Lecture notes of a course presented in the framework of the
"3ième cycle de physique de Suisse romande"

Inflationary cosmology

EPFL, March 2006

Julien Lesgourgues
LAPTH, Annecy-Le-Vieux, France

Contents

1 Introduction

1.1 Basic recalls on the standard cosmological model

In the standard cosmological model and at the level of background quantities (i.e., averaging over spatial fluctuations), the universe is described by the Friedmann-Lemaître-Robertson-Walker metric

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1 - k\, r^2} + r^2 (d\theta^2 + \sin^2\theta d\varphi^2) \right] \tag{1}$$

where k is a constant related to the spatial curvature (positive for closed models, negative for open ones, zero for a flat universe) and t is the proper time measured by a free-falling observer, that we will call *cosmological time*. Throughout this course, we adopt units such that $c = \hbar = 1$.

The Einstein equation relates curvature to matter. In the case of a homogeneous, isotropic universe, described by the Friedmann metric, the Einstein equation yields the Friedmann equation, which relates the total energy density ρ to the space-time curvature: on the one hand, the spatial curvature radius $R_k \equiv a\,|k|^{-1/2}$, and on the other hand the Hubble radius $R_H \equiv H^{-1} = a/\dot{a}$. The Friedmann equation reads

$$\frac{3}{R_H^2} \pm \frac{3}{R_k^2} = 8\pi\mathcal{G}\rho \tag{2}$$

where $\mathcal{G}$ is the Newton constant. The Friedmann equation is more commonly written as

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi\mathcal{G}}{3}\rho\ . \tag{3}$$

The Einstein equations also lead, through Bianchi identities, to the energy conservation equation

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p) \tag{4}$$

which applies to the total cosmological density ρ and pressure p (actually, this conservation equation can be derived by variation of the action for each individual homogeneous component in the universe).

Non-relativistic matter has vanishing pressure and gets diluted according to $\rho_m \propto a^{-3}$, while ultra-relativistic matter has $p_r = \frac{1}{3}\rho_r$ and follows $\rho_r \propto a^{-4}$. These dilution laws can be derived more intuitively by considering a comoving sphere of fixed comobile radius. The number of particles (ultra-relativistic or non-relativistic) is conserved inside the sphere (although non-relativistic particles are still in the comobile coordinate frame, while ultra-relativistic particles flow in and out). The individual energy $E \simeq m$ of a non-relativistic particle is independent of the expansion. Instead the energy of a photon is inversely proportional to its wavelength and is redshifted like a^{-1}. The volume of the sphere scales like a^3. Altogether these arguments lead to the above dilution laws. The curvature of the universe contributes to the expansion in the same way as an *effective curvature density*

$$\rho_k^{eff} \equiv -\frac{3}{8\pi\mathcal{G}}\frac{k}{a^2} \tag{5}$$

which scales like a^{-2}. Finally, the vacuum energy can never dilute, $\dot{\rho}_v = 0$ and $p_v = -\rho_v$. This is valid for the energy of a quantum scalar field in its fundamental state, as well as for a classical field in a state of equilibrium (its energy density density ρ_v is then given by the scalar potential $V(\varphi)$ at the equilibrium point). The vacuum energy is formally equivalent to a cosmological constant Λ which can be added to the Einstein equation without altering the covariance of the theory,

$$G^\nu_\mu + \Lambda\delta^\nu_\mu = 8\pi\mathcal{G}T^\nu_\mu\ , \tag{6}$$

with the identification $\rho_v = -p_v = \Lambda/(8\pi\mathcal{G})$. The critical density is defined for any given value of the Hubble parameter $H = \dot{a}/a$ as the total energy density that would be present in the universe if the spatial curvature was null,

$$H^2 = \frac{8\pi\mathcal{G}}{3}\rho_{crit}\ . \tag{7}$$

With such a definition, the Friedmann equation reads

$$\rho_{crit} - \rho_k^{eff} = \rho_{\rm tot}\ . \tag{8}$$

The contribution of spatial curvature to the expansion of the universe is parametrized by

$$\Omega_k \equiv \frac{\rho_k^{eff}}{\rho_{crit}} = -\frac{k}{(aH)^2} = \frac{R_H^2}{R_k^2} \ . \tag{9}$$

Whenever $|\Omega_k| \ll 1$, the universe can be seen as effectively flat.

Let us summarize very briefly how the cosmological model was built step by step:

- Einstein first proposed a solution for a static universe with $R_H = 0$, based on a non-zero cosmological constant, which was proved later to be unstable. The idea of a static (or even stationary) universe was then abandoned in favor of a nearly homogeneous, isotropic, expanding universe, corresponding in first approximation to the Friedmann-Lemaître-Robertson-Walker metric. This picture was sustained by the discovery of the homogeneous expansion by Hubble in 1929.

- the minimal assumption concerning the composition of the universe is that its energy density is dominated by the constituent of visible objects like stars, planets and inter-galactic gas: namely, non-relativistic matter. This is the Cold Big Bang scenario, in which the Friedmann equation

$$H^2 = \frac{8\pi\mathcal{G}}{3}\rho_{\rm m} \propto a^{-3} \tag{10}$$

  describes the expansion between some initial singularity ($a \rightarrow 0$) and now, caused by non-relativistic, pressureless matter ("cold matter") at a rate $a \propto t^{2/3}$. Gamow, Zel'dovitch, Peebles and others worked on scenarios for Nucleosynthesis in the Cold Big Bang scenario, and concluded that it was ruled out by the fact that the universe contains a significant amount of hydrogen.

- the next level of complexity is to assume that a radiation component (ultra-relativistic photons and neutrinos) dominated the universe expansion at early times. Nucleosynthesis taking place during radiation domination, when $a \propto t^{1/2}$, is in agreement with observations. Structure formation started after the time of equality between radiation and matter, when $\rho_r = \rho_m$. The detection of the Cosmic microwave background by Penzias and Wilson in the lated 60's beautifully confirmed this scenario cold the Hot Big Bang, due to the role of ultra-relativistic matter ("hot matter") after the initial singularity.

- for many years, people wondered whether this scenario should be completed with a recent stage of curvature and/or vacuum domination, starting after most structure have formed. It is now established that the current curvature fraction Ω_k should be very small, so the present universe is exactly or at least nearly flat. In addition, various recent experiments (and in particular supernovae observations) have settled since 1998 that a form of vacuum energy recently took over the matter density. Today, the ratio $\Omega_v \equiv \rho_v/\rho_{crit} \simeq \rho_v/(\rho_v + \rho_m)$ is measured to be close to 0.7.

- finally, some pioneers like Starobinsky and Guth suggested around 1979 that this scenario should be completed with a stage of early vacuum domination taking place much before Nucleosynthesis. After some time, this vacuum would decay mainly into ultra-relativistic particles, and the universe would enter into the radiation dominated phase. The purpose of this course is to explain this scenario, called inflation, and to review its motivations, its most important mechanisms and its observable consequences. As we shall see, the existence of inflation is established nowadays on a rather firm basis.

1.2 Motivations for inflation

1.2.1 Flatness problem

Today, Ω_k is measured to be at most of order 10^{-2}, possibly much smaller, while $\Omega_r \equiv \rho_r/\rho_{crit} \simeq \rho_r/(\rho_v + \rho_m)$ is of order 10^{-4}. Since ρ_k^{eff} scales like a^{-2}, while radiation scales like a^{-4}, the hierarchy between ρ_r and ρ_k^{eff} increases as we go back in time. If t_i is some initial time, t_0 is the time today, and we assume for simplicity that the ratio ρ_k^{eff}/ρ_r is at most equal to one today, we obtain

$$\frac{\rho_k^{eff}(t_i)}{\rho_r(t_i)} \leq \left(\frac{a(t_i)}{a(t_0)}\right)^2 = \left(\frac{\rho_r(t_0)}{\rho_r(t_i)}\right)^{1/2} \ . \tag{11}$$

Today, the radiation energy density $\rho_r(t_0)$ is of the order of $(10^{-4}\mathrm{e}V)^4$. If the early universe reached the order of the Planck density $(10^{18}\mathrm{Ge}V)^4$ at the Planck time t_P, then at that time the ratio was

$$\frac{\rho_k^{eff}(t_P)}{\rho_r(t_P)} = \frac{(10^{-4}\mathrm{e}V)^2}{(10^{18}\mathrm{Ge}V)^2} \sim 10^{-62} \ . \tag{12}$$

Even if the universe never reached such an energy, the hierarchy was already huge when ρ_r was of order, for instance, of $(1\,\mathrm{Te}V)^4$.

If we try to build a mechanism for the birth of the classical universe (when it emerges from a quantum gravity phase), we will be confronted to the problem of predicting an initial order of magnitude for the various terms in the Friedmann equation: matter, spatial curvature and expansion rate. The Friedmann equation gives a relation between the three, but the question of the relative amplitude of the spatial curvature with respect to the total matter energy density, i.e. of the hierarchy between ρ_k^{eff} and ρ_r, is an open question. We could argue that the most natural assumption is to start from contributions sharing the same order of magnitude; this is actually what one would expect from random initial conditions at the end of a quantum gravity stage. The flatness problem can therefore be formulated as: why should we start from initial conditions in the very early universe such that ρ_k^{eff} should be fine-tuned to a fraction 10^{-62} of the total energy density in the universe?

The whole problem comes from the fact that the ratio ρ_k^{eff}/ρ_r (or more generally $\Omega_k \equiv \rho_k^{eff}/\rho_{crit}$) increases with time: i.e., a flat universe is an unstable solution of the Friedmann equation. Is this a fatality, or can we choose a framework in which the flat universe would become an attractor solution? The answer to this question is yes, even in the context of ordinary general relativity. We noticed earlier that $|\Omega_k|$ is proportional to $(aH)^{-2}$, i.e. to $\dot{a}^{-2}$. So, as long as the expansion is decelerated, $\dot{a}$ decreases and $|\Omega_k|$ increases. If instead the expansion is accelerated, $\dot{a}$ increases and $|\Omega_k|$ decreases: the curvature is diluted and the universe becomes asymptotically flat.

Inflation is precisely defined as an initial stage during which the expansion is accelerated. One of the motivations for inflation is simply that if this stage is long enough, $|\Omega_k|$ will be driven extremely close to zero, in such way that the evolution between the end of inflation and today does not allow to reach again $|\Omega_k| \sim 1$.

We can search for the *minimal quantity of inflation* needed for solving the flatness problem. For addressing this issue, we should study a cosmological scenario where inflation takes place between times t_i and t_f such that $|\Omega_k| \sim 1$ at t_i, and $|\Omega_k| \sim 1$ again today at t_0. Let us compute the duration of inflation in this model. This will give us an absolute *lower bound* on the needed amount of inflation in the general case. Indeed, we could assume $|\Omega_k| \gg 1$ at t_i (since there could be a long stage of decelerated expansion before inflation); this would just require more inflation. Similarly, we could assume $|\Omega_k| \ll 1$ today at t_0, requiring again more inflation.

So, we assume that between t_i and t_f the scale factor grows from a_i to a_f, and for simplicity we will assume that the expansion is exactly De Sitter (i.e., exponential) with a constant Hubble rate H_i, so that the total density ρ_v is constant between t_i and t_f. We assume that at the end of inflation all the energy ρ_v is converted into a radiation energy ρ_r, which decreases like a^{-4} between t_f and t_0. Finally, we assume that ρ_k^{eff} (which scales like a^{-2}) is equal to ρ_v at t_i and to ρ_r at t_0. With such assumptions, we can write

$$\frac{\rho_k^{eff}(a_0)}{\rho_k^{eff}(a_i)} = \left(\frac{a_i}{a_0}\right)^2 = \frac{\rho_r(a_0)}{\rho_v(a_i)} = \frac{\rho_r(a_0)}{\rho_v(a_f)} = \frac{\rho_r(a_0)}{\rho_r(a_f)} = \left(\frac{a_f}{a_0}\right)^4 \tag{13}$$

and we finally obtain the relation

$$\frac{a_f}{a_i} = \frac{a_0}{a_f} \ . \tag{14}$$

In the general case, the condition becomes

$$\frac{a_f}{a_i} \geq \frac{a_0}{a_f} \ , \tag{15}$$

which can be summarized in one sentence: there should be as much expansion during inflation as after inflation. A convenient measure of expansion is the so-called *e-fold number* defined as

$$N \equiv \ln a \ . \tag{16}$$

The scale factor is physically meaningful up to a normalization constant, so the e-fold number is defined modulo a choice of origin. The amount of expansion between two times t_1 and t_2 is specified by the number of e-folds $\delta N = N_2 - N_1 = \ln(a_2/a_1)$. So, the condition on the absolute minimal duration of inflation reads

$$(N_f - N_i) \geq (N_0 - N_f) \tag{17}$$

i.e., the number of inflationary e-folds should greater or equal to the number of post-inflationary e-folds $\Delta N \equiv N_0 - N_f$. There is no upper bound on $(N_f - N_i)$: for solving the flatness problem, inflation could be arbitrarily long.

It is easy to compute ΔN as is a function of the energy density at the end of inflation, $\rho_r(a_f)$. We know that today $\rho_r(a_0)$ is of the order of $(10^{-4}\mathrm{e}V)^4$, and we will see in subsection 3.3 that the inflationary energy scale is at most of the order of $(10^{16}\mathrm{Ge}V)^4$, otherwise current observations of CMB anisotropies would have detected primordial gravitational waves. This gives

$$\Delta N = \ln \frac{a_0}{a_f} = \ln \left(\frac{\rho_r(a_f)}{\rho_r(a_0)} \right)^{1/4} \leq \ln 10^{29} \sim 67 \; . \tag{18}$$

We conclude that if inflation takes place around the $10^{16}\mathrm{Ge}V$ scale, it should last for a minimum of 67 e-folds. If it takes place at lower energy, the condition is weaker. The lowest scale for inflation considered in the literature (in order not to disturb too much the predictions of the standard inflationary scenario) is of the order of 1 TeV. In this extreme case, the number of post-inflationary e-folds would be reduced to

$$\Delta N \sim \ln 10^{16} \sim 37 \tag{19}$$

and the flatness problem can be solved with only 37 e-folds of inflation.

1.2.2 Horizon problem

The horizon $d_H(t_1, t_2)$ is defined as the physical distance at time t_2 between two particles emitted at the same point but in opposite directions at time t_1, and traveling at the speed of light. If the origin of spherical comobile coordinates is chosen to coincide with the point of emission, the physical distance at time t_2 can be computed by integrating over small distance elements dl between the origin and the position r_2 of one particle, and multiplying by two,

$$d_H(t_1, t_2) = 2 \int_0^{r_2} dl = 2 \int_0^{r_2} a(t_2) \frac{dr}{\sqrt{1 - k\, r^2}} \; . \tag{20}$$

In addition, the geodesic equation for ultra-relativistic particles gives $ds = 0$, i.e., $dt = a(t) dr / \sqrt{1 - k\, r^2}$, which can be integrated along the trajectory of the particles,

$$\int_{t_1}^{t_2} \frac{dt}{a(t)} = \int_0^{r_2} \frac{dr}{\sqrt{1 - k\, r^2}} \; . \tag{21}$$

We can now replace in the expression of d_H and get

$$d_H(t_1, t_2) = 2 a(t_2) \int_{t_1}^{t_2} \frac{dt}{a(t)} \; . \tag{22}$$

Usually, the result is presented in this form. However, for the following discussion, it is be particularly useful to eliminate the time from the integral by noticing that $dt = da/(aH)$,

$$d_H(a_1, a_2) = 2 a_2 \int_{a_1}^{a_2} \frac{da}{a^2 H(a)} \; , \tag{23}$$

where the Hubble parameter is seen now as a function of a. Let us assume that t_1 and t_2 are two times during Radiation Domination (RD). We know from the Friedmann equation that during RD on has $H \propto a^{-2}$, so we can parametrize the Hubble rate as $H(a) = H_2 \; (a_2/a)^2$. We obtain

$$d_H(a_1, a_2) = 2 a_2 \int_{a_1}^{a_2} \frac{da}{a_2^2 H_2} = \frac{2}{H_2} \frac{(a_2 - a_1)}{a_2} \; . \tag{24}$$

If the time t_2 is much after t_1 so that $a_2 \gg a_1$, the expression for the horizon does not depend on a_1,

$$d_H(a_1, a_2) \simeq \frac{2}{H_2} \; . \tag{25}$$

So, the horizon equals twice the Hubble radius at time t_2,

$$d_H(t_1, t_2) = 2R_H(t_2) \ . \tag{26}$$

The horizon represents the causal distance in the universe. Suppose that a physical mechanism is turned on at time t_1. Since no information can travel faster than light, the physical mechanism cannot affect distances larger than $d_H(t_1, t_2)$ at time t_2. So, the horizon provides the *coherence scale* of a given mechanism. For instance, if a phase transition creates bubbles or patches containing a given vacuum phase, the scale of homogeneity (i.e., the maximum size of the bubble, or the scale on which a patch is nearly homogeneous) is given by $d_H(t_1, t_2)$ where t_1 is the time at the beginning of the transition.

Before photon decoupling, the Planck temperature of photons at a given point depends on their local density. A priori, we can expect that the universe will emerge from a quantum gravity stage with random values of the local density. The coherence length, or characteristic scale on which the density is nearly homogeneous, is given by $d_H(t_1, t_2)$. We have seen that if t_1 and t_2 are two times during radiation domination, this quantity cannot exceed $2R_H(t_2)$, even in the most favorable limit in which t_1 is chosen to be infinitely close to the initial singularity. We conclude that at time t_2, the photon temperature should not be homogeneous on scales larger than $2R_H(t_2)$.

CMB experiments map the photon temperature on our last-scattering-surface at the time of photon decoupling, which roughly coincides with the time of equality between radiation and matter (actually decoupling takes place a bit after equality, but this is unimportant for our purpose). So, we expect CMB maps to be nearly homogeneous on a characteristic scale $2R_H(t_{\rm dec})$. This scale is very easy to compute: knowing that $H(t_0)$ is of the order of $(h/3000)\,{\rm M}pc^{-1}$ with $h \simeq 0.7$, we can extrapolate $H(t)$ back to the time of equality, and find that the distance $2R_H(t_{\rm dec})$ subtends an angle of order of a few degrees in the sky - instead of encompassing the diameter of the last scattering surface. So, it seems that the last scattering surface is composed of several thousands causally disconnected patches. However, the CMB temperature anisotropies are only of the order of 10^{-5}: in other words, the full last scattering surface is extremely homogeneous. This appears as completely paradoxical in the framework of the Hot Big Bang scenario.

What is the origin of this problem? When we computed the horizon, we integrated $(a^2H)^{-1}$ over da and found that the integral was converging with respect to the boundary a_1: so, even by choosing the initial time to be infinitely early, the horizon is bounded by a function of a_2. If the integral was instead divergent, we could obtain an infinitely large horizon at time t_2 simply by choosing a_1 to be small enough. The convergence of the integral

$$\int_{a_1}^{a_2} \frac{da}{a^2 H(a)} = \int_{a_1}^{a_2} \frac{da}{a\dot{a}} \tag{27}$$

with respect to $a_1 \to 0$ depends precisely on the fact that the expansion is accelerated or decelerated. For linear expansion, the integrand is $1/a$, the limiting case between convergence and divergence. If it is decelerated, $\dot{a}$ decreases and the integral converges. If it is accelerated, $\dot{a}$ increases and the integral diverges in the limit $a_1 \to 0$.

So, if the radiation dominated phase is preceded by an infinite stage of accelerated expansion, one can reach an arbitrarily large value for the horizon at the time of decoupling. In fact, in order to explain the homogeneity of the last scattering surface, we only need to boost the horizon by a factor of $\sim 10^3$ with respect to the Hubble radius at that time. This can be fulfilled with a rather small amount of accelerated expansion.

In order to determine the minimum number of inflationary e-folds, we assume a cosmological scenario in which two photons are emitted at a time t_i which coincides with the beginning of inflation. We then assume that the expansion is exactly exponential (i.e., with a constant Hubble parameter H_i) until t_f, and then we switch back to radiation domination. During inflation, the horizon will grow like

$$d_H(t_i, t_f) = 2a_f \int_{a_1}^{a_2} \frac{da}{a^2 H_i} = \frac{2}{H_i}\left(\frac{a_f}{a_i} - 1\right) \ , \tag{28}$$

and later on it goes on increasing as we already computed before in Eqs. (24,25), so that

$$d_H(t_i, t) = d_H(t_i, t_f) + 2R_H(t) \ . \tag{29}$$

The horizon problem can be solved if the *comoving scale* (i.e. the physical scale divided by a) corresponding to the horizon at the end of inflation is at least equal to the *comoving scale* corresponding to the diameter of the last scattering surface. This will ensure that the last scattering surface is entirely in causal contact. What is the physical diameter of the last scattering surface, or in other words, of the largest wavelength that we can observe today? In order to compute it, we must assume that two photons are emitted at the decoupling time t_{dec}, both in the direction of earth, but from opposite directions; we then assume that they reach us today. If the origin of comobile coordinates is chosen to coincide with our position, the goal is simply to compute the comobile coordinate $r_{\rm dec}$ of the photons at the time of emission and to express the comobile distance $2r_{\rm dec}$ in physical units today. This calculation is similar to that of $d_H(t_1, t_2)$ in Eqs. (20-22), excepted that the integral in r space is performed between $r_{\rm dec}$ and 0, the time integral between $t_{\rm dec}$ and t_0, and dr is replaced by $(-dr)$ in the equation of propagation of photons. So, the physical diameter of the last scattering surface (expressed in the units of today) is nothing but $d_H(t_{\rm dec}, t_0)$. If we neglect the recent stage of cosmological constant (or dark energy) domination, we can integrate $d_H(t_{\rm dec}, t_0)$ using the fact that $a \propto t^{2/3}$ during matter domination, and find that $d_H(t_{\rm dec}, t_0) \sim R_H(t_0)$ modulo a numerical factor of order one. Even if we were taking into account the recent Dark Energy (DE) domination, we would not find a significant difference between $d_H(t_{\rm dec}, t_0)$ and $R_H(t_0)$. So, the current value of the Hubble radius provides the order of magnitude of the largest scale observable today, i.e. the diameter of the last scattering surface expressed in units of today.

We just showed that the horizon problem can be solved if the comoving horizon at the end of inflation is at least equal to the comoving diameter of the last scattering surface. We can now write this as

$$\frac{d_H(t_i, t_f)}{a_f} \geq \frac{R_H(t_0)}{a_0} \; . \tag{30}$$

Using the fact that $H_i = H_f$ (the Hubble parameter is still assumed to be constant during inflation), and omitting numerical factors, this condition reads

$$\frac{1}{a_f H_f} \left(\frac{a_f}{a_i} - 1 \right) \geq \frac{1}{a_0 H_0} \; . \tag{31}$$

In the limit $a_f \gg a_i$, we obtain

$$\frac{a_f}{a_i} \geq \frac{a_f H_f}{a_0 H_0} \; . \tag{32}$$

In order to evaluate the right hand-side, we can us the fact that during radiation domination (aH) scales like a^{-1}. This is not true anymore during matter and dark energy domination, but the number of e-folds during these two periods equals respectively seven and one, while we have seen that radiation domination lasts between thirty and sixty e-folds. So, we can neglect matter and DE domination and approximate the ratio $(a_f H_f)/(a_0 H_0)$ by simply a_0/a_f. We finally obtain

$$\frac{a_f}{a_i} \geq \frac{a_0}{a_f} \; . \tag{33}$$

This condition is the same as for the flatness problem: the number of inflationary e-fold should be at least equal to that of post-inflationary e-folds. If it is larger, then the size of the observable Universe is even smaller with respect to the causal horizon.

1.2.3 Origin of perturbations

Since our universe is inhomogeneous, one should find a physical mechanism explaining the origin of cosmological perturbations. Inhomogeneities can be expanded in comoving Fourier space. Their physical wavelength

$$\lambda(t) = \frac{2\pi a(t)}{k} \tag{34}$$

is stretched with the expansion of the universe. During radiation domination, $a(t) \propto t^{1/2}$ and $R_H(t) \propto t$. So, the Hubble radius grows with time faster than the perturbation wavelengths. We conclude that observable perturbations were originally super-Hubble fluctuations (i.e., $\lambda > R_H \;\Leftrightarrow\; k < 2\pi a H$). Actually, the discussion of the horizon problem already showed that at decoupling the largest observable fluctuations are super-Hubble fluctuations. Even if we take a smaller scale, e.g. the typical size of a galaxy cluster $\lambda(t_0) \sim 1$ Mpc, we find that the corresponding fluctuations were clearly super-Hubble fluctuations for instance at the time of Nucleosynthesis. We have seen that in the Hot Big Bang scenario (without inflation) the Hubble radius $R_H(t_2)$ gives an upper bound on the causal horizon $d_H(t_1, t_2)$ for whatever value of t_1. So, super-Hubble fluctuations are expected to be out of causal contact. The problem is that it is impossible to find a mechanism for generating coherent fluctuations on acausal scales. There are two possible solutions to this issue:

- we can remain in the framework of the Hot Big Bang scenario and assume that perturbations are produced causally when a given wavelength enters into the horizon. In this case, there should be not coherent fluctuations on super-Hubble scales, i.e. the power spectrum of any kind of perturbation should fall like white noise in the limit $k \ll aH$. This is exactly what happens if cosmological fluctuations are assumed to be generated by topological defects (e.g. cosmic strings). This scenario is now ruled out for at least two reasons. First, the observation of CMB anisotropies on angular scales greater than one degree (i.e., super-Hubble scales at that time) is consistent with coherent fluctuations rather than white noise. Second, we shall see in subsection 3.2.4 that the observations of acoustic peaks in the power spectrum of CMB anisotropies is a clear proof that cosmological perturbations are generated much before Hubble crossing.

- we can modify the cosmological scenario in such way that all cosmological perturbations observable today were inside the causal horizon when they were generated at some early time (we will study a concrete generation mechanism in section 2).

So, our goal is to find a paradigm such that the largest wavelength observable today, which is $\lambda_{max}(t_0) \sim R_H(t_0)$ (see subsection 1.2.2), was already inside the causal horizon at some early time t_i. If before t_i the universe was in decelerated expansion, then the causal horizon at that time was of order $R_H(t_i)$. How can we have $\lambda_{max} \leq R_H$ at t_i and $\lambda_{max} \sim R_H$ today? If between t_i and t_0 the Universe is dominated by radiation or matter, it is impossible since the Hubble radius grows faster than the physical wavelengths. However, in general,

$$\frac{\lambda(t)}{R_H(t)} = \frac{2\pi a(t)}{k} \frac{\dot{a}(t)}{a(t)} = \frac{2\pi \dot{a}(t)}{k} , \tag{35}$$

so that during accelerated expansion the physical wavelengths grow faster than the Hubble radius. So, if between some time t_i and t_f the universe experiences some inflationary stage, it is possible to have $\lambda_{max} < R_H$ at t_i: the scale λ_{max} can then exit the Hubble radius during inflation and re-enter approximately today (see Figure 1).

In order to find a condition for the needed amount of inflation, let us assume that $\lambda_{max} \sim R_H$ both at t_i and t_0. This will provide an estimate of the number of inflationary e-folds, which should be regarded as a lower bound. Indeed, if we were assuming instead that $\lambda_{max}(t_i) < R_H(t_i)$, we would find that more inflation is needed. So, we can write

$$k_{max} = 2\pi a(t_i)/R_H(t_i) = 2\pi a(t_0)/R_H(t_0) \qquad \Leftrightarrow \qquad \frac{a_i}{a_0} = \frac{H_0}{H_i} , \tag{36}$$

where k_{max} is the comoving wavenumber associated to λ_{max}. Now, let us assume that during t_i and t_f the universe experiences inflation, and for simplicity let us suppose again that the expansion is De Sitter (exponential) with a constant Hubble parameter. In this case we can notice that $H_i = H_f$, and that between t_f and t_0 the Friedmann equation gives $H \propto a^{-2}$ during radiation domination or $H \propto a^{-3/2}$ during matter domination. Actually, if we want a rough estimate of H_0/H_f, we can do as if radiation domination was holding until today, since the universe experiences many more e-folds of radiation domination than matter (and dark energy) domination. So, we get $(H_0/H_f) \sim (a_f/a_0)^2$ and equation (36) becomes

$$\frac{a_i}{a_0} = \left(\frac{a_f}{a_0}\right)^2 \qquad \Leftrightarrow \qquad \frac{a_f}{a_i} = \frac{a_0}{a_f} . \tag{37}$$

So, once again, we find that the number of inflationary e-folds should be at least equal to that of post-inflationary e-folds.

One could argue that the argument on the origin of fluctuations is equivalent to that of the horizon problem, reformulated in a different way. Anyway, for understanding inflation it is good to be aware of the two arguments, even if they are not really independent from each other.

1.2.4 Monopoles

We will not enter here into the details of the monopole problem. Just in a few words, some phase transitions in the early universe are expected to create "dangerous relics" like magnetic monopoles, with a very large density which would dominate the total density of the universe. These relics are typically non-relativistic, with an energy density decaying like a^{-3}: so, they are not diluted, and the domination of radiation and ordinary matter can never take place.

Inflation can solve the problem provided that it takes place after the creation of dangerous relics. During inflation, monopoles and other relics will decay like a^{-3} (a^{-4} in the case of relativistic relics) while the leading vacuum energy is nearly constant: so, the energy density of the relics is considerably diluted, typically by a factor $(a_f/a_i)^3$, and today they are irrelevant. The condition on the needed amount of inflation is much weaker than the condition obtained for solving the flatness problem, since dangerous relics decay faster than the effective curvature density ($\rho_k^{eff} \propto a^{-2}$).

Let us conclude this section by a rather useless but fun discussion concerning the following question: should we summarize inflation as a mechanism producing *more expansion* or *less expansion*?

Usually people would say *more expansion*. The reason is that if we fix some initial conditions, and say that at time t_i the universe has given values of H_i and a_i, we can check that the universe will experience more expansion after t_i if the following stage is inflationary. This is obvious: the expansion rate will maintain itself to a nearly constant value instead of falling like $H = n/t$ for $a \propto t^n$; in the meantime the scale factor will accelerate, so at a given time $t > t_i$ it will have a larger value if there is inflation than if expansion is decelerated.

However, one can easily argue that the opposite answer (*less expansion*) makes a lot of sense. In cosmology, what is fixed is not the initial condition but the current expansion rate: we do not measure any H_i, but we do measure H_0, and given this we try to extrapolate back in time. Without inflation, the extrapolation gives $H(a) \propto a^{-2}$ when $a \to 0$ (assuming radiation domination at early times). So, back in time, the expansion rate grows to infinity. With inflation, it stops at a nearly constant value H_i. In this sense, the early universe experiences *less expansion*. If we assume that between t_i and t_f the universe is experiencing either inflation or radiation domination, and if we pick up any time $t < t_f$, we can easily check that $H(t)$ is smaller in the inflationary case, while $a(t)$ (normalized today to some arbitrary value, e.g. $a(t_0) = 1$) is larger. In other words, if we assume inflation, we see that our observable universe (corresponding to the comobile scale $k_{\max}$ equating the comobile Hubble radius today) is *larger at early times* if we assume inflation[1]. It is worth stressing this point, since many people have the opposite intuition.

1.3 Quick overview of scalar field inflation

We reviewed many cosmological problems which can be solved by an early stage of accelerated expansion. The Einstein equation should tell us which kind of matter can lead to such a stage. Let us start from the Friedmann equation in a flat universe, which can be written as

$$\dot{a} = \sqrt{\frac{8\pi\mathcal{G}}{3}} a\rho^{1/2} \ . \tag{38}$$

Taking the time derivative of this equation and replacing $\dot{\rho}$ according to the energy conservation equation, we obtain

$$\ddot{a} = -\sqrt{\frac{8\pi\mathcal{G}}{3}} \frac{\dot{a}}{2\rho^{1/2}} (\rho + 3p) \ . \tag{39}$$

As long as the universe is expanding, $\dot{a}$ is positive. So, the condition for accelerated expansion is simply

$$\rho + 3p < 0 \ . \tag{40}$$

We conclude that during inflation the pressure should be negative and smaller than $-\rho/3$. Which type of matter could fulfill this requirement? A cosmological constant could do the job since it has $p = -\rho$. During a fully Λ-dominated stage, the Hubble parameter remains constant: one has De Sitter (exponential) inflation. The problem is that a cosmological constant never decays, so inflation will be indefinite. If we want inflation to end, "something must happen", so there must be an arrow of time. Therefore the type of matter responsible for inflation cannot be exactly in equilibrium. The most simple possibility is to consider a scalar field (called the inflaton) slow-rolling in a very flat valley. Because it is rolling, there is an arrow of time, and something can happen that will end inflation. But because the valley is very flat and the rolling is very slow, the field can be seen at any time as in an "instantaneous vacuum state" sharing almost the properties of a true vacuum state: in particular, the energy of the field is diluted very slowly, and the pressure is very close to $-\rho$ (see subsection 2.1).

In the rest of this course, we will review this scenario, in which one can distinguish various interesting phases. The numbers below correspond to those of Figure 1, which summarizes in a sketchy way the evolution of perturbation wavelengths as a function of time in our universe.

[1] However, it is also true that in presence of inflation our observable universe is initially very small *in units of the Hubble radius*.

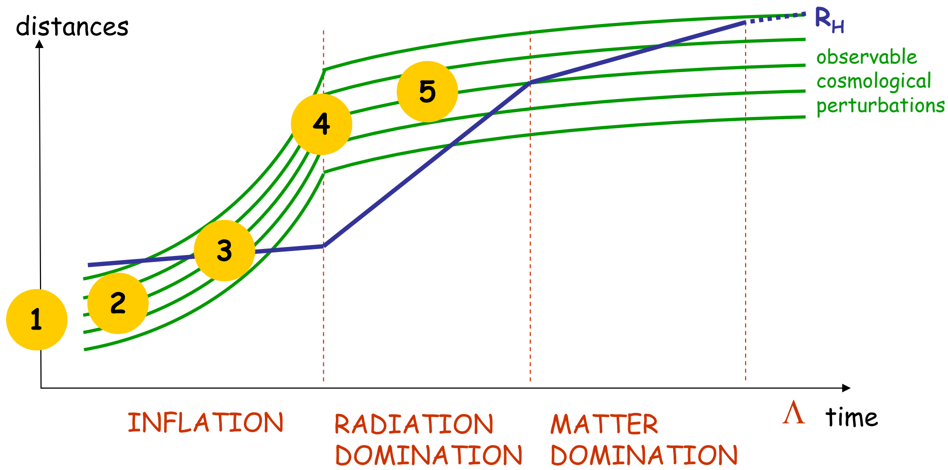

Figure 1: Evolution of a set of observable cosmological wavelengths as a function of time, during inflation, radiation domination, matter domination and cosmological constant (or dark energy) domination. The largest wavelength displayed is equal to the Hubble radius today, so it is the largest scale accessible to observations (it defines the size of the observable universe). We show at the end of subsection 1.3 that this scale exits the Hubble radius 37 to 67 e-folds before the end of inflation, depending on the inflationary energy scales. The numbers 1 to 5 refer to the five epochs described in subsection 1.3.

1. Inflation starts at some time t_i when the potential energy of a slow-rolling scalar field (the inflaton) starts to dominate the total energy density in the universe. At that time the curvature fraction Ω_k is either of order one, or eventually larger (due to a stage of decelerated expansion between the quantum gravity epoch and inflation). In any case, the curvature gets rapidly diluted and after a few e-folds one has $|\Omega_k| \ll 1$. In the next section we will always start studying inflation after that time, so we will only employ the metric of a flat Friedmann-Lemaître universe.

2. Assuming that the previously derived constraints on the minimal number of inflationary e-folds are satisfied, there exists a time at which all observable cosmological wavelengths are inside the Hubble radius (see subsection 1.2.3). In particular, the scalar field and the metric have small quantum fluctuations which can be described like in flat space-time for wavelengths deep inside the Hubble radius (see subsection 2.4).

3. One after each other, and starting from the largest one, the perturbation wavelengths exit the Hubble radius. At that time, the quantum fluctuations of the scalar field and of the metric undergo a semi-classical transition (see subsection 2.4.4).

4. At the end of inflation, the scalar field decays into particles which are either the ones that we observe today (photons, neutrinos, baryons, cold dark matter) or some intermediate particles which will decay later on. This stage is called reheating, and its description is beyond the scope of this course. During reheating, the scalar field decays, so its large wavelength perturbations vanish. However, metric perturbation survive. In addition, metric perturbations with observable cosmological wavelengths have $\lambda \gg R_H$ during reheating. As we shall see in subsections 2.5 and 2.6, this means that they are frozen, i.e. protected against the micro-physics taking place on much smaller scales.

5. after the decay of the scalar field, the universe is dominated by the energy of relativistic particles produced during reheating and enters into the usual radiation-dominated stage. The long-wavelength metric perturbations couple gravitationally to radiation and matter perturbations, which therefore inherit of the large wavelength perturbation spectrum generated during inflation. The evolution of all these perturbations during the radiation and matter dominated stages lead to the observed

CMB anisotropies and to the Large Scale Structure (LSS) of galaxies, clusters of galaxies, etc. A beautiful aspect of the inflationary paradigm is that these astrophysical observables are supposed to depend on the physics of quantum perturbations in the very early Universe!

The (linear) scales that we observe today in the CMB and LSS span roughly three decades in the space of comoving wavenumbers k. The largest observable wavelength λ_{max}, associated to the wavenumber k_{max}, corresponds to $\lambda_{max}(t_0) \sim R_H(t_0)$ (see subsection 1.2.2). We have seen in 1.2.3 that this scale exits the Hubble radius when the number of e-folds ΔN before the end of inflation is equal to that between the end of inflation and today. In subsection 1.2.1 we showed that ΔN lies between 67 and 37, depending on the energy scale of inflation. We can conclude that observable cosmological wavelengths exit the Hubble radius typically 60 e-folds before the end of inflation for a maximum inflationary scale of 10^{16} GeV, or typically 30 e-folds before the end of inflation for a minimum inflationary scale of 1 TeV (see subsection 1.2.1). This result plays an important role for comparing inflationary models with observations.

2 Inflation with a single scalar field

We recall that the general action for a scalar field in curved space-time

$$S = -\int d^4x \sqrt{|g|}\,(\mathcal{L}_g + \mathcal{L}_\varphi) \tag{41}$$

involves the Lagrangian of gravitation

$$\mathcal{L}_g = \frac{R}{16\pi\mathcal{G}} \tag{42}$$

and that of the scalar field

$$\mathcal{L}_\varphi = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V(\varphi) = \frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi)\ . \tag{43}$$

The variation of the action with respect to $g_{\mu\nu}$ enables to define the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - \mathcal{L}_\varphi g_{\mu\nu} \tag{44}$$

and the Einstein tensor $G_{\mu\nu}$, which are related through the Einstein equations

$$G_{\mu\nu} = 8\pi\mathcal{G}\,T_{\mu\nu}\ . \tag{45}$$

Instead, the variation of the action with respect to φ gives Klein-Gordon equation

$$\frac{1}{\sqrt{|g|}}\partial_\mu\left[\sqrt{|g|}\partial^\mu\varphi\right] + \frac{\partial V}{\partial\varphi} = 0\ . \tag{46}$$

The same equation could have been obtained using a particular combination of the components of $T_{\mu\nu}$ and their derivatives, which vanish by virtue of the Bianchi identities (in other word, the Klein-Gordon equation is contained in the Einstein equations).

2.1 Slow-roll conditions

We assume that the homogeneous Friedmann universe with flat metric

$$g_{\mu\nu} = \mathrm{d}iag\left(1, -a(t)^2, -a(t)^2, -a(t)^2\right) \tag{47}$$

is filled by a homogeneous classical scalar field $\bar\varphi(t)$ (here $x_0 = t$ is the proper time or cosmological time). *Exercise:* show that the corresponding energy-momentum tensor is diagonal, $T_\mu^\nu = \mathrm{d}iag(\rho, -p, -p, -p)$, with

$$\rho = \frac{1}{2}\dot{\bar\varphi}^2 + V(\varphi)\ , \tag{48}$$

$$p = \frac{1}{2}\dot{\bar\varphi}^2 - V(\varphi)\ . \tag{49}$$

The Friedmann equation reads

$$G_0^0 = 3H^2 = 8\pi\mathcal{G}\,\rho \tag{50}$$

and the Klein-Gordon equation

$$\ddot{\bar\varphi} + 3H\dot{\bar\varphi} + \frac{\partial V}{\partial\varphi}(\bar\varphi) = 0 \ . \tag{51}$$

These two independent equations specify completely the evolution of the system. However it is worth mentioning that the full Einstein equations provide another relation

$$G^i_i = \left(2\frac{\ddot a}{a} + \left(\frac{\dot a}{a}\right)^2\right) = -8\pi\mathcal{G}\, p \ . \tag{52}$$

The combination $\dot G^0_0 + 3H(\dot G^0_0 - G^i_i)$ vanishes (it is one of the Bianchi identities), and gives a conservation equation $\dot\rho + 3H(\rho+p) = 0$, which is nothing but the Klein-Gordon equation. Finally, the combination $G^i_i - G^0_0$ provides a very useful relation

$$\dot H = -4\pi\mathcal{G}\,\dot{\bar\varphi}^2 \tag{53}$$

which is consistent with the fact that the Hubble parameter can only decrease (since $\rho + p$ is positive, for the scalar field as for any kind of matter respecting the weak energy condition[2]).

The condition $p < -\rho/3$ (see subsection 1.3) reads $\dot{\bar\varphi}^2 < V$: when the potential energy dominates over the kinetic energy, the universe expansion is accelerated. In the limit of zero kinetic energy, the energy-momentum tensor would be that of a cosmological constant, and the expansion would be exponential (this is called "De Sitter expansion") and everlasting. For a long, finite stage of acceleration we must require that the *first slow-roll condition*

$$\frac{1}{2}\dot{\bar\varphi}^2 \ll V(\bar\varphi) \tag{54}$$

holds over an extended period. Since the evolution of the scalar field is given by a second-order equation, the above condition could apply instantaneously but not for an extended stage, in particular in the case of oscillatory solutions. If we want the first slow-roll condition to hold over an extended period, we must impose that the time-derivative of this condition also holds (in absolute value). This gives the *second slow-roll condition*

$$|\ddot{\bar\varphi}| \ll \left|\frac{\partial V}{\partial\varphi}(\bar\varphi)\right| \tag{55}$$

which can be rewritten, by virtue of the Klein-Gordon equation, as

$$|\ddot{\bar\varphi}| \ll 3H\,|\dot{\bar\varphi}| \ . \tag{56}$$

When these two conditions hold, the Friedmann and Klein-Gordon equations become

$$\begin{aligned} 3H^2 &\simeq 8\pi\mathcal{G}\,V(\bar\varphi) \ , & (57)\\ \dot{\bar\varphi} &\simeq -\frac{1}{3H}\frac{\partial V}{\partial\varphi}(\bar\varphi) \ . & (58) \end{aligned}$$

The two slow-roll conditions can be rewritten as conditions either on the slowness of the variation of H, or on the flatness of the potential.
Exercise: show that Eqs. (54, 56) are equivalent either to

$$-\dot H \ll 3H^2 \ , \qquad |\ddot H| \ll -6H\dot H \ll 18H^3 \tag{59}$$

or to

$$\left(\frac{\partial V/\partial\varphi}{V}\right)^2 \ll 48\pi\mathcal{G} \ , \qquad \left|\frac{\partial^2 V/\partial\varphi^2}{V}\right| \ll 48\pi\mathcal{G} \ . \tag{60}$$

Liddle and Lyth introduced the following dimensionless *slow-roll parameters*

$$\epsilon = \frac{1}{16\pi\mathcal{G}}\left(\frac{\partial V/\partial\varphi}{V}\right)^2 \ , \qquad \eta = \frac{1}{8\pi\mathcal{G}}\frac{\partial^2 V/\partial\varphi^2}{V} \ . \tag{61}$$

We see from Eqs. (60) that the slow-roll conditions read

$$\epsilon \ll 1 \ , \qquad |\eta| \ll 1 \ , \tag{62}$$

with numerical factors are a bit different from those in Eqs. (54, 56, 60).

[2] or at least the null energy condition.

2.2 Background evolution

The exact evolution of the background (the homogeneous field and the scale factor) can be found by solving the Friedmann and Klein-Gordon equations (50, 51). However, if we are sure that the slow-roll conditions are satisfied, we can solve simply the approximate first-order equation

$$\dot{\bar{\varphi}} = -\frac{1}{3H}\frac{\partial V}{\partial \varphi} = -\frac{1}{\sqrt{24\pi\mathcal{G}\,V(\bar{\varphi})}}\frac{\partial V}{\partial \varphi} \ . \tag{63}$$

It is necessary to check that the potential allows for a sufficient number of inflationary e-folds. By integrating over $dN = d\ln a$ and making use of Eqs. (57, 58), it is straightforward to show that the number of e-folds between the time t_i and the time t_f is

$$N = \int_{t_i}^{t_f} H dt = -8\pi\mathcal{G}\int_{\varphi_i}^{\varphi_f} \frac{V}{\partial V/\partial \varphi} d\varphi \ . \tag{64}$$

For a particular form of the potential, one can compute the value of the field φ_f at the end of inflation (generally, φ_f is such that $\max[\,\epsilon,\,|\eta|\,] = 1$). Then, the above relation provides a condition on the initial value φ_i in order to obtain a sufficient number of e-folds. We will employ this method for many concrete examples of potentials in section 3.4.

2.3 Perturbations and gauge freedom

We now decompose the metric and field into a homogeneous background plus spatial perturbations:

$$g_{\mu\nu}(t,\mathbf{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\mathbf{x}) \ , \qquad \varphi(t,\mathbf{x}) = \bar{\varphi}(t) + \delta\varphi(t,\mathbf{x}) \tag{65}$$

where $\bar{g}_{\mu\nu}$ is the Friedmann metric. In principle the symmetric 4×4 tensor $\delta g_{\mu\nu}(t,\mathbf{x})$ has ten degrees of freedom. They can be classified according to their behavior under (three-dimensional) spatial rotations as scalars, vectors and tensors. There are four scalar degrees of freedom describing the generalization of Newtonian gravity; four vector degrees of freedom describing gravito-magnetism; and two tensor degrees of freedom describing gravitational waves. At first order in perturbation theory, these three sectors are uncoupled and governed by independent equations of motion. The scalar metric perturbations couple with the scalar field perturbation $\delta\varphi$. The vector modes have only decaying solutions: so, they can be neglected in the present context. The tensor perturbations are uncoupled with the field but their equations of motion have non-decaying solutions: so, we will include them in our study.

There are more degrees of freedom in the metric than physical modes: some solutions of the equations of motion are gauge artifacts, which cannot be observed. This subtlety comes from the fact that there is not a unique way of defining the perturbations at a given point. Let us consider a perturbed quantity $\rho(t,\mathbf{x})$. Such a quantity is perfectly defined. A change of coordinates will re-map the field ρ but not change it. This is not true for the perturbation $\delta\rho(t,\mathbf{x}) = \rho(t,\mathbf{x}) - \langle\rho(t,\mathbf{x})\rangle_x$ (here $\langle\rho(t,\mathbf{x})\rangle_{\mathbf{x}}$ is the spatial average at time t). The definition of the perturbation is non-local since the average is performed along the hypersurface of simultaneity at time t. If the change of coordinate $x_\mu \mapsto x'_\mu$ changes the hypersurfaces of simultaneity, the new quantity $\delta\rho(t',\mathbf{x}')$ will be defined by comparing the local value $\rho(t',\mathbf{x}')$ with *different* physical points on a *different* hypersurface of simultaneity. In particular, it is possible that in one coordinate system $\delta\rho(t,\mathbf{x})$ is null everywhere in space-time, while in another system it is not.

Let us assume that the universe is slightly perturbed, i.e. that there exists one choice of coordinates such that all perturbations are small with respect to the actual inhomogeneous quantities. There is an infinite number of gauges, i.e. of ways to define the hypersurfaces of simultaneity, such that the perturbations remain small. Therefore, the metric perturbations can be changed by a group of gauge transformation, i.e. coordinate changes mixing space and time but keeping the perturbations small. Gauge transformations have four degrees of freedom (since there are four space-time coordinates). As a consequence the number of physical degrees of freedom in metric perturbations is equal to $10 - 4 = 6$: two scalars, two vectors and two tensors (tensors appear to be gauge-invariant).

One way to perform computations is to define gauge-invariant quantities obeying to gauge-invariant equations of motion. This is rather complicated and not strictly necessary. In fact, physically observable quantities are always gauge-invariant. So, one can simply fix a gauge, i.e. impose a prescription leading to a unique slicing of space-time into hypersurfaces of simultaneity. Then, the variables are not gauge-invariant, but at least their equation of motion have the correct number of independent solutions. In the Friedmann universe and for the scalar sector, we will use one of these prescriptions which amounts in requiring that the perturbed metric is diagonal:

$$g_{\mu\nu} = \mathrm{diag}\left((1+2\phi), -a^2(1-2\psi), -a^2(1-2\psi), -a^2(1-2\psi)\right) \ . \tag{66}$$

This gauge is called the Newtonian or longitudinal gauge. Tensor perturbations are gauge-invariant, and the gravitational waves are usually defined as the two independent components (or degrees of polarization) of the traceless transverse 3×3 tensor h_{ij} such that

$$\delta g_{ij} = -a(t)^2 h_{ij} \ , \tag{67}$$

where by definition $h^i_i = 0$ and $\forall j$, $\partial_i h^i_j = 0$. It is possible to decompose this tensor as

$$h_{ij} = h_1 e^1_{ij} + h_2 e^2_{ij} \tag{68}$$

where h_1 and h_2 are two independent functions of space and time (the two degrees of polarization of gravitational waves) and e^1_{ij}, e^2_{ij} are two orthogonal traceless transverse tensors of norm $1/2$ each, so that

$$\sum_{ij\lambda} e^\lambda_{ij} e^\lambda_{ij} = 1/2 + 1/2 = 1 \tag{69}$$

(attention: some authors use an orthonormal basis in which $\sum_{ij\lambda} e^\lambda_{ij} e^\lambda_{ij} = 1 + 1 = 2$, and some of the intermediate results of subsection 2.5 read differently in that case).

2.4 Quantization and semi-classical transition

2.4.1 Basic recalls on quantization of a free scalar field in flat space-time

Let us recall that for a quantum harmonic oscillator with equation of motion $\ddot{x} + \omega^2 x = 0$, the wave function of the fundamental state is a Gaussian of variance $\omega^{-1/2}$,

$$\Psi_0(x) = \mathcal{N} e^{-\frac{1}{2}\omega x^2} \ . \tag{70}$$

So, the probability $\mathcal{P}(x) = |\Psi_0(x)|^2$ to find the system in a position x is a Gaussian of variance

$$\sigma = \sqrt{\frac{1}{2\omega}} \ . \tag{71}$$

We know that a free massless real scalar field in flat space-time –i.e. with the Minkowski metric $g_{\mu\nu} = \mathrm{d}iag(1, -1, -1, -1)$– can be easily quantized, since each Fourier mode is analogous to the above harmonic oscillator. The Lagrangian

$$\mathcal{L} = \frac{1}{2} \partial_\mu \chi \partial^\mu \chi \tag{72}$$

leads to the Euler equation

$$\ddot{\chi} + \partial_i \partial^i \chi = 0 \ , \tag{73}$$

to the definition of the conjugate momentum

$$p = \frac{\partial \mathcal{L}}{\partial \dot{\chi}} = \dot{\chi} \tag{74}$$

and to the Hamiltonian

$$H = \frac{1}{2} \int d^3x \left[\dot{\chi}^2 + \partial_i \chi \partial^i \chi \right] \ . \tag{75}$$

The real field χ can be Fourier transformed to the complex field $\chi_k = \chi^*_{-k}$. In Fourier space the Euler equation reads

$$\ddot{\chi}_k + k^2 \chi_k = 0 \tag{76}$$

and the Hamiltonian becomes

$$H = \frac{1}{2} \int d^3k \left[\dot{\chi}_k \dot{\chi}^*_k + k^2 \chi_k \chi^*_k \right] \ . \tag{77}$$

So, each Fourier mode is a harmonic oscillator, and in the fundamental state the wave functional of a given mode is

$$\Psi_0(\chi_k) = \mathcal{N} e^{-\frac{1}{2} k |\chi_k|^2} \ . \tag{78}$$

So, the probability $\mathcal{P}(\chi_k) = |\Psi_0(\chi_k)|^2$ that a Fourier mode of wavenumber k has an amplitude χ_k is a Gaussian of variance

$$\sigma = \sqrt{\frac{1}{2k}} \ . \tag{79}$$

This picture becomes much more complicated either when one considers non-quadratic self-coupling terms (due to renormalization issues) or when one goes to curved space-time. In the context of inflation, the first complication does not appear, because instead of quantizing the field itself one quantizes its small perturbations. For such perturbations, non-quadratic self-coupling terms are always negligible (they vanish at first order in perturbation theory). However the curvature of space-time does play a crucial role.

2.4.2 Definition of the mode function

Quantum field theory (in flat or curved space-time) can be explained and understood in various formalisms: Heisenberg, Schrödinger, etc. For the purpose of inflation, it is extremely useful to rely on one particular quantity: the mode function. Let us recall its definition, first for a free scalar field in flat space-time. Any solution of the Euler equation

$$\ddot{\chi}_k + k^2 \chi_k = 0 \tag{80}$$

represents a particular classical solution or classical trajectory (in Fourier space). These solutions can be parametrized as

$$\chi_k = A_k e^{-ikt} + B_k e^{ikt} \ . \tag{81}$$

In quantum field theory (in flat-space time), there is no ambiguity in the definition of time because the metric is invariant under time translations: then one says that $(\partial/\partial t)$ is a Killing vector. Eigenfunctions of this Killing vector obeying to $(\partial/\partial t) f_k = -ik f_k$ with $k > 0$ are called positive frequency solutions, and are the only physical solutions. In the previous equation, the positive frequency solutions are those with $B_k = 0$. Among all possible positive frequency solutions, one plays a particular role, because it is normalized according to the commutation relation $[x, p] = [\chi, \dot{\chi}] = i$ (remember that we are using units with $\hbar = 1$). In Fourier space this relation gives

$$\chi_k \dot{\chi}_k^* - \dot{\chi}_k \chi_k^* = i \ . \tag{82}$$

The quantity on the left hand side is the Wronskian, which is always a conserved quantity for a solution of a second-order differential equation. This normalization condition gives in our case

$$ik|A_k|^2 - (-ik|A_k|^2) = i \ , \tag{83}$$

and up to an arbitrary phase one finds

$$A_k = \sqrt{\frac{1}{2k}} \ . \tag{84}$$

The positive frequency solution of the classical equation of motion, normalized to the commutation relation, is called the *mode function.* It represents a typical solution, i.e. a classical solution corresponding to one standard deviation for the quantum system in its fundamental state. We can check it explicitly. We found that for the free field in flat space-time the mode function reads

$$\chi_k = \sqrt{\frac{1}{2k}} e^{-ikt} \ . \tag{85}$$

The modulus of this function is equal to $(2k)^{-1/2}$, i.e. according to Eq. (79) to one standard deviation for the probability $\mathcal{P}(\chi_k)$.

2.4.3 Free field in curved space-time

In curved space-time, the time derivative is not anymore a Killing vector. It is impossible to build positive-frequency solutions. The Euler equation and Hamiltonian for each Fourier mode are those of a harmonic oscillator with a mass explicitly depending on time. At a given time, it is possible to build annihilation/creation operators and to define an "instantaneous fundamental state" and "instantaneous N-particle states". However this construction of a Fock space is not invariant with time. The annihilation/creation operators evolve according to the Bogolioubov transformations, and a state corresponding to the fundamental state at time t_1 will be an excited state at time t_2. As a consequence, quantization in curved space-time is generally ambiguous.

For the purpose of quantizing inflationary perturbations, we can avoid such ambiguities. There is a simple physical explanation for that. Today, in laboratory experiments, we can treat all quantum fields as if they where living in flat space-time. However we do leave in an expanding Friedmann universe. The reason for which space-time curvature effects can be neglected on small scales is that in everyday life we deal with distances and wavelengths much smaller than the curvature radius R_k and the Hubble radius R_H.

We have seen that during inflation, observable Fourier modes start inside the Hubble radius. Wavelengths grow much faster than R_H and at some point $\lambda \sim R_H$ (in terms of wavenumbers this corresponds to $k \sim aH$). If we define the initial state of the field much before that time, when $k \gg aH$, the system is effectively in flat space-time, exactly like for laboratory experiments today. So, at that time, the definition of the vacuum is not ambiguous[3]. The system can be assumed to stay in its fundamental state when $k \gg aH$. Later on, this state will evolve and will not be anymore the vacuum state: there will be particle creation from the vacuum, a usual feature in curved space-time.

It is actually possible to show that near horizon crossing and afterward, the wave function of the Fourier modes initially in their vacuum state develops an "imaginary squared variance",

$$\Psi_0(\chi_k) = \mathcal{N} e^{-\frac{\chi_k^2}{2\sigma_k^2(t)}[1+iF_k(t)]} . \tag{86}$$

In this parametrization, the real variance σ_k is equal to $(2k)^{-1/2}$ well-inside the Hubble radius, and remains equal to the modulus of the mode function at any later time. The real function $F_k(t)$ is negligible well-inside the Hubble radius, and then becomes very large. Because of $F_k(t)$, the wave function is not a Gaussian and does not represent an "instantaneous fundamental state" at late times. However, the probability $\mathcal{P}(\chi_k) = |\Psi_0(\chi_k)|^2$ is still a Gaussian with variance $\sigma_k(t)$. So, if we are only interested in $\mathcal{P}(\chi_k)$, we just need to compute the mode function in order to know all the properties of the system at any time.

2.4.4 Quantum to semi-classical transition

The reason for which we are only interested in $\mathcal{P}(\chi_k)$ is that the system undergoes a quantum to semi-classical transition. These words should be understood exactly in the following sense: at late times, the quantum perturbations are indistinguishable from the perturbations of a classical stochastic system.

This should not be confused with another mechanism called decoherence. When a quantum system interacts with an environment, it can be shown that generally this system will be effectively classical after a while. This can happen also in cosmology. However, the quantum to semi-classical transition during inflation has nothing to do with possible interactions with an environment: it is only an effect of the space-time dynamics. Even if the inflationary perturbations are a non-interacting system following a unitary evolution, they become equivalent to classical stochastic variables after Hubble crossing, in an unavoidable and irreversible way.

A classical stochastic system is described by an equation of motion and an initial distribution of probability in phase space. The distribution of probability at late time is equal to that at initial time "transported" by the equation of motion, which provides a mapping between each point in phase-space at two different time. This mapping is simply given by the solution of the Hamilton-Jacobi equation:

$$x(t) = \alpha(t)x(t_0) + \beta(t)p(t_0) \tag{87}$$

$$p(t) = \gamma(t)x(t_0) + \delta(t)p(t_0) \tag{88}$$

[3] An interesting question to study is: how small are the observable cosmological wavelengths with respect to the Hubble radius at the time when we quantize them, and how good is the approximation consisting in writing the initial fundamental state as if the fields where in flat space-time? If inflation was infinitely long, we could say that by quantizing the fields early enough we can have $k/(aH)$ as large as we want, and there would be no problem at all with the definition of the initial fundamental state. However, if we go back in time, there will be some point at which the physical wavelength is so small that ordinary quantum field theory does not apply –presumably, this happens when λ is of the order of the Planck length λ_P, but in some scenarios with extra dimensions the fundamental scale of gravity can be different.

Quantum perturbations emerge from the quantum gravity regime with a spectrum that we are unable to predict. Is this a problem for computing the inflationary power spectrum? The answer depends on the ratio between the fundamental scale of gravity and the Hubble scale during inflation. If this ratio is very large, then a given wavelength starts to be described by ordinary quantum field theory at a time when it is very deep inside the Hubble radius. At that time, it is legitimate to apply the ordinary machinery of quantum field theory in flat space-time and to define the vacuum state in the usual way. However, if this ratio is not so large, we can argue that when the modes emerge from quantum gravity, they will see a slightly curved space-time in which the definition of the vacuum state has some degree of arbitrariness. As a consequence, some signature of quantum gravity could survive, and the mode function could depart from Eq. (85) with significant corrections usually called *transplanckian corrections*. Such corrections are often considered in the literature, but one should keep in mind that they are not a generic prediction of inflation, since in usual models the large hierarchy between the Planck scale and the Hubble radius during inflation ensures that if they exist, they should be vanishingly small. Indeed, the fact that inflation takes place at most at the 10^{16} GeV scale implies that $R_H \geq 10^{-6} m_P^{-1} = 10^{-6}\lambda_P$.

and the initial phase-space distribution is some properly normalized function $\mathcal{P}(x,p)$. There exists a sub-class of these systems for which the Hamilton-Jacobi equations lead to

$$\lim_{t\to\infty} \beta(t)/\alpha(t) = \lim_{t\to\infty} \delta(t)/\gamma(t) = 0 \ . \tag{89}$$

In this case, at late time, the phase-space density concentrates along a line of equation $p = (\gamma(t)/\alpha(t))x$, and along this line it obeys to some probability distribution $\mathcal{P}(x,t)$ for the position. Then, if we want to compute an arbitrary statistical momentum $\langle x^n p^m \rangle$ for this system, we need to compute the integral

$$\langle x^n p^m \rangle = \int dx\, x^n \left(\frac{\gamma(t)}{\alpha(t)} x \right)^m \mathcal{P}(x,t) \ . \tag{90}$$

There is a class of quantum fields in curved space-time for which the Hamilton-Jacobi equations for each mode k obey precisely to Eqs. (89). The corresponding Wigner function (a generalization of the notion of phase-space distribution for quantum systems) does concentrate along a line of equation $p_{\chi k} = (\gamma(t)/\alpha(t))\chi_k$. Then, the mode is said to be in a *squeezed state*, and the momenta $\langle \chi_k^n p_{\chi k}^m \rangle$ are asymptotically indistinguishable from the classical expression of Eq. (90). This can be shown explicitly. In principle the quantum momenta should be computed as

$$\langle \chi_k^n p_{\chi k}^m \rangle = \int d\chi_k \ \Psi^*(\chi_k,t) \chi_k^n \left(\frac{\partial}{i\partial\chi_k} \right)^m \Psi(\chi_k,t) \tag{91}$$

and their semi-classical limit can be defined as

$$\langle \chi_k^n p_{\chi k}^m \rangle_{\rm cl} = \int d\chi_k \ \chi_k^n \left(\frac{\gamma(t)}{\alpha(t)} \chi_k \right)^m |\Psi(\chi_k,t)|^2 \ . \tag{92}$$

For a squeezed state, one can show that

$$\frac{\langle \chi_k^n p_{\chi k}^m \rangle - \langle \chi_k^n p_{\chi k}^m \rangle_{\rm cl}}{\langle \chi_k^n p_{\chi k}^m \rangle} \longrightarrow 0 \tag{93}$$

in the limit in which $\beta(t)/\alpha(t)$ and $\delta(t)/\gamma(t)$ are vanishingly small. This proves that the system becomes effectively classical stochastic.

Quantum inflationary perturbations fall exactly in this ballpark. It is possible to look for the two independent analytical solutions of the Euler equation *after* Hubble crossing: then, one generally finds a *decaying mode* and a *growing mode*. These two solutions can be matched to the positive frequency solution defined *before* horizon crossing. The matching provides normalization coefficient for the growing and decaying mode. Near horizon crossing the two modes have comparable amplitudes, but after a while the amplitude of the decaying mode becomes vanishingly small. This limit is exactly the one in which, in the previous notations, $\beta(t)/\alpha(t)$ and $\delta(t)/\gamma(t)$ are small. In this limit, one can also prove that the term $F_k(t)$ in the wave function of Eq. (86) becomes much larger than one. Technically, this is the reason for which Eq. (93) holds.

The condition of semi-classicality of Eq. (93) is a bit heavy to write down and explain. Switching from the Schrödinger to the Heisenberg representation, one can write down a nearly equivalent condition

$$\langle \chi_k | \left\{ \hat{\chi}_k, \hat{p}^\dagger_{\chi\, k} \right\} | \chi_k \rangle \gg \left| \langle \chi_k | \left[\hat{\chi}_k, \hat{p}^\dagger_{\chi\, k} \right] | \chi_k \rangle \right| \tag{94}$$

where ($\hat{\chi}_k$, $\hat{p}_{\chi\, k}$) are the time-dependent position and momentum operators. Equation (94) says that the semi-classical transition takes place when, in absolute value, the mean value of the anti-commutator becomes much larger than that of the commutator, which is equal to one (in units where $\hbar = 1$). Again, this condition says that the mean value of a product of operators can be effectively computed as if the operators did commute. Finally, note that if we introduce the *instantaneous number-of-particle operator* $\hat{N} = \int d^3k\, \hat{a}_k^\dagger \hat{a}_k$, where $\hat{a}_k$ is the annihilation operator at a given time, the previous condition is equivalent to $\langle \hat{N} \rangle \gg 1$. Hence, in the literature, it is often mentioned that the condition for semi-classicality is to get a number of particles (created from the vacuum) much larger than one, as is the case for inflationary perturbations after Hubble crossing.

The result of these considerations is that we can almost forget that we are dealing with quantum perturbations, excepted when we specify the normalization of the mode function, which is related to the positive frequency condition and to the canonical commutation relation. However, at late time, we can consider inflationary perturbations as stochastic, Gaussian variables with variance $|\chi_k(t)|$, where $\chi_k(t)$ is the mode function.

2.5 Tensor perturbations

2.5.1 General equations

We have seen that tensor perturbations are described by the two degrees of polarization h_1, h_2, with $h_{ij} = h_1 e^1_{ij} + h_2 e^2_{ij}$ and $\sum_{ij\lambda} e^\lambda_{ij} e^\lambda_{ij} = 1$. At first order in perturbations, the tensor and scalar sectors do not mix with each other, and the Lagrangian for (h_1, h_2) is contained in

$$\mathcal{L}_{grav} = \sqrt{|g|}\frac{R}{16\pi\mathcal{G}} \ . \tag{95}$$

In this section and in the next one, we will use conformal time η instead of proper (cosmological) time t. The two are related by the change of variable $dt = a d\eta$ which renders the metric conformally invariant, $ds^2 = a^2(\eta)[d\eta^2 - d\vec{x}^2]$. Then, $\sqrt{|g|} = a^4$ and the part of the Lagrangian describing tensor perturbations reads

$$\mathcal{L}_{tensors} = \frac{a^4}{16\pi\mathcal{G}}\left[\frac{1}{4}\partial_\mu h_1 \partial^\mu h_1 + \frac{1}{4}\partial_\mu h_2 \partial^\mu h_2 + \mathrm{d}iv\right] \tag{96}$$

where "div" stands for irrelevant total divergence terms[4]. We see from the Lagrangian that h_1 and h_2 share the same Lagrangian and equations of motion. After quantization and Hubble crossing, they will be two stochastic Gaussian variables with equal variance. So, we can forget one of them and concentrate on a single mode h_λ, provided that in the final gravitational wave power spectrum we do not forget to take into account both degrees of freedom. We could try to quantize directly h_λ but for pedagogical purposes we will change variable and introduce a rescaled field

$$y \equiv \frac{a\, h_\lambda}{\sqrt{32\pi\mathcal{G}}} \tag{97}$$

which has a canonically normalized kinetic terms. Indeed, the Lagrangian for y reads

$$\mathcal{L}_y = \frac{1}{2}\left\{\left(y' - \frac{a'}{a}y\right)^2 - (\partial_i y)^2\right\} \tag{98}$$

and starts with the usual term $\frac{1}{2}y'^2$ (the prime stands for the derivative with respect to conformal time). The conjugate momentum reads

$$p_y = \frac{\partial \mathcal{L}_y}{\partial y'} = y' - \frac{a'}{a}y \tag{99}$$

and the Hamiltonian is found to be

$$H = \frac{1}{2}\int d^3x \left[\left(y' - \frac{a'}{a}y\right)^2 + (\partial_i y)^2\right] \ . \tag{100}$$

In Fourier space the equation of motion reads

$$y_k'' + \left(k^2 - \frac{a''}{a}\right) y_k = 0 \tag{101}$$

and the Hamiltonian becomes

$$H = \frac{1}{2}\int d^3k \left[\left|y_k' - \frac{a'}{a}y_k\right|^2 + k^2 |y_k|^2\right] \ . \tag{102}$$

It is not a surprise to see that in the small wavelength limit $k \gg aH$ the Hamiltonian reduces to its flat space-time counterpart

$$H = \frac{1}{2}\int d^3k \left[|y_k'|^2 + k^2 |y_k|^2\right] \ . \tag{103}$$

This matches our expectation that modes deep inside the Hubble radius do not see the space-time curvature, and can be quantized in the usual way. In this limit the mode function is the one found in Eq. (85),

$$y_{k\gg aH} = \sqrt{\frac{1}{2k}} e^{-ik\eta} \ . \tag{104}$$

[4] Note that the authors who define the polarization tensors as orthonormal, $\sum_{ij\lambda} e^\lambda_{ij} e^\lambda_{ij} = 2$, find a factor 1/2 instead of 1/4 in the brackets of Eq. (96): i.e., they find canonically normalized kinetic terms for h_1 and h_2 in the Ricci scalar R.

2.5.2 Solution during De-Sitter stage

In the limit of an exact De Sitter stage with $a(t) = e^{H_i t}$, the relation between a and conformal time can be found by integrating over $d\eta = e^{-H_i t} dt$: one gets $\eta = -1/(aH_i)$. During the De Sitter stage, a goes from very small to very large values: this corresponds to η running from $-\infty$ to zero. Hubble crossing during inflation corresponds to $k \simeq aH_i$, i.e. to $k\eta \simeq -1$. Then the equation of motion becomes

$$y_k'' + \left(k^2 - \frac{2}{\eta}\right) y_k = 0 \tag{105}$$

with two solutions

$$y_k = A_k \left(1 - \frac{i}{k\eta}\right) e^{-ik\eta} + B_k \left(1 + \frac{i}{k\eta}\right) e^{ik\eta} \ . \tag{106}$$

A matching with Eq. (104) shows that the mode function reads

$$y_k = \sqrt{\frac{1}{2k}} \left(1 - \frac{i}{k\eta}\right) e^{-ik\eta} \ . \tag{107}$$

In the large wavelength limit $k \ll aH$ corresponding to $k\eta \to 0$, we get

$$y_{k \ll aH} = -\frac{i}{\sqrt{2k^3}\eta} = \frac{iaH}{\sqrt{2k^3}} \ . \tag{108}$$

For the variable h_λ, this gives

$$h_{\lambda\,k} = \frac{\sqrt{32\pi\mathcal{G}}}{a} \frac{iaH}{\sqrt{2k^3}} = iH\sqrt{\frac{16\pi\mathcal{G}}{k^3}} \ . \tag{109}$$

We have seen that the modulus of the mode function can be interpreted as the Gaussian variance of the classical stochastic mode after Hubble crossing. So, for $k \ll aH$, each degree of polarization has a squared variance

$$\langle |h_{\lambda\,k}|^2 \rangle = \frac{16\pi\mathcal{G}H_i^2}{k^3} \ . \tag{110}$$

Finally, the total squared variance of the gravitational wave tensor h_{ij} at that time is given by

$$\left\langle \left| \sum_{ij} h_{ij\,k} \right|^2 \right\rangle = \langle |h_{1\,k}|^2 \rangle \sum_{ij} e^1_{ij} e^1_{ij} + \langle |h_{2\,k}|^2 \rangle \sum_{ij} e^2_{ij} e^2_{ij} = \langle |h_{\lambda\,k}|^2 \rangle \left(\frac{1}{2} + \frac{1}{2}\right) = \frac{16\pi\mathcal{G}H_i^2}{k^3} \ . \tag{111}$$

2.5.3 Long-wavelength solution during and after inflation

The background of gravitational waves generated during inflation could be observed through its contribution to CMB anisotropies, or maybe directly in future space-based gravitational wave detectors. The evolution of gravitational waves after they re-enter inside the Hubble radius during radiation, matter or dark energy domination can be easily computed. The initial condition which is needed is the amplitude, or more precisely the Gaussian variance of gravitational waves on super-Hubble scales, for instance during radiation domination.

The evolution of super-Hubble modes during any stage (inflation, radiation domination, matter domination, etc.) is actually trivial since in the long-wavelength limit $k \longrightarrow 0$ the equation of motion (101) reduces to

$$y_k'' - \frac{a''}{a} y_k = 0 \tag{112}$$

with an obvious solution $y_k \propto a$ corresponding to a constant $h_{\lambda\,k}$. The other solution $y_k = a \int d\eta/a^2$ is a decaying mode, so we conclude that $h_{\lambda\,k}$ is approximately constant on super-Hubble modes during any stage [5].

[5] One could fear that this reasoning does not hold during radiation domination, for which $a(t) \propto t^{1/2}$, leading to $a(\eta) \propto \eta$ and $a'' = 0$. However in that case the general solutions reads $y_k = A_k \cos k\eta + B_k \sin k\eta$. In the long-wavelength limit this is equal to $y_k = A_k + B_k k\eta$ and a matching with the solution during inflation gives $A_k = 0$. So, here again, $y_k \propto a$ and $h_{\lambda\,k}$ is approximately constant.

2.5.4 Primordial spectrum of tensor perturbations

Since $h_{\lambda\, k}$ is constant on super-Hubble scales, the primordial spectrum needed for computing observable quantities can be evaluated at the end of inflation, and is given by Eq. (111) in the limit of exact De Sitter inflation with a Hubble parameter equal to H_i (we will study in subsection 2.7 how this result can be generalized to quasi-De Sitter inflation). Defining the power spectrum as the squared variance multiplied by k^3, we obtain

$$\mathcal{P}_h \equiv k^3 \langle | \sum_{ij} h_{ij\, k}|^2 \rangle = 16\pi \mathcal{G} H_i^2 \ . \tag{113}$$

We see that in this approximation $\mathcal{P}_h$ does not depend on k: such a power spectrum is called *scale-invariant*. For historical reasons, a scale-invariant spectrum is also called a Harrison-Zel'dovitch spectrum. We can also express this primordial power spectrum in terms of the scalar potential during inflation,

$$\mathcal{P}_h = \frac{2}{3}(8\pi\mathcal{G})^2 V_i \ . \tag{114}$$

Alternative notations. Different authors adopt different definitions for the power spectrum of a stochastic quantity X_k. Throughout the literature, one finds for instance $\mathcal{P}_X = \langle |X_k|^2 \rangle$, $\mathcal{P}_X = (2\pi)^{-3} \langle |X_k|^2 \rangle$ (as for instance in some Bertschinger and collaborator papers) or $\mathcal{P}_X = k^3 \langle |X_k|^2 \rangle$ (as in this course). However, one of the most popular conventions for defining the tensor power spectrum is that of Liddle and Lyth (that we will note with a tilde)

$$\tilde{\mathcal{P}}_h \equiv \frac{2k^3}{2\pi^2} \langle | \sum_{ij} h_{ij\, k}|^2 \rangle = \frac{1}{\pi^2} \mathcal{P}_h = \frac{16 \mathcal{G} H_i^2}{\pi}. \tag{115}$$

2.6 Scalar perturbations

2.6.1 General equations

In the tensor case, we quantized two degrees of freedom – the two degrees of polarization. For scalar perturbations and in the longitudinal (Newtonian) gauge we have three scalar variables: the diagonal metric perturbations ϕ and ψ, and the inflaton field perturbation $\delta\varphi$. Do we need to quantize these three fields independently from each other? The answer is suggested by the equations of motion. The Einstein equations provide a relation

$$\delta G_i^j = \partial_i \partial^j (\phi - \psi) = 8\pi\mathcal{G} \ \partial_i \delta\varphi \ \partial^j \delta\varphi \tag{116}$$

for $i \neq j$. The right hand-side vanishes at first order in perturbations. So $\phi - \psi$ is either null or a quadratic function of the spatial coordinate. The latter is impossible: far from the origin $\phi - \psi$ would become arbitrarily large, and the metric perturbations would not be small anymore. We conclude that at first order in perturbation theory there is an identity $\phi = \psi$. We are left with only two variables ϕ and $\delta\varphi$ evolving according to the Klein Gordon equation and Einstein equations,

$$\begin{aligned} \delta\ddot{\varphi}_k + 3H\, \delta\dot{\varphi}_k + \left(\frac{k^2}{a^2} + \frac{\partial^2 V}{\partial \varphi^2}(\bar{\varphi}) \right) \delta\varphi_k &= 4\dot{\bar{\varphi}}\, \dot{\phi}_k - 2\frac{\partial V}{\partial \varphi}(\bar{\varphi})\, \phi_k \ , & (117) \\ \dot{\phi}_k + H\phi_k &= 4\pi\mathcal{G}\dot{\bar{\varphi}}\, \delta\varphi_k \ . & (118) \end{aligned}$$

It is clear that the first equation is an equation of propagation, while the second is only a constraint equation. This confirms the intuition that unlike gravitational waves, scalar metric perturbations are not additional independent fields: they just follow matter perturbations, like the gravitational potential in Newtonian gravity. In absence of matter perturbations, scalar metric perturbations would simply vanish, unlike gravitational waves. Therefore, we only need to quantize one independent degree of freedom. It is actually possible to combine the Einstein equations in order to find a second-order differential equation involving a single master variable. This master variable is not unique. For instance, it is possible to write a second-order differential equation for the metric perturbation only:

$$\ddot{\phi}_k + \left(H - 2\frac{\ddot{\bar{\varphi}}}{\dot{\bar{\varphi}}} \right) \dot{\phi}_k - \left(2H\frac{\ddot{\bar{\varphi}}}{\dot{\bar{\varphi}}} + 8\pi\mathcal{G}\dot{\bar{\varphi}}^2 - \frac{k^2}{a^2} \right) \phi_k = 0 \ . \tag{119}$$

This equation is rather complicated. Actually we can find a much simpler one for a master variable called the Mukhanov variable. It was shown by Mukhanov that metric and scalar field perturbations can be combined into a gauge-invariant quantity which reduces, when written in the Newtonian gauge, to

$$\xi = a\, \delta\varphi + \frac{\bar{\varphi}'}{H}\phi \tag{120}$$

(we recall that the prime represents derivation with respect to conformal time). In addition, Mukhanov found that in any gauge, the part of the action (41) describing first-order scalar perturbations is entirely contained inside the Lagrangian

$$\mathcal{L}_\xi = \frac{1}{2}\left(\xi'^2 - (\partial_i \xi)^2 + \frac{z''}{z}\xi^2 + \mathrm{d}iv\right) \tag{121}$$

where "div" stands for irrelevant total divergence terms, and

$$z \equiv \frac{\bar{\varphi}'}{H} \ . \tag{122}$$

The conjugate momentum reads

$$p_\xi = \frac{\partial \mathcal{L}_\xi}{\partial \xi'} = \xi' \tag{123}$$

and the Hamiltonian is found to be

$$H = \frac{1}{2}\int d^3x \left[\xi'^2 + (\partial_i \xi)^2 - \frac{z''}{z}\xi^2\right] \ . \tag{124}$$

In Fourier space the equation of motion of the Mukhanov variable reads

$$\xi_k'' + \left(k^2 - \frac{z''}{z}\right)\xi_k = 0 \tag{125}$$

and the Hamiltonian becomes

$$H = \frac{1}{2}\int d^3k \left[{\xi_k'}^2 + \left(k^2 - \frac{z''}{z}\right)\xi^2\right] \ . \tag{126}$$

2.6.2 Solution during Quasi-De-Sitter stage

We could try to proceed like for tensor perturbations, and to quantize ξ in the limit of an exact De Sitter stage, i.e. exponential expansion. Strictly speaking this requires $\dot{H} = 0$ and therefore $\bar{\varphi}' = \dot{\bar{\varphi}} = 0$. This limit is actually singular since $z = 0$. The quantization of scalar perturbations in an exact De-Sitter space is not difficult, but it requires a special treatment. Anyway, the outcome is that in this limit the amplitude of metric perturbations $\langle|\phi|^2\rangle$ is damped at the end of inflation (there is only a decaying mode). This is excluded by the actual observation of density perturbations in our universe. In other words, the mechanism for generating scalar perturbations during inflation requires a small but non-zero derivative $\dot{\bar{\varphi}}$.

So, first-order calculations of the scalar power spectrum are based on the assumption that over the relevant range of time (a few e-folds before and after horizon crossing) the two quantities H and $\dot{\bar{\varphi}}$ are approximately constant and equal to H_i, $\dot{\bar{\varphi}}_i$. In this approximation, z is proportional to a, and the relation $\eta = -1/(aH_i)$ remains true. So $z''/z = a''/a = -2/\eta^2 = -2(aH_i)^2$, and in the sub-Hubble limit $k \gg aH_i$ the Hamiltonian reduces as expected to its flat space-time counterpart. The equation of motion for ξ,

$$\xi_k'' + \left(k^2 - \frac{a''}{a}\right)\xi_k = 0 \ , \tag{127}$$

is the same as that for y in the previous section, and therefore we know that the mode function reads

$$\xi_k = \sqrt{\frac{1}{2k}}\left(1 - \frac{i}{k\eta}\right)e^{-ik\eta} \ . \tag{128}$$

Let us go back to the variable ϕ, which is very useful for relating scalar perturbations during and after inflation (since ϕ does not decay like the inflaton field). By combining Eq. (120) with the Einstein equation, one obtains the exact differential relation

$$\phi_k = -\frac{4\pi\mathcal{G}}{k^2}\frac{\bar{\varphi}'^2}{aH}\left(\frac{H}{\bar{\varphi}'}\xi_k\right)' \ . \tag{129}$$

Inserting the mode function for ξ into this relation, we find that the mode function for ϕ reads (at any time)

$$\phi_k = \frac{4\pi\mathcal{G}}{\sqrt{2k^3}}\,\dot{\bar{\varphi}}_i\, ie^{-ik\eta} \ . \tag{130}$$

So, in the large wavelength limit $k \ll aH$ corresponding to $k\eta \to 0$, we get

$$\phi_{k \ll aH} = \frac{4\pi\mathcal{G}}{\sqrt{2k^3}}\,\dot{\bar{\varphi}}_i\, i\ . \tag{131}$$

We have seen that the modulus of the mode function can be interpreted as the Gaussian variance of the classical stochastic mode after Hubble crossing. So, the squared variance of the the metric perturbation reads

$$\langle|\phi_k|^2\rangle = \frac{(4\pi\mathcal{G})^2}{2k^3}\,\dot{\bar{\varphi}}_i^2\ . \tag{132}$$

2.6.3 Long-wavelength solution during and after inflation

It is possible to show that during any stage (inflation, radiation domination, matter domination, dark energy domination) the metric perturbation ϕ is described by the following exact analytical solution in the long-wavelength limit $k \longrightarrow 0$

$$\phi_k = C_1(k)\left(1 - \frac{H}{a}\int^t a(t)\, dt\right) - C_2(k)\frac{4\pi\mathcal{G}}{k^2}\frac{H}{a} \tag{133}$$

where C_1, C_2 are respectively the coefficient of the non-decaying and decaying modes. Changing the lower bound of the integral amounts in absorbing some of the decaying mode into the growing mode (in practice this minor ambiguity is unimportant). The integration can be performed explicitly during each stage:

- during inflation, in the exact De Sitter limit we find that the non-decaying mode vanishes: ϕ_k decays like H/a. So, in order to catch the dominant contribution to the growing mode, we need to go to the next order and approximate $H(t)$ near some time t_i as $H_i + \dot{H}_i(t - t_i)$, where H_i and $\dot{H}_i$ are two constants. The relation $H = d\ln a/dt$ implies

$$\begin{aligned} a(t) &= a(t_i)\exp\left(\int_{t_i}^t H(t)dt\right) && (134)\\ &\simeq a(t_i)\exp\left(\int_{t_i}^t (H_i + \dot{H}_i t)dt\right) && (135)\\ &= a(t_i)\exp\left(H_i(t-t_i) + \dot{H}_i\frac{(t^2 - t_i^2)}{2}\right) && (136)\end{aligned}$$

  Exercise : plug this expression into Eq.(133), and compute the result at first order in the small parameter $\dot{H}_i/H_i^2$. Show that

$$\phi_k = -C_1(k)\frac{\dot{H}_i}{H_i^2}\ . \tag{137}$$

  This shows that at any time during slow-roll inflation, the non-decaying mode is slowly-varying (since $\dot{H}(t)$ and $H(t)$ are slowly-varying), with a dominant contribution given by

$$\phi_k = -C_1(k)\frac{\dot{H}(t)}{H(t)}\ . \tag{138}$$

- during Radiation Domination (RD), $a(t) \propto t^{1/2}$ and the non-decaying mode reads

$$\phi_k = C_1(k)\left(1 - \frac{1}{2}\times\frac{2}{3}\right) = \frac{2}{3}C_1(k)\ . \tag{139}$$

  So, during radiation domination, ϕ_k is constant on super-Hubble scales.

- during Matter Domination (MD), $a(t) \propto t^{2/3}$ and the non-decaying mode reads

$$\phi_k = C_1(k)\left(1 - \frac{2}{3}\times\frac{3}{5}\right) = \frac{3}{5}C_1(k)\ . \tag{140}$$

  This shows that at the time of equality ϕ_k is damped by a factor $(3/5)/(2/3) = 9/10$, and then becomes constant again on super-Hubble scales.

- during dark energy domination ϕ_k is damped on super-Hubble scales, but we do not need to compute its detailed evolution here.

2.6.4 Primordial spectrum of scalar perturbations

We can deduce from Eqs. (137-140) that the variance of ϕ_k on super-Hubble scales during radiation domination is equal to the variance during inflation times a factor $2H_i^2/(3\dot{H}_i)$. Similarly, during matter domination, the factor is $3H_i^2/(5\dot{H}_i)$. Most author define the primordial spectrum of ϕ as the squared variance of the modes on super-horizon scales during matter domination. This is a bit misleading because at the time of equality, many observable modes are already inside the Hubble radius. However, for those modes, we can formally define the power spectrum as the amplitude on super-horizon scales during radiation domination times the factor (9/10), and obtain

$$\mathcal{P}_\phi^{\mathrm{MD}} \equiv k^3 \langle |\phi_k|^2 \rangle_{MD} = k^3 \left(\frac{3}{5} \frac{H_i^2}{\dot{H}_i} \right)^2 \langle |\phi_k|^2 \rangle_i = k^3 \left(\frac{3}{5} \frac{H_i^2}{\dot{H}_i} \frac{4\pi\mathcal{G}}{\sqrt{2k^3}} \dot{\bar{\varphi}}_i \right)^2 = \frac{9H_i^4}{50\dot{\bar{\varphi}}_i^2} \tag{141}$$

where we employed Eqs. (132) and (53). We see that in this approximation $\mathcal{P}_\phi^{MD}$ does not depend on k: it is a *scale-invariant* or Harrison-Zel'dovitch spectrum. With the help of Eqs. (57, 58), we can re-express the primordial spectrum in terms of the scalar potential during inflation,

$$\mathcal{P}_\phi^{MD} = \frac{3}{50} (8\pi\mathcal{G})^3 \frac{V_i^3}{(\partial V/\partial\varphi)_i^2} \ . \tag{142}$$

Using the definition (61) of the slow-roll parameter ϵ, that we will call ϵ_i in order to show explicitly that it is computed assuming constant values H_i and $\dot{\bar{\varphi}}_i$, we obtain

$$\mathcal{P}_\phi^{MD} = \frac{3}{100} (8\pi\mathcal{G})^2 \frac{V_i}{\epsilon_i} \ . \tag{143}$$

Note that the ratio of the tensor spectrum over the scalar one reads

$$\frac{\mathcal{P}_h}{\mathcal{P}_\phi^{\mathrm{MD}}} = \frac{200}{9} \epsilon_i \tag{144}$$

(if we were using the scalar power spectrum during radiation domination, $\mathcal{P}_\phi^{\mathrm{R}D} = (10/9)^2 \mathcal{P}_\phi^{\mathrm{M}D}$, we would find instead a ratio $\mathcal{P}_h/\mathcal{P}_\phi^{\mathrm{R}D} = 18\epsilon_i$).

Spectrum of curvature perturbations. Some authors prefer to parametrize the scalar perturbations not with the scalar metric perturbations ϕ (which is called the Bardeen potential or the generalized gravitational potential), but instead with the curvature perturbation $\mathcal{R}$, defined as the perturbation of the spatial curvature radius $R^{(3)}$ on comoving hypersurfaces. The inconvenient of $\mathcal{R}$ is that its definition is a bit complicated (it is related to ϕ through a differential equation). Its advantage is that as long as the universe does not contain isocurvature perturbations, and in particular, in the standard cosmological picture with single-field inflation, the long-wavelength solution for the curvature perturbation is $\dot{\mathcal{R}} = 0$: the curvature is a conserved quantity for super-Hubble adiabatic perturbations, while ϕ is not. However, during the radiation dominated era, we have seen that ϕ is constant, and at that time we can derive a proportionality factor $\mathcal{R} = (3/2)\, \phi^{\mathrm{R}D}$. The Bardeen potential is also constant during the matter dominated era, and then $\mathcal{R} = (5/3)\, \phi^{\mathrm{M}D}$. This gives

$$\mathcal{P}_\mathcal{R} = \left(\frac{5}{3} \right)^2 \mathcal{P}_\phi^{MD} = \frac{1}{12} (8\pi\mathcal{G})^2 \frac{V_i}{\epsilon_i} = \frac{2\pi\mathcal{G}H_i^2}{\epsilon_i} \tag{145}$$

and a tensor-to-curvature ratio

$$\frac{\mathcal{P}_h}{\mathcal{P}_\mathcal{R}} = 8\epsilon_i \ . \tag{146}$$

However, using the definitions of Lyth and Liddle for the power spectra,

$$\tilde{\mathcal{P}}_\mathcal{R} \equiv \frac{1}{2\pi^2} \mathcal{P}_\mathcal{R} \ , \qquad \tilde{\mathcal{P}}_h \equiv \frac{2}{2\pi^2} \mathcal{P}_h \ , \tag{147}$$

one gets a ratio that we will call r,

$$r \equiv \frac{\tilde{\mathcal{P}}_h}{\tilde{\mathcal{P}}_\mathcal{R}} = 16\epsilon_i \ . \tag{148}$$

This parameter r is the most common way of defining the tensor-to-scalar ratio, and we will use it abundantly when comparing inflationary models with observations.

2.7 Computing smooth spectra using a slow-roll expansion

So far, we computed the tensor spectrum assuming an exact De Sitter stage with a constant H, and the scalar spectrum assuming a quasi-De Sitter stage with constant H and $\dot{\bar{\varphi}}$ (implying a constant $\dot{H}$, which is not in contradiction with a constant H since during slow-roll $|\dot{H}| \ll H^2$). These assumptions are not unreasonable, because the details of scalar and tensor spectra are only sensitive to a small number of e-folds, during which slow-roll condition guarantee that variations in the background quantities are very small. Let us try to understand how many e-folds are typically relevant.

As long as $k \gg aH$, the mode function ξ_k (for scalars) and y_k (for tensors) are oscillatory, with a fixed amplitude $(2k)^{1/2}$ which does *not* depend on the background dynamics. Any variation in H and $\dot{H}$ at that time have no consequence for the mode function amplitude. In the opposite regime $k \ll aH$, we can match the mode function with an analytical long-wavelength approximation which consistently takes into account any variation in the background until the end of inflation and even later. For scalars, we did such a matching for ϕ_k using the solution of Eq. (133). If instead we were using the curvature perturbations we would just use the fact that $\mathcal{R}_k$ is time-invariant as soon as $k \ll aH$. For tensors we used the time-invariance of $h_{\lambda\, k}$ for $k \ll aH$.

So, the details of the primordial spectrum depend only on the small range of time during which the modulus of $(\xi_k,\ y_k)$ departs from $(2k)^{1/2}$ *and* Eq. (133), $\dot{\mathcal{R}}_k = 0$ and $\dot{h}_{\lambda\, k} = 0$ are not yet good approximations. This happens when k and aH have the same order of magnitude, typically between $k = 10\, aH$ and $k = 0.1\, aH$. During this time interval, the scale factor grows approximately by a factor 100, which corresponds to a number of e-folds

$$\Delta N = \Delta \ln a = \ln 100 \simeq 4.6\ . \tag{149}$$

So, the results that we already obtained are valid provided that the relative variation of H (and also of $\dot{H}$ for scalars) is negligible over typically five e-folds.

If we want to go beyond this approximation, but still avoiding the resolution of the exact mode function equations (101,125), we can work in a *slow-roll expansion scheme*. This consists in Taylor-expanding one background function, for instance $H(t)$, $H(a)$, $H(N)$, $\bar{\varphi}(t)$ or $V(\bar{\varphi})$ around a pivot value (that we will note t_*, or a_*, or N_*, or $\bar{\varphi}_*$). The expansion is cut at a given order, and the mode function equation is solved at this order, assuming that higher-order derivatives vanish. The approximation is optimal if the pivot value corresponds to the time at which a typical observable wavelength k_* (chosen to be roughly in the middle of the range probed by CMB and LSS experiments) crosses the Hubble radius during inflation: so, t_* is the time at which $k_* = a_* H_*$.

Technically, solving a second-order equation at a given order in the time-derivative of its coefficients can be done following the so-called WKB approximation. For instance, the first-order WKB solution of the equation $\ddot{x}+\omega(t)^2 x$ is given by $x \propto \omega(t)^{-1/2} \exp(\pm i\omega(t)t)$, and is a good approximation whenever the time-variation of $\omega(t)$ over one period of oscillation is negligible. Higher-order WKB solutions include higher order time-derivatives of $\omega(t)$. This scheme provides an appropriate way to compute the inflationary power spectra at a given order in slow-roll. However, if we want to extend our previous results to the next order in slow-roll (i.e., for tensors, keeping term in $\dot{H}$, and for scalars in $\ddot{H}$), we shall see that the correct results can be obtained in a much simpler way.

Since at first order the solutions for $\mathcal{P}_h$ and $\mathcal{P}_\phi$ are scale-invariant, the second order solution will provide the approximate value of the *scalar and tensor tilts*, defined as

$$n_t \equiv \left.\frac{d \ln \mathcal{P}_h}{d \ln k}\right|_{k=k*} , \tag{150}$$

$$n_s - 1 \equiv \left.\frac{d \ln \mathcal{P}_\phi}{d \ln k}\right|_{k=k*} . \tag{151}$$

Note that for a question of habits, the definition of n_t and n_s differ by a "minus one" term, so that a scale-invariant scalar spectrum corresponds to $n_s = 1$, while a scale-invariant tensor spectrum corresponds to $n_t = 0$. If at first order the spectra are scale-invariant, at second order the tilts are scale invariant and the power spectra are exact power laws,

$$\mathcal{P}_h(k) = \mathcal{P}_h(k_*) \left(\frac{k}{k_*}\right)^{n_t} , \tag{152}$$

$$\mathcal{P}_\phi(k) = \mathcal{P}_\Phi(k_*) \left(\frac{k}{k_*}\right)^{n_s-1} . \tag{153}$$

In order to compute the running of the tilts with k, one would need to go to the third order, i.e. keeping $\ddot{H}$ terms for tensors and (d^3H/dt^3) terms for scalars. If instead of reasoning on $H(t)$, we define the slow-roll expansion using $\bar{\varphi}(t)$ or $V(\bar{\varphi})$, Eqs. (53, 57, 58) show that it is equivalent to assume a constant H, $\bar{\varphi}$ or V; as well as to assume a constant $\dot{H}$, or $\dot{\bar{\varphi}}$, or $(\partial V/\partial\varphi)$. Finally it is straightforward to show that it is equivalent to stop the slow-roll expansion at a given order n in (d^nH/dt^n), $(d^n\bar{\varphi}/dt^n)$ or $(\partial^n V/\partial\varphi^n)$.

Simple method for computing the tilts. Let us now try to compute the tilts by starting from the lowest-order solution

$$
\begin{aligned}
\mathcal{P}_h(k) &= 16\pi\mathcal{G}H^2 \ , &(154)\\
\mathcal{P}_\phi(k) &= \frac{9H^4}{5\dot{\bar{\varphi}}^2} \ , &(155)
\end{aligned}
$$

that we obtained assuming that H and $\dot{\bar{\varphi}}$ were constant during inflation. Now, we will assume that these formulas also apply to the case where H and $\dot{\bar{\varphi}}$ are slightly varying, with the right-hand side evaluated precisely at the time at which $k = aH$. Given the previous discussion, this assumption sounds reasonable, but the fact that it is sufficient for computing the tilts at first order in slow-roll (as we will do in the following lines) is not something that we will prove rigorously[6]. We therefore admit that

$$
\begin{aligned}
\mathcal{P}_h(k) &= 16\pi\mathcal{G}H^2_{k=aH} \ , &(156)\\
\mathcal{P}_\phi(k) &= \frac{9H^4_{k=aH}}{5\dot{\bar{\varphi}}^2_{k=aH}} \ . &(157)
\end{aligned}
$$

Then, the tilts can be computed simply from the finite difference between $(H, \dot{\bar{\varphi}})$ at the time when k equals the pivot scale $k_* = a_*H_*$ and the time when it equals

$$
k_* + dk = (a_* + da)(H_* + dH) = a_*H_* + a_*\left(1 + \frac{H_*^2}{\dot{H}_*}\right) dH \ . \tag{158}
$$

During slow-roll, we have seen that $|\dot{H}| \ll H^2$, and we conclude that

$$
dk \simeq \frac{a_*H_*^2}{\dot{H}_*} dH \ . \tag{159}
$$

Tensor tilt. By definition, n_t reads

$$
n_t \simeq \frac{\ln \mathcal{P}_h(k_* + dk) - \ln \mathcal{P}_h(k_*)}{\ln(k_* + dk) - \ln(k_*)} \simeq \frac{\ln(H_* + dH)^2 - \ln H_*^2}{(dk/k_*)} \simeq \frac{2k_*\, dH}{H_*\, dk} \ . \tag{160}
$$

Using now Eq. (159), we obtain

$$
n_t \simeq \frac{2k_*\, \dot{H}_*}{a_*H_*^3} \ . \tag{161}
$$

Using the fact that $k_* = a_*H_*$, we finally get

$$
n_t \simeq \frac{2\dot{H}_*}{H_*^2} \ . \tag{162}
$$

This result can also be expressed as a function of the potential derivatives using the slow-roll relations (57, 58) and the exact equation (53),

$$
n_t \simeq -\frac{(\partial V/\partial\varphi)_*^2}{8\pi\mathcal{G}V_*^2} = -2\epsilon_* \ . \tag{163}
$$

Scalar tilt. We follow the same method for n_s which reads

$$
n_s - 1 = \left.\frac{d\ln \mathcal{P}_\phi}{d\ln k}\right|_{k_*} = \left.\frac{d\ln(H^4/\dot{\bar{\varphi}}^2)}{d\ln k}\right|_{k_*} = 4\left.\frac{d\ln H}{d\ln k}\right|_{k_*} - 2\left.\frac{d\ln\dot{\bar{\varphi}}}{d\ln k}\right|_{k_*} \ . \tag{164}
$$

[6] The rigorous proof follows from a WKB resolution scheme.

We already evaluated the first term for the tensor tilt and found

$$\left.\frac{d\ln H}{d\ln k}\right|_{k_*} = \frac{\dot{H}_*}{H_*^2} \ . \tag{165}$$

For the second term, we factorize out $(d\ln H/d\ln k)$ and obtain

$$\left.\frac{d\ln\dot{\bar\varphi}}{d\ln k}\right|_{k_*} = \left.\frac{d\ln H}{d\ln k}\right|_{k_*} \left.\frac{d\ln\dot{\bar\varphi}}{d\ln H}\right|_{H_*} = \left(\frac{\dot{H}_*}{H_*^2}\right) \times \left(\frac{H_*\ddot{\bar\varphi}_*}{\dot{H}_*\dot{\bar\varphi}_*}\right) = \frac{\ddot{\bar\varphi}_*}{H_*\dot{\bar\varphi}_*} \ . \tag{166}$$

The final result is then

$$n_s - 1 = 4\frac{\dot{H}_*}{H_*^2} - 2\frac{\ddot{\bar\varphi}_*}{H_*\dot{\bar\varphi}_*} \ , \tag{167}$$

and like for the tensor tilt, we can express it in function of potential derivatives, using mainly the slow-roll relation $\dot{\bar\varphi} = -(\partial V/\partial\varphi)/3H$ and its time-derivative

$$\ddot{\bar\varphi} = -\frac{(\partial^2 V/\partial\varphi^2)\dot{\bar\varphi}}{3H} + \frac{(\partial V/\partial\varphi)\dot{H}}{3H^2} \ . \tag{168}$$

Then, we obtain

$$n_s - 1 \ = \ 6\frac{\dot{H}_*}{H_*^2} + \frac{2(\partial^2 V/\partial\varphi^2)_*}{3H_*^2} \ = \ 6\frac{\dot{H}_*}{H_*^2} + \frac{2(\partial^2 V/\partial\varphi^2)_*}{8\pi\mathcal{G}V_*} \ = \ -6\epsilon_* + 2\eta_* \ . \tag{169}$$

Alternative slow-roll expansions. We presented the lowest-order results for n_s and n_t in a slow-roll expansion based on $V(\varphi)$ [Eqs (163), (169)]. Another popular expansion scheme nowadays is that in Hubble Slow-Roll (HSR) parameters defined as

$$\epsilon_0 \equiv \frac{H(N_i)}{H(N)} \ , \qquad \epsilon_n \equiv \frac{d\ln\epsilon_{n-1}}{dN} \ , \tag{170}$$

where N_i is the e-fold number, e.g., at the beginning of inflation (anyway, the normalization of ϵ_0 is unimportant). In this scheme, the first slow-roll parameter reads

$$\epsilon_1 = -\frac{d\ln H}{dN} = -\frac{d\ln H}{d\ln a} = -\frac{\dot{H}}{H^2} = \epsilon \ . \tag{171}$$

However, the second slow-roll parameter differs from η, since a short computation gives

$$\epsilon_2 = \frac{d\ln\epsilon_1}{dN} = \frac{1}{4\pi\mathcal{G}}\left[\left(\frac{\partial V/\partial\varphi}{V}\right)^2 - \frac{\partial^2 V/\partial\varphi^2}{V}\right] = -2\eta + 4\epsilon \ . \tag{172}$$

Many authors use these parameters, writing sometimes (ϵ_1, ϵ_2) as (ϵ, η). The scalar tilt as a function of HSR parameters reads

$$n_s = 1 - 2\epsilon_{1*} - \epsilon_{2*} \ . \tag{173}$$

In this course, we will not present results at the next order in slow-roll, and refer to the specialized literature for finding e.g. the expression of tilt running.

2.8 Computing Broken Scale Invariance spectra with analytical/numerical methods

For all single-field models such that slow-roll conditions are well-satisfied during the "observable e-folds", i.e. when observable modes cross the Hubble radius, the previous computation is sufficient for accurately describing the power spectrum and comparing with observations. Note that in many models, the slow-roll parameters increase with time (we will see this in subsection 3.4, when studying particular forms for the potential). Then, the fact that slow-roll parameters should still be smaller than one just before inflation ends usually guarantees that during the observable e-folds, the first-order slow-roll results provide an excellent approximation of the spectrum.

However, it is not impossible to build models in which the slow-roll conditions are only marginally satisfied during the observable e-folds. This can happen for instance:

1. in some models of hybrid inflation (defined in subsection 3.4) where at least one of the parameters ϵ or $|\eta|$ decreases with time, and the beginning of inflation is close to N_*. Then, the slow-roll parameters could be close to order one at $N = N_*$.

2. in mode complicated models such that during observable e-folds, for φ close to φ_*, the potential contains a sharp feature (this could be an effective description of a physical phenomenon triggered by some other fields, e.g. a phase transition).

In the first case, it is generally sufficient to extend the slow-roll computations to second order, and to find an expression for the running of the tilts. The running can be detectable if it is large enough and if we have some precise observations on a wide range of scales. This next-order approach can still fail to describe the power spectra when they inherit a particular shape at a given scale scale from a sharp feature in the effective potential. Then, the spectrum is said to be Broken Scale Invariant (BSI). In these models, the slow-roll conditions are almost violated, or even completely violated during a short time. The stronger is the slow-roll violation, the larger is the feature in the power spectrum (spike, well, break, oscillation, etc.). These situations have been widely studied in the past, but they receive a decreasing interest since precise cosmological observations, which are compatible with a smooth spectrum, leave less and less room for such features.

Some BSI spectra can be computed analytically, using various possible approximations e.g. for the equation of evolution (125), which experiences different regimes (usually, a first slow-roll stage, a transition and a second slow-roll stage). In any case, it is always quite easy to compute the inflationary power spectrum numerically. This method is the only possible one for the most complicated BSI models. It is also the best way to control the level of precision of the analytic slow-roll predictions. Numerical codes designed for computing the inflationary power spectra usually obey to the following scheme:

1. First, the background equations are integrated as a function of time for a given potential and a set of initial conditions $\bar{\varphi}(t_i)$, $\dot{\bar{\varphi}}(t_i)$. The results for $a(t)$, $H(t)$, $\bar{\varphi}(t)$ and $\dot{\bar{\varphi}}(t)$ are kept in memory.

2. Then, one performs a loop over each independent observable Fourier mode. For each mode, one integrates the equation of evolution of scalar perturbations: either the master equation (125) for ξ_k, or equation (119) for ϕ_k, or the coupled equations (117, 118) for ($\delta\varphi_k$, ϕ_k). The computation should start when $k \gg aH$ (typically, $k = 50\, aH$ gives very precise results) so that the initial mode function can be approximated by the flat space-time solution, up to an arbitrary initial phase. If this phase is chosen in such way that $\delta\varphi_k$ is real, the initial conditions read

$$\delta\varphi_k \simeq \frac{1}{a\sqrt{2k}} , \qquad \delta\dot{\varphi}_k \simeq -i\frac{k}{a}\delta\varphi_k , \tag{174}$$

$$\phi_k \simeq i\frac{4\pi\mathcal{G}\dot{\bar{\varphi}}}{\sqrt{2k^3}} , \qquad \dot{\phi}_k \simeq -i\frac{k}{a}\phi_k , \tag{175}$$

$$\xi_k \simeq \frac{1}{\sqrt{2k}} , \qquad \dot{\xi}_k \simeq -i\frac{k}{a}\xi_k . \tag{176}$$

The computation can be stopped as soon as the solution for ϕ_k is proportional to $\dot{H}/H^2$. Then, the coefficient $C_1(k)$ of Eq. (133) is found from

$$C_1(k) = -\phi_k(t)\,\frac{H(t)^2}{\dot{H}(t)} \tag{177}$$

and the power spectrum of ϕ_k e.g. during matter domination is given by Eqs. (140,141)

$$\mathcal{P}_\phi^{\mathrm{MD}} = k^3 \left(\frac{3}{5}\right)^2 |C_1(k)|^2, \tag{178}$$

while the super-horizon curvature perturbation spectrum at any time is given by

$$\mathcal{P}_\mathcal{R} = k^3 |C_1(k)|^2 . \tag{179}$$

Usually, it is sufficient to stop the integration at $k = aH/50$ for computing $C_1(k)$. The rigorous way to stop the calculation is to plot the quantity $\phi_k H^2/\dot{H}$ as a function of time and to wait until its time-derivative falls below some threshold.

3. in a similar loop over k values, one integrates the equation for y_k, starting from the initial condition

$$y_k = \frac{1}{\sqrt{2k}} \ , \qquad \dot{y}_k \simeq -i\frac{k}{a}y_k \ . \tag{180}$$

The computation can be stopped when the ratio y_k/a is approximately constant. Then, the ratio $\tilde{C}_1(k) \equiv \sqrt{32\pi\mathcal{G}}\,(y_k/a)$ provides the tensor spectrum, since

$$\mathcal{P}_h = k^3|\tilde{C}_1(k)|^2 \ . \tag{181}$$

3 Current constraints from observations

3.1 Relating the primordial spectra to CMB and LSS observables

If we want to understand CMB anisotropy and large scale structure formation, we need to follow the evolution of the Fourier modes of matter perturbations $\delta_m \equiv \delta\rho_m/\bar{\rho}_m$ and radiation perturbations $\delta_r \equiv \delta\rho_r/\bar{\rho}_r$. Initial conditions for these quantities can be defined at some time deep inside the radiation era ($\bar{\rho}_m \ll \bar{\rho}_r$), on super-Hubble scales $k \ll aH$ (i.e., before the causal evolution due to gravitational and electromagnetic forces can start), and after the various phase transitions leading to a universe composed of ordinary photons, neutrinos, baryons and cold dark matter particles. For simplicity, let us neglect the role of neutrinos and do as if all the radiation was in the form of photons. Then, in the longitudinal (Newtonian) gauge, the linearized Einstein equations relate the metric perturbation ϕ to the density perturbation of the dominant component, which is radiation:

$$\delta_r = -2\phi^{\mathrm{R}D} = -\frac{4}{3}\mathcal{R} = constant \ . \tag{182}$$

This provides an initial condition for radiation perturbations. What about matter perturbations? The answer actually depends on the scenario for producing matter particles in the early universe. However, there is a very generic class of scenarios such that all particles produced during reheating –or during subsequent phase transitions– share the same local number density contrast:

$$\forall(i,j), \qquad \frac{\delta n_i}{\bar{n}_i}(\vec{x}) = \frac{\delta n_j}{\bar{n}_j}(\vec{x}) \ . \tag{183}$$

In particular, this is always the case in cosmological scenarios where primordial perturbations are generated during single-field inflation. Indeed, in this case, all species are produced (possibly through many intermediate steps) from the decay of a unique particle, the inflaton. So, the local value of the inflaton density contrast at the end of inflation is the only relevant function of space allowing to compute the density contrast at later times: on super-horizon scales, no mechanism can introduce a shift between $\delta n/\bar{n}$ for one species and $\delta n/\bar{n}$ for another species. This would not be true in multiple inflation scenarios, possibly leading to entropy perturbations. Standard initial conditions, described by the condition of Eq. (183), are called isentropic or adiabatic conditions. The word adiabatic is justified in this context by the fact that Eq. (183) also guarantees that the total pressure perturbation as a function of space, $p(\vec{x})$, is proportional to the total density perturbation $\rho(\vec{x})$, as in any adiabatic fluid.

Focusing on two types of particles, that we call generically matter and radiation, we observe that the relation

$$\frac{\delta n_r}{\bar{n}_r}(\vec{x}) = \frac{\delta n_m}{\bar{n}_m}(\vec{x}) \tag{184}$$

can be translated in terms of density contrasts using the dilution laws $\rho_m \propto a^{-3} \propto n_m$ for non-relativistic matter, and $\rho_m \propto a^{-4} \propto {n_m}^{4/3}$ for ultra-relativistic species. After differentiation and use of Eq. (184), we get

$$\delta_m = \frac{3}{4}\delta_r \ . \tag{185}$$

Together with Eq. (182), this equation provide initial conditions for matter and radiation perturbations, namely

$$\delta_m = \frac{3}{4}\delta_r = -\frac{3}{2}\phi^{\mathrm{R}D} = -\mathcal{R} = constant \ . \tag{186}$$

In addition, we have seen in subsection 2.6 that $\mathcal{R}$ is frozen on super-Hubble scale, so the power spectrum of δ_m during radiation domination and on super-Hubble scales cab be readily deduced from the that of metric or curvature perturbations at the end of inflation, computed in subsection 2.6.

Matter perturbations. Matter perturbations δ_m which enter inside the Hubble radius during radiation domination start to evolve with time. During RD, baryons are strongly coupled with photons and undergo acoustic oscillations. Meanwhile, the CDM particles undergo a complicated evolution which can be summarized as small oscillations (driven by the metric perturbation) on top of an average logarithmic growth (corresponding to gravitational clustering). After photon decoupling, CDM perturbations grow like $\delta_m \propto a$, and the perturbations of decoupled baryons converge towards the CDM ones. Matter perturbations which enter during matter domination grow like the scale factor, both for baryons and CDM, which are then indistinguishable. In summary, total matter perturbations always grow with time because of gravitational clustering, and the modes entering earlier are those experiencing more growth. However, the growth is less efficient for modes entering during radiation domination. Following Eq. (186) the matter power spectrum

$$P(k) \equiv \langle |\delta_{m\,k}|^2 \rangle \tag{187}$$

is initially equal to

$$P(k) = \langle |\mathcal{R}|^2 \rangle = k^{-3} \mathcal{P}_{\mathcal{R}}(k) \ . \tag{188}$$

So, for $n_s \simeq 1$, $P(k)$ has a logarithmic slope close to -3. As a consequence of the previous described evolution, the slope changes from -3 to +1 for modes entering during inside the Hubble radius during matter domination. For smaller modes, the logarithmic slope is reduced because of the slow perturbation growth during radiation domination. The slope is actually negative and very much scale-dependent (i.e., this branch of the spectrum is not a power-law).

As long as the evolution inside the Hubble radius is linear, it can be summarized in terms of a transfer function $T(k)$,

$$\delta_{m\,k}(t_0) = T(k)\, \delta_{m\,k}(t_i) \ , \tag{189}$$

where t_0 is the time today and t_i is chosen during radiation domination when $k \ll aH$. The transfer function can easily be shown to be proportional to k^2 on the largest scales observable today,

$$T(k) \propto k^2 \qquad\qquad for \qquad k_{\max} \ll k \ll k_{\rm nr} \ , \tag{190}$$

where $k_{\max}$ is the scale entering the Hubble radius today, and $k_{\rm nr}$ the one entering the Hubble radius at the time of equality between radiation and matter. The linear matter power spectrum today can be written as

$$P(k) = \frac{T(k)^2}{k^3} \mathcal{P}_{\mathcal{R}}(k) \ . \tag{191}$$

In conclusion, the matter power spectrum which can be reconstructed from the observation of large scale structures (galaxies, clusters, fluctuations of the intergalactic medium, etc.) is directly proportional to the inflationary scalar power spectrum, with potentially a clear imprint of the scalar tilt or any other other feature generated by the inflationary dynamics.

CMB perturbations. Still neglecting neutrinos and discarding the (small) evolution of photon perturbation after decoupling, we can consider that observable CMB temperature anisotropies as related to the density perturbations $\delta_r(t_{\rm dec}, r_{\rm lss}\hat{n})$ at the time of decoupling $t_{\rm dec}$, in direction $\hat{n}$ and on the last scattering surface (a sphere of comobile radius $r_{\rm lss}$). The map of temperature perturbations can be expanded in spherical harmonics,

$$\frac{\delta T}{\bar{T}}(\hat{n}) = a_{lm}^T Y_{lm}(\hat{n}) \ . \tag{192}$$

If the perturbations $\delta_r(t_{\rm dec})$ are Gaussian, so are the multipoles a_{lm}. Their properties are therefore encoded in the power spectrum

$$C_l^T \equiv \langle |a_{lm}|^2 \rangle \tag{193}$$

which are related in first approximation to the Fourier power spectrum of $\delta_{r\,k}(t_{\rm dec})$ convolved with Bessel functions. The evolution of $\delta_{r\,k}(t)$ between some initial time t_i and t is linear, so it can be parametrized by a transfer function. Actually, tensor perturbation also contribute to temperature anisotropies, and one can also define a transfer function relating initial perturbations h_{ij} to temperature fluctuations at decoupling. In total, the multipole power spectrum C_l^T can be decomposed in

$$C_l^T = \int \frac{dk}{k} \left(\mathcal{P}_{\mathcal{R}}(k) \left| \Delta_l^{(S)}(k, t_{\rm dec}) \right|^2 + \mathcal{P}_h(k) \left| \Delta_l^{(T)}(k, t_{\rm dec}) \right|^2 \right) \tag{194}$$

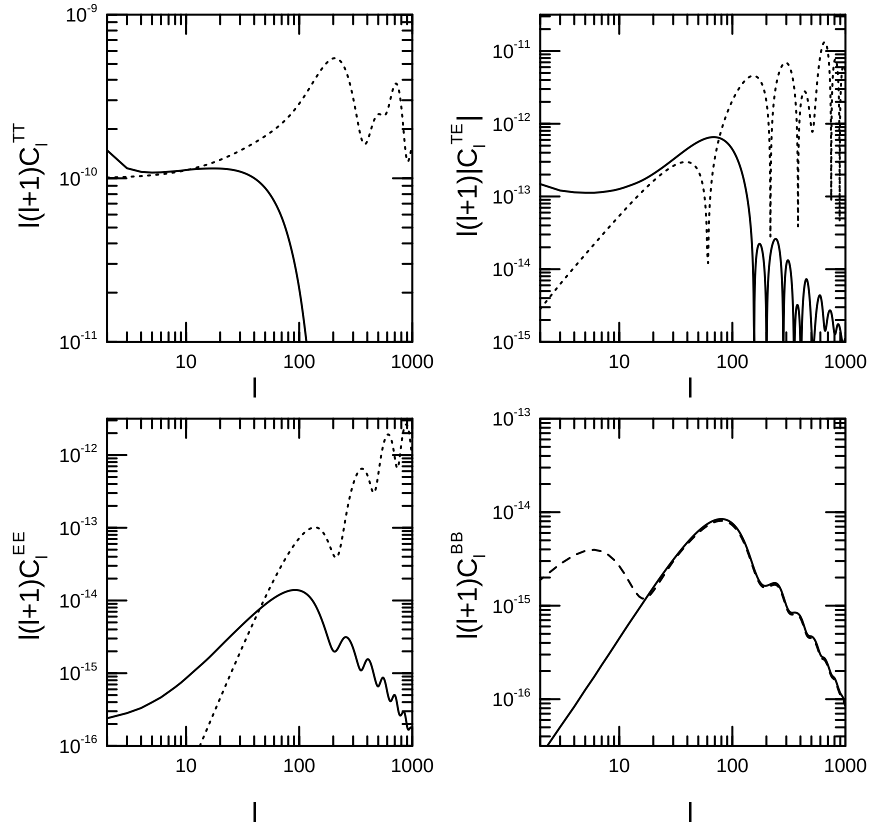

Figure 2: The four independent power spectra of primary CMB anisotropies for temperature (TT), E-polarization (EE), B-polarization (BB) and temperature×E-polarization cross-correlation. These curves are theoretical predictions for a standard cosmological model (with no reionization and no cosmological constant) obtained by Kamionkowski and Kosowsky in Phys.Rev.D57:685,1998 `[astro-ph/9705219]`. In each panel, the dotted line shows the contribution of primordial scalar perturbations, and the solid line that of primordial tensor perturbations (gravitational waves from inflation). In the BB case there is no scalar contribution, but in the corresponding plot the authors show the impact of reionization (dashed curve).

where $|\Delta_l^{(S,T)}|^2$ are the dimensionless transfer functions expanded in multipole space for respectively scalar and tensor modes. Observationally, the spectrum C_l^T is determined by mapping temperature anisotropies, expanding in spherical harmonics and averaging over m,

$$C_l^{obs} = \frac{\sum_m |a_{lm}|^2}{m} \ . \tag{195}$$

Cosmological parameters can be determined by fitting the theoretical power spectrum of Eq. (194) to C_l^{obs}.

All what we said for temperature anisotropies also applies for polarization anisotropies. Since the polarization map consists in a two-dimensional vector field, it can be decomposed in two scalar functions: a curl-free component $E(\hat{n})$ and a divergence-free component $B(\hat{n})$. The E mode is seeded both by scalar and tensor anisotropies, while the B-mode can only be generated by tensor anisotropies (this statement is true at the level of primary anisotropies, i.e. neglecting the additional distortion caused by foregrounds between the last scattering surface and ourselves). In total, four power spectra can be measured: the temperature and E-mode power spectra (C_l^T, C_l^E) related to $\mathcal{P}_\mathcal{R}$ and $\mathcal{P}_h$; the B-mode power spectrum C_l^B related to $\mathcal{P}_h$ only; and the cross-correlation power spectrum between temperature and E-modes, C_l^C, related to $\mathcal{P}_\mathcal{R}$ and $\mathcal{P}_h$. Other cross-correlation spectra can be shown to vanish by construction. Figure 2 shows some typical theoretical predictions for each of these spectra.

3.2 Overall agreement between observations and inflationary predictions

Before presenting the results of a detailed data fitting, we should mention that the overall agreement between the observed CMB anisotropy and the predictions of perturbation theory in inflationary cosmology is absolutely remarkable. Inflation makes various non-trivial predictions (which are not true for a variety of alternative models), each of which is verified by the observations of the last decade. Let us mention four confirmed predictions which can be seen as indirect evidences for inflation.

3.2.1 Gaussianity

The fact that standard inflationary cosmology predicts Gaussian CMB anisotropies comes from the fact:

- that each Fourier mode starts in a quantum fundamental state with a Gaussian wave function. So, the probability distribution $\mathcal{P} = |\Psi|^2$ (that we interpret as a classical probability after Hubble crossing) is also Gaussian.
- that the equations of evolution between inflation and photon decoupling are linear, so that the shape of the probability distribution is conserved (a probability distribution $\mathcal{P}(y)$ transported by linear equations can only change according to the rescaling of the y variable; in cosmology, the evolution of linear cosmological perturbations affects the mode variance, but the probability remain Gaussian).

Until the mid-1990's, it was difficult to discriminate between the main two paradigms for generating cosmological perturbations: inflation and topological defects. The latter mechanism is strongly non-linear, leading to significantly non-Gaussian perturbations. The first observation of CMB anisotropies by COBE (1990-1994) probed large angular scales, i.e. C_l^T on small multipoles ($l \leq 20$). It was then established that on these scales the perturbations do not exhibit strong deviations from Gaussianity, a conclusion favorable to inflation. With the most recent CMB observations, constraints on possible deviations from Gaussianity are very stringent.

The observation of Gaussian perturbations are an indirect evidence in favor of the most simplest inflationary scenarios (those described in this course). In order to produce large non-Gaussianities, inflationary scenarios should:

- either start from a non-vacuum initial state. This could be justified by non-minimal scenarios with e.g. two inflationary stages: a first one would excite the quantum states, and a second one would produce observable cosmological perturbations starting from non-Gaussian wave functions. However, in these scenarios, the Fourier modes remain independent from each other and the central limit theorem applies. The observable a_{lm} is given by a sum over independent Fourier modes, and this sum tends to render each a_{lm} nearly Gaussian.
- or force the metric and/or inflaton perturbations to be marginally non-linear at some stage, during which different Fourier modes couple with each other and significant non-Gaussianity can emerge.

In conclusion, the observation of nearly Gaussian CMB anisotropies is in agreement with the most simple and natural inflationary scenarios, and disfavors various inflationary or non-inflationary alternatives.

3.2.2 Scale-invariance

The approximate scale-invariance of the primordial power spectrum (for scalar and tensor perturbations) is a distinct prediction of inflation. The observation of temperature anisotropies on large angular scales by COBE brought evidence in favor of nearly scale-invariant primordial perturbations in the range $2 \leq l \leq 20$. The following generations of CMB observations (completed with the reconstruction of the matter power spectrum from the observation of LSS) now confirm the scale invariance of the primordial scalar spectrum with an impressive precision. As we shall mention here, the prediction of scale invariance by inflation is really non-trivial, and some alternatives to inflation prefer a completely different scaling law.

It is instructive to trace back the origin of the factors k contributing to the final prediction $\phi_k \sim k^{-3/2}$ at the end of inflation. Some factors come from the initial condition for the mode function, i.e. from the quantum commutation relation; other factors are related to the later evolution. Suppose, for instance, that we focus on the variable $\delta\varphi_k$, for which the discussion is more illuminating than for ξ_k. The initial mode function for $\delta\varphi_k$ reads

$$\delta\varphi_k = \frac{1}{\sqrt{2k}\,a} e^{-i\int \frac{k}{a} dt} \tag{196}$$

so that a factor $k^{-1/2}$ is present from the beginning. The amplitude then evolves quickly with time, like a^{-1}. However, it freezes out near the characteristic time of horizon crossing, when $k \sim a_* H_*$. So, later on, the amplitude remains close to $(\sqrt{2k}\, a_*)^{-1} \sim H_*/\sqrt{2k^3}$. Outside the Hubble radius, the relationship between $\delta\varphi_k$ and ϕ_k reduces to $H\phi_k \simeq 4\pi\mathcal{G}\dot{\bar{\varphi}}\,\delta\varphi_k$ and does not involve any power of k. So, the factor $k^{-3/2}$ is inherited by the metric perturbation ϕ_k. In order to reach this result, it was particularly important to assume that the inflaton field is light during inflation (if we quantize a heavy scalar field with an effective mass $m > H$, the mode function inside the Hubble radius will not contain the desired factor $k^{-1/2}$), and also that the inflaton field dominates the total energy density of the universe so that inflaton fluctuations are directly imprinted into metric fluctuations. Alternative scenarios to inflation like the Pre Big Bang or the Ekpyrotic paradigms usually have difficulties in predicting nearly scale-invariant perturbations, because some of the above requirements are not satisfied.

3.2.3 Adiabaticity

We have seen in subsection 3.1 that after single-field inflation, the number density contrasts of all species are related to a unique function: the inflaton density fluctuation at the end of inflation. This ensures that for any two species (i, j),

$$\frac{\delta n_i}{\bar{n}_i}(\vec{x}) = \frac{\delta n_j}{\bar{n}_j}(\vec{x}) \ . \tag{197}$$

This adiabatic initial condition plays a very important role in the study of cosmological perturbations. If there exists at least one component for which $\frac{\delta n_i}{\bar{n}_i}$ differs from the others, the universe contains entropy fluctuations (also called isocurvature modes) in addition to (or in replacement of) adiabatic fluctuations. This will change the temporal phase of photon-baryon acoustic oscillations before decoupling, as well as the k-dependence of the transfer functions for radiation and matter. The CMB peaks will appear at different angular scales (different multipoles l), and the hierarchy between the amplitude of the CMB peaks will change, as well as the slope of the matter power spectrum $P(k)$.

The most robust proof in favor of a plain adiabatic scenario came with the observation of the first CMB peaks (a detection of at least two peaks was performed by the Boomerang balloon in 2000). Today, bounds on entropy fluctuations are quite constraining.

The reason for which single-field inflation leads to adiabatic fluctuations is that all initial fluctuations are summarized by a single function; everything happens as if all points in comobile space where experiencing exactly the same post-inflationary cosmological evolution, excepted for an initial time shift due to the inflaton perturbations at the end of inflation. In multiple inflationary models, when the quantum fluctuations of more than one field play a role, one can have significant entropy fluctuations surviving after inflation; alternatives to inflation like Pre Big Bang models or topological defects also lead to isocurvature perturbations. So, once again, the observation of adiabatic fluctuations is in remarkable agreement with the predictions of the most simple inflationary scenario.

3.2.4 Coherence

The fact the we see acoustic peaks in the spectrum of CMB temperature and polarization anisotropies is also far from obvious. Before photon decoupling, we know that the baryon-photon fluid has all the required characteristic for the propagation of acoustic waves. However, these waves will be present only if the fluid is displaced from equilibrium at some time. Depending on the initial conditions, oscillations could take place in phase (like for stationary waves) or with random phases.

Among all possible initial conditions, one usually makes a distinction between *active* and *passive* mechanisms. Active mechanisms are such that at some very early time, the system is displaced from equilibrium, and then it evolves freely. The Einstein equation have no source term, and the evolution is similar to that of a free harmonic oscillator. The crucial point is that for a given wavelength, all modes enter inside the Hubble radius at the same time, when $k = aH$. It is this time which sets the initial condition for the phase: all perturbations with given wavenumber will evolve with different amplitudes (since they are stochastic) but with a common phase, passing through extrema at the same time. At decoupling, the oscillation are frozen, and because of the phase coherence the structure of the peaks appears in Fourier space, and therefore also in multipole space.

In passive mechanisms, the acoustic oscillation are displaced from equilibrium by some mechanism injecting stress at all times. The Einstein equation do have a source term, representing for instance topological defects. In the vicinity of a defect, strong gravitational interactions create an initial stress which then propagates away from the defect. These perturbations are easier to imagine in real space rather than in Fourier space. Interactions with topological defects happen at different times and locations, so

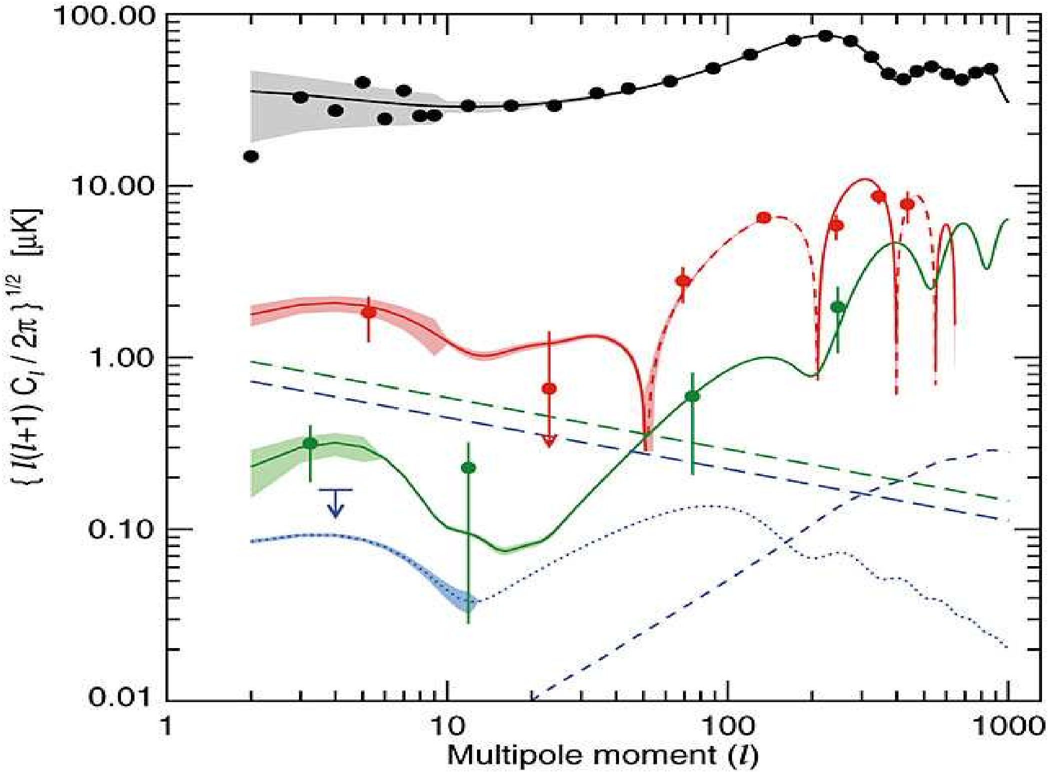

Figure 3: WMAP 3rd year results for the temperature power spectrum (black), E-polarization (green), B-polarization (blue), TE cross-correlation (red). In each case, the data is represented by dots and error-bars. For temperature, the error-bars are so small that they are invisible. For B-polarization the data consists in a single upper bound (blue arrow). The lines correspond to the prediction of the favored ΛCDM model. These results are taken from the WMAP mission's result webpage `http://map.gsfc.nasa.gov/m_mm.html`.

that Fourier modes with a common wavenumber have a priori random phases. At decoupling, for each wavenumber, the average over all phases results in a smooth spectrum: acoustic oscillations are averaged out.

The observation of acoustic peaks in the CMB excludes scenarios where topological defects are the main seed for cosmological perturbations. This is not a direct proof of inflation, but at least a strong indication that cosmological perturbations are generated very early, when modes are far outside the Hubble radius.

In total, the four predictions of inflation reviewed in this chapter are all compatible with observations. Together with the motivations presented in the introduction, they represent a variety of complementary, indirect proofs for inflation, in the sense that nobody could build so far a different mechanism explaining simultaneously all these features. It is particularly impressive that all observations converge towards the *most simple implementation of the inflationary paradigm*, rather than towards some particular, non-minimal realization.

3.3 Recent constraints on slow-roll parameters

Let us report the most recent constraints on inflation, based on three years of observation by the WMAP satellite (accurate measurements of C_l^T until $l \sim 1000$, and more preliminary measurements of C_l^E and C_l^C, see Figure 3), completed by various LSS observations.

First, there is no evidence at the moment in favor of a non-zero contribution of tensor perturbations to temperature and/or polarization anisotropies. The power spectra are well-fitted with a purely scalar contribution, with an amplitude (computed near the scale $k = 0.05\,\mathrm{M}pc^{-1}$, with an uncertainty of order

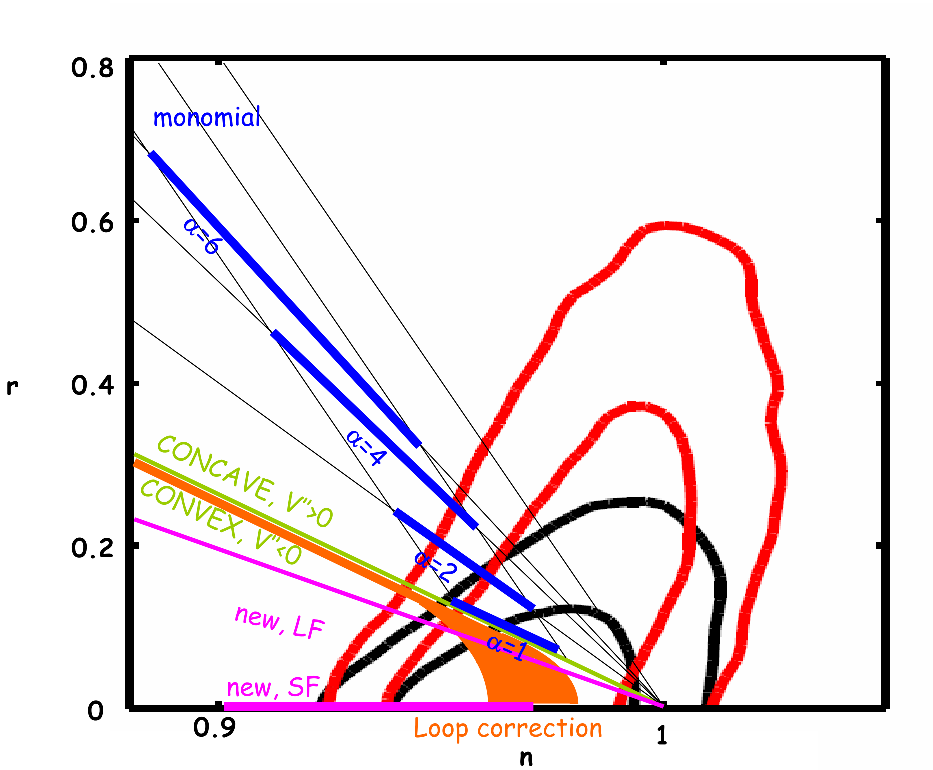

Figure 4: The (n_s, r) "phase-space diagram" of inflationary models. Red contours show the 1-σ and 2-σ preferred region in n_s, r) space, as obtained by Peiris and Easther (Figure 4 in `astro-ph/0603587`) using only WMAP 3-year data. Black contours add some information from the SDSS galaxy power spectrum. We superimpose the regions associated with the particular potentials described in section 3.4. According to the terminology of this course (different from that of other authors like Kinney and collaborators), inflationary models are divided into concave/convex and small/large field models in the following way: the green line separates concave and convex potentials; small field models are confined near the $r = 0$ axis; and the rest of parameter space corresponds to large field models. Hybrid inflation models can stand anywhere in this parameter space. Monomial potentials correspond to the blue segments. New inflation potentials stand in the lower left sector. Hybrid inflation with a logarithmic loop correction correspond to the orange band.

10%),

$$\mathcal{P}_\phi^{\mathrm{MD}} \equiv k^3 \langle |\phi_{k \ll aH}^{MD}|^2 \rangle = \frac{3}{50}(8\pi\mathcal{G})^3 \frac{V_*^3}{V_*'^2} = \frac{3}{100}(8\pi\mathcal{G})^2 \frac{V_*}{\epsilon_*} \simeq 2.10^{-8} \quad \Rightarrow \quad m_P^{-6} \frac{V_*^3}{V_*'^2} \simeq 2.10^{-11} \tag{198}$$

(we defined the Planck mass as $\mathcal{G} \equiv m_P^{-2}$). The 95% confidence limit on the tensor-to-curvature ratio $r = 16\epsilon_*$ is found to be

$$r < 0.55 \tag{199}$$

which can be combined with the previous constraint in order to get an upper bound on the inflationary energy scale,

$$V_* = \frac{100}{3\,(8\pi)^2 16}\; r\,\mathcal{P}_\phi^{\mathrm{MD}} m_P^4 < \left(2.5 \times 10^{-3} m_P\right)^4 = \left(3.7 \times 10^{16}\mathrm{GeV}\right)^4 \; . \tag{200}$$

The scalar primordial spectrum is compatible with a power-law (no evidence for tilt running) with a scalar tilt $n_s = 0.98 \pm 0.02$ (Peiris & Easther `astro-ph/0603587`). There is some parameter degeneracy between n_s and r, so that constraints on inflationary models are best seen in two-dimensional (n_s, r) likelihood plots.

3.4 Implications for various types of inflaton potentials

3.4.1 Practical method for constraining potentials

The study of a generic inflationary potential can be carried along the following steps:

1. compute the slow-roll parameters $\epsilon(\varphi)$, $\eta(\varphi)$ defined in Eq. (61).
2. solve the integral of Eq. (64) which provides the relation $N(\varphi)$, where N is the number of e-folds before the end of inflation. Re-express the slow-roll parameters as functions of N: $\epsilon(N)$, $\eta(N)$.
3. let us call N_* the number of e-folds between horizon crossing for observables scales and the end of inflation. For the allowed range $30 < N_* < 60$ (see subsection 1.3), compute at first order in slow-roll the tensor-to-scalar ratio $r = 16\epsilon_*$ [Eq. (148)] and the scalar tilt $n_s = 1 - 6\epsilon_* + 2\eta_*$ [Eq. (169)].
4. check the implications of observational constrains in the $(r,\, n_s)$ plane.
5. check the constraint of Eq. (198) on the scalar spectrum amplitude.

3.4.2 Monomial potentials

We first consider monomial potentials of the form

$$V = \lambda m_P^4 \left(\frac{\varphi}{m_P}\right)^\alpha \; . \tag{201}$$

For $\alpha = 2$ one recovers a usual quadratic potential with mass $m = \sqrt{2\lambda}\, m_P$.

Exercise 1: show that

$$\epsilon = \frac{\alpha}{2(\alpha-1)}\eta = \frac{1}{16\pi}\left(\frac{\alpha m_P}{\varphi}\right)^2 \; . \tag{202}$$

Deduce that

$$n_s - 1 = -2\epsilon\left(\frac{\alpha+2}{\alpha}\right) \tag{203}$$

and that a model with given value of α stands along the following line in the $(r,\, n_s)$ plane:

$$r = -\frac{8\alpha}{\alpha+2}(n_s - 1) \; . \tag{204}$$

Exercise 2: assuming that $\epsilon = 1$ is the condition for ending inflation, show that the corresponding field value is $\varphi_{end} = \alpha m_P/\sqrt{16\pi}$. Integrate dN between φ_* and φ_{end} and show that the integral is dominated by the φ_* boundary (at least for $\alpha \ll 4N_*$). Conclude that irrespectively of the details of inflation ending,

$$\varphi_* = \sqrt{\frac{\alpha N_*}{4\pi}} m_P \; . \tag{205}$$

Exercise 3: deduce that $r = 4\alpha/N_*$ and $n_s - 1 = -(\alpha+2)/(2N_*)$. Conclude that for a given N_* and any α, monomial inflationary models stand along the line

$$r = -8(n_s - 1) - \frac{8}{N_*} \; . \tag{206}$$

Summarize the result in the $(r,\, n_s)$ phase-space diagram.

The results are displayed in Fig. 4. In $(r,\, n_s)$ space, monomial inflationary models stand in a narrow band delimited by the curves $r = -8(n_s-1) - 0.26$ [for $N_* = 30$] and $r = -8(n_s-1) - 0.13$ [for $N_* = 60$]. Inside this band, models with a given value of α are along the line $r = -8[\alpha/(\alpha+2)](n_s-1)$. It turns out from Fig. 4 that the case $\alpha = 2$ is comfortably allowed, while $\alpha = 4$ raises some tension with observations and $\alpha \geq 6$ seems to be excluded.

Exercise 4: show that

$$m_P^{-6}\frac{V_*^3}{V_*'^2} = \frac{\lambda}{\alpha^2}\left(\frac{N_*\alpha}{4\pi}\right)^{\alpha/2+1} \simeq 2.10^{-11} \ , \tag{207}$$

implying

$$\lambda \simeq 7.10^{-14}\alpha^2\left(\frac{4\pi}{N_*\alpha}\right)^{\alpha/2+1} \ . \tag{208}$$

Conclude that for $\alpha = 2$, the mass must be tuned to $m \sim 5.10^{-8}m_P$; for $\alpha = 4$, the self-coupling should be $\lambda \sim 2.10^{-16}$; and for larger values of α, the tuning should be more severe, with $\lambda \to 0$ in the limit $\alpha \to \infty$.

Inflationary models with a monomial potential are often called chaotic inflationary models, for historical reasons: it is for this family of potentials that Linde discussed in 1983 the issue of **chaotic initial conditions**. At that time, some people objected that the universe would emerge from the quantum gravity period with field values necessarily smaller than the Planck mass, so that monomial inflation can never take place (Eq. (205) can never be fulfilled). Linde emphasized that, following the Heisenberg uncertainty principle, the universe is expected to emerge from the quantum gravity era with stochastic initial conditions such that the *energy densities* (and not the field itself) should have typically Planckian values: $\langle E\rangle \sim m_P^4$. Decomposing into potential, kinetic and gradient energy density, this gives

$$\langle V\rangle \sim \langle\dot{\varphi}^2\rangle \sim \langle(\vec{\nabla}\varphi)^2\rangle \sim m_P^4 \ . \tag{209}$$

With the potential of Eq. (201) and $\lambda \ll 1$, the condition $\langle V\rangle \sim m_P^4$ implies $\langle\varphi\rangle > m_P$. Statistically, there should exist some patches in which the potential energy V slightly dominates over the kinetic and gradient terms. In any such region, the slow-roll conditions are satisfied and inflation can take place.

3.4.3 New inflation

Instead of assuming that the field rolls down in a potential well, we will now assume that it rolls away from an unstable point of equilibrium. In the vicinity of this point, we can Taylor-expand the potential and keep the leading term

$$V = V_0\left[1 - \left(\frac{\varphi}{\mu}\right)^{\alpha} + \text{n}ext\ order\ terms\right] \tag{210}$$

where $\varphi < \mu$, and the Taylor approximation is expected to break down for some value of order of μ. This kind of potential will appear, for instance, in the case of a spontaneous symmetry breaking in the early universe. In principle there could be two cases. First, if μ is much smaller than the Planck mass, we have

$$\varphi < \mu \ll m_P \qquad (SFM) \tag{211}$$

throughout inflation. We will call such models Small Field models (SFM). Alternatively, if μ is of the same order as (or eventually larger than) the Planck mass, we will call the model a Large Field Model (LFM). We will see that SFM and LFM have different implications.

New inflation models were proposed by Starobinsky in 1979 and Linde in 1982. A full understanding of these models requires some knowledge of another paradigm called **stochastic inflation**, introduced by Starobinsky in 1986. Here we will only describe the underlying ideas of this paradigm.

We have seen that usually, the field can be decomposed into a homogeneous background $\bar{\varphi}$ obeying to the homogeneous Klein-Gordon equation, plus small perturbations which mode functions obey to the linearized Klein-Gordon equation. If we take the case of a field placed initially in an unstable equilibrium point, we see immediately that this description cannot be the good one. Indeed, in the unstable equilibrium point, $\partial V/\partial\varphi = 0$. The homogeneous Klein-Gordon equation tells us that $\bar{\varphi}$ will remain in this point forever. On the other hand, the linear perturbation equation has a mass term $[(k/a)^2 + \partial^2 V/\partial\varphi^2]$ which is negative for $k^2 \leq -a^2(\partial^2 V/\partial\varphi^2)$. If we want inflation to occur, we should assume that $|\partial^2 V/\partial\varphi^2| \leq H^2$ (this is the second slow-roll condition). So, the modes with a negative effective squared mass are all super-Hubble modes with $k < aH$. For these modes, the negative squared mass term causes a tachyonic instability: they are exponentially amplified. Modes inside the Hubble radius will feel the large super-Hubble modes as an "effective zero-mode" or effective background which is exponentially amplified with time. We see that the usual splitting into $\bar{\varphi}$ and $\delta\varphi_k$ variables cannot describe properly the physical evolution of this model: we must pass to a more complicated description in which the "background" is the sum of a classical zero-mode, plus the assembly of all super-Hubble modes, which contribute as a stochastic component. This is exactly the goal of stochastic inflation. The evolution of the effective background is not given anymore by the homogeneous Klein-Gordon equation, but by some equations usually employed in statistical mechanics, like the Fokker-Planck equation.

The fact that perturbations evolve in a stochastic background composed of a classical zero-mode plus the assembly of all super-Hubble modes is in principle *always* true. However, in most cases this can be safely neglected in first approximation, because the modes which exit the Hubble radius during the relevant inflationary e-folds contribute to the total effective zero-mode in a negligible way. The dominant part of the background, i.e. the classical zero-mode plus the assembly of all modes already beyond the Hubble radius at the beginning of the observable e-folds, obey exactly to the homogeneous Klein-Gordon equation. Therefore, the usual description applies.

At the beginning of new inflation, assuming that we start exactly from $\bar{\varphi}$, the zero-mode is initially sub-dominant and the effective background will consist entirely in large wavelength undergoing tachyonic amplification. It is possible to compute the evolution of each super-Hubble mode, to perform a statistical average and to derive the evolution of the effective zero-mode. After a short period, one can prove that the time evolution of the effective zero-mode gets closer and closer to that of a true zero-mode: i.e., the effective zero-mode starts to obey to the homogeneous Klein-Gordon equation. At that time we can forget about the complicated mechanism of "effective zero-mode assembly", and do as if we had a usual splitting into two decoupled sectors: the classical background and the small perturbations. Therefore, in this model, the machinery of stochastic inflation is relevant during the first few e-folds only. If the "observable e-folds" take place later, we can use the usual formalism, starting from some initial value $\bar{\varphi}_i \neq 0$.

In principle, it is possible to build more complicated models in which the stochastic inflation mechanism plays a more important role than just setting the initial conditions.

After this digression on the beginning of new inflation, we come back to the systematic study of the potential and its observational consequences.

Exercise 1: show that

$$\epsilon = \frac{\alpha^2}{16\pi}\left(\frac{m_P}{\varphi}\right)^2 f\left[\left(\frac{\varphi}{\mu}\right)^\alpha\right]^2 , \tag{212}$$

$$\eta = -\frac{\alpha(\alpha-1)}{8\pi}\left(\frac{m_P}{\varphi}\right)^2 f\left[\left(\frac{\varphi}{\mu}\right)^\alpha\right] , \tag{213}$$

where we defined $f(X) \equiv X/(X-1)$. Show that for SFM, the fact that slow-roll conditions must be satisfied during inflation implies that necessarily

$$f(X) \simeq X \ll 1 , \qquad \varphi \ll \mu , \qquad 0 < \epsilon \ll -\eta , \tag{214}$$

while for LFM we can have in principle φ of the same order as μ (although it is smaller by assumption) and the only robust prediction is $\eta < 0$. Deduce from this that on a phase diagram, SFM are very close to the $r = 0$ axis with $n_s < 1$, while LFM can occupy all the region in which $r \leq -\frac{8}{3}(n_s - 1)$.

Exercise 2: show that for $\alpha = 2$, the assumption that we have a SFM leads to

$$\eta = -\frac{\alpha(\alpha-1)}{8\pi}\left(\frac{m_P}{\mu}\right)^2 \gg 1 . \tag{215}$$

Conclude that the $\alpha = 2$ model can *only* belong to the LFM category.

Exercise 3: for a SFM and $\alpha \geq 4$, show that inflation ends when $\eta \sim -1$ and

$$\varphi_{end} \sim \left(\frac{8\pi\mu^\alpha}{\alpha(\alpha-1)m_P^2}\right)^{\frac{1}{\alpha-2}} . \tag{216}$$

Check that at the end of inflation

$$\left(\frac{\varphi_{end}}{\mu}\right)^\alpha \sim \frac{8\pi}{\alpha(\alpha-1)}\left(\frac{\varphi_{end}}{m_P}\right)^2 \ll 1 , \tag{217}$$

consistently with our assumption that in SFM one has $\varphi \ll \mu$ throughout inflation (including the very last inflationary e-fold). Integrate dN between φ_* and φ_{end} and show that the integral is dominated by the φ_* boundary. Conclude that irrespectively of the details of inflation ending,

$$\varphi_* = \left(\frac{8\pi\mu^\alpha}{\alpha(\alpha-2)N_* m_P^2} \right)^{\frac{1}{\alpha-2}} , \tag{218}$$

and deduce that

$$\eta_* = -\left(\frac{\alpha-1}{\alpha-2} \right) \frac{1}{N_*} \tag{219}$$

and

$$n_s - 1 = -\frac{2}{N_*} \frac{\alpha-1}{\alpha-2} . \tag{220}$$

For $\alpha \geq 4$ and $30 < N_* < 60$, check that $0.90 < n_s < 0.97$ and draw the new inflation SFM region in the $(r,\, n_s)$ space.

We can see on Fig. 4 that many new inflation model (both SFMs and LFMs) are still allowed by observations. Actually, the current best-fit model has $r \simeq 0$, $n_s \simeq 0.96$ or 0.97 (depending on the data set) and matches exactly the predictions of new inflation SFMs.

3.4.4 Hybrid inflation

Chaotic inflation and new inflation scenarios refer essentially to the way of entering into the inflationary stage. Instead, hybrid inflation is a generic name for models exiting the inflationary stage in a particular way: with a phase transition triggered by an auxiliary field, occurring either before or just after the breaking of slow-roll conditions. This second field doesn't play any role during inflation, so these models belong effectively to the category of single-field inflation rather than multiple-field inflation.

Hybrid inflation was first introduced by Lyth, Liddle and Stewart in 1990, and discussed in details by Linde in 1991. Any potential in which low values of the inflaton lead to a phase transition are good candidates for hybrid inflation. A typical example is provided by the potential

$$\begin{aligned} V_{tot}(\varphi,\chi) &= \lambda(M^2-\chi^2)^2 + V(\varphi) + \tilde{\lambda}\varphi^2\chi^2 \\ &= \lambda M^4 + (\tilde{\lambda}\varphi^2 - \lambda M^2)\chi^2 + \lambda\chi^4 + V(\varphi) \end{aligned} \tag{221}$$

where φ and χ are two real scalar fields (respectively the inflaton and the trigger field). Let us scrutinize the sign of the effective mass of χ, which appears explicitly in the second line of Eq. (221). For large inflaton values, the mass is positive. We assume that initial conditions are such that $\tilde{\lambda}\varphi^2 > \lambda M^2$, and that due to the positive squared mass the χ field quickly rolls down to $\chi = 0$ (the realization of this initial condition is less trivial than it appears at first sight, but here we won't go into these details). When $\chi = 0$, the effective potential for the inflaton reads

$$V_{eff}(\varphi) = \lambda M^4 + V(\varphi) . \tag{222}$$

If the potential $V(\varphi)$ is flat enough, inflation can eventually take place. There are two possibilities for ending inflation.

First, the slow-roll conditions can still be satisfied when the inflaton reaches the critical value such that the effective squared mass of the χ field vanishes (when $\tilde{\lambda}\varphi^2 = \lambda M^2$). Below this value, the χ field leaves the unstable equilibrium point $\chi = 0$ and the system rolls towards one of the absolute minima $(\chi,\varphi) = (\pm M, 0)$. During the phase transition, the effective mass of χ is of the order of $\sqrt{\lambda}M$. Usually this mass is larger than H by several orders of magnitude, which implies that the phase transition is very fast and does not support any accelerated expansion stage: inflation ends exactly when $\tilde{\lambda}\varphi^2 = \lambda M^2$, and the phase transition is often called the "waterfall". More exceptionally, some authors discuss the case in which $\sqrt{\lambda}M$ is smaller than the Hubble radius: then, the phase transition can add a few e-folds of inflation (like a second stage of inflaton). We will not discuss this situation here (actually, this case really belongs to multiple field inflation).

Second, the slow-roll condition can break *before* the inflaton reaches the critical value. Then, inflation ends, and it takes a very short time until the critical value is finally reached and the phase transition takes place.

In our example, there are two degenerate minima $(\chi,\varphi) = (\pm M, 0)$. The field can reach different minima in different region of the universe, separated by domain walls. In general, the phase transition at the end of inflation can represent the symmetry breaking of a larger symmetry than just the Z_2 symmetry of our example. Depending on this symmetry, the phase transition at the end of hybrid inflation can lead to various types of topological defects (e.g. cosmic strings). We will not focus here on this aspect, but we should mention that it is particularly interesting. Indeed, we have seen that the cosmological perturbations responsible for CMB anisotropies and for structure formation cannot be explained with topological defects. However, a small contribution of topological defects on top of the dominant inflationary perturbations is not excluded (although there are very strict bounds on the relative contribution). Scenarios in which the amplitude of inflationary and defect-generated perturbations differ by just one or two orders of magnitude are called *mixed perturbation models*. Usually, it can be objected that this possibility is not natural because it would result from some unlikely coincidence. However, in hybrid inflation scenarios, the parameters governing the density of topological defects at the end of inflation are the same as those governing the amplitude of the inflationary power spectrum: namely, these are the coupling constants and the mass scales in the two-field inflationary potential. Therefore, there are models which predict *naturally* some mixed perturbation, with a level of e.g. cosmic-string-generated fluctuations which could be detected in principle in future CMB observations. Such an observation would be a distinct signature of a certain class of hybrid models.

Generally speaking, a hybrid inflation model is specified by an effective potential of the form

$$V_{eff} = V_0[1 + f(\varphi)] \tag{223}$$

with $f(\varphi) > 0$, plus a value for inflation ending φ_{end} which is such that $\max[\epsilon(\varphi_{\rm end}), \eta(\varphi_{end})] \leq 1$, but not necessarily $\max[\epsilon(\varphi_{end}), \eta(\varphi_{\rm end})] \sim 1$ like in non-hybrid single-field models. We can classify the entire family of hybrid models according to the following criteria, resulting in different regions in the (r, n_s) diagram:

1. models can have a convex or concave potential, depending on the sign of $\frac{\partial^2 V}{\partial\varphi^2}$ or $\frac{\partial^2 f}{\partial\varphi^2}$ when observable scales exit the Hubble radius and $\varphi = \varphi_*$. Convex models ($\partial^2 V/\partial\varphi^2 \leq 0$) have $\eta_* < 0$ and belong to the region where $n_s \leq 1$ and $0 \leq r \leq -\frac{8}{3}(n_s - 1)$. Concave models ($\partial^2 V/\partial\varphi^2 \geq 0$) occupy the rest of the (r, N_s) plane.

2. like for new inflation, we can define Small Field Models (SFMs) as models for which $\varphi \ll m_P$ holds throughout inflation (until the very end) and Large Field Models (LFMs) as any other models. This distinction gives an information on the hierarchy between ϵ and η. Indeed, the slow-roll parameters read in the general case

$$\epsilon = \frac{m_P^2}{16\pi}\left(\frac{\partial f/\partial\varphi}{1+f}\right)^2 , \tag{224}$$

$$\eta = \frac{m_P^2}{8\pi}\left(\frac{\partial^2 f/\partial\varphi^2}{1+f}\right) . \tag{225}$$

   Usually, we can assume that $\partial f/\partial\varphi$ is of the same order of magnitude as f/φ, and that $\partial^2 f/\partial\varphi^2$ is of order f/φ^2. In particular, this is true as long as the "observable" inflationary potential can be approximated by a Taylor series truncated at a reasonable order. Then, the order of magnitude of the slow-roll parameters is given by

$$\epsilon \sim \frac{1}{16\pi}\left(\frac{m_P}{\varphi}\right)^2\left(\frac{f}{1+f}\right)^2 , \tag{226}$$

$$\eta \sim \frac{1}{8\pi}\left(\frac{m_P}{\varphi}\right)^2 \frac{f}{1+f} . \tag{227}$$

   For any small field model, the ratio m_P/φ is large: so, inflation can take place only if $f \ll 1$, which in turns imply $\epsilon \sim f\,|\eta| \ll |\eta|$. We conclude that $r = 16\epsilon_*$ is much smaller than $|n_s - 1| \simeq 2\eta_*$, and that in the (r, n_s) diagram SFMs are very close to the $r = 0$ axis. Instead, LFMs can have $f(\varphi_*)$ smaller or larger than one and ϵ smaller or larger than η: they can stand anywhere in the (r, n_s) plane.

Note that the previously studied "new inflation models" (respectively SFMs and LFMs) occupy the same region as the "convex hybrid models". This is not a surprise since the distinction between the two comes only from the overall sign of the function f, and in the above classification we made no assumption on this sign. Note also that the "monomial potential models" are included in the region of "concave hybrid LFMs". Again this is very reasonable since these models do have a concave potential, large field values, and can be seen as a limit of hybrid models with $f(\varphi) = (\lambda m_P^4/V_0)(\varphi/M_P)^\alpha$ for $V_0 \to 0$.

The class of hybrid models is so large that we cannot present a systematic study of all potentials. However, we will focus on three cases that will play a particular role in the section 4 (connection with high-energy physics).

Hybrid inflation with small quadratic term.

We will first study the potential

$$V_{eff}(\varphi) = V_0 + \frac{m^2}{2}\varphi^2 \tag{228}$$

which is concave, and can fall in the SFM or LFM category depending on parameter values. Here, let us focus only on the case where, by assumption, $\frac{m^2}{2}\varphi^2 \ll V_0$ throughout inflation (the opposite case is close to that of a monomial quadratic potential).

Exercise 1: show that the slow-roll parameters read

$$\epsilon = \frac{m_P^2}{16\pi}\frac{m^4\varphi^2}{V_0^2} , \qquad \eta = \frac{m_P^2}{8\pi}\frac{m^2}{V_0} , \tag{229}$$

that the assumption $\frac{m^2}{2}\varphi^2 \ll V_0$ implies $\epsilon \ll \eta$, and that the potential parameters must obey $m^2 m_P^2 \ll V_0$. Show that inflation ends not when slow-roll conditions break, but when the critical value φ_{end} triggering the phase transition is reached. Integrate dN between φ_* and φ_{end} and show that the integral is dominated neither by the φ_* boundary, neither by the φ_{end} boundary, but instead that

$$\varphi_* = \varphi_{end} \exp\left[\frac{m_P^2 m^2 N_*}{8\pi V_0}\right] . \tag{230}$$

Conclude that since φ_{end} is arbitrary, φ_* can take a priori any value.

So, in the $(r,\, n_s)$ plane, the category of models that we are studying stands close to the $r = 0$ axis, with $n_s \geq 1$. These models are not the most favored one at the moment, but they are not excluded with high significance. The preferred case is that with $n_s \simeq 1$, corresponding to the limit in which m^2 is smaller than V_0/m_P^2 by several orders of magnitude. If in the future the case $(r,\, n_s)$=(0,1) becomes excluded with high significance, these models will be ruled out.

Exercise 2: show that the constraint on the scalar power spectrum amplitude reads

$$\frac{V_0^3}{m_P^6 m^4 \varphi_*^2} \sim 2.10^{-11} . \tag{231}$$

By combining the above condition and the requirement that $\eta \ll 1$, show that for SFMs we can derive a bound for the mass, $m \ll 4.10^{-6} m_P$.

We can make a final comment on these models: there exists a branch of solution, called low-energy hybrid inflation, in which V_0 can be arbitrarily small and all the above constraints can still be satisfied. This is best seen by noticing that the amplitude constraint can be decomposed into

$$\frac{V_0}{m^2\varphi_*^2}\,\frac{V_0}{m_P^2 m^2}\,\frac{V_0}{m_P^4} \sim 2.10^{-11} . \tag{232}$$

On the left hand side, we have the product of three ratios. Suppose that we take an arbitrarily small value of V_0, i.e. of the third ratio. By adjusting m, we can keep the second ratio large enough so that η and $(n_s - 1)$ are as small as required by observations. The product of the three factors can be adjusted to 2.10^{-11} simply by choosing a small value of φ_{end} and φ_*, corresponding to a large value of the first ratio. This branch of solution has been used for building inflationary models at very low scales (with respect to the usual GUT scale) at the expense of very small, fine-tuned values of the mass and of φ_{end}.

Hybrid inflation with small quartic term.

We now turn to the case

$$V_{eff}(\varphi) = V_0 + \frac{\lambda}{4}\varphi^4 \tag{233}$$

which is concave, and can fall in the SFM or LFM category depending on parameter values. Here, let us focus only on the case where, by assumption, $\frac{\lambda}{4}\varphi^4 \ll V_0$ throughout inflation (the opposite case is close to that of a monomial quartic potential).

Exercise 1: show that the slow-roll parameters read

$$\epsilon = \frac{m_P^2}{16\pi}\left(\frac{\lambda\varphi^3}{V_0}\right)^2 , \qquad \eta = \frac{m_P^2}{8\pi}\,\frac{3\lambda\varphi^2}{V_0} , \tag{234}$$

and that the assumption $\frac{\lambda}{4}\varphi^4 \ll V_0$ implies $\epsilon \ll \eta$. Show that inflation must end with a waterfall transition, rather than by breaking the slow-roll conditions, and that we should impose

$$\varphi_*^2 \leq \frac{8\pi V_0}{3\lambda m_P^2} . \tag{235}$$

Integrate dN between φ_* and φ_{end} and show that the integral is dominated by the φ_{end} boundary, so that N_* and φ_{end} are related by

$$\varphi_{end}^2 \simeq \frac{4\pi V_0}{m_P^2 N_*} . \tag{236}$$

From the requirement that $\varphi_{end} < \varphi_*$, show that we get a weak constraint $\lambda < \frac{4}{3}N_*$.

So, in the $(r,\, n_s)$ plane, the category of models that we are studying stands close to the $r = 0$ axis, with $n_s \geq 1$. These models are not the most favored one at the moment, but they are not excluded with high significance. The preferred case is that with $n_s \simeq 1$, corresponding to the limit in which the arbitrary parameter φ_*^2 is chosen to be smaller than $V_0/(\lambda m_P^2)$ by several orders of magnitude. If in the future the case $(r,\, n_s)$=(0,1) becomes excluded with high significance, these models will be ruled out.

Exercise 2: show that the constraint on the scalar power spectrum amplitude reads

$$m_P^6 \frac{V_0^3}{(\lambda\varphi_*^3)^2} \sim \frac{\lambda}{\left(\frac{8\pi}{3}\eta_*\right)^3} \sim 2.10^{-11} . \tag{237}$$

By combining the above condition and the requirement that $\eta_* \ll 1$, show that we can derive a bound on the self-coupling constant, $\lambda \ll 10^{-8}$.

Here again, there exists a branch of solutions in which V_0 can be arbitrarily small.

Hybrid inflation with small logarithmic term.

The potential

$$V_{eff}(\varphi) = V_0\left[1 + \lambda \ln \frac{\varphi}{\varphi_{end}}\right] \tag{238}$$

with (by assumption) $\lambda \ln(\varphi/\varphi_{end}) \ll 1$ plays a particular role in models with spontaneously broken supersymmetry (as we shall see in section 4).

Exercise 1: show that the slow-roll parameters read

$$\epsilon = \frac{\lambda^2}{16\pi}\left(\frac{m_P}{\varphi}\right)^2 , \qquad \eta = -\frac{\lambda}{8\pi}\left(\frac{m_P}{\varphi}\right)^2 . \tag{239}$$

Integrate dN between φ_* and φ_{end} and show that the integral is dominated by the φ_* boundary. Conclude that irrespectively of the details of inflation ending,

$$\varphi_* \simeq \sqrt{\frac{\lambda N_*}{4\pi}}\, m_P \tag{240}$$

and conclude that

$$\epsilon_* \simeq \frac{\lambda}{4N_*} , \qquad \eta_* \simeq -\frac{1}{2N_*} . \tag{241}$$

Exercise 2: show that for SFMs we must assume $\lambda \ll 1$, obtaining

$$n_s \simeq 1 - \frac{1}{N_*} \; , \qquad r \ll |n_s - 1| \; . \tag{242}$$

Conclude that for $30 \leq N_* \leq 60$, all SFMs stand in the $(r,\, n_s)$ plane very close to the $r = 0$ axis with $0.96 \leq n_s \leq 0.98$.

Exercise 3: show that LFMs extrapolate between the above region and an asymptotic region close to the line $r = -\frac{8}{3}(n_s - 1)$.

We can see on Fig. 4 that this type of model is still allowed by observations, excepted in the large r region, which correspond to the large λ limit: actually, observations place an upper bound on λ, of the order of $\lambda < 1$.

Exercise 4: show that the constraint on the scalar power spectrum amplitude gives a relation between V_0 and λ

$$\frac{V_0}{\lambda} \sim \frac{8\pi . 10^{-11}}{N_*} m_P^4 \; , \tag{243}$$

and conclude that we obtain an upper limit on the energy scale of inflation

$$V_0^{1/4} < \frac{4.10^{-3}}{N_*} m_P \; . \tag{244}$$

4 Connection with high-energy physics

4.1 Fine-tuning issues

Generically, it is difficult to build inflationary models based on a Lagrangian motivated by high-energy physics, mainly for two reasons:

1. in all the examples studied in the last section, some of the potential parameters (dimensionless coupling constants, or mass scales in units of the Planck mass) must be fixed to very small values. The generic reason is the smallness of the scalar spectrum amplitude, associated to the constraint

$$m_P^{-6} \left. \frac{V^3}{[\partial V / \partial \varphi]^2} \right|_{\varphi_*} \simeq 2.10^{-11} \; . \tag{245}$$

   Unavoidably, the small number on the right-hand side has to come from small parameters in the potential. In high-energy physics, the problem with small parameters is that they have to be justified by some symmetry or some special mechanism, otherwise they appear as the result of some unrealistic fine-tuning. In addition (and this is related to the next point), naked parameters often receive some radiative corrections (i.e. need to be renormalized) in such way that plausible values, related to some renormalization scale, are usually far too large for inflation. Therefore, one should investigate possible symmetries justifying the small parameters and protecting them from large radiative corrections.

2. if one writes an inflationary potential assuming that it represents a sector of the ultimate theory describing Nature, one can in principle choose any potential provided that the theory remains renormalizable. However, in the real world, we usually assume that the theory governing the physics at the scale where inflation takes place (i.e., at most the GUT scale, given the limit on the tensor mode amplitude) is only an effective theory, i.e. a low-energy approximation of some theory valid at higher energy (like string theory). In this case, one should assume assume that the effective potential contains *a priori* all terms in the Taylor expansion

$$V(\varphi) = V_0 + \frac{1}{2} m^2 \varphi^2 + \lambda^4 \varphi^4 + \sum_{\alpha \geq 6} \lambda_\alpha m_P^4 \left(\frac{\varphi}{m_P} \right)^\alpha \; . \tag{246}$$

Terms with $\alpha \geq 6$ are non-renormalizable by construction but they do appear in the effective theory. In the above expansion, we can assume that the mass and/or some of the coupling constants are small –or exactly zero– *if and only if this is justified* by some symmetry or some argument based on the high-energy theory. We can then distinguish two cases:

- for Small Field Models (SFMs) with $\varphi \ll m_P$ throughout inflation, the series formed by the non-renormalizable terms can converge in the large-α limit even if the coupling constant λ_α are not particularly small. In that case, the theory can be kept under control, and the potential can match all inflationary criteria provided that a finite number of parameters are fine-tuned to small values. This can be eventually justified by some symmetry.
- for Large Field Models (LFMs) with $\varphi \geq m_P$, the series diverges and the potential is not appropriate for inflation, unless an infinite number of parameters λ_α are tuned to very small values. The model is then completely unrealistic from the point of view of particle physics, unless a given symmetry has the power to keep all λ_α couplings vanishingly small.

4.2 Global Supersymmetry models

Supersymmetry is a symmetry between fermions and bosons, introduced in order to solve various problems in particle physics, like the unification of gauge couplings for the electroweak and strong interactions, and the size of radiative corrections to the the Higgs mass. However, when imposed as a global symmetry, supersymmetry does not address the problem of the unification of gravity with other interactions: this is one of the motivations in favor of the more general theory called supergravity, in which supersymmetry is realized locally (like a gauge symmetry) rather than globally. In this section, we will assume that global symmetry is the "ultimate high-energy theory" and explore the consequences for inflation. In the next sections, we will adopt the more realistic point of view that supersymmetry is only an approximation of supergravity.

In "global SUperSYmmetry" (SUSY), the scalar potential is not a free function: it must be built in a very specific way, and can be split in two contributions called the F-term and D-term,

$$V = V_F + V_D \ . \tag{247}$$

The F-term is related to a function called the superpotential $W(\Phi_1, ..., \Phi_n)$, which must be a holomorphic function of order at most three in the complex scalar fields Φ_i:

$$W(\Phi_1, ..., \Phi_n) = \mu^3 + \sum_i \mu_i^2 \Phi_i + \sum_{i,j} \mu_{ij} \Phi_i \Phi_j + \sum_{i,j,k} \mu_{ijk} \Phi_i \Phi_j \Phi_k \ . \tag{248}$$

In presence of additional symmetries (like gauge symmetries), the expression of the superpotential W is usually very constrained and depends only on a small number of free parameters. By definition, the F-term is equal to

$$V_F = \sum_n \left| \frac{\partial W}{\partial \Phi_n} \right|^2 \tag{249}$$

where the sum runs over all the scalar fields Φ_n present in the theory. The D-term is related to symmetries. For instance, for a U(1) symmetry, one obtains a contribution

$$V_D = \frac{1}{2} g^2 \left(\sum_n q_n |\Phi_n|^2 \right)^2 \ , \tag{250}$$

where q_n is the charge of Φ_n under the U(1) symmetry. Actually, nothing prevents from adding a constant term ξ –called a Fayet-Illiopoulos term– inside the parenthesis,

$$V_D = \frac{1}{2} g^2 \left(\sum_n q_n |\Phi_n|^2 + \xi \right)^2 \ . \tag{251}$$

At the level of global supersymmetry, the motivation for ξ is the same as that for a cosmological constant from the point of view of Einstein: it is a term which can be put by hand, and respects the covariance of the theory.

It has been realized in the 1990's that global supersymmetry is an ideal framework for building inflationary models. The deep reason for that is twofold:

1. in supersymmetry, the scalar potential generally exhibits some *flat directions*: this means that when one field is staying at its equilibrium value, the potential does not depend on the other fields. So, along some directions in field space, the potential is constant.

2. supersymmetry leads to a cancellation between bosonic and fermionic loops resulting in the absence of radiative corrections for the scalar field parameters: there is a consequence of the *non-renormalization theorem* valid in SUSY.

Let us be a bit more precise:

1. when *supersymmetry is realized* in Nature, the cancellation between bosonic and fermionic loops is exact, leading to null radiative correction. Therefore, the flat directions are exactly flat. In addition, in this case, it can be shown that classically the fields must stand in a point where $V = 0$. So, flat directions appear as degenerate lines in field space where the potential exactly vanishes.

2. when *supersymmetry is spontaneously broken*, the difference between the mass of standard model particles and their superpartners leads to the apparition of small radiative corrections at the one-loop order. Flat directions are not exactly flat anymore. In addition, the tree-level contribution to the potential cannot be zero: SUSY breaking implies that classically the fields must stand in a point where $V \neq 0$ at tree-level.

In our universe, in order to obtain masses for the superpartners compatible with observational bounds, we must assume that supersymmetry is broken in a very special way called *soft supersymmetry breaking*, involving a hidden sector. The breaking terms are such that the scalar potential does not offer a good framework for inflation. At high energy, supersymmetry is restored, and the scalar fields must be in an equilibrium configuration where $V = 0$: again this is not suitable for inflation. The general idea declined in the literature is that inflation describes the dynamics of the fields rolling to this minimum – assuming that they start e.g. from chaotic initial conditions. When the rolling takes place along flat directions where the tree-level potential is constant and non-zero, and the dynamics is governed by small loop corrections, inflation can take place. The non-renormalization theorem (i.e., the flatness at tree level) enforces very small value of the quantities (m, λ, λ_α) in the Taylor expansion of Eq. 246: in this way, in global supersymmetry, one can obtain inflation in a natural way, without severe fine-tunings.

F-term inflation. A famous toy-model for supersymmetric inflation is based on three scalar fields Φ^+, Φ^- and S, assumed to have charges $+1$, -1 and 0 under a U(1) symmetry. The Fayet-Illiopoulos term is assumed to vanish. The most general superpotential $W(\Phi^+, \Phi^-, S)$ compatible with the above U(1) symmetry, plus another simple symmetry often used in SUSY model building (a continuous R-symmetry) reads

$$W = \alpha \Phi^+ \Phi^- S - \mu^2 S \ , \tag{252}$$

and leads to the following terms in the scalar tree-level potential

$$\begin{aligned} V_F &= |\alpha \Phi^+ \Phi^- - \mu^2|^2 + \alpha^2 |S|^2 \left(|\Phi^+|^2 + |\Phi^-|^2 \right) \ , & (253) \\ V_D &= \frac{g^2}{2} \left(|\Phi^+|^2 - |\Phi^-|^2 \right)^2 \ . & (254) \end{aligned}$$

This potential appears to be perfect for implementing hybrid inflation. Indeed, when $|S|$ is above some critical value,

$$|S|^2 \geq \frac{\mu^2}{\alpha} \ , \tag{255}$$

the charged fields remain in the equilibrium configuration in which

$$\Phi^+ = \Phi^- = 0 \tag{256}$$

and the tree-level potential is constant: $V = \mu^4$. This is a typical flat direction, lifted only by loop corrections. The Coleman-Weinberg formula predicts that, at the one-loop order, the effective potential reads

$$V_{eff} = \mu^4 + \frac{\alpha^2 \mu^4}{16\pi^2} \left(\frac{3}{2} + \ln \frac{\alpha^2 |S|^2}{\Lambda^2} \right) \tag{257}$$

where Λ is a renormalization scale, which should be taken of order $\Lambda \sim \mu\sqrt{\alpha}$. This is nothing but the previously considered potential of Eq. (238), with the role of the inflaton played by the canonically normalized modulus of S, defined to be $\varphi \equiv \sqrt{2}|S|$. The terms V_0 and λ of Eq. (238) can be identified as

$$V_0 \equiv \mu^4 \ , \qquad \lambda \equiv \frac{\alpha^2}{8\pi^2} \ . \tag{258}$$

At some point, the modulus of the S field will fall below the critical value

$$|S|_{crit} = \frac{\mu}{\sqrt{\alpha}} \tag{259}$$

triggering a "waterfall transition" to the true minimum

$$\Phi^+ \longrightarrow \frac{\mu}{\sqrt{\alpha}} e^{i\theta} \ , \ \Phi^- \longrightarrow \frac{\mu}{\sqrt{\alpha}} e^{-i\theta} \ , \ S \longrightarrow 0 \ , \tag{260}$$

where $V = 0$ and SUSY is restored (θ is an arbitrary phase). We already studied this model in subsection 3.4.4, and showed that it is compatible with observations for $\lambda \leq 1$, which just requires $\alpha \leq 2\sqrt{2}\pi$ (this sounds like a natural condition), and

$$\mu \sim \sqrt{\frac{\alpha}{2\sqrt{2}\pi} \frac{4.10^{-3}}{N_*}} m_P \ . \tag{261}$$

Note that if we assume that α is not very small with respect to one, the model belongs to the class of LFMs such that $\varphi_* \simeq \sqrt{\lambda N_*/(4\pi)} m_P \sim m_P$, $\mu \sim 10^{-3} m_P$ and inflationary gravitational waves could be detected with future experiments.

This toy model is a particularly simple illustration of SUSY F-term inflation. However, more complicated and better motivated scenarios have been built, relying sometimes on the breaking of the SUper-SYmmetric Grand Unified Theory (SUSY GUT) group, for instance SU(5); or involving more complicated schemes for breaking supersymmetry than the above spontaneous breaking (the one-loop potential can then be more complicated that above); or assuming that inflation takes place at low energy, around $V^{1/4} \sim 10^{10}$ GeV, in order to get a unified description of the mechanisms responsible for inflation and for the SUSY breaking still realized in our observable universe.

We will see later that when supergravity corrections are included, the situation is not as promising as one could think from these lines.

D-term inflation. It is also easy to build global supersymmetric model in which the term V_0 breaking supersymmetry is not contained in V_F, but in V_D. Taking the same three fields as in the previous toy-model, and assuming the same charges under the U(1) symmetry but a different R-symmetry, one is lead to the superpotential

$$W = \alpha \Phi^+ \Phi^- S \tag{262}$$

which is equal to the previous one with $\mu = 0$. However, we now put by hand a non-vanishing Fayet-Illiopoulos term ξ so that the two components of the scalar potential read

$$\begin{aligned} V_F &= \alpha^2 |\Phi^+ \Phi^-|^2 + \alpha^2 |S|^2 \left(|\Phi^+|^2 + |\Phi^-|^2 \right) \ , & (263) \\ V_D &= \frac{g^2}{2} \left(|\Phi^+|^2 - |\Phi^-|^2 + \xi \right)^2 \ . & (264) \end{aligned}$$

Again, this potential is perfect for hybrid inflation. When $|S|$ is above some critical value,

$$|S|^2 \geq \frac{g^2 \xi}{\alpha^2} \ , \tag{265}$$

the charged fields remain in the equilibrium configuration in which

$$\Phi^+ = \Phi^- = 0 \tag{266}$$

and the tree-level potential $V = \frac{1}{2} g^2 \xi^2$ is constant as a function of $|S|$. This is again a flat direction. The one-loop corrections which give a slope to V can be computed using the Coleman-Weinberg formula, and the effective potential along the (nearly) flat direction reads

$$V_{\mathrm{e}ff} = \frac{g^2 \xi^2}{2} + \frac{g^4 \xi^2}{16\pi^2} \left(\frac{3}{2} + \ln \frac{\alpha^2 |S|^2}{\Lambda^2} \right) \tag{267}$$

where Λ is a renormalization scale, which should be taken of order $\Lambda \sim g\sqrt{\xi}$. This is nothing but the previously considered potential of Eq. (238), with the role of the inflaton played by $\varphi \equiv \sqrt{2}|S|$. The terms V_0 and λ of Eq. (238) can be identified as

$$V_0 \equiv \frac{g^2 \xi^2}{2} \ , \qquad \lambda \equiv \frac{g^2}{8\pi^2} \ . \tag{268}$$

At some point, the modulus of the S field will fall below the critical value

$$|S|_{crit} = \frac{g\sqrt{\xi}}{\alpha} \tag{269}$$

triggering a "waterfall transition" to the true minimum

$$\Phi^+ \longrightarrow 0\ ,\ \Phi^- \longrightarrow \sqrt{\xi}e^{i\alpha}\ ,\ S \longrightarrow 0\ , \tag{270}$$

where $V = 0$ and supersymmetry is restored (θ is an arbitrary phase). We already studied this model in subsection 3.4.4, and showed that it is compatible with observations for $\lambda \leq 1$, which just requires $g \leq 2\sqrt{2}\pi$ (this sounds like a natural condition), and

$$\xi^{1/2} \sim \left(\frac{20}{\pi N_*}\right)^{1/4} 10^{-3} m_P\ . \tag{271}$$

Note that if we assume that g is not very small with respect to one, the model belongs to the class of LFMs such that $\varphi_* \simeq \sqrt{\lambda N_*/(4\pi)}\, m_P \sim m_P$ and inflationary gravitational waves could be detectable.

We will see in the next sections that this models resists better than F-term inflation to supergravity corrections, while from the point of view of string theory both models are under considerable pressure.

4.3 Supergravity models

SUperGRAvity (SUGRA) is a generalization of SUSY in which supersymmetry is local. This theory naturally includes gravitons, and is considered as a very promising step forward in view of unifying gravity with other interactions. In particle physics, one often encounters situations in which the formalism of global supersymmetry is sufficient for deriving results which are good approximations of the exact supergravity problem. We will see that this is hardly the case for studying inflation.

In supergravity, the scalar potential can still be split into an F-term and a D-term, obeying to construction rules which are similar to those of global supersymmetry, although slightly more complicated. In particular, it is still necessary to start from a superpotential W which is a holomorphic function of the fields. In order to obtain a renormalizable theory, one should limit W to be of order three in the fields; however, if supergravity is only an effective theory, terms are present at all orders. In addition, one should specify two more functions: the gauge kinetic function f, which is also a holomorphic function of the fields, and the Kähler potential K, which can be an arbitrary real function of the fields *and* their complex conjugates (with the dimension of a squared mass).

Here, we will not write the full expression of the scalar potential as a function of W, f and K, because our goal is not to enter into a detailed discussion of SUGRA-motivated inflationary models. It is sufficient to point out a few salient features.

F-term inflation. In supergravity, the potential V_F is always of the form

$$V_F = e^{8\pi K/m_P^2} [...]\ , \tag{272}$$

where the brackets contain a function of the fields related to (W, K) as well as their first and second order derivatives. Note that if we Taylor-expand the Kähler potential in the vicinity of the origin field space, we can always cancel the lowest-order terms through appropriate transformations and obtain without loss of generality

$$K = \sum_{mn} K_{mn} \Phi_m \Phi_n^* + \text{h}\textit{igher order terms}, \tag{273}$$

where the coefficients of the matrix K_{mn} are typically of order one. Let us assume that the radial field $\varphi \equiv \sqrt{2}|\Phi_1|$ plays the role of an inflaton. Then, the inflaton potential will start with

$$V_F = e^{4\pi K_{11}(\varphi/m_P)^2 + \cdots} [...] \tag{274}$$

and its derivatives will include the terms

$$\frac{\partial V_F}{\partial \varphi} = \frac{8\pi K_{11}\,\varphi}{m_P^2}\, V_F + ...\ , \tag{275}$$

$$\frac{\partial^2 V_F}{\partial \varphi^2} = \frac{8\pi K_{11}}{m_P^2}\, V_F + ...\ , \tag{276}$$

with a contribution to the second slow-roll parameter of the form

$$\eta = \frac{m_P^2}{8\pi}\frac{8\pi K_{11}}{m_P^2} + ... = K_{11} + ... \ . \tag{277}$$

Since K_{11} is generically of order one, η cannot be small. This very famous issue is called the *η-problem.* We conclude that in supergravity, the potential V_F violates the second slow-roll condition by construction, unless the expression of K and W is such that some extra terms cancel the one above almost exactly.

If supergravity is considered as the ultimate theory in Nature, rather than a low-energy effective theory, it is possible to choose a simple renormalizable superpotential, and to assume an arbitrary form for the Kälher potential. Then, it is not difficult to arrange for a cancellation leading to $|\eta| \ll |K_{11}|$. A famous example is called *minimal supergravity*, corresponding to the prescription $K = \sum_n |\Phi_n|^2$, i.e. $K_{mn} = \delta_{mn}$. In minimal supergravity and with the same superpotential as for global SUSY F-term inflation, the contribution to η of order one does vanish. So, at the level of supergravity, the problems affecting F-term inflation can be eliminated in a reasonable way; we will see however that they come back in all known string-motivated frameworks.

D-term inflation. In supergravity, the potential V_D receives corrections depending on the form of the Kähler potential and on the gauge kinetic function f. For instance, for a U(1) symmetry with a non-zero Fayet-Illiopoulos term,

$$V_D = \frac{1}{2}(\mathrm{Re}f)^{-1}g^2\left(\sum_n q_n \frac{\partial K}{\partial \Phi_n}\Phi_n + \xi\right)^2 \ . \tag{278}$$

We have seen that the Kähler potential tends to spoil F-term inflation; instead, it is completely irrelevant for D-term inflation (at the level of supergravity). Indeed, taking the same toy model as before with $W = \alpha\Phi^+\Phi^- S$ and a completely arbitrary function K, one finds that during inflation, when $\Phi^+ = \Phi^- = 0$, the superpotential and all its derivatives $\partial W/\partial \Phi_n$ vanish, leading to $V_F = 0$: in other words, the flat direction is preserved in the F-term. Meanwhile, the D-term reads

$$V_D = \frac{1}{2}(\mathrm{Re}f)^{-1}g^2\xi^2 \ . \tag{279}$$

So, at tree level, the flat direction is only lifted by the gauge kinetic function f: the slow-roll parameters are given by $(\mathrm{Re}f)^{-1}$ and its derivatives with respect to the inflaton field. This is not as problematic as the role of K in F-term inflation, first because f is holomorphic and can be easily constrained with some symmetries, and second because the potential depends on f^{-1} instead of e^f, so the contribution to ϵ and η is usually small provided that the inflaton field is slightly smaller than m_P during inflation. The order of magnitude of the inflaton is given by Eqs. (240, 268),

$$\varphi_* \sim \frac{g\, m_P}{4\pi}\sqrt{\frac{N_*}{2\pi}} \ , \tag{280}$$

so for g slightly smaller than one the f-correction is small and D-term inflation can take place like in global supersymmetry, with the same unique requirement

$$\xi^{1/2} \sim \left(\frac{20}{\pi N_*}\right)^{1/4} 10^{-3} m_P \ . \tag{281}$$

Again, we will see that problems come back in string-motivated frameworks.

4.4 Compatibility with string theory

If supergravity is an effective theory, i.e. a low-energy approximation of a more general theory like string theory, we expect that non-renormalizable terms will be present in W and f, and that the Kähler potential will include all terms compatible with the symmetries imposed by the underlying string theory.

F-term inflation. We have seen that this paradigm is plagued by the η problem, unless K and W have special forms leading to some exact cancellations (like, for instance, in minimal supergravity). People have searched for string theory constructions in which the low-energy SUGRA limit would have the required properties; however, it is very difficult to constrain the form of the effective Kähler potential, and no natural framework for string-motivated F-term inflation has ever been proposed: in other words, the current understanding is simply that symmetries that would cure the η problem do not exist in string theory.

D-term inflation. If there are non-renormalizable terms in the gauge kinetic function, we must assume that inflation takes place for $\varphi \ll m_P$ in order to avoid the resurrection of the problems that motivated the introduction of supersymmetric inflation: namely, the appearance of a polynomial tree-level potential for the inflaton, deriving from the $(\mathrm{R}ef)^{-1}$ factor in V_D, and requiring an infinite number of fine-tunings. The condition $\varphi \ll m_P$ is fulfilled in our toy-model by requiring that g is slightly smaller than one. However, in more realistic examples, the one-loop potential often receives a correction of the type

$$V_{\mathrm{e}ff} = \frac{g^2\xi^2}{2} + C\frac{g^4\xi^2}{16\pi^2}\left(\frac{3}{2} + \ln\frac{\alpha^2|S|^2}{\Lambda^2}\right) \tag{282}$$

where the correction factor C can be as large as one hundred or so. In this case, g must be fine-tuned by an extra factor $\sqrt{C}$ in order to get $\varphi \ll m_P$, and this is difficult to motivate from string theory. Even more problematic is the origin of the Fayet-Illiopoulos term. There exists a known mechanism for generating $\xi \neq 0$. In string theory, one is often lead to introduce anomalous U(1) symmetry, i.e. symmetries with $\sum_n q_n \neq 0$. In that case, the so-called Green-Schwartz mechanism of anomaly cancellation generates a Fayet-Illiopoulos term slightly smaller than m_P^2. This term is generically a bit too large for satisfying the normalization condition of Eq. (283), but the small discrepancy could be solved in presence of the previously mentioned correction factor C, since in that case we just need to obtain

$$\xi^{1/2} \sim \left(\frac{20C}{\pi N_*}\right)^{1/4} 10^{-3} m_P \ . \tag{283}$$

This framework has been though for a while to be appropriate for inflation. Unfortunately, it was realized that in these models the dilaton field (which is always present in string theory) always get a runaway potential, and must be stabilized in order to avoid a large running of fundamental constants. The stabilization mechanism generically involves an F-term, and one is back to the original η problem of F-term inflation.

In order to avoid the runaway dilaton potential, one could try to introduce the Fayet-Illiopoulos term by hand, for a non-anomalous U(1) symmetry. In this case however, $\xi^{1/2}$ is expected to be naturally of order m_P. Then, in order to match the normalization constraint, one should motivate a huge number C. So it is impossible to get the correct order of magnitude for $\xi^{1/2}C^{-1/4}$ without invoking some strong fine-tuning.

In summary, all attempts to implement inflation in a string-motivated framework in which supersymmetry would preserve flat directions have failed (or require some non-negligible amount of fine-tuning). The generic problem is that the non-zero potential energy V_0 necessary for inflation breaks supersymmetry in such a way that flat directions can be preserved at the level of SUSY or even SUGRA when these theories are regarded as fundamental ones, but not when they are derived as effective theories.

4.5 State of the art of inflationary model building

Currently, a fraction of the community still investigates new string theory frameworks in which supersymmetric flat directions could be preserved at inflationary energy scales in the low-energy, four-dimensional effective theory. However, many people are pessimistic about this direction, and retain that supersymmetry fails in imposing flat direction during inflation. So, it is more fashionable in these days to investigate the consequences of other types of symmetries for inflation. For instance, various symmetry breaking mechanisms lead naturally to the existence of Pseudo Nambu-Goldstone Bosons (PNGB). Usual Goldstone bosons appear when the scalar field potential has a degenerate minimum (in the case of a Mexican hat potential, the degenerate minimum is a circle). The minimum valley can be seen as an exactly flat direction. In the case of a PNGB, the symmetry of the potential is slightly broken, and the valley of degenerate minima receives small periodic corrections, typically of the form

$$V = V_0[1 + \cos(\varphi/\mu)] \tag{284}$$

where the scalar field φ is the phase of a complex field, and μ is a mass scale. This potential is reminiscent of that of new inflation with a leading quadratic term,

$$V = V_0 \left[2 - \frac{1}{2} \left(\frac{\varphi}{\mu} \right)^2 + \mathcal{O}(\varphi^4) \right] . \qquad (285)$$

This framework opens an opportunity for slow-roll inflation, since the slope of the potential arises from a symmetry breaking term which is predicted to be small in many set-ups, ensuring a nearly flat potential. However, we have seen in subsection 3.4.4 that that "quadratic new inflation" falls necessarily in the category of large-fields models. Then, the flatness of the potential at tree-level tends to be spoiled by radiative corrections, and in particular by the infinite hierarchy of large, non-renormalizable terms.

A current direction of investigation is to find realistic models from the point of view of high-energy physics in which the inflaton is a PNGB, and in which, for some reason, the quadratic term vanishes. If this is the case, the model can be assumed to be a SFMs with negligible non-renormalizable terms. Some frameworks like the Little Higgs mechanism could lead to such a nice situation. On can also combine the advantages of PNGB and supersymmetry by building supersymmetric PNGB models. Finally, one can study PNGBs in the context of large extra dimension models. In that case, the interpretation of the Planck mass is radically different, and the fundamental scale of gravity can be lower than m_P: then, the fact that a model appears as a large-field model in the effective theory does not necessarily imply large radiative corrections. Various inflationary models based on large extra dimension have been proposed; for instance, the inflaton could be a radion, i.e. a scalar field representing some inter-brane distance.

4.6 Prospects for observations and theoretical developments

The future of inflationary model building will depend very much on the observation of primordial gravitational waves. In summary, we can make the following statements. Future experiments are expected to detect gravitational waves provided that $r \geq 0.01$ for future CMB experiments optimized for B-mode detection, i.e. beyond *Planck*, or provided that $r \geq 10^{-3} - 10^{-4}$ for the next generation of spatial gravitational wave interferometers beyond *LISA*. We have seen that such levels correspond typically to LFMs where the inflaton expectation value is of the same order of magnitude as the Planck mass, either during observable inflation, or at least at the end of inflation. So, future experiments should be able to discriminate between SFMs and LFMs. Let us review the possible scenarios for the future, as well as their implications:

- primordial gravitational waves might never be detected. From the point of view of high-energy physics, this would sound like good news, because it would mean that inflation is implemented as a SFM and can be described with the laws of standard quantum field theory, with sub-Planckian fields and small radiative corrections. But in this case, we could never measure more than two inflationary parameters: the scalar tilt and the scalar amplitude, already constrained today. The inflationary energy scale could not be measured. Evidences for inflation would remain, like today, strong but indirect, and presumably, one could only make vague conjectures about the implementation of inflation in particle physics models.

- primordial gravitational waves might be detected. We can then distinguish too sub-cases:

  - either the gravitational waves are just around the corner, with $r \sim 0.1$, and we are about to see them. In that case it would be possible to measure also the tensor tilt with reasonable precision. That would be fantastic because we could compare the value of n_t with that of r. Single-field slow-roll inflation predicts a relation $r = 16\epsilon = -8n_t$ called the *inflationary self-consistency relation*. By confirming this relation experimentally, one would have a very clear, direct proof of inflation. If $-8n_t$ was found to be a bit different from r, there would be some evidence for multiple-field inflation. In any case this situation would be extremely rich and interesting for early universe physics.
  - either r is in the range between 10^{-4} and 10^{-2}, so that primordial gravitational waves will be detect but their tilt will be hardly measured. In this case, the self-consistency relation could not be probed, but still we would know the energy scale of inflation, and get independent measurements of V_*, V'_* and V''_*.

  So, in these two sub-cases we would get a lot of new information, but paradoxically inflationary model building could be stuck by the evidence for new physics. Indeed, we have seen that r can be detectable only in LFMs, with field values around or greater than the Planck mass. In ordinary quantum field theory this raises the problem of large radiative correction and non-renormalizable terms, and we have seen that in effective theories supersymmetry cannot solve this problem. So, this situation would really require some new high-energy physics set-up, which would actually be a rather exciting situation. In particular, the detection of inflationary gravitational waves could be seen as an indication in favor of large extra dimensions.

Acknowledgments

I wish to thank Ruth Durrer for her kind invitation, as well as all participating students for their very stimulating interest.