2 Inflation

The FRW cosmology described in the previous chapter is incomplete. It doesn't explain why the universe is homogeneous and isotropic on large scales. In fact, the standard cosmology predicts that the early universe was made of many causally disconnected regions of space. The fact that these apparently disjoint patches of space have very nearly the same densities and temperatures is called the *horizon problem*. In this chapter, I will explain how inflation—an early period of accelerated expansion—drives the primordial universe towards homogeneity and isotropy, even if it starts in a more generic initial state.

Throughout this chapter, we will trade Newton's constant for the (reduced) Planck mass,

$$M_{\rm pl} \equiv \sqrt{\frac{\hbar c}{8\pi G}} = 2.4 \times 10^{18}\,{\rm GeV}\ ,$$

so that the Friedmann equation (1.3.131) is written as $H^2 = \rho/(3M_{\rm pl}^2)$.

2.1 The Horizon Problem

2.1.1 Light and Horizons

The size of a causal patch of space it determined by how far light can travel in a certain amount of time. As we mentioned in §1.1.3, in an expanding spacetime the propagation of light (photons) is best studied using conformal time. Since the spacetime is isotropic, we can always define the coordinate system so that the light travels purely in the radial direction (i.e. $\theta = \phi = const.$). The evolution is then determined by a two-dimensional line element[1]

$$\mathrm{d}s^2 = a^2(\tau)\left[\mathrm{d}\tau^2 - \mathrm{d}\chi^2\right]\ . \tag{2.1.1}$$

Since photons travel along null geodesics, $\mathrm{d}s^2 = 0$, their path is defined by

$$\Delta\chi(\tau) = \pm\Delta\tau\ , \tag{2.1.2}$$

where the plus sign corresponds to outgoing photons and the minus sign to incoming photons. This shows the main benefit of working with conformal time: light rays correspond to straight lines at 45° angles in the χ-τ coordinates. If instead we had used physical time t, then the light cones for curved spacetimes would be curved. With these preliminaries, we now define two different types of cosmological horizons. One which limits the distances at which past events can be observed and one which limits the distances at which it will ever be possible to observe future events.

[1] For the radial coordinate χ we have used the parameterisation of (1.1.23), so that (2.1.1) is conformal to two-dimensional Minkowski space and the curvature k of the three-dimensional spatial slices is absorbed into the definition of the coordinate χ. Had we used the regular polar coordinate r, the two-dimensional line element would have retained a dependence on k. For flat slices, χ and r are of course the same.

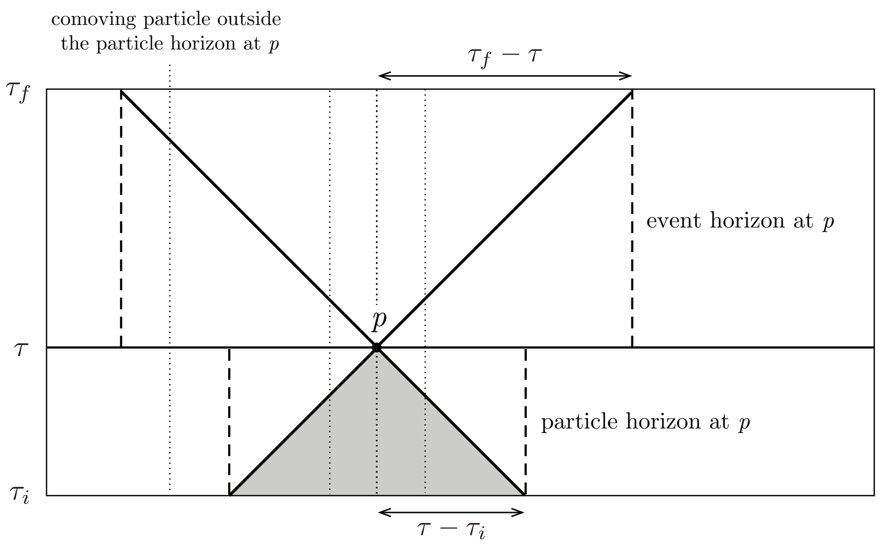

Figure 2.1: *Spacetime diagram illustrating the concept of horizons. Dotted lines show the worldlines of comoving objects. The event horizon is the maximal distance to which we can send signal. The particle horizon is the maximal distance from which we can receive signals.*

- *Particle horizon.*—Eq. (2.1.2) tells us that the maximal comoving distance that light can travel between two times τ_1 and $\tau_2 > \tau_1$ is simply $\Delta\tau = \tau_2 - \tau_1$ (recall that $c \equiv 1$). Hence, if the Big Bang 'started' with the singularity at $t_i \equiv 0$,[2] then the greatest comoving distance from which an observer at time t will be able to receive signals travelling at the speed of light is given by
$$\chi_{\rm ph}(\tau) = \tau - \tau_i = \int_{t_i}^{t} \frac{{\rm d}t}{a(t)} \ . \qquad (2.1.3)$$
This is called the (comoving) particle horizon. The size of the particle horizon at time τ may be visualised by the intersection of the past light cone of an observer p with the spacelike surface $\tau = \tau_i$ (see fig. 2.1). Causal influences have to come from within this region. Only comoving particles whose worldlines intersect the past light cone of p can send a signal to an observer at p. The boundary of the region containing such worldlines is the particle horizon at p. Notice that every observer has his of her own particle horizon.

- *Event horizon.*—Just as there are past events that we cannot see now, there may be future events that we will never be able to see (and distant regions that we will never be able to influence). In comoving coordinates, the greatest distance from which an observer at time t_f will receive signals emitted at any time later than t is given by
$$\chi_{\rm eh}(\tau) = \tau_f - \tau = \int_{t}^{t_f} \frac{{\rm d}t}{a(t)} \ . \qquad (2.1.4)$$
This is called the (comoving) event horizon. It is similar to the event horizon of black holes. Here, τ_f denotes the 'final moment of (conformal) time'. Notice that this may be finite even if physical time is infinite, $t_f = +\infty$. Whether this is the case or not depends on the form of $a(t)$. In particular, τ_f is finite for our universe, if dark energy is really a cosmological constant.

[2]Notice that the Big Bang singularity is a *moment in time*, but **not** *a point space*. Indeed, in figs. 2.1 and 2.2 we describe the singularity by an extended (possibly infinite) spacelike hypersurface.

2.1.2 The Growing Hubble Sphere

It is the particle horizon that is relevant for the horizon problem of the standard Big Bang cosmology. Eq. (2.1.3) can be written in the following illuminating way

$$\chi_{\rm ph}(\tau) = \int_{t_i}^{t} \frac{{\rm d}t}{a} = \int_{a_i}^{a} \frac{{\rm d}a}{a\dot{a}} = \int_{\ln a_i}^{\ln a} (aH)^{-1}\, {\rm d}\ln a\ , \tag{2.1.5}$$

where $a_i \equiv 0$ corresponds to the Big Bang singularity. The causal structure of the spacetime can hence be related to the evolution of the *comoving Hubble radius* $(aH)^{-1}$. For a universe dominated by a fluid with constant equation of state $w \equiv P/\rho$, we get

$$(aH)^{-1} = H_0^{-1}\, a^{\frac{1}{2}(1+3w)}\ . \tag{2.1.6}$$

Note the dependence of the exponent on the combination $(1+3w)$. All familiar matter sources satisfy the strong energy condition (SEC), $1+3w>0$, so it used to be a standard assumption that the comoving Hubble radius increases as the universe expands. In this case, the integral in (2.1.5) is dominated by the upper limit and receives vanishing contributions from early times. We see this explicitly in the example of a perfect fluid. Using (2.1.6) in (2.1.5), we find

$$\chi_{\rm ph}(a) = \frac{2H_0^{-1}}{(1+3w)}\left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)}\right] \equiv \tau - \tau_i\ . \tag{2.1.7}$$

The fact that the comoving horizon receives its largest contribution from late times can be made manifest by defining

$$\tau_i \equiv \frac{2H_0^{-1}}{(1+3w)}\, a_i^{\frac{1}{2}(1+3w)} \xrightarrow{\ a_i\to 0\,,\, w>-\frac{1}{3}\ } 0\ . \tag{2.1.8}$$

The comoving horizon is finite,

$$\chi_{\rm ph}(t) = \frac{2H_0^{-1}}{(1+3w)}\, a(t)^{\frac{1}{2}(1+3w)} = \frac{2}{(1+3w)}\,(aH)^{-1}\ . \tag{2.1.9}$$

We see that in the standard cosmology $\chi_{\rm ph} \sim (aH)^{-1}$. This has lead to the confusing practice of referring to both the particle horizon and the Hubble radius as the "horizon" (see §2.2.2).

2.1.3 Why is the CMB so uniform?

About 380 000 years after the Big Bang, the universe had cooled enough to allow the formation of hydrogen atoms and the decoupling of photons from the primordial plasma (see §3.3.3). We observe this event in the form of the cosmic microwave background (CMB), an afterglow of the hot Big Bang (see Chapter 7). Remarkably, this radiation is almost perfectly isotropic (see fig. 7.2), with anisotropies in the CMB temperature being smaller than one part in ten thousand.

A moment's thought will convince you that the finiteness of the conformal time elapsed between $t_i = 0$ and the time of the formation of the CMB, t_{rec}, implies a serious problem: it means that most spots in the CMB have non-overlapping past light cones and hence never were in causal contact. This is illustrated by the spacetime diagram in fig. 2.2. Consider two opposite directions on the sky. The CMB photons that we receive from these directions were emitted at the points labelled p and q in fig. 2.2. We see that the photons were emitted sufficiently close to the Big Bang singularity that the past light cones of p and q don't overlap. This implies that no point lies inside the particle horizons of both p and q. So the puzzle is: how do the photons coming from p and q "know" that they should be at almost exactly the same temperature? The same question applies to any two points in the CMB that are separated by more than 1 degree in the sky. The homogeneity of the CMB spans scales that are much larger than the particle horizon at the time when the CMB was formed. In fact, in the standard cosmology the CMB is made of about 10^4 disconnected patches of space. If there wasn't enough time for these regions to communicate, why do they look so similar? This is the *horizon problem*.

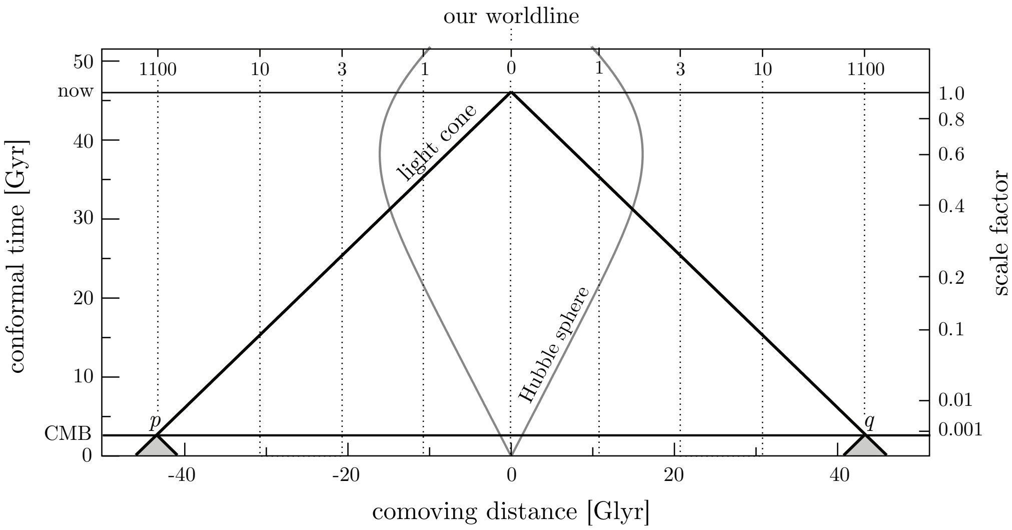

Figure 2.2: *The horizon problem in the conventional Big Bang model. All events that we currently observe are on our past light cone. The intersection of our past light cone with the spacelike slice labelled CMB corresponds to two opposite points in the observed CMB. Their past light cones don't overlap before they hit the singularity, $a = 0$, so the points appear never to have been in causal contact. The same applies to any two points in the CMB that are separated by more than 1 degree on the sky.*

2.2 A Shrinking Hubble Sphere

Our description of the horizon problem has highlighted the fundamental role played by the growing Hubble sphere of the standard Big Bang cosmology. A simple solution to the horizon problem therefore suggests itself: let us conjecture a phase of *decreasing Hubble radius* in the early universe,

$$\frac{d}{dt}(aH)^{-1} \ < \ 0 \ . \tag{2.2.10}$$

If this lasts long enough, the horizon problem can be avoided. Physically, the shrinking Hubble sphere requires a SEC-violating fluid, $1 + 3w < 0$.

2.2.1 Solution of the Horizon Problem

For a shrinking Hubble sphere, the integral in (2.1.5) is dominated by the lower limit. The Big Bang singularity is now pushed to *negative conformal time*,

$$\tau_i = \frac{2H_0^{-1}}{(1+3w)}\, a_i^{\frac{1}{2}(1+3w)} \ \xrightarrow{a_i \to 0\,,\, w < -\frac{1}{3}} \ -\infty \ . \tag{2.2.11}$$

This implies that there was "much more conformal time between the singularity and decoupling than we had thought"! Fig. 2.3 shows the new spacetime diagram. The past light cones of

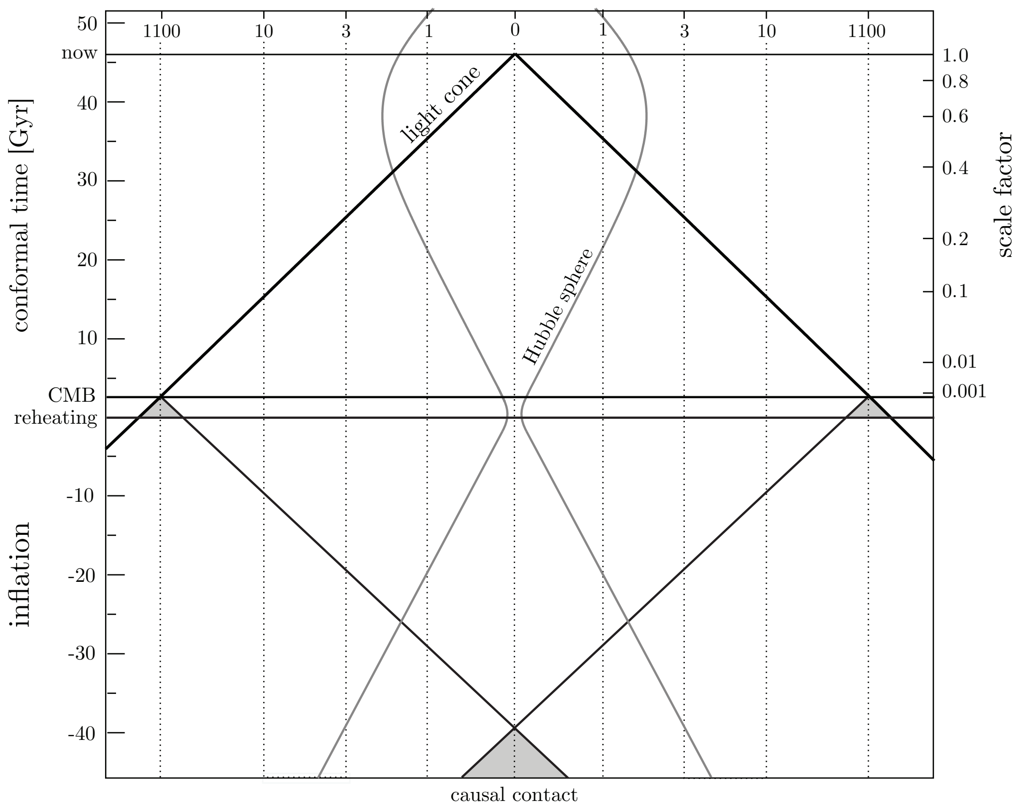

Figure 2.3: *Inflationary solution to the horizon problem. The comoving Hubble sphere shrinks during inflation and expands during the conventional Big Bang evolution (at least until dark energy takes over at $a \approx 0.5$). Conformal time during inflation is negative. The spacelike singularity of the standard Big Bang is replaced by the reheating surface, i.e. rather than marking the beginning of time it now corresponds simply to the transition from inflation to the standard Big Bang evolution. All points in the CMB have overlapping past light cones and therefore originated from a causally connected region of space.*

widely separated points in the CMB now had enough time to intersect before the time τ_i. The uniformity of the CMB is not a mystery anymore. In inflationary cosmology, $\tau = 0$ isn't the initial singularity, but instead becomes only a transition point between inflation and the standard Big Bang evolution. There is time both before and after $\tau = 0$.

2.2.2 Hubble Radius vs. Particle Horizon

A quick word of warning about bad (but unfortunately standard) language in the inflationary literature: Both the particle horizon $\chi_{\rm ph}$ and the Hubble radius $(aH)^{-1}$ are often referred to simply as the "horizon". In the standard FRW evolution (with ordinary matter) the two are roughly the same—cf. eq. (2.1.9)—so giving them the same name isn't an issue. However, the whole point of inflation is to make the particle horizon much larger than the Hubble radius.

The Hubble radius $(aH)^{-1}$ is the (comoving) distance over which particles can travel in the course of one expansion time.[3] It is therefore another way of measuring whether particles are causally connected with each other: comparing the comoving separation λ of two particles with $(aH)^{-1}$ determines whether the particles can communicate with each other *at a given moment* (i.e. within the next Hubble time). This makes it clear that $\chi_{\rm ph}$ and $(aH)^{-1}$ are conceptually very different:

[3]The expansion time, $t_H \equiv H^{-1} = dt/d\ln a$, is roughly the time in which the scale factor doubles.

- if $\lambda > \chi_{\rm ph}$, then the particles could *never* have communicated.
- if $\lambda > (aH)^{-1}$, then the particles cannot talk to each other *now*.

Inflation is a mechanism to achieve $\chi_{\rm ph} \gg (aH)^{-1}$. This means that particles can't communicate now (or when the CMB was created), but were in causal contact early on. In particular, the shrinking Hubble sphere means that particles which were initially in causal contact with another—i.e. separated by a distance $\lambda < (a_I H_I)^{-1}$—can no longer communicate after a sufficiently long period of inflation: $\lambda > (aH)^{-1}$; see fig. 2.4. However, at any moment before horizon exit (careful: I really mean exit of the Hubble radius!) the particles could still talk to each other and establish similar conditions. Everything within the Hubble sphere at the beginning of inflation, $(a_I H_I)^{-1}$, was causally connected.

Since the Hubble radius is easier to calculate than the particle horizon it is common to use the Hubble radius as a means of judging the horizon problem. If the entire observable universe was within the comoving Hubble radius at the beginning of inflation—i.e. $(a_I H_I)^{-1}$ was larger than the comoving radius of the observable universe $(a_0 H_0)^{-1}$—then there is no horizon problem. Notice that this is more conservative than using the particle horizon since $\chi_{\rm ph}(t)$ is always bigger than $(aH)^{-1}(t)$. Moreover, using $(a_I H_I)^{-1}$ as a measure of the horizon problem means that we don't have to assume anything about earlier times $t < t_I$.

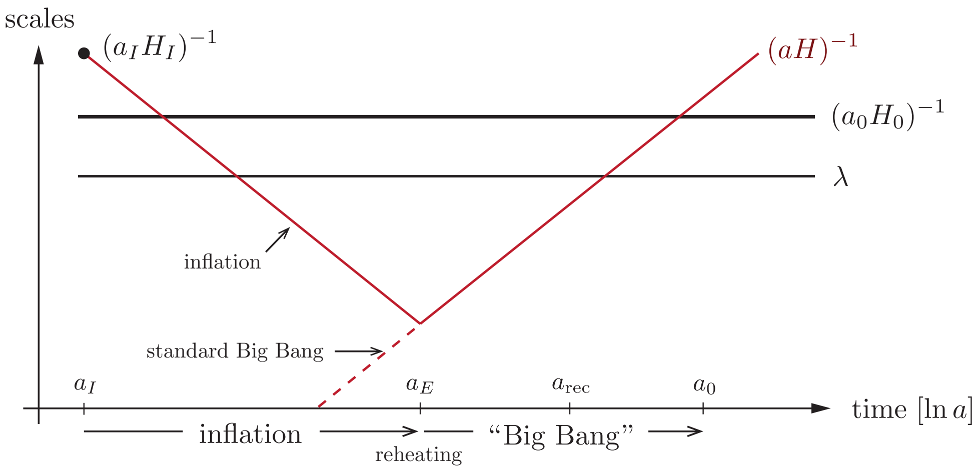

Figure 2.4: *Scales of cosmological interest were larger than the Hubble radius until $a \approx 10^{-5}$ (where today is at $a(t_0) \equiv 1$). However, at very early times, before inflation operated, all scales of interest were smaller than the Hubble radius and therefore susceptible to microphysical processing. Similarly, at very late times, the scales of cosmological interest are back within the Hubble radius. Notice the symmetry of the inflationary solution. Scales just entering the horizon today, 60 e-folds after the end of inflation, left the horizon 60 e-folds before the end of inflation.*

Duration of inflation.—How much inflation do we need to solve the horizon problem? At the very least, we require that the observable universe today fits in the comoving Hubble radius at the beginning of inflation,

$$(a_0 H_0)^{-1} < (a_I H_I)^{-1} \ . \tag{2.2.12}$$

Let us assume that the universe was radiation dominated since the end of inflation and ignore the relatively recent matter- and dark energy-dominated epochs. Remembering that $H \propto a^{-2}$ during radiation domination, we have

$$\frac{a_0 H_0}{a_E H_E} \sim \frac{a_0}{a_E}\left(\frac{a_E}{a_0}\right)^2 = \frac{a_E}{a_0} \sim \frac{T_0}{T_E} \sim 10^{-28} \ , \tag{2.2.13}$$

where in the numerical estimate we used $T_E \sim 10^{15}$ GeV and $T_0 = 10^{-3}$ eV (~ 2.7 K). Eq. (2.2.12) can then be written as

$$(a_I H_I)^{-1} > (a_0 H_0)^{-1} \sim 10^{28} (a_E H_E)^{-1} \ . \tag{2.2.14}$$

For inflation to solve the horizon problem, $(aH)^{-1}$ should therefore shrink by a factor of 10^{28}. The most common way to arrange this it to have $H \sim const.$ during inflation (see below). This implies $H_I \approx H_E$, so eq. (2.2.14) becomes

$$\frac{a_E}{a_I} > 10^{28} \quad \Rightarrow \quad \ln\left(\frac{a_E}{a_I}\right) > 64 \ . \tag{2.2.15}$$

This is the famous statement that the solution of the horizon problem requires about 60 e-folds of inflation.

2.2.3 Conditions for Inflation

I like the shrinking Hubble sphere as the fundamental definition of inflation since it relates most directly to the horizon problem and is also key for the inflationary mechanism of generating fluctuations (see Chapter 5). However, before we move on to discuss what physics can lead to a shrinking Hubble sphere, let me show you that this definition of inflation is equivalent to other popular ways of describing inflation.

- *Accelerated expansion.*—From the relation

$$\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}(\dot{a})^{-1} = -\frac{\ddot{a}}{(\dot{a})^2} \ , \tag{2.2.16}$$

  we see that a shrinking comoving Hubble radius implies accelerated expansion

$$\ddot{a} > 0 \ . \tag{2.2.17}$$

  This explains why inflation is often defined as a period of acceleration.

- *Slowly-varying Hubble parameter.*—Alternatively, we may write

$$\frac{d}{dt}(aH)^{-1} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1-\varepsilon) \ , \qquad \text{where} \quad \varepsilon \equiv -\frac{\dot{H}}{H^2} \ . \tag{2.2.18}$$

  The shrinking Hubble sphere therefore also corresponds to

$$\varepsilon = -\frac{\dot{H}}{H^2} < 1 \ . \tag{2.2.19}$$

- *Quasi-de Sitter expansion.*—For perfect inflation, $\varepsilon = 0$, the spacetime becomes de Sitter space

$$\mathrm{d}s^2 = \mathrm{d}t^2 - e^{2Ht}\mathrm{d}\boldsymbol{x}^2 \ , \tag{2.2.20}$$

  where $H = \partial_t \ln a = const.$ Inflation has to end, so it shouldn't correspond to perfect de Sitter space. However, for small, but finite $\varepsilon \neq 0$, the line element (2.2.20) is still a good approximation to the inflationary background. This is why we will often refer to inflation as a quasi-de Sitter period.

- *Negative pressure.*—What forms of stress-energy source accelerated expansion? Let us consider a perfect fluid with pressure P and density ρ. The Friedmann equation, $H^2 = \rho/(3M_{\rm pl}^2)$, and the continuity equation, $\dot{\rho} = -3H(\rho + P)$, together imply

$$\dot{H} + H^2 = -\frac{1}{6M_{\rm pl}^2}(\rho + 3P) = -\frac{H^2}{2}\left(1 + \frac{3P}{\rho}\right) . \qquad (2.2.21)$$

We rearrange this to find that

$$\varepsilon = -\frac{\dot{H}}{H^2} = \frac{3}{2}\left(1 + \frac{P}{\rho}\right) < 1 \quad \Leftrightarrow \quad w \equiv \frac{P}{\rho} < -\frac{1}{3} , \qquad (2.2.22)$$

i.e. inflation requires negative pressure or a violation of the strong energy condition. How this can arise in a physical theory will be explained in the next section. We will see that there is nothing sacred about the strong energy condition and that it can easily be violated.

- *Constant density.*—Combining the continuitiy equation, $\dot{\rho} = -3H(\rho+P)$, with eq. (2.2.21), we find

$$\left|\frac{d\ln\rho}{d\ln a}\right| = 2\varepsilon < 1 . \qquad (2.2.23)$$

For small ε, the energy density is therefore nearly constant. Conventional matter sources all dilute with expansion, so we need to look for something more unusual.

2.3 The Physics of Inflation

We have shown that a given FRW spacetime with time-dependent Hubble parameter $H(t)$ corresponds to cosmic acceleration if and only if

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = -\frac{d\ln H}{dN} < 1 . \qquad (2.3.24)$$

Here, we have defined $\mathrm{d}N \equiv \mathrm{d}\ln a = H\mathrm{d}t$, which measures the number of e-folds N of inflationary expansion. Eq. (2.3.24) implies that the fractional change of the Hubble parameter per e-fold is small. Moreover, in order to solve the horizon problem, we want inflation to last for a sufficiently long time (usually at least $N \sim 40$ to 60 e-folds). To achieve this requires ε to remain small for a sufficiently large number of Hubble times. This condition is measured by a second parameter

$$\eta \equiv \frac{d\ln\varepsilon}{dN} = \frac{\dot{\varepsilon}}{H\varepsilon} . \qquad (2.3.25)$$

For $|\eta| < 1$, the fractional change of ε per Hubble time is small and inflation persists. In this section, we discuss what microscopic physics can lead to the conditions $\varepsilon < 1$ and $|\eta| < 1$.

2.3.1 Scalar Field Dynamics

As a simple toy model for inflation we consider a scalar field, the *inflaton* $\phi(t, \boldsymbol{x})$. As indicated by the notation, the value of the field can depend on time t and the position in space $\boldsymbol{x}$. Associated with each field value is a potential energy density $V(\phi)$ (see fig. 2.5). If the field is dynamical (i.e. changes with time) then it also carries kinetic energy density. If the stress-energy associated with the scalar field dominates the universe, it sources the evolution of the FRW background. We want to determine under which conditions this can lead to accelerated expansion.

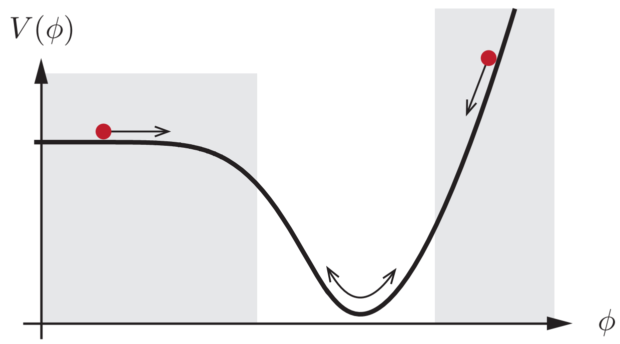

Figure 2.5: *Example of a slow-roll potential. Inflation occurs in the shaded parts of the potential.*

The stress-energy tensor of the scalar field is[4]

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - V(\phi)\right) . \tag{2.3.26}$$

Consistency with the symmetries of the FRW spacetime requires that the background value of the inflaton only depends on time, $\phi = \phi(t)$. From the time-time component $T^0{}_0 = \rho_\phi$, we infer that

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) . \tag{2.3.27}$$

We see that the total energy density, ρ_ϕ, is simply the sum of the kinetic energy density, $\frac{1}{2}\dot{\phi}^2$, and the potential energy density, $V(\phi)$. From the space-space component $T^i{}_j = -P_\phi\,\delta^i_j$, we find that the pressure is the *difference* of kinetic and potential energy densities,

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) . \tag{2.3.28}$$

We see that a field configuration leads to inflation, $P_\phi < -\frac{1}{3}\rho_\phi$, if the potential energy dominates over the kinetic energy.

Next, we look in more detail at the evolution of the inflaton $\phi(t)$ and the FRW scale factor $a(t)$. Substituting ρ_ϕ from (2.3.27) into the *Friedmann equation*, $H^2 = \rho_\phi/(3M_{\rm pl}^2)$, we get

$$\boxed{H^2 = \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\dot{\phi}^2 + V\right]} . \tag{F}$$

Taking a time derivative, we find

$$2H\dot{H} = \frac{1}{3M_{\rm pl}^2}\left[\dot{\phi}\ddot{\phi} + V'\dot{\phi}\right] , \tag{2.3.29}$$

where $V' \equiv dV/d\phi$. Substituting ρ_ϕ and P_ϕ into the second Friedmann equation (2.2.21), $\dot{H} = -(\rho_\phi + P_\phi)/(2M_{\rm pl}^2)$, we get

$$\dot{H} = -\frac{1}{2}\frac{\dot{\phi}^2}{M_{\rm pl}^2} . \tag{2.3.30}$$

Notice that $\dot{H}$ is sourced by the kinetic energy density. Combining (2.3.30) with (2.3.29) leads to the *Klein-Gordon equation*

$$\boxed{\ddot{\phi} + 3H\dot{\phi} + V' = 0} . \tag{KG}$$

[4]You can derive this stress-energy tensor either from Noether's theorem or from the action of a scalar field. You will see those derivations in the QFT course: David Tong, *Part III Quantum Field Theory.*

This is the evolution equation for the scalar field. Notice that the potential acts like a *force*, V', while the expansion of the universe adds *friction*, $H\dot{\phi}$.

2.3.2 Slow-Roll Inflation

Substituting eq. (2.3.30) into the definition of ε, eq. (2.3.24), we find

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2} \ . \tag{2.3.31}$$

Inflation ($\varepsilon < 1$) therefore occurs if the kinetic energy, $\frac{1}{2}\dot{\phi}^2$, only makes a small contribution to the total energy, $\rho_\phi = 3M_{\rm pl}^2 H^2$. This situation is called *slow-roll inflation*.

In order for this condition to persist, the acceleration of the scalar field has to be small. To assess this, it is useful to define the dimensionless acceleration per Hubble time

$$\delta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}} \ . \tag{2.3.32}$$

Taking the time-derivative of (2.3.31),

$$\dot{\varepsilon} = \frac{\dot{\phi}\ddot{\phi}}{M_{\rm pl}^2 H^2} - \frac{\dot{\phi}^2 \dot{H}}{M_{\rm pl}^2 H^3} \ , \tag{2.3.33}$$

and comparing to (2.3.25), we find

$$\eta = \frac{\dot{\varepsilon}}{H\varepsilon} = 2\frac{\ddot{\phi}}{H\dot{\phi}} - 2\frac{\dot{H}}{H^2} = 2(\varepsilon - \delta) \ . \tag{2.3.34}$$

Hence, $\{\varepsilon, |\delta|\} \ll 1$ implies $\{\varepsilon, |\eta|\} \ll 1$.

Slow-roll approximation.—So far, no approximations have been made. We simply noted that in a regime where $\{\varepsilon, |\delta|\} \ll 1$, inflation occurs and persists. We now use these conditions to simplify the equations of motion. This is called the *slow-roll approximation*. The condition $\varepsilon \ll 1$ implies $\frac{1}{2}\dot{\phi}^2 \ll V$ and hence leads to the following simplification of the Friedmann equation (F),

$$\boxed{H^2 \approx \frac{V}{3M_{\rm pl}^2}} \ . \tag{F$_{\rm SR}$}$$

In the slow-roll approximation, the Hubble expansion is determined completely by the potential energy. The condition $|\delta| \ll 1$ simplifies the Klein-Gordon equation (9.2.76) to

$$\boxed{3H\dot{\phi} \approx -V'} \ . \tag{KG$_{\rm SR}$}$$

This provides a simple relationship between the gradient of the potential and the speed of the inflaton. Substituting (F$_{\rm SR}$) and (KG$_{\rm SR}$) into (2.3.31) gives

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2} \approx \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 \ . \tag{2.3.35}$$

Furthermore, taking the time-derivative of (KG$_{\rm SR}$),

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} = -V''\dot{\phi} \ , \tag{2.3.36}$$

leads to

$$\delta + \varepsilon = -\frac{\ddot{\phi}}{H\dot{\phi}} - \frac{\dot{H}}{H^2} \approx M_{\rm pl}^2 \frac{V''}{V} \ . \tag{2.3.37}$$

Hence, a convenient way to assess whether a given potential $V(\phi)$ can lead to slow-roll inflation is to compute the *potential slow-roll parameters*[5]

$$\boxed{\epsilon_{\rm v} \equiv \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2 \quad , \quad |\eta_{\rm v}| \equiv M_{\rm pl}^2 \frac{|V''|}{V}} \ . \tag{2.3.38}$$

Successful slow-roll inflation occurs when these parameters are small, $\{\epsilon_{\rm v}, |\eta_{\rm v}|\} \ll 1$.

Amount of inflation.—The total number of 'e-folds' of accelerated expansion are

$$N_{\rm tot} \equiv \int_{a_I}^{a_E} {\rm d}\ln a = \int_{t_I}^{t_E} H(t)\,{\rm d}t \ , \tag{2.3.39}$$

where t_I and t_E are defined as the times when $\varepsilon(t_I) = \varepsilon(t_E) \equiv 1$. In the slow-roll regime, we can use

$$H{\rm d}t = \frac{H}{\dot{\phi}}{\rm d}\phi = \frac{1}{\sqrt{2\varepsilon}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \approx \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \tag{2.3.40}$$

to write (2.3.39) as an integral in the field space of the inflaton[6]

$$N_{\rm tot} = \int_{\phi_I}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \ , \tag{2.3.41}$$

where ϕ_I and ϕ_E are defined as the boundaries of the interval where $\epsilon_{\rm v} < 1$. The largest scales observed in the CMB are produced about 60 e-folds before the end of inflation

$$N_{\rm cmb} = \int_{\phi_{\rm cmb}}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|{\rm d}\phi|}{M_{\rm pl}} \approx 60 \ . \tag{2.3.42}$$

A successful solution to the horizon problem requires $N_{\rm tot} > N_{\rm cmb}$.

Case study: $m^2\phi^2$ inflation.—As an example, let us give the slow-roll analysis of arguably the simplest model of inflation: single-field inflation driven by a mass term

$$V(\phi) = \frac{1}{2}m^2\phi^2 \ . \tag{2.3.43}$$

The slow-roll parameters are

$$\epsilon_{\rm v}(\phi) = \eta_{\rm v}(\phi) = 2\left(\frac{M_{\rm pl}}{\phi}\right)^2 \ . \tag{2.3.44}$$

To satisfy the slow-roll conditions $\epsilon_{\rm v}, |\eta_{\rm v}| < 1$, we therefore need to consider super-Planckian values for the inflaton

$$\phi > \sqrt{2}M_{\rm pl} \equiv \phi_E \ . \tag{2.3.45}$$

[5]In contrast, the parameters ε and η are often called the *Hubble slow-roll parameters*. During slow-roll, the parameters are related as follows: $\epsilon_{\rm v} \approx \varepsilon$ and $\eta_{\rm v} \approx 2\varepsilon - \frac{1}{2}\eta$.

[6]The absolute value around the integration measure indicates that we pick the overall sign of the integral in such a way as to make $N_{\rm tot} > 0$.

The relation between the inflaton field value and the number of e-folds before the end of inflation is

$$N(\phi) = \frac{\phi^2}{4M_{\rm pl}^2} - \frac{1}{2} \ . \tag{2.3.46}$$

Fluctuations observed in the CMB are created at

$$\phi_{\rm cmb} = 2\sqrt{N_{\rm cmb}}\, M_{\rm pl} \sim 15 M_{\rm pl} \ . \tag{2.3.47}$$

2.3.3 Reheating*

During inflation most of the energy density in the universe is in the form of the inflaton potential $V(\phi)$. Inflation ends when the potential steepens and the inflaton field picks up kinetic energy. The energy in the inflaton sector then has to be transferred to the particles of the Standard Model. This process is called *reheating* and starts the *Hot Big Bang*. We will only have time for a very brief and mostly qualitative description of the absolute basics of the reheating phenomenon.[7] This is non-examinable.

Scalar field oscillations.—After inflation, the inflaton field ϕ begins to oscillate at the bottom of the potential $V(\phi)$, see fig. 2.5. Assume that the potential can be approximated as $V(\phi) = \frac{1}{2}m^2\phi^2$ near the minimum of $V(\phi)$, where the amplitude of ϕ is small. The inflaton is still homogeneous, $\phi(t)$, so its equation of motion is

$$\ddot{\phi} + 3H\dot{\phi} = -m^2\phi \ . \tag{2.3.48}$$

The expansion time scale soon becomes much longer than the oscillation period, $H^{-1} \gg m^{-1}$. We can then neglect the friction term, and the field undergoes oscillations with frequency m. We can write the energy continuity equation as

$$\dot{\rho}_\phi + 3H\rho_\phi = -3HP_\phi = -\frac{3}{2}H(m^2\phi^2 - \dot{\phi}^2) \ . \tag{2.3.49}$$

The r.h.s. averages to zero over one oscillation period. The oscillating field therefore behaves like pressureless matter, with $\rho_\phi \propto a^{-3}$. The fall in the energy density is reflected in a decrease of the oscillation amplitude.

Inflaton decay.—To avoid that the universe ends up empty, the inflaton has to couple to Standard Model fields. The energy stored in the inflaton field will then be transferred into ordinary particles. If the decay is slow (which is the case if the inflaton can only decay into fermions) the inflaton energy density follows the equation

$$\dot{\rho}_\phi + 3H\rho_\phi = -\Gamma_\phi\rho_\phi \ , \tag{2.3.50}$$

where Γ_ϕ parameterizes the inflaton decay rate. If the inflaton can decay into bosons, the decay may be very rapid, involving a mechanism called *parametric resonance* (sourced by Bose condensation effects). This kind of rapid decay is called *preheating*, since the bosons thus created are far from thermal equilibrium.

Thermalisation.—The particles produced by the decay of the inflaton will interact, create other particles through particle reactions, and the resulting particle soup will eventually reach thermal

[7] For more details see Baumann, *The Physics of Inflation*, DAMTP Lecture Notes.

equilibrium with some temperature $T_{\rm rh}$. This reheating temperature is determined by the energy density $\rho_{\rm rh}$ at the end of the reheating epoch. Necessarily, $\rho_{\rm rh} < \rho_{\phi,E}$ (where $\rho_{\phi,E}$ is the inflaton energy density at the end of inflation). If reheating takes a long time, we may have $\rho_{\rm rh} \ll \rho_{\phi,E}$. The evolution of the gas of particles into a thermal state can be quite involved. 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 so 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. After thermalisation of at least the baryons, photons and neutrinos is complete, the standard Hot Big Bang era begins.