harmonic gauge, that remain entirely non-singular throughout, and it is safest to perform the calculation in that gauge. Such calculations confirm the results of the present simplified analysis.)

Thus, for these bounce models, one can obtain a reliable history for the cosmological perturbations of interest, from their generation during a contacting phase until the present.