

Contents

7 Inflation: background	113
7.1 Motivation	113
7.1.1 Homogeneity and isotropy problem, or the horizon problem	113
7.1.2 The flatness problem	115
7.1.3 The relic problem	115
7.1.4 What is needed	116
7.2 Inflation introduced	116
7.3 The inflaton field	121
7.4 Slow-roll inflation	122
7.5 Reheating	125
7.5.1 Scalar field oscillations	125
7.5.2 Inflaton decay	126
7.5.3 Thermalisation	127
7.6 Scales of inflation	127
7.6.1 E-folds of inflation	127
7.6.2 Evolution of scales	127
7.7 Before inflation	131

Preface

These are lecture notes for the course Cosmology II at the University of Helsinki. They owe a heavy debt to lecture notes prepared by Hannu Kurki-Suonio. Credit for some figures belongs to Reijo Keskitalo. These lecture notes continue the lecture notes of the course Cosmology I, which is assumed as a prerequisite. These notes refer to equations in those notes and use the same notation.

7 Inflation: background

7.1 Motivation

Inflation is a scenario in which there was a period of accelerated expansion in the very early universe. Like dark matter, inflation has not been established beyond reasonable doubt. However, inflationary models have explained observations well and made several detailed predictions that have been observationally verified, and no competing scenario has had the same success.

The word “inflation” also refers to the period of accelerated expansion, as well as to the accelerated expansion itself. For example, we say that the “the universe inflates” to mean that the expansion accelerates. The motivation for inflation came from some unresolved issues of the “big bang model”, i.e. the homogeneous and isotropic FLRW model with matter consisting of a gas of particles, which was considered in Cosmology I.

7.1.1 Homogeneity and isotropy problem, or the horizon problem

One question concerns the origin of the symmetry of the FLRW model. There is no unique (or even natural and generally accepted) way to define a measure on the set

of spacetimes, but the homogeneous and isotropic universes seem very special. Homogeneity and isotropy of the universe have two distinct aspects. First, departures from homogeneity and isotropy are small, i.e. the differences in any physical quantity calculated at two different spatial points are small. This is the case in the early universe, as we know from the fact that the amplitude of the CMB perturbations is only $\sim 10^{-5}$ (apart from the dipole component which is 10^{-3} and presumably overwhelmingly due to our motion)¹. Today, perturbations are locally large in the sense that the relative differences in the local value of the energy density, expansion rate and so on are large. The density of a typical galaxy is around 10^6 times the mean density, and the expansion rate inside is 100% smaller than the mean (as bound structures do not expand); a void can be 90% underdense, and expand tens of % faster than the mean. However –and this is the second aspect of homogeneity and isotropy– the universe is still statistically homogeneous and isotropic today, because the initial distribution of small perturbations had this symmetry. So the homogeneity and isotropy question is related to the origin of the seeds of structure. Where did the small deviations come from, and why is their distribution the observed one? One aspect of this is the *horizon problem*. The CMB anisotropies are correlated on all observed scales. However, at the time of last scattering any regions which are today separated by more than about 1° had not had time to interact assuming the FLRW metric and ordinary matter.

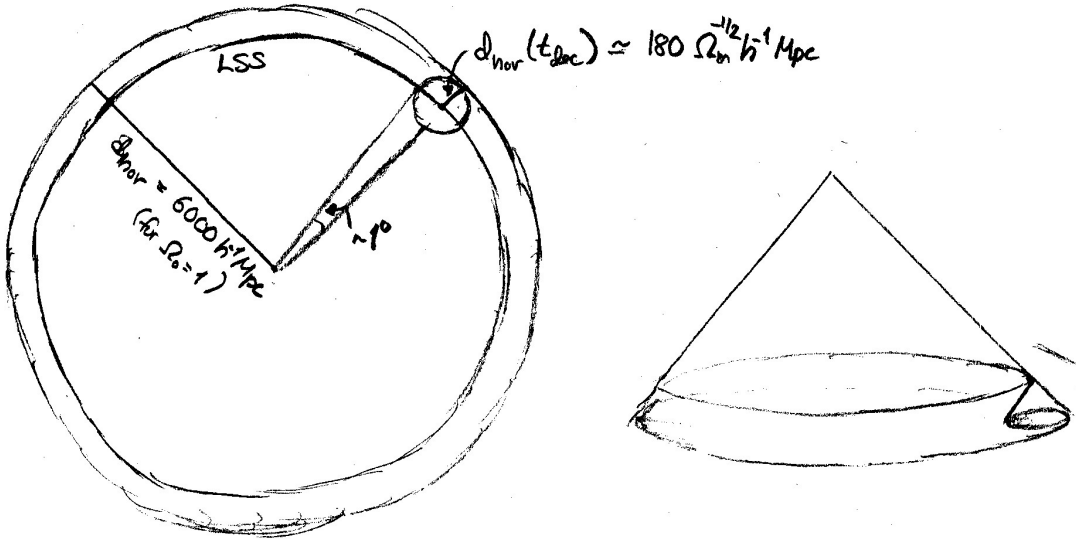


Figure 1: The horizon problem.

¹Strictly speaking, this is the amplitude of deviations in the distribution of photons. As baryonic matter was in equilibrium with the photons, it was also close to homogeneous and isotropic. However, departures from homogeneity and isotropy at last scattering were larger in the dark matter, which is decoupled from photons. This is crucial for structure formation, as we will discuss in chapter 9.

7.1.2 The flatness problem

Another issue is the spatial flatness of the universe. The density parameter is

$$\Omega - 1 = \frac{K}{(aH)^2} . \quad (7.1)$$

If $\Omega = 1$ at some time, it is always unity. If $\Omega \neq 1$ at any time, it evolves with time. Assuming that the spatial curvature is initially small, we have

$$\text{mat. dom.} \quad a \propto t^{2/3}, \quad H \propto t^{-1} \Rightarrow \frac{1}{aH} \propto t^{1/3} \quad \Rightarrow |1 - \Omega| \propto t^{2/3} \propto a \quad (7.2)$$

$$\text{rad. dom.} \quad a \propto t^{1/2}, \quad H \propto t^{-1} \Rightarrow \frac{1}{aH} \propto t^{1/2} \quad \Rightarrow |1 - \Omega| \propto t \propto a^2 . \quad (7.3)$$

If the spatial curvature is positive, it will dominate over matter or radiation, and the expansion will stop and turn around, and the universe will collapse. If the spatial curvature is negative, the universe will become empty and cold. The flatness problem is thus also called the oldness problem or the age problem. In the Λ CDM model, the curvature today is very small, $\Omega_0 - 1 = -0.001 \pm 0.002$ [1]. Therefore, at BBN (around $z = 10^{10}$) we have $|\Omega(t_{\text{BBN}}) - 1| \lesssim 10^{-20}$, which seems like a strong tuning. Of course, if the spatial curvature is initially zero, it will remain zero. However, this is a special value, for which we would like to have an explanation.

7.1.3 The relic problem

For matter consisting of a gas of particles, at early times the temperature is very high. In grand unified theories of particle physics there are phase transitions at high temperatures (in the Standard Model, there are no such phase transitions, only the QCD and electroweak crossovers). These phase transitions can generate topological defects such as magnetic monopoles, cosmic strings and domain walls, which correspond to (approximately) zero-, one-, and two-dimensional topological defects, respectively. Just like (in a specific model) we can calculate the relic density of dark matter particles, we can calculate the density of these relics. In some models the energy density of monopoles today would be much higher than the observed energy density. This is related to the fact that monopoles are typically very massive, with masses of the order of the grand unified scale. The presence of cosmic strings and domain walls would also be problematic, as they are also typically very heavy, and their energy density relative to ordinary matter increases with time (i.e. it goes down more slowly). In supersymmetric models one particular problem is the overproduction of gravitinos, the supersymmetric partners of the graviton. If gravitinos are not stable, their lifetime is very long, because they interact only gravitationally, so they typically decay after BBN and ruin its observational success. It is also possible that the gravitinos are stable and form the dark matter. The constraint on the temperature of the universe from gravitinos is of the order $T \lesssim 10^7$ GeV. However, it may be that grand unified theories and supersymmetric models do not describe reality (we have no observational evidence for them), in which case there is no such problem. Observationally, from BBN, we know only that the universe was at thermal equilibrium when the energy density was about $(1 \text{ MeV})^4$, we have no observations of equilibrium for earlier times.

7.1.4 What is needed

The above problems are solved if we have a mechanism that produces an initial condition for the universe at $T > 1$ MeV where the universe is homogeneous and isotropic up to small perturbations that are correlated on all observable scales, where the spatial curvature is very small and matter is in thermal equilibrium (or at least the part that consists of baryons, photons and neutrinos). In specific theories of particle physics, there may be an upper limit on the temperature.

7.2 Inflation introduced

Inflation is defined as accelerating expansion:

$$\text{inflation} \Leftrightarrow \ddot{a} > 0 . \quad (7.4)$$

We can write this equivalently as

$$\text{inflation} \Leftrightarrow -\frac{\dot{H}}{H^2} < 1 . \quad (7.5)$$

Often the term inflation is used to refer only to a period of acceleration expansion in the early universe, and not to the recent phase of accelerated expansion.

The origin of the flatness problem is that $|\Omega - 1| = |K|/(aH)^2 = |K|\dot{a}^{-2}$ grows with time because $\dot{a}$ falls, i.e. the universe decelerates. If the expansion instead accelerates, Ω is driven towards unity starting from any value. (This is the case for an expanding universe. If the universe contracts, the behaviour is reversed.)

Consider now the horizon problem. The problem is that in the FLRW model with matter and radiation the horizon at the time of photon decoupling is small compared to the part of the universe we can see today. In standard picture the universe is first radiation-dominated and then becomes matter-dominated somewhat before photon decoupling. (Recall that $t_{\text{eq}} \approx 50\,000$ years and $t_{\text{dec}} \approx 380\,000$ years.) In the radiation-dominated era, the horizon is $d_{\text{hor}}(t) = 2t = H^{-1}$. In the matter-dominated era, we have $d_{\text{hor}}(t) = 3t = 2H^{-1}$. The horizon at decoupling is between these values, $H^{-1} < d_{\text{hor}}(t_{\text{dec}}) < 2H^{-1}$. The size of the observable universe today is of the order of the present Hubble length $d_{\text{hor}}(t_0) \sim H_0^{-1}$ (the precise prefactor depends on the vacuum energy density). The presently observable universe was a factor of a_{dec}/a_0 smaller at decoupling.

In order to compare size of the horizons at different times we can use comoving lengths, where this change is taken into account. The comoving horizon at decoupling is

$$d_{\text{hor}}^c(t_{\text{dec}}) = (1 + z_{\text{dec}})d_{\text{hor}}(t_{\text{dec}}) \sim (1 + z_{\text{dec}})H_{\text{dec}}^{-1} = (a_{\text{dec}}H_{\text{dec}})^{-1} , \quad (7.6)$$

and $d_{\text{hor}}(t_0) = d_{\text{hor}}^c(t_0) \sim H_0^{-1}$. The horizon problem arises because the first horizon is much smaller than the second,

$$\frac{d_{\text{hor}}^c(t_{\text{dec}})}{d_{\text{hor}}^c(t_0)} \sim \frac{a_0 H_0}{a_{\text{dec}} H_{\text{dec}}} \sim (1 + z_{\text{dec}}) \frac{t_{\text{dec}}}{t_0} \approx 0.03 \ll 1 , \quad (7.7)$$

where we have for clarity inserted a_0 , even though it is equal to unity, and used $t_{\text{dec}} = 380\,000$ yr, $t_0 = 14 \times 10^9$ yr and $z_{\text{dec}} = 1090$. In other words, the presently

observed universe was about 30 times larger than the particle horizon at decoupling, so it contained about 10^5 regions that had never been in causal contact with each other. Thus the problem is again that aH decreases with time,

$$\frac{d}{dt}(aH) = \ddot{a} < 0 , \quad (7.8)$$

so a period with $\ddot{a} > 0$ might solve the problem.

Recall that the particle horizon refers to the maximum distance that light could in principle have travelled from the beginning of the universe until time t . If we add a new period in the early universe to the matter-dominated era and the radiation-dominated era, such as accelerating expansion, the calculation of d_{hor}^c will depend on it. We always have $d_{\text{hor}}^c(t_0) > d_{\text{hor}}^c(t_{\text{dec}})$ since $t_0 > t_{\text{dec}}$, and the interval $(0, t_{\text{dec}})$ is included in the interval $(0, t_0)$. However, in the horizon problem, the relevant present-day quantity is not actually the distance from which we could in principle have received signals, but the distance from which we actually measure signals. Because the universe is opaque before decoupling, the size of the presently observable² universe is given by the distance photons have travelled in the interval (t_{dec}, t_0) , and this is not affected by what happens before t_{dec} . Thus the relevant present-day scale is always $\sim H_0^{-1}$.

The comoving Hubble parameter is equal to the conformal Hubble parameter,

$$aH = \frac{1}{a} \frac{da}{d\eta} = \dot{a} , \quad (7.9)$$

where η is conformal time. The Hubble length is $l_H \equiv H^{-1}$, and the comoving Hubble length is

$$l_H^c \equiv \frac{1}{a} l_H = \frac{1}{aH} = \frac{1}{\dot{a}} . \quad (7.10)$$

If aH decreases, then $(aH)^{-1}$ increases, and vice versa. So we can say that inflation is any epoch when the comoving Hubble length shrinks:

$$\text{inflation} \Leftrightarrow \frac{d}{dt} \left(\frac{1}{aH} \right) < 0 . \quad (7.11)$$

It has unfortunately become customary in cosmology to use the word “horizon” also for the Hubble distance, particularly with regard to inflation. We will try to be careful not to confuse the two concepts.

Let us consider a case of accelerated expansion that we are already familiar with: exponential expansion $a(t) \propto e^{Ht}$, corresponding to the vacuum energy equation of state $w = -1$, with constant H . We will shortly see that this is a first approximation to the expansion law during inflation. The physical horizon distance is

$$d_{\text{hor}}(t) = a(t) \int_0^t \frac{dt'}{a(t')} = H^{-1} e^{Ht} (1 - e^{-Ht}) \simeq H^{-1} e^{Ht} , \quad (7.12)$$

²That is, observable with electromagnetic signals. In principle, we can receive neutrinos from times down to 1 s and gravitational waves all the way to the beginning (or up to the possible quantum gravity era), because the universe is transparent to them. No early universe neutrinos have been directly observed, and there are no observations at all of primordial gravitational waves yet.

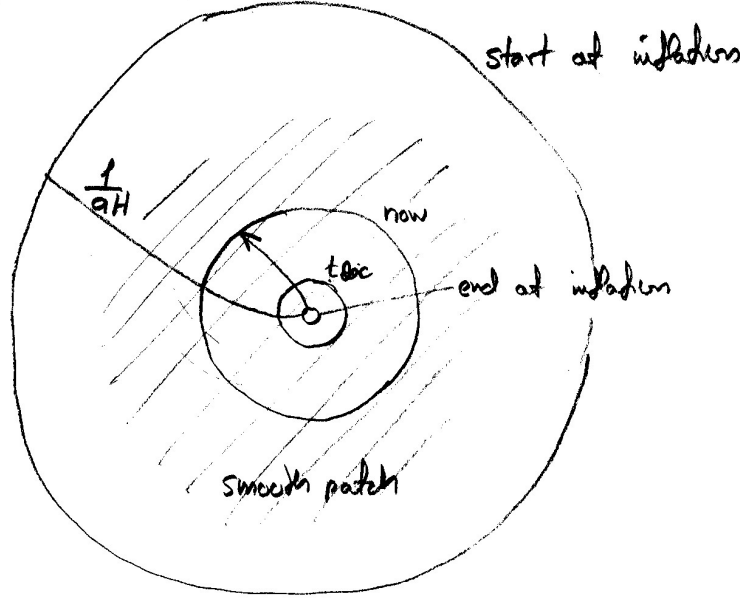


Figure 2: Evolution of the comoving Hubble radius (length, distance) during and after inflation (schematic).

where the last limit is for $t \gg H^{-1}$. So in contrast to the radiation- and matter-dominated eras, the physical particle horizon grows exponentially, and the comoving particle horizon stays almost constant, $d_{\text{hor}}^c(t) \simeq H^{-1}$. The presently observable universe has evolved from a small patch of a much larger causally connected region. See figure 2.

However, even though the particle horizon grows exponentially, the distance over which it is possible to send signals does not grow. If we consider a light ray, we have $0 = ds^2 = -dt^2 + a(t)^2 dr^2$, so the comoving coordinate separation (which is equal to the comoving proper distance, since the universe is spatially flat) between emission at t_1 and reception at t_2 is (with $a = e^{Ht}$)

$$\Delta r = \int_{t_1}^{t_2} \frac{dt'}{a(t')} = H^{-1}(e^{-Ht_1} - e^{-Ht_2}) < H^{-1}. \quad (7.13)$$

If the coordinate separation between two points is more than the Hubble length, it is not possible to send signals between them. In this sense, the Hubble length gives the comoving size of the region during inflation over which it is possible to retain causal connection. During inflation a signal sent at t_1 cannot travel a longer coordinate distance than $H^{-1}e^{-Ht_1}$, and this distance gets smaller as t_1 grows, so causal connection is lost during inflation. Note that the particle horizon, which expresses the maximum distance at which parts of the universe can have been in causal contact always grows as a function of time, never shrinks. But during inflation regions that once were in causal contact cannot send signals to each other any more.

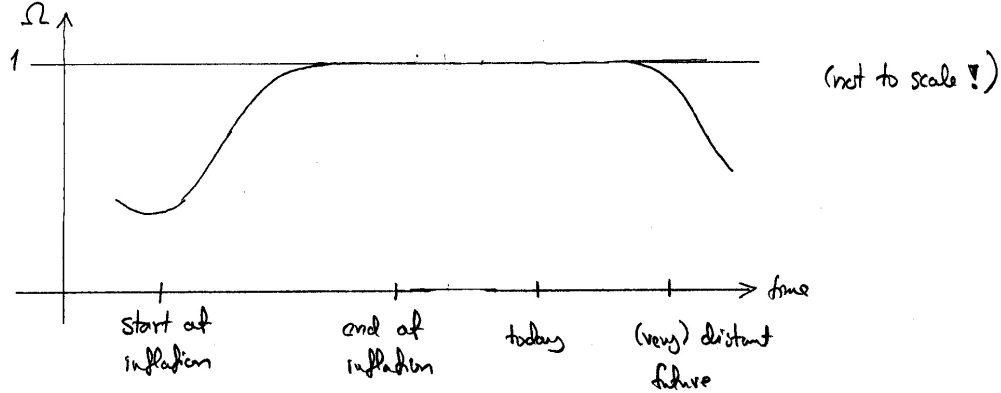


Figure 3: Solving the flatness problem. This figure is for a universe with no dark energy, where the expansion keeps decelerating after inflation ended in the early universe. Present observational evidence indicates that actually the expansion began accelerating again a few billion years ago. Thus the universe is, technically speaking, inflating again, and Ω is again driven towards 1.

The Friedmann equations are

$$3\frac{\dot{a}^2}{a^2} = 8\pi G_N \rho - 3\frac{K}{a^2} \quad (7.14)$$

$$3\frac{\ddot{a}}{a} = -4\pi G_N(\rho + 3p) . \quad (7.15)$$

Thus, in general relativity and assuming the FLRW metric, accelerating expansion requires negative pressure:

$$\text{inflation} \Leftrightarrow \rho + 3p < 0 . \quad (7.16)$$

Note that the energy density of matter for which $p/\rho < -\frac{1}{3}$ decreases in an expanding universe slower than a^{-2} , i.e. it grows relative to the spatial curvature. (If $p/\rho < -1$, the energy density actually grows as the universe expands.) The flatness problem is simply the feature that for matter composed of a gas of free particles we have $p \geq 0$, so the energy density falls at least as fast as a^{-3} , and the curvature term will at some point overtake the energy density. In inflationary models, the energy density typically remains nearly constant during a period in which the scale factor grows by a huge factor, typically by a factor e^{60} or more. Thus inflation predicts that $\Omega_0 = 1$ to extremely high accuracy.³ See figure 3.

As for the relic problem, if unwanted relics are produced before inflation, they are diluted to negligible density, just like spatial curvature. However, we have to be careful that they are not produced after inflation (if the theory of particle physics has them), i.e. the *reheating* temperature (see below) has to be small enough. This

³If it were discovered by observations that $\Omega_0 \neq 1$, this would be a blow to the credibility of inflation. However, there is a version of inflation, called *open inflation*, for which $\Omega_0 < 1$. The existence of such models of inflation have led critics of inflation to complain that inflation is unfalsifiable in the sense that no matter what the observation, a model of inflation can be found that agrees with it. Nevertheless, most models of inflation give similar generic predictions, including $\Omega_0 = 1$ to great accuracy, and thus far the observations have been in good agreement with them.

is one constraint on models of inflation. At the end of inflation, matter is produced in reheating⁴, which produces the gas of particles that eventually (through the transitions discussed in Cosmology I) becomes the matter today.

Like dark matter, inflation is more properly called a scenario rather than a theory. It is an idea of a certain behaviour of the early universe, which is realised in hundreds of different models (see e.g. [2] for a listing of more than 100 of the simplest models). Some models are closely related to otherwise motivated extensions of the Standard Model of particle physics or extensions of general relativity, while others are just toy models that have the right features and are presumably at most an approximate description of some more complicated physics. One noteworthy inflationary model is based on the Standard Model Higgs boson coupled to gravity in a non-standard way.

The important point is that inflation makes many generic predictions, i.e. predictions that are independent of the particular model of inflation, for most models. (Exceptional models can be found that would violate one or more of these general features.) There are also numerical predictions of cosmological observables that differ from one model of inflation to another, allowing observations to rule out models, as has already happened to many models. Present observational data agrees with the generic predictions (which were made before any observations of the CMB anisotropies, which are the most direct way of testing the models), while alternatives to inflation have not managed to explain the observations in an equally simple and successful way, nor make predictions that disagree with those of inflation that would have been verified by observation. Most cosmologists thus consider it likely that inflation took place in the primordial universe, but this is not beyond reasonable doubt. In the words of cosmologist Douglas Scott, “*something like inflation is something like proven*”.

In the above discussion, we have assumed that we can use the FLRW metric, i.e. that the universe is already homogeneous and isotropic. In order to explain how inflation produces homogeneity and isotropy and solves the horizon problem, we should consider how inflation gets started from some generic initial conditions. Inflation certainly makes the homogeneity and isotropy problem “exponentially smaller” in the sense that it produces a large homogeneous and isotropic causally connected patch from a small one. However, how inflation starts remains an open question. What is required is that starting from inhomogeneous initial conditions, there is at least one region where inflation starts, and its volume will then overwhelm the other regions. There are some ideas and studies of how this can happen, but as we have no solid theoretical understanding (and no observations at all) of the pre-inflationary era, the issue remains uncertain. We will comment on this a bit more after discussing simple inflationary models.

We will assume that sufficient inflation has already taken place to make the universe (within a Hubble radius) spatially flat, homogeneous and isotropic, and follow inflation in detail after that. We thus work in the flat FLRW universe. In any case, from the modern point of view, the most important (and testable) aspect of inflation is the generation of the seeds of structure, as it makes deviation from homogeneity and isotropy quantitative, as will discuss in the next chapter.

⁴“Reheating” may turn out to be as much a misnomer as “recombination”: we don’t know whether matter was ever in a thermal state before inflation.

7.3 The inflaton field

As we saw in section 7.2, inflation requires negative pressure. In chapter 4 we considered a gas of particles where interaction energies can be neglected. For such a system the pressure is always non-negative.⁵ However, the particle picture is not fundamental. In the early universe, at high energy densities, we have to consider the more fundamental entities, *fields*. Particles are just excitations of fields. The mean value of a field can have negative pressure, even if a gas consisting of the particles corresponding to the field does not. The simplest form of matter with negative pressure is a scalar field, so the simplest inflationary models involve just a single scalar field. The field responsible for inflation (and the corresponding spin 0 particle) is called the *inflaton*.

In the FLRW model, the energy density and pressure of the scalar field (as measured by an observer comoving with the matter) are

$$\rho = \frac{1}{2}\dot{\varphi}^2 + V \quad (7.17)$$

$$p = \frac{1}{2}\dot{\varphi}^2 - V, \quad (7.18)$$

The pressure is negative when the potential dominates over the kinetic term, i.e. when the field moves slowly. The equation of state parameter $w \equiv p/\rho$ is

$$w = \frac{\dot{\varphi}^2 - 2V(\varphi)}{\dot{\varphi}^2 + 2V(\varphi)} = \frac{1 - 2V/\dot{\varphi}^2}{1 + 2V/\dot{\varphi}^2}, \quad (7.19)$$

so

$$-1 \leq w \leq 1. \quad (7.20)$$

If the kinetic term $\frac{1}{2}\dot{\varphi}^2$ dominates, $w \approx 1$; if the potential term $V(\varphi)$ dominates, $w \approx -1$. Different inflaton models have different potentials $V(\varphi)$. From (7.17), we can form the useful combinations

$$\begin{aligned} \rho + p &= \dot{\varphi}^2 \\ \rho + 3p &= 2(\dot{\varphi}^2 - V). \end{aligned} \quad (7.21)$$

Inserting the energy density and pressure from (7.17) into the continuity equation

$$\dot{\rho} = -3H(\rho + p) \quad (7.22)$$

gives the equation of motion for the field:

$$\ddot{\varphi} + 3H\dot{\varphi} = -V'. \quad (7.23)$$

Note the analogy with a Newtonian particle moving in one dimension: the energy density corresponds to the total (kinetic plus potential) energy, and the equation of motion corresponds to Newton's second law. The effect of expansion is to add the *Hubble friction* term $3H\dot{\varphi}$, which slows down φ .

Let us assume that φ is initially far from the minimum of $V(\varphi)$. The potential then pulls φ towards the minimum (see figure 4). If the potential has a suitable (sufficiently flat) shape, the friction term soon makes $\dot{\varphi}$ small enough that the potential dominates, even if it did not do so initially.

⁵A gas of interacting particles could have negative pressure.

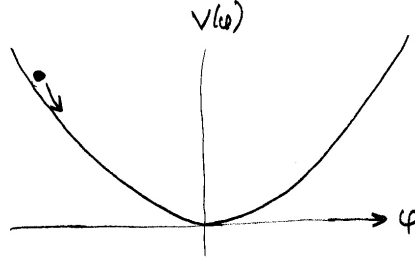


Figure 4: An example of inflaton potential.

We also need the Friedmann equation,

$$3H^2 = \frac{\rho}{M_{\text{Pl}}^2} . \quad (7.24)$$

Inserting the energy density from (7.17), we have

$$3H^2 = \frac{1}{M_{\text{Pl}}^2} \left(\frac{1}{2} \dot{\phi}^2 + V \right) . \quad (7.25)$$

The second Friedmann equation reads

$$3 \frac{\ddot{a}}{a} = \frac{1}{M_{\text{Pl}}^2} (V - \dot{\phi}^2) . \quad (7.26)$$

The condition for inflation is thus satisfied if

$$V > \dot{\phi}^2 . \quad (7.27)$$

We have ignored other contributions to the energy density and pressure besides the inflaton. During inflation, the inflaton moves slowly, so the inflaton energy density, which is dominated by $V(\phi)$, also changes slowly. If there are matter and radiation components in the energy density, they decrease fast, $\rho \propto a^{-3}$ or $\propto a^{-4}$, and soon become negligible, even faster than the spatial curvature. The presence of extra matter can put some constraints on the initial conditions for inflation to get started and the inflaton to become dominant. But once inflation begins, we can soon forget components other than the inflaton.

7.4 Slow-roll inflation

The Hubble friction term $3H\dot{\phi}$ tends to slow down the evolution of ϕ , so the system easily reaches a situation where the *slow-roll conditions* hold:

$$\dot{\phi}^2 \ll V \quad (7.28)$$

$$|\ddot{\phi}| \ll 3H|\dot{\phi}| . \quad (7.29)$$

As (7.27) shows, the universe always inflates in slow-roll. If the slow-roll conditions are valid, we may approximate (the *slow-roll approximation*) (7.23) and (7.25) by the *slow-roll equations*:

$$H^2 = \frac{V}{3M_{\text{Pl}}^2} \quad (7.30)$$

$$3H\dot{\phi} = -V' . \quad (7.31)$$

The shape of the potential $V(\varphi)$ determines the *slow-roll parameters*, defined as

$$\varepsilon(\varphi) \equiv \frac{1}{2} M_{\text{Pl}}^2 \left(\frac{V'}{V} \right)^2 \quad (7.32)$$

$$\eta(\varphi) \equiv M_{\text{Pl}}^2 \frac{V''}{V} . \quad (7.33)$$

The slow-roll conditions (7.28) and (7.29) imply $\varepsilon \ll 1$ and $|\eta| \ll 1$. (**Exercise.** Show this.) The implication goes only in this direction. The conditions $\varepsilon \ll 1$ and $|\eta| \ll 1$ are necessary but not sufficient for the slow-roll approximation (i.e. the slow-roll conditions) to be valid. They are not sufficient, because they only constrain the form of the potential in the slow-roll section, where the slow-roll approximation *may* be valid. Since the field equation (7.23) is second order, it accepts arbitrary φ and $\dot{\varphi}$ as initial conditions. Thus (7.28) and (7.29) may not hold initially, even if φ is in the slow-roll section. However, the slow-roll solution, the solution of the slow-roll equations (7.30) and (7.31), is an *attractor* of the full equations, (7.23) and (7.25). This means that the solution of the full equations rapidly approaches it, if the initial conditions that are in the *basin of attraction*. In practice, if the initial velocity is large, the field needs to be deep in the slow-roll section so that there is time for the field to slow down while remaining there.

Once the system has reached the attractor, where (7.30) and (7.31) hold, $\dot{\varphi}$ is determined by φ . In fact everything is determined by φ (assuming a fixed potential $V(\varphi)$). The value of φ is the single parameter describing the state of the universe, and φ evolves down the potential $V(\varphi)$ as specified by the slow-roll equations.

The ideas “attractor” and “basin of attraction” can be taken further. **If** the universe (or a region of it) finds itself initially (or enters) the basin of attraction of slow-roll inflation, meaning that: there is a sufficiently large region, where the curvature is sufficiently small, the inflaton makes a sufficient contribution to the total energy density, the inflaton is sufficiently homogeneous, and lies sufficiently deep in the slow-roll section, **then** this region begins inflating, it becomes rapidly very homogeneous and flat, all other contributions to the energy density besides the inflaton become negligible, and the inflaton begins to follow the slow-roll solution.

Thus inflation erases all memory of initial conditions, and we can predict the later history of the universe just from the shape of $V(\varphi)$ and the assumption that φ started out far enough in the slow-roll part of it. Note the similarity to thermal equilibrium. In the stages of the universe we discussed earlier, things were calculable because in thermal equilibrium, it is sufficient to know the temperature, masses of particles and conserved quantum numbers in order to have full information about the system. In the case of inflation, knowing the inflaton field value (and the shape of the potential) is enough, because of a rather different kind of attractor behaviour.

Example.

$$V(\varphi) = \frac{1}{2} m^2 \varphi^2 \quad \Rightarrow \quad V'(\varphi) = m^2 \varphi , \quad V''(\varphi) = m^2 \quad (7.34)$$

$$\left. \begin{aligned} \varepsilon(\varphi) &= \frac{1}{2} M_{\text{Pl}}^2 \left(\frac{2}{\varphi} \right)^2 \\ \eta(\varphi) &= M_{\text{Pl}}^2 \frac{2}{\varphi^2} \end{aligned} \right\} \Rightarrow \quad \varepsilon = \eta = 2 \left(\frac{M_{\text{Pl}}}{\varphi} \right)^2 \quad (7.35)$$

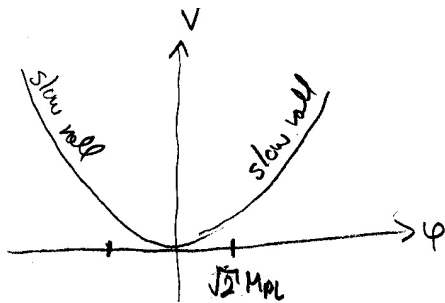


Figure 5: The potential $V(\varphi) = \frac{1}{2}m^2\varphi^2$ and its two slow-roll sections.

and

$$\varepsilon, \eta \ll 1 \quad \Rightarrow \quad \varphi^2 \gg 2M_{\text{Pl}}^2 \quad (7.36)$$

See figure 5.

We can show that in slow-roll

$$-\frac{\dot{H}}{H} = \epsilon. \quad (7.37)$$

So, recalling the definition of inflation (7.5), we see that if the slow-roll approximation is valid, the field is deep in the inflationary regime. This result also shows that during slow-roll inflation, the Hubble parameter changes slowly (while the scale factor changes almost exponentially). Slow-roll is not necessary for inflation, it is possible to have inflation even when the slow-roll parameters are not small (this is called fast-roll inflation). When we come to perturbations, we will see that slow-roll inflation (unlike fast-roll inflation) automatically produces a spectrum of scalar perturbations close to what is observed.

A scalar field model of inflation consists of the potential for the inflation and its couplings to other fields. In most models, couplings to other fields don't matter during inflation, and only the inflaton is dynamically important. Restricting to single-field slow-roll models, [2] catalogues over 100 of them. However, these couplings to other fields usually come into play when inflation ends. Inflation can end because the slow-roll approximation is no longer valid, as the field has rolled down the potential. The slow-roll approximation stops being valid when either $\varepsilon(\varphi)$ or $|\eta(\varphi)|$ becomes unity, whichever comes first. Usually inflation ends quickly after the end of slow-roll. Another possibility is that inflation ends while the inflaton undergoes slow-roll, because other fields coupled to the inflaton become dynamically important and terminate inflation. An example of this is *hybrid inflation*, where there is an extra scalar field in addition to the inflaton. The single-field models can be divided into the following two classes:

1. Small field inflation: $\Delta\varphi < M_{\text{Pl}}$ in the slow-roll section.
2. Large field inflation: $\Delta\varphi > M_{\text{Pl}}$ in the slow-roll section.

Here $\Delta\varphi$ is the change of φ during (the observationally relevant part of) inflation.

Example. Consider a simple potential of the form $V(\varphi) = A\varphi^n$. This is a large field model, since $V'/V = n/\varphi \Rightarrow \varepsilon \ll 1$ requires $\varphi^2 \gg \frac{1}{2}n^2M_{\text{Pl}}^2$.

See figure 6 for typical shapes of potentials of large field and small field models.

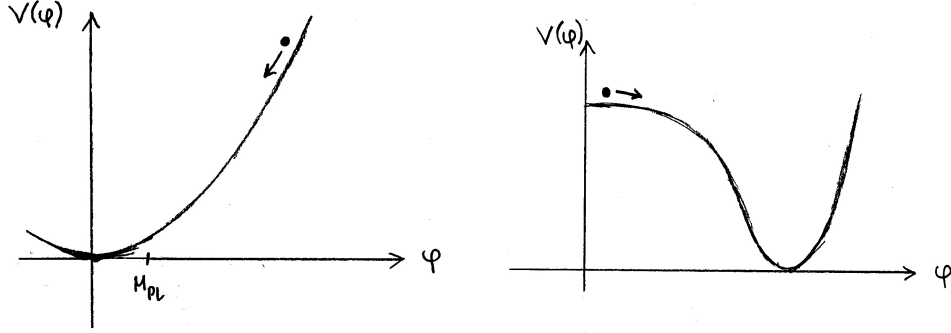


Figure 6: Potential for (a) large field and (b) small field inflation. For a typical small-field model, $\varphi \ll M_{\text{Pl}}$ in the entire range shown.

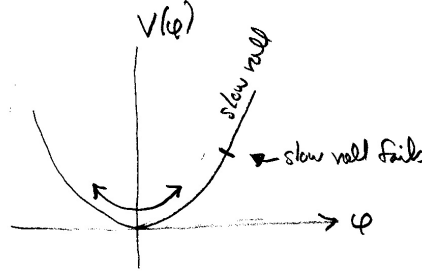


Figure 7: After inflation, the inflaton field is left oscillating at the bottom.

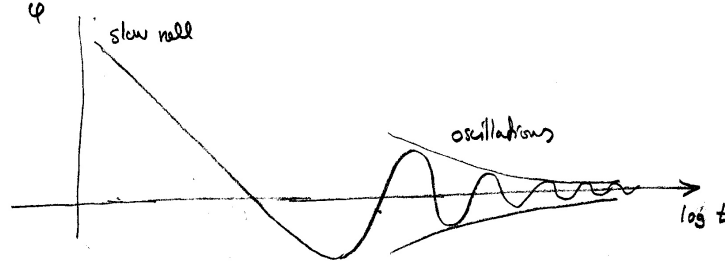
7.5 Reheating

During slow-roll inflation, practically all energy density in the universe is in the inflaton potential $V(\varphi)$. As inflation ends, this energy is transferred to the kinetic term, and in the reheating process to a bath of particles. So reheating all the stuff there is in the later universe is created from the coherent inflaton field. The conversion of the inflaton energy density into a thermal gas of particles does not affect the spectrum of density perturbations in single field models of inflation (at least on super-Hubble scales; see section 7.6 below). (It does change the relation between φ_k and k/H_0 given in (7.47), i.e. the amount that the perturbations are stretched between the end of inflation and today.) The main constraint on reheating is that the reheating temperature must be above 1 MeV, but sufficiently low so as not to produce unwanted relics, where “sufficiently” depends on the theory under consideration.

7.5.1 Scalar field oscillations

After inflation, the inflaton field φ begins to oscillate at the bottom of the potential $V(\varphi)$, see figure 7. In a simple approximation the inflaton field is still homogeneous, $\varphi(t, \vec{x}) = \varphi(t)$, so it oscillates in the same phase everywhere (the oscillation is *coherent*). The oscillation period soon becomes much shorter than the expansion time scale H^{-1} .

Assume the potential can be approximated as $V(\varphi) = \frac{1}{2}m^2\varphi^2$ near the minimum

Figure 8: The time evolution of φ as inflation ends.

of $V(\varphi)$, where the amplitude of φ is small. The equation of motion is then

$$\ddot{\varphi} + 3H\dot{\varphi} = -m^2\varphi. \quad (7.38)$$

In the limit $m \gg H$, we can neglect the friction term, and the field undergoes oscillations with frequency m . We can write the energy continuity equation as

$$\dot{\rho} + 3H\rho = -3Hp = -\frac{3}{2}H(m^2\varphi^2 - \dot{\varphi}^2). \quad (7.39)$$

The oscillating factor on the right hand side averages to zero over one oscillation period (in the limit where the period is $\ll H^{-1}$), so on average the energy density goes like $\rho \propto a^{-3}$, just like in a matter-dominated universe. Falling energy density leads to decreasing oscillation amplitude, see figure 8.

7.5.2 Inflaton decay

When the inflaton field oscillates around the minimum of the potential, the energy stored in the inflaton field is transferred into particles, both by decay into quanta of the inflaton field, which subsequently decay, and direct decay into other fields via coupling between them and the inflaton.⁶

If the decay is slow, inflaton energy density satisfies the equation

$$\dot{\rho}_\varphi + 3H\rho_\varphi = -\Gamma_\varphi\rho_\varphi, \quad (7.40)$$

where $\Gamma_\varphi = 1/\tau_\varphi$, the *decay width*, is the inverse of the inflaton decay time τ_φ , and the term $-\Gamma_\varphi\rho_\varphi$ represents energy transfer to other particles.

If the inflaton can decay into bosons, the decay may be very rapid, involving a mechanism called *parametric resonance*. The produced particles are far from thermal equilibrium (only certain bands in momentum space become populated, and their occupation numbers are huge). There are also decay mechanisms. In realistic models of inflation, the inflaton can decay via a mixture of different decay methods. The process in which the inflaton transfers its energy into particles is called *preheating* and the thermalisation of the gas of particles is called *reheating*. However, terminology varies, and often the term reheating is used to refer just to the energy transfer, even if the final state is not in thermal equilibrium.

⁶If the scale of inflation is sufficiently high, it is possible to reheat without any couplings between the inflaton and the Standard Model degrees of freedom by producing particles gravitationally out of the vacuum. This is called *gravitational reheating*, and it is one of the many delicacies of inflation we will not have time to sample.

7.5.3 Thermalisation

The particles produced from the inflaton will interact, create other particles through reactions, and the resulting soup will eventually reach thermal equilibrium with some *reheating temperature* T_{reh} . It is determined by the energy density ρ_{reh} at the end of the reheating epoch:

$$\rho_{\text{reh}} = \frac{\pi^2}{30} g_*(T_{\text{reh}}) T_{\text{reh}}^4 . \quad (7.41)$$

Necessarily $\rho_{\text{reh}} < \rho_{\text{end}}$ (end = end of inflation). If reheating takes a long time, we can have $\rho_{\text{reh}} \ll \rho_{\text{end}}$. The evolution of the gas of particles into a thermal state can be quite involved, and it has been studied in various models. Usually it is just assumed that it happens eventually, since the particles are able to interact. However, it is possible that some particles (such as gravitinos) never reach thermal equilibrium, since their interactions are too weak. In any case, as long as the momenta of the particles are much higher than their masses, the energy density of the universe behaves like radiation, regardless of the momentum space distribution. So the background expansion rate is the same. After thermalisation of at least the Standard Model particles is complete, the period sometimes called the hot big bang era (i.e. the FLRW metric plus a gas of particles in thermal equilibrium) starts.

Exercise. Assume that at the beginning of inflation we have $|\Omega_K| = 0.1$. Calculate, as a function of the reheating temperature T_{reh} , how many e-folds of inflation are required to reduce present-day spatial curvature to $|\Omega_{K0}| < 10^{-3}$. (Assume $h = 0.7$ and that neutrinos are massless.) Approximate that the expansion rate at the beginning of inflation is completely dominated by the inflaton, that the inflaton field value does not change during inflation and that reheating happens instantaneously. In which directions do the above approximations change the result? What is the number of e-folds for $T_{\text{reh}} = 10^7$ GeV?

7.6 Scales of inflation

7.6.1 E-folds of inflation

During inflation, the scale factor $a(t)$ grows by a huge factor. We define the *number of e-foldings* from time t to end of inflation t_{end} as

$$N(t) \equiv \ln \frac{a(t_{\text{end}})}{a(t)} . \quad (7.42)$$

We can calculate $N(t) \equiv N(\varphi(t)) \equiv N(\varphi)$ from the shape of the potential $V(\varphi)$ and the value of φ at time t :

$$N(\varphi) = \ln \frac{a(t_{\text{end}})}{a(t)} = \int_t^{t_{\text{end}}} H(t) dt = \int_{\varphi}^{\varphi_{\text{end}}} \frac{H}{\dot{\varphi}} d\varphi \stackrel{\text{slow roll}}{\approx} \frac{1}{M_{\text{Pl}}^2} \int_{\varphi_{\text{end}}}^{\varphi} \frac{V}{V'} d\varphi , \quad (7.43)$$

where we have used $\frac{da}{a} = d \ln a = H dt = H \frac{d\varphi}{\dot{\varphi}}$.

7.6.2 Evolution of scales

When discussing the evolution of density perturbations and formation of structures in the universe (which will get to in the next two chapters), we will be interested in

the history of each comoving distance scale, or each *comoving wave number* k (from Fourier expansion in comoving coordinates).

$$k = \frac{2\pi}{\lambda}, \quad k^{-1} = \frac{\lambda}{2\pi}$$

An important question is whether a distance scale is larger or smaller than the Hubble length at a given time. A scale is said to be

- super-Hubble, when $k < aH$ ($k^{-1} > (aH)^{-1}$)
- at the Hubble radius (exiting or entering the Hubble radius), when $k = aH$
- sub-Hubble, when $k > aH$ ($k^{-1} < (aH)^{-1}$) .

The terms “sub-horizon”, “entering/exiting the horizon” and “superhorizon” are also often used, although they are technically wrong, as in inflationary models the horizon is much larger than the Hubble radius⁷. Note that large length scales (large k^{-1}) correspond to small wavenumbers k , and vice versa, although we often talk about “scale k ”. This can cause confusion, so be careful with wording! Recall that $(aH)^{-1}$ shrinks during inflation and grows during all other eras. See figures 9 and 10.

We will later find that the amplitude of primordial density perturbations on a given comoving scale freezes as the scale exits the Hubble radius during inflation. The largest observable scales are of the size of the Hubble radius today. (Since the universe has recently started to accelerate again, these scales have just barely entered, and are now exiting again.)

To identify the distance scales during inflation with the corresponding distance scales in the present universe, we need a complete history from inflation to the present. We divide it into the following periods:

1. **From** the time the scale k of interest exits the Hubble radius during inflation **to** the end of inflation (t_k to t_{end}).
2. **From** the end of inflation **to** the time when thermal equilibrium at high temperature is achieved, i.e. reheating. We assume that the universe behaves as if matter-dominated, $\rho \propto a^{-3}$, during this period, as discussed in section 7.5.1 (t_{end} to t_{reh}).
3. **From** reheating **to** today (radiation energy density goes like $\rho_r \propto a^{-4}$) (t_{reh} to t_0).

Consider a scale k that exits at $t = t_k$, when $a = a_k$ and $H = H_k$

$$\Rightarrow k = (aH)_k = a_k H_k .$$

⁷As discussed in the first part of the course, there are (at least) three different usages for the word “horizon”:

1. particle horizon
2. event horizon
3. Hubble length.

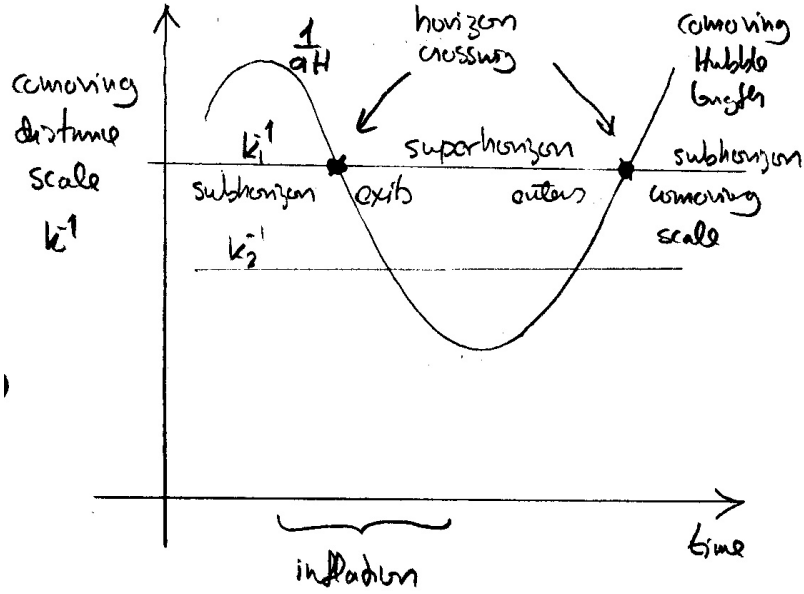


Figure 9: The evolution of the Hubble length, and two scales, k_1^{-1} and k_2^{-1} in comoving coordinates.

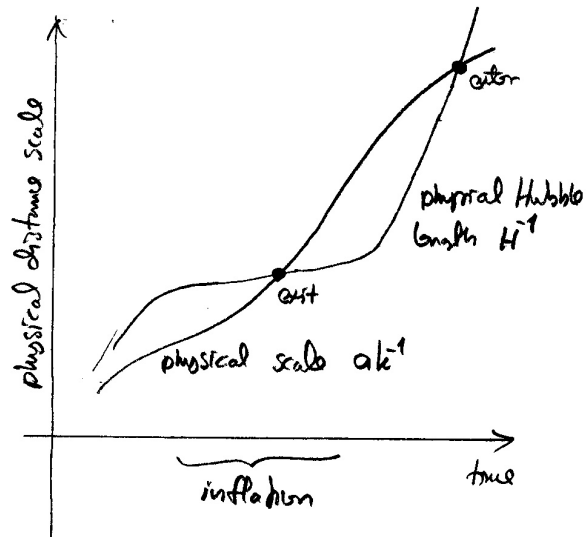


Figure 10: The evolution of the Hubble length, and the scale k^{-1} in terms of physical distance.

To find out how large this scale is today, we relate it to the a reference scale k_* :

$$\begin{aligned}
\frac{k}{k_*} &= \frac{a_k H_k}{k_*} \\
&= \frac{a_k}{a_{\text{end}}} \frac{a_{\text{end}}}{a_{\text{reh}}} \frac{a_{\text{reh}}}{a_0} \frac{H_k}{k_*} \\
&= e^{-N(k)} \left(\frac{\rho_{\text{end}}}{\rho_{\text{reh}}} \right)^{-\frac{1}{3}} \left(\frac{\rho_{\text{reh}}}{\rho_{\text{r0}}} \right)^{-\frac{1}{4}} \frac{V_k^{\frac{1}{2}}}{\sqrt{3} M_{\text{Pl}} k_*}, \tag{7.44}
\end{aligned}$$

where we have in the second equality inserted $a_0 = 1$. In the second equality we have used the relation $\frac{a_k}{a_{\text{end}}} = e^{-N(k)}$, where $N(k)$ is the number of e-foldings until the end of inflation after the scale k exits the Hubble radius and taken into account that from the end of inflation until reheating approximately $\rho \propto a^{-3}$, and that from reheating until today the radiation component (but not the total energy density) evolves like $\rho_r \propto a^{-4}$. This is slightly inaccurate, since $\rho_r \propto a^{-4}$ does not take into account the change in g_* . However, the approximation is good enough for us, as the error only enters logarithmically in the number of e-foldings⁸. (We assume that almost all of the energy density goes into particles with masses smaller than the reheating temperature.) Finally, we have used $3H^2 = \rho/M_{\text{Pl}}^2$, which follows from the Friedmann equation, and noted that during slow-roll inflation $\rho_k \approx V_k$, where the subscript k again refers to the time when the mode with wavenumber k exits the Hubble radius.

Shuffling the terms, we can write (7.44) as

$$\frac{k}{k_*} = e^{-N(k)} \frac{10^{16} \text{ GeV} \rho_{\text{r0}}^{\frac{1}{4}}}{\sqrt{3} M_{\text{Pl}} k_*} \frac{V_k^{\frac{1}{4}}}{10^{16} \text{ GeV}} \left(\frac{V_k}{V_{\text{end}}} \right)^{\frac{1}{4}} \left(\frac{\rho_{\text{reh}}}{\rho_{\text{end}}} \right)^{\frac{1}{12}}, \tag{7.46}$$

where we have inserted the comparison scale 10^{16} GeV and taken into account that $\rho_{\text{end}} \approx V_{\text{end}}$ (if inflation ends due to the slow-roll approximation being violated, this will only be true up to factors of order unity, which we neglect). We don't know the energy scale of inflation, but there is an upper limit of approximately 10^{16} GeV from the lack of observation of primordial gravity waves, whose amplitude provides a measure of the inflationary energy scale. Noting that $\rho_{\text{r0}} = 3H_0^2 h^{-2} M_{\text{Pl}}^2 \omega_r$, using

⁸Accurately this would go as:

$$g_* a^3 T^3 = \text{const.} \quad \Rightarrow \quad \frac{a_{\text{reh}}}{a_0} = \left[\frac{g_*(T_0)}{g_*(T_{\text{reh}})} \right]^{\frac{1}{3}} \frac{T_0}{T_{\text{reh}}}. \tag{7.45}$$

We approximated this with

$$\left(\frac{\rho_{\text{r0}}}{\rho_{\text{reh}}} \right)^{\frac{1}{4}} = \left[\frac{g_*(T_0)}{g_*(T_{\text{reh}})} \right]^{\frac{1}{4}} \frac{T_0}{T_{\text{reh}}}$$

Taking $g_*(T_{\text{reh}}) = g_*(T_{\text{reh}}) \sim 100$, the ratio of these two becomes

$$\frac{g_*(T_0)^{\frac{1}{3}}}{g_*(T_0)^{\frac{1}{4}} g_*(T_{\text{reh}})^{\frac{1}{12}}} \approx \frac{3.909^{\frac{1}{3}}}{3.363^{\frac{1}{4}} 100^{\frac{1}{12}}} = 0.79 \sim 1.$$

Note that $a \propto \rho_r^{-1/4}$ is a better approximation than $a \propto T^{-1}$, since these two differ by

$$\left[\frac{g_*(T_{\text{reh}})}{g_*(T_0)} \right]^{\frac{1}{4}} \sim \left(\frac{100}{3.363} \right)^{\frac{1}{4}} \sim 2.33.$$

$\omega_r = 4.18 \times 10^{-5}$, choosing the comparison scale $k_* = 0.05 \text{ Mpc}^{-1}$,⁹ converting between GeV and Mpc^{-1} , and taking the logarithm, we obtain the number of e-folds

$$N(\varphi_k) = -\ln \frac{k}{k_*} + 56 - \frac{1}{3} \ln \frac{\rho_{\text{end}}^{1/4}}{\rho_{\text{reh}}^{1/4}} + \ln \frac{V_k^{1/4}}{V_{\text{end}}^{1/4}} - \ln \frac{10^{16} \text{ GeV}}{V_k^{1/4}}, \quad (7.47)$$

where $\varphi_k \equiv \varphi(t_k)$. The terms have been arranged so that the quantities inside the logarithms are bigger than unity. For any given present scale, (7.47) identifies the value φ_k the inflaton had when this scale exited the Hubble radius during inflation. The last three terms give the dependence on the energy scales connected with inflation and reheating. In typical inflation models, they are relatively small (< 1 or < 10). The second term depends on the efficiency of reheating: if all of the inflaton potential energy is converted into radiation degrees of freedom instantaneously, it is zero. The third term is expected to be small, since the potential varies slowly during slow-roll: the dependence on k in the first term is expected to dominate. The last factor can however be large if the inflation scale is much lower than 10^{16} GeV . For example, inflation at the TeV scale would give -30 .

Anyway, we see that typically (for high scale inflation) about 60 e-foldings of inflation occur after the largest observable scales exit the Hubble radius. There is no similar constraint on the number of e-folds before these scales exited the Hubble radius, and the number varies from a few to 10^8 (or more) between different models.

Exercise. Assume that at the beginning of inflation we have $|\Omega_K| = 0.1$. a) Calculate, as a function of the reheating temperature T_{reh} , how many e-folds of inflation are required to reduce present-day spatial curvature to $|\Omega_{K0}| < 10^{-3}$. (Assume $h = 0.7$ and that neutrinos are massless.) Approximate that the expansion rate at the beginning of inflation is completely dominated by the inflaton, that the inflaton field value does not change during inflation, and that reheating happens instantaneously. b) In which directions do the above approximations change the result? c) What is the number of e-folds for $T_{\text{reh}} = 10^7 \text{ GeV}$?

7.7 Before inflation

As we discussed earlier, inflation erases all memory of the initial conditions, and we do not have a good theoretical understanding of what happened before inflation. However, there are some ideas. During inflation the universe expands and the energy density decreases.¹⁰ Sometimes it is assumed that inflation begins right after the Planck scale, $\rho \sim M_{\text{Pl}}^4$, the highest energy density to which we can extend a description based on classical general relativity, as quantum gravity is expected to become important at the Planck scale. One idea is that the universe at that time, in the *Planck era*, is a “spacetime foam”, where the spacetime has large quantum fluctuations. When the energy density of some region, larger than H^{-1} , falls below M_{Pl}^4 , spacetime in that region begins to behave in a classical manner. See figure 11.

The initial conditions, i.e. conditions at the time when our Universe emerges from the spacetime foam, are usually assumed *chaotic* (this word does not refer to

⁹This comparison scale is convenient because it is logarithmically in the middle of the scales we observe in the CMB; more on this in the next chapter and in chapter 10.

¹⁰At least in the simplest simple field models we consider. In some models the energy density can increase at least for some part of inflation.

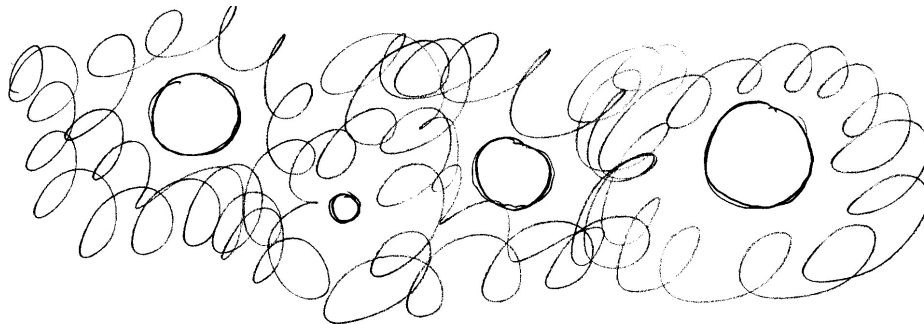


Figure 11: Classical regions emerging from spacetime foam.

chaos theory!), i.e. φ takes random values at different regions. Since $\rho \geq \rho_\varphi$, and

$$\rho_\varphi = \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}\frac{1}{a^2}\sum_i \partial_i\varphi\partial_i\varphi + V, \quad (7.48)$$

we must have

$$\dot{\varphi}^2 \lesssim M_{\text{Pl}}^4, \quad \frac{1}{a^2}\sum_i \partial_i\varphi\partial_i\varphi \lesssim M_{\text{Pl}}^4, \quad V \lesssim M_{\text{Pl}}^4 \quad (7.49)$$

in a region for it to emerge from the spacetime foam.

Inflation may begin at many different parts of the spacetime foam. Our observable universe is just one small part of one such region which has inflated to a huge size. Within the classical description, starting inflation from an inhomogeneous state has been studied both analytically and with relativistic simulations. Typically, regions where the initial field configuration is overdense and spatial gradients are significant collapse into black holes, and inflation starts only in regions that are already somewhat smooth over one Hubble radius and where the field is sufficiently displaced from its minimum (i.e. close enough to the smooth slow-roll attractor regime). However, if we weigh things by volume, it is sufficient that inflation starts in one region, as its volume will grow exponentially and dominate the total volume. It is not clear if this is the right way to think about the issue.

It is also possible that during inflation, in some part of the potential, quantum fluctuations of the inflaton dominate over the classical evolution and push φ higher in some regions. These regions will then expand faster, and dominate the volume. This gives rise to *eternal inflation*, where, at any given time, most of the volume of the universe is inflating. (Whether or not this can happen depends on the shape of the potential and the field value during inflation.) But our observable Universe is part of a region where φ rolled down and came to a region of the potential, where the quantum fluctuations of φ were small and the classical behaviour began to dominate and inflation ended.

Thus the ultra-large scale structure of the universe may be very complicated. However, this is not observable to us, and all the features of the universe we see can be explained in terms of what happened in our patch during and after inflation. These ideas of spacetime foam and eternal inflation are rather speculative, and there are also different suggestions for the early stages of the universe.

References

- [1] PLANCK collaboration, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [[1807.06209](#)].
- [2] J. Martin, C. Ringeval and V. Vennin, *Encyclopædia Inflationaris: Opiparous Edition*, *Phys. Dark Univ.* **5-6** (2014) 75–235, [[1303.3787](#)].