# Cosmic inflation and new physics

Jinn-Ouk Gong

*Asia Pacific Center for Theoretical Physics*
*Pohang 790-784, Korea*

September 25, 2014

**Abstract**

The purpose of this note, based on the lectures delivered at the 20th Vietnam School of Physics, Quy Nhon, Vietnam, 12 - 14 Aug, 2014, is to provide comprehensive and explicit accounts on cosmic inflation and the cosmological perturbations produced during inflation in the early universe, and the new physics we can extract from observations.

## Contents

We adopt the sign convention of the textbook by Misner, Thorne & Wheeler (1973), i.e.

$$\eta^{\mu\nu} = (-+++)\,, \tag{1}$$

$$R^\mu{}_{\alpha\beta\gamma} = \Gamma^\mu_{\alpha\gamma,\beta} - \Gamma^\mu_{\alpha\beta,\gamma} + \Gamma^\mu_{\sigma\beta}\Gamma^\sigma_{\gamma\alpha} - \Gamma^\mu_{\sigma\gamma}\Gamma^\sigma_{\beta\alpha}\,, \tag{2}$$

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}\,, \tag{3}$$

$$G_{\mu\nu} = \frac{T_{\mu\nu}}{m_{\rm Pl}^2}\,, \tag{4}$$

where $m_{\rm Pl} \equiv (8\pi G)^{-1/2} \sim 10^{18}$ GeV is the Planck mass. Also, we set $c = \hbar = k_B = 1$ so that energy, mass and temperature all have the same dimension, usually described in terms of GeV.

# 1 Inflation

In this series of lectures, I would like to deliver explicit accounts on the cosmic inflation and the cosmological perturbations at linear order produced during inflation. They give rise to a number of observable quantities, from which we can extract the clues on the elusive new physics relevant for the early universe where the energy scale is far larger than what terrestrial accelerator experiments can ever reach. In the first part, therefore, I will first concentrate on inflation: what it is, why it is attractive, how it occurs, and so on.

## 1.1 Background equations

We begin with the so-called Friedmann-Robertson-Walker metric of a flat universe

$$ds^2 = -dt^2 + a^2(t)\delta_{ij}dx^idx^j\,. \tag{5}$$

This metric describes a flat, expanding universe parametrized by the "scale factor" $a(t)$. The spatial distance with the scale factor being singled out is described by $\delta_{ij}dx^idx^j$, which is called "comoving" distance. On the contrary, the "physical" distance is multiplied by the scale factor.

We first want the key equations which we will use throughout the lectures. These are given by the Einstein equation (4). At the moment it is sufficient to consider background equations.

### 1.1.1 Einstein tensor

We first consider the LHS of the Einstein equation, namely, the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\,. \tag{6}$$

We can immediately write each component of the metric tensor $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ as

$$\begin{aligned} g_{00} &= -1\,, \quad g_{ij} = a^2\delta_{ij}\,, \\ g^{00} &= -1\,, \quad g^{ij} = a^{-2}\delta^{ij}\,. \end{aligned} \tag{7}$$

We want to calculate the Christoffel symbol, the Ricci tensor and the Ricci scalar, each given by

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\sigma} \left( g_{\mu\sigma,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma} \right) , \tag{8}$$

$$R_{\mu\nu} = \Gamma^\alpha_{\mu\nu,\alpha} - \Gamma^\alpha_{\mu\alpha,\nu} + \Gamma^\alpha_{\sigma\alpha}\Gamma^\sigma_{\mu\nu} - \Gamma^\alpha_{\sigma\nu}\Gamma^\sigma_{\mu\alpha} , \tag{9}$$

$$R = g^{\mu\nu} R_{\mu\nu} , \tag{10}$$

explicitly. The non-zero components of the Christoffel symbols are, after some calculations,

$$\Gamma^0_{ij} = a^2 H \delta_{ij} , \tag{11}$$

$$\Gamma^i_{0j} = \Gamma^i_{j0} = H {\delta^i}_j , \tag{12}$$

with $H = \dot{a}/a$ being the Hubble parameter, otherwise zero. Then, easily we have

$$R_{00} = -3 \left( H^2 + \dot{H} \right) , \tag{13}$$

$$R_{ij} = a^2 \left( 3H^2 + \dot{H} \right) \delta_{ij} , \tag{14}$$

$$R = 6 \left( \dot{H} + 2H^2 \right) . \tag{15}$$

Thus, the non-zero components of the Einstein tensor (6), or more frequently ${G^\mu}_\nu = g^{\mu\rho} G_{\rho\nu}$, are

$$G_{00} = 3H^2 , \tag{16}$$

$$G_{ij} = -a^2 \left( 2\dot{H} + 3H^2 \right) \delta_{ij} , \tag{17}$$

$${G^0}_0 = -3H^2 , \tag{18}$$

$${G^i}_j = -\left( 2\dot{H} + 3H^2 \right) {\delta^i}_j . \tag{19}$$

### 1.1.2 Energy-momentum tensor

As can be read from the Einstein equation (4), the Einstein tensor which describes the structure of the space-time should be matched with the energy-momentum tensor which describes the matter residing in the space-time. On the assumption of the homogeneous and isotropic background, we may regard at the background level that the energy-momentum tensor is that of perfect fluid[1], i.e.

$${T^\mu}_\nu = \mathrm{diag}(-\rho, p, p, p) . \tag{20}$$

[1]In terms of the general (hydrodynamical) matter fluid, the energy-momentum tensor is written as

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p g_{\mu\nu} ,$$

where $u^\mu$ is the fluid 4-velocity which satisfies

$$u^\mu u_\mu = g_{\mu\nu} u^\mu u^\nu = -1 ,$$

so that $u^\mu$ is a time-like, unit 4-vector. Thus we can set $u^\mu = (1, 0, 0, 0)$. Using these we can trivially find (20).

### 1.1.3 Einstein equation

Now we can write each component of the Einstein equation (4):

$$\text{00 component:}\ H^2 = \frac{\rho}{3m_{\rm Pl}^2}\,, \tag{21}$$

$$ij\ \text{component:}\ -3H^2 - 2\dot{H} = \frac{p}{m_{\rm Pl}^2}\,. \tag{22}$$

(21) is called the Friedmann equation, which relates the Hubble parameter to the energy density. Using (21) for (22) to replace $H^2$ with $\rho$, we can find the time variation of $H$ as

$$\dot{H} = -\frac{\rho + p}{2m_{\rm Pl}^2}\,. \tag{23}$$

Or, explicitly in terms of the time derivatives of the scale factor,

$$\frac{\ddot{a}}{a} = -\frac{\rho + 3p}{6m_{\rm Pl}^2}\,. \tag{24}$$

We will refer to this equation soon. Note that by taking a time derivative of (21) and using (22) to eliminate $\dot{H}$, we can derive energy conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0\,. \tag{25}$$

This is what we can find from the conservation of energy-momentum tensor: from

$$T^\mu{}_{\nu;\mu} = T^\mu{}_{\nu,\mu} - \Gamma^\rho_{\mu\nu}T^\mu{}_\rho + \Gamma^\mu_{\rho\mu}T^\rho{}_\nu = 0\,, \tag{26}$$

we can trivially check that $\nu = 0$ component gives (25). $\nu = i$ component vanishes identically.

## 1.2 Cosmic microwave background

### 1.2.1 Generation of the CMB

With the necessary background equations, now let us see what happened in the past when the temperature was high enough. First, we note that from the conservation equation (25) that different species scale differently: ordinary particles (electron, proton, neutron...) have very large rest energy compared to the kinetic energy, so they are called pressureless matter and $p = 0$. Meanwhile, photons, or more generally relativistic particles, have $p = \rho/3$ and are called radiation. Plugging these relations into (25), we find

$$\rho_{\rm matter} \propto a^{-3}\,, \tag{27}$$

$$\rho_{\rm radiation} \propto a^{-4}\,. \tag{28}$$

We may understand that the energy density of pressureless matter is inversely proportional to the volume $\sim a^3$ which contains the matter particles, and for radiation the energy density is also proportional to the frequency, or the inverse of the wavelength, so we have one more power of the scale factor. What this tells us is that, in the past, the universe was dominated by radiation.

More radiation in the past means, of course, the universe was hotter. It was too hot to maintain neutral molecules, like hydrogen: because of the very hot temperature, electrons were energetic enough to overcome the binding energy to protons, so that the universe was filled by radiation (mostly photons), free electrons and nuclei (and dark matter). During this stage, the mean free path of photons was very short because of the Thomson scattering between free electrons and photons, maintaining thermal equilibrium. Thus, the universe was very "foggy" for photons: exactly like we cannot see very far away when the weather is very foggy. This stage continued until the universe was cooled to a critical temperature $T_c \sim 3000K$. Below this temperature, the binding energy between electrons and protons could overcome thermal background and there remained no free electron. Thus, from this time on, the universe has become transparent to photons and they could reach us after propagating for a long long time. This situation is depicted in Figure 1. These very old photons, which have traveled all the time since the moment of this "last scattering", are the cosmic microwave background (CMB). It was observed in 1965 by Penzias and Wilson by chance.

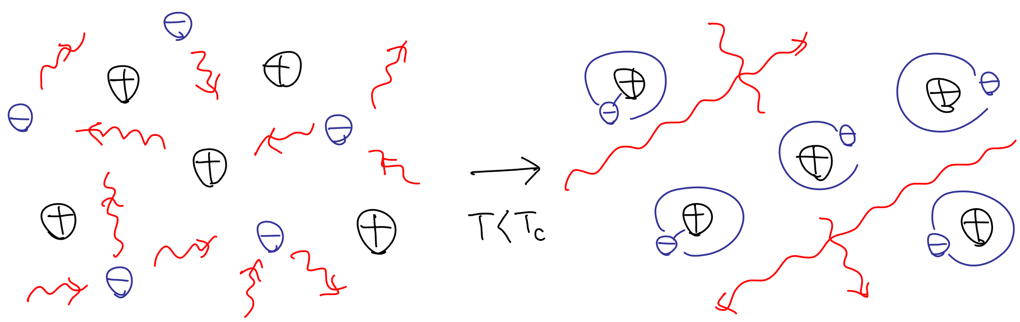


Figure 1: When $T > T_c$, electrons were free and constantly scattered off photons, so that the universe was "foggy". After the temperature drops below $T_c$, electrons are all captured by protons and photons can propagate without scattering.

The observations tell us that the CMB is extremely homogeneous and isotropic, i.e. we observe the same average temperature $T_0 \sim 2.7K$ no matter which part or direction of the sky we observe. Since photons were constantly scattering off free electrons and thus in thermal equilibrium, the temperature spectrum of the CMB exhibits that of almost perfect blackbody radiation. Moreover, the CMB could be generated only when the universe was hotter in the past. Thus the discovery of the CMB was the knockdown blow for the steady state cosmology which was competing against the hot big bang model in 60's. Note that, after removing all the contaminations and foreground effects, we have genuine temperature fluctuations of the magnitude $\delta T/T_0 \sim 10^5$. We will return to this point later. In Figure 2 we show the background and fluctuation temperature maps of the CMB.

### 1.2.2 Horizon problem

The CMB has brought, with the triumph of the hot big bang cosmology, big mysteries at the same time. Let us consider 1 of them, namely, why the CMB is so much homogeneous. For this, it is very convenient to introduce the conformal time $\tau$, defined by

$$d\tau \equiv \frac{dt}{a} \,. \tag{29}$$

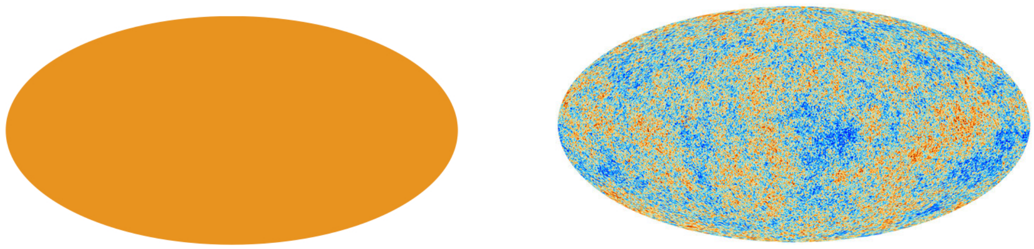

Figure 2: (Left) the cosmic microwave background is observed to be extremely homogeneous and isotropic with the average temperature $T_0 \sim 2.7K$. (Right) however, it contains genuine temperature fluctuations with respect to $T_0$ of the magnitude $\delta T/T_0 \sim 10^{-5}$. The temperature fluctuation map is taken by the Planck satellite.

With $\tau$, the line element (5) is written as

$$ds^2 = a^2(\tau)\left(-d\tau^2 + \delta_{ij}dx^i dx^j\right) , \tag{30}$$

so that the metric is written as a product of the static Minkowski metric times the scale factor. What does the conformal time mean? Let us consider the radial propagation of light, which is the null geodesic $ds^2 = 0$. Then, using the spherical coordinate we can write the radial distance $r$ a photon has traveled from some initial moment in terms of the conformal time as

$$r = \tau , \tag{31}$$

i.e. the conformal time measures the (comoving) distance a photon has traveled.

Then what's the trouble with the CMB? We can straightforwardly find that from an initial moment $i$ till some later time 0, the conformal time (i.e. the distance photons have traveled)

$$\tau = \int_i^0 \frac{dt}{a} = \int_i^0 \frac{1}{a}\frac{dt}{da}da = \int_i^0 \frac{1}{aH} d\log a \propto a^{1/2}|_i^0 , \tag{32}$$

where for each equality we have used 1) the scale factor is a function of time solely, $a = a(t)$, 2) the definition of the Hubble parameter, $\dot{a} = aH$, and 3) assumption of a matter dominated universe, $H \sim \rho_{\text{matter}}^{1/2} \sim a^{-3/2}$. Now, without loss of generality, we can take initial moment as the initial singularity $a(t_i) = 0$, where also $\tau = 0$, so that simply $\tau \propto a^{1/2}$. Further, using the relation between the scale factor which is normalized to $a_0 = 1$ at present and the redshift $z$

$$a = \frac{1}{1+z} , \tag{33}$$

we can find $\tau \propto (1+z)^{-1/2}$. Using $z_0 = 0$ and $z_{\text{CMB}} \sim 1100$, we can easily find

$$\frac{\tau_{\text{CMB}}}{\tau_0} \sim \frac{1}{\sqrt{1100^3}} \sim 0.03 . \tag{34}$$

Thus, at the moment when the CMB was generated, the past light cones stemming from the two end points do not have any overlapping region initially, i.e. those two points were never in causal communication and thus there is no reason they should have the same temperature with the accuracy of $10^{-5}$: we must impose a heavy fine tuning over $10^4 - 10^5$ causally disconnected patches at the moment of the last scattering unless we provide a natural way for them to have the same temperature. This is the so-called horizon problem. It is depicted in the left panel of Figure 3. Note that the spatial distance shown in the figure is the comoving one, thus the physical distance is obtained by multiplying the scale factor $a(t)$ which vanishes as we approach the cosmic singularity, currently at $\tau = 0$.

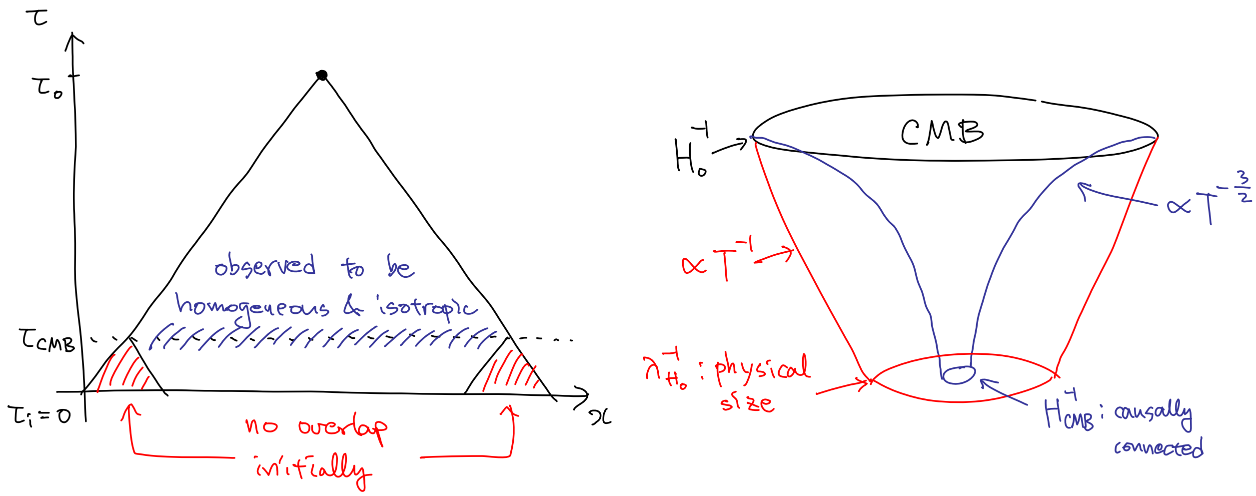


Figure 3: (Left) conformal diagram of the universe. From the cosmic singularity ($\tau_i = 0$) until the moment of the CMB generation ($\tau_{\mathrm{CMB}}$) there was no time for the CMB to achieve causal communication to have the same temperature $T_0$. (Right) As a sample calculation, we can see that at that time the universe was filled with $10^4 - 10^5$ causally disconnected patches.

To have a better idea, let us assume that the observable CMB size coincides with the current Hubble patch $1/H_0$, within which causal communications are possible. Then let us ask whether they were the same when the CMB was generated, or if different how much they were different. First, what is $\lambda_{H_0^{-1}}$, the *physical* size that corresponds to $1/H_0$? Physical sizes simply scale with the scale factor $a(t)$, which is inversely proportional to the temperature $T$. Thus, we can easily find

$$\lambda_{H_0^{-1}} = H_0^{-1}\frac{a_{\mathrm{CMB}}}{a_0} = H_0^{-1}\frac{T_0}{T_{\mathrm{CMB}}} \,. \tag{35}$$

Meanwhile, $H$ evolves according to the Friedmann equation (21). It is important to notice at this moment that $H$ depends on the energy density, i.e. which types of matter contents are there. For simplicity we assume the universe is dominated by matter that is inversely proportional to the physical volume as can be read from (27). Thus, we can find $H_{\mathrm{CMB}}^{-1}$, the Hubble horizon radius when the CMB was generated, as

$$H^2 \propto \rho_{\mathrm{matter}} \propto a^{-3} \propto T^3 \longrightarrow H_{\mathrm{CMB}}^{-1} = H_0^{-1}\left(\frac{T_0}{T_{\mathrm{CMB}}}\right)^{3/2} \,. \tag{36}$$

Thus, if we compare the ratio of these volumes,

$$\frac{\lambda_{H_0^{-1}}^3}{\left(H_{\mathrm{CMB}}^{-1}\right)^3} = \left(\frac{T_{\mathrm{CMB}}}{T_0}\right)^{3/2} \sim 4\times 10^4 \,. \tag{37}$$

That is, assuming that at present the Hubble horizon size and the CMB scale are the same, when the CMB was generated, the corresponding physical volume was filled with $10^4$ - $10^5$ causally disconnected patches: see the right panel of Figure 3. Thus, it is a tremendous fine tuning that these disconnected patches all turn out to have the same temperature with the accuracy of $10^{-5}$ as the current observations on the CMB demand!

## 1.3 Inflation

### 1.3.1 Inflation: what and how

Thus, we see that at the heart of the horizon problem lies the fact that the Hubble horizon $1/H = 1/(\dot{a}/a)$ always expands faster than the physical length scale $\lambda \sim a$,

$$\frac{d}{dt}\left(\frac{\lambda}{H^{-1}}\right) \sim \frac{d}{dt}\left[\frac{a}{(\dot{a}/a)^{-1}}\right] = \ddot{a} < 0\,, \tag{38}$$

irrespective of whether the universe is dominated by matter or radiation. Thus, we can just turn upside down and make the physical size expands faster than the Hubble horizon: then physical scales expand faster than the horizon so causal communication could be possible during this stage. This tells us

$$\frac{d}{dt}\left(\frac{\lambda}{H^{-1}}\right) > 0 \longleftrightarrow \ddot{a} > 0\,. \tag{39}$$

That is, the universe experiences an *accelerated* expansion. This period of accelerated expansion is called "*inflation*".

How can we more quantitatively say if it's inflation or not? We can rewrite (24) as

$$\frac{\ddot{a}}{a} = \frac{2\rho}{6m_{\rm Pl}^2} - \frac{3\rho + 3p}{6m_{\rm Pl}^2} = H^2 + \dot{H} > 0\,, \tag{40}$$

where the 2nd equality follows by applying (21) and (23), and the last inequality is the definition of inflation (39). Thus, inflation occurs when the following condition is satisfied:

$$\epsilon \equiv -\frac{\dot{H}}{H^2} < 1\,. \tag{41}$$

This parameter, which tells whether it's inflation or not, is called "slow-roll" parameter, in the context of slow-roll inflation: see the next section.

So with what kind of matter can we have inflation? From (24), we see that to have $\ddot{a} > 0$ we need a special form of matter which has a negative pressure,

$$p < -\frac{\rho}{3} \longleftrightarrow w \equiv \frac{p}{\rho} < -\frac{1}{3}\,. \tag{42}$$

Clearly usual pressureless matter ($w = 0$) or radiation ($w = 1/3$) cannot support inflation. The simplest candidate is the so-called cosmological constant $\Lambda$, which has

$$p_\Lambda = -\rho_\Lambda \quad (w_\Lambda = -1)\,. \tag{43}$$

Then the Friedmann equation (21) is trivially solved: since $\Lambda$ is, as the name suggests, a constant thus

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3m_{\rm Pl}^2} = \text{constant} \longleftrightarrow a = a_i \exp\left(\sqrt{\frac{\Lambda}{3m_{\rm Pl}^2}}t\right) . \tag{44}$$

Thus we can see that the scale factor increases exponentially during inflation.

### 1.3.2 Horizon problem revisited

So the question is: how does inflation solve the horizon problem? Now we can move to the conformal time to see a clear visualization how inflation solves the horizon problem. During inflation, for convenience driven by a cosmological constant $\Lambda$ so that $H$ is constant, the conformal time is given by

$$\tau = \int \frac{dt}{a} = \int \frac{e^{-Ht}}{a_0}dt = -\frac{1}{aH} < 0 . \tag{45}$$

That is, the conformal time is *negative* during inflation. Further, now the cosmic singularity $a = 0$ can be pushed to $\tau = -\infty$. Thus, even the two end points at $\tau = \tau_{\rm CMB}$ have no overlap at $\tau = 0$, now $\tau$ can be negatively indefinite so that there could be ample overlapping region enough to explain the homogeneity of the CMB.

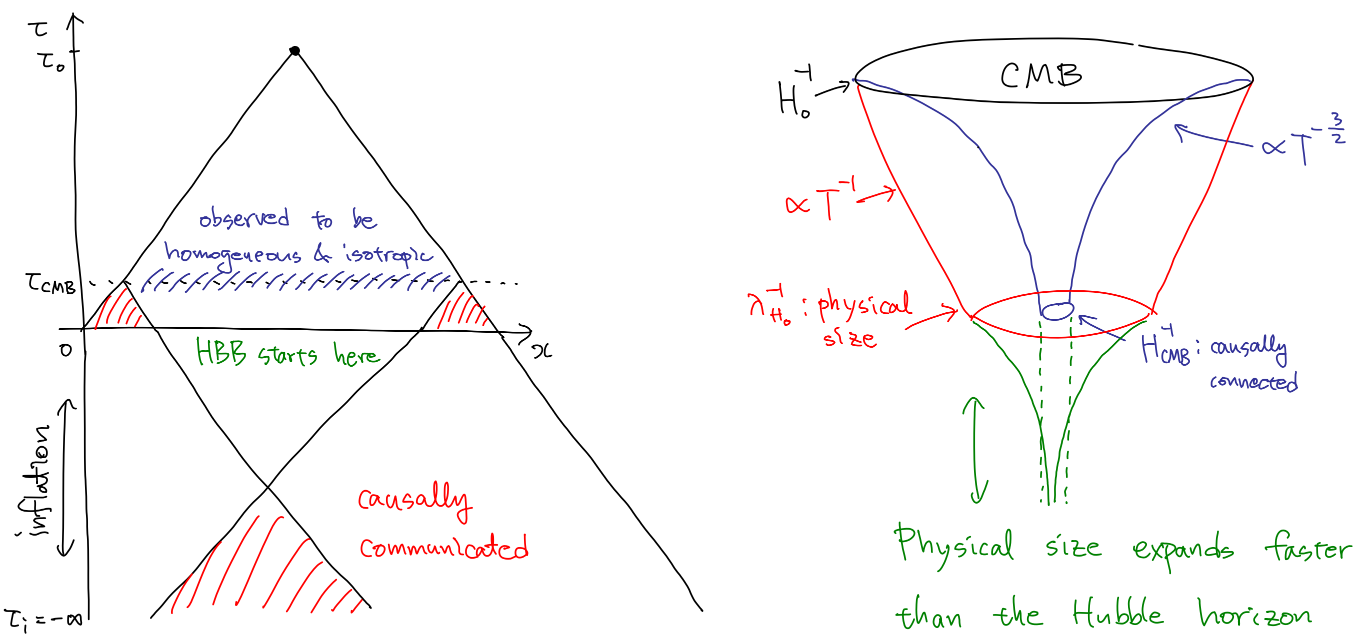


Figure 4: (Left) conformal diagram of the universe, this time including inflation. Inflation extends $\tau$ to $-\infty$, giving ample room for causal communication well before the onset of hot big bang evolution at $\tau = 0$. (Right) Inflation corresponds to the period when the physical size $\lambda \sim a$ expands faster than the Hubble horizon $1/H$.

As is clear from Figure 4, the longer inflation last, the larger the overlapping region becomes. Thus we need a certain duration of of inflation to explain the homogeneous CMB. The amount of inflation is quantified by the number of $e$-folds $N$ between some initial ($i$) and final ($f$) moments, which is given by

$$N = \int_i^f H dt = \int_i^f \frac{da}{a} = \log\left(\frac{a_f}{a_i}\right) . \tag{46}$$

Thus, with a given $N$, the final scale factor is related to the initial scale factor by $a_f = a_i e^N$, i.e. the universe has expanded by $e^N$ times. Now we can compute how large $N$ should be for the CMB. The most natural way is that at the beginning of inflation (or the part of inflation relevant for our observable universe) the physical length scale $\lambda_{H_0^{-1}}$ is smaller than the Hubble horizon during inflation $H_I$ so that causal communication has been established within $\lambda_{H_0^{-1}}$ to have the same temperature. This gives

$$\lambda_{H_0^{-1}} = H_0^{-1}\frac{a_i}{a_0} = H_0^{-1}\frac{a_f}{a_0}\frac{a_i}{a_f} = H_0^{-1}\frac{T_0}{T_f}e^{-N} < H_I^{-1} . \tag{47}$$

Thus, solving for $N$ from the last inequality, we obtain

$$N > \log\left(\frac{T_0}{H_0}\right) - \log\left(\frac{T_f}{H_I}\right) \sim 67 - \log\left(\frac{T_f}{H_I}\right) , \tag{48}$$

where we have used $H_0 \sim 10^{-42}$GeV and $T_0 \sim 10^{-13}$GeV. Thus, assuming that the logarithmic term which includes two unknown factors give a number of $\mathcal{O}(1)$, we require that

$$N \gtrsim 60 . \tag{49}$$

That is, to explain the homogeneity of the CMB, i.e. to solve the horizon problem, we need 60 $e$-folds of expansion: during inflation the universe should have expanded by $e^{60} \sim 10^{26}$ times.

### 1.3.3 Slow-roll inflation

The cosmological constant is obviously the simplest candidate that drives inflation, but the problem is that if this is the case, inflation never ends and we cannot recover the universe in which we leave with stars, galaxies, clusters of galaxies and so on. Thus, we need some different material which can mimic the cosmological constant and at the same time provide a "graceful exit" from inflation. This is usually achieved by a scalar field $\phi$. For simplicity here we assume that this scalar field, named "inflaton" in the sense that it drives inflation, is minimally coupled to gravity and has canonical kinetic term. Then the action is the sum of the gravitational sector, which we take the Einstein-Hilbert action, and the matter sector:

$$S = \int d^4x\sqrt{-g}\frac{m_{\rm Pl}^2}{2}R + \int d^4x\sqrt{-g}\left[\underbrace{-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)}_{\equiv\mathcal{L}_m}\right] . \tag{50}$$

The corresponding energy-momentum tensor $T_{\mu\nu}$ of $\phi$ can be obtained by perturbing the matter Lagrangian with respect to $g^{\mu\nu}$,

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta\left(\sqrt{-g}\mathcal{L}_m\right)}{\delta g^{\mu\nu}} = g_{\mu\nu}\mathcal{L}_m - 2\frac{\delta\mathcal{L}_m}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi + V(\phi)\right] . \tag{51}$$

Then we can easily compute 00 and $ii$ components which can then be matched to the energy density and pressure respectively [see (20)]:

$$\rho = T^0{}_0 = \frac{1}{2}\dot{\phi}^2 + V\,, \tag{52}$$

$$p = \frac{1}{3}T^i{}_i = \frac{1}{2}\dot{\phi}^2 - V\,. \tag{53}$$

Thus, if potential dominates over the kinetic energy ($\dot{\phi}^2 \ll V$) these simplify to $\rho \approx V \approx -p$, thus the inflaton provides a nearly cosmological constant, leading to an exponential expansion of the universe – inflation!

Let us first write the background equation of motion for $\phi$. From this we can find a number of useful formulae which do not resort to the dynamics of $\phi$ but to $V$ and its derivatives only. The equation of motion for $\phi$ can be found from the Euler-Lagrange equation,

$$\partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right] = \frac{\partial \mathcal{L}}{\partial \phi}\,. \tag{54}$$

This gives

$$-\Box\phi + \frac{\partial V}{\partial \phi} = 0\,, \tag{55}$$

where

$$\Box \equiv \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\right) = -\frac{\partial^2}{\partial t^2} - 3H\frac{\partial}{\partial t} + \frac{\Delta}{a^2}\,, \tag{56}$$

with $\Delta \equiv \delta^{ij}\partial_i\partial_j$ being the spatial Laplacian operator. Here for the last equality we have taken the background metric. Thus, the background field $\phi = \phi(t)$ follows the equation of motion

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial \phi} = 0\,. \tag{57}$$

So how this equation for $\phi$ simplifies? From (24), using (52) and (53) we require

$$\frac{\ddot{a}}{a} = -\frac{\dot{\phi}^2 - V}{3m_{\rm Pl}^2} > 0\,. \tag{58}$$

[Note that we can again precisely find (41) by using (52) and (53) for (21) and (23)] Then this means

$$\dot{\phi}^2 < V\,. \tag{59}$$

Taking a time derivative on both sides, this says $\ddot{\phi} < \partial V/\partial\phi$. Thus, (57) is simplified to

$$3H\dot{\phi} + \frac{\partial V}{\partial \phi} = 0\,. \tag{60}$$

Thus we can replace $\dot{\phi}$, or more generally the dynamics of $\phi$, with the derivatives of the potential $V$.

Then now let us consider the slow-roll parameter $\epsilon$, (41). Applying (60), we find

$$\epsilon = -\frac{\dot{H}}{H^2} \approx \frac{\dot{\phi}^2/(2m_{\rm Pl}^2)}{V/(3m_{\rm Pl}^2)} \approx \frac{3}{2}\frac{1}{V}\frac{V'^2}{9H^2} \approx \frac{m_{\rm Pl}^2}{2}\left(\frac{V'}{V}\right)^2\,, \tag{61}$$

where $V' \equiv \partial V/\partial\phi$. Thus, $\epsilon$ in the slow-roll approximation tells us how steep the potential slope is. We can introduce another important slow-roll parameter $\eta$, which describes how quickly $\epsilon$ evolves:

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} \approx \left[H\frac{m_{\rm Pl}^2}{2}\left(\frac{V'}{V}\right)^2\right]^{-1} m_{\rm Pl}^2\frac{V'}{V}\left[\frac{V''}{V} - \left(\frac{V'}{V}\right)^2\right]\dot{\phi} \approx 2m_{\rm Pl}^2\frac{V''}{V} + 4\epsilon\,. \tag{62}$$

Also note that in the slow-roll approximation the $e$-fold $N$ can be written in terms of the potential solely:

$$N = \int_i^f H dt = \int_i^f H\frac{dt}{d\phi}d\phi = \int_i^f \frac{H}{\dot{\phi}}d\phi \approx \frac{1}{m_{\rm Pl}^2}\int_f^i \frac{V}{V'}d\phi\,. \tag{63}$$

# 2 Perturbation equations

Having discussed about background, now we move to the linear cosmological perturbations in the context of single field inflation. The purpose is to derive the equations of motion for the relevant perturbations.

## 2.1 Perturbed Einstein equation

We first consider the LHS of the Einstein equation. Here the perturbations are in the metric tensor $g_{\mu\nu}$. Including (linear) perturbations in (5), the most general perturbed metric is written as

$$ds^2 = -(1+2A)dt^2 + 2a\mathcal{B}_i dt dx^i + a^2\left[(1+2\varphi)\delta_{ij} + 2\mathcal{E}_{ij}\right] dx^i dx^j \,. \tag{64}$$

The inverse metric can be found by requiring $g^{\mu\rho}g_{\rho\nu} = \delta^\mu{}_\nu$. For example, if we assume the form $g^{00} = -1+\alpha$ and $g^{0i} = \beta^i$ with $\alpha$ and $\beta^i$ being the perturbations,

$$g^{0\mu}g_{\mu 0} = 1 = g^{00}g_{00} + g^{0i}g_{i0} = (-1+\alpha)(-1-2A) + \beta^i a\mathcal{B}_i \approx 1 + 2A - \alpha \,, \tag{65}$$

so that at linear order we can find $\alpha = 2A$. In the similar way, after some calculations, we can find all the components of the inverse metric as

$$\begin{aligned} &\delta g_{00} = -2A\,, \quad \delta g_{0i} = a\mathcal{B}_i\,, \quad \delta g_{ij} = 2a^2\left(\varphi\delta_{ij} + \mathcal{E}_{ij}\right)\,, \\ &\delta g^{00} = 2A\,, \quad \delta g^{0i} = a^{-1}\mathcal{B}^i\,, \quad \delta g^{ij} = 2a^{-2}\left(-\varphi\delta^{ij} - \mathcal{E}^{ij}\right)\,, \end{aligned} \tag{66}$$

where the index of $\mathcal{B}_i$ and $\mathcal{E}_{ij}$ is raised and lowered by $\delta_{ij}$. Then, what's left is a straightforward but a bit tedious calculation, and we obtain

$$\delta G_{00} = 6H\dot\varphi + \frac{2}{a^2}\left\{-\Delta\varphi - aH\mathcal{B}^i{}_{,i} + a^2\left[H\dot{\mathcal{E}}^i{}_i + \frac{1}{2}\left(\mathcal{E}^{ij}{}_{,ij} - \Delta\mathcal{E}^i{}_i\right)\right]\right\}\,, \tag{67}$$

$$\delta G_{0i} = 2\left(-\dot\varphi + HA\right)_{,i} - a\left(2\dot H + 3H^2\right)\mathcal{B}_i + \frac{1}{2a}\left(\mathcal{B}^j{}_{,ij} - \Delta\mathcal{B}_i\right) + \left(\dot{\mathcal{E}}^j{}_{i,j} - \dot{\mathcal{E}}^j{}_{j,i}\right)\,, \tag{68}$$

$$\begin{aligned} \delta G_{ij} = {} & a^2\left(-2\ddot\varphi + 2H\left(\dot A - 3\dot\varphi\right) + 2\left(2\dot H + 3H^2\right)(A-\varphi)\right. \\ & \left. + \frac{1}{a^2}\left\{\Delta(A+\varphi) + a^2\left(\dot{\mathcal{B}}^k + 2H\mathcal{B}^k\right)_{,k} - a^2\left[\ddot{\mathcal{E}}^k{}_k + 3H\dot{\mathcal{E}}^k{}_k + \left(\mathcal{E}^{kl}{}_{,kl} - \Delta\mathcal{E}^k{}_k\right)\right]\right\}\right) \\ & + \left\{-(A+\varphi)_{,ij} - a\left[\frac{\dot{\mathcal{B}}_{i,j} + \dot{\mathcal{B}}_{j,i}}{2} + H\left(\mathcal{B}_{i,j} + \mathcal{B}_{j,i}\right)\right]\right. \\ & \left. + a^2\left[\ddot{\mathcal{E}}_{ij} + 3H\dot{\mathcal{E}}_{ij} - 2\left(2\dot H + 3H^2\right)\mathcal{E}_{ij} + \frac{1}{a^2}\left(\mathcal{E}^k{}_{j,ik} + \mathcal{E}^k{}_{i,jk} - \mathcal{E}^k{}_{k,ij} - \Delta\mathcal{E}_{ij}\right)\right]\right\}\,. \end{aligned} \tag{69}$$

At this point, it is very convenient to decompose the vector and the tensor components of the metric perturbations $\mathcal{B}_i$ and $\mathcal{E}_{ij}$ into pure scalar, transverse vector and transverse and traceless tensor components such as[2]

$$\begin{aligned}\mathcal{B}_i &= B_{,i} + S_i\,, \\ \mathcal{E}_{ij} &= E_{,ij} + \frac{1}{2}\left(F_{i,j} + F_{j,i}\right) + \frac{1}{2}h_{ij}\,,\end{aligned} \tag{70}$$

where the pure vectors $S_i$ and $F_i$ and the pure tensor $h_{ij}$ satisfy

$$S^i{}_{,i} = F^i{}_{,i} = 0\,, \tag{71}$$

$$h^i{}_i = h^i{}_{j,i} = 0\,. \tag{72}$$

Then, initially $S^i$ and $F^i$ have 3 degrees of freedom each, but the transverse consitions remove 1 each so that in the vector perturbations $S^i$ and $F^i$ there are total 4 degrees of freedom. Likewise, while the symmetric 3×3 matrix (or rank-2 tensor) $h_{ij}$ has 6 degrees of freedom, but after applying the transverse and traceless conditions 4 of them are removed, and we are left with 2 degrees of freedom. Note that there are 4 scalar degrees of freedom in the metric, $A$, $B$, $\varphi$ and $E$. These sum up to have total 10 degrees of freedom for the metric perturbations. This is in agreement with the observation that the 4×4 matrix $g_{\mu\nu}$ has 10 independent components.

After applying the decomposition of the scalar, vector and tensor components (70), we can after some calculations obtain each component of the Einstein equation as

$$\delta G^0{}_0 = 6H\left(-\dot{\varphi} + HA\right) - 2\frac{\Delta}{a^2}\left[-\varphi - H\left(aB - a^2\dot{E}\right)\right]\,, \tag{73}$$

$$\delta G^0{}_i = -2\left(-\dot{\varphi} + HA\right)_{,i} + \frac{\Delta}{2a}S_i - \frac{\Delta}{2}\dot{F}_i\,, \tag{74}$$

$$\delta G^i{}_i = 6\left[\frac{d}{dt}\left(-\dot{\varphi} + HA\right) + 3H\left(-\dot{\varphi} + HA\right) + \dot{H}A + \frac{\Delta}{3a^2}D\right]\,, \tag{75}$$

$$\begin{aligned}\delta G^{T\,i}{}_j &= \delta G^i{}_j - \frac{1}{3}\delta^i{}_j \delta G^k{}_k \\ &= -\frac{1}{a^2}\left(\partial^i\partial_j - \delta^i{}_j\frac{\Delta}{3}\right)D \\ &\quad + \left[-\frac{1}{2a}\left(\dot{S}^i + 2HS^i\right) + \frac{1}{2}\left(\ddot{F}^i + 3H\dot{F}^i\right)\right]_{,j} + \left[-\frac{1}{2a}\left(\dot{S}_j + 2HS_j\right) + \frac{1}{2}\left(\ddot{F}_j + 3H\dot{F}_j\right)\right]^{,i} \\ &\quad + \frac{1}{2}\left(\ddot{h}^i{}_j + 3H\dot{h}^i{}_j - \Delta h^i{}_j\right)\,,\end{aligned} \tag{76}$$

[2]Note that we may from the beginning include the scale factor in such a way that

$$\begin{aligned}\mathcal{B}_i &= \frac{\widetilde{B}_{,i}}{a} + S_i\,, \\ \mathcal{E}_{ij} &= \frac{\widetilde{E}_{,ij}}{a^2} + \frac{1}{2a}\left(\widetilde{F}_{i,j} + \widetilde{F}_{j,i}\right) + \frac{1}{2}h_{ij}\,.\end{aligned}$$

This will eliminate additional scale factor dependence in the Einstein equation and in that sense more convenient than the above notation. However, many literature adopt the notation without the scale factor dependence so we keep presenting the results in that notation.

where

$$\begin{aligned} D &\equiv (A+\varphi) + a\left(\dot{B}+2HB\right) - a^2\left(\ddot{E}+3H\dot{E}\right) \\ &= (A+\varphi) + \frac{d}{dt}\left(aB - a^2\dot{E}\right) + H\left(aB-a^2\dot{E}\right) \,. \end{aligned} \tag{77}$$

Note that we have decomposed the $ij$ spatial component into trace and traceless parts. When $i=j$, we can trivially see that $\delta {G^{T}}^i{}_j = 0$ as it should be.

For the RHS of the Einstein equation, we have an additional scalar perturbation, that is, the perturbation of the inflaton field $\delta\phi$. From the expression of the energy-momentum tensor of $\phi$ (51), including the metric perturbations we can straightforwardly find each component as

$$\delta T^0{}_0 = -\left[\dot{\phi}\left(\dot{\delta\phi} - \dot{\phi}A\right) + V'\delta\phi\right] \,, \tag{78}$$

$$\delta T^0{}_i = -\dot{\phi}\delta\phi_{,i} \,, \tag{79}$$

$$\delta T^i{}_i = \frac{1}{3}\left[\dot{\phi}\left(\dot{\delta\phi} - \dot{\phi}A\right) - V'\delta\phi\right] \,, \tag{80}$$

$$\delta {T^{T}}^i{}_j = 0 \,. \tag{81}$$

There is an important remark at this point. As we can see, at linear order the scalar, vector and tensor components of the perturbations are all *decoupled*. Thus we can consider each component independent of the others.

Now we can write each component of the scalar, vector and tensor perturbations separately. They are given by

$$\begin{aligned} \text{00 component: } & 3H\left(-\dot{\varphi}+HA\right) - \frac{\Delta}{a^2}\left[-\varphi - H\left(aB-a^2\dot{E}\right)\right] \\ & = -\frac{1}{2m_{\rm Pl}^2}\left[\dot{\phi}\left(\dot{\delta\phi}-\dot{\phi}A\right) + V'\delta\phi\right] \,, \end{aligned} \tag{82}$$

$$\text{Scalar } 0i \text{ component: } -\dot{\varphi} + HA = \frac{1}{2m_{\rm Pl}^2}\dot{\phi}\delta\phi \,, \tag{83}$$

$$\text{Vector } 0i \text{ component: } S_i - a\dot{F}_i = 0 \,, \tag{84}$$

$$\begin{aligned} \text{Trace } ij \text{ component: } & \frac{d}{dt}\left(-\dot{\varphi}+HA\right) + 3H\left(-\dot{\varphi}+HA\right) + \dot{H}A + \frac{\Delta}{3a^2}D \\ & = \frac{1}{2m_{\rm Pl}^2}\left[\dot{\phi}\left(\dot{\delta\phi}-\dot{\phi}A\right) - V'\delta\phi\right] \,, \end{aligned} \tag{85}$$

$$\text{Traceless scalar } ij \text{ component: } D = 0 \,, \tag{86}$$

$$\text{Traceless vector } ij \text{ component: } \frac{d}{dt}\left(S^i - a\dot{F}^i\right) + 2H\left(S^i - a\dot{F}^i\right) = 0 \,, \tag{87}$$

$$\text{Traceless tensor } ij \text{ component: } \ddot{h}_{ij} + 3H\dot{h}_{ij} - \frac{\Delta}{a^2}h_{ij} = 0 \,. \tag{88}$$

Thus, we can find an important fact: if inflation is driven by a single (canonical) inflaton field, there exists only scalar component so that the vector and tensor metric perturbations are unsourced. Also there exist no anisotropic stress either, which gives $D=0$.

## 2.2 Gauge transformations

Before we begin the discussion on the scalar perturbations, we consider the issue of "background" and "perturbation". In the background universe $U$, there is no ambiguity in choosing the time coordinate on the homogeneous and isotricpic spatial hypersurfaces in such a way that time is constant: $t = t_1$ corresponds to the moment when the homogeneous scalar field has a specific value of $\phi(t = t_1)$, and so on. However, in a perturbed universe $\widehat{U}$, our choice of time is arbitrary in the sense that we can choose arbitrary coordinate system where the deviation from homogeneity and isotropy is small. In different coordinate systems, the notion of perturbations is different too. For example, we can choose spatial hypersurfaces on which the density perturbation vanishes. Thus, just saying that the density perturbation is such and such is not enough. We have to also specify the coordinate system in describing the density perturbation.

Let us consider in a more detail. How can we define the perturbation in a scalar quantity $\widehat{\phi}$ at a point $p$ in the perturbed universe $\widehat{U}$? To define the perturbation, we need to specify the corresponding background value $\phi_0$: the difference between $\widehat{\phi}$ and $\phi_0$ is the perturbation $\delta\phi(p)$. But what is the corresponding background $\phi_0$? For this, we have to specify a coordinate system, or *mapping* in such a way that each point in the perturbed universe $\widehat{U}$ is associated with the corresponding point $x^\mu$ in the background universe $U$. Once this mapping is specified, the perturbation

$$\delta\phi(p) = \widehat{\phi}(p) - \phi_0(x^\mu) \tag{89}$$

is meaningful.

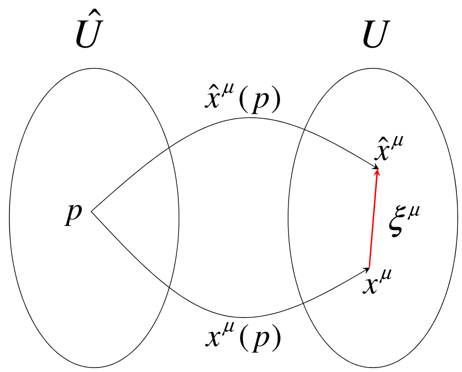


Figure 5: A schematic image of gauge transformation.

So to specify perturbations we only need to specify the coordinate system, or the mapping between $\widehat{U}$ and $U$. The problem is, as stated before, there is no natural choice of this mapping and one is as good (or bad) as the others. Thus we need to know how one mapping is related to another. It's very important to note that any change induced by a change in the mapping is *not* physical: it is simply a transformation because we have changed the coordinate, or "gauge", to describe the same thing. In this sense, this non-physical change is called *gauge transformation*. Suppose 2 coordinate systems, $x^\mu$ and $\hat{x}^\mu$, which map $p$ in $\widehat{U}$ to the corresponding *different* points in $U$, are related by

$$\widehat{x}^\mu(p) = x^\mu(p) + \xi^\mu(x^\nu(p)) \,. \tag{90}$$

If this transformation is infinitesimal, we have

$$\begin{aligned}\widehat{\delta\phi}(p) &= \widehat{\phi}(p) - \phi_0(\hat{x}^\mu(p)) \\ &= \delta\phi(p) - [\phi_0(\hat{x}^\mu(p)) - \phi_0(x^\mu(p))] \\ &= \delta\phi(p) - \xi^\nu \frac{\partial\phi_0}{\partial x^\nu}(x^\mu(p)) \,. \end{aligned} \tag{91}$$

From below, we drop the subscript 0 to denote the background quantities. Since the background universe $U$ is spatially homogeneous and isotropic, we simply have

$$\widehat{\delta\phi}(p) = \delta\phi(p) - \dot{\phi}(t(p))\xi^0(x^\mu(p)) \,, \tag{92}$$

where we have taken $x^0 = t$. It is very important to note that we are comparing 2 different mappings, $x^\mu(p)$ and $\hat{x}^\mu(p)$, from the *same* point $p$ in $\widehat{U}$. It is schematically shown in Fig. 5.

Note that we can extract the gauge transformations of the metric perturbations by requiring that $ds^2$ be invariant under the gauge transformation[3]: with the coordinate transformations

$$t \to \widehat{t} = t + \xi^0(t, \mathbf{x}) \,, \tag{93}$$

$$x^i \to \widehat{x^i} = x^i + \xi^i(t, \mathbf{x}) \,, \tag{94}$$

we can easily see that in the linear order

$$a\left(\hat{t}\right) \equiv \widehat{a} = \left(1 + H\xi^0\right) a(t) \,, \tag{95}$$

$$\widehat{dt} = \left(1 + \dot{\xi}^0\right) dt + \xi^0_{,i} dx^i \,, \tag{96}$$

$$\widehat{dx^i} = dx^i + \dot{\xi}^i dt + \xi^i{}_{,j} dx^j \,. \tag{97}$$

From the fact that the line element in space-time is the same irrespective of the coordinate transformation, we can write

$$\begin{aligned}\widehat{ds}^2 &= -\left(1 + 2\widehat{A}\right)\widehat{dt}^2 + 2\widehat{a}\widehat{\mathcal{B}}_i\widehat{dt}\widehat{dx^i} + \widehat{a}^2\left[\left(1 + 2\widehat{\varphi}\right)\delta_{ij} + 2\widehat{\mathcal{E}}_{ij}\right]\widehat{dx^i}\widehat{dx^j} \\ &= -\left[1 + 2\left(\widehat{A} + \dot{\xi}^0\right)\right]dt^2 + 2a\left(\widehat{\mathcal{B}}_i - \frac{\xi^0{}_{,i}}{a} + a\dot{\xi}_i\right)dtdx^i \\ &\quad + a^2\left\{\left[1 + 2\left(\widehat{\varphi} + H\xi^0\right)\right]\delta_{ij} + 2\left(\widehat{\mathcal{E}}_{ij} + \frac{\xi_{i,j} + \xi_{j,i}}{2}\right)\right\}dx^idx^j \,, \end{aligned} \tag{98}$$

where $\xi_i = \delta_{ij}\xi^j$. Equating this expression with (64), we can find that under the coordinate transformation given by (93) and (94), the new metric perturbations are given by

$$\widehat{A} = A - \dot{\xi}^0 \,, \tag{99}$$

$$\widehat{\mathcal{B}}_i = \mathcal{B}_i + \frac{\xi^0_{,i}}{a} - a\dot{\xi}_i \,, \tag{100}$$

$$\widehat{\varphi} = \varphi - H\xi^0 \,, \tag{101}$$

$$\widehat{\mathcal{E}}_{ij} = \mathcal{E}_{ij} - \frac{\xi_{i,j} + \xi_{j,i}}{2} \,. \tag{102}$$

[3]Generic gauge transformation law for an arbitrary tensor can be written in terms of the Lie derivatives, but we do not consider this approach here but follow simpler one.

Further, we can decompose the spatial gauge transformation vector $\xi^i$ into the scalar and transverse vector components as we did for the metric perturbation,

$$\xi^i = \delta^{ij}\xi_{,j} + \xi^{(\mathrm{tr})i}\,, \tag{103}$$

where $\xi^{(\mathrm{tr})i}{}_{,i} = 0$. Then, we can find trivially that the scalar, vector and tensor components of $\mathcal{B}_i$ and $\mathcal{E}_{ij}$ transform as

$$\widehat{B} = B + \frac{\xi^0}{a} - a\dot{\xi}\,, \tag{104}$$

$$\widehat{S}_i = S_i - a\dot{\xi}_i^{(\mathrm{tr})}\,, \tag{105}$$

$$\widehat{E} = E - \xi\,, \tag{106}$$

$$\widehat{F}_i = F_i - \xi_i^{(\mathrm{tr})}\,, \tag{107}$$

$$\widehat{h}_{ij} = h_{ij}\,. \tag{108}$$

Notice that the tensor perturbation $h_{ij}$ as well as the combination $S_i - a\dot{F}_i$ remain the same under the gauge transformation, i.e. it is *gauge invariant*. Thus, when we consider the vector and tensor perturbations, we need not worry about the gauge ambiguity because the variables we are dealing with are from the beginning gauge invariant. Gauge ambiguity only matters for scalar perturbations, and we will explicitly discuss this issue in the following section.

Also, it is fruitful to consider the gauge transformation property of the scalar components of the Einstein equation more closely before we write the relevant equation for the scalar perturbations. As we can see, they include all the metric perturbations, but $B$ and $E$ only appear in the specific combination $aB - a^2\dot{E}$: see (77) and (82). Now, from (104) and (106), we can see that

$$a\widehat{B} - a^2\dot{\widehat{E}} = a\left(B + \frac{\xi^0}{a} - a\dot{\xi}\right) - a^2\frac{d}{dt}(E-\xi) = aB - a^2\dot{E} + \xi^0\,, \tag{109}$$

so that although the transformations of $B$ and $E$ include the spatial component of the gauge transformation $\xi$, in practice only $\xi^0$, the time translation matters.

## 2.3 Scalar perturbations

There are 2 strategies one could take to deal with this gauge ambiguity in scalar perturbations. First, one could fix the gauge by choosing the perturbation in some physical quantity of interest to vanish. For example, one could choose the perturbation in $\widehat{\phi}$ to vanish. Temporarily treating $\xi^0$ as a finite gauge transformation, and working to 1st order in perturbation variable, from (92)

$$\widehat{\delta\phi}(p) = 0 = \delta\phi(p) - \dot{\phi}(t(p))\xi^0(x^\mu(p))\,, \tag{110}$$

and $\xi^0$ is fixed to be

$$\xi^0 = \frac{\delta\phi}{\dot{\phi}} \equiv \xi^0_{\delta\phi}\,. \tag{111}$$

This is the time translation from an arbitrary hypersurface to the one where $\delta\phi$ vanishes, with $\delta\phi$ evaluated at that arbitrary hypersurface: this gauge fixing of $\xi^0$ fixes the constant time hypersurfaces in $\widehat{U}$ to be constant $\widehat{\phi}$ hypersurfaces with the time parametrization $t(p)$ fixed by $\phi(t(p)) = \widehat{\phi}(p)$.

Alternatively, one could just use gauge invariant quantities like $h_{ij}$ and the combination $S_i - a\dot{F}_i$. For example, let us consider

$$\mathcal{R} = \varphi - \frac{H}{\dot{\phi}}\delta\phi \,, \tag{112}$$

which is gauge invariant:

$$\begin{aligned}\widehat{\mathcal{R}} &= \widehat{\varphi} - \frac{H}{\dot{\phi}}\widehat{\delta\phi} \\ &= \left(\varphi - H\xi^0\right) - \frac{H}{\dot{\phi}}\left(\delta\phi - \xi^0\dot{\phi}\right) \\ &= \varphi - \frac{H}{\dot{\phi}}\delta\phi = \mathcal{R} \,.\end{aligned} \tag{113}$$

The physical interpretation of this gauge invariant quantity is clear: if we insert (111), we immediately find

$$\mathcal{R} = \varphi - H\xi^0_{\delta\phi} \equiv \varphi_{\delta\phi} \,, \tag{114}$$

i.e. $\mathcal{R}$ is the perturbation in the spatial curvature $\varphi$ on the hypersurfaces where $\delta\phi = 0$. When the universe is dominated by $\phi$, we can see from (79) that $T^0{}_i = 0$ so we do not see momentum flux. In this sense, this gauge condition is called "comoving", and correspondingly $\mathcal{R}$ is called *comoving curvature perturbation*. Note that $\mathcal{R}$ can be evaluated on arbitrary hypersurfaces as it is independent of gauge.

Now we find the equation of motion for $\mathcal{R}$ in 3 different ways. The reason why we consider $\mathcal{R}$, not other scalar perturbations, is twofold and will be obvious soon. Let us just proceed at the moment.

1. First we choose a different gauge and find the equation for $\mathcal{R}$. This is possible since $\mathcal{R}$ is gauge invariant. Let us choose the so-called Newtonian gauge, or sometimes called zero shear gauge. In this gauge, we choose

   $$B = E = 0 \,. \tag{115}$$

   Note that from (106) $E = 0$ fixes the spatial gauge $\xi$, and in turn also fixes the temporal gauge $\xi^0$. The reason why it is called zero shear gauge is that in this gauge $aB - a^2\dot{E} = 0$ [which from (109) fixes $\xi^0$] which is identified as the *shear*. Also, from (77) and (86) we can set

   $$A = -\varphi \equiv \Phi \,. \tag{116}$$

   Then the metric is written as

   $$ds^2 = -(1+2\Phi)dt^2 + a^2(1-2\Phi)\delta_{ij}dx^i dx^j \,, \tag{117}$$

   which coincides with the weak field limit where the Newtonian gravitatoinal potential is identified as the perturbation in the 00 component. This is the reason why it is called

"Newtonian" gauge. Also note that until this stage, we have eliminated 3 scalar degrees of freedom.

Now, the scalar 00, $0i$ and trace $ij$ equations are respectively reduced to

$$3H\left(\dot{\Phi}+3H\Phi\right)-\frac{\Delta}{a^2}\Phi=-\frac{1}{2m_{\rm Pl}^2}\left[\dot{\phi}\left(\dot{\delta\phi}-\dot{\phi}\Phi\right)+V'\delta\phi\right]\,, \tag{118}$$

$$\dot{\Phi}+3H\Phi=\frac{1}{2m_{\rm Pl}^2}\dot{\phi}\delta\phi\,, \tag{119}$$

$$\frac{d}{dt}\left(\dot{\Phi}+3H\Phi\right)+3H\left(\dot{\Phi}+3H\Phi\right)+\dot{H}\Phi=\frac{1}{2m_{\rm Pl}^2}\left[\dot{\phi}\left(\dot{\delta\phi}-\dot{\phi}\Phi\right)-V'\delta\phi\right]\,. \tag{120}$$

The remaining 2 degrees of freedom are $\Phi$ and $\delta\phi$, and we have 3 equations. Thus 1 of these equations are redundant and we can use only 2 of them to solve $\Phi$ and $\delta\phi$. We use 00 and $0i$ components, since they are simpler. Note that $\mathcal{R}$ is written as, with the RHS of (112) evaluated on this gauge,

$$\mathcal{R}=-\Phi-\frac{H}{\dot{\phi}}\delta\phi\,. \tag{121}$$

From 00 equation (118),

$$\begin{aligned}
3H\left(\dot{\Phi}+3H\Phi\right)-\frac{\Delta}{a^2}\Phi &= -\frac{1}{2m_{\rm Pl}^2}\left[\dot{\phi}\dot{\delta\phi}-\dot{\phi}^2\Phi-\left(\ddot{\phi}+3H\dot{\phi}\right)\delta\phi\right] \\
&= -\frac{1}{2m_{\rm Pl}^2}\left(\frac{\dot{\phi}\dot{\delta\phi}-\ddot{\phi}\delta\phi}{\dot{\phi}^2}\dot{\phi}^2-\dot{\phi}^2\Phi-3H\dot{\phi}\delta\phi\right) \\
&= -\frac{\dot{\phi}^2}{2m_{\rm Pl}^2}\left[\frac{d}{dt}\left(\frac{\delta\phi}{\dot{\phi}}\right)-\Phi\right]+3H\left(\dot{\Phi}+3H\Phi\right)\,,
\end{aligned} \tag{122}$$

where for the 1st equality we have used (57) to eliminate $V'$, and for the 2nd equality (119) to eliminate $\dot{\phi}\delta\phi$. Thus, we have

$$\frac{d}{dt}\left(\frac{\delta\phi}{\dot{\phi}}\right)=\Phi-\frac{1}{\dot{H}}\frac{\Delta}{a^2}\Phi\,. \tag{123}$$

Further, taking a derivative for (121) we find

$$\begin{aligned}
\dot{\mathcal{R}} &= -\dot{\Phi}-\dot{H}\frac{\delta\phi}{\dot{\phi}}-H\frac{d}{dt}\left(\frac{\delta\phi}{\dot{\phi}}\right) \\
&= -\dot{\Phi}-\frac{\dot{H}}{\dot{\phi}}2m_{\rm Pl}^2\left(\dot{\Phi}+H\Phi\right)-H\left(\Phi-\frac{1}{\dot{H}}\frac{\Delta}{a^2}\Phi\right) \\
&= \frac{H}{\dot{H}}\frac{\Delta}{a^2}\Phi\,,
\end{aligned} \tag{124}$$

where for the 2nd equality we have used (119) and (123). Thus we can write $\Phi$ in terms of $\mathcal{R}$ as

$$\Phi=\frac{\dot{H}}{H}a^2\Delta^{-1}\dot{\mathcal{R}}\,, \tag{125}$$

where $\Delta^{-1}$ is the inverse Laplacian operator. Finally, we put all these into (119). The LHS becomes, using (125),

$$\dot{\Phi} + H\Phi = a^2 \frac{\dot{H}}{H} \Delta^{-1} \left[ \ddot{\mathcal{R}} + \left( 3H + \frac{2\ddot{\phi}}{\dot{\phi}} - \frac{\dot{H}}{H} \right) \dot{\mathcal{R}} \right] . \tag{126}$$

Meanwhile, the RHS becomes

$$\begin{aligned} \frac{1}{2m_{\text{Pl}}^2} \dot{\phi} \delta\phi &= \frac{\dot{H}}{H} (\mathcal{R} + \Phi) \\ &= \frac{\dot{H}}{H} \left( \mathcal{R} + a^2 \frac{\dot{H}}{H} \Delta^{-1} \dot{\mathcal{R}} \right) \\ &= a^2 \frac{\dot{H}}{H} \Delta^{-1} \left( \frac{\Delta}{a^2} \mathcal{R} + \frac{\dot{H}}{H} \dot{\mathcal{R}} \right) , \end{aligned} \tag{127}$$

where for the 1st equality we have used (121), and for the 2nd equality (125). Thus, equating the LHS and RHS, we find finally the equation of motion for $\mathcal{R}$ as

$$\ddot{\mathcal{R}} + \left( 3H + \frac{2\ddot{\phi}}{\dot{\phi}} - \frac{2\dot{H}}{H} \right) \dot{\mathcal{R}} - \frac{\Delta}{a^2} \mathcal{R} = \frac{1}{a^3 \epsilon} \frac{d}{dt} \left( a^3 \epsilon \dot{\mathcal{R}} \right) - \frac{\Delta}{a^2} \mathcal{R} = 0 . \tag{128}$$

2. Next we work in the comoving gauge from the beginning to write the equation for $\mathcal{R}$. This is as we will soon see much simpler than the 1st approach. But we place this as the 2nd approach, since it has some illuminating aspects for the 3rd approach we will consider next.

   Comoving gauge, as we have already considered, requires $\delta\phi = 0$. This fixes the temporal gauge, thus we need to fix the spatial gauge by imposing $E = 0$. (this is not explicitly stated in some literatures) Writing $\varphi = \mathcal{R}$, the 00, $0i$ and traceless $ij$ components respectively become

$$3H \left( -\dot{\mathcal{R}} + HA \right) + \frac{\Delta}{a^2} (\mathcal{R} + aHB) = \frac{\dot{\phi}^2}{2m_{\text{Pl}}^2} A , \tag{129}$$

$$-\dot{\mathcal{R}} + HA = 0 , \tag{130}$$

$$(A + \mathcal{R}) + \frac{d}{dt}(aB) + aHB = 0 . \tag{131}$$

   Note that the trace $ij$ equation vanishes identically upon imposing (130) and (131). Now, from (130) and then (129), we can solve for $A$ and $B$ as

$$A = \frac{\dot{\mathcal{R}}}{H} , \tag{132}$$

$$aB = -\frac{\mathcal{R}}{H} + a^2 \epsilon \Delta^{-1} \dot{\mathcal{R}} . \tag{133}$$

The fact that 00 and $0i$ equations can be *algebraically* solved in terms of $\mathcal{R}$ is not a mere coincidence. We will return to this point in the next approach. Then, plugging the solutions for $A$ and $B$ into (131), we have

$$\frac{\dot{\mathcal{R}}}{H} + \mathcal{R} + \frac{d}{dt}\left(-\frac{\mathcal{R}}{H} + a^2\epsilon\Delta^{-1}\dot{\mathcal{R}}\right) + H\left(-\frac{\mathcal{R}}{H} + a^2\epsilon\Delta^{-1}\dot{\mathcal{R}}\right) = 0\,. \tag{134}$$

Applying the Laplacian operator $\Delta$ to remove the spurious inverse Laplacian operator $\Delta^{-1}$, we find

$$\frac{d}{dt}\left(a^2\epsilon\dot{\mathcal{R}}\right) + a^2\epsilon H\dot{\mathcal{R}} - \epsilon\Delta\mathcal{R} = a^2\epsilon\left[\frac{1}{a^3\epsilon}\frac{d}{dt}\left(a^3\epsilon\dot{\mathcal{R}}\right) - \frac{\Delta}{a^2}\mathcal{R}\right] = 0\,, \tag{135}$$

so we reach the same equation of motion for $\mathcal{R}$.

3. Finally, we resort to the action approach which will be more directly related to the quantization procedure in the next section. We begin with the action (50), and including the 5 scalar perturbations we write the action quadratic in these perturbations. This is because the linear equation of motion follows from the quadratic action. The steps to write the quadratic action are straightforward but a bit tedious. Those interested are strongly recommended to refer to the seminal review by Mukhanov, Feldman & Brandenberger (1992), Section 10 there, for the detailed steps. We start from the obtained quadratic action, in the conformal time,

$$\begin{aligned} S_2^{(s)} = \int d^4x \frac{m_{\text{Pl}}^2}{2}a^2 \Big[ & -6\varphi'^2 + 12\mathcal{H}A\varphi' - 2\left(\mathcal{H}' + 2\mathcal{H}^2\right)A^2 - 2(2A+\varphi)\delta\phi \\ & + \frac{1}{m_{\text{Pl}}^2}\left(\delta\phi'^2 + \delta\phi\Delta\delta\phi - a^2V_{\phi\phi}\delta\phi^2\right) + \frac{2}{m_{\text{Pl}}^2}\left(-3\phi'\varphi'\delta\phi - \phi'A\delta\phi' - a^2V_\phi A\delta\phi\right) \\ & + 4\left(\frac{1}{2m_{\text{Pl}}^2}\phi'\delta\phi + \varphi' - \mathcal{H}A\right)\Delta(B-E')\Big]\,, \end{aligned} \tag{136}$$

where a prime denotes a derivative with respective to the conformal time (instead we write $\partial V/\partial\phi \equiv V_\phi$) and $\mathcal{H} \equiv a'/a$. As we have noted before, at this order the scalar perturbations are not mixed with vector and tensor ones and we can separately consider them at linear order. We can note that $\varphi$, $\delta\phi$ and $E$ have time derivatives, while $A$ and $B$ not. Hence $A$ and $B$ do not give dynamical evolution, and their equations of motion are *constraints* which can be solved at any time and then can be plugged back into the action, since they are always satisfied. That is, after some arrangement we should be able to write $\varphi$, $\delta\phi$ and $E$ in the canonical form with the Hamiltonian $\mathfrak{H}^{(s)}$ while $A$ and $B$ are multiplied by their equations of motion, viz. constraints,

$$\mathcal{L}_2^{(s)} = \Pi_\varphi\varphi' + \Pi_{\delta\phi}\delta\phi' + \Pi_E E' - \mathfrak{H}^{(s)} - \mathcal{C}_A A - \mathcal{C}_B B\,. \tag{137}$$

Thus we can understand why 00 and $0i$ equations in the previous approach were solved algebraically: it is because of the structure of the Lagrangian. By construction 00 and $0i$ equations are constraints[4].

[4]This point becomes more transparent if the metric is written in the so-called Arnowitt-Deser-Misner form, which in this lecture is not covered.

We can easily find the conjugate momentum of $\varphi$, $\delta\phi$ and $E$ as

$$\Pi_\varphi \equiv \frac{\delta}{\delta\phi'}\mathcal{L}_2^{(s)} = \frac{m_{\rm Pl}^2}{2}a^2\left[-12\varphi' + 12\mathcal{H}A - \frac{6}{m_{\rm Pl}^2}\phi'\delta\phi + 4\Delta(B-E')\right], \tag{138}$$

$$\Pi_{\delta\phi} \equiv \frac{\delta}{\delta(\delta\phi')}\mathcal{L}_2^{(s)} = a^2\left(\delta\phi' - \phi' A\right), \tag{139}$$

$$\Pi_E \equiv \frac{\delta}{\delta E'}\mathcal{L}_2^{(s)} = \frac{m_{\rm Pl}^2}{2}a^2\Delta\left(-4\varphi' + 4\mathcal{H}A - \frac{2}{m_{\rm Pl}^2}\phi'\delta\phi\right). \tag{140}$$

Combining $\Pi_\varphi$ and $\Pi_E$ we can write

$$\Delta(B-E') = \frac{1}{2m_{\rm Pl}^2a^2}\left(\Pi_\varphi - 3\Delta^{-1}\Pi_E\right), \tag{141}$$

and from $\Pi_{\delta\phi}$ we can write $\delta\phi'$ as

$$\delta\phi' = \frac{\Pi_{\delta\phi}}{a^2} + \phi' A\,. \tag{142}$$

Then, after some straightforward but tedious calculations, we find

$$\begin{aligned}
\mathcal{L}_2^{(s)} = {} & \Pi_\varphi\varphi' + \Pi_{\delta\phi}\delta\phi' + \Pi_E E' \\
& - \left\{\frac{1}{2a^2m_{\rm Pl}^2}\left[-\Pi_\varphi\Delta^{-2}\Pi_E + \frac{3}{2}\left(\Delta^{-1}\Pi_E\right)^2 + m_{\rm Pl}^2\Pi_{\delta\phi}^2\right] - \frac{1}{2m_{\rm Pl}^2}\phi'\Pi_\varphi\delta\phi \right.\\
& \left. + a^2m_{\rm Pl}^2\left[\varphi\Delta\varphi - \frac{3}{4m_{\rm Pl}^2}\phi'^2\delta\phi^2 - \frac{1}{2m_{\rm Pl}^2}\left(\delta\phi\Delta\delta\phi - a^2V_{\phi\phi}\delta\phi^2\right)\right]\right\} \\
& - \left[\mathcal{H}\Pi_\varphi + \phi'\Pi_{\delta\phi} + 2a^2m_{\rm Pl}^2\Delta\varphi + a^2\left(3\mathcal{H}\phi' + a^2V_\phi\right)\delta\phi\right]A - \Pi_E B\,.
\end{aligned} \tag{143}$$

This Lagrangian is of the form (137). The equations of motion for $A$ and $B$ are simply the constraints $\mathcal{C}_A = 0$ and $\mathcal{C}_B = 0$. From $\mathcal{C}_B = 0$, $E$ disappears, and from $\mathcal{C}_A = 0$, $\Pi_{\delta\phi}$ is written in terms of $\Pi_\varphi$, $\varphi$ and $\delta\phi$ (or $\Pi_\varphi$ can be replaced as well).

After plugging back the solutions of the constraints and rearrangement, the quadratic Lagrangian becomes

$$\begin{aligned}
\mathcal{L}_2^{(s)} = {} & \left(\Pi_\varphi + \frac{2a^2m_{\rm Pl}^2}{\phi'}\Delta\delta\phi\right)\left(\varphi - \mathcal{H}\frac{\delta\phi}{\phi'}\right)' - \frac{2m_{\rm Pl}^2\mathcal{H}}{\phi'^2}\left(\Pi_\varphi + \frac{2a^2m_{\rm Pl}^2}{\phi'}\Delta\delta\phi\right)\Delta\left(\varphi - \mathcal{H}\frac{\delta\phi}{\phi'}\right) \\
& - \frac{\mathcal{H}^2}{2a^2\phi'^2}\left(\Pi_\varphi + \frac{2a^2m_{\rm Pl}^2}{\phi'}\Delta\delta\phi\right)^2 - \frac{2a^2m_{\rm Pl}^2}{\phi'^2}\left[\Delta\left(\varphi - \mathcal{H}\frac{\delta\phi}{\phi'}\right)\right]^2 \\
& - a^2m_{\rm Pl}^2\left(\varphi - \mathcal{H}\frac{\delta\phi}{\phi'}\right)\Delta\left(\varphi - \mathcal{H}\frac{\delta\phi}{\phi'}\right).
\end{aligned} \tag{144}$$

We can redefine another set of canonical variables that combine $\Pi_\varphi$ and $\varphi$ with $\delta\phi$ as

$$\mathcal{R} \equiv \varphi - \frac{\mathcal{H}}{\phi'}\delta\phi\,, \tag{145}$$

$$\Pi_\mathcal{R} \equiv \Pi_\varphi + \frac{2a^2m_{\rm Pl}^2}{\phi'}\Delta\delta\phi\,, \tag{146}$$

and we finally find the quadratic Lagrangian without any constraint

$$\mathcal{L}_2^{(s)} = \Pi_{\mathcal{R}}\mathcal{R}' - \underbrace{\left[ \frac{2a^2m_{\rm Pl}^2}{\phi'^2}\left(\Delta\mathcal{R} + \frac{\mathcal{H}}{2a^2m_{\rm Pl}^2}\right)^2 + a^2m_{\rm Pl}^2\mathcal{R}\Delta\mathcal{R}\right]}_{\equiv \mathfrak{H}_{\mathcal{R}}} . \tag{147}$$

This Lagrangian has no constraint and is of canonical form, thus now we are left with a single physical variable – the comoving curvature perturbation! We can write this in a more well-known form by eliminating $\Pi_{\mathcal{R}}$ in favour of $\mathcal{R}'$. From the Hamiltonian equation of motion,

$$\frac{\delta\mathfrak{H}_{\mathcal{R}}}{\delta\Pi_{\mathcal{R}}} = \mathcal{R}' , \tag{148}$$

we can find

$$\frac{\mathcal{H}}{2a^2m_{\rm Pl}^2}\Pi_{\mathcal{R}} = \frac{\phi'^2}{2m_{\rm Pl}^2\mathcal{H}}\mathcal{R}' - \Delta\mathcal{R} . \tag{149}$$

Then we trivially find

$$\mathcal{L}_2^{(s)} = \frac{1}{2}\left(\frac{a\phi'}{\mathcal{H}}\right)^2\left[\mathcal{R}'^2 - (\nabla\mathcal{R})^2\right] . \tag{150}$$

At this point, we return to the cosmic time then we can immediately find that

$$S_2^{(s)} = \int d\tau d^3x \frac{1}{2}\left(\frac{a\phi'}{\mathcal{H}}\right)^2\left[\mathcal{R}'^2 - (\nabla\mathcal{R})^2\right] = \int dtd^3xa^3\epsilon m_{\rm Pl}^2\left[\dot{\mathcal{R}}^2 - \frac{(\nabla\mathcal{R})^2}{a^2}\right] , \tag{151}$$

and the equation that follows from this action is

$$\frac{1}{a^3\epsilon}\frac{d}{dt}\left(a^3\epsilon\dot{\mathcal{R}}\right) - \frac{\Delta}{a^2}\mathcal{R} = 0 , \tag{152}$$

thus again we recover the same equation of motion for $\mathcal{R}$!

So it's a good moment to come up with 1 reason why we consider $\mathcal{R}$. As we have seen, initially we begin with 5 scalar perturbations: $A$, $B$, $\varphi$, $E$ and $\delta\phi$. But eventually we can eliminate 4 of them and are left with only 1 single physical degree of freedom: in the 1st approach, we eliminate $B$ and $E$, and $A$ and $\varphi$ are the same, and $\delta\phi$ can be replaced by using the Einstein equation. In the 2nd approach, $E$ and $\delta\phi$ are set to be zero, and $A$ and $B$ are solved in terms of $\varphi = \mathcal{R}$. In the 3rd approach, $A$ and $B$ are solved and $E$ is found to be vanishing, and we can eliminate $\delta\phi$ in favour of $\varphi$ or vice versa. In fact this has deeper reason. In general relativity, we have 2 scalar gauge transformation functions, $\xi^0$ and $\xi$, as well as 2 constraint equations. Thus, solving each eliminates 1 degree of freedom, so we are after all left with a single degree of freedom: $\mathcal{R}$. Of course this gives the reason why we *can* work with $\mathcal{R}$, but not why we *should* (or are highly recommended to) work with $\mathcal{R}$, not with any other perturbation variable. This reason will be clear in the next section by solving the equation of motion for $\mathcal{R}$ on very large scales.

## 2.4 Vector and tensor perturbations

### 2.4.1 Vector perturbations

The vector equation (84) simply says there is no vector perturbation! Even if $S^i - a\dot{F}^i \equiv X^i \neq 0$ initially, from (87)

$$\dot{X}_i + 2HX_i = 0\,. \tag{153}$$

This immediately gives a solution

$$X_i \propto a^{-2}\,. \tag{154}$$

During inflation $a \sim e^{Ht}$, even if initially non-zero, vector perturbation decays exponentially. This is why usually vector perturbation is neglected in the context of inflation[5].

Let us now be more formal to look into the vector perturbations. As we did for the scalar perturbations in the previous section, we can find the 2nd order action for the vector perturbations after expanding the action up to 2nd order. The result is

$$S_2^{(v)} = \int d^4x a^2 m_{\rm Pl}^2 \left(S^i - F^{i\prime}\right)^{,j} \left(S_i - F_i'\right)_{,j}\,, \tag{155}$$

where $F^{i\prime} = a\dot{F}^i$. Thus, only $F^i$ has time derivative and the associated conjugate momentum exists,

$$\Pi^i \equiv \frac{\delta}{\delta F_i'}\mathcal{L}_2^{(v)} = 2a^2 m_{\rm Pl}^2 \Delta \left(S^i - F^{i\prime}\right)\,. \tag{156}$$

Then the Lagrangian is written as

$$\mathcal{L}_2^{(v)} = \Pi^i F_i' - \mathfrak{H}^{(v)} - S_i\Pi^i\,, \tag{157}$$

$$\mathfrak{H}^{(v)} = -\frac{1}{a^2 m_{\rm Pl}^2}\Pi_i \Delta^{-1}\Pi^i\,. \tag{158}$$

Thus, equation of motion for $S_i$, i.e. the constraint gives $\Pi^i = 0$. Plugging this solution back into the action, we find simply $\mathcal{L}_2^{(v)} = 0$. Thus there is no relevant vector perturbation during inflation driven by a single inflaton field.

### 2.4.2 Tensor perturbations

In fact, we have already found the relevant equation of motion for the tensor perturbations: (88), and this is all! We may just close this section here and proceed to solve the scalar and tensor perturbation equations, but nevertheless let us spend some time to see if we have already extracted all we need for tensor perturbations without worrying about gauge issues. From perturbing the action, we can find the tensor 2nd order action as

$$S_2^{(t)} = \int d^4x \frac{a^2 m_{\rm Pl}^2}{8}\left(h^{ij\prime}h_{ij}' + h^{ij}\Delta h_{ij}\right)\,. \tag{159}$$

[5] However, this is not always the case if there exists background vector field, such as the case of vector inflation.

The conjugate momentum is

$$\Pi^{ij} \equiv \frac{\delta}{\delta h'_{ij}} \mathcal{L}_2^{(t)} = \frac{a^2 m_{\rm Pl}^2}{4} h^{ij\prime} \,, \tag{160}$$

and the action is simply

$$S_2^{(t)} = \int d^4x \left( \Pi^{ij} h'_{ij} - \mathfrak{H}^{(t)} \right) \,, \tag{161}$$

$$\mathfrak{H}^{(t)} = \frac{2}{a^2 m_{\rm Pl}^2} \Pi^{ij} \Pi_{ij} - \frac{a^2 m_{\rm Pl}^2}{8} h^{ij} \Delta h_{ij} \,. \tag{162}$$

There is no constraint, so $h_{ij}$ is directly relevant. As mentioned shortly after the beginning of this section, note that after imposing the transverse and traceless conditions, there are 2 independent degrees of freedom for $h_{ij}$. These are usually called the 2 "polarization states" of the gravitational waves. If we consider the Hamiltonian equation of motion from $\mathfrak{H}^{(t)}$

$$\frac{\delta \mathfrak{H}^{(t)}}{\delta h^{ij}} = -\Pi'_{ij} \,, \tag{163}$$

we can easily obtain (88) as expected. The other Hamiltonian equation of motion, $\delta\mathfrak{H}^{(t)}/\delta\Pi^{ij} = h'_{ij}$, is a trivial identity.

# 3 Power spectra of perturbations

Finally we are now able to proceed to quantize the perturbations of our interest, $\mathcal{R}$ and $h_{ij}$, and compute their power spectra. We first consider the curvature perturbation.

## 3.1 Quantization for the curvature perturbation

### 3.1.1 Asymptotic solutions

Our starting point is the quadratic action (150). By introducing

$$z \equiv \frac{a\phi'}{\mathcal{H}}\,, \tag{164}$$

$$u \equiv z\mathcal{R} = a\left(\delta\phi - \frac{\phi'}{\mathcal{H}}\varphi\right)\,, \tag{165}$$

after partial integrations the action becomes

$$S_2 = \int d^4x \frac{1}{2}\left[u'^2 - (\nabla u)^2 + \frac{z''}{z}u^2\right]\,. \tag{166}$$

Thus, the resulting equation of motion for $u$ is

$$u'' - \Delta u - \frac{z''}{z}u = 0\,. \tag{167}$$

We can write the Fourier mode $u(\tau, \boldsymbol{k})$, which will be more convenient for the subsequent study, as

$$u(\tau, \boldsymbol{x}) = \int \frac{d^3k}{(2\pi)^3}e^{i\boldsymbol{k}\cdot\boldsymbol{x}}u(\tau, \boldsymbol{k})\,, \tag{168}$$

the equation becomes [from now on $u$ is the Fourier mode unless specified, $u = u(\tau, \boldsymbol{k})$]

$$u'' + \left(k^2 - \frac{z''}{z}\right)u = 0\,. \tag{169}$$

Let us consider the equation (169) more closely. Upon the identification

$$k^2 - \frac{z''}{z} \equiv \omega_k^2(\tau)\,, \tag{170}$$

(169) describes a harmonic oscillator with time dependent frequency $\omega_k(\tau)$. Let us consider the frequency in more detail. With (164), we have the *exact* expression

$$\frac{z''}{z} = 2a^2H^2\left(1 - \frac{\epsilon}{2} + \frac{3}{4}\eta - \frac{\epsilon\eta}{4} + \frac{\eta^2}{8} + \frac{\dot{\eta}}{4H}\right) \equiv \frac{1}{\tau^2}\left(\nu^2 - \frac{1}{4}\right)\,, \tag{171}$$

where in the last expression $\nu$ is defined by

$$\nu^2 = \frac{9}{4} + 3\epsilon + \frac{3}{2}\eta + \cdots\,. \tag{172}$$

Thus, for a given $k$, we can think of 2 exreme cases: either $k^2$ is much more dominant than $z''/z \sim (aH)^2$ in $\omega_k^2$, or the other way round. The equation is simplified to

$$u'' + \left(k^2 - \frac{z''}{z}\right) u \longrightarrow \begin{cases} u'' + k^2 u = 0 & \text{for } k \gg aH \,, \\ u'' - \dfrac{z''}{z} u = 0 & \text{for } k \ll aH \,. \end{cases} \tag{173}$$

For $k \gg aH$, i.e. when a mode with typical length scale $\lambda \sim 1/k$ is much smaller than the comoving Hubble horizon $1/(aH)$ so that it is deep inside the horizon ("sub-horizon"), the mode function behaves like a plane wave, $u \sim e^{\pm ik\tau}$. Meanwhile, when the mode is far outside the horizon, i.e. on super-horizon scales, we have a simple solution $u \propto z$. This means

$$\mathcal{R} = \frac{u}{z} \sim \text{constant} \,, \tag{174}$$

viz. the comoving curvature perturbation well outside the horizon is frozen and its value is *conserved*. This is another important reason why we consider $\mathcal{R}$: it is conserved on very large scales once it exits the horizon during inflation, until it enters the horizon after inflation. This is a very nice property of $\mathcal{R}$ and many other perturbation variables, such as $\delta\phi$, continue evolution even on super-horizon scales.

### 3.1.2 Canonical quantization

Now we return to the actio (166). Considering the Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ and the effective mass $m_{\text{eff}}^2 \equiv -z''/z$, we can rewrite the action as

$$S_2 = \int d^4x \left( -\frac{1}{2}\eta^{\mu\nu}\partial_\mu u \partial_\nu u - \frac{1}{2} m_{\text{eff}}^2 u^2 \right) \,. \tag{175}$$

This form of the action is identical to that of a 1) free and 2) canonical scalar field in the Minkowski space, thus the quantization procedure is standard. That is, we promote $u$ and the conjugate momentum $\Pi_u = \delta\mathcal{L}/\delta u' = u'$ to operators $\widehat{u}$ and $\widehat{\Pi}_u$ and imposes the canonical commutation relation between them.

Before we proceed, one comment is in order. One may worry that the effective mass $m_{\text{eff}}^2$ is intrinsically negative, and even worse, becomes indefinitely large as inflation proceeds ($\tau \to 0$). Does this mean any pathology, since it is a badly behaving tachyon? The answer is no: the real physical quantity of our interest is the comoving curvature perturbation $\mathcal{R}$. In late time, where $m_{\text{eff}}^2$ is negatively diverging, $\mathcal{R}$ is perfectly well behaved as (174).

1. Since $u$ is a *free* field, we can expand the operator $\widehat{u}$ in terms of the creation and annihilation operators in the Fourier space. That is, the Fourier mode given by (168) is promoted to the operator $\widehat{u}(\tau, \boldsymbol{k})$, which can be expanded in terms of the creation and annihilation operators

$$\widehat{u}(\tau, \boldsymbol{k}) = a_{\boldsymbol{k}} u_k(\tau) + a_{-\boldsymbol{k}}^\dagger u_k^*(\tau) \,, \tag{176}$$

where the creation and annihilation operators satisfy the standard commutation relation

$$\left[a_{\boldsymbol{k}}, a_{\boldsymbol{q}}^\dagger\right] = (2\pi)^3 \delta^{(3)}(\boldsymbol{k} - \boldsymbol{q}) \,, \tag{177}$$

otherwise zero.

2. Now we require that the canonical conjugate variables $\widehat{u}$ and $\widehat{\Pi}_u$ satisfy the equal time canonical commutation relation,

$$\left[\widehat{u}(\tau,\boldsymbol{x}),\widehat{\Pi}_u(\tau,\boldsymbol{y})\right] = i\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})\,. \tag{178}$$

Using the Fourier mode (168) and the expansion (176) with the relation (177),

$$\begin{aligned}\left[\widehat{u}(\tau,\boldsymbol{x}),\widehat{\Pi}_u(\tau,\boldsymbol{y})\right] &= \int\frac{d^3k}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}e^{i\boldsymbol{k}\cdot\boldsymbol{x}}e^{i\boldsymbol{q}\cdot\boldsymbol{y}}\left\{\left[a_{\boldsymbol{k}},a^\dagger_{-\boldsymbol{q}}\right]u_k\frac{du_q^*}{d\tau}-\left[a_{\boldsymbol{q}},a^\dagger_{-\boldsymbol{k}}\right]\frac{du_q}{d\tau}u_k^*\right.\\ &\qquad\left.+\left[a_{\boldsymbol{k}},a_{\boldsymbol{q}}\right]u_k\frac{du_q}{d\tau}+\left[a^\dagger_{-\boldsymbol{k}},a^\dagger_{-\boldsymbol{q}}\right]u_k^*\frac{du_q^*}{d\tau}\right\}\\ &= \int\frac{d^3k}{(2\pi)^3}e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})}\left(u_k\frac{du_k^*}{d\tau}-\frac{du_k}{d\tau}u_k^*\right)\,,\end{aligned} \tag{179}$$

which should match the delta function. Thus, the mode function $u_k$ satisfies the normalization condition

$$u_k\frac{du_k^*}{d\tau}-\frac{du_k}{d\tau}u_k^* = i\,. \tag{180}$$

### 3.1.3 Vacuum state

In the previous section we have set up the canonical commutation relations for the operators. But we need to determine the mode function $u(\tau,\boldsymbol{k})$, which amounts to fix the vacuum state $|0\rangle$ defined by

$$a_{\boldsymbol{k}}|0\rangle = 0 \quad \text{for all } k\,. \tag{181}$$

In the Minkowski space, the vacuum state is such that the Hamiltonian operator of the system is minimized. In fact in our case we have only 1 single sensible situation to do so: when $k \gg aH$, as can be seen from (173) the frequency is time-independent. Thus we can straightly apply the standard procedure to find the mode function solution, i.e. the vacuum state. The Lagrangian in this limit, say $\tau=\tau_0$, is approximated by

$$\mathcal{L} = \frac{1}{2}\left[u'^2-(\nabla u)^2\right]\,, \tag{182}$$

which gives the Hamiltonian operator

$$\begin{aligned}\widehat{\mathfrak{H}} &= \int d^3x\frac{1}{2}\left[\widehat{\Pi}_u^2+(\nabla\widehat{u})^2\right]\\ &= \int\frac{d^3k}{(2\pi)^3}\left\{a_{\boldsymbol{k}}a_{-\boldsymbol{k}}\left(\widehat{u}_k'^{\,2}+k^2\widehat{u}_k^2\right)+c.c.+\left[2a^\dagger_{\boldsymbol{k}}a_{\boldsymbol{k}}+(2\pi)^3\delta^{(3)}(0)\right]\left(|\widehat{u}_k'|^2+k^2|\widehat{u}_k|^2\right)\right\}\,,\end{aligned} \tag{183}$$

where for the 2nd equality we have used (176). Evaluating the expectation value of $\widehat{\mathfrak{H}}$ with respect to the vacuum state $|0\rangle_0$, we find

$${}_0\langle 0|\widehat{\mathfrak{H}}|0\rangle_0 = \frac{1}{2}\int d^3k\delta^{(3)}(0)\left(|\widehat{u}_k'|^2+k^2|\widehat{u}_k|^2\right)\,. \tag{184}$$

Thus our task is to find the mode function $u_k$ that minimizes this expression. Since we already know the solution is a plane wave, let us assume that $u_k$ takes the form

$$u_k = \psi_k e^{i\theta_k} \,, \tag{185}$$

with $\psi_k$ and $\theta_k$ being real without loss of generality. Also we assume $\psi_k$ is constant, since the maximum amplitude in the Minkowski space is preserved. First, from the normalization condition for $u_k$ (180), we find

$$-2i\psi_k^2\theta_k' = i \,. \tag{186}$$

Then,

$$|\widehat{u}_k'|^2 + k^2|\widehat{u}_k|^2 = {\psi_k'}^2 + k^2\psi_k^2 + \psi_k^2{\theta_k'}^2 = {\psi_k'}^2 + k^2\psi_k^2 + \frac{1}{4\psi_k^2} \,, \tag{187}$$

where we have used (186) for the 2nd equality. Thus, the constant $\psi_k$ that minimizes the above expression is

$$\psi_k = \frac{1}{\sqrt{2k}} \,. \tag{188}$$

Then, from (186) we can determine $\theta_k$ as

$$\theta_k = -k\tau \,, \tag{189}$$

where without loss of generality we have dropped the integration constant. Thus, the mode function solution is

$$u_k = \frac{1}{\sqrt{2k}} e^{-ik\tau} \,, \tag{190}$$

which corresponds to the vacuum state with the frequency $\omega_k = k$. In fact, we can see that this is exactly the solution of a massless scalar field, $E_k = \omega_k = k$. Moreover, using (190) the Hamiltonian (183) is written as

$$\widehat{\mathfrak{H}} = \int \frac{d^3k}{(2\pi)^3} \left[ a_{\boldsymbol{k}}^\dagger a_{\boldsymbol{k}} + \frac{1}{2}(2\pi)^3\delta^{(3)}(0) \right] \omega_k \,, \tag{191}$$

which is precisely that of a harmonic oscillator! (barring the factors coming from the Fourier mode convention and infinite spatial volume)

This mode function solution, or the vacuum state, however, does not remain as the solution (or vacuum state) all the time. Remember that the mode function solution (190) is found when the frequency is simply $k$. In general, as we can see from (170), the frequency is time dependent. Thus, the Hamiltonian operator is time-dependent and in turn the mode function solution (or the vacuum state) which minimizes the Hamiltonian is no longer the same. Let us write the Fourier mode expansion at some later time $\tau_1 > \tau_0$ in terms of a new set of creation and annihilation operators as well as new mode function,

$$\widehat{u}(\tau, \boldsymbol{k}) = b_{\boldsymbol{k}} v_k(\tau) + b_{-\boldsymbol{k}}^\dagger v_k^*(\tau) \,, \tag{192}$$

and we can define a new vacuum state as

$$b_{\boldsymbol{k}}|0\rangle_1 = 0 \,. \tag{193}$$

In general, the new mode function $v_k$ is related to another mode function $u_k$ via a linear transformation, the so-called Bogoliubov transformation,

$$v_k = \alpha_k u_k + \beta_k u_k^* \,. \tag{194}$$

Note that (192) is also the solution of the equation (169), provided that the complex coefficient $\alpha_k$ and $\beta_k$ is normalized to

$$|\alpha_k|^2 - |\beta_k|^2 = 1 \tag{195}$$

to satisfy (180), given that $u_k$ is a solution. That is, in general the vacuum state is time-dependent,

$$|0\rangle_0 \neq |0\rangle_1 \,. \tag{196}$$

What this tells us is: the notion of vacuum state is dependent on time and there is no *unique* vacuum state throughout all the time.

This has a profound consequence. Let us consider that at $\tau_1 > \tau_0$ we can expand the Fourier mode of the rescaled curvature perturbation $u(\tau, \boldsymbol{k})$ as (192), with the mode function $v_k$ being related to the one at $\tau_0$, $u_k$, by (194). If we evaluate the expectation value of the number operator $N_k^{(b)} \equiv b_{\boldsymbol{k}}^\dagger b_{\boldsymbol{k}}$ with respect to $|0\rangle_0$, the vacuum at $\tau_0$, we find

$${}_0\langle 0|N_k^{(b)}|0\rangle_0 = {}_0\langle 0| \left( \alpha_k a_{\boldsymbol{k}}^\dagger - \beta_k a_{-\boldsymbol{k}} \right) \left( \alpha_k^* a_{\boldsymbol{k}} - \beta_k^* a_{-\boldsymbol{k}}^\dagger \right) |0\rangle_0 = (2\pi)^3 |\beta_k|^2 \delta^{(3)}(0) \,, \tag{197}$$

where we have used the commutation relation (177). That is, even if we have started with a vacuum state $|0\rangle_0$ which contains no particle at an initial time $\tau_0$, at a later time $\tau_1$ we find that $|0\rangle_0$ contains a non-vanishing number of $b$-particles. That is, *we have something out of nothing*. This is how quantum fluctuations are generated in the gravitational background.

## 3.2 General solution for the curvature perturbation

Now we can write the general solution of the mode function: the solution satisfies the equation (169), the normalization condition (180), and the boundary condition (190). The general solution can be written in terms of the Bessel functions, and for later convenience we use the Hankel functoin,

$$u_k(\tau) = \sqrt{-\tau} \left[ c_1(k) H_\nu^{(1)}(-k\tau) + c_2(k) H_\nu^{(2)}(-k\tau) \right] \,, \tag{198}$$

where $c_1(k)$ and $c_2(k)$ are coefficients to be determined. Also note that writing the solution, we have assumed $\nu$ to be constant.

To fix the coefficients, we require that on sub-horizon limit $k \gg aH$ we recover the usual Minkowski massless vacuum solution (190). This can be found by taking the argument of the Hankel function $z \equiv -k\tau \approx k/(aH)$ very large,

$$H_\nu^{(1)}(z) \underset{z\gg 1}{\longrightarrow} \sqrt{\frac{2}{\pi z}} e^{i(z - \pi\nu/2 - \pi/4)} \,, \tag{199}$$

with $H_\nu^{(2)}$ being the complex conjugate of $H_\nu^{(1)}$. Thus, to match (190), we can write $c_1(k)$ and $c_2(k)$ as

$$c_1(k) = \frac{\sqrt{\pi}}{2} e^{i(\nu+1/2)\pi/2} \,, \tag{200}$$

$$c_2(k) = 0 \,. \tag{201}$$

One can easily chech that with these coefficients, by taking the limit $-k\tau \gg 1$ we can reproduce (190).

A particularly important and simple case is when $\nu = 3/2$ exactly. As can be read from (172), this corresponds to $\epsilon = 0$, i.e. the perfect de Sitter case. In this case

$$\frac{z''}{z} = 2a^2H^2\,, \tag{202}$$

$$\tau = \frac{-1}{aH}\,, \tag{203}$$

and the Hankel function is given by

$$H^{(1)}_{3/2}(x) = -\sqrt{\frac{2}{\pi z}}\left(1+\frac{i}{x}\right)e^{ix}\,. \tag{204}$$

Then we can find the mode function solution as

$$u_k(\tau) = \frac{1}{\sqrt{2k}}\left(1-\frac{i}{k\tau}\right)e^{-ik\tau}\,. \tag{205}$$

### 3.3 Power spectrum of the curvature perturbation

Power spectrum is the Fourier transformation of the 2-point correlation function. In terms of the Fourier mode, power spectrum of the comoving curvature perturbation is defined by

$$\langle\mathcal{R}(\boldsymbol{k})\mathcal{R}(\boldsymbol{q})\rangle \equiv (2\pi)^3\delta^{(3)}(\boldsymbol{k}+\boldsymbol{q})P_{\mathcal{R}}(k)\,. \tag{206}$$

Here the average is interpreted as the vacuum expectation value with respect to the initial vacuum. Also note that this power spectrum $P_{\mathcal{R}}(k)$ is not dimensionless but has the mass dimension $-3$. It is custumary to define the dimensionless power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ by

$$\mathcal{P}_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2}P_{\mathcal{R}}(k)\,. \tag{207}$$

Using $u = z\mathcal{R}$ and (176), and matching the definition (206), we can find

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2}\left|\frac{u_k}{z}\right|^2\,. \tag{208}$$

Here we note that $\mathcal{R}(\boldsymbol{k}) = \mathcal{R}(\tau,\boldsymbol{k})$ as is obvious from the equation of motion for $\mathcal{R}$ derived in the previous section. Thus the natural question is: when do we evaluate the power spectrum? In fact the answer is very obvious. We have seen in (174) that $\mathcal{R}$ becomes constant on the super-horizon scales $k \ll aH$, i.e. $-k\tau \to 0$, and maintains the value until the moment of horizon entry. Thus it is natural to evaluate the power spectrum at that moment. In this limit, the Hankel function is approximated to

$$H^{(1)}_\nu(z) \underset{z\ll 1}{\longrightarrow} \sqrt{\frac{2}{\pi}}e^{-i\pi/2}2^{\nu-3/2}\frac{\Gamma(\nu)}{\Gamma(3/2)}z^{-\nu}\,. \tag{209}$$

Thus, the power spectrum is written as

$$
\begin{aligned}
\mathcal{P}_{\mathcal{R}}(k) &= \lim_{-k\tau\to 0} \frac{k^3}{2\pi^2}\left|\frac{u_k}{z}\right|^2 \\
&= \lim_{-k\tau\to 0} 2^{2\nu-3}\left[\frac{\Gamma(\nu)}{\Gamma(3/2)}\right]^2 (1+\epsilon)^{1-2\nu}\left(\frac{H}{2\pi}\right)^2\left(\frac{H}{\dot{\phi}}\right)^2\left(\frac{k}{aH}\right)^{3-2\nu} .
\end{aligned}
\tag{210}
$$

With $\nu = 3/2 + \epsilon + \eta/2 + \cdots$, we may expand the coefficients to find

$$
\mathcal{P}_{\mathcal{R}}(k) = \lim_{-k\tau\to 0} [1 + 2(\alpha - 1)\epsilon + \alpha\eta]\left(\frac{H}{2\pi}\right)^2\left(\frac{H}{\dot{\phi}}\right)^2\left(\frac{k}{aH}\right)^{3-2\nu} , \tag{211}
$$

where

$$
\alpha \equiv 2 - \log 2 - \gamma \approx 0.729637 . \tag{212}
$$

Practically, we can evaluate the RHS of (211) at any time around horizon crossing. For definiteness, we evaluate it at horizon crossing $k = aH$, then

$$
\mathcal{P}_{\mathcal{R}}(k) = [1 + 2(\alpha - 1)\epsilon + \alpha\eta]\left(\frac{H}{2\pi}\right)^2\left(\frac{H}{\dot{\phi}}\right)^2\Bigg|_{k=aH} . \tag{213}
$$

An important property of $\mathcal{P}_{\mathcal{R}}(k)$ is how it scales with $k$. Assuming a simple power-law form, it has the form

$$
\mathcal{P}_{\mathcal{R}} \propto k^{n_{\mathcal{R}}-1} , \tag{214}
$$

where $n_{\mathcal{R}}$ is called the spectral index of the power spectrum. It is straightly read from (211), as

$$
n_{\mathcal{R}} - 1 \equiv \frac{d\log\mathcal{P}_{\mathcal{R}}}{d\log k} = 3 - 2\nu = -2\epsilon - \eta|_{k=aH} . \tag{215}
$$

The amplitude of the power spectrum and spectral index are very well constrained by most recent observations on the CMB as $\mathcal{P}_{\mathcal{R}} \sim 2.5 \times 10^{-9}$ and $n_{\mathcal{R}} \sim 0.96$.

## 3.4 Tensor perturbations

Now we consider the tensor perturbations. Before we proceed more rigorous discussions, we can quickly see the tensor perturbations also become conserved on super-horizon scales: if we neglect the spatial gradient in (88), we immediately find that one of the solution is a constant. The other solution can be found, by regarding $\dot{h}_{ij}$ as a variable, that $\dot{h}_{ij} \propto a^{-3}$. Thus,

$$
h_{ij} \sim \begin{cases} \text{constant} , \\ \int a^{-3}dt \sim \int e^{-3Ht}dt . \end{cases} \tag{216}
$$

The 2nd solution is exponentially decaying, hence the constant mode quickly becomes dominant. In other words, on the super-horizon scales, the amplitude of the tensor perturbation remains constant. Thus, once it is primordially produced, it maintains more or less the same magnitude throughout the history of the universe.

Now we study the tensor perturbations more closely. Our starting point is the tensor quardatic action (159). But as we did for the curvature perturbation, we can rescale $h_{ij}$ to obtain the action for canonically normalized fields. For this, in the Fourier mode we introduce the polarization tensor $e_{ij}(\boldsymbol{k},\lambda)$ with $\lambda$ denoting 2 different polarization states in such a way that

$$h_{ij}(\tau,\boldsymbol{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \sum_{\lambda=1}^{2} \psi_\lambda(\tau,\boldsymbol{k}) e_{ij}(\boldsymbol{k},\lambda)\,, \tag{217}$$

where $e_{ij}$ satisfies the following properties:

$$e_{ij} = e_{ji}\,, \tag{218}$$

$$e^i{}_i = 0\,, \tag{219}$$

$$k^i e_{ij} = 0\,, \tag{220}$$

$$e^*_{ij}(\lambda) e^{ij}(\lambda') = 2\delta_{\lambda\lambda'}\,, \tag{221}$$

$$e_{ij}(-\boldsymbol{k}) = e^*_{ij}(\boldsymbol{k})\,, \tag{222}$$

which represent the symmetry between the spatial indices, traceless and transverseness, the existence of 2 independent polarizations and the realness of $h_{ij}$. Then, the action becomes

$$S_2^{(t)} = \int d\tau \frac{a^2 m_{\rm Pl}^2}{4} \int \frac{d^3k}{(2\pi)^3} \sum_\lambda \left( {\psi'_\lambda}^2 - k^2 \psi_\lambda^2 \right) . \tag{223}$$

Further, by introducing

$$v_\lambda \equiv \frac{a m_{\rm Pl}}{\sqrt{2}} \psi_\lambda\,, \tag{224}$$

we have

$$S_2^{(t)} = \sum_\lambda \int d\tau \frac{d^3k}{(2\pi)^3} \frac{1}{2} \left( {v'_\lambda}^2 - k^2 v_\lambda^2 + \frac{a''}{a} v_\lambda^2 \right) . \tag{225}$$

Notice that the only 2 differences from the case of $\mathcal{R}$ are 1) there are 2 copies of the identical action of a canonically normalized scalar field $v_\lambda$ for each polarization state, and 2) the effective mass of $v_\lambda$ contains not $z = a\phi'/\mathcal{H}$ but $a''/a$, which is much simpler than that of the curvature perturbation. The following equation of motion for each $\lambda$ is

$$v''_\lambda + \left( k^2 - \frac{a''}{a} \right) v_\lambda = 0\,. \tag{226}$$

Now we can follow virtually identical steps. Thus we do not show all the explicit calculational details except that because the effective mass contains not $z''/z$ but $a''/a$, the index of the Hankel function solution becomes different. More explicitly, we can write

$$\frac{a''}{a} = 2a^2H^2 \left( 1 - \frac{\epsilon}{2} \right) \equiv \frac{1}{\tau^2} \left( \mu^2 - \frac{1}{4} \right) , \tag{227}$$

where as before the 1st equality is exact, and we have defined

$$\mu^2 = \frac{9}{4} + 3\epsilon + \cdots\,. \tag{228}$$

Thus the properly normalized solution for $v_\lambda$ to match the Bunch-Davies vacuum state is

$$v_\lambda(\tau) = \frac{\sqrt{\pi}}{2} e^{i(\mu+1/2)\pi/2} \sqrt{-\tau} H_\mu^{(1)}(-k\tau) \,. \tag{229}$$

The power spectrum is defined by the sum of each polarization mode,

$$\sum_\lambda \langle \psi_\lambda(\boldsymbol{k}) \psi_\lambda(\boldsymbol{q}) \rangle \equiv (2\pi)^3 \delta^{(3)}(\boldsymbol{k}+\boldsymbol{q}) P_T(k) \,, \tag{230}$$

but for the dimensionless power spectrum $\mathcal{P}_T(k)$ it is conventional to define it without the factor 1/2, i.e.

$$\mathcal{P}_T(k) \equiv \frac{k^2}{\pi^2} P_T(k) = \frac{k^2}{\pi^2} \sum_\lambda \left| \frac{\sqrt{2} v_\lambda}{a m_{\text{Pl}}} \right|^2 \,. \tag{231}$$

From (229), we can easily find

$$\lim_{-k\tau \to 0} \left| \frac{\sqrt{2} v_\lambda}{a m_{\text{Pl}}} \right|^2 = \frac{1}{k^3} \frac{H^2}{m_{\text{Pl}}^2} \left( \frac{k}{aH} \right)^{3-2\mu} 2^{2\mu-3} \left[ \frac{\Gamma(\mu)}{\Gamma(3/2)} \right]^2 \,. \tag{232}$$

Thus, the dimensionless power spectrum becomes[6]

$$\begin{aligned} \mathcal{P}_T(k) &= \lim_{-k\tau \to 0} \left[ 1 + 2(\alpha - 1)\epsilon \right] \frac{8}{m_{\text{Pl}}^2} \left( \frac{H}{2\pi} \right)^2 \left( \frac{k}{aH} \right)^{3-2\mu} \\ &= \left[ 1 + 2(\alpha - 1)\epsilon \right] \frac{8}{m_{\text{Pl}}^2} \left( \frac{H}{2\pi} \right)^2 \Bigg|_{k=aH} \,. \end{aligned} \tag{237}$$

This power spectrum is assumed to have a power-law form as

$$\mathcal{P}_T \propto k^{n_T} \,, \tag{238}$$

---

[6] In many literatures, the tensor perturbation is introduced in the metric with the factor 2, i.e. writing the tensor contributions only,

$$ds^2 = -dt^2 + a^2 \left( \delta_{ij} + 2h_{ij} \right) dx^i dx^j \,. \tag{233}$$

In this case, the tensor quadratic action has of course an overall factor of not 1/8 but 1/2,

$$S_2^{(t)} = \int d^4x \frac{a^2 m_{\text{Pl}}^2}{2} \left( h^{ij\prime} h_{ij}^\prime + h^{ij} \Delta h_{ij} \right) \,. \tag{234}$$

To compensate this factor in the subsequent process of quantization in terms of the canonically normalized scalar field, the polarization tensor is normalized as

$$e_{ij}^*(\lambda) e^{ij}(\lambda') = \delta_{\lambda\lambda'} \,, \tag{235}$$

and $v_\lambda$ is defined by

$$v_\lambda \equiv a m_{\text{Pl}} \psi_\lambda \,. \tag{236}$$

Then after all we end up with the identical quadratic action for $v_\lambda$, (225). The final power spectrum of $h_{ij}$ is now multiplied by the factor 1/4 correspondingly. However, the spectral index is the same.

where note that we conventionally do not include $-1$ here. So the corresponding spectral index is given by

$$n_T \equiv \frac{d\log\mathcal{P}_T}{d\log k} = 3 - 2\mu = -2\epsilon|_{k=aH} \,. \tag{239}$$

Before we proceed, let us take a look at (237) to see why the primordial tensor perturbations are important. As we can see, the power spectrum of the primordial tensor perturbations is directly proportional to $H^2$, which is by the Friedmann equation (21) directly proportional to the energy density. This is not surprising, since the tensor perturbations are after all gravity which only cares the total energy in the universe, irrespective of the model-dependent detailed dynamics during inflation. Thus, by detecting the power spectrum of the tensor perturbations we can directly determine the energy scale during inflation!

Another important quantity related to the tensor perturbations is the so-called tensor-to-scalar ratio. As the name stands, it is the ratio of the tensor power spectrum to the scalar one and denoted by $r$,

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_\mathcal{R}} \,. \tag{240}$$

And using (213) and (237), to leading order in the slow-roll paramters we find

$$r = 16\epsilon = -8n_T \,. \tag{241}$$

This relation is valid for *any* single field inflation model with canonical kinetic term, so it is called a *consistency* relation. Thus if we are lucky enough to test this relation, that amounts to testing *all* canonical single field inflation models at one shot. There is another profound meaning. If we write (241) in terms of the derivative with respect to the number of $e$-folds $N$ using $dN = Hdt$, which follows from (46), we find

$$r = \frac{8}{m_{\rm Pl}^2}\left(\frac{d\phi}{dN}\right)^2 \,. \tag{242}$$

If we limit our interest to the regime relevant for the large scale CMB observations, which spans $N_1 < N < N_2$, we can recast this equation as

$$\frac{\Delta\phi}{m_{\rm Pl}} = \int_{N_1}^{N_2} dN\sqrt{\frac{r}{8}} \,, \tag{243}$$

where $\Delta\phi$ is the field range $\phi$ excurses during $\Delta N \equiv N_2 - N_1 = \mathcal{O}(1)$. As we can see from the spectral indices for both scalar and tensor perturbations, during slow-roll inflation $r$ does not change appreciably for a small interval of $\Delta N$. Thus, we may pull $r$ out of the integral to obtain

$$\frac{\Delta\phi}{m_{\rm Pl}} \sim \left(\frac{r}{0.01}\right)^{1/2} \,. \tag{244}$$

That is, if we ever detect a large tensor-to-scalar ratio $r \gtrsim 0.01$, that means the field excursion is super-Planckian. This seemingly trivial relation in fact raises an important question in inflation model building as we will see in the next section.

## 3.5 Simple example and beyond

### 3.5.1 Quadratic potential case

Finally, let us consider a simple example where the canonical inflaton field has the simple quadratic potential

$$V(\phi) = \frac{1}{2}m^2\phi^2 \,. \tag{245}$$

The derivatives of the potential are simply read $V' = m^2\phi$ and $V'' = m^2$. The first quantity to compute is the number of $e$-folds $N$, that is, to check whether we can have 60 $e$-folds or not. Using the slow-roll approximation, from (63)

$$N = \frac{\phi_i^2 - \phi_f^2}{4m_{\rm Pl}^2} = 60 \,. \tag{246}$$

The final value $\phi_f$ can be found by requiring $\epsilon(\phi_f) = 1$, i.e. at $\phi_f$ inflation stops. This gives

$$\phi_f = \sqrt{2}m_{\rm Pl} \,. \tag{247}$$

This is very small in (246), thus ignoring this contribution for simplicity we find the initial value $\phi_i$ to have 60 $e$-folds as

$$\phi_i = \sqrt{240}m_{\rm Pl} \sim 15m_{\rm Pl} \,. \tag{248}$$

Now we proceed to compute the amplitude of the scalar power spectrum and the spectral index. In (213), for simplicity we neglect the slow-roll terms in the coefficients. Then, using (61)

$$\mathcal{P}_\mathcal{R} = \frac{V^3}{12\pi^2 m_{\rm Pl}^6 V'^2} = \frac{m^2\phi_i^4}{96\pi^2 m_{\rm Pl}^6} \sim \left(10\frac{m}{m_{\rm Pl}}\right)^2 \,. \tag{249}$$

Thus, we can constrain the effective mass of the inflaton as

$$m \sim 5\times 10^{-6} m_{\rm Pl} \sim 10^{13}{\rm GeV} \,. \tag{250}$$

The spectral index can be similarly found using (62) as

$$n_\mathcal{R} = 1 - 8\left(\frac{m_{\rm Pl}}{\phi_i}\right)^2 \approx 0.96 \,. \tag{251}$$

Note that using (246) we can find a simple relation

$$n_\mathcal{R} = 1 - \frac{2}{N} \,. \tag{252}$$

This simple relation, i.e. $n_\mathcal{R} - 1 \propto 1/N$ is common to the inflation model with power-law potential. For the tensor perturbations we can find

$$\mathcal{P}_T = \frac{2V}{3\pi^2 m_{\rm Pl}^4} = \frac{1}{3\pi^2}\left(\frac{m}{m_{\rm Pl}}\right)^2\left(\frac{\phi_i}{m_{\rm Pl}}\right)^2 \sim 2\times 10^{-10} \,, \tag{253}$$

$$n_T = -4\frac{m_{\rm Pl}^2}{\phi_i^2} = -\frac{1}{N} \sim -0.017 \,, \tag{254}$$

$$r = \frac{8}{N} \sim 0.1 \,. \tag{255}$$

### 3.5.2 Digression: effective theory

The simple harmonic oscillator potential we have considered in the previous section surprisingly satisfy almost all the observed constraints deduced from the most recent CMB observations. This class of model, where during inflation $\phi$ spans a large field range compared to $m_{\rm Pl}$, is called "large field model". From (244), such a model typically gives a large tensor-to-scalar ratio as we have checked in the previous section. In other words, to be consistent with the current observations on the CMB, we must ensure that over a field excursion range greater than $m_{\rm Pl}$, both the monomial potential and the canonical kinetic sector describe the dynamics well. But is it true?

For this, let us return to a very basic wisdom. In reality, the universe spans a huge range of various scales: we can think of, for example, the size of an atom and that of a galaxy, or the mass of an electron and that of the sun. Thus, it seems that to describe a physical phenomenon we must take into account all the physics relevant over all scales, from (say) quantum mechanics to (say) general relativity. But in fact, when we do a table-top experiment, we hardly resort to any of them but just classical mechanics works fine. Why is it so? This is because the effects of the scales too much different from relevant one are suppressed by powers of the ratio of scales in the problem. Thus we need not worry too much about quantum mechanical effects ($\sim 10^{10}$m) when we perform a table-top experiment ($\sim 10^0 - 10^1$m). This separation of scales is more formally elaborated in quantum field theory as *effective field theory*.

We can obtain an effective field theory in 2 ways. If the mother theory is known that contains (for simplicity) both a light degree of freedom $\phi$ relevant for low-energy physics and a heavy one $\Phi$ whose mass is so larger than the energy scale of our interest that its effects are not important, formally we can integrate out $\Phi$ by performing a path integral. This results in an effective action for $\phi$ solely,

$$e^{iS_{\rm eff}(\phi)} = \int [D\Phi] e^{iS(\phi,\Phi)} \, . \tag{256}$$

Typically, this gives rise to non-local, higher dimensional terms since the propagators of $\Phi$ intervene those of $\phi$ in such a way that, for example,

$$\phi \left(-\Box + M^2\right)^{-1} \phi = \frac{\phi}{M^2}\left(1 + \frac{\Box}{M^2} + \cdots\right)\phi \, , \tag{257}$$

where $M$ is the mass scale of $\Phi$ which satisfies (regime of our interest) $\ll M \lesssim m_{\rm Pl}$. But in many cases we cannot expect to have such a luxury of knowing the mother theory. Rather, usually we only know an effective Lagrangian at a low-energy with a cutoff scale $\Lambda$ which satisfies (regime of our interest) $\ll \Lambda \lesssim m_{\rm Pl}$, and parametrize our ignorance based on symmetry principles. That is, with a number of symmetries (Lorentz, gauge, global...) that survive at low-energies, we write down all the possible operators consistent with those symmetries. In doing so we necessarily take into account the effects of the integrated out heavy physics in a model independent way. In this case, we also face higher dimensional terms. That is,

$$\mathcal{L}_{\rm eff} = \mathcal{L}_\phi + \sum_i c_i \frac{\mathcal{O}_i}{M^{n_i-4}} \quad \left(\text{or } \Lambda^{-(n_i-4)}\right) \, , \tag{258}$$

with $\mathcal{O}_i$ being an operator of dimension $n_i > 4$. These operators include not only $\phi$ itself but also its derivatives and cross-terms with $\phi$, e.g.

$$\phi^n \,,\ (\partial_\mu \phi)^n \,,\ (\partial_\mu \phi)^n \phi^m \,, \tag{259}$$

and so on. Thus, from the effective theory point of view, we expect many sub-Planckian structures that may well interrupt otherwise successful large-field inflation with super-Planckian field excursions.

A good example of how these higher dimensional operators affect the predictions from naive Lagrangian is the so-called "$\eta$-problem". In its original context, this tells us that when constructing an inflation model in supergravity, because of the overall exponential factor for the Kähler potential the inflaton mass receives $\mathcal{O}(H^2)$ correction thus spoiling the slow-roll condition $m_{\rm Pl}^2 V''/V \sim \mathcal{O}(1)$ (this parameter is called $\eta$ here). However, the problem is not only the corrections to the potential. Typically, the terms allowed in effective field theory spoil the otherwise smooth structure (both potential and kinetic sectors) of the theory, hence interrupting successful long enough period of inflation. This is a key challenge not easily surmountable in realistic inflation model building in the context of particle physics.

Thus, we usually need a non-trivial mechanism to protect and/or prevent certain classes of operators. For example, if we impose an exact shift symmetry, i.e. the Lagrangian is invariant under a constant shift of $\phi$,

$$\phi \to \phi + \text{constant} \,, \tag{260}$$

all non-derivative terms are forbidden by this symmetry, i.e. there is no potential term. Usually an exact symmetry is weakly broken by loop effects, i.e. potential is induced by radiative effects. Thus in this case small mass of $\phi$ can be explained.

# 4 Non-Gaussianity

In the previous sections we have concentrated on the linear perturbations and the power spectra. As we have seen, the quadratic action for scalar and tensor perturbations are (after some manipulations) essentially that of a quantum harmonic oscillator. Thus we can resort to the conventional wisdom of quantum physics to solve the system, and the solutions are free ones – thus following Gaussian properties. This is because, during inflation, the expansion is so fast that any higher order (i.e. beyond quadratic order) interactions that lead to deviations from Gaussianity are diluted to a very small level. Indeed, current observations on the CMB indicate that the properties of the primordial perturbations are very close to Gaussian. That is, power spectra are all we need: any odd-order correlation function disappears, and even-order ones can be written in terms of the product of power spectra.

In other words, if we can ever probe any deviation from Gaussianity, i.e. *non-Gaussianity*, we have a very powerful probe to study the inflationary dynamics since we can investigate *interactions* beyond free field limit. Moreover, as power spectra did, we can use the non-trivial higher order correlation function, especially the first one – 3-point correlation function, or "bispectrum" – to distinguish different models of inflation. Thanks to the rapid developments in observational instruments and techniques, we may well hope to constrain severely the degree of non-Gaussianity from various cosmological observations, including the CMB and the distribution pattern of galaxies on large scales.

In this section, we study how to specify and quantify primordial non-Gaussianity. First we quickly sketch the so-called in-in formalism which we will use to compute quantum mechanical non-Gaussianity. The existence of interaction terms demands the vacuum state different from the one of free field theory, which is taken into account when we apply the in-in formalism. Then, we concentrate on the so-called squeezed configuration of the bispectrum. In this limit we can extract another powerful probe to constrain all classes of single field inflation models universally.

## 4.1 Bispectrum and non-Gaussianity

In the statistical point of view, a quantity with Gaussian distribution is completely described by its power spectrum. That is, all the correlation function with odd power vanish, and the connected part of the correlation functions vanish. The "connected part" means the part of a correlation function which cannot be expressed in terms of the correlation functions with lower order[7]: let us consider a zero-mean random field $\delta(t,\boldsymbol{x})$, i.e. $\langle\delta\rangle = 0$. Then, up to 3rd order

[7] At this point, it would be illustrative to remind of the Wick's theorem. There are many different description of the Wick's theorem, e.g. in quantum field theory, but the most appropriate form for the current purpose would be this:

$$\text{The correlation of order } m = \sum_l \begin{array}{l}\text{product of the connected (unreducible)}\\ \text{correlation of order } l \le m\end{array} .$$

For example,

$$\langle\delta_1\delta_2\delta_3\rangle_{\text{full}} = \langle\delta_1\rangle_c\langle\delta_2\rangle_c\langle\delta_3\rangle_c + \langle\delta_1\rangle_c\langle\delta_2\delta_3\rangle_c + (\text{cyclic}) + \langle\delta_1\delta_2\delta_3\rangle_c\,. \tag{261}$$

correlation functions, we find [with $\delta(t, \boldsymbol{x}_1) \equiv \delta_1$]

$$\langle \delta_1 \delta_2 \rangle_{\text{full}} = \langle \delta_1 \rangle \langle \delta_2 \rangle + \langle \delta_1 \delta_2 \rangle_c = \langle \delta_1 \delta_2 \rangle_c \,, \tag{262}$$

$$\langle \delta_1 \delta_2 \delta_3 \rangle_{\text{full}} = \langle \delta_1 \rangle \langle \delta_2 \delta_3 \rangle + \langle \delta_2 \rangle \langle \delta_1 \delta_3 \rangle + \langle \delta_3 \rangle \langle \delta_1 \delta_2 \rangle + \langle \delta_1 \delta_2 \delta_3 \rangle_c = \langle \delta_1 \delta_2 \delta_3 \rangle_c \,. \tag{263}$$

That is, the 2nd and 3rd order connected correlation functions coincide with the full correlation functions themselves. Meanwhile, for 4th order,

$$\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle = \langle \delta_1 \delta_2 \rangle \langle \delta_3 \delta_4 \rangle + \langle \delta_1 \delta_3 \rangle \langle \delta_2 \delta_4 \rangle + \langle \delta_1 \delta_4 \rangle \langle \delta_2 \delta_3 \rangle + \langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle_c + \text{terms with } \langle \delta_i \rangle \,. \tag{264}$$

For this to be completely described by 2nd order correlation functions, the connected 4th order correlation function $\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle_c$ should disappear. More generally, $\langle \delta_1 \cdots \delta_N \rangle_c = 0$ for $N > 2$.

Hence, non-vanishing higher order connected correlation function indicates that the corresponding quantity is non-Gaussian distributed. Especially, the 3-point function, or its Fourier transform, the bispectrum, represents the lowest order statistics with which we are able to distinguish non-Gaussian from Gaussian perturbations.

## 4.2 Prescriptions of in-in formalism

Higher order correlation functions are generated when the scalar field has some interaction with itself or other fields. This amounts to saying that the action contains higher order contributions beyond the quadratic order. In this respect, the total Hamiltonian $H$ can be written as a combination of the free Hamiltonian $H_0$ and the interaction Hamiltonian $H_{\text{int}}$,

$$H = H_0 + H_{\text{int}} \,. \tag{265}$$

To find the interaction Hamiltonian usually requires that only dynamical degrees of freedom remain in the action. Once the action is given, we can construct the Hamiltonian by defining conjugate momenta, and separating out the quadratic from the higher order parts, i.e.

$$H_0 \supset \{\text{quadratic terms in terms of the perturbative degrees of freedom}\} \,, \tag{266}$$

$$H_{\text{int}} \supset \{\text{higher order terms}\} \,. \tag{267}$$

Now, the crucial point is that the system is *not* free but contains interactions, and hence any (vacuum) expectation values should be taken with respect to the interaction vacuum state $|\Omega\rangle$, i.e. the *actual* vacuum state of the theory, *not* the free vacuum state $|0\rangle$ defined by $a_{\boldsymbol{k}}|0\rangle = 0$ for all $\boldsymbol{k}$.

For this description, we resort to the *interaction picture*. In quantum mechanics, the interaction picture (or Dirac picture) is an intermediate between the Schrödinger picture and the Heisenberg picture. Whereas in the other two pictures either the state vector or the operators carry time dependence, in the interaction picture both carry part of the time dependence of observables. The purpose of the interaction picture is to shunt all the time dependence due to $H_0$ onto the operators, leaving only $H_{\text{int}}$ affecting the time-dependence of the state vectors: reviving for the moment the Planck constant $\hbar$, the time evolution of state vectors and operators are given by

$$i\hbar \frac{d}{dt} |\psi_{\text{int}}(t)\rangle = H_{\text{int}}(t) |\psi_{\text{int}}(t)\rangle \,, \tag{268}$$

$$i\hbar \frac{d}{dt} \widehat{A}_{\text{int}}(t) = \left[ \widehat{A}_{\text{int}}(t), H_0 \right] \,, \tag{269}$$

respectively.

Now, we denote by $\left\langle \widehat{\mathcal{O}}(t) \right\rangle$ the expectation value evaluated at a time $t$ of a time dependent operator

$$\widehat{\mathcal{O}}(t) = \left( e^{-i\int_{t_{\rm in}}^t H_0(t')dt'} \right)^\dagger \widehat{\mathcal{O}} \left( e^{-i\int_{t_{\rm in}}^t H_0(t'')dt''} \right) , \tag{270}$$

where $t_{\rm in}$ is some early "in" time when the interaction is turned on. This expectation value is taken with respect to the vacuum state at that time $|\Omega(t)\rangle$, which has evolved from an "in" state $|{\rm in}\rangle$ according to

$$|\Omega(t)\rangle = e^{-i\int_{t_{\rm in}}^t H_{\rm int}(t')dt'} |{\rm in}\rangle \equiv U_{\rm int}(t, t_{\rm in})|{\rm in}\rangle . \tag{271}$$

Then, we can write

$$\left\langle \widehat{\mathcal{O}}(t) \right\rangle = \frac{\left\langle \Omega(t) \left| \widehat{\mathcal{O}}(t) \right| \Omega(t) \right\rangle}{\langle \Omega(t) \mid \Omega(t) \rangle} = \frac{\left\langle {\rm in} \left| U_{\rm int}^\dagger(t, t_{\rm in}) \widehat{\mathcal{O}}(t) U_{\rm int}(t, t_{\rm in}) \right| {\rm in} \right\rangle}{\left\langle {\rm in} \left| U_{\rm int}^\dagger(t, t_{\rm in}) U_{\rm int}(t, t_{\rm in}) \right| {\rm in} \right\rangle} . \tag{272}$$

To specify the "initial" conditions in the context of quantum field theory, we usually first find the eigenstates $|n\rangle$ of the free Hamiltonian $H_0$ and stipulate that the system begins in one or some combination of $|n\rangle$: if the system begins in the quantum mechanical vacuum, this amounts to putting our system in the vacuum state of $H_0$, which we denote by $|0\rangle$, at the initial time $t_{\rm in}$, in the sense that initially the interaction is turned off. Hence, Assuming that $|0\rangle$ is properly normalized, we obtain

$$\left\langle \widehat{\mathcal{O}}(t) \right\rangle = \left\langle 0 \left| U_{\rm int}^\dagger(t, t_{\rm in}) \widehat{\mathcal{O}}(t) U_{\rm int}(t, t_{\rm in}) \right| 0 \right\rangle . \tag{273}$$

One further technical manipulation is the contour of the time integral. To explicitly this, we first consider the complete set of the full interacting theory $|n\rangle$, which is the eigenstates of $H$, with $|n = 0\rangle = |\Omega\rangle$ being the interacting vacuum state. To obtain $|\Omega\rangle$, we evolve $|0\rangle$ for some time from the initial time $t_{\rm in}$, and use the eigenstates of the theory such that

$$e^{-iH(t-t_{\rm in})} = e^{-iE_0(t-t_{\rm in})}|\Omega\rangle\langle\Omega|0\rangle + \sum_{n\geq 1} e^{-iE_n(t-t_{\rm in})}|n\rangle\langle n|0\rangle . \tag{274}$$

If we slightly distort the contour to include small imaginary part, $t_{\rm in} \to -\infty(1-i\delta)$, we see that there appears an exponential suppression factor $\exp(+iE_0 t_{\rm in}) = \exp\left[-iE_0(1-i\delta)\infty\right] \sim e^{-\infty}$. Thus all terms from the sum over $n \geq 1$ become exponentially small compared to the leading term involving $|\Omega\rangle$. It thus follows that $|\Omega\rangle$, the vacuum of the full theory at a time $t$, is given by

$$|\Omega\rangle = \lim_{t_{\rm in} \to -\infty(1-i\delta)} \frac{e^{-iH(t-t_{\rm in})}}{e^{-iE_0(t-t_0)}\langle\Omega|0\rangle} |0\rangle . \tag{275}$$

Therefore, going back to (272) and using (270) and (271), we obtain

$$\begin{aligned} \left\langle \widehat{\mathcal{O}}(t) \right\rangle &= \lim_{t_{\rm in} \to -\infty(1-i\delta)} \frac{\left\langle 0 \left| \left( e^{-i\int_{t_{\rm in}}^t H(t')dt'} \right)^\dagger \widehat{\mathcal{O}} \left( e^{-i\int_{t_{\rm in}}^t H(t'')dt''} \right) \right| 0 \right\rangle}{\langle 0 \mid 0 \rangle} \\ &= \lim_{t_{\rm in} \to -\infty(1-i\delta)} \left\langle 0 \left| U_{\rm int}^\dagger(t, t_{\rm in}) \widehat{\mathcal{O}}(t) U_{\rm int}(t, t_{\rm in}) \right| 0 \right\rangle . \end{aligned} \tag{276}$$

Note that this is the same as (273), with the prescription on the time integral contour being specified. In (276) which is being evaluated in the complex time plane, from right to left, time starts from infinite past with slightly positive imaginary part, or shortly $-\infty^+$, to some arbitrary time $t$ when the expectation value is evaluated, then back to $-\infty^-$. This time contour, which is shown in Figure 6, forms a closed time path. This is why the in-in formalism is sometimes called closed time path formalism. It is also very important to note that the time forward contour does not coincide with the time backward contour: the vacuum specification has broken the time symmetry of the forward and backward time integrals.

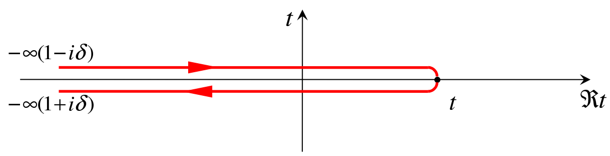


Figure 6: Closed time path in "in-in" formalism.

Now, let us consider $H_{\rm int}$ as a small perturbation to the free Hamiltonian $H_0$. Expanding the exponential in (276) up to 2nd order in terms of $H_{\rm int}$, we obtain[8]

$$
\begin{aligned}
\left\langle \widehat{\mathcal{O}}(t) \right\rangle = & \left\langle 0 \middle| \left[ 1 - i\int_{t_{\rm in}}^t H_{\rm int}(t')dt' + \frac{1}{2}\left( -i\int_{t_{\rm in}}^t H_{\rm int}(t')dt' \right)^2 \right]^\dagger \widehat{\mathcal{O}}(t) \right. \\
& \left. \times \left[ 1 - i\int_{t_{\rm in}}^t H_{\rm int}(t')dt' + \frac{1}{2}\left( -i\int_{t_{\rm in}}^t H_{\rm int}(t'')dt'' \right)^2 \right] \middle| 0 \right\rangle \\
= & \left\langle 0 \middle| \left[ 1 + \left( -i\int_{t_{\rm in}}^t H_{\rm int}(t')dt' \right)^\dagger + \frac{1}{2}\left( -\int_{t_{\rm in}}^t H_{\rm int}(t_1)dt_1 \int_{t_{\rm in}}^t H_{\rm int}(t_2)dt_2 \right)^\dagger \right] \widehat{\mathcal{O}}(t) \right. \\
& \left. \times \left[ 1 + \left( -i\int_{t_{\rm in}}^t H_{\rm int}(t')dt' \right) + \frac{1}{2}\left( -\int_{t_{\rm in}}^t H_{\rm int}(t_1)dt_1 \int_{t_{\rm in}}^t H_{\rm int}(t_2)dt_2 \right) \right] \middle| 0 \right\rangle \\
= & \left\langle 0 \middle| \widehat{\mathcal{O}}(t) + i\left\{ \int_{t_{\rm in}}^t \left[ H_{\rm int}(t')\widehat{\mathcal{O}}(t) - \widehat{\mathcal{O}}(t)H_{\rm int}(t') \right] dt' \right\} \right. \\
& - \frac{1}{2}\left\{ \int_{t_{\rm in}}^t dt_2 \int_{t_{\rm in}}^t dt_1 \left[ H_{\rm int}(t_2)H_{\rm int}(t_1)\widehat{\mathcal{O}}(t) - 2H_{\rm int}(t_2)\widehat{\mathcal{O}}(t)H_{\rm int}(t_1) \right. \right. \\
& \left. \left. \left. + \widehat{\mathcal{O}}(t)H_{\rm int}(t_2)H_{\rm int}(t_1) \right] \right\} \middle| 0 \right\rangle \\
= & \left\langle 0 \middle| \widehat{\mathcal{O}}(t) \middle| 0 \right\rangle + i\int_{t_{\rm in}}^t dt_1 \left\langle 0 \middle| \left[ H_{\rm int}(t'), \widehat{\mathcal{O}}(t) \right] \middle| 0 \right\rangle \\
& + i^2 \int_{t_{\rm in}}^t dt_2 \int_{t_{\rm in}}^{t_2} dt_1 \left\langle 0 \middle| \left[ H_{\rm int}(t_1), \left[ H_{\rm int}(t_2), \widehat{\mathcal{O}}(t) \right] \right] \middle| 0 \right\rangle \,,
\end{aligned}
\tag{277}
$$

where we have used for the last equality, the commutator identity

$$
AAB - 2ABA + BAA = [A,[A,B]] \,, \tag{278}
$$

[8] Here, we omit for simplicity $\lim_{t_{\rm in}\to-\infty^+}$ but $t_{\rm in}$ is evaluated in this limit after all.

and to reduce the double integral inside the curly brackets, the fact that for any symmetric, holomorphic function, $f(t_1,t_2)=f(t_2,t_1)$,

$$\int_a^b dt_1 \int_a^b dt_2 f(t_1,t_2) = 2\int_a^b dt_1 \int_a^{t_1} dt_2 f(t_1,t_2)\,. \tag{279}$$

This is because $f(t_1,t_2)$, for the current case $f(t_1,t_2)=H_{\text{int}}(t_1)H_{\text{int}}(t_2)$, is symmetric under the exchange of its arguments, the integral over the square region on the LHS is twice the RHS integral over the lower shaded triangle, where $t_2<t_1$. In this way, we can show that (276) is formally consistent with the infinite sum of the nested commutators,

$$\left\langle \widehat{\mathcal{O}}(t)\right\rangle = \sum_{n=0}^{\infty} i^n \int_{t_{\text{in}}}^{t} dt_n \int_{t_{\text{in}}}^{t_n} dt_{n-1}\cdots\int_{t_{\text{in}}}^{t_2} dt_1 \left\langle 0\left|\left[H_{\text{int}}(t_1),\left[H_{\text{int}}(t_2),\cdots\left[H_{\text{int}}(t_n),\widehat{\mathcal{O}}(t)\right]\cdots\right]\right]\right|0\right\rangle\,. \tag{280}$$

## 4.3 Bispectrum of the curvature perturbation

### 4.3.1 Cubic interaction and bispectrum

Now we return to our main interest, the 3-point correlation function of the comoving curvature perturbation $\mathcal{R}$. According to the prescription of the in-in formalism, the operator whose expectation value we want to compute is the 3 copies of the curvature perturbation, i.e.

$$\widehat{\mathcal{O}}(t) = \mathcal{R}(\boldsymbol{k}_1,t)\mathcal{R}(\boldsymbol{k}_2,t)\mathcal{R}(\boldsymbol{k}_3,t)\,. \tag{281}$$

Thus, to leading order in $H_{\text{int}}$, we can define the bispectrum $B_{\mathcal{R}}(k_1,k_2,k_3)$ as

$$\begin{aligned}\langle\mathcal{R}(\boldsymbol{k}_1,t)\mathcal{R}(\boldsymbol{k}_2,t)\mathcal{R}(\boldsymbol{k}_3,t)\rangle &\equiv (2\pi)^3\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)B_{\mathcal{R}}(k_1,k_2,k_3)\\ &= i\int_{t_{\text{in}}}^{t} dt'\,\langle 0\,|[H_{\text{int}}(t'),\mathcal{R}(\boldsymbol{k}_1,t)\mathcal{R}(\boldsymbol{k}_2,t)\mathcal{R}(\boldsymbol{k}_3,t)]|\,0\rangle\,.\end{aligned} \tag{282}$$

Thus, to have non-zero bispectrum, we need to find at least 3rd order Hamiltonian, $H_3=-L_3$: it contains 3 copies $\mathcal{R}$, thus we can connect every $\mathcal{R}$ to another, leaving no disconnected piece. The computation of the 3rd order action is essentially the same as what we have done to obtain the 2nd order action, but only more lengthy and tedious. For more detail, see e.g. Maldacena (2003) where the quantum calculation is presented for the first time. The result is

$$S_3 = \int d^4x a^3 m_{\text{Pl}}^2\left[-\epsilon\mathcal{R}\frac{(\nabla\mathcal{R})^2}{a^2}+3\epsilon\dot{\mathcal{R}}^2\mathcal{R}-\frac{\epsilon}{H}\dot{\mathcal{R}}^3+\frac{1}{2a^4}\left\{\left(3\mathcal{R}-\frac{\dot{\mathcal{R}}}{H}\right)\left[\psi^{,ij}\psi_{,ij}-(\Delta\psi)^2\right]-4\mathcal{R}^{,i}\psi_{,i}\Delta\psi\right\}\right]\,, \tag{283}$$

where $\psi$ is the solution of the $0i$ component of the scalar metric perturbation $B$ already presented in (133),

$$\psi \equiv aB = -\frac{\dot{\mathcal{R}}}{H}+\underbrace{a^2\epsilon\Delta^{-1}\dot{\mathcal{R}}}_{\equiv\chi}\,. \tag{284}$$

This form of the action is not bad, but one may suspect that while $S_2$ is overall suppressed by $\epsilon$, $S_3$ seems to have unsuppressed contributions and thus the bispectrum may be even larger than the power spectrum. In fact, one may use $\delta\phi$ instead of $\mathcal{R}$ (remember that when we reduce the quadratic Lagrangian, from $\mathcal{C}_A = 0$ we could eliminate $\Pi_\varphi$ in favour of $\delta\phi$) where $S_3$ is obviously suppressed by $\epsilon^2$. We can explicitly check this by performing a number of partial integrations, and the result is

$$
\begin{aligned}
S_3 = \int d^4x \Bigg[ & a^3 m_{\rm Pl}^2 \left\{ \epsilon^2 \dot{\mathcal{R}}^2 \mathcal{R} + \epsilon^2 \mathcal{R} \frac{(\nabla\mathcal{R})^2}{a^2} - 2\epsilon\dot{\mathcal{R}} \frac{\mathcal{R}^{,i}\chi_{,i}}{a^2} + \frac{\epsilon}{2}\dot{\eta}\dot{\mathcal{R}}\mathcal{R}^2 + \frac{\epsilon}{2a^4}\mathcal{R}^{,i}\chi_{,i}\Delta\chi + \frac{1}{4a^4}\Delta\mathcal{R}(\nabla\chi)^2 \right\} \\
& + 2 \left.\frac{\delta L}{\delta\mathcal{R}}\right|_1 \left\{ \frac{\eta}{4}\mathcal{R}^2 + \frac{1}{H}\dot{\mathcal{R}}\mathcal{R} + \frac{1}{4a^2H^2}\left[ -(\nabla\mathcal{R})^2 + \Delta^{-1}\left(\mathcal{R}^{,i}\mathcal{R}^{,j}\right)_{,ij} \right] \right. \\
& \left. + \frac{1}{2a^2H}\left[ \mathcal{R}^{,i}\chi_{,i} - \Delta^{-1}\left(\mathcal{R}^{,i}\chi^{,j}\right)_{,ij} \right] \right\} \Bigg] \,, 
\end{aligned}
\tag{285}
$$

where $\delta L/\delta\mathcal{R}|_1$ denotes the linear equation of motion for $\mathcal{R}$ derived from $S_2$,

$$
\left.\frac{\delta L}{\delta\mathcal{R}}\right|_1 = a^3\epsilon m_{\rm Pl}^2 \left[ \frac{1}{a^3\epsilon}\frac{d}{dt}\left(a^3\epsilon\dot{\mathcal{R}}\right) - \frac{\Delta}{a^2}\mathcal{R} \right] . \tag{286}
$$

Clearly, in (285) barring the terms multiplied by $\delta L/\delta\mathcal{R}|_1$, which we now denote by $f(\mathcal{R})$, all terms are suppressed by $\epsilon^2$. Then, how to deal with $f(\mathcal{R})$? Fortunately, at this order we can eliminate $f(\mathcal{R})$ by a simple field redefinition. If we redefine $\mathcal{R}$ non-linearly [remember that $f(\mathcal{R}) = \mathcal{O}(\mathcal{R}^2)$] as

$$
\mathcal{R} \to \mathcal{R} + f(\mathcal{R}) \,, \tag{287}
$$

the quadratic action (151) becomes, keeping up to 3rd order in $\mathcal{R}$,

$$
\begin{aligned}
S_2 \to & \int d^4x a^3\epsilon m_{\rm Pl}^2 \left[ \dot{\mathcal{R}}^2 + 2\dot{\mathcal{R}}\dot{f} - \frac{(\nabla\mathcal{R})^2 + 2(\nabla\mathcal{R})(\nabla f)}{a^2} \right] \\
= & \int d^4x a^3\epsilon m_{\rm Pl}^2 \left[ \dot{\mathcal{R}}^2 - \frac{(\nabla\mathcal{R})^2}{a^2} \right] + \underbrace{\int d^4x \left[ 2a^3\epsilon m_{\rm Pl}^2 \dot{\mathcal{R}}\dot{f} - 2a\epsilon m_{\rm Pl}^2 (\nabla\mathcal{R})(\nabla f) \right]}_{= \int d^4x \left[ -2 \frac{\delta L}{\delta\mathcal{R}}\big|_1 f(\mathcal{R}) \right]} \,,
\end{aligned}
\tag{288}
$$

thus considering the whole action up to 3rd order $S_2 + S_3$, $f(R)$ terms precisely cancel each other.

The calculation of the full bispectrum is now straightforward, a bit tedious though. One thing we should not forget is that, unlike in quantum field theory in static background, the redefined field $\mathcal{R} + f(\mathcal{R})$ is *different* from the original one $\mathcal{R}$. This distinction is important since the background is time evolving. Thus, at the last stage we have to compensate this redefinition by including the contributions coming from this redefinition. Schematically, if we write

$$
\mathcal{R} \to \mathcal{R} + f(\mathcal{R}) = \mathcal{R} + \lambda\mathcal{R}^2 \,, \tag{289}
$$

where $\lambda$ denotes the quadratic order operators that appear in $f(\mathcal{R})$, the bispectrum we need to calculate is [with $\mathcal{R}_1 \equiv \mathcal{R}(\boldsymbol{k}_1)$ etc]

$$
\langle \mathcal{R}_1\mathcal{R}_2\mathcal{R}_3 \rangle \to \langle \mathcal{R}_1\mathcal{R}_2\mathcal{R}_3 \rangle + \lambda \left[ \langle \mathcal{R}_1\mathcal{R}_2 \rangle \langle \mathcal{R}_1\mathcal{R}_3 \rangle + 2 \text{ perm} \right] , \tag{290}
$$

where the 1st term comes from $H_3$ while the 2nd contribution from the field redefinition.

### 4.3.2 Consistency relation

As we can expect, the bispectrum in general exhibits very rich configurations, since the conservation of the 3 momenta allows infinitely many shapes of the triangle the 3 momenta can form. Thus, usually we think of convenient and representative "templates", i.e. certain configurations of the triangle. A convenient and physically very important one is the so-called squeezed configuration, which corresponds to the local-type non-Gaussianity.

Let us assume that $\mathcal{R} = \mathcal{R}(\boldsymbol{x})$ is a local function and we can expand the fully non-linear curvature perturbation $\mathcal{R}$ *locally* as[9]

$$\mathcal{R} = \mathcal{R}^{(1)} + \frac{3}{5} f_{\rm NL} \left(\mathcal{R}^{(1)}\right)^2 + \cdots \,, \tag{291}$$

where $\mathcal{R}^{(1)}$ is the linear, Gaussian part of the curvature perturbation which gives the (leading) power spectrum $\mathcal{P}_\mathcal{R}$,

$$\left\langle \mathcal{R}_k^{(1)} \mathcal{R}_q^{(1)} \right\rangle = (2\pi)^3 \delta^{(3)}(\boldsymbol{k}+\boldsymbol{q}) P_\mathcal{R}(k) \,, \tag{292}$$

and the 2nd order expansion coefficient, "non-linear parameter", $f_{\rm NL}$ is a constant. Moving to the Fourier space, we have

$$\mathcal{R}_k = \mathcal{R}_k^{(1)} + \underbrace{\frac{3}{5} f_{\rm NL} \int \frac{d^3q_1 d^3q_2}{(2\pi)^3} \delta^{(3)}(\boldsymbol{k} - \boldsymbol{q}_1 - \boldsymbol{q}_2) \mathcal{R}_{q_1}^{(1)} \mathcal{R}_{q_2}^{(1)}}_{\equiv \mathcal{R}_k^{(2)}} + \cdots \,, \tag{293}$$

the bispectrum is immediately computed as

$$\langle \mathcal{R}_1 \mathcal{R}_2 \mathcal{R}_3 \rangle = \left\langle \mathcal{R}_1^{(1)} \mathcal{R}_2^{(1)} \mathcal{R}_3^{(1)} \right\rangle + \left\langle \mathcal{R}_1^{(1)} \mathcal{R}_2^{(1)} \mathcal{R}_3^{(2)} \right\rangle + 2 \text{ perm} + \cdots \,, \tag{294}$$

where the 1st term obviously vanishes, and the 1st non-zero contribution comes from the term that contain one $\mathcal{R}^{(2)}$. Explicitly,

$$\langle \mathcal{R}_1 \mathcal{R}_2 \mathcal{R}_3 \rangle = (2\pi)^3 \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) \frac{6}{5} f_{\rm NL} \left[ P_\mathcal{R}(k_1) P_\mathcal{R}(k_2) + 2 \text{ perm} \right] . \tag{295}$$

It is very important to note that this form of the bispectrum is obtained by *assuming* local expansion of the curvature perturbation. But nevertheless this "local form" of the bispectrum provides an important and useful parametrization as follows. A particularly interesting limit is obtained if we take 1 of the 3 momenta, (say) $k_3$, very smaller than the other 2, (say) $k_1 \approx k_2$. In this case, the shape of the triangle is highly squeezed in $k_3$, so it is called "squeezed" configuration. Then, we can pull $2P_\mathcal{R}(k_1)P_\mathcal{R}(k_3)$ from the squared brackets to have

$$P_\mathcal{R}(k_1) P_\mathcal{R}(k_2) + 2 \text{ perm} = 2 P_\mathcal{R}(k_1) P_\mathcal{R}(k_3) \left[ 1 + \frac{P_\mathcal{R}(k_1)}{2P_\mathcal{R}(k_3)} \right] . \tag{296}$$

[9] The coefficient 3/5 in front of $f_{\rm NL}$ is added because originally $f_{\rm NL}$ is defined in terms of the Newtonian potential $\Phi$. It is related to the comoving curvature perturbation $\mathcal{R}$ by $\Phi = 3\mathcal{R}/5$ on large scales, for which the corresponding modes entere the horizon during matter dominated era.

Since $P_\mathcal{R} \sim k^{-3}\mathcal{P}_\mathcal{R} \propto k^{n_\mathcal{R}-4}$, the last term in the square brackets above behaves as $P_\mathcal{R}(k_1)/P_\mathcal{R}(k_3) \propto (k_3/k_1)^{4-n_\mathcal{R}} \to 0$, thus

$$\langle \mathcal{R}_1\mathcal{R}_2\mathcal{R}_3 \rangle \underset{k_3\to 0}{\longrightarrow} (2\pi)^3\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\frac{12}{5}f_\text{NL}P_\mathcal{R}(k_1)P_\mathcal{R}(k_3)\,. \tag{297}$$

Matching this expression with the definition of the bispectrum (282), we can find the local non-linear parameter in the squeezed configuration as

$$f_\text{NL} = \frac{5}{12}\lim_{k_3\to 0}\frac{B_\mathcal{R}(k_1,k_2,k_3)}{P_\mathcal{R}(k_1)P_\mathcal{R}(k_3)}\,. \tag{298}$$

Physically, in the squeezed configuration, the mode corresponding to small $k$ has much longer wavelength than the other 2. In other words, the long wavelength mode is already outside the horizon while the other 2 remain sub-horizon. Thus, essentially the long wavelength mode can be seen for the small scale modes very smooth, slowly varying background. This statement is very generic independent of the detail, so one may expect a universal relation irrespective of the detail of inflationary dynamics in the squeezed bispectrum. Indeed there exists such a relation valid universally for *all* single field inflation models.

It is more convenient to work in the configuration space rather than the Fourier space. The correlation functions are related in the standard manner, e.g.

$$\begin{aligned}\langle \mathcal{R}(\boldsymbol{x}_1)\mathcal{R}(\boldsymbol{x}_2)\rangle &= \int \frac{d^3k_1}{(2\pi)^3}\frac{d^3k_2}{(2\pi)^3}e^{i\boldsymbol{k}_1\cdot\boldsymbol{x}_1}e^{i\boldsymbol{k}_2\cdot\boldsymbol{x}_2}\langle \mathcal{R}(\boldsymbol{k}_1)\mathcal{R}(\boldsymbol{k}_2)\rangle \\ &= \int \frac{d^3k}{(2\pi)^3}e^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}P_\mathcal{R}(k)\,,\end{aligned} \tag{299}$$

and so on. Now, the correlation of 3 $\mathcal{R}$'s with 2 small scale modes (denoted by a subscript $S$) being under the influence of a long scale one (denoted by a subscript $L$) can be written as

$$\langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2)\mathcal{R}_L(\boldsymbol{x}_3)\rangle \approx \left\langle \langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2)\rangle_L \mathcal{R}_L(\boldsymbol{x}_3)\right\rangle\,, \tag{300}$$

where the subscript $L$ for the 2-point correlation of the small scale modes means under the existence of the long wavelength modes. For this, we can absorb the long wavelength mode $\mathcal{R}_L$ into the metric by redefine the spatial coordinates as

$$\tilde{x}^i \equiv e^{\mathcal{R}_L}x^i\,, \tag{301}$$

so that now the coordinates themselves include the influence of the long wavelength mode. Thus,

$$\langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2)\rangle_L = \langle \mathcal{R}_S(\tilde{\boldsymbol{x}}_1)\mathcal{R}_S(\tilde{\boldsymbol{x}}_2)\rangle \approx \langle \mathcal{R}_S(\boldsymbol{x}_1+\mathcal{R}_L\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2+\mathcal{R}_L\boldsymbol{x}_2)\rangle_L\,. \tag{302}$$

To proceed further, for simplicity without loss of generality let us set $\boldsymbol{x}_2 = 0$. Then, expanding the above equation gives

$$\left\langle \left[\mathcal{R}_S(\boldsymbol{x}_1) + \frac{\partial\mathcal{R}_S(\boldsymbol{x}_1)}{\partial x^i}\mathcal{R}_Lx^i + \cdots\right]\mathcal{R}_S(\boldsymbol{x}_2=0)\right\rangle \approx \left\langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2) + \mathcal{R}_L(\boldsymbol{x}_1)x_1^i\frac{\partial}{\partial x_1^i}\mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2) + \cdots\right\rangle\,. \tag{303}$$

Thus, the 3-point correlation function becomes

$$\begin{aligned}\left\langle \left\langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2)\right\rangle_L \mathcal{R}_L(\boldsymbol{x}_3)\right\rangle \approx{}& \left\langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2)\right\rangle \left\langle \mathcal{R}_L(\boldsymbol{x}_3)\right\rangle \\ &+ \left\langle \mathcal{R}_L(\boldsymbol{x}_1)\mathcal{R}_L(\boldsymbol{x}_3)\right\rangle x_1^i \frac{\partial}{\partial x_1^i}\left\langle \mathcal{R}_S(\boldsymbol{x}_1)\mathcal{R}_S(\boldsymbol{x}_2)\right\rangle \,, \end{aligned} \tag{304}$$

where the 1st term on the RHS vanishes. We now see there are power spectra of short and long wavelength modes separately.

It's now almost done. Using the Fourier modes, the above equation becomes

$$\begin{aligned} &\int \frac{d^3k_L}{(2\pi)^3} e^{i\boldsymbol{k}_L\cdot(\boldsymbol{x}_1-\boldsymbol{x}_3)} P_{\mathcal{R}}(k_L) \int \frac{d^3k_S}{(2\pi)^3} P_{\mathcal{R}}(k_S)\, \underbrace{x_1^i \frac{\partial}{\partial x_1^i} e^{i\boldsymbol{k}_S\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}}_{=k_S^i \frac{\partial}{\partial k_S^i} e^{i\boldsymbol{k}_S\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}} \\ &= \int \frac{d^3k_L}{(2\pi)^3} e^{i\boldsymbol{k}_L\cdot(\boldsymbol{x}_1-\boldsymbol{x}_3)} P_{\mathcal{R}}(k_L) \int \frac{d^3k_S}{(2\pi)^3} \underbrace{P_{\mathcal{R}}(k_S) k_S^i \frac{\partial}{\partial k_S^i} e^{i\boldsymbol{k}_S\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}}_{f'=\frac{\partial}{\partial k_S^i} e^{i\boldsymbol{k}_S\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}\,,\ g=k_S^i P_{\mathcal{R}}(k_S)} \\ &= \int \frac{d^3k_L}{(2\pi)^3} \frac{d^3k_S}{(2\pi)^3} e^{i\boldsymbol{k}_L\cdot(\boldsymbol{x}_1-\boldsymbol{x}_3)} e^{i\boldsymbol{k}_S\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)} P_{\mathcal{R}}(k_L) \frac{\partial}{\partial k_S^i}\left[-k_S^i P_{\mathcal{R}}(k_S)\right] . \end{aligned} \tag{305}$$

Let us consider the derivative inside the integral a bit more closely. Since there are 3 spatial directions, the derivative acting on $k_S^i$ gives just a number 3. Meanwhile, since $P_{\mathcal{R}}(k_S)$ is only dependent on $k_S = \sqrt{(k_S^i)^2}$, $\partial P_{\mathcal{R}}(k_S)/\partial k_S^i = (\partial k_S/\partial k_S^i)[\partial P_{\mathcal{R}}(k_S)/\partial k_S] = (k_S^i/k_S)[\partial P_{\mathcal{R}}(k_S)/\partial k_S]$. Thus,

$$\frac{\partial}{\partial k_S^i}\left[k_S^i P_{\mathcal{R}}(k_S)\right] = 3P_{\mathcal{R}}(k_S) + k_S \frac{\partial P_{\mathcal{R}}(k_S)}{\partial k_S} = P_{\mathcal{R}}(k_S) \frac{d\log\left[k_S^3 P_{\mathcal{R}}(k_S)\right]}{d\log k_S} = (n_{\mathcal{R}}-1) P_{\mathcal{R}}(k_S) \,. \tag{306}$$

Thus, we finally reach

$$\int \frac{d^3k_L}{(2\pi)^3} \frac{d^3k_S}{(2\pi)^3} e^{i\boldsymbol{k}_L\cdot(\boldsymbol{x}_1-\boldsymbol{x}_3)} e^{i\boldsymbol{k}_S\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)} (1-n_{\mathcal{R}}) P_{\mathcal{R}}(k_L) P_{\mathcal{R}}(k_S) \,. \tag{307}$$

Finally, making use of the identification (remember that the orientation of a 3-dimensional vector is arbitrary so its sign does not matter)

$$\boldsymbol{k}_L = -\boldsymbol{k}_3 \,, \tag{308}$$

$$\boldsymbol{k}_S = -\boldsymbol{k}_2 = +\boldsymbol{k}_1 \,, \tag{309}$$

we have $\boldsymbol{k}_L + \boldsymbol{k}_S = -\boldsymbol{k}_2 - \boldsymbol{k}_3 = +\boldsymbol{k}_1$. Further, replacing $e^{-i\boldsymbol{k}_S\cdot\boldsymbol{x}_2}$ with

$$e^{-i\boldsymbol{k}_S\cdot\boldsymbol{x}_2} = \int d^3k_2 e^{i\boldsymbol{k}_2\cdot\boldsymbol{x}_2} \delta^{(3)}(\boldsymbol{k}_2+\boldsymbol{k}_S) = \int d^3k_2 e^{i\boldsymbol{k}_2\cdot\boldsymbol{x}_2} \delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2) \approx \int \frac{d^3k_2}{(2\pi)^3} e^{i\boldsymbol{k}_2\cdot\boldsymbol{x}_2} (2\pi)^3 \delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3) \,, \tag{310}$$

we finally obtain the desired expression

$$\int \frac{d^3k_1}{(2\pi)^3} \frac{d^3k_2}{(2\pi)^3} \frac{d^3k_3}{(2\pi)^3} e^{i\boldsymbol{k}_1\cdot\boldsymbol{x}_1} e^{i\boldsymbol{k}_2\cdot\boldsymbol{x}_2} e^{i\boldsymbol{k}_3\cdot\boldsymbol{x}_3} (2\pi)^3 \delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)(1-n_{\mathcal{R}}) P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3) \,. \tag{311}$$

Thus, we can identify the bispectrum in the squeezed limit as

$$B_{\mathcal{R}}(k_1, k_2, k_3) \underset{k_3 \to 0}{\longrightarrow} (1 - n_{\mathcal{R}}) P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3)\,, \tag{312}$$

or, from (298) the local non-linear parameter is given by

$$f_{\text{NL}} = \frac{5}{12}(1 - n_{\mathcal{R}})\,. \tag{313}$$

Since $n_{\mathcal{R}} \approx 0.96$ by current observations, for *any* single field inflation we have

$$f_{\text{NL}} \sim \mathcal{O}(\epsilon) \ll 1\,. \tag{314}$$

Had we observed large (say, $f_{\text{NL}} \sim 10$) $f_{\text{NL}}$, we would have been able to rule out all single field inflation model. But fortunately or unfortunately, the most recent observations made by the Planck satellite reported

$$f_{\text{NL}} = 2.7 \pm 5.8\,, \tag{315}$$

so $f_{\text{NL}} \ll 1$, i.e. single field inflation, is still mostly consistent with observations.