Cosmology II

Hannu Kurki-Suonio

Fall 2020

Preface

In Cosmology I we discussed the universe in terms of a homogenous and isotropic approximation to it. In Cosmology II we add the inhomogeneity and anisotropy (Chapters 8 and 9). The mathematical background required includes Fourier analysis (taught in Fysiikan matemaattiset menetelmät I) and spherical harmonic analysis (taught in Fysiikan matemaattiset menetelmät II). We will take some results from Quantum Field Theory and Cosmological Perturbation Theory, but students are not expected to have them as background – they are more advanced courses. We begin with Inflation, but postpone the discussion of generation of perturbations during it to after we have discussed inhomogeneity in general and its later evolution – the chapter on Structure Formation. Thus in the Inflation chapter we still assume the homogeneous FRW model. We end with the Cosmic Microwave Background Anisotropy, which forms an important part of observational data in cosmology.

7 Inflation

7.1 Motivation

In Cosmology I we discussed how the universe began with a Hot Big Bang. This leaves open the question of initial conditions – how did the Hot Big Bang begin, and why did it begin with such a state of high density and temperature, rapid expansion, and a high level of isotropy and homogeneity. Inflation is a *scenario* to address this question, at least to some extent. Inflation is a period in the very early universe, before the events discussed in Cosmology I, when the expansion of the universe was *accelerating*.

Inflation is not really a specific theory; rather it is a more general idea of a certain kind of behavior (i.e., a "scenario") for the universe. It is not known for sure whether inflation occurred, but it makes a number of predictions that agree with observations. Inflation has been more successful than competing ideas for the very early universe and it has become part of the standard model for cosmology. The most important property of inflation is that it provides a mechanism for generating the initial density fluctuations, the *primordial perturbations*, from which the structure of the universe, stars and galaxies, grew. However, this property was discovered later, and the original motivation for inflation was to explain the initial flatness and homogeneity of the universe and the lack of certain relics that could have been produced at the very high temperatures of the very early universe[1][1]. This chapter discusses inflation in the homogeneous and isotropic approximation. Perturbations are discussed in Chapter 8.

Much of this chapter follows Chapter 3 of the book by Liddle&Lyth.[2]

7.1.1 Flatness problem

The Friedmann equation can be written as

$$\Omega - 1 = \frac{K}{a^2 H^2}\,. \tag{1}$$

If the universe has the critical density, $\Omega = 1$, it stays that way (since $K = 0$). But if $\Omega \neq 1$, it evolves in time. The difference $\Omega_k = 1 - \Omega$ grows with time during both the radiation-dominated and matter-dominated epochs. If Ω_k is small, its time evolution takes the form

$$\text{mat.dom} \qquad a \propto t^{2/3},\ H \propto t^{-1} \Rightarrow \frac{1}{aH} \propto t^{1/3} \qquad \Rightarrow \Omega_k \propto t^{2/3} \tag{2}$$

$$\text{rad.dom} \qquad a \propto t^{1/2},\ H \propto t^{-1} \Rightarrow \frac{1}{aH} \propto t^{1/2} \qquad \Rightarrow \Omega_k \propto t\,. \tag{3}$$

Since today, and at the end of the matter-dominated epoch, $\Omega_0 = \mathcal{O}(1)$ – and it is not essential here that Ω_0 is very close to 1, it would be enough that, say, $0.1 < \Omega_0 < 10$ – we can calculate backwards in time to, e.g., Big Bang Nucleosynthesis (BBN) and we find that the density parameter must have been extremely close to 1 then:

$$|\Omega(t_{\rm BBN}) - 1| \;=\; |\Omega_k(t_{\rm BBN})| \;\lesssim\; 10^{-16} \tag{4}$$

Thus we get as an initial condition to Big Bang, that Ω must have been initially extremely close to 1. The flatness problem is to explain why it was so. Otherwise, if we start the FRW universe in a radiation-dominated state with some initial value of the density parameter Ω_i not extremely close to 1, one of two things happens:

- $\Omega_i > 1 \quad \Rightarrow$ the universe recollapses almost immediately

[1] Guth[1] was not the first to propose a period of accelerating expansion in the early universe, but it was his proposal that became widely known and made inflation popular.

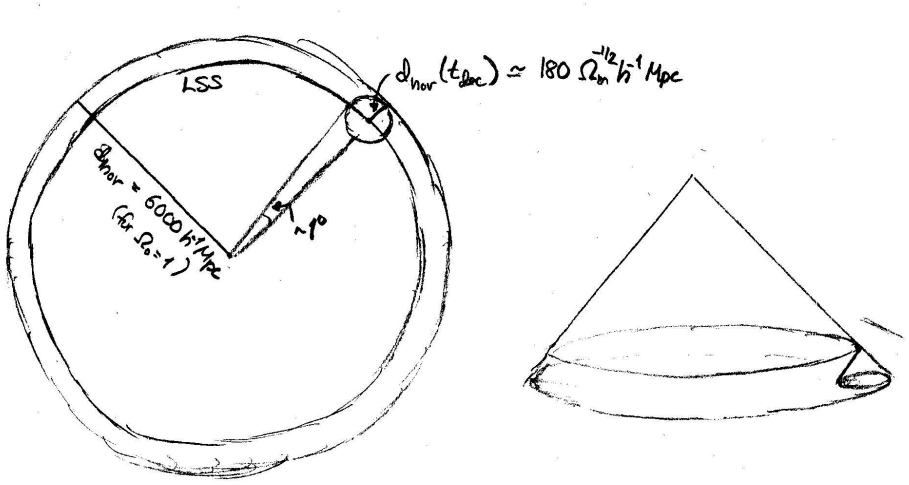


Figure 1: The horizon problem: regions on the CMB sky separated by more than about 1° had not had time to interact, yet their temperature is the same with an accuracy of $\lesssim 10^{-4}$.

- $\Omega_i < 1 \quad \Rightarrow$ the universe expands very fast and cools to $T < 3\,\mathrm{K}$ in a very short time.

Thus the flatness problem can also be called the oldness problem: why did it take so long, 14 billion years, for the universe to cool to $T = 2.7\,\mathrm{K}$.

Exercise: Oldness problem. Assume $\Omega_k(T = 100\,\mathrm{keV}) = 0.1$. (BBN takes place near $T = 100\,\mathrm{keV}$). Include just curvature and radiation (with $g_* = 3.384$) in the Friedmann equation. How long does it take for the universe to cool to $T = 2.7\,\mathrm{K}$? Why would inclusion of matter (with, say $\eta = 6\times10^{-10}$, and $\rho_m = 6\rho_b$) not change the answer?

7.1.2 Horizon problem

The horizon problem can also be called the homogeneity problem. The cosmic microwave background (CMB), which shows the universe at $z = 1090$ (age 370 000 years), is remarkably isotropic, the relative temperature variations being only $\mathcal{O}(10^{-4})$. This implies that density variations at that time must have been also very small, so the early universe was very homogeneous. Calculated according to the standard Hot Big Bang model, the horizon distance at that time was much smaller than the part of the early universe we see in the CMB, corresponding to only about 1° on the sky. Thus there could not have been any process to homogenize conditions over scales larger than this. This implies that this level of homogeneity must have been an initial condition.

Even the small CMB anisotropies show correlations at larger scales that 1°, a fact discovered after inflation was proposed.

7.1.3 Unwanted relics

If the Hot Big Bang begins at very high T it may produce objects surviving to the present, that are ruled out by observations.

- **Gravitino.** The supersymmetric partner of the graviton. $m \sim 100\,\mathrm{GeV}$. They interact very weakly (gravitational strength) $\Rightarrow$ they decay late, after BBN, and ruin the success of BBN.

- **Magnetic monopoles.** If the symmetry of a Grand Unified theory (GUT) is broken in a spontaneous symmetry breaking phase transition, magnetic monopoles are produced. These are point-like *topological defects* that are stable and very massive, $m \sim T_{\rm GUT} \sim 10^{14}\,{\rm GeV}$. Their expected number density is such that their contribution to the energy density today $\gg$ the critical density.

- **Other topological defects** (cosmic strings, domain walls). These may also be produced in a GUT phase transition, and may also be a problem, but this is model-dependent. On the other hand, *cosmic strings* had been suggested as a possible explanation for the initial density perturbations—but this scenario fell later in trouble with the observational data (especially the anisotropy of the CMB).

These relics are produced very early, at extremely high temperatures, typically $T \gtrsim 10^{14}\,{\rm GeV}$. From BBN, we only know that we should have standard Hot Big Bang for $T \lesssim 1\,{\rm MeV}$.

7.1.4 What is needed

The word "problem" in the preceding is not to be taken to imply that the Hot Big Bang theory for the early universe would be in trouble. The theory by itself just does not contain answers to some questions one may pose about its initial conditions, for which we thus need additional ideas. We are perfectly happy if we can produce as an "initial condition" for Big Bang a universe with temperature $1\,{\rm MeV} < T < 10^{14}\,{\rm GeV}$, which is almost homogeneous and has $\Omega = 1$ with extremely high precision.

7.2 Inflation introduced

7.2.1 Accelerated expansion

Inflation is not a replacement for the Hot Big Bang, but an addition to it, occurring at very early times (e.g., $t \sim 10^{-35}$ s), without disturbing any of its successes. Thus we have first inflation, then Hot Big Bang; so that inflation produces the initial conditions for the Hot Big Bang.

The origin of the flatness problem is that $|\Omega - 1| = |K|/(aH)^2$ grows with time. Now

$$\frac{d}{dt}|\Omega - 1| \;=\; |K|\frac{d}{dt}\left(\frac{1}{a^2H^2}\right) \;=\; |K|\frac{d}{dt}\left(\frac{1}{\dot{a}^2}\right) \;=\; \frac{-2|K|}{\dot{a}^3}\ddot{a}\,. \tag{5}$$

For an expanding universe, $aH = a\left(\frac{\dot{a}}{a}\right) = \dot{a} > 0$. Thus $\dot{a}^3 > 0$, and

$$\frac{d}{dt}|\Omega - 1| > 0 \qquad \Leftrightarrow \qquad \ddot{a} < 0\,. \tag{6}$$

Thus the reason for the flatness problem is that the expansion of the universe is decelerating, i.e., slowing down. If we had an early period in the history of the universe, where the expansion was accelerating, it could make an initially arbitrary value of $|\Omega - 1| = |K|/(aH)^2$ very small.

Definition: Inflation = any epoch when the expansion is accelerating.

$$\text{Inflation} \;\Leftrightarrow\; \ddot{a} > 0 \tag{7}$$

Consider then the horizon problem. The horizon at photon decoupling, $d^p_{\rm hor}(t_{\rm dec})$ is somewhere between the radiation-dominated and matter-dominated values, H^{-1} and $2H^{-1}$. For comparing sizes of regions at different times, we should use their comoving sizes, $d^c \equiv d^p/a$. We have

$$d^c_{\rm hor}(t_{\rm dec}) \sim \frac{1}{a_{\rm dec}H_{\rm dec}}\,, \tag{8}$$

whereas the size of the observable universe today is of the order of the present Hubble length

$$d^c_{\rm hor}(t_0) \sim H_0^{-1}\,. \tag{9}$$

The horizon problem arises because the first is much smaller than the second,

$$\frac{d^c_{\rm hor}(t_{\rm dec})}{d^c_{\rm hor}(t_0)} \sim \frac{a_0 H_0}{a_{\rm dec} H_{\rm dec}} \ll 1\,. \tag{10}$$

Thus the problem is that aH, whose inverse gives roughly the comoving size of the horizon, decreases with time,

$$\frac{d}{dt}(aH) = \frac{d}{dt}(\dot a) = \ddot a < 0\,. \tag{11}$$

Having a period with $\ddot a > 0$ could solve the problem.

In the preceding we referred to the (comoving) horizon distance at some time t, defined as the comoving distance light has traveled from the beginning of the universe until time t. If there are no surprises at early times, we can calculate or estimate it; like in the preceding where we assumed radiation-dominated or matter-dominated behavior (standard Big Bang). If we now start adding other periods, like accelerating expansion at early times, the calculation of $d_{\rm hor}$ will depend on them. In principle, $d^c_{\rm hor}(t_0) > d^c_{\rm hor}(t_{\rm dec})$ always, since $t_0 > t_{\rm dec}$, so $(0, t_{\rm dec}) \subset (0, t_0)$. But note that in the horizon problem, the relevant present horizon is how far we can *see*: the observable universe is given just by the integrated comoving distance the photon has traveled in the interval $(t_{\rm dec}, t_0)$, which is not affected by what happens before $t_{\rm dec}$. Thus the relevant present horizon is still $\sim H_0^{-1}$.

What is the relation between $d^c_{\rm hor}(t)$ and $1/(aH)$ for arbitrary expansion laws? Introduce the comoving, or conformal, Hubble parameter,

$$\mathcal{H} \equiv aH = \frac{1}{a}\frac{da}{d\eta} \equiv \dot a\,, \tag{12}$$

where η is the conformal time, defined by $d\eta = dt/a$. The Hubble length is

$$l_H \equiv H^{-1}\,, \qquad \text{where } H \equiv \frac{\dot a}{a}\,, \tag{13}$$

and the *comoving Hubble length* is

$$l^c_H \equiv \frac{l_H}{a} = \frac{1}{aH} = \frac{1}{\dot a} = \mathcal{H}^{-1}\,. \tag{14}$$

Roughly speaking, $\mathcal{H}^{-1}$ gives the comoving distance light travels in a "cosmological timescale", i.e., the Hubble time. This statement cannot be exact, since both the comoving Hubble length and the Hubble time change with time. However, if $\mathcal{H}^{-1}$ is increasing with time, the comoving distances traveled at earlier "epochs" are shorter, and thus $\mathcal{H}^{-1}(t)$ is a good estimate for the total comoving distance light has traveled since the beginning of time (the horizon). On the other hand, if $\mathcal{H}^{-1}$ is shrinking, then at earlier epochs light was traveling longer comoving distances, and we expect the horizon at time t to be larger than $\mathcal{H}^{-1}$. In any case

$$d^c_{\rm hor}(t) \gtrsim \mathcal{H}(t)^{-1}\,. \tag{15}$$

Since the Hubble length is more easily "accessible" (less information needed to figure it out) than the horizon distance it has become customary in cosmology to use the word "horizon" also for the Hubble distance. We shall also adopt this practice. The Hubble length gives the distance over which we have causal interaction in cosmological timescales. The comoving Hubble length gives this distance in comoving units.

If aH is decreasing (Eq. 11) then $\mathcal{H}^{-1}$ increases, and vice versa.

$\therefore$ Inflation = any epoch when the comoving Hubble length is shrinking.

$$\text{Inflation} \;\Leftrightarrow\; \frac{d}{dt}\mathcal{H}^{-1} < 0 \tag{16}$$

Thus the comoving distance over which we have causal connection is *decreasing* during inflation: causal contact to other parts of the Universe is being lost.

Inflation can be discussed either 1) in terms of physical distances or 2) in terms of comoving distances.

1) In terms of physical distance, the distance between any two points in the Universe is increasing, with an accelerating rate. The distance over which causal connection can be maintained is increasing (much) more slowly.

2) In terms of comoving distance, i.e., viewed in comoving coordinates, the distance between two points stays fixed; regions of the universe corresponding to present structures maintain fixed size. From this viewpoint (the one normally adopted), the region causally connected to a given location in the Universe is shrinking.

To connect with dynamics, look at the second Friedmann equation,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \tag{17}$$

$$\therefore \text{Inflation} \;\Leftrightarrow\; \rho + 3p < 0 \;\Leftrightarrow\; w < -\tfrac{1}{3} \tag{18}$$

Thus inflation requires negative pressure, $p < -\frac{1}{3}\rho$ (we assume $\rho \geq 0$).

There is a huge class of models to realize the inflation scenario. These models rely on so-far-unknown physics of very high energies. Some models are just "toy models", with a hoped-for resemblance to the actual physics of the early universe. Others are connected to proposed extensions (like supersymmetry) to the standard model of particle physics.

The important point is that inflation makes many *generic*[2] predictions, i.e., predictions that are independent of the particular model of inflation. Present observational data agrees with these predictions. Thus it is widely believed—or considered probable—by cosmologists that inflation indeed took place in the very early universe. There are also numerical predictions of cosmological observables that differ from one model of inflation to another, allowing future observations to rule out classes of such models. (Many inflation models are already ruled out.)

7.2.2 Solving the problems

Inflation can solve[3] all the problems discussed in Sec. 7.1. The idea is that during inflation the universe expands by a large factor (at least by a linear factor of something like $\sim e^{70} \sim 10^{30}$ to solve the problems). This cools the universe to $T \sim 0$ (if the concept of temperature is applicable). When inflation ends, the universe is heated to a high temperature, and the usual Hot Big Bang history follows. This heating at the end of inflation is called *reheating*, since originally the thinking was that inflation started at an earlier hot epoch, but it is actually not clear whether that was the case.

[2] I was once in a conference where a speaker began his talk on inflation by promising not to use the words "generic" or "scenario". He failed in one but not the other.

[3] This is not to be taken too rigorously. The problems are related to the question of initial conditions of the universe at some very early time, whose physics we do not understand. Thus theorists are free to have different views on what kind of initial conditions are "natural". Inflation makes the flatness and horizon problems "exponentially smaller" in some sense, but inflation still places requirements—on the level of homogeneity—for the initial conditions *before* inflation, so that inflation can begin.

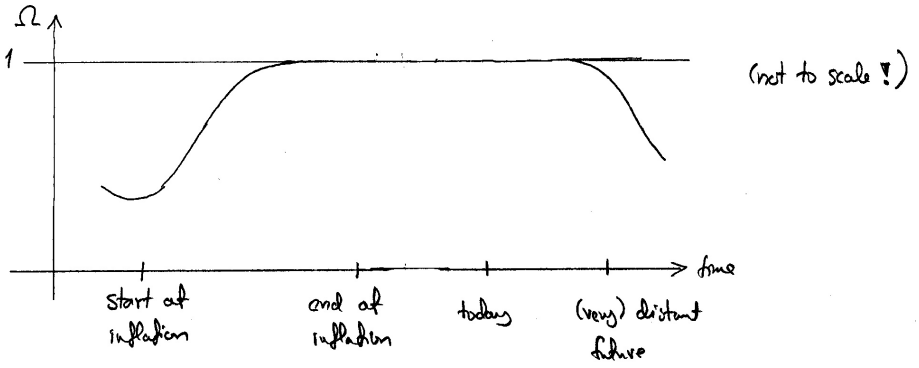


Figure 2: Solving the flatness problem. This figure is for a universe with no dark energy, where the expansion keeps decelerating after inflation ended in the early universe. Present observational evidence indicates that actually the expansion began accelerating again (supposedly due to the mysterious dark energy) a few billion years ago. Thus the universe is, technically speaking, inflating again, and Ω is again being driven towards 1. However, this current epoch of inflation is not enough to solve the flatness problem, or the other problems, since the universe has only expanded by about a factor of 2 during it.

Solving the flatness problem: The flatness problem is solved, since during inflation

$$|1-\Omega| = \frac{|K|}{\mathcal{H}^2} \quad \text{is shrinking.} \tag{19}$$

Thus inflation drives $\Omega \to 1$. Starting with an arbitrary Ω, inflation drives $|1-\Omega|$ so small that, although it has grown all the time from the end of inflation to the recent onset of dark energy domination, it is still very small today. See Fig. 2. In fact, inflation predicts that $\Omega_0 = 1$ to high accuracy, since it would be an unnatural coincidence for inflation to last just the right amount so that Ω would begin to deviate from 1 just at the current epoch.[4]

Solving the horizon problem: The horizon problem is solved, since during inflation the causally connected region is shrinking. It was very large before inflation; much larger than the present horizon. Thus the present observable universe has evolved from a small patch of a much larger causally connected region; and it is natural that the conditions were (or became) homogeneous in that patch then. See Fig. 3.

Getting rid of relics: If unwanted relics are produced before inflation, they are diluted to practically zero density by the huge expansion during inflation. We just have to take care they are not produced after inflation, i.e., the reheating temperature has to be low enough. This is an important constraint on models of inflation.

Did we really solve the problem of initial conditions? Actually solving the flatness and horizon problems is more complicated. We discussed them in terms of a FRW universe, which by assumption is already homogeneous. In fact, for inflation to get started, a sufficiently large region which is not too inhomogeneous and not too curved, is needed. We shall not discuss this in more detail, since the solution of these problems is not the most important aspect of inflation.

If inflation happened, we expect that the early universe after inflation was very homogeneous except for fluctuations generated during inflation and that $\Omega_0 = 1$. Thus inflation leads to predictions that can be tested with observations. More important than flatness and homogeneity

[4]Thus, if it were discovered by observations, that actually $\Omega_0 \neq 1$, this would be a blow to the credibility of inflation. However, there is a version of inflation, called *open inflation*, for which it is natural that $\Omega_0 < 1$. The existence of such models of inflation have led critics of inflation to complain that inflation is "unfalsifiable"—no matter what the observation, there is a model of inflation that agrees with it. Nevertheless, most models of inflation give the same "generic" predictions, including $\Omega_0 = 1$.

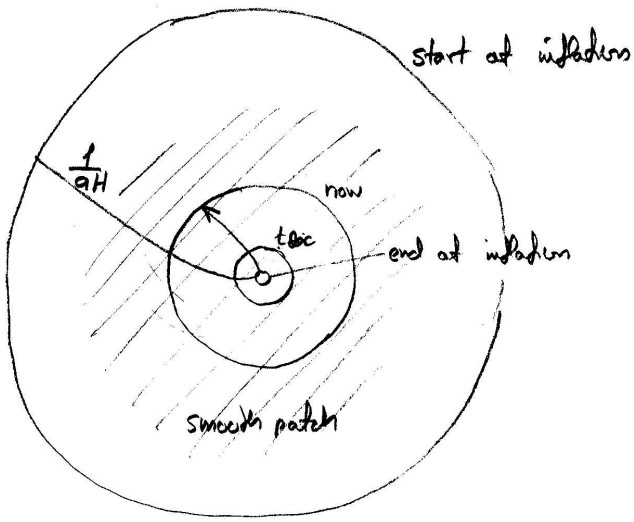


Figure 3: Evolution of the comoving Hubble radius (length, distance) during and after inflation (schematic).

are the predictions inflation makes about primordial perturbations, the "seeds" for structure formation, discussed in the next chapter.

Thus we assume that sufficient inflation has already taken place to make the universe (within a horizon volume) flat and homogeneous, and follow the inflation in detail after that, working in the flat FRW universe.

7.3 Quantum field theory for children

The theories (known and hypothetical) needed to describe the (very) early universe are *quantum field theories* (QFT). The fundamental entities of these theories are *fields*, i.e., functions of space and time. For each particle species, there is a corresponding field, having at least as many (real) components φ_i as the particle has internal degrees of freedom. For example, for the photon, the corresponding field is the vector field $A^\mu = (A^0, A^1, A^2, A^3) = (\phi, \vec{A})$, which you are probably familiar with from electrodynamics.[5] The photon has two internal degrees of freedom. The larger number of components in A^μ is related to the *gauge freedom* of electrodynamics. Since A^μ is a (Lorentz) vector field, it has the same number of components as there are spacetime dimensions, but other types of fields do not have this correspondence.

In classical field theory the evolution of the field is governed by the *field equation*. From the field equation one can identify a field potential, an expression in terms of the field, which helps to understand the field dynamics. Quantizing a field theory gives a quantum field theory. *Particles are quanta of the oscillations of the field around the minimum of its potential.* The state where the field values are constant at the potential minimum is called the *vacuum*. Up to now, we have described the events in the early universe in terms of the *particle picture*. However, the particle picture is not fundamental, and can be used only when the fields are doing small oscillations. For many possible events and objects in the early universe (inflation, topological defects, spontaneous symmetry breaking phase transitions) the field behavior is different, and we need to describe them in terms of field theory. In some of these topics classical field theory is already sufficient for a reasonable and useful description.

In this section we discuss "low-temperature" field theory in Minkowski space, i.e., we forget high-temperature effects and the curvature of spacetime.

The starting point in field theory is the *Lagrangian density*, a function of space and time, which is a scalar quantity constructed from the fields and their derivatives:

$$\mathcal{L}(\varphi_i, \partial_\mu \varphi_i)\,. \tag{20}$$

The Lagrangian density can be expressed as a sum of two parts, the *kinetic term*, which depends on field derivatives (gradients), and the *field potential* $V(\varphi_1, \ldots, \varphi_N)$ (for a theory with N fields), which does not. This expression for the Lagrangian density as a function of the fields and their derivatives defines the field theory, and one can derive the field equations (differential equations governing the field evolution) and the energy-momentum tensor (energy density and pressure of the fields) from the Lagrangian density. Usually the kinetic term has a simple form, called the canonical kinetic term, and we assume that here. The remaining freedom in defining the field theory is in defining the potential.

The simplest case is a theory with one scalar field φ, for which

$$\mathcal{L} \;=\; -\tfrac{1}{2}\partial_\mu\varphi\,\partial^\mu\varphi - V(\varphi)\,. \tag{21}$$

(We use the Einstein summation convention, where a repeated index implies summation over it, here $\mu = 0, 1, 2, 3$. Also $\partial_\mu \equiv \partial/\partial x^\mu$, where $x^0 = t$. Here we are in Minkowski space and use Cartesian coordinates, so that $\partial^0 = -\partial_0$ and $\partial^j = \partial_j$ for $j = 1, 2, 3$.) We write

$$V'(\varphi) \equiv \frac{dV}{d\varphi} \qquad \text{and} \qquad V''(\varphi) \equiv \frac{d^2V}{d\varphi^2}\,. \tag{22}$$

The *field equation*, which determines the classical evolution of the field, is obtained from the Lagrangian by minimizing (or extremizing) the action

$$\int \mathcal{L}\, d^4x\,, \tag{23}$$

[5] ϕ and $\vec{A}$ are the electromagnetic scalar and vector potentials. This is a different use of the word "potential" than what we use here. In our terminology, $A^\mu = (\phi, \vec{A})$ are fields.

which leads to the *Euler–Lagrange equation*

$$\frac{\partial \mathcal{L}}{\partial \varphi_i(x)} - \partial_\mu \frac{\partial \mathcal{L}}{\partial[\partial_\mu \varphi_i(x)]} = 0\,. \tag{24}$$

For the above scalar field we get the field equation

$$\partial_\mu \partial^\mu \varphi - V'(\varphi) = 0\,. \tag{25}$$

where $\partial_\mu \partial^\mu \varphi = -\ddot{\varphi} + \nabla^2 \varphi$, so that the field equation is

$$\boxed{\ddot{\varphi} - \nabla^2 \varphi \;=\; -V'(\varphi)\,.} \tag{26}$$

Here we use the overdot to denote partial derivative with respect to time: $\dot{} = \partial_0 = \partial/\partial t$.

The Lagrangian also gives us the energy tensor

$$T^{\mu\nu} = -\frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi)} \partial^\nu \varphi + g^{\mu\nu} \mathcal{L}\,. \tag{27}$$

For the scalar field

$$T^{\mu\nu} = \partial^\mu \varphi \partial^\nu \varphi - g^{\mu\nu} \big[\frac{1}{2} \partial_\rho \varphi \partial^\rho \varphi + V(\varphi)\big]\,. \tag{28}$$

In particular, the energy density $\rho = T^{00}$ and pressure $p = \frac{1}{3}(T^{11} + T^{22} + T^{33})$ of a scalar field are

$$\rho \;=\; \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}(\nabla\varphi)^2 + V(\varphi) \tag{29}$$

$$p \;=\; \frac{1}{2}\dot{\varphi}^2 - \frac{1}{6}(\nabla\varphi)^2 - V(\varphi)\,. \tag{30}$$

(We are in Minkowski space, so that $g^{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$). We see that the pressure due to a scalar field may be negative. The minimum value of $V(\varphi)$ is the *vacuum energy*. In principle it could be negative, acting like a negative cosmological constant. Any other contribution to ρ is positive. Since there is no evidence for a negative vacuum energy or cosmological constant, let us assume that $V(\varphi) \geq 0$.

If $V(\phi) \equiv const$ (typically assumed to be 0) the field equation becomes the wave equation

$$\ddot{\varphi} \;=\; \nabla^2 \varphi\,, \tag{31}$$

whose solutions are waves propagating at the speed of light.

For the corresponding quantum theory, the potential gives information about the masses and interactions of the particles that are the quanta of the field oscillations. The particles corresponding to scalar fields are spin-0 bosons. Spin-$\frac{1}{2}$ particles correspond to spinor fields and spin-1 particles to vector fields. The case $V(\varphi) \equiv 0$ corresponds to massless noninteracting particles. If the potential has the form

$$V(\varphi) = \frac{1}{2} m^2 \varphi^2\,, \tag{32}$$

the particle corresponding to the field φ will have mass m and it will have no interactions. In general, the mass of the particle is given by $m^2 = V''(\varphi)$.

Interactions between particles of two different species are due to terms in the Lagrangian which involve both fields. For example, in the Lagrangian of quantum electrodynamics (QED) the term

$$-ie\psi^\dagger \gamma^0 \gamma^\mu A_\mu \psi \tag{33}$$

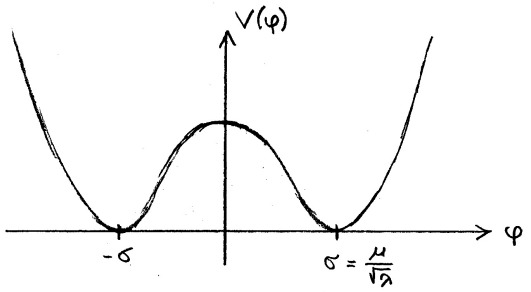
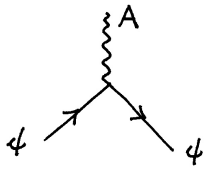


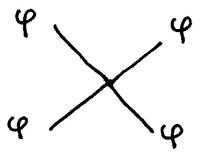
Figure 4: Potential giving rise to spontaneous symmetry breaking.

is responsible for the interaction between photons (A^μ) and electrons (ψ). (The γ^μ are Dirac matrices). A graphical representation of this interaction is the Feynman diagram



A higher power (third or fourth) of a field, e.g.,

$$V(\varphi) = \frac{1}{4}\lambda\varphi^4\,, \tag{34}$$



represents self-interaction, i.e., φ particles interacting with each other directly (as opposed to, e.g., electrons, who interact with each other indirectly, via photons). In QCD, gluons have this property.

Some theories exhibit *spontaneous symmetry breaking* (SSB). For example, the potential

$$V(\varphi) = V_0 - \frac{1}{2}\mu^2\varphi^2 + \frac{1}{4}\lambda\varphi^4 \tag{35}$$

has two minima, at $\varphi = \pm\sigma$, where $\sigma = \mu/\sqrt{\lambda}$. At low temperatures, the field is doing small oscillations around one of these two minima (see Fig. 4). Thus the vacuum value of the field is nonzero. If the Lagrangian has interaction terms, $c\varphi\psi^2$, with other fields ψ, these can now be separated into a mass term, $c\sigma\psi^2$, for ψ and an interaction term, by redefining the field φ as

$$\varphi = \sigma + \tilde{\varphi} \quad \Rightarrow \quad c\varphi\psi^2 = c\sigma\psi^2 + c\tilde{\varphi}\psi^2\,. \tag{36}$$

Thus spontaneous symmetry breaking gives the ψ particles a mass $\sqrt{2c\sigma}$. This kind of a field φ is called a *Higgs field*. In electroweak theory the fermion masses[6] are due to a Higgs field.

[6]The origin of neutrino masses is not clear.

7.4 Inflaton field

As we saw in Sec. 7.2, inflation requires negative pressure. In Chapter 4 we considered systems of particles where interaction energies can be neglected (ideal gas approximation). For such systems the pressure is always nonnegative. However, negative pressure is possible in systems with attractive interactions. In the field picture, negative pressure comes from the potential term. In many models of inflation, the inflation is caused by a scalar field. This scalar field (and the corresponding spin-0 particle) is called the *inflaton*.

Historical note. The idea of scalar fields playing an important role in the very early universe was very natural at the time inflation was proposed by Guth[1]. We already mentioned how a scalar field, the Higgs field, is responsible for the electroweak phase transition at $T \sim 100\,\mathrm{GeV}$. It is thought that at a much higher temperature, $T \sim 10^{14}\,\mathrm{GeV}$, another spontaneous symmetry breaking phase transition occurred, the GUT (Grand Unified Theory) phase transition, so that above this temperature the strong and electroweak forces were unified. This GUT phase transition gives rise to the monopole problem. Guth realized that the Higgs field associated with the GUT transition might lead the universe to "inflate" (the term was coined by Guth), solving this monopole problem. It was soon found out, however, that inflation based on the GUT Higgs field is not a viable inflation model, since in this model too strong inhomogeneities were created. So the inflaton field must be some other scalar field. The supersymmetric extensions of the standard model contain many inflaton field candidates.

During inflation the inflaton field is *almost* homogeneous.[7] The energy density and pressure of the inflaton are thus those of a homogeneous scalar field,

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\varphi}^2 + V(\varphi) \\ p &= \frac{1}{2}\dot{\varphi}^2 - V(\varphi)\,, \end{aligned} \tag{37}$$

where $V(\varphi) \geq 0$. For the equation-of-state parameter $w \equiv p/\rho$ we have

$$w = \frac{\dot{\varphi}^2 - 2V(\varphi)}{\dot{\varphi}^2 + 2V(\dot{\varphi})} = \frac{1 - (2V/\dot{\varphi}^2)}{1 + (2V/\dot{\varphi}^2)}\,, \tag{38}$$

so that

$$-1 \leq w \leq 1\,. \tag{39}$$

If the kinetic term $\dot{\varphi}^2$ dominates, $w \approx 1$; if the potential term $V(\varphi)$ dominates, $w \approx -1$.

For the present discussion, the potential $V(\varphi)$ is some arbitrary non-negative function. Different inflaton models correspond to different $V(\varphi)$. From Eq. (37), we get the useful combinations

$$\begin{aligned} \rho + p &= \dot{\varphi}^2 \\ \rho + 3p &= 2\left[\dot{\varphi}^2 - V(\varphi)\right]\,. \end{aligned} \tag{40}$$

We already had the field equation for a scalar field in Minkowski space,

$$\ddot{\varphi} - \nabla^2\varphi = -V'(\varphi)\,. \tag{41}$$

For the homogenous case it is just

$$\ddot{\varphi} = -V'(\varphi)\,. \tag{42}$$

We get a working mental picture of the evolution of a homogeneous field by comparing it to the classical mechanics equation for a particle in a gravitational potential $V(\mathbf{r})$ whose accelaration

[7] Inflation makes the inflaton field homogenous. Again, a sufficient level of initial homogeneity of the field is required to get inflation started. We start our discussion when a sufficient level of inflation has already taken place to make the gradients negligible.

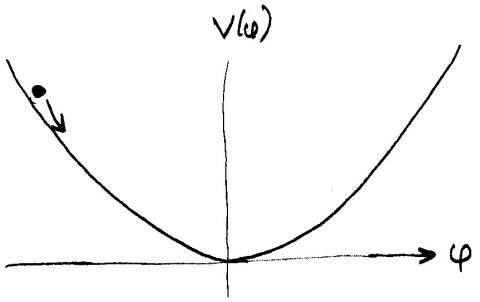


Figure 5: The inflaton and its potential.

is given by $\ddot{\mathbf{r}} = -\nabla V(\mathbf{r})$. Thus we can think of the field "rolling down" its potential like a stone rolling down a hillside, see Fig. 5; this motion is governed by Eq. (42).

We need to modify (42) for the expanding universe. We do not need to go to the GR formulation of field theory, since the modification for the present case can be simply obtained by sticking the ρ and p from Eq. (37) into the energy continuity equation

$$\dot{\rho} = -3H(\rho + p)\,. \tag{43}$$

This gives

$$\dot{\varphi}\ddot{\varphi} + V'(\varphi)\dot{\varphi} = -3H\dot{\varphi}^2 \;\Rightarrow\; \boxed{\ddot{\varphi} + 3H\dot{\varphi} = -V'(\varphi)}\,, \tag{44}$$

the field equation for a homogeneous φ in an expanding (FRW) universe. We see that the effect of expansion is to add the term $3H\dot{\varphi}$, which acts like a *friction term*, slowing down the evolution of φ.

The condition for inflation, $\rho + 3p = 2\dot{\varphi}^2 - 2V(\dot{\varphi}) < 0$, is satisfied, if

$$\dot{\varphi}^2 < V(\varphi)\,. \tag{45}$$

The idea of inflation is that φ is initially far from the minimum of $V(\varphi)$. The potential then pulls φ towards the minimum. See Fig. 5. If the potential has a suitable (sufficiently flat) shape, the friction term soon makes $\dot{\varphi}$ small enough to satisfy Eq. (45), even if it was not satisfied initially.

We shall also need the Friedmann equation for the flat universe,

$$H^2 = \frac{8\pi G}{3}\rho = \frac{1}{3M_{\rm Pl}^2}\rho\,. \tag{46}$$

where we have introduced the *reduced Planck mass*

$$M_{\rm Pl} \equiv \frac{1}{\sqrt{8\pi}} m_{\rm Pl} \equiv \frac{1}{\sqrt{8\pi G}} = 2.4353 \times 10^{18}\,{\rm GeV}\,. \tag{47}$$

Inserting Eq. (37), this becomes

$$\boxed{H^2 = \frac{1}{3M_{\rm Pl}^2}\left[\frac{1}{2}\dot{\varphi}^2 + V(\varphi)\right]}\,. \tag{48}$$

We have ignored other components to energy density and pressure besides the inflaton. During inflation, the inflaton φ moves slowly, so that the inflaton energy density, which is dominated by $V(\varphi)$ also changes slowly. If there are matter and radiation components to the energy density, they decrease fast, $\rho \propto a^{-3}$ or $\propto a^{-4}$ and soon become negligible. Again, this puts some initial conditions for inflation to get started, for the inflaton to become dominant. But once inflation gets started, we can soon forget the other components to the universe besides the inflaton.

7.5 Slow-roll inflation

The friction (expansion) term tends to slow down the evolution of φ, so that we may easily reach a situation where:

$$\dot{\varphi}^2 \ll V(\varphi) \tag{49}$$
$$|\ddot{\varphi}| \ll |3H\dot{\varphi}| \tag{50}$$

These are the *slow-roll conditions*.

If the slow-roll conditions are satisfied, we may approximate (the *slow-roll approximation*) Eqs. (48) and (44) by the *slow-roll equations*:

$$H^2 = \frac{V(\varphi)}{3M_{\rm Pl}^2} \tag{51}$$
$$3H\dot{\varphi} = -V'(\varphi) \tag{52}$$

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

$$\varepsilon(\varphi) \equiv \frac{1}{2}M_{\rm Pl}^2\left(\frac{V'}{V}\right)^2$$
$$\eta(\varphi) \equiv M_{\rm Pl}^2\frac{V''}{V} \tag{53}$$

Exercise: Show that

$$\varepsilon \ll 1 \quad \text{and} \quad |\eta| \ll 1 \qquad \Leftarrow \qquad \text{Eqs. (49) and (50)} \tag{54}$$

Note that 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 to be valid (i.e., the slow-roll conditions to be satisfied).

The conditions $\varepsilon \ll 1$ and $|\eta| \ll 1$ are just *conditions on the shape of the potential*, and identify from the potential a *slow-roll section*, where the slow-roll approximation *may* be valid. Since the initial field equation, Eq. (44) was second order, it accepts arbitrary φ and $\dot{\varphi}$ as initial conditions. Thus Eqs. (49) and (50) may not hold initially, even if φ is in the slow-roll section. However, it turns out that the *slow-roll solution*, the solution of the slow-roll equations (51) and (52), is an *attractor* of the full equations, (48) and (44). This means that the solution of the full equations rapidly approaches it, starting from arbitrary initial conditions. Well, not fully arbitrary, the initial conditions need to lie in the *basin of attraction*, from which they are then attracted into the attractor. To be in the basin of attraction, means that φ must be in the slow-roll section, and that if $\dot{\varphi}$ is very large, φ needs to be deeper in the slow-roll section.

Once we have reached the attractor, where Eqs. (51) and (52) hold, $\dot{\varphi}$ is determined by φ (since we replaced the second-order differential equation with a first-order one). In fact everything is determined by φ (assuming a known form of $V(\varphi)$). 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.

This language of "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,

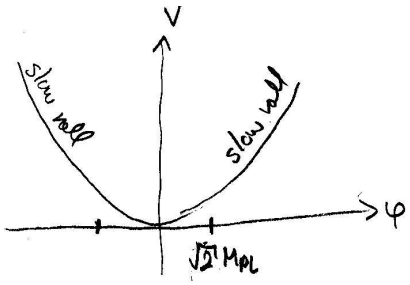


Figure 6: The potential $V(\varphi) = \frac{1}{2}m^2\varphi^2$ and its two slow-roll sections.

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 the 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.

Example: The simplest model of inflation (see Fig. 6) is the one where

$$V(\varphi) = \frac{1}{2}m^2\varphi^2 \quad \Rightarrow \quad V'(\varphi) = m^2\varphi\,, \quad V''(\varphi) = m^2\,. \tag{55}$$

The slow-roll parameters are

$$\left.\begin{aligned} \varepsilon(\varphi) &= \frac{1}{2}M_{\rm Pl}^2\left(\frac{2}{\varphi}\right)^2 \\ \eta(\varphi) &= M_{\rm Pl}^2\frac{2}{\varphi^2} \end{aligned}\right\} \quad \Rightarrow \quad \varepsilon = \eta = 2\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \tag{56}$$

and the slow-roll section of the potential is given by the condition

$$\varepsilon, \eta \ll 1 \quad \Rightarrow \quad \varphi^2 \gg 2M_{\rm Pl}^2\,. \tag{57}$$

7.5.1 Relation between inflation and slow roll

From the definition of the Hubble parameter,

$$H = \frac{\dot{a}}{a} \quad \Rightarrow \quad \dot{H} = \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} \quad \Rightarrow \quad \boxed{\frac{\ddot{a}}{a} = \dot{H} + H^2} \tag{58}$$

Thus the condition for inflation is $\dot{H} + H^2 > 0$. This would be satisfied, if $\dot{H} > 0$, but this is not possible here, since it would require $p < -\rho$, i.e., $w \equiv p/\rho < -1$, which is not allowed by Eq. (37).[8] Thus

$$\boxed{\dot{H} \leq 0} \tag{59}$$

and

$$\text{Inflation} \;\Leftrightarrow\; -\frac{\dot{H}}{H^2} < 1 \tag{60}$$

[8] From the Friedmann eqs.,

$$\left.\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{K}{a^2} \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p) \end{aligned}\right\} \quad \Rightarrow \quad \dot{H} \;=\; \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} \;=\; -4\pi G(\rho + p) + \frac{K}{a^2}$$

In the above, we are assuming space is already flat, i.e., $K = 0$. Then $\dot{H} > 0 \;\Rightarrow\; \rho + p < 0$.

If the slow-roll approximation is valid,

$$H^2 = \frac{V}{3M_{\rm Pl}^2} \quad \Rightarrow \quad 2H\dot{H} = \frac{V'\dot{\varphi}}{3M_{\rm Pl}^2} \quad \Rightarrow \quad H^2\dot{H} = \frac{V'H\dot{\varphi}}{6M_{\rm Pl}^2} \overset{3H\dot{\varphi}=-V'}{=} -\frac{V'^2}{18M_{\rm Pl}^2}$$
$$\Rightarrow \quad -\frac{\dot{H}}{H^2} = \frac{V'^2}{18M_{\rm Pl}^2}\frac{9M_{\rm Pl}^4}{V^2} = \frac{1}{2}M_{\rm Pl}^2\left(\frac{V'}{V}\right)^2 = \varepsilon \ll 1$$

Therefore, *if the slow-roll approximation is valid, inflation is guaranteed*. This is a sufficient, not necessary condition. The above result for slow-roll inflation, $-\dot{H}/H^2 \ll 1$ can also be written as

$$\left|\frac{\dot{H}}{H}\right| \ll \frac{\dot{a}}{a}\,. \tag{61}$$

During slow-roll inflation, the Hubble parameter H changes much more slowly than the scale factor a. For a constant H, the universe expands exponentially, since

$$\frac{\dot{a}}{a} = \frac{d\ln a}{dt} = H = const \quad \Rightarrow \quad \ln\frac{a}{a_1} = H(t-t_1) \quad \Rightarrow \quad a \propto e^{Ht}\,. \tag{62}$$

Thus, in slow-roll inflation, the universe expands "almost exponentially".

Note that accelerated expansion, which is defined to mean that $\ddot{a} > 0$, does not mean that the *expansion rate*, as given by H, would increase. Even during inflation, $\dot{H} < 0$, so the expansion rate decreases. (There may be some ambiguity in what is meant by an increasing/decreasing expansion rate. The Hubble parameter is a better quantity to be called the expansion rate than $\dot{a}$, since the value of the latter depends on the normalization of a_0. With the normalization $a_0 = 1$, $H = \dot{a}$ "today".)

Sometimes it is carelessly said that inflation was a period of very rapid expansion. Rapid compared to what? Certainly the expansion rate was larger than today, or indeed larger than during any period after inflation (since $\dot{H} < 0$ always). But note that in the original Hot Big Bang picture $H \to \infty$ (and also $\dot{a} \to \infty$) as $t \to 0$. When we replace the earliest part of Hot Big Bang with inflation, we replace it with *slower* expansion, H almost constant (and $\dot{a}$ becoming smaller towards earlier times—this is what acceleration means), instead of $H \to \infty$.

It is possible to have inflation without the slow-roll parameters being small (fast-roll inflation), but we will see that slow-roll inflation produces the observed primordial perturbation spectrum naturally (unlike fast-roll inflation).

7.5.2 Models of inflation

A model of inflation[9] consists of

1. a potential $V(\varphi)$
2. a way of ending inflation

There are two ways of ending inflation:

1. Slow-roll approximation is no more valid, as φ approaches the minimum of the potential with $V(\varphi_{\rm min}) = 0$ or very small. For a reasonable approximation we can assume inflation ends, when $\varepsilon(\varphi) = 1$ or $|\eta(\varphi)| = 1$. Denote this value of the inflaton field by $\varphi_{\rm end}$.
2. Extra physics intervenes to end inflation (e.g., *hybrid* inflation). In this case inflation may end while the slow-roll approximation is valid.

[9] There are also models of inflation which are not based on a scalar field.

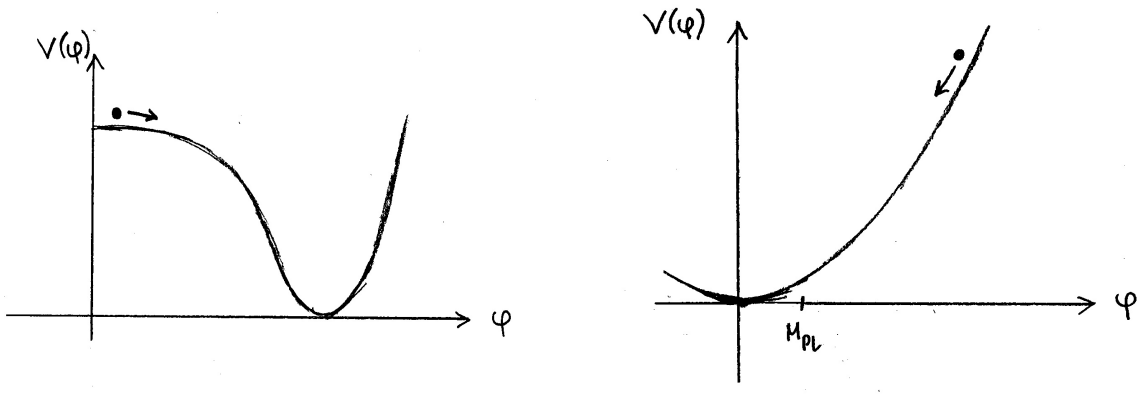


Figure 7: Potential for small-field (a) and large-field (b) inflation. For a typical small-field model, the entire range of φ shown is $\ll M_{\rm Pl}$.

Inflation models can be divided into two classes:

1. small-field inflation, $\Delta\varphi < M_{\rm Pl}$ in the slow-roll section
2. large-field inflation, $\Delta\varphi > M_{\rm Pl}$ in the slow-roll section

Here $\Delta\varphi$ is the range in which φ varies during (the observationally relevant part of) inflation. See Fig. 7 for typical shapes of potentials for large-field and small-field models.

Example of small-field inflation:

$$V(\varphi) = V_0 \left[1 - \frac{\lambda}{4}\left(\frac{\varphi}{M_{\rm Pl}}\right)^4 + \dots\right], \tag{63}$$

where the omitted terms, responsible for keeping $V \geq 0$ for larger φ are assumed negligible in the region of interest. We assume further that the second term is small in the slow-roll section, so that we can approximate $V(\varphi) \approx V_0$ except for its derivatives. The slow-roll parameters are then

$$\varepsilon = \frac{1}{2}\lambda^2 \left(\frac{\varphi}{M_{\rm Pl}}\right)^6 \qquad \text{and} \qquad \eta = -3\lambda \left(\frac{\varphi}{M_{\rm Pl}}\right)^2, \tag{64}$$

so that

$$\frac{\varepsilon}{|\eta|} = \frac{1}{6}\lambda \left(\frac{\varphi}{M_{\rm Pl}}\right)^4 \ll 1\,. \tag{65}$$

Thus $\eta < 0$ and $\varepsilon \ll |\eta|$, which is typical for small-field inflation, and inflation ends when

$$|\eta| = 1 \quad \Rightarrow \quad \varphi_{\rm end} = \frac{M_{\rm Pl}}{\sqrt{3\lambda}}\,. \tag{66}$$

The assumption that the second term in the potential is still small at $\varphi_{\rm end}$, requires that $\lambda \gtrsim 1$, and thus $|\eta| \ll 1$ requires $\varphi \ll M_{\rm Pl}/\sqrt{3}$, so this is indeed a small-field model.

Example of large-field inflation: A simple monomial potential of the form

$$V(\varphi) = A\varphi^n \qquad (n > 1)\,. \tag{67}$$

The slow-roll parameters are

$$\varepsilon = \frac{n^2}{2}\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \qquad \text{and} \qquad \eta = n(n-1)\left(\frac{M_{\rm Pl}}{\varphi}\right)^2, \tag{68}$$

so that $\eta > 0$ and ε and η are of similar size, typical for large-field inflation. This is a large-field model, since $\varepsilon \ll 1$ requires $\varphi^2 \gg \frac{1}{2}n^2 M_{\rm Pl}^2$.

For the special case of $V(\varphi) = \frac{1}{2}m^2\varphi^2$, $\varepsilon = \eta$, and inflation ends at $\varphi_{\rm end} = \sqrt{2}M_{\rm Pl}$. To get inflation to end, e.g., at energy scale $V(\varphi_{\rm end}) \equiv m^2 M_{\rm Pl}^2 = \left(10^{14}\,{\rm GeV}\right)^4$, we need $m = \left(10^{14}\,{\rm GeV}\right)^2/M_{\rm Pl} \approx 4 \times 10^9\,{\rm GeV}$.

7.5.3 Exact solutions

Usually the slow-roll approximation is sufficient. It fails near the end of inflation, but this just affects slightly our estimate of the total amount of inflation. It is much easier to solve the slow-roll equations, (51) and (52), than the full equations, (44) and (48). However, it is useful to have some exact solutions to the full equations, for comparison. For some special cases, exact analytical solutions exist.

One such case is *power-law inflation*, where the potential is

$$V(\varphi) = V_0 \exp\left(-\sqrt{\frac{2}{p}}\frac{\varphi}{M_{\rm Pl}}\right), \qquad p > 1\,, \tag{69}$$

where V_0 and p are constants.

An exact solution for the full equations, (44) and (48), is (**exercise**)

$$a(t) \propto t^p \tag{70}$$

$$\varphi(t) = \sqrt{2p} M_{\rm Pl} \ln\left(\sqrt{\frac{V_0}{p(3p-1)}}\frac{t}{M_{\rm Pl}}\right). \tag{71}$$

The general solution approaches rapidly this particular solution (i.e., it is an attractor). You can see that the expansion, $a(t)$, is power-law, giving the model its name.

The slow-roll parameters for this model are

$$\varepsilon = \frac{1}{2}\eta = \frac{1}{p}\,, \tag{72}$$

independent of φ. In this model inflation never ends, unless other physics intervenes.

7.6 Reheating

During inflation, practically all the energy in the universe is in the inflaton potential $V(\varphi)$, since the slow-roll condition says $\frac{1}{2}\dot\varphi^2 \ll V(\varphi)$. When inflation ends, this energy is transferred in the reheating process to a thermal bath of particles produced in the reheating. Thus reheating creates, from $V(\varphi)$, all the stuff there is in the later universe!

Note that *re*heating may be a misnomer, since we don't know whether the universe was in a thermodynamical equilibrium ever before.

In single-field models of inflation, reheating does not affect the primordial density perturbations,[10] except that it affects the relation of φ_k and k/H_0 given in (85), i.e., how much the distance scale of the perturbations is stretched between inflation and today (these will be discussed later).

Reheating is important for the question of whether unwanted—or wanted—relics are produced after inflation. The reheating temperature must be high enough so that we get standard Big Bang Nucleosynthesis (BBN) after reheating, but sufficiently low so that we do not produce unwanted relics. The latter constraint depends on the extended theory, but it should at least be below the GUT scale. Thus we can take that

$$1\,{\rm MeV} < T_{\rm reh} < 10^{14}\,{\rm GeV}\,. \tag{73}$$

[10] In more complicated models of inflation, involving several fields, reheating may also change the nature of primordial density perturbations.

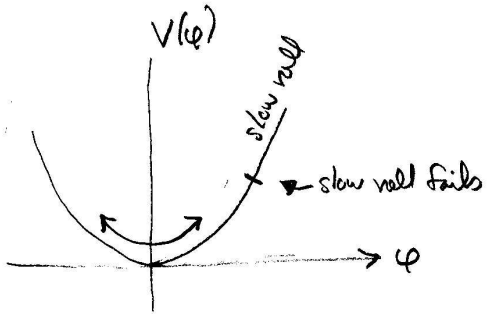


Figure 8: After inflation, the inflaton field is left oscillating at the bottom.

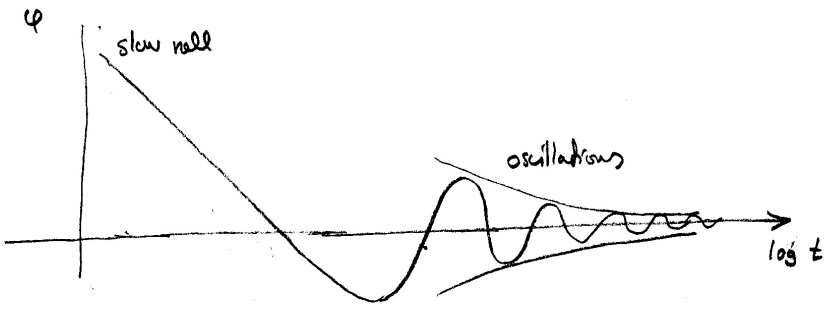


Figure 9: The time evolution of φ as inflation ends.

7.6.1 Scalar field oscillations

After inflation, the inflaton field φ begins to oscillate at the bottom of the potential $V(\varphi)$, see Fig. 8. The inflaton field is still homogeneous, $\varphi(t,\vec{x}) = \varphi(t)$, so it oscillates in the same phase everywhere (we say the oscillation is *coherent*). The expansion time scale H^{-1} soon becomes much longer than the oscillation period.

Assume the potential can be approximated as $\propto \varphi^2$ near the minimum of $V(\varphi)$, so that we have a harmonic oscillator. Write $V(\varphi) = \frac{1}{2}m^2\varphi^2$:

$$\left.\begin{aligned} \ddot{\varphi} + 3H\dot{\varphi} &= -V'(\varphi) \\ \rho &= \frac{1}{2}\dot{\varphi}^2 + V(\varphi) \end{aligned}\right\} \quad \text{become} \quad \begin{cases} \ddot{\varphi} + 3H\dot{\varphi} = -m^2\varphi \\ \rho = \frac{1}{2}\left(\dot{\varphi}^2 + m^2\varphi^2\right) \end{cases}$$

What is $\rho(t)$?

$$\dot{\rho} + 3H\rho \;=\; \dot{\varphi}\overbrace{\left(\ddot{\varphi} + m^2\varphi\right)}^{-3H\dot{\varphi}} + 3H\cdot\frac{1}{2}\left(\dot{\varphi}^2 + m^2\varphi^2\right) \;=\; \frac{3}{2}H\overbrace{\left(m^2\varphi^2 - \dot{\varphi}^2\right)}^{\text{oscillates}}$$

The oscillating factor on the right hand side averages to zero over one oscillation period (in the limit where the period is $\ll H^{-1}$).

Averaging over the oscillations, we get that the long-time behavior of the energy density is

$$\dot{\rho} + 3H\rho = 0 \quad \Rightarrow \quad \rho \propto a^{-3}\,, \tag{74}$$

just like in a matter-dominated universe (we use this result in Sec. 7.7.2). The fall in the energy density shows as a decrease of the oscillation amplitude, see Fig. 9.

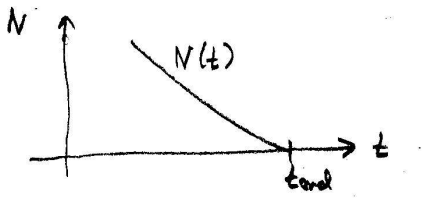


Figure 10: Remaining number of e-foldings $N(t)$ as a function of time.

7.6.2 Inflaton decays

Now that the inflaton field is doing small oscillations around the potential minimum, the particle picture becomes appropriate, and we can consider the energy density ρ_φ to be due to inflaton particles. These inflatons decay into other particles, once the Hubble time ($\sim$ the time after inflation ended) reaches the inflaton decay time.

If the decay is slow (which is the case if the inflaton can only decay into fermions) the inflaton energy density follows the equation

$$\dot{\rho}_\varphi + 3H\rho_\varphi \;=\; -\Gamma_\varphi \rho_\varphi \,, \tag{75}$$

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*. This kind of rapid decay is called *preheating*, since the bosons thus created are far from thermal equilibrium (occupation numbers of states are huge—not possible for fermions).

7.6.3 Thermalization

The particles produced from the inflatons will interact, create other particles through particle reactions, and the resulting particle soup will eventually reach thermal equilibrium with some temperature $T_{\rm reh}$. This *reheating temperature* is determined by the energy density $\rho_{\rm reh}$ at the end of the reheating epoch:

$$\rho_{\rm reh} = \frac{\pi^2}{30} g_*(T_{\rm reh}) T_{\rm reh}^4 \,. \tag{76}$$

Necessarily $\rho_{\rm reh} < \rho_{\rm end}$ (end = end of inflation). If reheating takes a long time, we may have $\rho_{\rm reh} \ll \rho_{\rm end}$. After reheating, we enter the standard Hot Big Bang history of the universe.

7.7 Scales of inflation

7.7.1 Amount 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 ($t_{\rm end}$) by

$$N(t) \equiv \ln \frac{a(t_{\rm end})}{a(t)} \tag{77}$$

See Fig. 10.

As we saw in Sec. 7.5.1, $a(t)$ changes much faster than $H(t)$ (when the slow-roll approximation is valid), so that the comoving Hubble length $\mathcal{H}^{-1} = 1/aH$ shrinks by almost as many e-foldings. ($a(t)$ grows fast, $H(t)$ decreases slowly.)

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:

$$N(\varphi) \equiv \ln \frac{a(t_{\rm end})}{a(t)} = \int_t^{t_{\rm end}} H(t)dt = \int_\varphi^{\varphi_{\rm end}} \frac{H}{\dot\varphi} d\varphi \overset{\text{slow roll}}{\approx} \boxed{\frac{1}{M_{\rm Pl}^2} \int_{\varphi_{\rm end}}^{\varphi} \frac{V}{V'} d\varphi} . \tag{78}$$

where we used

$$d\ln a = \frac{da}{a} = Hdt = H\frac{d\varphi}{\dot\varphi} . \tag{79}$$

Example: For the simple inflation model $V(\varphi) = \frac{1}{2}m^2\varphi^2$,

$$N(\varphi) = \frac{1}{M_{\rm Pl}^2} \int_{\varphi_{\rm end}}^{\varphi} \frac{V}{V'} d\varphi = \frac{1}{M_{\rm Pl}^2} \int_{\varphi_{\rm end}}^{\varphi} \frac{\varphi}{2} = \frac{1}{4M_{\rm Pl}^2} \left(\varphi^2 - {\varphi_{\rm end}}^2\right) = \frac{1}{4}\left(\frac{\varphi}{M_{\rm Pl}}\right)^2 - \frac{1}{2} . \tag{80}$$

The largest initial value of φ we may contemplate is that which gives the Planck density, $V(\varphi) = M_{\rm Pl}^4 \quad \Rightarrow \quad \varphi = \sqrt{2}M_{\rm Pl}^2/m$. Starting from this value we get $\frac{1}{2}\left[(M_{\rm Pl}/m)^2 - 1\right]$ e-foldings of inflation. With $m = 4 \times 10^9\,\text{GeV}$ (see the earlier example with this model), this gives 1.85×10^{17} e-foldings, i.e., expansion by a factor $e^{1.85\times10^{17}} \sim 10^{8\times10^{16}} = 10^{80\,000\,000\,000\,000\,000}$. That's quite a lot!

7.7.2 Evolution of scales

When discussing (next chapter) evolution of density perturbations and formation of structure in the universe, we will be interested in the history of each comoving distance scale, or each *comoving wave number* k (from a Fourier expansion in comoving coordinates).

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

An important question is, whether a distance scale is larger or smaller than the Hubble length at a given time.

We define a scale to be

- superhorizon, when $k < \mathcal{H} \quad \left(k^{-1} > \mathcal{H}^{-1}\right)$
- at horizon (exiting or entering horizon), when $k = \mathcal{H}$
- subhorizon, when $k > \mathcal{H} \quad \left(k^{-1} < \mathcal{H}^{-1}\right)$

Note that *large* scales (large k^{-1}) correspond to *low* k, and *vice versa*, although we often talk about "scale k". This can easily cause confusion, so watch for this, and be careful with wording! To avoid confusion, use the words *high*/*low* instead of large/small for k. Notice also that we are here using the word "horizon" to refer to the Hubble length.[11] Recall that:

$$\text{Inflation} \Rightarrow \mathcal{H}^{-1} \text{ shrinking}$$
$$\text{All other times} \Rightarrow \mathcal{H}^{-1} \text{ growing}$$

See Fig. 11.

[11] As discussed in Cosmology I, there are (at least) three different usages for the word "horizon":

1. particle horizon
2. event horizon (not used in Cosmology I/II)
3. Hubble length

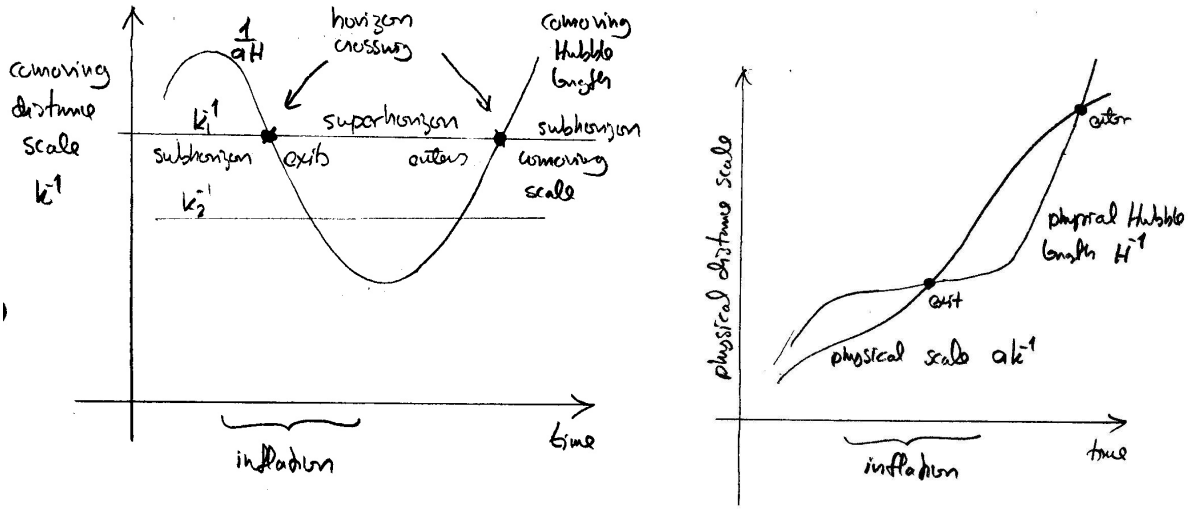


Figure 11: The evolution of the Hubble length, and two scales, k_1^{-1} and k_2^{-1}, seen in comoving coordinates (left) and in terms of physical distance (right).

We shall later find that the amplitude of primordial density perturbations at a given comoving scale is determined when this scale exits the horizon during inflation. The largest observable scales, $k \approx \mathcal{H}_0 = H_0$, are "at horizon" today. (Since the universe has recently began accelerating again, these scales have just barely entered, and are actually 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 horizon during inflation **to** the end of inflation (t_k to $t_{\rm end}$).

2. **From** the end of inflation **to** reheating. We assume (as discussed in Sec. 7.6.1) that the universe behaves as if matter-dominated, $\rho \propto a^{-3}$, during this period ($t_{\rm end}$ to $t_{\rm reh}$).

3. **From** reheating **to** the present time ($t_{\rm reh}$ to t_0).

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

$$\Rightarrow \quad k = \mathcal{H}_k = a_k H_k \,.$$

To find out how large this scale is today, we relate it to the present "horizon", i.e., the Hubble scale:

$$\frac{k}{H_0} \;=\; \frac{a_k H_k}{a_0 H_0} \;=\; \frac{a_k}{a_{\rm end}}\frac{a_{\rm end}}{a_{\rm reh}}\frac{a_{\rm reh}}{a_0}\frac{H_k}{H_0} \;=\; e^{-N(k)}\left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{\frac{1}{3}}\left(\frac{\rho_{r0}}{\rho_{\rm reh}}\right)^{\frac{1}{4}}\left(\frac{\rho_k}{\rho_{\rm cr0}}\right)^{\frac{1}{2}}, \qquad (81)$$

where $\rho_k \approx V(\varphi_k) \equiv V_k$ (since $\frac{1}{2}\dot{\varphi}^2 \ll V(\varphi)$ during slow roll) is the energy density when scale k exited and $N(k) \equiv$ number of e-foldings of inflation after that. Eq. (78) allows us to relate φ_k to $N(k)$. The factor $a_{\rm reh}/a_0 = a_{\rm reh}$ is related to the change in energy density from $t_{\rm reh} \to t_0$. The behavior of the total energy density as a function of a changed from the radiation-dominated to the matter-dominated to the dark-energy-dominated era, but we can keep things simpler by considering just the radiation component ρ_r, which was equal to the total energy density $\rho_{\rm reh}$ at end of reheating and behaves after that as $\rho_r \propto a^{-4}$. This is slightly inaccurate, since $\rho_r \propto a^{-4}$

does not take into account the change in g_*. However, the $\propto a^{-4}$ approximation is good enough[12] for us—we are making other comparable approximations also. From end of inflation to reheating

$$\rho \propto a^{-3} \quad \Rightarrow \quad \frac{a_{\rm end}}{a_{\rm reh}} = \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{\frac{1}{3}} ,$$

and the ratio H_k/H_0 we got from

$$H_k = \sqrt{\frac{8\pi G}{3}\rho_k}\,,\ H_0 = \sqrt{\frac{8\pi G}{3}\rho_{\rm cr0}} \quad \Rightarrow \quad \frac{H_k}{H_0} = \left(\frac{\rho_k}{\rho_{\rm cr0}}\right)^{\frac{1}{2}} .$$

Thus we get that

$$\frac{k}{H_0} = e^{-N(k)} \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{1/12} \left(\frac{V_k}{\rho_{\rm end}}\right)^{1/4} \frac{V_k^{1/4}\rho_{r0}^{1/4}}{\rho_{\rm cr0}^{1/2}} .$$

We can now relate $N(k)$ to k/H_0 as

$$N(k) \;=\; -\ln\frac{k}{H_0} - \frac{1}{3}\ln\frac{\rho_{\rm end}^{1/4}}{\rho_{\rm reh}^{1/4}} + \ln\frac{V_k^{1/4}}{\rho_{\rm end}^{1/4}} + \ln\frac{V_k^{1/4}}{10^{16}\,{\rm GeV}} + \ln\frac{10^{16}\,{\rm GeV}\cdot\rho_{r0}^{1/4}}{\rho_{\rm cr0}^{1/2}}\,, \tag{84}$$

where $10^{16}\,{\rm GeV}$ serves as a reference scale for V_k. This is roughly an upper limit to V_k due to lack of observation of primordial gravitational waves (discussed in the next chapter). Sticking in the known values of $\rho_{r0}^{1/4} = 2.4\times10^{-13}\,{\rm GeV}$ (assuming massless neutrinos; however, neutrino masses will not change the result for $N(\varphi_k)$) and $\rho_{\rm cr0}^{1/4} = 3.000\times10^{-12}\,{\rm GeV}\cdot h^{1/2}$, the last term becomes $60.85 - \ln h \approx 61$.

The final result is

$$N(\varphi_k) \;=\; -\ln\frac{k}{H_0} + 61 + \ln\frac{V_k^{1/4}}{\rho_{\rm end}^{1/4}} - \frac{1}{3}\ln\frac{\rho_{\rm end}^{1/4}}{\rho_{\rm reh}^{1/4}} - \ln\frac{10^{16}\,{\rm GeV}}{V_k^{1/4}}\,, \tag{85}$$

where the terms have been arranged so that they are all positive (when the sign in front of them is not included). Since the potential V_k changes slowly during slow roll, the k-dependence is dominated by the first term and the third term is small. The fourth term depends on how fast the reheating was. If it was instantaneous, this term is zero. The last term can be large if the inflation scale is much lower than $10^{16}\,{\rm GeV}$.

[12] Accurately this would go as:

$$g_{*s}a^3T^3 = \text{const.} \quad \Rightarrow \quad \frac{a_{\rm reh}}{a_0} = \left[\frac{g_{*s}(T_0)}{g_{*s}(T_{\rm reh})}\right]^{\frac{1}{3}} \frac{T_0}{T_{\rm reh}} \tag{82}$$

Eq. (81) approximates this with

$$\left(\frac{\rho_{r0}}{\rho_{\rm reh}}\right)^{\frac{1}{4}} = \left[\frac{g_*(T_0)}{g_*(T_{\rm reh})}\right]^{\frac{1}{4}} \frac{T_0}{T_{\rm reh}} \tag{83}$$

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

$$\frac{(82)}{(83)} = \frac{g_{*s}(T_0)^{\frac{1}{3}}}{g_*(T_0)^{\frac{1}{4}} g_*(T_{\rm reh})^{\frac{1}{12}}} \approx \frac{3.909^{\frac{1}{3}}}{3.363^{\frac{1}{4}}100^{\frac{1}{12}}} = 0.79 \sim 1$$

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

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

For any given present scale, given as a fraction of the present Hubble distance,[13] Eq. (85) identifies the value φ_k the inflaton had, when this scale exited the horizon 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. Usually, the precise value of N is not that important; we are more interested in the derivative dN/dk, or rather $d\varphi_k/dk$. We can see that typically (for high-energy-scale inflation) about 60 e-foldings of inflation occur *after* the largest observable scales exit the horizon. The number of e-foldings *before* that can be very large (e.g., billions or much more), depending on the inflation model and how the inflation is assumed to begin.

7.8 Initial conditions for inflation

Inflation provides the initial conditions for the Hot Big Bang. What about initial conditions for inflation? As we discussed earlier, inflation erases all memory of these initial conditions, removing this question from the reach of observational verification. However, a complete picture of the history of the universe should also include some idea about the conditions before inflation. To weigh how plausible inflation is as an explanation we may contemplate how easy it is for the universe to begin inflating.

Although inflation differs radically from the other periods of the history of the universe we have discussed, two qualitative features still hold true also during inflation: 1) the universe is expanding and 2) the energy density is decreasing[14] (although slowly during inflation).

Thus the energy density should be higher before inflation than during it or after it. Often it is assumed that inflation begins right at the Planck scale, $\rho \sim M_{\rm Pl}^4$, which is the limit to how high energy densities we can extend our discussion, which is based on classical GR. Consider one such scenario:

When $\rho > M_{\rm Pl}^4$, quantum gravitational effects should be important. We can imagine that the universe at that time, the *Planck era*, is some kind of "spacetime foam", where the spacetime itself is subject to large quantum fluctuations. When the energy density of some region, larger than H^{-1}, falls below $M_{\rm Pl}^4$, spacetime in that region begins to behave in a classical manner. See Fig. 12. The initial conditions, i.e, conditions at the time when "our universe" (referring to one such region) emerges from the spacetime foam, are usually assumed *chaotic* (term due to Linde [3], does not refer to chaos theory), i.e., φ takes different, random, values at different regions. Since $\rho \geq \rho_\varphi$, and

$$\rho_\varphi = \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}\nabla\varphi^2 + V(\varphi)\,, \tag{86}$$

we must have

$$\dot{\varphi}^2 \lesssim M_{\rm Pl}^4\,, \quad \nabla\varphi^2 \lesssim M_{\rm Pl}^4\,, \quad V(\varphi) \lesssim M_{\rm Pl}^4 \tag{87}$$

in a region for it to emerge from the spacetime foam. If the conditions are suitable such a region may then begin to inflate. Thus inflation may begin at many different parts of the spacetime foam. Our observable universe would be just a small part of one such region which has inflated to a huge size.

It is also possible that during inflation, for some part of the potential, quantum fluctuations of the inflaton (not of the spacetime) dominate over the classical evolution, pushing 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. (This possibility depends on the shape of the potential.) But our observable universe would be a

[13] For example, $k/H_0 = 10$ means that we are talking about a scale corresponding to a wavelength λ, where $\lambda/2\pi$ is one tenth of the Hubble distance.

[14] Not necessarily true in all cases, e.g., eternal inflation.

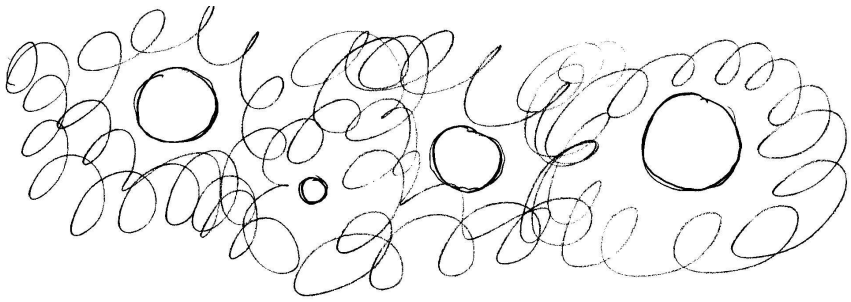


Figure 12: Spacetime foam and some regions emerging from it.

part of a region, where φ came down to a region of the potential where the quantum fluctuations of φ were small and the classical behavior began to dominate and eventually inflation ended.

Thus we see that the very, very, very large scale structure of the universe may be very complicated. But we will never discover this, since our entire observable universe is just a small homogeneous part of a patch which inflated, and then the inflation ended in that patch. All the observable features of the Universe can be explained in terms of what happened in this patch during and after inflation.

These ideas of spacetime foam and eternal inflation are rather speculative and there are also other suggestions for the initial stages of the universe.

7.9 Inflation and Big Bang

There is some confusion about what exactly is meant by the term "Big Bang". Initially the term was introduced to refer to the beginning of the Universe (specifically, Hoyle compared "the hypothesis that all matter of the universe was created in one big bang at a particular time in the remote past"[4] to his own steady-state model of the universe which required a continuous creation of matter). Matter- and/or radiation-dominated FRW models have a beginning where $a = 0$, and we can set this to correspond to the origin ($t = 0$) of the time coordinate, so it became customary to refer to this $t = 0$ as the Big Bang. In classical GR this is a singularity (the spacetime curvature becomes infinite) with infinite energy density. It is clear that classical GR must break down near this singularity, so instead of singularity, there should be something else, which is unknown. Moreover, we have well-tested physics theories only below the energy scale of the electroweak transition. This fuzziness about the very beginning was not an obstacle to calling it a Big Bang and thinking that it referred to a specific time, for as long as it was thought that the expansion was slowing down, because that made the time scale of the earlier events shorter, so that the "fuzzy part" was thought to be an extremely short time compared to the times for which there was rigorous scientific discussion.

Inflation changed this. In inflation the expansion is accelerating and the earlier times are no longer negligibly short compared to the time scales under discussion (e.g., of the late part of inflation during which the observable scales exit the horizon). In fact, we have no upper limit on how long inflation may have lasted, and no way to know what happened before inflation. Also, the conditions during inflation are very different from the hot dense particle soup that was thought to be created in the Big Bang. When Guth proposed inflation, he had in mind the earlier idea of Big Bang, where at first the temperature and density were extremely high so that cooling lead to a GUT transition giving rise to inflation. Thus he still had a "Big Bang" before inflation. This order of events has persisted in many popular presentations of the early universe. However, the modern view is that this is not correct. We have no knowledge of what happened before inflation, no evidence of an initial singularity or a Planck epoch, and thus we cannot place with confidence anything before inflation. If we go back to Hoyle's original phrase about

creation of all matter, then in the inflation scenario the "Big Bang" would be the reheating. On the other hand, we cannot be certain that inflation really happened, it is just a favorite scenario, not something proved beyond doubt. The clear evidence about "Big Bang" is the abundances of light element isotopes, which tell us about big bang nucleosynthesis, and the cosmic microwave background, which tell us about recombination and photon decoupling. When we say that we know that there was a Big Bang in the early universe, we really mean that we know that these events took place, in a way that we described in Cosmology I. Thus my personal favorite for the modern meaning of "Big Bang" is that it refers, not to a specific instant of time, but to this epoch in the early universe, ending at photon decoupling, when the universe was filled with an almost homogeneous hot soup of interacting particles. Not all cosmologists use the term this way. But in any case, within modern cosmology, "Big Bang" can sensibly refer only to something after inflation, not before it.

References

[1] A.H. Guth, *Inflationary universe: A possible solution to the horizon and flatness problems*, Phys. Rev. D **23**, 347 (1981)

[2] A.R. Liddle and D.H. Lyth: Cosmological Inflation and Large-Scale Structure (Cambridge University Press 2000)

[3] A.D. Linde, *Chaotic Inflation*, Phys. Lett. **129B**, 177 (1983)

[4] H. Kragh, *Big Bang: the etymology of a name*, Astronomy & Geophysics, **54**, Issue 2, p 2.28–30 (2013), `https://doi.org/10.1093/astrogeo/att035`