TASI 2012 Lectures on Inflation

Leonardo Senatore

Stanford Institute for Theoretical Physics
Department of Physics, Stanford University, Stanford, CA 94306

Kavli Institute for Particle Astrophysics and Cosmology,
Stanford University and SLAC, Menlo Park, CA 94025

Abstract

Planning to explore the beginning of the Universe? A lightweight *guide du routard* for you.

1 Introduction

The purpose of these TASI lectures on Inflation is to introduce you to the currently preferred theory of the beginning of the universe: the theory of Inflation. This is one of the most fascinating theories in physics. Starting from the shortcomings of the standard big bang theory, we will see how a period of accelerated expansion solves these issues. We will then move on to explain how inflation can give such an accelerated expansion (**lecture 1**). We will then move on to what is the most striking prediction of inflation, which is the possibility that quantum fluctuations during this epoch are the source of the cosmological perturbations that seed galaxies and all structures in the universe (**lecture 2**). We will then try to generalize the concept of inflation to develop a more modern description of this theory. We will introduce the Effective Field Theory of Inflation. We will learn how to compute precisely the various cosmological observables, and how to simply get the physics out of the Lagrangians (**lecture 3**). Finally, in the last lecture (**lecture 4**), we will discuss one of the most important observational signatures of inflation: the possible non-Gaussianity of the primordial density perturbation. We will see how a detection of a deviation from gaussianity would let us learn about the inflationary Lagrangian and make the sky a huge particle detector. Time permitting, we will discuss one of the conceptually most beautiful regimes of inflation, the regime of eternal inflation, during which quantum effects become so large to change the asymptotics of the whole space-time.

Notation

$$c = \hbar = 1\ , \qquad M_{\rm Pl}^2 = \frac{1}{8\pi G}\ . \tag{1}$$

2 Lecture 1

Intro on Inflation: how brave it was.

Notice that we will perform calculations more explicitly when they are less simple. So in this first lecture we will skip some passages. General homework of this class: fill in the gaps.

2.1 FRW cosmology

We begin by setting up the stage with some basic concepts in cosmology to highlight the short-coming of the standard big bang picture. Some of this lecture has overlap with Prof. Bertshinder's lectures.

The region of universe that we see today seems to be well described by an homogenous and isotropic metric. The most general metric satisfying these symmetries can be put in the following form

$$ds^2 = -dt^2 + a(t)^2\left(\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right) \tag{2}$$

We see that this metric represents a slicing of space-time with spatial slices Σ that are simply rescaled by the scale factor a as time goes on. If $k = 0$, we have a flat space, if $k = +1$, the space Σ describes the sphere, while if $k = -1$ we have an hyperbolic space. A fundamental quantity is of course the Hubble rate

$$H = \frac{\dot{a}}{a} \tag{3}$$

which has units of inverse time. It is useful for us to put the metric (2) into the following form

$$ds^2 = -dt^2 + a(t)^2\left(d\chi^2 + S_k(\chi^2)\left(d\theta^2 + \sinh^2\theta d\phi^2\right)\right) \tag{4}$$

where

$$r^2 = S_k(\chi^2) = \begin{cases} \sin^2\chi & \text{if } k = -1, \\ \chi^2 & \text{if } k = 0, \\ \sin^2\chi & \text{if } k = +1. \end{cases} \tag{5}$$

χ plays the role of a radius. Let us now change coordinates in time (it is General Relativity at the end of the day!) to something called conform time

$$\tau = \int^\tau \frac{dt}{a(t)}\ . \tag{6}$$

Now the FRW metric becomes

$$ds^2 = a(\tau)^2\left[-d\tau^2 + d\chi^2 + S_k(\chi^2)\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \tag{7}$$

In these coordinates it is particular easy to see the casual structure of space. This is determined by how light propagates on null geodesic $ds^2 = 0$. Since the space is isotropic, geodesic solutions have constant θ and ϕ. In this case we have

$$\chi(\tau) = \pm\tau + \text{const.} \tag{8}$$

These geodesics move at 45 degree in the $\tau - \chi$ plane, as they would in Monkowsky space. This is so because apart for the angular part, the metric in (7) is conformally flat: light propagates as in Minwkosky space in the coordinates $\tau - \chi$. Notice that this is not so if we had used t, the proper time for comoving (i.e. fixed FRW-slicing spatial coordinates) observers.

figure

It is interesting to notice that is we declare that the universe started at some time t_i, then there is a maximum amount of time for light to have travelled. A point sitting at the origin of space (remember that we are in a translation invariant space), by the time t could have sent a signal at most to a point at coordinate χ_p given by

$$\chi_p(\tau) = \tau - \tau_i = \int_{t_i}^{t} \frac{dt}{a(t)} \tag{9}$$

The difference is conformal time is equal to the maximum coordinate-separation a particle could have travelled. Notice that the geodesic distance on the spacial slice between two point one particle horizon apart is obtained by multiplying the coordinate distance with the scale factor:

$$d_p(t) = a(\tau)\chi_p(\tau) \tag{10}$$

The presence of an horizon for cosmologies that begin at some definite time will be crucial for the motivation of inflation.

It will be interesting for us to notice that there is a different kind of horizon, called event horizon. If we suppose that time ends at some point t_{end} (sometimes this t_{end} can be taken to ∞), then there is a maximum coordinate separation between two points beyond which no signal can be sent from the first point to reach the second point by the time t_{end}. This is called event horizon, and it is the kind of horizon associated to a Schwartshild black hole. From the same geodesic equation, we derive

$$\chi_e(\tau) = \tau_{end} - \tau = \int_{\tau}^{\tau_{\rm end}} \frac{dt}{a(t)} \tag{11}$$

Clearly, as $\tau \to \tau_{\rm end}$, $\chi_e \to 0$.

We have seen that the casual structure of space-time depends on when space-time started and ended, and also on the value of $a(t)$ at the various times, as we have to do an integral. In order to understand how $a(t)$ evolves with time, we need to use the equations that control the *dynamics* of the metric. These are the Einstein equations

$$G_{\mu\nu} = \frac{T_{\mu\nu}}{M_{\rm Pl}^2} \ . \tag{12}$$

These in principle 10 equations reduce for an FRW metric to just two. Indeed, by the symmetries of space-time, in FRW slicing, we must have

$$T^\mu{}_\nu = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix} \tag{13}$$

and the Einstein equations reduce to

$$H^2 = \left(\frac{\dot a}{a}\right)^2 = \frac{1}{3M_{\rm Pl}^2}\rho - \frac{k}{a^2} \tag{14}$$

$$\dot H + H^2 = \frac{\ddot a}{a} = -\frac{1}{6}\left(\rho + 3p\right) \ . \tag{15}$$

The first equation is known as Friedamnn equation. These two equations can be combined to give the energy conservation equation (this follows from the Bianchi identity $0 = \nabla_\mu G^\mu_\nu = \nabla_\mu T^\mu_\nu$):

$$\frac{d\rho}{dt} + 3H(\rho + p) = 0 \tag{16}$$

This is a general-relativistic generalization of energy conservation. (Homework: make sense of it by considering dilution of energy.) By defining a constant equation of state w

$$p = w\rho \ , \tag{17}$$

energy conservation gives

$$\rho \propto a^{-3(1+w)} \tag{18}$$

and

$$a(t) \propto \begin{cases} t^{\frac{2}{3(1+w)}} & w \neq 1 \\ e^{Ht} & w = -1 \ . \end{cases} \tag{19}$$

Notice that indeed $\rho_{matter} \propto a^{-3}$, $\rho_{radiation} \propto a^{-4}$. Notice also that is a is power low with t to an order one power, than $H \sim 1/t$. That is, the proper time sets the scale of H at each time.

The standard big bang picture is the one in which it is hypothesized that the universe was always dominated by 'normal' matter, with $w > 0$. In order to see the shortcomings of this picture, it is useful to define the present energy fractions of the various constituents of the universe. If we have various components in the universe

$$\rho = \sum_i \rho_i \ , \qquad p = \sum_i p_i \ , \quad w_i = \frac{p_i}{\rho_i}. \tag{20}$$

We can define the present energy fraction of the various components by dividing each density by the 'critical density' ρ_{cr}(the density that would be required to make the universe expand with rate H_0 without the help of anything else)

$$\Omega_{i,0} = \frac{\rho_0^i}{\rho_{cr,0}} \tag{21}$$

We also define

$$\Omega_{k,0} = -\frac{k}{a(t_0)^2 H_0^2} \tag{22}$$

as a measure of the relative curvature contribution. By setting as it us usually done $a(t_0) = a_0 = 1$, we can recast the Friedmann equation in the following form

$$\left(\frac{H^2}{H_0^2}\right) = \sum_i \Omega_{i,0} a^{-3(1+w_i)} + \Omega_{k,0} a^{-2} \tag{23}$$

At present time we have $\sum_i \Omega_{i,0} + \Omega_{k,0} = 1$.

One can define also time dependent energy fractions

$$\Omega_i(a) = \frac{\rho_i(a)}{\rho_{cr}(a)} \ , \qquad \Omega_k(a) = -\frac{k}{a^2 H^2(a)} \tag{24}$$

Notice that $\rho_{cr} = 3M_{\rm Pl}^2 H^2$ is indeed time dependent. The Friedmann equation becomes

$$\Omega_k(a) = 1 - \sum_i \Omega_i(a) \tag{25}$$

2.2 Big Bang Shortcomings

We are now going to highlight some of the shortcoming of the big bang picture that appear if we assume that its history has always been dominated by some form of matter with $w \geq 0$. We will see that upon this assumptions, we are led to very unusual initial conditions. Now, this leads us to a somewhat dangerous slope, which catches current physicists somewhat unprepared. Apart for cosmology, Physics is usually the science that predicts the evolution of a certain given initial state. No theory is general given for the initial state. Physicists claim that if you tell them on which state you are, they will tell you what will be your evolution (with some uncertainties). The big bang puzzles we are going to discover are about the very peculiar initial state the universe should have been at the beginning of the universe if 'normal' matter was always to dominate it. Of course, it would be nice to see that the state in which the universe happens to begin in is a natural state, in some not-well defined sense. Inflation was indeed motivated by providing an attractor towards those peculiar looking initial conditions [1]. We should keep in mind that there could be other reasons for selecting a peculiar initial state for the universe.

[1] Luckly, we will see that inflation does not do just this, but it is also a predictive theory.

2.2.1 Flatness Problem

Let us look back at

$$\Omega_k(a) = -\frac{k}{a^2H^2(a)} \ , \tag{26}$$

and let us assume for simplicity that the expansion is dominated by some form of matter with equation of state equal to w. We have then $a \sim t^{\frac{2}{3(1+w)}}$ and we have

$$\dot{\Omega}_k = H\Omega_k(1+3w) \ , \qquad \frac{\partial \Omega_k}{\partial \log a} = \Omega_k(1+3w) \tag{27}$$

If we assume that $w > -1/3$, then this shows that the solution $\Omega_k = 0$ is un unstable point. If $\Omega_k > 0$ at some point, Ω_k keeps growing. Viceversa, if $\Omega_k < 0$ at some point, it keeps decreasing. Of corse at most $\Omega_k = \pm 1$, in which case $w \to -1/3$ if $k < 0$, or otherwise the universe collapses if $k > 0$.

The surprising fact is that Ω_k is now observed to be smaller than about 10^{-2}: very close to zero. Given the content of matter of current universe, this mean that in the past it was even closer to zero. For example, at the BBN epoch, it has to be $|\Omega_k| \lesssim 10^{-18}$, at the Planck scale $|\Omega_k| \lesssim 10^{-63}$. In other words, since curvature redscifts as a^{-2}, it tends to dominate in the future with respect to other forms of matter (non relativistic matter redshifts as a^{-3}, radiation as a^{-4}). So, if today curvature is not already dominating, it means that it was very very very negligible in the past. The value of Ω_k at those early times represents a remarkable small number. Why at that epoch Ω_k was so small?

Of course one solution could be that $k = 0$ in the initial state of the universe. It is unknown why the universe should choose such a precise state initially, but it is nevertheless a possibility. A second alternative would be to change at some time the matter content of the universe, so that we are dominated by some matter content with $w < -1/3$. We will see that inflation provides this possibility in a very simple way [2].

2.2.2 Horizon Problem

An even more dramatic shortcoming of the standard big bang picture is the horizon problem. Let us assume again that the universe is dominated by some form of matter with equation of state w. Let us compute the particle horizon:

$$\chi_p(\tau) = \tau - \tau_i = \int_{\tau_i(t_i)}^{\tau(t)} \frac{dt'}{a(t')} = \int_{a_i}^{a} \frac{da}{Ha^2} \sim a^{(1+3w)/2} - a_i^{(1+3w)/2} \tag{28}$$

We notice that if $w > -1/3$ (notice, the same $-1/3$ as in the flatness problem), then in an expanding universe the horizon grows with time. This is very bad. It means that at every instant of time, new regions that had never been in causal contact before come into contact for the first time. This means that they should look like very different from one another (unless the universe did not decide to start in a homogenous state). But if we look around

[2]Another possibility would be to imagine the universe underwent a period of contraction, like in the bouncing cosmologies. Curvature becomes subdominant in a contracting universe.

us, the universe seems to be homogenous on scales that came into causal contact only very recently. Well, maybe they simply equilibrate very fast? Even if this unlikely possibility were to be true, we can make the problem even sharper when we look at the CMB. In this case we can take a snapshot of casually disconnected regions (at the time at which they were still disconnected), and we see that they look like the same. This is the horizon problem.

Notice that if $w > -1/3$ the particle horizon is dominated by late times, and so we can take $a_i \simeq 0$ in its expression. In this way we have that the current physical horizon is

$$d_p \sim a\tau \sim t \sim \frac{1}{H} \ . \tag{29}$$

For this kind of cosmologies where $w > -1/3$ at all times, the Hubble length is of order of the horizon. This is what has led the community to often use the ill-fated name 'horizon' for 'Hubble'. 'Hubble is the horizon' is true only for standard cosmologies, it is not true in general. We will try to avoid calling Hubble as the horizon in all of these lectures, even though sometimes habit will take a toll.

Notice however that the horizon problem goes away if we assume the universe sit there for a while at the singularity.

Let us look again at the CMB. Naive Horizon scale is one degree ($l \sim 200$), and fluctuations are very small on larger scales. How was that possible?

Apart for postulating an ad hoc initial state, we would need also to include those perturbations in the initial state. this is getting crazy! (though possible) We will see that inflation will provide an attractive solution.

The problem of the CMB large scale fluctuations is a problem as hard as the horizon one.

figure

2.2.3 Solving these problems: conditions

In order to solve these two problems, we need to have some form of energy with $w < -1/3$. We can say it somewhat differently, by noticing that in order for Ω_k to decrease with time, since

$$\Omega_k = -\frac{k}{(aH)^2} \tag{30}$$

we want an epoch of the universe in which aH increases with time. Equivalently, $1/(aH)$ decreases with time. $1/(aH)$ is sometimes called 'comoving Horizon', ... a really bad name in my humble opinion. You can notice that since $1/H$ is the particle horizon in standard cosmologies, $1/(aH)$ identifies the comoving coordinate distance between two points one naive-Horizon apart. If this decreases with time, then one creates a separation between the true particle horizon, and the naive particle horizon. Two points that naively are separated by a $1/(aH)$ comoving distantce are no more separated by a particle horizon. Even more simply, the formula for the particle horizon reads

$$\tau = \int_{t_i}^{t} \frac{da}{(aH)^2} \tag{31}$$

If $(aH)^{-1}$ is large in the past, then the integral is dominated by the past, and the actual size of the horizon has nothing to do with present time quantities such as the Hubble scale at present. in standard cosmologies the opposite was happening: the integral was dominated by late times.

Let us formulate the condition for $(aH)^{-1}$ to decrease with time in equivalent forms.

- Accelerate expansion: it looks like that this condition implies that the universe must be accelerating in that epoch:

$$\frac{\partial \frac{1}{(aH)}}{\partial t} < 0 \quad \Rightarrow \quad \ddot{a} > 0 \tag{32}$$

  This implies that $k/(aH)$ decreases: physical wavelengths become longer than H^{-1}.

- As we stressed, this should imply $w < -1$. Let us verify it. From Friedman equation

$$0 < \ddot{a} = -\frac{a}{6}(\rho + 3p) = \frac{a\rho}{6}(1+3w) \quad \Rightarrow \quad w < -1/3 \quad \text{if } \rho > 0 \tag{33}$$

Inflation, in its most essential definition, is the postulation of a phase with $w < -1/3$ in the past of our universe [3].

Is it possible to see more physically what is going on? In a standard cosmology, the scale factor goes to zero at finite conformal time. For $w > -1/3$, we have that

$$a \sim \tau^{2/(1+3w)} \tag{34}$$

implying the existence of a singularity $a \to 0, H \to \infty$ as $\tau \to 0$. This is why we had to stop there. This is the big bang moment in standard cosmology. This however implies that

[3] If there is only one field involved, than scale invariance of the perturbations forces $w \simeq -1$. This is a theorem.

there is a beginning of time, and that the particle horizon is order τ. This is the source of the problems we discussed about.

However, if we have a phase in which $w < -1/3$, then the singularity in the past is pushed way further back, and the actual universe is much longer than what τ indicates. For example, for inflation $H \sim$ const. and $a(\tau) = -\frac{1}{H\tau}$, with $\tau \in [-\infty, \tau_{end}]$, $\tau_{end} \leq 0$. In general τ can be extended to negative times, in this way making the horizon much larger than $1/H$.

figure

2.3 The theory of Inflation

Inflation is indeed a period of the history of the universe that is postulated to have happened before the standard big bang history. Direct observation of BBN products tell us that the universe was radiation dominated at $t \sim 1 - 100$ sec, which strongly suggests that inflation had to happen at least earlier than this. More specifically, inflation is supposed to be a period dominated by a form of energy with $w \simeq -1$, or equivalently $H \simeq$ const. How can this be achieved by some physical means?

2.3.1 Simplest example

The simplest example of a system capable of driving a period of inflation is a scalar field on top a rather flat potential. These kinds of models are called ‘slow roll inflation’ and were the ones initially discovered to drive inflation. Let us look at this

figure

The scalar field plus gravity has the following action

$$S = \int d^4x\sqrt{-g}\left[\frac{M_{\rm Pl}^2}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] \tag{35}$$

The first term is the Einstein Hilbert term of GR. The second and third terms represent the action of a scalar field S_ϕ. The idea of inflation is to fill a small region of the initial universe with an homogeneously distributed scalar field sitting on top of its potential $V(\phi)$. Let us see what happens, by looking at the evolution of the space-time. We need the scalar field stress tensor:

$$T^{(\phi)}_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_\phi}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}\partial_\rho\phi\partial^\rho\phi + V(\phi)\right) \tag{36}$$

For an homogenous field configuration, this leads to the following energy density and pressure

$$\rho_\phi = \frac{1}{2}\dot\phi^2 + V(\phi) \quad \text{obviously} \tag{37}$$

$$p_\phi = \frac{1}{2}\dot\phi^2 - V(\phi) \quad \text{notice the sign of } V \tag{38}$$

Therefore the equation of state is

$$w_\phi = \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot\phi^2 - V(\phi)}{\frac{1}{2}\dot\phi^2 + V(\phi)}. \tag{39}$$

We see that if the potential energy dominates over the kinetic energy, we have

$$\dot\phi^2 \ll V(\phi) \quad \Rightarrow \quad w_\phi \simeq -1 < -\frac{1}{3} \tag{40}$$

Notice that this means that

$$\epsilon = -\frac{\dot H}{H^2} \sim \frac{\dot\phi^2}{V} \ll 1\ . \tag{41}$$

The equation of motion for the scalar field is

$$\frac{\delta S}{\delta\phi} = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\partial^\mu\phi) + V_{,\phi} = 0 \quad \Rightarrow \quad \ddot\phi + 3H\dot\phi + V_{,\phi} = 0 \tag{42}$$

This equation of motion is the same as the one of a particle rolling down its potential. This particle is subject to friction though the $H\dot\phi$ term. Like for a particle trajectory, this means that the solution where $\dot\phi \simeq V_\phi/(3H)$ is an attractor 'slow-roll' solution if friction is large enough. Being on this trajectory requires

$$\eta = -\frac{\ddot\phi}{H\dot\phi} \ll 1 \tag{43}$$

We have therefore found two 'slow roll parameters'.

$$\epsilon = -\frac{\dot H}{H^2} \ll 1\ , \qquad \eta = -\frac{\ddot\phi}{H\dot\phi} \ll 1 \tag{44}$$

The first parameters being much smaller than one means that we are on a background solution where the Hubble rate changes very slowly with time. The second parameter means that we are on an attractor solution (so that the actual solution does not depend much from the initial conditions), and also that this phase of accelerated expansion ($w \simeq -1$, $a \sim$ Exp(Ht)) will last for a long time. Indeed, one can cheek that

$$\frac{\dot{\epsilon}}{H\epsilon} \sim \mathcal{O}(\epsilon, \eta) \ . \tag{45}$$

We will see that the smallness of η is really forced on us by the scale invariance of the cosmological perturbations.

Once we assume we are on the slow roll solution, then we can express them in terms of the potential terms. We have

$$\epsilon \simeq \frac{M_{\rm Pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \ , \qquad \eta \simeq M_{\rm Pl}^2 \frac{V_{,\phi\phi}}{V} - \frac{M_{\rm Pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \ . \tag{46}$$

On this solution we also have

$$\dot{\phi} \simeq \frac{V_{,\phi}}{3H} \ , \qquad H \simeq \frac{V(\phi)}{3M_{\rm Pl}^2} \simeq \text{const} \ , \qquad a \sim e^{3Ht} \ . \tag{47}$$

When does inflation end? By definition, inflation ends when w ceases to be close to -1. This means that

$$\epsilon \sim \eta \sim 1 \ . \tag{48}$$

More concretely, we see that the field that starts on top of his potential will slowly roll down until two things will happen: Hubble will decrease, providing less friction, the potential will become too steep to guaranteed that the kinetic energy is negligible with respect to potential energy. We call the point in field space where this happens ϕ_{end}. At that point, a period dominated by a form of energy with $w > -1/3$ is expected to begin. We will come back in a second on it.

Duration of Inflation: For the moment, let us see how long inflation needs to last. The number of e-foldings of inflation is defined as the logarithm of the ratio of the scale factor at the end of inflation and at the beginning of inflation. For a generic initial point ϕ, we have

$$N^{to\ end}(\phi) = \log\left(\frac{a_{end}}{a}\right) \simeq \int_t^{t_{end}} H dt = \int_\phi^{\phi_{end}} \frac{H}{\dot{\phi}} d\phi \simeq \int_{\phi_{end}}^{\phi} \frac{V}{V_{,\phi}} d\phi \ , \tag{49}$$

where in the third passage we have used that $a \sim e^{Ht}$, and in the last passage we have used the slow roll solutions.

The horizon and flatness problems are solved in inflation very simply. During inflation

$$\Omega_k = -\frac{k}{a^2 H^2} \propto \frac{1}{a^2} \to 0. \tag{50}$$

So, if we start with $\Omega_k \sim 1$ at the onset of inflation, and we wish to explain why $\Omega_k(a_{BBN}) \sim 10^{-18}$, we need about 20 e-foldings of inflation. This is so because at the end of inflation we have

$$\Omega_k(a_{end}) \simeq \Omega_k(a_{in}) \frac{a_{in}^2}{a_{end}^2} \sim \frac{a_{in}^2}{a_{end}^2} = 2\log N \tag{51}$$

and this must be equal to curvature we expect at the beginning of the FRW phase (that we can assume to be equal to the end of inflation)

$$\Omega_k(a_{end}) = \Omega_k(a_0)\frac{a_{end}^2 H_I^2}{a_0^2 H_0^2} \sim 10^{-2}\frac{a_{end}^2 H_I^2}{a_0^2 H_0^2} \quad \Rightarrow \quad N = \log\left(\frac{a_{end} H_I}{a_0 H_0}\right) \ . \tag{52}$$

In this case however we would need the hot-big-bang period to be start after inflation directly with BBN-like temperatures. If the universe started at higher temperatures, say the GUT scale, we would need about 60 e-foldings of inflation. So, you see that the required number of e-foldings depends on the starting temperature of the universe, but we are in the realm of several tens.

The horizon problem is solved by asking that the region we see in the CMB was well inside the horizon. Since the contribution to the particle horizon from the radiation and the matter dominated eras is too small to account for the isotropy of the CMB, we can can assume that the integral that defines the particle horizon is dominated by the period of inflation. If t_L is the time of the last scattering surface, we have

$$d_p = a(t_L)\int_{t_in}^{t_{end}} \frac{dt}{a(t)} \simeq \frac{a(t_L)}{a_{end} H_I} e^N \ , \tag{53}$$

where we have used that $a(t) = a(t_{end})e^{H_I(t_{end}-t)}$. The particle horizon has to be bigger than the region that we can see now of the CMB. This is given by the angular diameter distance of the CMB last scattering surface. It is simply the physical distance between two points that now are one Hubble radios far apart, at the time t_L:

$$d_L = \frac{a(t_L)}{H_0 a_0} \tag{54}$$

To solve the horizon problem we need

$$d_p \gtrsim d_L \quad \Rightarrow \quad N \gtrsim \log\left(\frac{a_I H_I}{a_0 H_0}\right) \tag{55}$$

This is the same number as we need to solve the flatness problem, so we find the same number of e-foldings is needed to solve the horizon problem as are necessary to solve the flatness problems.

figure

2.4 Reheating

But we still miss a pice of the story. How inflation ends. So far, we have simply seen that as $\epsilon \sim 1$ the accelerated phase stops. At this point, typically the inflaton begins to oscillate around the bottom of the potential. In this regime it drives the universe as if it were dominated by non-relativistic matter. The equation for the inflation indeed reads

$$\frac{\partial \rho_\phi}{\partial t} + (3H + \Gamma)\,\rho_\phi = 0 \tag{56}$$

Homework: derive this expression. For $\Gamma = 0$, this is the dilution equation for non-relativistic matter. Γ represents the inflation decay rate. Indeed, in this period of time the inflation is supposed to decay into other particles. These thermalize and, once the inflation has decayed enough, start dominating the universe. This is the start of the standard big-bang universe.

2.5 Simplest Models of Inflation

2.5.1 Large Field Inflation

The simplest versions of inflation are based on scalar fields slowly rolling down their potential. These typically fall into two categories: large fields and small fields. Large field models are those characterized by a potential of the form

$$V(\phi) = \frac{\phi^\alpha}{M^{\alpha-4}}. \tag{57}$$

figure

For any M and α, if we put the scalar field high enough, we can have an inflationary solution. Let us see how this happens by imposing the slow roll conditions.

$$\epsilon \sim M_{\rm Pl}^2 \left(\frac{V_{,\phi}}{V}\right)^2 \sim \alpha^2 \frac{M_{\rm Pl}^2}{\phi^2} \tag{58}$$

For $\alpha \sim 1$, we have

$$\epsilon \ll 1 \qquad \Rightarrow \qquad \phi \gg M_{\rm Pl} \ . \tag{59}$$

The field dev has to be super planckian. Further, notice that the field travels an amount of order

$$\Delta\phi = \int_{\phi_{in}}^{\phi_{end}} d\phi = \int_{t_{in}}^{t_{end}} \dot\phi dt \simeq \frac{\dot\phi}{H}\int_{Ht_{in}}^{Ht_{end}} d(Ht) = \frac{\dot\phi}{H} N_e \sim \epsilon^{1/2} N_e \ M_{\rm Pl} \tag{60}$$

For $\epsilon \sim 1/N_e$ and not too small, the field excursion is of order $M_{\rm Pl}$. This is a pretty large field excursion (this explains the name large field models). But notice that in principle there is absolutely nothing bad about this. The energy density is of the field is of order $\phi^\alpha/M^{\alpha-4} \sim \left(\frac{M_{\rm Pl}}{M}\right)^\alpha M^4$ and needs to be smaller than $M_{\rm Pl}^4$ for us to be able to trust general relativity and the semiclassical description of space-time. This is realized once $M \gg M_{\rm Pl}$ (for $\alpha = 4$ we have $V = \lambda\phi^4$ and we simply require $\lambda \ll 1$). So far so good from the field theory point of view. Now, ideally some of us would like to embed inflationary theories in UV complete theories of gravity such as string theory. In this case the UV complete model need to be able to control all $M_{\rm Pl}$ suppressed operators. This is possible, though sometimes challenging, depending on the scenario considered. This is a lively line of research.

2.5.2 Small Field Inflation

From (60) we see that if we wish to have a $\Delta\phi \ll M_{\rm Pl}$, we need to have ϵ very very small. This is possible to achieve in models of the form

$$V(\phi) = V_0\left(1 - \left(\frac{\phi}{M}\right)^2\right) \tag{61}$$

figure

In this case, we have

$$\epsilon \simeq \frac{M_{\rm Pl}^2 \phi^2}{M^4} \tag{62}$$

that becomes smaller and smaller as we send $\phi \to 0$. Of course, we need to guarantee a long enough duration of inflation, which means that $\phi \sim \Delta\phi \sim \epsilon^{1/2} M_{\rm Pl} N_e$. Both conditions are satisfied by taking $M \gtrsim M_{\rm Pl} N_e$.

2.5.3 Generalizations

Over the thirty years since the discovery of the first inflationary models, there have been a very large number of generalizations. From fields with a non-trivial kinetic terms, such as DBI inflation and Ghost Inflation, to theories with multiple fields or with dissipative effects. We will come back to these models later, when we will offer a unified description.

2.6 Summary of lecture 1

- Standard Big Bang Cosmology has an horizon and a flatness problem. Plus, who created the density fluctuations in the CMB?
- A period of early acceleration solve the problems
- Inflation, here for the moment presented in the simplest form of a scalar field rolling downs a flat potential, solves them.

3 Lecture 2: Generation of density perturbations

This is the most **exciting, fascinating and predicting part**. It is the most predicting part, because we will see that this is what makes inflation predictive. While the former cosmological shortcomings that we saw so far were what motivated scientists such as Guth to look for inflation, cosmological perturbations became part of the story well after inflation was formulated. The fact that inflation could source primordial perturbations was indeed realized only shortly after the formulation of inflation. At that time, CMB perturbations were not yet observed, but the fact that we observed galaxies today, and the fact that matter grows as $\delta \propto a$ in a matter dominated universe predicted that some perturbations had to exist on the CMB. The way inflation produces these perturbations is both exciting and beautiful. It is simply beautiful because it shows that quantum effects, that are usually relegated to the hardly experiencable world of the small distances, can be exponentiated in the peculiar inflationary space-time to become actually the source of all the cosmological perturbations, and ultimately of the galaxies and of all the structures that are present in our universe. With inflation, quantum effects are at the basis of the formation of the largest structures in the universe. This part is also when inflation becomes more intellectually exciting. We will see that there is a very interesting quantum field theory that happens when we put some field theory in a accelerating space-time. And this is not just for fun, it makes predictions that we are actually testing right now in the universe!

The calculation of the primordial density perturbations can be quite complicated. Historically, it has taken some time to outstrip the description of all the irrelevant parts and make the story simple. This is typical of all parts of science and of all discoveries. Therefore, I will give you what I consider the simplest and most elegant derivation. Even with this, the calculation is quite complicated. Therefore we will first see how we can estimate the most important characteristics of the perturbations without doing any calculations. Only later, we will do the rigorous, and now simple, calculation [4].

3.1 Simple Derivation: real space

In this simple derivation we will drop all numerical factors. We will concentrate on the physics **Homework for you: derive all numerical factors.**.

Let us expand the field around the background solution. Since the world is quantum mechanical, if the lowest energy state is not an eigenstate of the field operator $\hat{\phi}|0\rangle \neq \phi|0\rangle$, then

$$\phi = \phi_0(t) + \delta\phi(\vec{x}, t) \tag{63}$$

Notice that if we change coordinates

$$x^\mu \to x'^\mu = x^\mu + \xi^\mu \tag{64}$$

then

$$\delta\phi(\vec{x}, t) \to \widetilde{\delta\phi}(\vec{x}'^\mu) - \dot{\phi}_0(t)\xi^0 \tag{65}$$

$\delta\phi$ does not transform as a scalar, it shifts under time diffeomorphisms (diffs.). The actual definition of $\delta\phi$ *depends* on the coordinates chosen. This has been the problem that has terrified the community for a long time, and made the treatment of perturbations in inflation very complicated [5]. Instead, we will simply ignore this subtlety, as it is highly irrelevant. Indeed, we are talking about a scalar field, very much like the Higgs field. When we study the Higgs field we do not bother about specifying the coordinates.

figure

---

[4] General lesson I think I have learned from my teachers: always know the answer you have to get before starting a difficult calculation.

[5] Of course, at the beginning things were new, and it was very justified not to get things immediately in the simplest way.

So why we should do it now? We do not even bother of writing down the metric perturbations, so why we should do it now? Let us therefore proceed, and expand the action for the scalar field at quadratic order in an unperturbed FRW metric:

$$S = \int d^4x \mathcal{L}_0 + \left.\frac{\delta \mathcal{L}}{\delta\phi}\right|_0 \delta\phi + \left.\frac{\delta^2\mathcal{L}}{\delta\phi^2}\right|_0 \delta\phi^2 = S_0 + \int d^4x e^{3Ht}\left[-g^{\mu\nu}\partial_\mu\delta\phi\partial_\nu\phi\right] \ , \tag{66}$$

Notice that the term linear in $\delta\phi$ is called the tadpole term, and if we expand around the solution of the background equations $\delta S/\delta\phi|_0 = 0$ it vanishes. We have used that $\sqrt{-g} = a^3 = e^{3Ht}$. The action contains simply a kinetic term for the inflation. The potential terms are very small, because the potential is very flat, so that we can neglect it.

figure on length scales

- Let us concentrate on very small wavelengths (high-frequencies). $\omega \gg H$. $\Delta\vec{x} \ll H^{-1}$. In that regime, we can clearly neglect the expansion of the universe, as we do when we do LHC physics (this is nothing but the equivalence principle at work: at distances much shorter than the curvature of the universe we live in flat space). We are like in Minkowski space, and therefore

$$\langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle_{vac.} \sim \text{ something} \sim [\text{length}]^{-2} \ , \tag{67}$$

  just by dimensional analysis. Since there is no length scale or mass scale in the Lagrangian (remember that H is negligible), then the only length in the system is $\Delta\vec{x}$. We have

$$\langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle_{vac.} \sim \frac{1}{|\Delta\vec{x}|^2} \tag{68}$$

  Notice that the two point function decreases as we increase the distance between the two points: this is why usually quantum mechanics is segregated to small distances.

- But the universe is slowly expanding, so the physical distance between to comoving points grows (slowly) with a:

$$|\Delta\vec{x}| \quad \to \quad |\Delta\vec{x}(t)| \propto a(t) \qquad \Rightarrow \qquad \langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle_{vac.} \sim \frac{1}{|\Delta\vec{x}|^2(t)} \tag{69}$$

- Since H is constant (it would be enough for the universe to be accelerating), at some point we will have

$$|\Delta\vec{x}|(t) \sim H^{-1} \tag{70}$$

and keeps growing. At this point, the Hubble expansion is clearly not a slow time scale for the system, it is actually very important. In particular, if two points are one Hubble far apart, then we have

$$v_{\rm relative} \gtrsim v_{\rm light} \tag{71}$$

Notice that this is not in contradiction with the principle of relativity: the two points simply stop communicating. But then gradients are irrelevant, and the value of ϕ and $\vec{x}$ should be unaffected by the value of ϕ at $\vec{x}'$. Since any value of $\delta\phi$ is as good as the others (if you look at the action, there is no potential term that gives difference in energy to different values of $\delta\phi$). Because of this, the two point function stops decreasing and becomes constant

$$\langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle_{vac.} \sim \frac{1}{|\Delta\vec{x}|^2 = H^{-2}} \sim H^2 \quad \text{as} \quad \Delta\vec{x} \to \infty \tag{72}$$

So, we see that the two point function stops decreasing and as $\Delta\vec{x}$ becomes larger than H^{-1}, and it remains basically constant of order H^2. This means that there is no scale in the two point functions, once the distance is larger than H^{-1}. An example of a scale dependent two point function that we could have found could have been: $\langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle \sim H^2\vec{x}$. This does not happen, and we have a scale invariant spectrum.

3.2 Simple Derivation: Fourier space

Let us see at the same derivation, working this time in Fourier space. The action reads

$$S = \int d^4x e^{3Ht}\left[-g^{\mu\nu}\partial_\mu\delta\phi\partial_\nu\delta\phi\right] = \int dt d^3k\ a^3\left(\dot{\delta\phi}_{\vec{k}}\dot{\delta\phi}_{-\vec{k}} - \frac{k^2}{a^2}\delta\phi_{\vec{k}}\delta\phi_{-\vec{k}}\right)\ , \tag{73}$$

- Each Fourier mode evolves independently. This is a quadratic Lagrangian!
- Each Fourier mode represents an quantum mechanical harmonic oscillator (apart for the overall factor of a^3), with a time-dependent frequency

$$\omega(t) \sim \frac{k}{a(t)} \tag{74}$$

The canonically normalized harmonic oscillator is $\delta\phi_{can} \sim a^{-3/2}\delta\phi$

- Let us focus on one Fourier mode. At sufficiently early times, we have

$$\omega(t) \simeq \frac{k}{a} \gg H\ . \tag{75}$$

In this regime, as before, we can neglect the expansion of the universe and therefore any time dependence. Then we are as if we were in Minkowski space, and therefore we must have, for a canonically normalized scalar field (i.e. harmonic oscillator)

$$\langle\delta\phi^2_{can,k}\rangle \sim \frac{1}{\omega(t)} \qquad \Rightarrow \qquad \langle\delta\phi^2_k\rangle \sim \frac{1}{a^3}\cdot\frac{1}{\omega(t)} \tag{76}$$

- While $\omega \gg H$, ω slowly decreases with time $\dot{\omega}/\omega \sim H \ll \omega$, so the two point function follows adiabatially the value on the vacuum. This happens until $\omega \sim H$ and ultimately $\omega \ll H$. At this transition, called freeze-out, the adiabatic approximation breaks down. What happens is that no more evolution is possible, because the two points are further away than an Hubble scale, and so they are beyond the event horizon. Equivalently the harmonic oscillator now has an overdamping friction term $\ddot{\delta\phi}_{\vec{k}} + 3H\dot{\delta\phi}_{\vec{k}} = 0$ that now is relevant. Since this happen when

$$\omega \sim \frac{k}{a(t_{freeze_out})} \sim H \qquad \Rightarrow \qquad a_{freeze-out} \sim \frac{k}{H} \tag{77}$$

By substituting in the two point function, we obtain

$$\langle\delta\phi^2_k\rangle \sim \frac{1}{a^3_{freeze-out}}\cdot\frac{1}{\omega(t_{freeze-out})} \sim \frac{H^2}{k^3} \tag{78}$$

This is how a scale invariant two-point function spectrum looks like in Fourier space. It is so because in Fourier space the phase space goes as $d^3k \sim k^3$, so, if the power spectrum goes as $1/k^3$, we have that each logarithmic interval in k-space contributes equally to the two-point function in real space. In formulae

$$\langle\delta\phi(\vec{x})^2\big|^{E_2}_{E_1}\rangle \sim \int_{E_1}^{E_2} d^3k\langle\delta\phi^2_k\rangle \sim H^2\log\left(\frac{E_1}{E_2}\right) \tag{79}$$

This is simply beautiful, at least in my opinion. In Minkowski space quantum mechanics is segregated to small distances because

$$\langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle_{vac.} \sim \frac{1}{|\Delta\vec{x}|^2} \tag{80}$$

In an inflationary space-time (it looks like a de Sitter space, but, contrary to de Sitter space, it ends), we have that on very large distances

$$\langle\delta\phi(\vec{x},t)\delta\phi(\vec{x}',t)\rangle_{vac.} \sim H^2 \gg \frac{1}{|\Delta\vec{x}|^2} \qquad \text{for} \qquad \Delta\vec{x} \gg H^{-1} \tag{81}$$

At a given large distance, quantum effects are on large scales much larger than what they would have naively been in Minkowski space, and this by a huge amount once we consider that in inflation scales are stretched out of the horizon by a factor of order e^{60}.

Since we are all physicists here, we can say that this is a remarkable story for the universe.

Further, it tells us that trough this mechanisms, by exploring cosmological perturbations we are studying quantum mechanics, and so fundamental physics.

But still, we need to make contact with observations.

3.3 Contact with observation: Part 1

In the former subsection we have seen that the scalar field develops a large scale-invariant two-point function at scales longer than Hubble during inflation. How these become the density perturbations that we see in the CMB and that then grow to become the galaxies?

Let us look at what happens during inflation. Let us take a box full of inflation up in the potential, and let inflation happen. In each point in space, the inflation will roll down the potential and inflation will end when the inflation at each location will reach a point $\phi(\vec{x}, t_{end}) = \phi_{end}$. We can therefore draw a surface of constant field $\phi = \phi_{end}$. Reheating will start, and in every point in space reheating will happen in the same way: the only thing that changes between the various points is the value of the gradient of the fields, but for the modes we are interested in, these are much much longer than the horizon, and so gradients are negligible; also the velocity of the field matters, but since we are on an attractor solution, we have the same velocity everywhere. At this point there is no difference between the various points, and so reheating will happen in the same way in every location. In the approximation in which re-heating happens instantaneously, the surfaces $\phi = \phi_{end}$ are equal temperature surfaces (if reheating is not instantaneous, then the equal temperature surface will be displaced later, but nothing will change really), and so equal energy density surfaces. Now, is this surface an equal time surface? In the limit in which there no quantum fluctuations for the scalar field, it would be so, but quantum fluctuations make it perturbed. How a quantum fluctuation will affect the duration of inflation at each point? Well, a jump $\delta\phi$ will move the inflaton towards or far away from the end of inflation. This means that the duration of inflation in a given location will be perturbed, and so the overall expansion of the universe when $\phi = \phi_{end}$ will be different. We therefore have a $\phi = \phi_{end}$ surface which locally looks like an unperturbed universe, the only difference is that the have a difference local scale factor [6]. These are the curvature perturbations that we call ζ. In formulas

$$\delta\phi \quad \Rightarrow \quad \delta t_{inflation} \sim \frac{\delta\phi}{\dot{\phi}} \quad \Rightarrow \delta\text{expansion} \sim \zeta \sim \frac{\delta a}{a} \sim H\delta t_{inflation} \sim \frac{H}{\dot{\phi}}\delta\phi \tag{82}$$

So, the power spectrum of the curvature perturbation is given by

$$\langle \zeta_{\vec{k}}\zeta_{\vec{k}'}\rangle = \frac{H^2}{\dot{\phi}^2}\langle \phi_{\vec{k}}\phi_{\vec{k}'}\rangle = (2\pi)^3\delta^3(\vec{k}+\vec{k}')\frac{H^4}{\dot{\phi}^2}\cdot\frac{1}{k^3} \equiv (2\pi)^3\delta^3(\vec{k}+\vec{k}')P_\zeta \tag{83}$$

$$P_\zeta = \frac{H^4}{\dot{\phi}^2}\cdot\frac{1}{k^3} \simeq \frac{H^2}{M_{\rm Pl}^2\epsilon} \tag{84}$$

where in the second passage we have used the slow roll expressions.

It is the time-delay, stupid! [7]. It is important to realize that the leading mechanism through which inflation generates perturbations is by the time delay induced by the inflation fluctuations, not by the fluctuations in energy during inflation. Let us be sure about this. In

[6] Notice that since this surface has the same energy but different overall expansion: by GR, there are must a curvature for space.

[7] A famous quote from Bill Clinton in his campaign to become president in 1992.

slow roll inflation the potential needs to be very flat, we can therefore work by expanding in the smallness of the slow roll parameters. How large are the metric perturbations? Well, the difference in energy associated to a jump of the inflation is about

$$\delta\rho \sim V'\delta\phi \sim \sqrt{\epsilon}H^3 M_{\rm Pl} \quad \Rightarrow \quad \delta g^{\mu\nu} \sim \frac{\delta\rho}{\rho} \sim \sqrt{\epsilon}\frac{H}{M_{\rm Pl}} \tag{85}$$

This means that the curvature perturbation due to this effects has actually an ϵ upstairs, so, in the limit that ϵ is very small, this is a subleading contribution. Notice indeed that the time-delay effect has an ϵ *downstairs*: the flatter is the potential, the longer it takes to make-up for the loosed or gained ϕ-distance, and so the more δexpansion you get. This is ultimately the justification of why we could do the correct calculation without having to worry *at all* about metric perturbations.

3.3.1 ζ conservation for modes longer than the horizon

Why we cared to compute the power spectrum of $\zeta \sim \delta a/a$? Why do we care of ζ and not of something else? The reason is that this is the quantity that it is conserved during all the history of the universe from when a given mode becomes longer than H^{-1}, to when it becomes shorter the H^{-1} during the standard cosmology. This is very very important. We know virtually nothing about the history of the universe from when inflation ends to say BBN. In order to trust predictions of inflation, we need something to be constant during this epoch, so that we can connect to when we know something about the universe. Proving this constancy in a rigorous way requires some effort, and it is a current topic of research to prove that this conservation holds at quantum level. For the moment, it is easy to give an heuristic argument. The ζ fluctuation is defined as the component of the metric that represents the perturbation to the scale factor $a_{\rm eff} = a(1+\zeta)$. Let us consider the regime in which all modes are longer than the Hubble scale. The universe looks locally homogenous, with everywhere the same energy density, exactly the same universe, with the only difference that in each place the scale factor is valued $a(1+\zeta)$ instead of a. But remember that the metric, apart for tensor modes, is a constrained variable fully determined by the matter fluctuations. Since matter is locally unperturbed, how can it change in a time dependent way the evolution of the scale factor? Impossible. The scale factor will evolve as in an unperturbed universe, and therefore ζ will be constant in time. This will happen until gradients will become shorter than Hubble again, so that local dynamics will be able to feel that the universe is not really unperturbed, and so ζ will start evolving.

We should think that it is indeed ζ that sources directly the temperature perturbations we see in the CMB. We should think that $P_\zeta \sim 10^{-10}$.

3.4 Scale invariance and tilt

As we saw, the power spectrum of ζ is given by

$$P_\zeta = \frac{H^4}{\dot\phi^2}\cdot\frac{1}{k^3} \simeq \frac{H^2}{M_{\rm Pl}^2\epsilon} \tag{86}$$

This is a scale invariant power spectrum. The reason why it is scale invariant is because every Foureir mode sees exactly the same history: it starts shorter than H^{-1}, becomes longer than H^{-1}, and becomes constant. In the limit in which H and $\dot\phi$ are constant (we are in an attractor solution, so $\ddot\phi$ is just a function of ϕ), then every Fourier mode sees the same history and so the power in each mode is the same. In reality, this is only an approximation. Notice that the value of H and of $\dot\phi$ depend slightly on the position of the scalar field. In order to account of this, the best approximation is to evaluate for each mode H and $\dot\phi$ at the time when the mode crossed Hubble and became constant. This happens at the k-dependent $t_{f.o.}(k)$ freezing time defined by

$$\omega(t_{f.o.}) \simeq \frac{k}{a(t_{f.o.})} = H(t_{f.o.}) \tag{87}$$

$$\Rightarrow \quad t_{f.o.}(k) \simeq \frac{1}{H(t_{f.o.}(k))} \log\left(\frac{H(t_{f.o.}(k))}{k}\right)$$

This leads to a deviation from scale invariance of the power spectrum. Our improved version now reads

$$P_\zeta = \frac{H(t_{f.o.}(k))^4}{\dot\phi(t_{f.o.}(k))^2} \cdot \frac{1}{k^3} \tag{88}$$

A measure of the scale dependence of the power spectrum is given by the tilt, defined such that the k-dependence of the power spectrum is approximated by the form

$$P_\zeta \sim \frac{1}{k^3}\left(\frac{k}{k_0}\right)^{n_s-1} \tag{89}$$

where k_0 is some pivot scale of reference. We therefore have

$$n_s - 1 = \frac{\partial \log(k^3 P_k)}{\partial \log k} = \left.\frac{d\log\left(\frac{H^4}{\dot\phi^2}\right)}{d\log k}\right|_{k/a\sim H} = \left.\frac{d\log\left(\frac{H^4}{\dot\phi^2}\right)}{Hdt}\frac{Hdt}{d\log k}\right|_{k/a\sim H} \tag{90}$$

where we have used the fact that the solution is a function of k though the ratio k/a as this is the physical wavenumber. At this point we can use that

$$d\log k = d\log(aH) \simeq Hdt \tag{91}$$

to obtain

$$n_s - 1 \simeq -2\frac{\dot H}{H^2} + 2\frac{\ddot\phi}{H\dot\phi} = 4\epsilon - 2\eta \tag{92}$$

The tilt of the power spectrum is of order of the slow roll parameters, as expected. How come we were able to compute the tilt of the power spectrum that is slow roll suppressed, though we neglected metric fluctuations, that are also slow roll suppressed? The reason is that the correction to the power spectrum due to the tilt become larger and larger as k becomes more and more different from k_0. Metric fluctuations are expected to give a finite correction of order slow roll to the power spectrum, but not one that is enhanced by the difference of wave numbers considered. This is the same approximation we do in Quantum Field Theory when we use the running of the couplings (which is log enhanced), without bothering of the finite corrections. The pivot scale k_0 is in this context analogous to the renormalization scale.

3.5 Energy scale of Inflation

We can at this point begin to learn something about inflation. Remember that the power spectrum and its tilt are of order

$$P_\zeta \sim \frac{H^2}{M_{\rm Pl}^2}\epsilon\ , \qquad n_s - 1 = 4\epsilon - 2\eta\ , \tag{93}$$

with, for slow roll inflation

$$H^2 \simeq \frac{V(\phi)}{M_{\rm Pl}^2} \tag{94}$$

From observations of the CMB, we know that

$$P_\zeta \sim 10^{-10}\ , \qquad n_s - 1 \sim 10^{-2}\ . \tag{95}$$

Knowledge of these two numbers is not enough to reconstruct the energy scale of inflation. However, if we assume for the moment that $\eta \sim \epsilon$, a reasonable assumption that however it is sometimes violated (we could have $\epsilon \ll \eta$), then we get

$$\frac{H}{M_{\rm Pl}} \sim 10^{-6}\ , \qquad H \sim 10^{-13}{\rm GeV}\ , \qquad V \sim \left(10^{15}{\rm GeV}\right)^4 \tag{96}$$

These are remarkably large energy scales. This is the energy scale of GUT, not very distance from the Plank scale. Inflation is really beautiful. Not only it has made quantum fluctuations the origin of all the structures of the universe, but it is likely that these are generated by physics at very high energy scales. These are energy scales that unfortunately we will probably never be able to explore at particle accelerators. But these are energy scales that we really would like to be able to explore. We expect very interesting new physics to lie there: new particles, possibly GUT theories, and even maybe string theory. We now can explore them with cosmological observations!

3.6 Statistics of the fluctuations: Approximate Gaussianity

Let us go back to our action of the fluctuations of the scalar field. Let us write again the action in Foureir space, but this time it turns out to be simpler to work in a finite comoving box of volume V. We have

$$\phi(x) = \frac{1}{V}\sum_{\vec k}\phi_k e^{i\vec k\cdot\vec x} \tag{97}$$

Notice that the mass dimensions of $\phi_{\vec k}$ is -2. To get the action, we need the following manipulation

$$\int d^3x\ \phi(x)^2 = \frac{1}{V^2}\sum_{k,k'}\int d^3x\ e^{i(\vec k+\vec k')\cdot\vec x}\phi_k\ \phi_{k'} \simeq \frac{1}{V^2}\sum_{k,k'}\delta^3(\vec k+\vec k')\phi_k\ \phi_{k'} \tag{98}$$

$$\simeq \frac{1}{V}\sum_{k,k'}\delta_{\vec k,-\vec k'}\phi_k\ \phi_{k'} = \frac{1}{V}\sum_{k,k'}\phi_k\ \phi_{-k'}$$

The action therefore reads

$$S_2 = \frac{1}{V}\sum_k a^3 \left(\dot\phi_{\vec k}\dot\phi_{-\vec k} + \frac{k^2}{a^2}\phi_{\vec k}\phi_{-\vec k}\right) \tag{99}$$

Let us find the Hamiltonian. We need the momentum conjugate to $\phi_{\vec k}$.

$$\Pi_{\vec k} = \frac{\delta S_2}{\delta \dot\phi_{\vec k}} = \frac{a^3}{V}\dot\phi_{-\vec k} \tag{100}$$

The Hamiltonian reads

$$\begin{aligned} H &= \sum_{\vec k}\Pi_{\vec k}\dot\phi_{\vec k} - \frac{1}{V}\sum_k a^3\left(\dot\phi_{\vec k}\dot\phi_{-\vec k} + \frac{k^2}{a^2}\phi_{\vec k}\phi_{-\vec k}\right) \\ &= \sum_{\vec k}\frac{V}{a^3}\Pi_{\vec k}\Pi_{-\vec k} + \frac{1}{V}\frac{k^2}{a^2}\phi_{\vec k}\phi_{-\vec k} \end{aligned} \tag{101}$$

If we concentrate on early times where the time dependence induced by Hubble expansion is negligible, we have, for each $\vec k$ mode, the same Hamiltonian as an Harmonic oscillator, which reads (again, remember that I am dropping all numerical factors)

$$H = \frac{P^2}{m} + m\omega^2 x^2 \tag{102}$$

We can therefore identify

$$m = \frac{a^3}{V}\ , \qquad \phi_{\vec k} = x\ , \qquad \omega = \frac{k}{a}. \tag{103}$$

The vacuum wave function for an harmonic oscillator is a Gaussian

$$|0\rangle = \int dx\ e^{-m\omega x^2}|x\rangle \tag{104}$$

which tells us that the vacuum wave function for each Fourier mode $\vec k$ reads

$$|0\rangle_{k/a\gg H} = \sum_{\phi_{\vec k}}\ e^{-\frac{a^3}{V}\frac{k}{a}\phi^2_{\vec k}}|\phi_k\rangle \tag{105}$$

Since all Fourier mode evolve independently, for the set of Fourier modes that have $k/a \gg H$, we can write

$$|0\rangle_{k_i/a\gg H} = \prod_{\vec k_i\gg Ha}\sum_{\phi_{\vec k}}\ e^{-\frac{a^3}{V}\frac{k_i}{a}\phi^2_{\vec k_i}}|\phi_{\vec k_i}\rangle \tag{106}$$

For each Fourier mode, at early time we have a Gaussian wave function with width $V^{1/2}/(k^{1/2}a)$.

Let us follow the evolution of the wave function with time. As discussed, at early times when $k/a \gg H$, the wave functions follows adiabatically the wave function of the would be harmonic oscillator with those time dependent mass and frequency given by (103). However, as the frequency drops below the Hubble rate, the natural time scale of the harmonic oscillator becomes too slow to keep up with Hubble expansion. The state gets frozen on the parameters that it had when $\omega(t) \sim H$. Bu substituting $k/a \to H$, $a \to k/H$,the wave function at late times becomes

$$|0\rangle_{k_i/a \ll H} = \prod_{\vec{k}_i \ll Ha} \sum_{\phi_{\vec{k}}} e^{-\frac{1}{V}\frac{k_i^3}{H^2}\phi_{\vec{k}_i}^2} |\phi_{\vec{k}_i}\rangle \tag{107}$$

This is a Gaussian in field space. Its width is given by

$$\langle \phi_{\vec{k}} \phi_{-\vec{k}} \rangle = \delta_{\vec{k},-\vec{k}} V \frac{H^2}{k^3} \simeq (2\pi)^3 \delta^3(\vec{k}+\vec{k}') \frac{H^2}{k^3} \quad \text{as} \quad V \to \infty \tag{108}$$

We recover the same result of before for the power spectrum. We additionally see that the distribution of values of $\phi_{\vec{k}}$ are Gaussianly distributed. Notice that we are using a quite unusual base of the Hilbert space of a quantum field theory (more used when one talks about the path integral), which is the $|\phi\rangle$ eigenstates base instead of the usual Fock base with occupation numbers. This base is sometimes more useful, as we see here.

So, we learn that the distribution is Gaussian. This result could have been expected. At the end, (so far!), we started with a quadratic Lagrangian, the field theory is free, and so equivalent to an harmonic oscillator, which, in its vacuum, is Gaussianly distributed. We will see in the last lecture that when we consider interacting field theories the distribution will not be Gaussian anymore! Indeed, the statement that cosmological perturbations are so far Gaussian simply means that the field theory describing inflation is a weakly coupled quantum field theory in its vacuum. We will come back to this.

3.7 Why does the universe looks classical?

So far we have seen that the cosmological fluctuations are produced by the quantum fluctuations of the inflation in its vacuum state. But then, why does the universe looks classical? The reason is the early vacuum state for each wave number becomes a very classical looking state at late times. Let us see how this happens.

The situation is very simple. We saw in the former subsection that the vacuum state at early times is the one of an harmonic oscillator with frequency $k/a \gg H$. However the frequency is red shifting, and at some point it becomes too small to keep up with Hubble expansion. At that point, while the frequency goes to zero, the state remains trapped in the vacuum state of the would-be harmonic oscillator with frequency $k/a \sim H$. The situation is very similar to what happens to the vacuum state of an harmonic oscillator when one opens up very abruptly the width of the potential well.

This is an incredibly squeezed state with respect to the ground state of the harmonic oscillator with frequency $\omega \sim e^{-60}H$. This state is no more the vacuum state of the late time harmonic oscillator. It has a huge occupation number, and it looks classical.

figure

Let us check that indeed that wave function is semiclassical. The typical condition to check if a wavefunction is well described by a semiclassical approximation is to check if the ϕ-length scale over which the amplitude of the wavefunction changes is much longer than the ϕ-length scale over which the phase changes. To obtain the wavefunction at late times, we performed the sudden approximation of making the frequency instantaneously zero. This corresponds to make an expansion in $k/(aH)$. In our calculation we obtained a real wavefunction (109). This means that the phase must have been higher order in $k/(aH) \ll 1$, in the sense that it should be much more squeezed than the width of the magnitude, much more certain the outcome: the time-dependent phase has decayed away. We therefore can write approximately

$$|0\rangle_{k_i/a\ll H,guess} \sim \prod_{\vec{k}_i\ll Ha} \int d\phi_{\vec{k}_i}\ e^{-\frac{1}{V}\frac{k^3}{H^2}\phi^2_{\vec{k}_i}\left[1+i\frac{aH}{k}\right]}|\phi_{\vec{k}_i}\rangle \tag{109}$$

We obtain:

$$\Delta\phi_{\text{Amplitude Variation}} \sim \frac{H}{k^{3/2}}\frac{1}{V^{1/2}}\ , \qquad \Delta\phi_{\text{Phase Variation}} \sim \frac{H}{k^{3/2}}\frac{1}{V^{1/2}}\left(\frac{k}{aH}\right)^{1/2}\ , \tag{110}$$

So

$$\frac{\Delta\phi_{\text{Phase Variation}}}{\Delta\phi_{\text{Amplitude Variation}}} \sim \left(\frac{k}{aH}\right)^{1/2} \to 0 \tag{111}$$

So we see that the semiclassicality condition is satisfied at late times.

3.8 Tensor

Before moving on, let us discuss briefly the generation of tensor modes. In order to do that, we need to discuss about the metric. (Remarkably, this is the first time we have to do that).

3.8.1 Helicity Decomposition of metric perturbations

A generically perturbed FRW metric can be put in the following form

$$ds^2 = -(1+2\Phi)dt^2 + 2a(t)B_i dx^i + a(t)^2\left[(1-2\Psi)\delta_{ij} + E_{ij}\right] \tag{112}$$

For background space-times that have simple transformation rules under rotation (FRW for example is invariant), it is useful to decompose these perturbations according to their transformation properties under rotation under one axis. A perturbation of wavenumber $\vec{k}$ has elicit λ if under a rotation along the $\hat{k}$ of angle θ, transforms simply by multiplication by $e^{i\lambda\theta}$:

$$\delta g \to e^{i\lambda\theta}\delta g \tag{113}$$

Scalars have helicity zero, vectors have helicity one, and tensors have helicity two. It is possible to decompose the various components of $\delta g_{\mu\nu}$ in the following way:

$$\Phi,\ \Psi \tag{114}$$

have helicity zero. We can then write

$$B_i = \partial_i B_S + \tilde{B}_{V,i} \tag{115}$$

where $\partial^i \tilde{B}_{V,i} = 0$. B_S is a scalar, B_V is a vector. Finally

$$E_{ij} = E_{ij}^S + E_{ij}^V + \gamma_{ij} \tag{116}$$

where

$$\begin{aligned} E_{ij}^S &= \frac{1}{\partial^2}\left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\partial^2\right)\tilde{E}^S \\ E_{ij}^V &= \frac{1}{2\partial^2}\left(\partial_i\tilde{E}_j^V + \partial_j\tilde{E}_i^V\right) \;, \quad \text{with} \quad \partial_i\tilde{E}^i = 0 \\ \partial_i\gamma_{ij} &= 0, \quad \gamma_i{}^i = 0 \;. \end{aligned} \tag{117}$$

with $\partial^2 = \delta^{ij}\partial_i\partial_j$. $\tilde{E}^S$ is a scalar, $\tilde{E}^V$ is a vector, and γ is a tensor.

Now, it is possible to show that at linear level, in a rotation invariant background, scalar, vector and tensor modes do not couple and evolve independently (you can try to contract the vectors together it does not work: you cannot make it).

Under a change of coordinate

$$x^\mu \to \tilde{x}^\mu = x^\mu + \xi^\mu \tag{118}$$

these perturbations change according to the transformation law of the metric

$$\tilde{g}^{\mu\nu} = \frac{\partial\tilde{x}^\mu}{\partial x^\rho}\frac{\partial\tilde{x}^\nu}{\partial x^\sigma}g^{\rho\sigma} \tag{119}$$

The change of coordinates ξ^μ can also be decomposed into a scalar and a vector component

$$\xi_S^0 \;, \quad \xi_S^i = \partial^i\xi \tag{120}$$

$$\xi_V^0 = 0 \quad \partial_i\xi_V^i = 0 \tag{121}$$

At linear level, different helicity metric perturbations do not get mixed and they are transformed only by the change coordinates with the same helicity (for the same reasons as before). For this reasons, we see that tensor perturbations are invariant. They are gauge invariant. This is not so for scalar and vector perturbations. For example, scalar perturbations transform as the following

$$\Phi \to \Phi - \dot{\xi}_S^0 \tag{122}$$

$$B_S \to B_S + \frac{1}{a}\xi_S^0 - a\dot{\xi} \tag{123}$$

$$E \to E - B_S \tag{124}$$

$$\Psi \to \Psi - H\alpha \tag{125}$$

The fact that tensor modes are gauge invariant and uncoupled (at linear level!) means that we can write the metric for them as

$$g_{ij} = a^2 \left(\delta_{ij} + \gamma_{ij}\right) \ , \tag{126}$$

and set to zero all other perturbations (including $\delta\phi$). By expanding the action for the scalar field plus GR at quadratic order, one obtains an action of the form (actually only the GR part contributes, and the following action could just be guessed)

$$S = \int d^4x \ a^3 \ M_{\rm Pl}^2 \left[(\dot{\gamma}_{ij})^2 - \frac{1}{a^2}(\partial_l \gamma_{ij})^2\right] = \sum_{s=+,\times} \int dt d^3k \ a^3 \ M_{\rm Pl}^2 \left[\dot{\gamma}^s_{\vec{k}} \dot{\gamma}^s_{-\vec{k}} - \frac{k^2}{a^2} \gamma^s_{\vec{k}} \gamma^s_{-\vec{k}}\right] \tag{127}$$

where in the last passage we have decomposed the generic tensor mode in the two possible polarization state

$$\gamma_{ij}^{+,\times} = \gamma_s(t) e_{ij}^{(+,\times)} \tag{128}$$

In matrix form, for a mode in the $\hat{k} = \hat{z}$ direction

$$\gamma = \begin{pmatrix} \gamma^\times & \gamma^+ & 0 \\ \gamma^+ & -\gamma^\times & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{129}$$

$$\gamma_{ij} = \int d^3k \sum_{s=+,\times} e^s_{ij}(k) \gamma^s_{\vec{k}}(t) e^{i\vec{k}\cdot\vec{x}} \tag{130}$$

$$\epsilon^s_{ii} = k^i \epsilon_{ij} = 0 \quad \epsilon^s_{il} \epsilon^{s'}_{lj} = \delta_{ij} \tag{131}$$

We see that the action for each polarization is the same as for a normal scalar field, just with a different canonical normalization. The two polarization are also independent (of course), and therefore, without having to do any calculation, we obtain the power spectrum for gravity waves to be

$$\langle \gamma^s_{\vec{k}} \gamma^{s'}_{\vec{k}'} \rangle = (2\pi)^3 \delta^3(\vec{k} + \vec{k}') \delta_{s,s'} \frac{H^2}{M_{\rm Pl}^2} \frac{1}{k^3} \tag{132}$$

Notice that the power spectrum depends only on one unknown quantity H. This means that if we detect gravitational waves from inflation, we could measure the energy scale of inflation. … Actually, this was a 'theorem' that was believed to hold until last september. At that time new mechanisms further than the vacuum fluctuations have been identified that could dominate the ones produced by vacuum fluctuations and that could be detectable.

By now we are expert: the tilt of gravity waves power spectrum is given by

$$n_t - 1 = -2\epsilon \tag{133}$$

as only the variation of H is involved.

The measurement of this tilt would give us a measurement of ϵ. Again, until recently this was thought to be true, and unfortunately (and luckily) things have changed now, and the above formula for the tilt holds only for the simplest models of inflation.

Notice further that if we were to measure the amplitude of the gravitational waves and their tilt, then, under the hypothesis of standard slow roll inflation, we would know H and ϵ. In this same hypothesis therefore we would therefore predict the size of the ζ power spectrum. If this would hold, we would discover that inflation happened in the slow roll inflation way. This is called consistency condition for slow roll single field inflation.

Notice that, in standard slow roll inflation (this is true only for the simple inflationary scenarios), the power in gravity waves is smaller than the one in scalars by a factor of $\epsilon \ll 1$. This means that if gravity waves are detected, ϵ cannot be too small, and therefore the field excursion during inflation is over planckian: $\Delta\phi \gtrsim M_{\rm Pl}$. This is known as the lyth's bound.

Finally, notice that this signal is proportional to $\hbar$. Such a measurement would be the first direct evidence that GR is quantized. We have never seen this (frankly there are no doubt that gravity is quantized but still better to see it in experiments.)

3.9 Summary of Lecture 2

- the quantum fluctuations of the scalar field naturally produce a scale invariant spectrum of perturbations
- they become curvature perturbations at the end of inflation
- they look like classical and (quasi) Guassian
- Quantum mechanical effects are at the source of the largest structures in the universe
- The Energy scale of inflation could be as high as the GUT scale, opening the possiblity to explore the most fundamental laws of physics from the cosmological observations
- Tensor modes are also produced. If seen, first evidence of quantization of gravity.
- Everything is derived without hard calculations

Now we are ready to see how we check for this theory in the data.

4 Lecture 3: contact with observations (part 2)

Absolutely, the best way we are testing inflation is by the observation of the cosmological perturbations

You had already several classes on the evolution of perturbations in the universe and how they connect to observations. Here I will simply focus on the minimum amount of information that we need to establish what this observations are really telling us about Inflation. I will focus just on CMB, for brevity. The story is very similar also for large scale structures.

4.1 CMB basics

For a given perturbation $\delta X(k,\tau)$ at a given time τ and with Fourier mode k, we can define its transfer function for the quantity X at that time τ and for the Fourier mode k as

$$\delta X = T(k,\tau,\tau_{in})\zeta_k(\tau_{in}) \tag{134}$$

This must be so in the linear approximation. We can take τ_{in} early enough so that the mode k is smaller than aH, in this way $\zeta_k(\tau_{in})$ represents the constant value ζ took at freeze out during inflation.

For the CMB temperature, we perform a spherical harmonics decomposition

$$\frac{\delta T}{T}(\tau_0,\hat{n}) = \sum_{l,m} a_{lm} Y_{lm}(\hat{n}) \tag{135}$$

and the by statistical isotropy the power spectra reads

$$\langle a_{lm} a_{l'm'} \rangle = C_l^{TT} \delta_{ll'} \delta_{mm'} \tag{136}$$

Since the temperature anisotropy are dominated by scalar fluctuations, we have

$$a_{lm} = \int d^3k \Delta_l(k) \zeta_k Y_{lm}(\hat{k}) \ , \quad \Rightarrow \quad C_l = \int dk\ k^2\ \Delta_l(k)^2 \ \ P_\zeta(k) \ , \tag{137}$$

$\Delta_l(k)$ contains both the effect of the transfer functions and also of the projection on the sky.

- **large scales:** If we look at very large scales, we find modes that were still outside H^{-1} at the time of recombination. Nothing could have happened to them.

  figure on scales

  As in Julien's class you have been told, there has been no evolution and only projection effects.

$$\Delta_l(k) \simeq j_l(k(\tau_0 - \tau_{\rm rec})) \quad \Rightarrow \quad C_l \simeq \int dk\, k^2 P_\zeta\ j_l^2(k(\tau_0 - \tau_{\rm rec})) \tag{138}$$

$j_l^2(k(\tau_0 - \tau_{rec}))$ is sharply peaked at $k(\tau_0 - \tau_{\rm rec}) \sim l$, so we can approximately perform the integral, to obtain

$$C_l \simeq k^3 P_\zeta\big|_{k=l/(\tau_0-\tau_{\rm rec})} \times \int d\log x \; j_l^2(x) \sim k^3 P_\zeta\big|_{k=l/(\tau_0-\tau_{\rm rec})} \times \frac{1}{l(l+1)} \quad (139)$$

$$\Rightarrow \quad l(l+1)C_l \;\text{ is flat }. \quad (140)$$

- **small scales**. On short scales, mode entered inside H^{-1} and begun to feel both the gravitational attraction of denser zones, but also their pressure repulsion. This leads to oscillatory solutions.

$$\ddot{\delta}T + c_s^2\nabla^2\delta T \simeq F_{\rm gravity}(\zeta) \quad (141)$$

$$\Rightarrow \qquad \delta T_k \simeq A_{\vec k}\,\cos(k\eta) + B_{\vec k}\,\sin(k\eta) = \tilde A_{\vec k}\,\cos(k\eta + \phi_{\vec k}) \quad (142)$$

Here $A_{\vec k}$ and $B_{\vec k}$ depend on the initial conditions. In inflation, we have

$$\tilde A_{\vec k} \simeq \frac{1}{k^3}\ , \qquad \phi_{\vec k} = 0\ . \quad (143)$$

All the modes are in phase coherence. Notice, dynamics and wavenumber force all mode of a fixed wavenumber to have the same frequency. However, they need not have necessarily the same phase. Inflation, or superHubble fluctuations, forces $\zeta \simeq \frac{\delta T}{T}$ =const on large scales, which implies $\phi_{\vec k} = 0$. This is what leads to acoustic oscillations in the CMB

$$\delta T(\vec k, \eta) \sim \delta T_{in}(\vec k) \times \cos(k\eta) \quad \Rightarrow \quad \delta T(\vec k, \eta_0) \sim \delta T_{in}(\vec k) \times \cos(k\eta_{rec}) \quad (144)$$

$$\Rightarrow \quad \langle \delta T(\vec k, \eta_0)\rangle \sim \langle \delta T^2_{\vec k}\rangle \cos^2(k\eta_{rec}) \quad (145)$$

we get the acoustic oscillations.

figure on CMB oscillations

figure on pahses

This is the greatest verification of inflation so far. Acoustic oscillations told us that the horizon was much larger than H^{-1} at recombination and that there were constant superHubble perturbation before recombination. This is very very non-trivial prediction of inflation. Notice that scale invariance of the fluctuations was already observed in the sky (Harrison-Zeldovich spectrum) at the time of formulation of inflation, but nobody knew of the acoustic oscillations at that time. Boomerang found them!

This is a very important *qualitative* verification of inflation that we get from the CMB. But it is not a quantitative confirmation. Information on the quantitative part is very limited.

4.2 What did we verify of Inflation so far?

figure on CMB

Let us give a critical look at what we learnt about inflation so far form the observational point of view.

There have been three *qualitative* theoretical predictions of inflation that have been verified so far. One is the oscillations in the CMB, another is the curvature of the universe, of order $\Omega_k \sim 10^{-2}$. At the time inflation was formulated, Ω_k could have been of order one. It is a natural prediction of inflation that lasts a little more than the necessary amount to have $\Omega_k \ll 1$. The third is that the perturbations are Gaussian to a very good approximation: the signature of a weakly coupled field theory.

But what did we learn at a *quantitative* level about inflation so far? Just two numbers, not so much in my opinion unfortunately. This is so because all the beautiful structures of the peaks in the CMB (and also in Large Scale Structures) is just controlled by well known Standard Model physics at 1 eV of energy. The input from inflation are the qualitative initial conditions for each mode, and quantitatively the power spectrum and its tilt

$$P_\zeta \simeq \frac{H^2}{M_{\rm Pl}^2 \epsilon} \sim 10^{-10} \ , \quad n_s - 1 \simeq 4\epsilon - 2\eta \simeq -6 \times 10^{-2} \ , \qquad (146)$$

just two numbers fit it all.

This is a pity, because clearly cosmological data have much more information inside them. Is it there something more to look for?

4.3 CMB Polarization

One very interesting observable is the CMB polarization. The CMB has been already observed to be partially polarized. Polarization of the CMB can be represented as the set of vectors tangent to the sphere, the direction of each vector at each angular point representing the direction of the polarization coming from the point, and its length the fractional amount.

CMB polarization in induced by Thomson scattering in the presence of a quadruple perturbation. Information on cosmological perturbations is carried over by the correlation of polarization (very much the same as the correlation of temperature). It is useful to define two scalar fields that live on the sphere.

Polarization can be decomposed into the sum of the fields, E and B, that have very different angular patter.

figure polarization in the sky

figure polarization E and B

Scalar perturbations induce E polarization, and they are being measured with greater and greater accuracy. However tensor perturbations induce both E and B polarization. This means that a discovery of B modes would be a detection of tensor modes produced during inflation (there are some B modes produced by lensing, but they are only on small angular scales). So far there is no evidence of them, but even if we saw them, what we would learn about inflation?

We will learn a great qualitative point. Producing scale invariant tensor perturbations is very hard, because tensor perturbation tend to depend only on the nature of the space-time. Scale invariant tensor modes would represent most probably that inflation did happen.

At a quantitative level, however, we would just learn two numbers: the amplitude and the tilt of the power spectrum. In the simplest models of inflation, the amplitude of the power spectrum gives us direct information about H, and if the signal is detectable, it would teach us about the energy scale of inflation. Its scale invariance would teach us that H is constant with time: this is the definition of inflation.

However recently new mechanism for produce large and detectable tensor modes have been found, which disentangle the measurement of B modes from a measurement of H, at least in principle.

So, the question really remains: is it there something more to look for?

4.4 Many more models of inflation

Indeed, there are many more models of inflation than standard slow roll that we discussed.

DBI Inflation: One remarkable example is DBI inflation. This described the motion of a brane in ADS space. Since the brane has a speed limit, an inflationary solution happens when the brane is moving at the speed of light. At that point special relativistic effects slow down the brane, and you have inflation, even though the brane is moving at the speed of light. The brane fluctuations in this case play the role of the inflaton.

figure on DBI

This model, though it happens in a totally different regime than slow roll inflation, it is totally fine with the observations we looked at so far. It turns out that the power spectrum scales in a different way that in slow roll models. We have a speed of sound $c_s \ll 1$

$$\omega^2 \sim c_s^2 k^2 \,. \tag{147}$$

This affects the power spectrum in the following way

$$P_\zeta \sim \frac{H^2}{\epsilon M_{\rm Pl}^2 c_s} \sim 10^{-10} \,, \qquad n_s - 1 \simeq 4\epsilon - 2\eta + \frac{\dot{c}_s}{H c_s} \sim 10^{-2} \tag{148}$$

Given than to match the CMB we need just these two inputs from the inflationary model, it is pretty expectable that they can be fixed. And indeed this happens.

This inflationary model had the remarkable features that non-gaussianities were detectably large. The skewness of the distribution of the fluctuations was

$$\frac{\langle\zeta^3\rangle}{\langle\zeta^2\rangle} \sim \frac{1}{c_s^2} \gg 1 \tag{149}$$

For comparison, the same number is standard inflation is of order $\epsilon \ll 1$. While for standard slow roll inflation this is undetectably small, it is detectable for DBI inflation.

This opens up a a totally new possible observational signature, and the possibility to distinguish and to learn about models that would be indistinguishable at the level of the two point function.

Non-Gaussianity!!

Ghost inflation: Ghost inflation is another peculiar looking model. It consists of a scalar field with the wrong sign kinetic term (a ghost).

figure on Ghost inflation

This triggers an instability that condensate in a different vacuum, where $\dot{\phi}$ =const even in the absence of potential. This leads to inflation. The fluctuations have a dispersion relation of the form

$$\omega^2 \simeq \frac{k^4}{M^2} \tag{150}$$

which is extremely non-relativistic.

Again, this model is totally fine in fitting observations of the power spectrum, but it produces a large and detectable non-Gaussianity.

This are new models, some inspired by string theory. But they have new signatures. So, the question is: how generic are these signatures? What are the generic signatures of inflation?

In order to do that, we need an approach that is very general, and looks at inflation in its most essential way: we go to the Effective Field Theory approach.

4.5 The Effective Field Theory of Inflation

Effective Field Theories (EFTs) have played the role of the guiding principle for particle physics and even condensed matter physics. EFTs have the capacity of synthesizing the relevant physics at the energy scale of interests. Effects of higher energy, largely irrelevant, physics are encoded in the coefficients of the higher dimension operators. It is *the way* to explore the phenomenology at a given energy scale. What we are going to do next is to develop the effective field theory of inflation. In doing so, we can look at inflation as the theory of a Goldstone boson: the Goldstone boson of time translations.

Review of Goldstone bosons: Goldstone bosons are ubiquitous in particle physics (they got Nambu the well deserved 2008 nobel prize!). Let consider the simplest theory of a $U(1)$ global symmetry $\phi \to e^{i\alpha}\phi$ that is spontaneously broken because of a mexican hat like potential $\phi \to \langle\phi\rangle$. Then there is Goldstone boson π that non-linearly realizes the symmetry $\pi \to \pi + \alpha$.

Figure of Mexican Hat

$$\mathcal{L} = \partial_\mu\phi^\star\partial^\mu\phi - m^2\phi^\star\phi + \lambda\phi^{\star 2}\phi^2 \quad \to \quad \phi = \frac{m}{\lambda^{1/2}} + i\,\pi(\vec{x},t) \tag{151}$$

The action for the field π is therefore the one of a massless scalar field endowed with a shift symmetry

$$\mathcal{L}_\pi = (\partial\pi)^2 + \frac{1}{m/\lambda^{1/2}}(\partial\pi)^4 + \ldots \tag{152}$$

the higher derivative operators being suppressed by powers of the high energy scale $m/\lambda^{1/2}$.

A famous example of Goldstone bosons are the pions of the Chiral Lagrangian, that represent the Goldstone boson that non-linearly realize the $SU(2)$ chiral flavor symmetry, and they represent in the UV theory of QCD bound states of quark and antiquark.

Inflation as the theory of a Goldstone boson: How do we build the EFT of Inflation. In order to do that, we need to think of inflation in its most essential way. What we really know about inflation is that it is a period of accelerated expansion, where the universe was quasi de sitter. However, it could not be exactly de Sitter, because it has to end. This means that time-translation is spontaneously broken, and there is a physical clock measuring time and forcing inflation to end.

No matter what this clock is, we can use coordinate invariance of GR to go to the frame where these physical clock is set to zero. This can be done by choosing spatial slices where the fluctuations of the clock are zero, by performing a proper time diffs from any coordinate frame. As an example, if the inflation was a scalar field (we are not assuming that, but just to make example) and we are in a coordinate frame where $\delta\phi(\vec{x},t) \neq 0$, we can perform a time diff. $t \to \tilde{t} = t + \delta t(\vec{x},t)$, such that (at linear order, it can be generalized to arbitrary non-linear order)

$$0 = \tilde{\delta\phi}(\vec{x},t) = \delta\phi(\vec{x},t) - \dot{\phi}_0(t)\delta t(\vec{x},t) \qquad (153)$$

Now, suppose we are in this frame. We follow the rule of EFT. This says we have to write the action with the degrees of freedom that are available to us. This is just the metric fluctuations. We have to expand in fluctuations, and write down all operators compatible with the symmetries of the problem. In our case we can arbitrarily change spatial coordinates within the various spacial slices, on each spatial slice in a different way. This means that the residual gauge symmetry is time-dependent spatial diff.s:

$$x^i \to \tilde{x}^i = x^i + \xi^i(t,\vec{x}) \ . \qquad (154)$$

figure on slices

Further, still following EFT procedure, we expand in perturbations and go to the order up to which we are interested (for example, quadratic order for 2-point functions, cubic order for 3-point functions, quartic order for 4-point functions, and so on), and then expand, at each order in the fluctuations, in derivatives, higher derivative terms being suppressed by the ratio of the energy scale E of the problem versus some high energy scale Λ.

4.5.1 Construction of the action in unitary gauge

What is the most general Lagrangian in this unitary gauge? One must write down operators that are functions of the metric $g_{\mu\nu}$, and that are invariant under the (linearly realized) time dependent spatial diffeomorphisms $x^i \to x^i + \xi^i(t, \vec{x})$. Spatial diffeomorphisms are in fact unbroken. Besides the usual terms with the Riemann tensor, which are invariant under all diffs, many extra terms are now allowed, because of the reduced symmetry of the system. They describe the additional degree of freedom eaten by the graviton. For example it is easy to realize that g^{00} is a scalar under spatial diffs, so that it can appear freely in the unitary gauge Lagrangian.

$$\tilde{g}^{00} = \frac{\partial \tilde{t}}{\partial x^\mu}\frac{\partial \tilde{t}}{\partial x^\nu} g^{\mu\nu} = \delta^0_\mu \delta^0_\nu g^{\mu\nu} = g^{00} \tag{155}$$

Polynomials of g^{00} are the only terms without derivatives. Given that there is a preferred slicing of the spacetime, one is also allowed to write geometric objects describing this slicing. For instance the extrinsic curvature $K_{\mu\nu}$ of surfaces at constant time is a tensor under spatial diffs and it can be used in the action. If n^μ is the vector orthogonal to the equal time slices, we have

$$K_{\mu\nu} = h_\nu{}^\sigma \nabla_\sigma n_\nu \ , \tag{156}$$

with ∇ being the covariant derivative, and $h_{\mu\nu}$ the induced metric on the spatial slices

$$h_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu \ . \tag{157}$$

Notice that generic functions of time can multiply any term in the action. In appendix A and B we reproduce the proof that the most generic Lagrangian can be written as

$$\begin{aligned} S \ = \ & \int d^4x \, \sqrt{-g} \Big[\frac{1}{2} M_{\rm Pl}^2 R - c(t) g^{00} - \Lambda(t) + \frac{1}{2!} M_2(t)^4 (\delta g^{00})^2 + \frac{1}{3!} M_3(t)^4 (\delta g^{00})^3 + \\ & - \frac{\bar{M}_1(t)^3}{2} (\delta g^{00}) \delta K^\mu{}_\mu - \frac{\bar{M}_2(t)^2}{2} \delta K^\mu{}_\mu{}^2 - \frac{\bar{M}_3(t)^2}{2} \delta K^\mu{}_\nu \delta K^\nu{}_\mu + ... \Big] \ , \end{aligned} \tag{158}$$

where the dots stand for terms which are of higher order in the fluctuations or with more derivatives. $\delta g^{00} = g^{00} + 1$. We denote by $\delta K_{\mu\nu}$ the variation of the extrinsic curvature of constant time surfaces with respect to the unperturbed FRW: $\delta K_{\mu\nu} = K_{\mu\nu} - a^2 H h_{\mu\nu}$ with $h_{\mu\nu}$ is the induced spatial metric. Notice that only the first three terms in the action above contain linear perturbations around the chosen FRW solution, all the others are explicitly quadratic or higher. Therefore the coefficients $c(t)$ and $\Lambda(t)$ will be fixed by the requirement of having a given FRW evolution $H(t)$, *i.e.* requiring that tadpole terms cancel around this solution. Before fixing these coefficients, it is important to realize that this simplification is not trivial. One would expect that there are an infinite number of operators which give a contribution at first order around the background solution. However one can write the action as a polynomial of linear terms like $\delta K_{\mu\nu}$ and $g^{00}+1$, so that it is evident whether an operator starts at linear, quadratic or higher order. All the linear terms besides the ones in eq. (158) will contain derivatives and they can be integrated by parts to give a combination of the three linear terms we considered plus covariant terms of higher order. This construction is explicitly carried out in appendix B. We conclude that *the unperturbed history fixes $c(t)$ and $\Lambda(t)$, while the difference among different models will be encoded into higher order terms.*

We can now fix the linear terms imposing that a given FRW evolution is a solution. As we discussed, the terms proportional to c and Λ are the only ones that give a stress energy tensor

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\rm matter}}{\delta g^{\mu\nu}} \tag{159}$$

which does not vanish at zeroth order in the perturbations and therefore contributes to the right hand side of the Einstein equations. During inflation we are mostly interested in a flat FRW Universe (see Appendix B for the general case)

$$ds^2 = -dt^2 + a^2(t)d\vec{x}^2 \tag{160}$$

so that Friedmann equations are given by

$$H^2 = \frac{1}{3M_{\rm Pl}^2}\left[c(t) + \Lambda(t)\right] \tag{161}$$

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = -\frac{1}{3M_{\rm Pl}^2}\left[2c(t) - \Lambda(t)\right] . \tag{162}$$

Solving for c and Λ we can rewrite the action (158) as

$$S = \int d^4x\, \sqrt{-g}\Big[\frac{1}{2}M_{\rm Pl}^2 R + M_{\rm Pl}^2 \dot{H} g^{00} - M_{\rm Pl}^2(3H^2 + \dot{H}) + \frac{1}{2!}M_2(t)^4(\delta g^{00})^2 + \frac{1}{3!}M_3(t)^4(\delta g^{00})^3 + \\ -\frac{\bar{M}_1(t)^3}{2}(\delta g^{00})\delta K^\mu{}_\mu - \frac{\bar{M}_2(t)^2}{2}\delta K^\mu{}_\mu{}^2 - \frac{\bar{M}_3(t)^2}{2}\delta K^\mu{}_\nu \delta K^\nu{}_\mu + ...\Big] . \tag{163}$$

As we said all the coefficients of the operators in the action above may have a generic time dependence. However we are interested in solutions where H and $\dot{H}$ do not vary significantly in one Hubble time. Therefore it is natural to assume that the same holds for all the other operators. With this assumption the Lagrangian is approximately time translation invariant [8]. Therefore the time dependence generated by loop effects will be suppressed by a small breaking parameter [9]. This assumption is particularly convenient since the rapid time dependence of the coefficients can win against the friction created by the exponential expansion, so that

[8]The limit in which the time shift is an exact symmetry must be taken with care because $\dot{H} \to 0$. This implies that the spatial kinetic term for the Goldstone vanishes, as we will see in the discussion of Ghost Inflation.

[9]Notice that this symmetry has nothing to do with the breaking of time diffeomorphisms. To see how this symmetry appears in the ϕ language notice that, after a proper field redefinition, one can always assume that $\dot{\phi} = $ const. With this choice, invariance under time translation in the unitary gauge Lagrangian is implied by the shift symmetry $\phi \to \phi +$ const. This symmetry and the time translation symmetry of the ϕ Lagrangian are broken down to the diagonal subgroup by the background. This residual symmetry is the time shift in the unitary gauge Lagrangian.

inflation may cease to be a dynamical attractor, which is necessary to solve the homogeneity problem of standard FRW cosmology.

It is important to stress that this approach does describe the most generic Lagrangian not only for the scalar mode, but also for gravity. High energy effects will be encoded for example in operators containing the perturbations in the Riemann tensor $\delta R_{\mu\nu\rho\sigma}$. As these corrections are of higher order in derivatives, we will not explicitly talk about them below.

Let us give some examples of how to write simple models of inflation in this language. A model with minimal kinetic term and a slow-roll potential $V(\phi)$ can be written in unitary gauge as

$$\int d^4x \sqrt{-g}\left[-\frac{1}{2}(\partial\phi)^2 - V(\phi)\right] \to \int d^4x \sqrt{-g}\left[-\frac{\dot\phi_0(t)^2}{2}g^{00} - V(\phi_0(t))\right] . \tag{164}$$

As the Friedmann equations give $\dot\phi_0(t)^2 = -2M_P^2\dot H$ and $V(\phi(t)) = M_{\rm Pl}^2(3H^2 + \dot H)$ we see that the action is of the form (163) with all but the first three terms set to zero. Clearly this cannot be true exactly as all the other terms will be generated by loop corrections: they encode all the possible effects of high energy physics on this simple slow-roll model of inflation.

A more general case includes all the possible Lagrangians with at most one derivative acting on each ϕ: $L = P(X, \phi)$, with $X = g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$. Around an unperturbed solution $\phi_0(t)$ we have

$$S = \int d^4x \sqrt{-g}\; P(\dot\phi_0(t)^2 g^{00}, \phi(t)) \tag{165}$$

which is clearly of the form above with $M_n^4(t) = \dot\phi_0(t)^{2n}\partial^n P/\partial X^n$ evaluated at $\phi_0(t)$. Terms containing the extrinsic curvature contain more than one derivative acting on a single scalar and will be crucial in the limit of exact de Sitter, $\dot H \to 0$. They reproduce ghost inflation and new models that are discovered in this set up.

4.5.2 Action for the Goldstone Boson

The unitary gauge Lagrangian is very general, but it is clearly not very intuitive. For example, in a particular limit, it contains standard slow roll inflation. But where is the scalar degree of freedom? This is so complicated because it is the unitary gauge Lagrangian of a spontaneously broken gauge symmetry.

Goldstone boson equivalence theorem: The unitary gauge Lagrangian describes three degrees of freedom: the two graviton helicities and a scalar mode. This mode will become explicit after one performs a broken time diffeomorphism (Stückelberg trick) as the Goldstone boson which non-linearly realizes this symmetry. In analogy with the equivalence theorem for the longitudinal components of a massive gauge boson [**?**], we expect that the physics of the Goldstone decouples from the two graviton helicities at short distance, when the mixing can be neglected. Let us review briefly what happens in a non-Abelian gauge theory before applying the same method in our case.

The unitary gauge action for a non-Abelian gauge group A_μ^a is

$$S = \int d^4x \; -\frac{1}{4}{\rm Tr}\, F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}m^2{\rm Tr}\, A_\mu A^\mu \ , \tag{166}$$

where $A_\mu = A_\mu^a T^a$. Under a gauge transformation we have

$$A_\mu \to U A_\mu U^\dagger + \frac{i}{g} U \partial_\mu U^\dagger \equiv \frac{i}{g} U D_\mu U^\dagger \ . \tag{167}$$

The action therefore becomes

$$S = \int d^4x \ -\frac{1}{4} \mathrm{Tr}\, F_{\mu\nu} F^{\mu\nu} - \frac{1}{2}\frac{m^2}{g^2} \mathrm{Tr} D_\mu U^\dagger D_\mu U \ . \tag{168}$$

The mass term was not gauge invariant, and so we have factors of U in that term. The gauge invariance can be "restored" writing $U = \exp\left[iT^a \pi^a(t,\vec{x})\right]$, where π^a are scalars (the Goldstones) which transform non-linearly under a gauge transformation Λ as

$$e^{iT^a \tilde{\pi}^a(t,\vec{x})} = \Lambda(t,\vec{x})\, e^{iT^a \pi^a(t,\vec{x})} \tag{169}$$

Notice that if for a moment we consider the case in which the gauge theory is a $U(1)$ theory, we would have

$$\Lambda = e^{i\alpha(\vec{x},t)} \ , \qquad \Rightarrow \qquad \pi \to \tilde{\pi} = \pi + \alpha \tag{170}$$

π shifts under a gauge transformation. This is a non-linear transformation because 0 is not mapped into 0. Gauge invariance has been restored by reintroducing a particle that however, transforms non-linearly. Gauge invariance is non-linearly realized.

Going to canonical normalization

$$\frac{m^2}{g^2} (\partial \pi)^2 \quad \Rightarrow \quad \pi_c \equiv m/g \cdot \pi \tag{171}$$

we see that the Goldstone boson self-interactions become strongly coupled at the scale $4\pi m/g$, which is parametrically higher than the mass of the gauge bosons. The advantage of reintroducing the Goldstones is that for energies $E \gg m$ the mixing between them and the transverse components of the gauge field becomes irrelevant, so that the two sectors decouple. Mixing terms in eq. (167) are in fact of the form

$$\frac{m^2}{g} A_\mu^a \partial^\mu \pi^a = m A_\mu^a \partial^\mu \pi_c^a \tag{172}$$

which are irrelevant with respect to the canonical kinetic term $(\partial \pi_c)^2$ for $E \gg m$.

Notice that from expanding the term $D_\mu U D^\mu U$ we obtain irrelevant (i.e. non-renormalizable) terms of the form

$$\frac{m^2}{g^2} \pi^2 (\partial \pi)^2 \sim \frac{1}{m^2/g^2} \pi_c^2 (\partial \pi_c)^2 \tag{173}$$

This is an operator that becomes strongly coupled and leads to unitarity violation at energies $E \sim 4\pi m/g$.

In the window $m \ll E \ll 4\pi m/g$ the physics of the Goldstone π is weakly coupled and it can be studied neglecting the mixing with transverse components.

Let us follow the same steps for our case of broken time diffeomorphisms. Let us concentrate for instance on the two operators:

$$\int d^4x\,\sqrt{-g}\left[A(t)+B(t)g^{00}(x)\right]\ . \tag{174}$$

Under a broken time diff. $t\to\widetilde{t}=t+\xi^0(x)$, $\vec{x}\to\widetilde{\vec{x}}=\vec{x}$, g^{00} transforms as:

$$g^{00}(x)\to\widetilde{g}^{00}(\widetilde{x}(x))=\frac{\partial\widetilde{x}^0(x)}{\partial x^\mu}\frac{\partial\widetilde{x}^0(x)}{\partial x^\nu}g^{\mu\nu}(x)\,. \tag{175}$$

The action written in terms of the transformed fields is given by:

$$\int d^4x\,\sqrt{-\widetilde{g}(\widetilde{x}(x))}\left|\frac{\partial\widetilde{x}}{\partial x}\right|\left[A(t)+B(t)\frac{\partial x^0}{\partial\widetilde{x}^\mu}\frac{\partial x^0}{\partial\widetilde{x}^\nu}\widetilde{g}^{\mu\nu}(\widetilde{x}(x))\right]\ . \tag{176}$$

Changing integration variables to $\widetilde{x}$, we get:

$$\int d^4\widetilde{x}\,\sqrt{-\widetilde{g}(\widetilde{x})}\left[A(\widetilde{t}-\xi^0(x(\widetilde{x})))+B(\widetilde{t}-\xi^0(x(\widetilde{x})))\frac{\partial(\widetilde{t}-\xi^0(x(\widetilde{x})))}{\partial\widetilde{x}^\mu}\frac{\partial(\widetilde{t}-\xi^0(x(\widetilde{x})))}{\partial\widetilde{x}^\nu}\widetilde{g}^{\mu\nu}(\widetilde{x})\right] \tag{177}$$

The procedure to reintroduce the Goldstone is now similar to the gauge theory case. Whenever ξ^0 appears in the action above, we make the substitution

$$\xi^0(x(\widetilde{x}))\to-\widetilde{\pi}(\widetilde{x})\,. \tag{178}$$

This gives, dropping the tildes for simplicity:

$$\int d^4x\,\sqrt{-g(x)}\left[A(t+\pi(x))+B(t+\pi(x))\frac{\partial(t+\pi(x))}{\partial x^\mu}\frac{\partial(t+\pi(x))}{\partial x^\nu}g^{\mu\nu}(x)\right]. \tag{179}$$

One can check that the action above is invariant under diffs at all orders (and not only for infinitesimal transformations) upon assigning to π the transformation rule

$$\pi(x)\to\widetilde{\pi}(\widetilde{x}(x))=\pi(x)-\xi^0(x)\ . \tag{180}$$

With this definition π transforms as a scalar field plus an additional shift under time diffs. Notice that diff. invariant terms did not get a π.

Applying this procedure to the unitary gauge action (163) we obtain

$$\begin{aligned}S=\int d^4x\,\sqrt{-g}\quad&\Big[\frac{1}{2}M_{\rm Pl}^2R-M_{\rm Pl}^2\left(3H^2(t+\pi)+\dot{H}(t+\pi)\right)+\\&+M_{\rm Pl}^2\dot{H}(t+\pi)\left((\partial_\mu(t+\pi)\partial_\nu(t+\pi)g^{\mu\nu}\right)+\\&\frac{M_2(t+\pi)^4}{2!}\left(\partial_\mu(t+\pi)\partial_\nu(t+\pi)g^{\mu\nu}+1\right)^2+\\&\frac{M_3(t+\pi)^4}{3!}\left(\partial_\mu(t+\pi)\partial_\nu(t+\pi)g^{\mu\nu}+1\right)^3+...\Big]\ ,\end{aligned} \tag{181}$$

where for the moment we have neglected for simplicity terms that involve the extrinsic curvature.

This action is rather complicated, and at this point it is not clear what is the advantage of reintroducing the Goldstone π from the unitary gauge Lagrangian. In analogy with the gauge theory case, the simplification occurs because, at sufficiently short distances, the physics of the Goldstone can be studied neglecting metric fluctuations (this is nothing but the equivalence principle). As for the gauge theory case, the regime for which this is possible can be estimated just looking at the mixing terms in the Lagrangian above. In eq.(181) we see in fact that quadratic terms which mix π and $g_{\mu\nu}$ contain fewer derivatives than the kinetic term of π so that they can be neglected above some high energy scale. In general the answer will depend on which operators are present. Let us here just do the simplest case in which only the tadpole terms are relevant ($M_2 = M_3 = \ldots = 0$). This includes the standard slow-roll inflation case. The leading mixing with gravity will come from a term of the form

$$\sim M_{\rm Pl}^2 \dot{H} \dot{\pi} \delta g^{00} \ . \tag{182}$$

We see that

$$\begin{array}{lcccl} \text{Kinetic term} \sim M_{\rm Pl}^2 \dot{H} \delta g^{00} & \rightarrow & M_{\rm Pl}^2 \dot{H} \left(\partial_\mu (t+\pi) \partial_\nu (t+\pi) g^{\mu\nu} \right) & \supset & M_{\rm Pl}^2 \dot{H} \dot{\pi}^2 \\ \text{Mixing term} \sim M_{\rm Pl}^2 \dot{H} \delta g^{00} & \rightarrow & M_{\rm Pl}^2 \dot{H} \left(\partial_\mu (t+\pi) \partial_\nu (t+\pi) g^{\mu\nu} \right) & \supset & M_{\rm Pl}^2 \dot{H} \delta g^{00} \dot{\pi} \end{array} \tag{183}$$

δg^{00} is a constrained variable, it is the gravitational potential, and it is determined by π. At short distances, the Newtonian approximation holds:

$$M_{\rm Pl}^2 H \partial_i \delta g^{00} \sim\sim \dot{H} M_{\rm Pl}^2 \partial_i \pi \quad \Rightarrow \quad \delta g^{00} \sim \frac{\dot{H}}{H} \pi \ . \tag{184}$$

We have

$$\frac{\text{Mixing term}}{\text{Kinetic term}} \sim \frac{\delta g^{00}}{\dot{\pi}} \sim \frac{\dot{H}}{H} \frac{\pi}{\dot{\pi}} \sim \frac{\dot{H}}{EH} \ll 1 \quad \Rightarrow \quad E \gg \epsilon H \ , \tag{185}$$

where we have used that at energies of order E, $\partial_t \sim E$. The mixing term is negligible in the UV (GR equivalence principle). The actual scale $E_{\rm mix}$ at which the mixing can be neglected depends on the actual operators turned on, but it is guaranteed that at energies $E \gg E_{\rm mix}$ we can neglect the mixing terms.

In the regime $E \gg E_{\rm mix}$ the action dramatically simplifies to

$$S_\pi = \int d^4x \sqrt{-g} \left[\frac{1}{2} M_{\rm Pl}^2 R - M_{\rm Pl}^2 \dot{H} \left(\dot{\pi}^2 - \frac{(\partial_i \pi)^2}{a^2} \right) + 2 M_2^4 \left(\dot{\pi}^2 + \dot{\pi}^3 - \dot{\pi} \frac{(\partial_i \pi)^2}{a^2} \right) - \frac{4}{3} M_3^4 \dot{\pi}^3 + \ldots \right] \tag{186}$$

Given an inflationary model, one is interested in computing predictions for present cosmological observations. From this point of view, it seems that the decoupling limit (186) is completely irrelevant for these extremely infrared scales. However, as for standard single field slow-roll inflation, one can prove that there exists a quantity, the usual ζ variable, which is constant out of the horizon at any order in perturbation theory

Therefore the problem is reduced to calculating correlation functions just after horizon crossing. We are therefore interested in studying our Lagrangian with an IR energy cutoff of order H. If the decoupling scale $E_{\rm mix}$ is smaller than H, the Lagrangian for π (186) will give the correct predictions up to terms suppressed by $E_{\rm mix}/H$. When this is not the case, nothing dramatic happens: we simply have to keep also the metric fluctuations.

This is the justification of the calculations we did in lecture 2.

As we discussed, we are assuming that the time dependence of the coefficients in the unitary gauge Lagrangian is slow compared to the Hubble time, that is, suppressed by some generalized slow roll parameters. This implies that the additional π terms coming from the Taylor expansion of the coefficients are small. In particular, the relevant operators, *i.e.* the ones which dominate moving towards the infrared, like the cubic term, are unimportant at the scale H and have therefore been neglected in the Lagrangian (186).

In conclusion, with the Lagrangian (186) one is able to compute all the observables which are not dominated by the mixing with gravity, like for example the non-Gaussianities in standard slow-roll inflation [**?**, **?**]. Notice however that the tilt of the spectrum can be calculated, at leading order, with the Lagrangian (186). As we will see later, its value can in fact be deduced simply by the power spectrum at horizon crossing computed neglecting the mixing terms. It is important to stress that our approach does not lose its validity when the mixing with gravity is important so that the Goldstone action is not sufficient for predictions. The action (163) contains all the information about the model and can be used to calculate all predictions even when the mixing with gravity is large.

Let us stress a few points

- The above Lagrangian is very simple, and it unifies all single-degree-of-freedon inflationary models.
- It describes the theory of the fluctuations, which is what we are actually testing.
- It is analogous to the Chiral Lagrangian of particle physics. Indeed, it is telling us that from the experimental point of view, inflation is the theory of a Goldstone boson
- Since it encodes all possible models on inflation, it allows to prove theorems on the possible signals.
- It also allows to explore all possible signatures.
- What is forced by symmetries, what are the allowed operators and what is possible to do is made clear. For example, the coefficient of $(\partial_i\pi)^2$ is fixed to be $\dot{H}M_{\rm Pl}^2$. This is not the case for $\dot{\pi}^2$. This tells us that at leading order in derivatives it is impossible to violate the null energy condition. $\dot{H}>0$ implies that the spatial kinetic term for π has the negative-energy sign, and so it leads to an uncontrollable instability. The EFT also tells you how this problem can be fixed, by adding higher derivative terms. Indeed all currently known ways to violate NEC that are currently known have been found in this context.

- This formalism is very prone to do with it whist we do for the beyond the standard model physics: one can add symmetries to enhance operators with respect to others, or one can try to UV complete some specific models.

- Being explicitly a theory for the fluctuations, it allows to assess the important of operators very easily. For example, in the standard treatment with scalar fields, an operators $(\partial\phi)^8$ contributes to the quadratic action with $\dot{\phi}_0^6(\partial\delta\phi)^2$. This is also very useful for studying loop corrections. At a fixed order in fluctuations and derivatives, in the EFT there is a finite number of counter terms, while this is not so with the scalar field theory. Indeed the EFT formalism was crucial to prove the constancy of ζ at quantum level.

4.6 Rigorous calculation of the power spectrum in Unitary gauge

We are now ready to see the new spectacular signatures of inflation. But I really feel that it is time for us to do a rigorous calculation. Notice that we got so far without having to do one at all. Pretty good I would say. However, there is little more rewarding that seeing your simple estimates being confirmed by a somewhat tricky calculation.

We just saw that we could neglect metric perturbations for standard slow roll inflation. And indeed we did the correct calculation neglecting them. Additionally, we saw that using π makes it explicit this fact. in order to see that we did not loose anything, we will now do the calculation in an un-intuitive gauge. The so called ζ-gauge or Maldacena-gauge. This is one of the gauges that are possible in our unitary gauge. Even though it is unintuitive, it is due to Maldacena, so, it must be good for something! Indeed, it is the absolutely best gauge to study the tricky infrared properties of ζ, the variable we ultimately need to compute. In this gauge, we will see that in the infrared ζ becomes constant. Unfortunately, as we discussed, unitary gauges are the worst possible gauges to see the decoupling of matter perturbations from metric perturbations. I am not aware of a gauge which is equally nice both in the UV and the IR at the same time. I will however show you later how to do the calculation using π.

We said that we want to compute the correlation function of ζ. Let us write the metric in the so-called ADM parametrization

$$ds^2 = -N^2 dt^2 + h_{ij}\left(dx^i + N^i dt\right)\left(dx^j + N^j dt\right) \tag{187}$$

We have to quantize a system with Gauge redundancy. In our case the gauge freedom (sometimes historically and wrongly called gauge symmetry) is time-dependent spatial diff invariance. The quantization is tricky, but it is the same as for gauge theories. Just a different symmetry group. The procedure is the following (see Weinberg's QFT I and II books).

- Expand the action. In ADM parametrization, it reads

$$S = \frac{1}{2}\int \sqrt{h}\left[NR^{(3)} + \frac{1}{N}\left(E_{ij}E^{ij} - E_i^{i2}\right) - 2M_{\rm Pl}^2\dot{H}\cdot\frac{1}{N} - M_{\rm Pl}^2(3H^2+\dot{H})\cdot N\right] \tag{188}$$

where

$$E_{ij} = \frac{1}{2}\left[\partial_t h_{ij} + \nabla_i N_j + \nabla_j N_i\right] \tag{189}$$

and ∇ is the covariant derivative with respect to h_{ij}.

Now take equations of motions with respect to all fluctuating degrees of freedom.

$$\frac{\delta S}{\delta\, \delta g^{\mu\nu}} = 0\ ,\ , \tag{190}$$

- For semplicity, we do the calculation for $M_{2,\ldots} = 0$ (this included slow roll inflation). In this case the equations of motion for N and N_i take the following form

$$\nabla_i\left[N^{-1}\left(E^i_j - \delta E\right)\right] = 0 \tag{191}$$
$$M^2_{\rm Pl}\left[R^{(3)} - \frac{1}{N^2}\left(E_{ij}E^{ij} - E^{i2}_i\right)\right] - (3H^2 + \dot{H}) + 2M^2_{\rm Pl}\dot{H}\cdot\frac{1}{N^2} = 0 = 0$$

  These two equations are extremely important. Notice that no time derivative acts on N nor on N_i. These tells us that N and N_i are *constrained* variables: they are known once you specify what the other degrees of freedom do. They are *not* degree of freedom. They are very much (and not by chance) like the gravitation potential in Newtonian gravity, or the Electric potential in electrostatic.

- Let us count the degrees of freedom. We started with the metric, which has 10 components. But we have 3-independent gauge generators for the spatial diffs. This means that we can set of these components to any value we want (including 0). This means that they are not degrees of freedom. For example we can set to zero 3 components of h_{ij}. Then from above, we see that N, N^i are 4 constrained variables. So they are also not degrees of freedom. We are left with

$$\text{number of degrees of freedom} = 10 - 3 - 4 = 3 \tag{192}$$

  Does this work? We should have the two elicities of the graviton and the matter degree of freedom (equivalent to π): 3. Ok, we are on!

- Fix a gauge. Fix the spatial diffs by fixing the spatial metric to be

$$h_{ij} = a^2\delta_{ij}e^{2\zeta} \tag{193}$$

  I am neglecting tensor perturbations here, because as said at quadratic level they do not mix. This gauge is called ζ-gauge or Maldacena-gauge. My recommendation: *do not use gauge invariant variables*. This is my humble opinion, though I should say many people like them. To me, they are just a complicated way to do the calculation in Newtonian gauge, which I do not even find it to be a particularly nice gauge. They were historically extremely useful and important, but for us let us choose a gauge, and let us get the best one. ζ-gauge has a good chance of being one of the best (we will see later that another very good one is the π-gauge.).

- In this gauge you can clearly see why $\zeta = \delta a/a$. Assuming that N and N_i go to their unperturbed value when $k/(aH) \to 0$ [10], then you see that you are in an perturbed FRW (as $\delta\phi = 0$), with just a δa.

- The constrained variables N and N^i are constrained, and so we can solve for them in terms of the only remaining degree of freedom: ζ. The solution reads

$$N = 1 + \frac{\dot{\zeta}}{H} \; , \qquad N_i = \partial_i \left(-\frac{1}{a^2}\frac{\dot{\zeta}}{H} - \frac{\dot{H}}{H^2}\frac{1}{\partial^2}\dot{\zeta} \right) \tag{194}$$

- Plug back this values for N and N^i in the action. Notice, you can do this only because they are constrained variables. The action now reads

$$S = \int d^4x \, a^3 \left(-\frac{\dot{H}}{H^2} \right) M_{\rm Pl}^2 \left[\dot{\zeta}^2 - \frac{1}{a^2}(\partial_i \zeta)^2 \right] \tag{195}$$

- Let us quantize the system. Follow textbook: find

$$\Pi_\zeta = \frac{\delta \mathcal{L}}{\delta \dot{\zeta}} = -2a^3 \left(\frac{\dot{H}}{H^2} \right) \dot{\zeta} \tag{196}$$

$$[\zeta, P_\zeta] = i \tag{197}$$

This is a quadratic Lagrangian, so we simply expand the fourier components of ζ in annihilation and creation operators

$$\hat{\zeta}_{\vec{k}}(t) = \zeta^{cl}_{\vec{k}}(t) a^\dagger_{\vec{k}} + \zeta^\star_{cl}(t) a_{-\vec{k}} \tag{198}$$

with ζ^{cl} satisfying the equation of motion (Heisemberg equation for $\hat{\zeta}$)

$$0 = \frac{\delta L}{\delta \zeta} = \frac{d\left(a^3 \left(-\frac{\dot{H}}{H^2} \dot{\zeta}^{cl}_k \right) \right)}{dt} + \frac{\dot{H}}{H^2} a k^2 \zeta^{cl}_k \tag{199}$$

This is a second order equation, that requires two initial conditions:=. Imposing the normalization condition

$$[a, a^\dagger] = (2\pi)^3 \delta^{(3)}(\vec{k} + \vec{k}') \tag{200}$$

gives a normalization condition for ζ^{cl}:

$$-a^3 \frac{\dot{H}}{H^2} M_{\rm Pl}^2 {\rm Im} \left[\zeta^{cl\star}_k \dot{\zeta}^{cl}_k \right] = 1 \tag{201}$$

The second condition has to do with the definition of the vacuum state. We define the vacuum state as the state annihilated by $a_{\vec{k}}$:

$$a_{\vec{k}} |0\rangle = 0 \tag{202}$$

[10] Indeed this will be true because N, N_i are constrained variables sourced by the gradients of ζ.

but what this state is actually depends on what we choose as ζ^{cl}. How do we choose it? Well, we know that at early times, the mode $k/a \gg 1$, so we would like the solution to be the same as in Minkowski space (this is GR!). In other words, the vacuum state for modes well inside H^{-1} should be the same as in flat space. This give the following condition

$$\zeta_k^{cl}(-k\eta \gg 1) \sim \frac{-i}{(2\epsilon)^{1/2} M_{\rm Pl} a(\eta)^3} \frac{1}{(2k/a(\eta))^{1/2}} e^{ik\eta} \quad \text{for} \quad \frac{k}{aH} = -k\eta \gg 1 \tag{203}$$

Notice that the exponential reads $k\eta \simeq \frac{k}{a} a\eta \simeq k_{\rm pays} t$. The pre factors come from the canonical normalization. This is the solution that we would get for an harmonic oscillator $1/\sqrt{2\omega}$ after we take into account of the rescaling to make the field canonical.

- Now we can solve the linear equation. Since at early times the Hubble expansion is negligible, and at late times ζ goes to a constant, we can neglect the time dependence of $H, \dot{H}$, and evaluate those terms at freeze out (it is possible to solve that equation exactly at first order in slow roll parameters. You can do this yourself). Using Mathematica, the solution reads

$$\zeta_k^{cl}(\eta) = \frac{1}{(2\epsilon)^{1/2} M_{\rm Pl}} \frac{1}{(2k)^{3/2}} (1 - ik\eta) e^{ik\eta} \ , \tag{204}$$

- We can now compute the power spectrum:

$$\langle 0|\zeta_{\vec{k}}(\eta)\zeta_{\vec{k}'}(\eta')|0\rangle = (2\pi)^3\delta^3(\vec{k}+\vec{k}')\times \frac{1}{(2\epsilon)^{1/2} M_{\rm Pl}} \frac{1}{(2k)^{3/2}} (1-ik\eta)e^{ik\eta} \frac{1}{(2\epsilon)^{1/2} M_{\rm Pl}} \frac{1}{(2k)^{3/2}} (1+ik\eta')e^{-ik\eta'} \tag{205}$$

when $k\eta \ll 1$ and $k\eta' \ll 1$, we obtain

$$\langle \zeta_{\vec{k}}\zeta_{\vec{k}'}\rangle_{\rm late} = (2\pi)^3\delta^3(\vec{k}+\vec{k}')\ \frac{1}{k^3} \cdot \frac{H^4}{4(-\dot{H})M_{\rm Pl}^2} \tag{206}$$

Which nicely reproduces the results we found with our estimates (but now we even got the factor of 4!).

- One can compute correlation functions not on the vacuum state. Vacuum is somewhat better justified, though generalizations have been found.

- We could have done exactly the same calculation in a gauge where π is zero, and we fix space and time diffs so that $h_{ij} = a^2\delta_{ij}$. The only subtlety is that at late times π is *not* constant, but ζ is the constant quantity. We therefore need a relationship between π and ζ. This is given by performing a time-diff $\delta t = \pi$ to go from π-gauge to ζ-gauge. Quite intuitively, the relationship is

$$\zeta = -H\pi \tag{207}$$

So, one computed $\langle \pi\pi\rangle$ up to freeze out, and then one switches to ζ. Very simply

$$\langle \zeta_{\vec{k}}\zeta_{\vec{k}'}\rangle_{\rm late} = H^2\langle \pi_{\vec{k}}\pi_{\vec{k}'}\rangle_{\rm f.o.} \tag{208}$$

and one has never to talk about constraint equations and metric variables.

4.6.1 The various limits of single field inflation

Slow-roll inflation and high energy corrections

The simplest example of the general Lagrangian (163) is obtained by keeping only the first three terms, which are fixed once we know the background Hubble parameter $H(t)$, and setting to zero all the other operators of higher order: $M_2 = M_3 = \bar{M}_1 = \bar{M}_2 \ldots = 0$. In the ϕ language, this corresponds to standard slow-roll inflation, with no higher order terms. We have already done this case, both using π or using ζ.

Notice however that not all observables can be calculated from the π Lagrangian (186): this happens when the leading result comes from the mixing with gravity or is of higher order in the slow-roll expansion. For example, as the first two terms of eq. (186) do not contain self-interactions of π, the 3-point function $\langle\zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\rangle$would be zero. One is therefore forced to look at subleading corrections, taking into account the mixing with gravity in eq. (181).

Obviously our choice of setting to zero all the higher order terms cannot be exactly true. At the very least they will be radiatively generated even if we put them to zero at tree level. The theory is non-renormalizable and all interactions will be generated with divergent coefficients at sufficiently high order in the perturbative expansion. As additional terms are generated by graviton loops, they may be very small. For example it is straightforward to check that starting from the unitary gauge interaction $M_{\rm Pl}^2 \dot{H} g^{00}$ a term of the form $(\delta g^{00})^2$ will be generated with a logarithmically divergent coefficient $M_2^4 \sim \dot{H}^2 \log\Lambda$. This implies that one should assume $M_2^4 \gtrsim \dot{H}^2$ ([11]). This lower limit is however very small. For example the dispersion relation of π will be changed by the additional contribution to the time kinetic term: this implies, as we will discuss thoroughly below, that the speed of π excitations deviates slightly from the speed of light, by a relative amount $1-c_s \sim M_2^4/(|\dot{H}|M_{\rm Pl}^2) \sim |\dot{H}|/M_{\rm Pl}^2$. Using the normalization of the scalar spectrum, we see that the deviation from the speed of light is $\gtrsim \epsilon^2 \cdot 10^{-10}$. A not very interesting lower limit.

The size of the additional operators will be much larger if additional physics enters below the Planck scale. In general this approach gives the correct parametrization of all possible effects of new physics. As usual in an effective field theory approach, the details of the UV completion of the model are encoded in the higher dimension operators. This is very similar to what happens in physics beyond the Standard Model. At low energy the possible effects of new physics are encoded in a series of higher dimensional operators compatible with the symmetries [**?**]. The detailed experimental study of the Standard model allows us to put severe limits on the size of these higher dimensional operators. The same can be done in our case, although the set of conceivable observations is unfortunately much more limited.

Small speed of sound and large non-Gaussianities

The Goldstone action (186) shows that the spatial kinetic term $(\partial_i\pi)^2$ is completely fixed by the background evolution to be $M_{\rm Pl}^2 \dot{H}(\partial_i\pi)^2$. In particular only for $\dot{H} < 0$, it has the "healthy"

[11] The explicit calculation of logarithmic divergences in a theory of a massless scalar coupled to gravity has been carried out a long time ago in [**?**].

negative sign. This is an example of the well studied relationship between violation of the null energy condition, which in a FRW Universe is equivalent to $\dot{H} < 0$, and the presence of instabilities in the system [?, ?]. Notice however that the wrong sign of the operator $(\partial_i \pi)^2$ is not enough to conclude that the system is pathological: higher order terms like $\delta K^{\mu}{}_{\mu}{}^2$ may become important in particular regimes, as we will discuss thoroughly below. Reference [?] studies examples in which $\dot{H} > 0$ can be obtained without pathologies.

The coefficient of the time kinetic term $\dot{\pi}^2$ is, on the other hand, not completely fixed by the background evolution, as it receives a contribution also from the quadratic operator $(\delta g^{00})^2$. In eq. (186) we have

$$\left(-M_{\rm Pl}^2 \dot{H} + 2M_2^4\right) \dot{\pi}^2 \, . \tag{209}$$

To avoid instabilities we must have $-M_{\rm Pl}^2 \dot{H} + 2M_2^4 > 0$. As time and spatial kinetic terms have different coefficients, π waves will have a "speed of sound" $c_s \neq 1$. This is expected as the background spontaneously breaks Lorentz invariance, so that $c_s = 1$ is not protected by any symmetry. As we discussed in the last section, deviation from $c_s = 1$ will be induced at the very least by graviton loops [12]. The speed of sound is given by

$$c_s^{-2} = 1 - \frac{2M_2^4}{M_{\rm Pl}^2 \dot{H}} \, . \tag{210}$$

This implies that in order to avoid superluminal propagation we must have $M_2^4 > 0$ (assuming $\dot{H} < 0$). Superluminal propagation would imply that the theory has no Lorentz invariant UV completion [?]. In the following we will concentrate on the case $c_s \leq 1$, see [?] for a phenomenological discussion of models with $c_s > 1$.

Using the equation above for c_s^2 the Goldstone action can be written at cubic order as

$$S_\pi = \int d^4x \, \sqrt{-g} \left[-\frac{M_{\rm Pl}^2 \dot{H}}{c_s^2} \left(\dot{\pi}^2 - c_s^2 \frac{(\partial_i \pi)^2}{a^2} \right) + M_{\rm Pl}^2 \dot{H} \left(1 - \frac{1}{c_s^2} \right) \left(\dot{\pi}^3 - \dot{\pi} \frac{(\partial_i \pi)^2}{a^2} \right) - \frac{4}{3} M_3^4 \dot{\pi}^3 ... \right] \tag{211}$$

From the discussion in section (4.5.2) we know that the mixing with gravity can be neglected at energies $E \gg E_{\rm mix} \simeq \epsilon H$.

The calculation of the 2-point function follows closely the case $c_s = 1$ if we use a rescaled momentum $\bar{k} = c_s k$ and take into account the additional factor c_s^{-2} in front of the time kinetic term. We obtain

$$\langle \zeta(\vec{k}_1) \zeta(\vec{k}_2) \rangle = (2\pi)^3 \delta(\vec{k}_1 + \vec{k}_2) \frac{1}{c_{s*}} \cdot \frac{H_*^4}{4 M_{\rm Pl}^2 |\dot{H}_*|} \frac{1}{k_1^3} = (2\pi)^3 \delta(\vec{k}_1 + \vec{k}_2) \frac{1}{c_{s*}} \cdot \frac{H_*^2}{4 \epsilon_* M_{\rm Pl}^2} \frac{1}{k_1^3} \, . \tag{212}$$

[12] If we neglect the coupling with gravity and the time dependence of the operators in the unitary gauge Lagrangian (so that $\pi \to \pi + \text{const}$ is a symmetry), $c_s = 1$ can be protected by a symmetry $\partial_\mu \pi \to \partial_\mu \pi + v_\mu$, where v_μ is a constant vector. Under this symmetry the Lorentz invariant kinetic term of π changes by a total derivative, while the operator proportional to M_2^4 in eq. (186) is clearly not invariant, so that $c_s = 1$. Notice that the theory is not free as we are allowed to write interactions with more derivatives acting on π. This symmetry appears in the study of the brane bending mode of the DGP model [?].

The variation with time of the speed of sound introduces an additional contribution to the tilt

$$n_s = \frac{d}{d\log k}\log\frac{H_*^4}{|\dot H_*|c_{s*}} = \frac{1}{H_*}\frac{d}{dt_*}\log\frac{H_*^4}{|\dot H_*|c_{s*}} = 4\frac{\dot H_*}{H_*^2} - \frac{\ddot H_*}{\dot H_* H_*} - \frac{\dot c_{s*}}{c_{s*}H_*}\ . \tag{213}$$

From the action (211) we clearly see that the same operator giving a reduced speed of sound induces cubic couplings of the Goldstones of the form $\dot\pi(\nabla\pi)^2$ and $\dot\pi^3$. The non-linear realization of time diffeomorphisms forces a relation between a reduced speed of sound and an enhanced level of the 3-point function correlator, *i.e.* non-Gaussianities. Indeed remember that the ϕ-wavefunction was a Gaussian in the vacuum state simply because the action was quadratic in the fields. Interactions will lead to deviation from a Gaussian wavefunction: i.e. non-Gaussianities.

To estimate the size of non-Gaussianities, one has to compare the non-linear corrections with the quadratic terms around freezing, $\omega\sim H$. In the limit $c_s \ll 1$, the operator $\dot\pi(\nabla\pi)^2$ gives the leading contribution, as the quadratic action shows that a mode freezes with $k/a\sim H/c_s$, so that spatial derivatives are enhanced with respect to time derivatives. Notice indeed that

$$H\sim\omega\sim c_s\frac{k}{a}\ ,\quad\Rightarrow\quad \frac{k}{a(t_{f.o.})}\sim\frac{H}{c_s}\gg H\ . \tag{214}$$

The level of non-Gaussianity will thus be given by the ratio:

$$\frac{\mathcal{L}_{\dot\pi(\nabla\pi)^2}}{\mathcal{L}_2}\sim\frac{H\pi\left(\frac{H}{c_s}\pi\right)^2}{H^2\pi^2}\sim\frac{H}{c_s^2}\pi\sim\frac{1}{c_s^2}\zeta\ , \tag{215}$$

where in the last step we have used the linear relationship between π and ζ. Taking $\zeta\sim 10^{-5}$ we have an estimate of the size of the non-linear correction . Usually the magnitude of non-Gaussianities is given in terms of the parameters $f_{\rm NL}$, which are parametrically of the form:

$$\frac{\mathcal{L}_{\dot\pi(\nabla\pi)^2}}{\mathcal{L}_2}\sim f_{\rm NL}\zeta \tag{216}$$

The leading contribution will thus give

$$f^{\rm equil.}_{{\rm NL},\ \dot\pi(\nabla\pi)^2}\sim\frac{1}{c_s^2}\ . \tag{217}$$

The superscript "equil." refers to the momentum dependence of the 3-point function, which in these models is of the so called equilateral form [?]. This is physically clear in the Goldstone language as the relevant π interactions contain derivatives, so that they die out quickly out of the horizon; the correlation is only among modes with comparable wavelength.

In the Goldstone Lagrangian (211) there is an additional independent operator, $-\frac{4}{3}M_3^4\dot\pi^3$, contributing to the 3-point function, coming from the unitary gauge operator $(\delta g^{00})^3$. We thus have two contributions of the form $\dot\pi^3$ which give

$$f^{\rm equil.}_{{\rm NL},\ \dot\pi^3}\sim 1-\frac{4}{3}\frac{M_3^4}{M_{\rm Pl}^2|\dot H|c_s^{-2}}\ . \tag{218}$$

The size of the operator $-\frac{4}{3}M_3^4\dot{\pi}^3$ is not constrained by the non-linear realization of time diffeomorphisms: it is a free parameter. In DBI inflation [?] we have $M_3^4 \sim M_{\rm Pl}^2|\dot{H}|c_s^{-4}$, so that its contribution to non-Gaussianities is of the same order as the one of eq. (217). The same approximate size of the M_3^4 is obtained if we assume that both the unitary gauge operators $M_2^4(\delta g^{00})^2$ and $M_3^4(\delta g^{00})^3$ become strongly coupled at the same energy scale.

It is interesting to look at the experimental limits on non-Gaussianities as a constraint on the size of the unitary gauge operator $(\delta g^{00})^2$ and therefore on the speed of sound. The explicit calculation [?] gives the contribution of the operator $\dot{\pi}(\nabla\pi)^2$ to the experimentally constrained parameter $f_{\rm NL}^{\rm equil.}$; at leading in order in c_s^{-1} we have

$$f_{\rm NL}^{\rm equil.} = \frac{85}{324}\cdot\frac{1}{c_s^2}\ . \qquad (219)$$

Cutoff and Naturalness

As discussed, for $c_s < 1$ the Goldstone action contains non-renormalizable interactions. Therefore the self-interactions among the Goldstones will become strongly coupled at a certain energy scale, which sets the cutoff of our theory. This cutoff can be estimated looking at tree level partial wave unitarity, *i.e.* finding the maximum energy at which the tree level scattering of πs is unitary. The calculation is straightforward, the only complication coming from the non-relativistic dispersion relation. The cutoff scale Λ turns out to be

$$\Lambda^4 \simeq 16\pi^2 M_2^4\frac{c_s^7}{(1-c_s^2)^2} \simeq 16\pi^2 M_{\rm Pl}^2|\dot{H}|\frac{c_s^5}{1-c_s^2}\ . \qquad (220)$$

The same result can be obtained looking at the energy scale where loop corrections to the $\pi\pi$ scattering amplitude become relevant. As expected the theory becomes more and more strongly coupled for small c_s, so that the cutoff scale decreases. On the other hand, for $c_s \to 1$ the cutoff becomes higher and higher. This makes sense as there are no non-renormalizable interactions in this limit and the cutoff can be extended up to the Planck scale. This cutoff scale is obtained just looking at the unitary gauge operator $(\delta g^{00})^2$; depending on their size the other independent operators may give an even lower energy cutoff. Notice that the scale Λ indicates the maximum energy at which our theory is weakly coupled and make sense; below this scale new physics must come into the game. However new physics can appear even much below Λ.

If we are interested in using our Lagrangian for making predictions for cosmological correlation functions, then we need to use it at a scale of order the Hubble parameter H during inflation. We therefore need that this energy scale is below the cutoff, $H \ll \Lambda$. Using the explicit expression for the cutoff (220) in the case $c_s \ll 1$ one gets

$$H^4 \ll M_{\rm Pl}^2|\dot{H}|c_s^5 \qquad (221)$$

which can be rewritten using the spectrum normalization (212) as an inequality for the speed of sound

$$c_s \gg P_\zeta^{1/4} \simeq 0.003\ . \qquad (222)$$

A theory with a lower speed of sound is strongly coupled at $E \simeq H$. Not surprisingly this value of the speed of sound also corresponds to the value at which non-Gaussianity are of order one: the theory is strongly coupled at the energy scale H relevant for cosmological predictions.

Let us comment on the naturalness of the theory. One may wonder whether the limit of small c_s is natural or instead loop corrections will induce a larger value. The Goldstone self-interactions, $\dot{\pi}(\nabla\pi)^2$ and $(\nabla\pi)^4$ for example, will induce a radiative contribution to $(\nabla\pi)^2$. It is easy to estimate that these contributions are of order $c_s^{-5}\Lambda^4/(16\pi^2 M_2^4)$, where Λ is the UV cutoff, *i.e.* the energy scale at which new physics enters in the game. We can see that it is impossible to have large radiative contribution; even if we take Λ at the unitarity limit (220), the effect is of the same order as the tree level value. This makes sense as the unitarity cutoff is indeed the energy scale at which loop corrections become of order one.

We would like also to notice that the action (186) is natural from an effective field theory point of view [**?**]. The relevant operators are in fact protected from large renormalizations if we assume an approximate shift symmetry of π. In this case the coefficients of the relevant operators are sufficiently small and they will never become important for observations as cosmological correlation functions probe the theory at a fixed energy scale of order H: we never go to lower energy. Clearly here we are only looking at the period of inflation, where an approximate shift symmetry is enough to make the theory technically natural; providing a graceful exit from inflation and an efficient reheating are additional requirements for a working model which are not discussed in our formalism.

De-Sitter Limit and the Ghost Condensate

In the previous section we saw that the limit $c_s \to 0$ is pathological as the theory becomes more and more strongly coupled. However we have neglected in our discussion the higher derivative operators in the unitary gauge Lagrangian (163)

$$\int d^4x\sqrt{-g}\left(-\frac{\bar{M}_2(t)^2}{2}\delta K^\mu{}_\mu{}^2 - \frac{\bar{M}_3(t)^2}{2}\delta K^\mu{}_\nu \delta K^\nu{}_\mu\right) . \tag{223}$$

These operators give rise in the Goldstone action to a spatial kinetic term of the form

$$\int d^4x\,\sqrt{-g}\left[-\frac{\bar{M}^2}{2}\,\frac{1}{a^4}(\partial_i^2\pi)^2\right] , \tag{224}$$

where $\bar{M}^2 = \bar{M}_2^2 + \bar{M}_3^2$. Notice that we obtain the very non-relativistic dispersion relation

$$\omega^2 \sim \frac{k^4}{M^2} . \tag{225}$$

This models naturally leads to large non-Gaussianities.

De-Sitter Limit without the Ghost Condensate

In this section we want to study the effect of the operator beginequation

$$\int d^4x\,\sqrt{-g}\,\left(-\frac{\bar{M}_1(t)^3}{2}\delta g^{00}\delta K^\mu{}_\mu\right) . \tag{226}$$

on the quadratic π action. We will see that, if the coefficient of this operator is sufficiently large, we obtain a different de Sitter limit, where the dispersion relation at freezing is of the form $\omega^2 \propto k^2$, instead of the Ghost Condensate behavior $\omega^2 \propto k^4$.

For simplicity we can take $\bar{M}_1$ to be time independent. Reintroducing the Goldstone we get a 3-derivative term of the form $-\bar{M}_1^3 \dot{\pi} \nabla^2 \pi / a^2$ ([13]). This would be a total time derivative without the time dependence of the scale factor $a(t)$ and of the metric determinant. Integrating by parts we get a standard 2-derivative spatial kinetic term

$$-\int d^4x \sqrt{-g}\, \frac{\bar{M}_1^3 H}{2} \left(\frac{\partial_i}{a}\pi\right)^2 . \tag{227}$$

In the exact de Sitter limit, $\dot{H} = 0$, and taking $M_2 \sim \bar{M}_1 \sim M$, this operator gives a dispersion relation of the form

$$c_s^2 = \frac{H}{M} \ll 1 . \tag{228}$$

and naturally to large non-Gaussianities.

This and the Ghost condensate case are finally the only known ways to violet the Null Energy Condition in a stable way []

5 Lecture 4. . . 4.5: Non-Gaussianity: who are you?

we have seen in the former section that we can have inflationary models with large self-interactions. We said that they produce some non-Gaussianity. Indeed we saw that in the limit of free-theory the vacuum wavefunction was a Gaussian. This was because the Lagrangian was quadratic and each fourier mode was like an harmonic oscillator. But if the action is slightly non-linear, than we can imagine some slight non-Gaussianity. Something like, just symbolically:

$$|0\rangle_{k_i/a \ll H} \sim \prod_{\{\zeta_{\vec{k}}\}} e^{-\frac{\zeta^2_{\vec{k}_i}}{\sigma_{\zeta_{\vec{k}_i}}} - \frac{\zeta_{\vec{k}_i}\zeta_{\vec{k}_j}\zeta_{\vec{k}_i+\vec{k}_j}}{C(\vec{k}_1,\vec{k}_2,\vec{k}_1+\vec{k}_2)}} |\{\zeta_{\vec{k}}\}\rangle \tag{229}$$

This would me that a signal like the three-point function

$$\langle \zeta_{\vec{k}_1}\zeta_{\vec{k}_2}\zeta_{\vec{k}_3}\rangle = (2\pi^3)\delta^{(3)}(\vec{k}_1+\vec{k}_2+\vec{k}_3)\bar{F}(\vec{k}_1,\vec{k}_2,\vec{k}_3) \tag{230}$$

would not be zero. Current limits sets

$$\frac{\langle \zeta^3 \rangle}{\langle \zeta^2 \rangle} \lesssim 10^{-2} \sim \frac{1}{N_{pix}^{1/2}} \tag{231}$$

[13] The operator gives also a contribution to $\dot{\pi}^2$ proportional to H. We will assume that this is small compared to $M_2^4 \dot{\pi}^2$. In Minkowski space the operator we are studying can be forbidden by a $\phi \to -\phi$ symmetry, which is equivalent to time reversal in unitary gauge [?]. In a de Sitter background this symmetry is broken by the metric, so that this operator cannot be set to zero.

which is a very small number! Look at the plot.

figure on two gaussians

Being a limit on a statistics, the limit scale as $N_{pix}^{-1/2}$. For WMAP, we have indeed about 10^5 modes.

But what this tells us is that a detection of non-Gaussianities would be associated to the interacting part of the Lagrangian, which is really the interesting part of the Lagrangian! And we are talking of interactions at extremely high energies! Interactions contain so much more information that they would allow to learn about the real dynamics that drove inflation.

Clearly, since non-Gaussianities are small, it is expectable that the leading signature will appear in the 3-point function.

Let us look at the function F. So far it depends on 9 variables. But let us use the symmetries of the problem. By the cyclic invariance of the correlation function (remember that at late times we are semiclassical), we can set $k_1 \geq k_2 \geq k_3$. Translation invariance forces the sum of the three momenta to be zero: they must form a closed triangle. We are down to 6. We can use to ration invariance to point $\vec{k}_1$ in the $\hat{x}$ direction, and $\vec{k}_2$ in the $x-y$ plane. We are down to 3 variables. Additionally, the 3-point function should be scale invariant, because two triplet of modes, one an overall rescaling of the other, see approximately the same history. We can use this to set the modulus of $k_1 = 1$. The overall k_1 dependence has to be $1/k_1^6$, so that the real space 3-point function

$$\langle \zeta(x)^3 \rangle = \int d^3k_1 d^3k_2 d^3k_3 \; \langle \zeta_{\vec{k}_1} \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \rangle \tag{232}$$

receives the same contribution from each logarithmic interval.

So, in terms of degrees of freedom, we are down to two variables

$$\begin{aligned} \langle \zeta_{\vec{k}_1} \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \rangle &= (2\pi^3)\delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)\frac{1}{k_1^6} F(x_2, x_3) \qquad (233) \\ x_2 = k_2/k_1 \;, &\quad x_3 = k_3/k_1 \;, \quad x_3 \leq x_2 \leq 1 \;, \quad x_3 \leq x_2 \end{aligned}$$

This is a *huge* amount of information. Remember that because of the various symmetries, the 2-point function had to go as $1/k^3$, and so it dependent only on one *number*. Because of the slight deviation of scale invariance, we had also the tilt, which is just a second *number*. Here with non-Gaussianities, we are talking about a function of 2 *parameters*. This is ∞ numbers! this is a huge amount of information, incomparable with respect to the information contained in the 2-point function. Indeed, it has the same amount of information as a 2-2 scattering as a function of angles. And this is not little thing: we learn about spin of particles and nature of interactions from this. Let us plot F. A useful quantity to plot is a quantity the resembles the signal to noise ratio in each triangular configuration. It is

$$\left.\frac{S}{N}\right|_{triangle} \sim x_2^2 x_3^2 F(1, x_2, x_3) \tag{234}$$

which is a function of the triangular shape. A typical shape is the following:

figure on shape

Isn't this a beauty? It has a lot of information. Such a detection would really make us confident that something very non-trivial was going on in the sky. It would also teach a lot about the dynamics that drove inflation.

5.1 Computation of 3-point function

Let us see how to compute this F. In the EFT of inflation, we have seen that at leading order in derivatives we have two interaction operators: $\dot{\pi}^3$ and $\dot{\pi}(\partial_i\pi)^2$. Let us compute the shape due to the first, as an example.

This is nothing by a QFT exercise, just follow the rules.

- We have an interacting theory. Very much as we do when computing scattering amplitudes or correlation functions in Minkowski, we go to the interaction picture. We split the Hamiltonian in

$$H = H_0 + H_{int} \tag{235}$$

and evolve the operators with H_0 and the state with H_{int}. Since the evolution under H_0 is completely understood, we need simply to evolve the state with the interaction picture evolutor

$$U_{int}(t, t_{in}) = T e^{-i \int_{t_{in}}^{t} dt' H_{int}(t')} \tag{236}$$

where T denotes time ordering.

- What we would like to compute is the expectation value of $\zeta_{\vec{k}_1}\zeta_{\vec{k}_2}\zeta_{\vec{k}_3}$ evaluated on the initial state of the theory, which is the vacuum $|\Omega(t_{in})\rangle$, evolved to time t.

  $$|\Omega(t)\rangle = U_{int}(t, t_{in})|\Omega(t_{in})\rangle \ . \tag{237}$$

  We then have:

  $$\langle\Omega(t)|\zeta_{\vec{k}_1}\zeta_{\vec{k}_2}\zeta_{\vec{k}_3}|\Omega(t)\rangle = \langle\Omega(t_{in})| \left(\bar{T} e^{i \int_{t_{in}}^{t} dt' H_{int}(t')}\right) \zeta^{int}_{\vec{k}_1}\zeta^{int}_{\vec{k}_2}\zeta^{int}_{\vec{k}_3} \left(T e^{-i \int_{t_{in}}^{t} dt' H_{int}(t')}\right) |\Omega(t_{in})\rangle \tag{238}$$

  with $\bar{T}$ representing one-time ordering and ζ^{int} the interaction picture operator.

  Notice that this expectation value is taken between two *in* states. This is why it is called in-in formalism. Notice that this is difference than what one usually do in scattering amplitudes, where one computes in-out correlation functions. This expelling the $\bar{T}$ appearing in the above expression. This is the source of a series of differences with scattering amplitude. For example, the results are not independent of field redefinitions. we wish to compute correlation functions of ζ.

- How do we compute the vacuum state? We know how to express well states in the Fock base, so, it would be good to express $|\Omega(t)\rangle$ in this base. It is possible to express $|\Omega(t)\rangle$ in terms of the free theory Bunch Davies vacuum with a simple rotation in the complex plane of the contour of integration of the evaluator operator. To see this, let us expand the free theory vacuum state $|0(t_1)\rangle$ in Eigenstates of the true Hamiltonian. As we will see, we need to impose the vacuum condition at early times, so that we can neglect about the time dependence of the Hamiltonian. Let us evolve $|0(t_1)\rangle$ from t_1 to t_2

  $$\begin{aligned}|0(t_2)\rangle = e^{H(t_2-t_1)}|0(t_1)\rangle = \sum_n e^{iE_n(t_2-t_1)}|n(t_1)\rangle\ \langle n(t_1)|0(t_1)\rangle = e^{iE_0(t_2-t_1)}|\Omega(t_1)\rangle\ \langle\Omega(t_1)|0(t_1)\rangle \\ + \sum_{n\neq 0} e^{iE_n(t_2-t_1)}|n(t_1)\rangle\ \langle n(t_1)|0(t_1)\rangle \end{aligned} \tag{239}$$

  We see that by sending $t_1 \to -\infty(1+i\epsilon)$, with a very small ϵ, we are projecting away all the states apart for Ω. So we can write

  $$|\Omega\rangle_{in} = \lim_{\epsilon\to 0}\lim_{t_1\to-\infty} \frac{1}{\langle\Omega(t_1)|0(t_1)\rangle e^{i(t_2-t_1^+)}} U(t_2, t_1^+)|0(t_1)\rangle \tag{240}$$

  where the t_1^+ reminds us how the t_1 count our has been rotated. Analogously:

  $$_{in}\langle\Omega| = \lim_{\epsilon\to 0}\lim_{t_1\to-\infty} \frac{1}{\langle 0(t_1)|\Omega(t_1)\rangle e^{-i(t_2-t_1^-)}} U^\dagger(t_2, t_1^-)\langle 0(t_1)| \tag{241}$$

- We are therefore led to compute

$$\langle\Omega(t)|\zeta_{\vec{k}_1}(t)\zeta_{\vec{k}_2}(t)\zeta_{\vec{k}_3}(t)|\Omega(t)\rangle = \lim_{\epsilon\to 0}\frac{1}{|\langle 0(t_1)|\Omega(t_1)\rangle|^2 e^{-i(t_2-t_1^-)}e^{+i(t_2-t_1^+)}} \times \tag{242}$$
$$\times_{in}\langle 0|\left(\bar{T}e^{i\int_{-\infty(1+i\epsilon)}^t dt' H_{int}(t')}\right)\zeta^{int}_{\vec{k}_1}(t)\zeta^{int}_{\vec{k}_2}(t)\zeta^{int}_{\vec{k}_3}(t)\left(Te^{-i\int_{-\infty(1-i\epsilon)}^t dt' H_{int}(t')}\right)|0\rangle_{in}$$

We can divide by 1:

$$1 = \langle\Omega|\Omega\rangle = \lim_{\epsilon\to 0}\frac{1}{|\langle 0(t_1)|\Omega(t_1)\rangle|^2 e^{-i(t_2-t_1^-)}e^{+i(t_2-t_1^+)}}\langle 0(t_1)|U^\dagger(t_2,t_1^+)U(t_2,t_1^-)|0(t_1)\rangle =$$
$$= \lim_{\epsilon\to 0}\frac{1}{|\langle 0(t_1)|\Omega(t_1)\rangle|^2 e^{-i(t_2-t_1^-)}e^{+i(t_2-t_1^+)}} \tag{243}$$

- We obtain

$$\langle\Omega(t)|\zeta_{\vec{k}_1}(t)\zeta_{\vec{k}_2}(t)\zeta_{\vec{k}_3}(t)|\Omega(t)\rangle = \tag{244}$$
$$\lim_{\epsilon\to 0}{}_{in}\langle 0|\left(\bar{T}e^{i\int_{-\infty(1+i\epsilon)}^t dt' H_{int}(t')}\right)\zeta^{int}_{\vec{k}_1}(t)\zeta^{int}_{\vec{k}_2}(t)\zeta^{int}_{\vec{k}_3}(t)\left(Te^{-i\int_{-\infty(1-i\epsilon)}^t dt' H_{int}(t')}\right)|0\rangle_{in}$$

At leading order we can Taylor expand the exponential to obtain

$$\langle\Omega(t)|\zeta_{\vec{k}_1}\zeta_{\vec{k}_2}\zeta_{\vec{k}_3}|\Omega(t)\rangle \simeq -2\mathrm{Re}\left[\int_{-\infty(1-i\epsilon)}^{\tau} d\tau'\langle\zeta^{int}_{\vec{k}_1}(\tau)\zeta^{int}_{\vec{k}_2}(\tau)\zeta^{int}_{\vec{k}_3}(\tau)H_{int}(\tau')\rangle\right] \tag{245}$$

- At this order in perturbation theory, $H_{int} = -\int d^3x\mathcal{L}_{int}$. Pay attention, this is partially non trivial! Our $\mathcal{L}_{int}$ is given by

$$\mathcal{L}_{int} = -\frac{4}{3}M_3^4\int d^3x\ a^4\left(\frac{1}{a(\tau)}\frac{\partial\pi(\vec{x},\tau)}{\partial\tau}\right)^3 = \tag{246}$$
$$= -\frac{4}{3}M_3^4\int d^3k_1\,d^3k_2\,d^3k_3\ a\ \delta^3(\vec{k}_1+\vec{k}_2+\vec{k}_3)\ \pi^{int\prime}_{\vec{k}_1}(\tau)\pi^{int\prime}_{\vec{k}_2}(\tau)\pi^{int\prime}_{\vec{k}_3}(\tau)$$

The factor a^4 is due to the fact that we are integrating in conformal time.

- Use that $\zeta = -H\pi$ and that

$$\pi^{int}_{\vec{k}}(\tau) = \pi^{cl}_{\vec{k}}(\tau)a^\dagger_{\vec{k}} + \pi^{cl\star}_{\vec{k}}(\tau)a_{-\vec{k}} \tag{247}$$

with

$$\pi^{cl}_k(\tau) = -\frac{1}{H}\frac{c_s}{(2\epsilon)^{1/2}M_{\rm Pl}}\frac{1}{(2c_sk)^{3/2}}(1-ic_sk\tau)e^{ic_sk\tau}\ , \tag{248}$$

- Perform the Wick contraction, and then perform the integral. The integral reads:

$$\langle\Omega(t)|\zeta_{\vec{k}_1}\zeta_{\vec{k}_2}\zeta_{\vec{k}_3}|\Omega(t)\rangle = (-H^3)(-6)\times 2\times\frac{4}{3}M_3^4 \tag{249}$$
$$\times\mathrm{Re}\left[\pi^{cl}_{\vec{k}_1}(\tau)^\star\pi^{cl}_{\vec{k}_2}(\tau)^\star\pi^{cl}_{\vec{k}_3}(\tau)^\star\int_{-\infty(1-i\epsilon)}^{\tau}\pi^{cl}_{\vec{k}_1}{}'(\tau')\pi^{cl}_{\vec{k}_2}{}'(\tau')\pi^{cl}_{\vec{k}_3}{}'(\tau')a(\tau')\,d\tau'\right] \tag{250}$$

The results gives

$$\langle \Phi_{\vec{k}_1} \Phi_{\vec{k}_2} \Phi_{\vec{k}_3} \rangle = (2\pi)^3 \delta^{(3)}(\sum_i \vec{k}_i) F(k_1, k_2, k_3) \ . \tag{251}$$

$$F_{\dot{\pi}^3}(k_1, k_2, k_3) = \frac{20}{3}\left(1 - \frac{1}{c_s^2}\right) \tilde{c}_3 \ \cdot \Delta_\Phi^2 \cdot \frac{1}{k_1 k_2 k_3 (k_1 + k_2 + k_3)^3} \ . \tag{252}$$

where (sorry, this is congenital, it is not my fault!)

$$\Phi = \frac{3}{5}\zeta \ , \tag{253}$$

$$\Delta_\Phi = \frac{9}{25} \frac{H^2}{4\epsilon\, c_s M_{\rm Pl}^2} \ , \qquad M_3^4 = \frac{\dot{H} M_{\rm Pl}^2}{c_s^4} \tilde{c}_3 \ . \tag{254}$$

For $\tilde{c}_3 \sim 1$, we have that the unitarity bound associated to the operator in M_3 is the same as the one from the operator in M_2.

- The standard definition of f_{NL} is

$$F(k, k, k) = f_{NL} \cdot \frac{6\Delta_\Phi^2}{k^6} \ , \tag{255}$$

This allows us to define

$$f_{NL}^{\dot{\pi}(\partial_i \pi)^2} = \frac{85}{324}\left(1 - \frac{1}{c_s^2}\right) \ , \tag{256}$$

$$f_{NL}^{\dot{\pi}^3} = \frac{10}{243}\left(1 - \frac{1}{c_s^2}\right)\left(\tilde{c}_3 + \frac{3}{2}c_s^2\right) \ ,$$

5.1.1 Shape of Non-Gaussianities

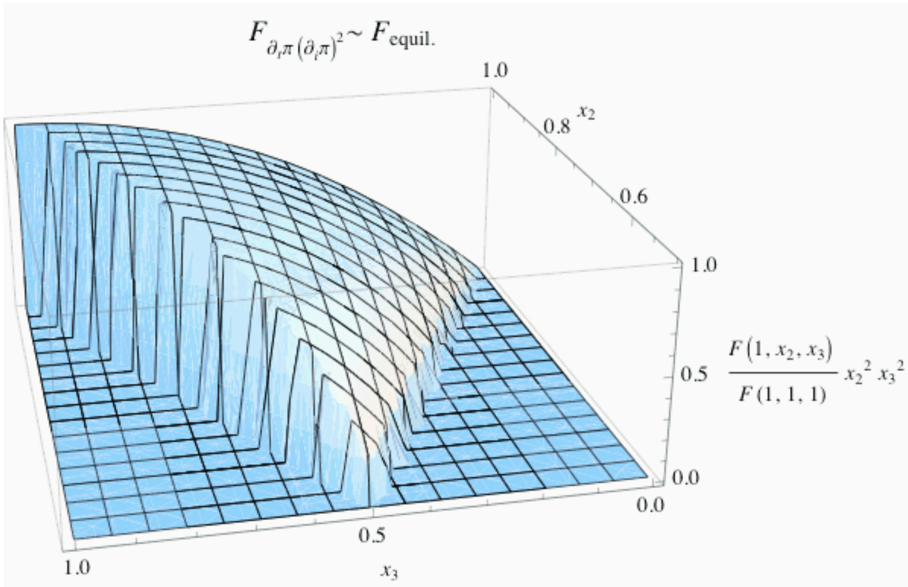
- **Huge information**

  We see that at leading order in derivatives we have two operators $\dot{\pi}^3$ and $\dot{\pi}(\partial_i \pi)^2$.. Let us see the plots.

  figure shapes

- **Local Shape:**

  As we can see, the non-Gaussian signal from these models is always very small. This is indeed a theorem due to Maldacena, In reality, in some humble sense we are now beyond that theorem, because we have the Lagrangian for any single-degree-of-freedom inflationary model. We have therefore access to all the shapes that single-clock inflation can do: if we see something different, we exclude single-degree-of-freedom inflation. But still it is a remarkable feature of single degree of freedom inflation that in the squeezed limit the signal is so small. Can there be inflationary models that give large 3-point function in that limit? Yes, multi filed inflation can do that.



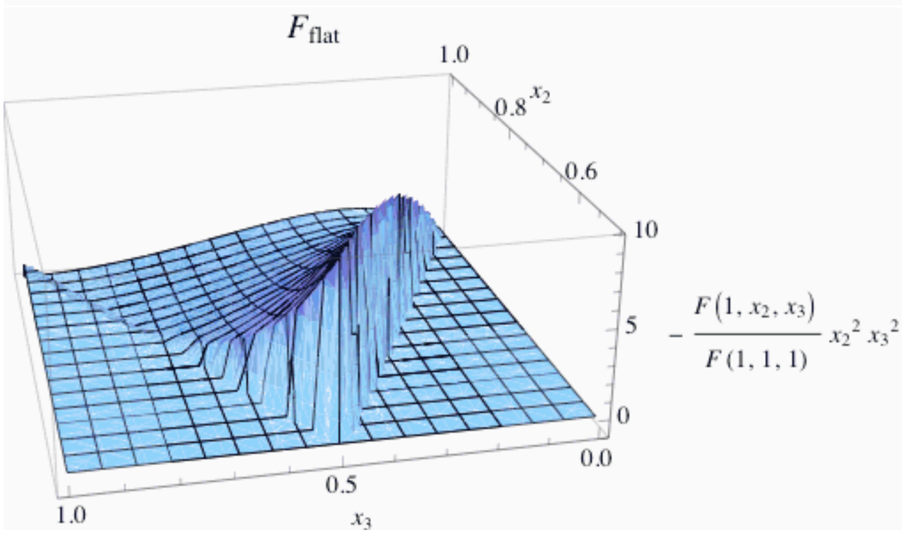


figure shapes

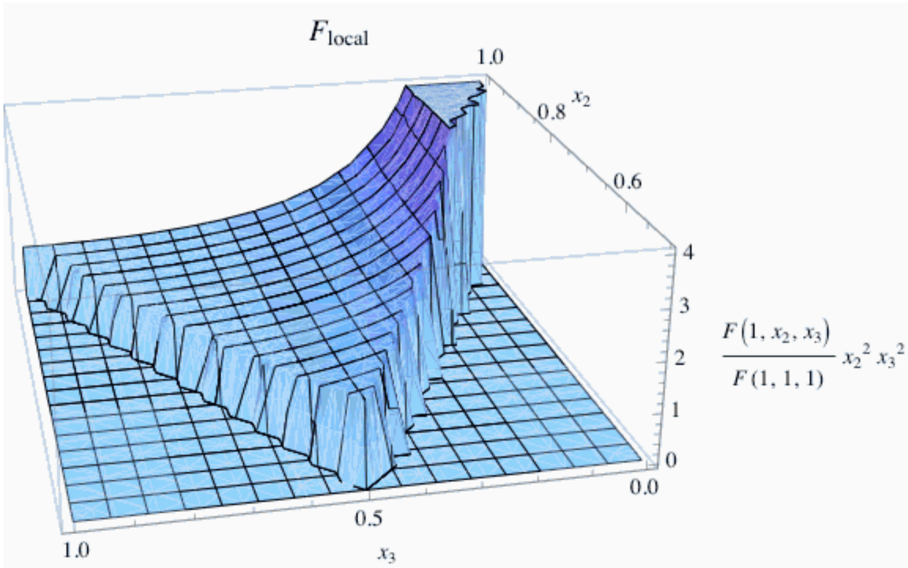
A shape with a lot of signal there is a shape where the fluctuation ζ is defined in real space with the help of an auxiliary gaussian field:

$$\zeta(\vec{x}) = \zeta_{gaussian}(\vec{x}) + \frac{6}{5} f_{NL}^{local} \left(\zeta_{gaussian}(\vec{x})^2 - \langle \zeta_{gaussian}(\vec{x})^2 \rangle \right) \tag{257}$$

Its F reads something like

$$F_{local}(k_1, k_2, k_3) = \frac{1}{k_1^3 k_2^3} + \frac{1}{k_2^3 k_3^3} + \frac{1}{k_1^3 k_3^3} \tag{258}$$

Such a non-Guassianity is generated for example when the duration of inflation depends on a second field which fluctuates during inflation. For example. this could happen if



the decay rate γ of the inflation is determined by a coupling that depends in turns from a light field σ.

figure wave potential

In this way:

$$\frac{\delta a}{a} = \zeta(\vec{x}) = f(\Gamma(\{\sigma\})) \tag{259}$$

Since the conversion of the σ fluctuations into $\delta a/a$ happens when all the interesting modes are outside of H^{-1}, the relation above must be local in real space:

$$\zeta(\vec{x}) = f(\Gamma(\sigma(\vec{x}))) \tag{260}$$

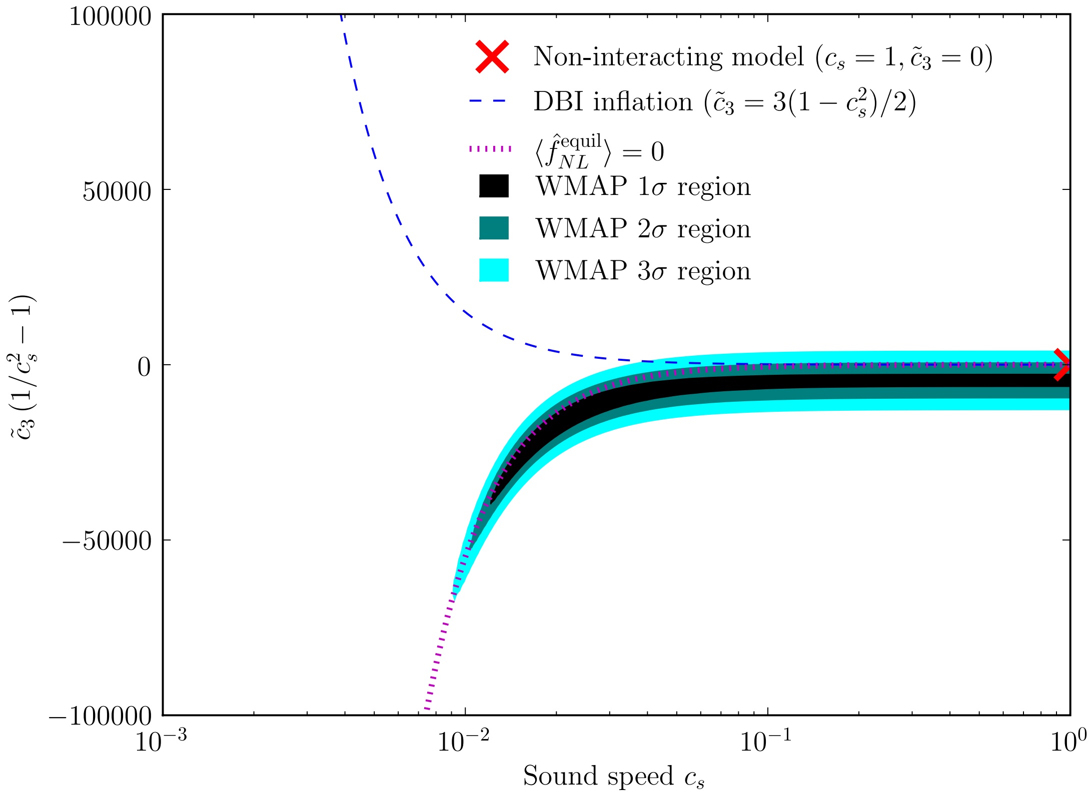
Since the non-gaussianities are quite small, the linear term must dominate. We can taylor expand f:

$$\zeta(\vec{x}) \simeq a_0 + a_1\sigma(\vec{x}) + a_2\sigma(\vec{x})^2 \equiv \zeta_{gaussian}(\vec{x}) + \frac{3}{5} f_{NL}^{local}\left(\zeta_{gaussian}(\vec{x})^2 - \langle\zeta_{gaussian}(\vec{x})^2\rangle\right) \tag{261}$$

- **Particle Physics Knowledge**

  Limits on non-Gaussian signatures get translated into limits onto limits of the parameters of the inflationary Lagrangian. Show plots. Cosmological observations are mapped directly into parameters of a fundamental physics Lagrangian. the sky is like a particle accelerator!

  figure Lagrangian constraints



The is really a lot more to say about non-Guassianities and the EFT of Inflation. Non-Gaussianities have really become a large field in inflationary cosmology, and maybe this is happening also for the EFT of inflation, as this is the ideal set up to study interactions. Indeed, many additional developments have been made in this field, that I have no time to mention: EFT of multi field inflation, impose additional symmetries on π. such as Supersymmetry, discrete shift symmetry, parity, etc. .. roughly, all what we have been doing in Beyond the Standard Model physics has now motivation to be applied to inflation and the EFT of inflation offers the simple connection.

6 Summary

This is all Guys.

In these lecture we have started from the shortcomings of Big Bang Cosmology that motivated inflation. We have seen how a period of accelerated expansion fixes all this problem. With simple estimates that are helpful to develop intuition, we have seen how inflation produce a quasi scale-invariant, quasi-Gaussian, stochastic but classical, spectrum of density perturbations, and how some qualitative predictions of inflation have been confirmed in the data. We have also seen that it would be great to have something more to look for. For this reason, we have introduced the Effective Field Theory of Inflation, which shows that Inflation is essentially a theory of a Goldstone boson. We have seen that there are new spectacular signatures in inflation: the non-Gaussianity of the density perturbation. They contain a huge amount of information, and they represent the interactions, and therefore the non-trivial dynamics, of the inflationary Lagrangian.

Inflationary physics is very ample, and there are many aspects that we could not touch. For example we did not discuss how some inflationary models are embedded in string theory, or what the beautiful phase called eternal inflation, according to which quantum effects change the asymptotic of the space-time, arise.

In any event, for all what concerns the phenomenology of Inflation and its connection to the data, you should be good to go.

Thank you very much for your attention and your interactions. Teaching at TASI has been a wonderful experience for me, and it has been a pleasure to have you around and discuss with you. I hope you'll find these lecture useful for your future research in Physics and Cosmology. It is a great moment for our field.

My best wishes.

Appendix

A The most general Lagrangian in unitary gauge

Let us study what are the rules for writing down the most general Lagrangian in unitary gauge. In a theory which is only invariant under spatial diffeomorphisms there is a preferred slicing of spacetime given by a function $\tilde{t}(x)$ (with time-like gradient), which non-linearly realizes time diffeomorphisms. For example if the breaking is given by a time evolving scalar, surfaces of constant $\tilde{t}$ are also of constant value of the scalar. Unitary gauge is the one in which the time coordinate t is chosen to coincide with $\tilde{t}$, so that the additional degree of freedom $\tilde{t}$ does not explicitly appear in the action. One can therefore build various terms:

1. Terms which are invariant under all diffeomorphisms: these are just polynomials of the Riemann tensor $R_{\mu\nu\rho\sigma}$ and of its covariant derivatives, contracted to give a scalar[14].

2. A generic function of $\tilde{t}$ becomes $f(t)$ in unitary gauge. We are therefore free to use generic functions of time in front of any terms in the Lagrangian.

3. The gradient $\partial_\mu \tilde{t}$ becomes δ^0_μ in unitary gauge. Thus in every tensor we can always leave free an upper 0 index. For example we can use g^{00} (and functions of it) in the unitary gauge Lagrangian, or the component of the Ricci tensor R^{00}.

4. It is useful to define a unit vector perpendicular to surfaces of constant $\tilde{t}$

$$n_\mu = \frac{\partial_\mu \tilde{t}}{\sqrt{-g^{\mu\nu}\partial_\mu \tilde{t}\partial_\nu \tilde{t}}} \, . \tag{262}$$

   This allows to define the induced spatial metric on surfaces of constant $\tilde{t}$: $h_{\mu\nu} \equiv g_{\mu\nu} + n_\mu n_\nu$. Every tensor can be projected on the surfaces using $h_{\mu\nu}$. In particular we can use in our action the Riemann tensor of the induced 3d metric ${}^{(3)}R_{\alpha\beta\gamma\delta}$ and covariant derivatives with respect to the 3d metric.

5. Additional possibilities will come from the covariant derivatives of $\partial_\mu \tilde{t}$. Notice that we can equivalently look at covariant derivatives of n_μ: the derivative acting on the normalization factor just gives terms like $\partial_\mu g^{00}$ which are covariant on their own and can be used in the unitary gauge Lagrangian. The covariant derivative of n_μ projected on the surfaces of constant $\tilde{t}$ gives the extrinsic curvature of these surfaces

$$K_{\mu\nu} \equiv h^\sigma_\mu \nabla_\sigma n_\nu \, . \tag{263}$$

   The index ν is already projected on the surface because $n^\nu \nabla_\sigma n_\nu = \frac{1}{2}\nabla_\sigma (n^\nu n_\nu) = 0$. The covariant derivative of n_ν perpendicular to the surface can be rewritten as

$$n^\sigma \nabla_\sigma n_\nu = -\frac{1}{2}(-g^{00})^{-1} h^\mu_\nu \partial_\mu (-g^{00}) \tag{264}$$

[14] The metric and the completely antisymmetric tensor $(-g)^{-1/2}\epsilon^{\mu\nu\rho\sigma}$ can be used to contract indices.

so that it does not give rise to new terms. Therefore all covariant derivatives of n_μ can be written using the extrinsic curvature $K_{\mu\nu}$ (and its covariant derivatives) and derivatives of g^{00}.

6. Notice that using at the same time the Riemann tensor of the induced 3d metric and the extrinsic curvature is redundant as ${}^{(3)}R_{\alpha\beta\gamma\delta}$ can be rewritten using the Gauss-Codazzi relation as [**?**]
$$ {}^{(3)}R_{\alpha\beta\gamma\delta} = h^\mu_\alpha h^\nu_\beta h^\rho_\gamma h^\sigma_\delta R_{\mu\nu\rho\sigma} - K_{\alpha\gamma}K_{\beta\delta} + K_{\beta\gamma}K_{\alpha\delta} \; . \qquad (265)$$
Thus one can forget about the 3d Riemann tensor altogether. We can also avoid using the induced metric $h_{\alpha\beta}$ explicitly: written in terms of the 4d metric and n_μ one gets only terms already discussed above[15]. Finally also the use of covariant derivatives with respect to the induced 3d metric can be avoided: the 3d covariant derivative of a projected tensor can be obtained as the projection of the 4d covariant derivative [**?**].

We conclude that the most generic action in unitary gauge is given by

$$S = \int d^4x\sqrt{-g}\; F(R_{\mu\nu\rho\sigma}, g^{00}, K_{\mu\nu}, \nabla_\mu, t) \; , \qquad (266)$$

where all the free indices inside the function F must be upper 0's.

B Expanding around a given FRW solution

In this Section we want to prove that the most generic theory with broken time diffeomorphisms around a given FRW background (with $k = -1, 0, +1$ depending of the spatial curvature) can be written as

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{2}M_{\rm Pl}^2 R + M_{\rm Pl}^2\Big(\dot H - \frac{k}{a^2}\Big)\cdot g^{00} - M_{\rm Pl}^2\Big(3H^2 + \dot H + 2\frac{k}{a^2}\Big) + \ldots\right] \qquad (267)$$

where the dots stand for terms which are invariant under spatial diffeomorphisms and of quadratic (or higher) order in the fluctuations around the given FRW background[16].

As such this statement is trivial. We know that the displayed terms give rise to the wanted FRW evolution so that, if we do not want to move away from it, the additional operators must start quadratic around this solution. What we want to say is that *each one* of the additional invariant terms is quadratic (or of higher order) without cancellation of linear contributions among various operators. These terms will be written as polynomials (quadratic and higher) of linear operators like $g^{00}+1$, $\delta K_{\mu\nu} = K_{\mu\nu} - K^{(0)}_{\mu\nu}$, $\delta R_{\mu\nu\rho\sigma} \equiv R_{\mu\nu\rho\sigma} - R^{(0)}_{\mu\nu\rho\sigma}$ and so on. Notice that these terms start linear in the perturbations as we have explicitly removed their value

[15] Notice that the determinant of the induced metric is related to the one of the full 4d metric by $h = g^{00}\cdot g$ and that the completely antisymmetric 3d tensor can be rewritten in terms of the 4d one as $h^{-1/2}\epsilon^{ijk} = (-g)^{-1/2}(-g^{00})^{-1/2}\epsilon^{0ijk}$.

[16] We can always make the coefficient in front of R time independent through an appropriate field redefinition $g_{\mu\nu} \to g_{\mu\nu}\cdot f(t)$. This corresponds, in the usual formalism, to going to Einstein frame.

evaluated on the given FRW solution. Given the symmetries of a FRW metric, every tensor evaluated on the background ($K^{(0)}_{\mu\nu}$, $R^{(0)}_{\mu\nu\rho\sigma}$, $(\nabla_\alpha R_{\mu\nu\rho\sigma})^{(0)}$ ...) can be written just in terms of $g_{\mu\nu}$, n_μ and functions of time. For example

$$K^{(0)}_{\mu\nu} = a^2 H h_{\mu\nu} \tag{268}$$

$$R^{(0)}_{\mu\nu\rho\sigma} = 2(H+k)h_{\mu[\rho}h_{\sigma]\nu} + (\dot{H}+H^2)a^2 h_{\mu\sigma}\delta^0_\nu\delta^0_\rho + \text{perm.} \tag{269}$$

where k is a constant which depends on the curvature of the spatial slices and the permutations are acting only on the last term. As such all the operators evaluated on the FRW background are themselves covariant operators, so that operators like $\delta K_{\mu\nu}$ and $\delta R_{\mu\nu\rho\sigma}$ are well defined covariant operators which vanish on the given FRW background and start linear in the perturbations. We stress that this possibility of rewriting the tensors evaluated on the background holds only because of the high degree of symmetry of the FRW background and it would not be true if one were interested in expanding around a less symmetric solution, *e.g.* a non-homogeneous background.

In equation (267) only the displayed operators contain linear terms in the fluctuations, so that the coefficients of $\sqrt{-g}\, g^{00}$ and $\sqrt{-g}$ are uniquely determined by the background FRW solution.

Let us now see how the Lagrangian can always be cast in the form (267). If we take an operator composed by the contraction of two tensors T and G (the generalization with more tensors is straightforward) we can write

$$TG = \delta T \delta G + T^{(0)}G + TG^{(0)} - T^{(0)}G^{(0)}\ . \tag{270}$$

Let us discuss each term of the sum. The first one starts explicitly quadratic in the perturbation as we want. As we said, given the symmetries of the FRW background, the unperturbed tensors $T^{(0)}$ and $G^{(0)}$ can be written as functions of $g_{\mu\nu}$, n_μ and t. Therefore the last term $T^{(0)}G^{(0)}$ is just a polynomial of g^{00} with time dependent coefficients; it contains the terms $\sqrt{-g}\, g^{00}$ and $\sqrt{-g}$ plus operators which start explicitly quadratic in the perturbations. We are left with tensors of the form $T^{(0)}G$. We want to prove that also these terms can be written as the linear operators in eq. (267) plus operators that start quadratic in the fluctuations. By construction G will be linear either in $K_{\mu\nu}$ or $R_{\mu\nu\rho\sigma}$ with covariant derivatives acting on them. Covariant derivatives can be dealt with by successive integration by parts, letting them act on $T^{(0)}$ and the time dependent coefficient of the operator. In doing so we can generate extrinsic curvature terms. In this case we can reiterate eq. (270) until no covariant derivatives are left[17]. We are thus left with the only possible scalar linear terms with no covariant derivatives: $K^\mu{}_\mu$ and R^{00}. Both of them can be rewritten in a more useful form. We can integrate by parts the extrinsic curvature term

$$\int d^4x\sqrt{-g}\; f(t)K^\mu{}_\mu = \int d^4x\sqrt{-g}\; f\; \nabla_\mu n^\mu = -\int d^4x\sqrt{-g}\; n^\mu\partial_\mu f = \int d^4x\sqrt{-g}\sqrt{-g^{00}}\dot{f} \tag{271}$$

[17]There can be also powers of g^{00} from $T^{(0)}$. We can deal with them by writing $g^{00} = -1 + \delta g^{00}$ and thus generating additional contributions to the the g^{00} operator in eq. (267) plus terms which are explicitly quadratic or more in the perturbations.

While we can deal with R^{00} using the following relationship [?]:

$$(-g^{00})^{-1}R^{00} = R_{\mu\nu}n^\mu n^\nu = K^2 - K_{\mu\nu}K^{\mu\nu} - \nabla_\mu(n^\mu\nabla_\nu n^\nu) + \nabla_\nu(n^\mu\nabla_\mu n^\nu)\,. \tag{272}$$

The last two terms can again be integrated by parts:

$$\int d^4x\sqrt{-g}f(t)\nabla_\mu(n^\mu\nabla_\nu n^\nu) = -\int d^4x\sqrt{-g}\;\partial_\mu f\; n^\mu K^\nu{}_\nu\,, \tag{273}$$

$$\int d^4x\sqrt{-g}f(t)\nabla_\nu(n^\mu\nabla_\mu n^\nu) = -\int d^4x\sqrt{-g}\;\partial_\nu f\; n^\mu\nabla_\mu n^\nu = 0 \tag{274}$$

where in the last passage we have used that $\partial_\nu f \propto n_\nu$. This shows that K^μ_μ and R^{00} can be written in terms of the linear operators of eq. (267) plus invariant terms that starts quadratically in the fluctuations.

In conclusion, we have shown that the most general Lagrangian of a theory with broken time diffeomorphisms around a given FRW background can be written in the form:

$$\begin{aligned} S \;=\; &\int d^4x\sqrt{-g}\;\Big[\frac{1}{2}M_{\rm Pl}^2 R + M_{\rm Pl}^2\Big(\dot H - \frac{k}{a^2}\Big)\cdot g^{00} - M_{\rm Pl}^2\Big(3H^2 + \dot H + 2\frac{k}{a^2}\Big) + \\ &F^{(2)}(g^{00}+1, \delta K_{\mu\nu}, \delta R_{\mu\nu\rho\sigma}; \nabla_\mu; t)\Big] \end{aligned} \tag{275}$$

where $F^{(2)}$ starts quadratic in the arguments $g^{00}+1$, $\delta K_{\mu\nu}$ and $\delta R_{\mu\nu\rho\sigma}$.