

Inflation and the primordial perturbations

Hael Collins
The Niels Bohr International Academy

February 6, 2009

Conventions

We shall apply a convention often used in theoretical particle physics, where factors of the speed of light (c) and Planck's constant (\hbar) are not written explicitly. We shall, however, keep track of the factors of Newton's constant (G).

In our notation for the coordinates describing space-time, lowercase Greek letters will range over all the coordinates (for example, $\mu, \nu, \lambda, \dots = 0, 1, 2, 3$) while lowercase English letters will only label the spatial dimensions, (as in $i, j, k, \dots = 1, 2, 3$). Any time that the same index appears as both an upper and a lower index in some factor, we shall be implicitly summing over all the appropriate values for that index. For example,

$$x_\mu x^\mu = x_0 x^0 + x_1 x^1 + x_2 x^2 + x_3 x^3. \quad (1)$$

We shall often write spatial vectors as \vec{x} when the spatial dimensions are flat. In that case, the scalar product is the usual one,

$$\vec{k} \cdot \vec{x} = \delta_{ij} k^i x^j = k^1 x^1 + k^2 x^2 + k^3 x^3. \quad (2)$$

Many different conventions are used in different treatments of the theory of relativity. Here we shall use the $[+, -, -, -]$ convention, so that the Minkowski, or flat space, metric would be

$$ds^2 = dt^2 - d\vec{x} \cdot d\vec{x}. \quad (3)$$

We shall also apply a few abbreviations when writing derivatives. A derivative with respect to all the space-time coordinates will be denoted by

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}. \quad (4)$$

In these notes we shall define the time coordinate in two different ways; derivatives with respect to the ordinary time will be represented with a dot ($\dot{}$), for example,

$$\dot{a} \equiv \frac{da}{dt}, \quad (5)$$

while a derivative with respect to the conformal time (which we shall explain later) will be noted by a prime (\prime), as in

$$a' \equiv \frac{da}{d\eta}. \quad (6)$$

Finally, ∇_μ is the covariant derivative. When written with a vector sign, however, $\vec{\nabla}$ represents an ordinary vector derivative in flat space; that is,

$$\vec{\nabla} \equiv \left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right). \quad (7)$$

In special relativity, the speed at which a signal propagates and the rate at which the size of the visible part of the universe grows are equal. Both propagate at the speed of light. In a universe governed by general relativity, however, this simple equality no longer holds. For example, it is possible for space to expand in such a way that the light of a sufficiently distant object will never reach us. As its light propagates toward us through space, space itself can be expanding so that the distance that the light must still traverse is growing ever larger. Depending on exactly *how* the universe is expanding, there can be a maximal distance beyond which a signal cannot outrun our expanding separation from it. This maximal distance is called an observer's *horizon*. Like a black hole's horizon, as long as the universe continues to expand in the same way, whatever passes beyond this horizon can no longer be seen by us, and the size of the visible part of the universe remains nearly fixed. But unlike a black hole's horizon, each observer has a slightly different horizon—each observer sits at the center of a spherical region of the universe whose edge is the horizon. And also unlike the case of a black hole, if the expansion of the universe changes dramatically, this horizon starts to grow, and what was until then hidden behind the horizon slowly starts to become visible again.

Because in such a universe we can send signals beyond the horizon (we just cannot receive any response to them once they have crossed the horizon) the region of the universe that can be influenced by a particular observer grows larger and larger even though, paradoxically, the “visible” universe hardly seems to grow at all. Putting the universe into this peculiar state is the essential idea behind the theory of *inflation*.

A universe composed entirely of matter and radiation behaves in exactly the opposite way. Whenever the light from some distant object reaches us, thereafter it remains always visible to us and it never slips beyond the horizon. Over time, we see ever more and more of the universe. However, despite seeming so sensible, such a universe has one disturbing property, especially if the universe had a beginning. If we can look sufficiently far back in time (which for a finite signal speed just means looking far away in space), we shall see parts of the universe which could not have been in causal contact at those earlier times, when the light that we are seeing today was first emitted. A signal can propagate only so far since the beginning, and today we can view far more of the universe than was visible at those times.

The idea of inflation is to combine these two phases, to first have the universe pass through an inflationary phase which is then followed by a more conventional radiation and matter-dominated phase. Structures can then be hidden behind the horizon during the inflationary expansion to allow more of the universe to have been causally connected at earlier times than would otherwise have been the case. The elegance of this picture is that the peculiarities of one phase help to resolve a paradox of the other.

I. The background

To begin, we need to describe the expansion of the universe. In these notes, we shall not consider the most general expansion possible, but rather one that is suggested by the universe as we observe it, with a few further assumptions along the way to simplify the analysis. Observations of the early universe, especially those of the relic cosmic microwave background radiation, show great spatial uniformity. So as a first approximation, let us imagine that the universe expands uniformly with time, without any spatial dependence at all. Such a universe is described by the metric,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) d\vec{x} \cdot d\vec{x}. \quad (8)$$

The *scale factor* $a(t)$ encodes how space inherently stretches or contracts over time. We can also write this metric in an alternate form by defining a *conformal time* coordinate through

$$\eta(t) = \int \frac{dt}{a(t)}, \quad (9)$$

in terms of which the metric becomes even simpler—the flat space (Minkowski) metric multiplied by an overall time-dependent factor,

$$ds^2 = a^2(\eta) [dt^2 - d\vec{x} \cdot d\vec{x}]. \quad (10)$$

The conformal time has an important physical significance. Let us restore an explicit factor of the speed of light c for the moment. If we consider a null path where $ds^2 = 0$ and look at the radial propagation $r(t)$ of this signal,

$$ds^2 = 0 = c^2 dt^2 - a^2(t) dr^2, \quad (11)$$

we then find that

$$r(t) = c \int_{t_1}^{t_2} \frac{dt}{a(t)}. \quad (12)$$

Up to the constant factor c , this distance is exactly the same as the conformal time η . So we see that the conformal time actually describes the maximal causal distance between t_1 and t_2 , since we derived $r(t)$ by assuming that the local speed of the signal was always equal to c .

The action that determines the evolution of the universe can be divided into two parts. One depends solely on the metric,

$$S_g = \frac{1}{16\pi G} \int d^4x \sqrt{-g} R, \quad (13)$$

while the other, S_ϕ , contains any non-gravitational ingredients of our theory. The *energy-momentum tensor* corresponds to the variation of this second piece with respect to the metric,

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta S_\phi}{\delta g^{\mu\nu}}. \quad (14)$$

Varying the total action $S_g + S_\phi$ yields Einstein's equation,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu}. \quad (15)$$

In general, this expression can contain up to ten separate equations; but for very simple backgrounds—those with much symmetry—most of these components either vanish or are redundant. For example, in the coordinates that we have introduced above, only the time (tt) and the space (ij) components contain any unique information,

$$R_{tt} = -3\frac{\ddot{a}}{a}, \quad R_{ij} = [a\ddot{a} + 2\dot{a}^2]\delta_{ij}, \quad R = -6\frac{\ddot{a}}{a} - 6\frac{\dot{a}^2}{a^2}. \quad (16)$$

To complete the picture, we must include some material to provide the energy-momentum that drives the expansion, described by the time-evolution of $a(t)$.

Inflation corresponds to the very general idea that the universe underwent a stage of *accelerated expansion*, when $\ddot{a} > 0$, during some early epoch. There are in practice many ways to implement this basic idea. Here we shall look only at the simplest, where the accelerated expansion is produced by a single scalar field ϕ with the appropriate properties. We start with a general action with a kinetic term and a potential $V(\phi)$,

$$S_\phi = \int d^4x \sqrt{-g} \left[\frac{1}{2}g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right]. \quad (17)$$

Because the background space-time is approximately spatially uniform, the field should also be more or less independent of the position, $\phi = \phi(t)$. For a general metric the energy-momentum tensor is

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2}g_{\mu\nu}g^{\lambda\rho} \partial_\lambda \phi \partial_\rho \phi + g_{\mu\nu}V(\phi); \quad (18)$$

but in our case the only nonvanishing components are

$$T_{tt} = \frac{1}{2}\dot{\phi}^2 + V(\phi), \quad T_{ij} = a^2\delta_{ij} \left[\frac{1}{2}\dot{\phi}^2 - V(\phi) \right]. \quad (19)$$

We have one further equation, the equation of motion obtained by varying this action with respect to ϕ itself,

$$\nabla^2 \phi + \frac{\delta V}{\delta \phi} = 0; \quad (20)$$

in our spatially independent case,

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \frac{\delta V}{\delta\phi} = 0. \quad (21)$$

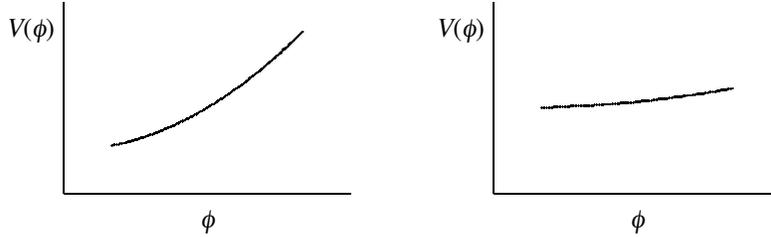
So far we have not specified the potential at all. But if inflation is to produce a period of accelerated expansion, this potential must satisfy a few conditions. The time (tt) and space (ij) components of Einstein's equation read

$$\begin{aligned} 3\frac{\dot{a}^2}{a^2} &= 8\pi G \left[\frac{1}{2}\dot{\phi}^2 + V \right] \\ -2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} &= 8\pi G \left[\frac{1}{2}\dot{\phi}^2 - V \right]. \end{aligned} \quad (22)$$

Using the first of these equations to eliminate the \dot{a}^2/a^2 term in the second, we find

$$\frac{\ddot{a}}{a} = \frac{8\pi G}{3} [V - \dot{\phi}^2]. \quad (23)$$

Since $a > 0$, an accelerating universe requires a kinetic energy that is small in comparison with the potential energy, $V > \dot{\phi}^2$. However, by itself this condition is not sufficient, since it is not enough to have inflation last only for a short while. We need to maintain the inflationary era long enough that a tiny causally connected patch of the early universe grows to something of the order of the size of the visible universe today. The problem is that if we merely start with $\dot{\phi} \approx 0$ at some point, then if the field can convert its potential energy into kinetic energy quickly enough, the $V > \dot{\phi}^2$ condition will be violated before enough inflation has occurred.



As an analogy, consider a marble, whose position is given by ϕ , rolling on a slope, whose height is given by $V(\phi)$. If the slope is very steep, as in the figure on the left, it will speed up too quickly as it rolls down the slope. Something much more level, as in the figure to the right, would be better. If it starts with $\dot{\phi} \approx 0$ at some time, it will only very gradually pick up its speed

as it rolls along. Because of this analogy, many inflationary models are said to have a “slowly rolling” phase during which most of the actual inflation occurs.

The ideal case is a universe with a perfectly constant, positive potential, $V(\phi) = V_0 > 0$. Such a universe is called *de Sitter space*. In de Sitter space, the universe expands at a constant rate too,

$$H(t) \equiv \frac{\dot{a}}{a} = \sqrt{\frac{8\pi G}{3} V_0} . \quad (24)$$

Solving for the behavior of the scale factor $a(t)$, we see that space expands very rapidly in de Sitter space, in fact at an exponential rate, $a(t) \propto e^{Ht}$.

Actually, this ideal is too much of a good thing. For subtle reasons explained later, de Sitter space cannot produce the *primordial perturbations* to the background space-time that are needed as an initial input for a more conventional evolution of the universe. How far a particular model differs from this ideal is encoded by two dimensionless “slow-roll” parameters, ε and δ , defined by

$$\varepsilon \equiv \frac{d}{dt} \frac{1}{H}, \quad \delta \equiv \frac{1}{H} \frac{\ddot{\phi}}{\dot{\phi}} . \quad (25)$$

The slow-roll conditions assume $\varepsilon \ll 1$ and $\delta \ll 1$. The first condition implies that the expansion rate is not changing much (in de Sitter space it does not change at all, as we have seen) while the second implies that the field is not accelerating too quickly, which would otherwise allow the kinetic energy to grow too large before enough inflation had occurred.

Although de Sitter space is ultimately not suitable for describing our universe, it is still a good approximation to models that can produce the needed primordial perturbations. Just what these primordial perturbations are, and how they are generated in inflation, will be explained next.

II. A baby version: fluctuations in de Sitter space

So far we have been describing the universe as though it were always perfectly spatially uniform. Certainly, this is a very poor approximation today. Yet, even if it was a better approximation at earlier times, the universe could never have been perfectly uniform. Even at the very earliest of times there must have been some tiny fluctuations in the shape of space-time and in the distribution of the material—the matter and radiation—spread throughout the universe. Over time, as matter was attracted to the regions where the pull of gravity was slightly stronger and away from the places where it was a tiny bit weaker, more and more dense regions would form, which in turn could exert a still stronger pull on the matter surrounding it. With enough time, these regions would become sufficiently dense to be able to collapse into stars and galaxies.

So, beginning with only just a tiny amount of spatial variation in an otherwise highly uniform universe, we could arrive at the universe as it appears to us today.

If we correctly understand the behavior of gravity as well as the contents of the universe, we can reverse this picture and trace the growth of these structures backward in time. We could then ask how large the original, or *primordial*, *perturbations* would need to have been to produce the properties of the universe today. If we do so, we find that a spatial variation of about one part in one hundred thousand, relative to the uniform part, would have been enough. We could also try to infer the size of these perturbations by what would have been needed to generate the acoustic oscillations observed in the cosmic microwave background radiation, and the answer is of the same size.

In inflation, the mechanism for generating these primordial perturbations relies on the quantum properties of the field. A quantum field is always fluctuating. Just as for a quantum harmonic oscillator, a quantum field satisfies a Heisenberg uncertainty relation. Having a perfectly uniform field at any particular time would require knowing its spatial properties to an infinite precision, which would mean in turn that we could know nothing about its momentum, which is its time derivative. One important difference between a simple harmonic oscillator and a quantum field is that the latter depends on the space-time location where it is evaluated. So in terms of creation and annihilation operators, a quantum field at first sight looks like an infinite set of harmonic oscillators, one for each value of the momentum, \vec{k} ,

$$[a, a^\dagger] = 1 \quad \Rightarrow \quad [a_{\vec{k}}, a_{\vec{k}'}^\dagger] = (2\pi)^3 \delta^3(\vec{k} - \vec{k}'). \quad (26)$$

In the last section, we treated the field ϕ as though it were purely classical. That is still a good approximation for the early universe, as far as we can observe its properties, so let us add a small quantum mechanical piece to the uniform behavior analyzed before,

$$\phi(t, \vec{x}) \rightarrow \phi(\eta, \vec{x}) = \phi_0(\eta) + \varphi(\eta, \vec{x}). \quad (27)$$

In this expression, the term $\phi_0(\eta)$ is still a classical field and is identical with what we called simply $\phi(t)$ in the previous section; $\varphi(\eta, \vec{x})$ is a quantum field. Notice that we have now switched to the conformal time; we do so since some of the equations that we shall encounter appear a little simpler in this coordinate. In general relativity it is not consistent to add a spatially dependent piece in the field without also doing so in the metric—since the field itself is the source for the gravitational field, and any of its spatial dependence must necessarily spill over to the curvature of space-time. However, because the metric $g_{\mu\nu}$ is a tensor and depends on how we have chosen our coordinates, the treatment of a small spatial variation of the metric is far more complicated. Therefore we shall not perturb the metric at all in this section.

We shall make one further assumption before deriving what inflation predicts for the initial spatial dependence of the universe. We shall work in the de Sitter limit. The potential $V(\phi)$ is then a constant V_0 as is the expansion rate, $H(\eta)$, defined by

$$H = \frac{\dot{a}}{a} = \frac{a'}{a^2}. \quad (28)$$

Since H is a constant, we can use this equation to solve for the conformal time dependence of the scale factor $a(\eta)$,

$$H = \frac{1}{a^2} \frac{da}{d\eta} \quad \Rightarrow \quad \frac{da}{a^2} = H d\eta, \quad (29)$$

to find

$$a(\eta) = -\frac{1}{H\eta} \quad (30)$$

In these coordinates, η does not range over the entire real line, but only from $-\infty$ (the far past) to 0 (the far future). So η will always be negative.

Because the potential is constant, the $\delta V/\delta\phi$ term in the equation of motion for ϕ vanishes, and the equation becomes just $\nabla^2\phi = 0$. This equation holds for both its classical and its quantum parts separately,

$$\nabla^2\phi_0 = 0 \quad \text{and} \quad \nabla^2\varphi = 0. \quad (31)$$

We have already studied the classical part, so let us look at the quantum part. In conformally flat coordinates, this equation becomes more generally,

$$\varphi'' + 2aH\varphi' - \vec{\nabla} \cdot \vec{\nabla}\varphi = 0, \quad (32)$$

while substituting the form of $a(\eta)$ appropriate for de Sitter space, we find,

$$\varphi'' - \frac{2}{\eta}\varphi' - \vec{\nabla} \cdot \vec{\nabla}\varphi = 0. \quad (33)$$

At this stage we still have a differential equation that depends on all four coordinates. We can reduce it to an equation of a single variable by Fourier expanding the field in plane waves for the spatial part, $e^{i\vec{k}\cdot\vec{x}}$. $\varphi(\eta, \vec{x})$ is a quantum field—that is, an operator—and not a classical function. So when we Fourier expand the field, we include a creation ($a_{\vec{k}}^\dagger$) or an annihilation operator ($a_{\vec{k}}$) associated with creating or annihilating one of these plane-waves,

$$\varphi(\eta, \vec{x}) = \int \frac{d^3\vec{k}}{(2\pi)^3} [\varphi_{\vec{k}}(\eta)e^{i\vec{k}\cdot\vec{x}}a_{\vec{k}} + \varphi_{\vec{k}}^*(\eta)e^{-i\vec{k}\cdot\vec{x}}a_{\vec{k}}^\dagger], \quad (34)$$

where $k = |\vec{k}|$ is the magnitude of the spatial momentum.

As we mentioned, aside from the additional label for the momentum, the operators $a_{\vec{k}}$ and $a_{\vec{k}}^\dagger$ resemble the creation and annihilation operators that occur when we quantize a harmonic oscillator. Here the meaning of having a separate oscillator for each value of the momentum becomes a little clearer. Note that only operators associated with the same momenta do not commute with each other,

$$[a_{\vec{k}}, a_{\vec{k}'}^\dagger] = (2\pi)^3 \delta^3(\vec{k} - \vec{k}'), \quad (35)$$

while

$$[a_{\vec{k}}, a_{\vec{k}'}] = 0 = [a_{\vec{k}}^\dagger, a_{\vec{k}'}^\dagger]. \quad (36)$$

There is also a ground state $|0\rangle$. It is annihilated by $a_{\vec{k}}$, for any momentum \vec{k} ,

$$a_{\vec{k}}|0\rangle = 0, \quad (37)$$

and similarly its dual $\langle 0|$ is annihilated by $a_{\vec{k}}^\dagger$, which is the dual of $a_{\vec{k}}$,

$$\langle 0|a_{\vec{k}}^\dagger = (a_{\vec{k}}|0\rangle)^\dagger = 0. \quad (38)$$

The ground state is normalized to satisfy $\langle 0|0\rangle = 1$.

There remains only the time-dependent part of the eigenmode expansion $\varphi_k(\eta)$ to explain. Once we have solved for this function, we shall completely understand what inflation predicts for the spatial variation—at least in this simplified example, which is actually not so very different from the result of a more careful analysis. Inserting the Fourier expansion for the field $\varphi(\eta, \vec{x})$ into its equation of motion yields a differential equation for the eigenmodes,

$$\varphi_k'' - \frac{2}{\eta} \varphi_k' + k^2 \varphi_k = 0, \quad (39)$$

The general solution of this equation is given by

$$\varphi_k(\eta) = A_k(i - k\eta)e^{-ik\eta} + B_k(i + k\eta)e^{ik\eta}. \quad (40)$$

To specify the solution completely still requires two further conditions to fix the constants of integration, A_k and B_k .

Since the field $\varphi(\eta, \vec{x})$ is a quantum mechanical object, it satisfies a relation very similar to the Heisenberg relation for the position (\hat{X}) and momentum (\hat{P}) operators for a harmonic oscillator, $[\hat{X}, \hat{P}] = i$, except that here too we must account for the spatial dependence of the field,

$$[\hat{X}, \hat{P}] = i \quad \Rightarrow \quad [\varphi(\eta, \vec{x}), \pi(\eta, \vec{y})] = i\delta^3(\vec{x} - \vec{y}). \quad (41)$$

This equation describes the *canonical commutation relation* between φ and its conjugate momentum π , defined in the usual way that should be familiar from Hamiltonian dynamics,

$$\pi \equiv \frac{\delta \mathcal{L}}{\delta \varphi'} = a^2 \varphi'. \quad (42)$$

Here we have used the Lagrangian \mathcal{L} implied by the action S_ϕ ,

$$\mathcal{L} = \frac{1}{2} a^2 \phi'^2 - \frac{1}{2} a^2 \vec{\nabla} \phi \cdot \vec{\nabla} \phi. \quad (43)$$

The δ -function just means that fields commute except when they are evaluated at the same point.

This commutation relation will provide our first condition on the modes $\varphi_k(\eta)$ which in turn will allow us to fix one of the constants of integration. To find this condition, substitute the Fourier expansion for φ and for π ,

$$\pi(\eta, \vec{x}) = a^2(\eta) \varphi'(\eta, \vec{x}) = a^2(\eta) \int \frac{d^3 \vec{k}}{(2\pi)^3} [\varphi'_k(\eta) e^{i\vec{k} \cdot \vec{x}} a_{\vec{k}} + \varphi_k'^*(\eta) e^{-i\vec{k} \cdot \vec{x}} a_{\vec{k}}^\dagger], \quad (44)$$

into the commutator and apply the properties of the creation and annihilation operators. In the end, we arrive at the following condition,

$$a^2 [\varphi_k \varphi_k'^* - \varphi_k'^* \varphi_k] = i. \quad (45)$$

For our general solution for the modes, this condition becomes,

$$A_k A_k^* - B_k B_k^* = \frac{H^2}{2k^3}. \quad (46)$$

At first this result might seem only a minor technical point, but it is actually quite important. It sets the overall size of $\varphi_k(\eta)$ and consequently the amplitude of the spatial fluctuations, which for the space-time fluctuations of the next section is what is measured experimentally.

Applying this condition eliminates one of the constants and only one more remains, which we shall now write as f_k ,

$$\varphi_k(\eta) = \frac{1}{\sqrt{1 - f_k f_k^*}} \frac{H}{\sqrt{2k^3}} [(i - k\eta) e^{-ik\eta} + f_k (i + k\eta) e^{ik\eta}]. \quad (47)$$

The second condition which we shall use to fix f_k is more of an assumption about what we think should be true of the state during inflation. The reasoning that underlies this assumption is safe enough for the setting that we are using throughout this section — that of a quantum field in a classical space-time

background—but it becomes a little more doubtful in a fully consistent treatment where the space-time has quantum fluctuations too. Such a setting can be dangerously close to the regime of quantum gravity, where we do not have an adequate theoretical understanding for how nature behaves. However, in a classical background, only the field has quantum properties, and these we more or less understand.

The curvature of de Sitter space is

$$R = -12H^2, \quad (48)$$

and because H is constant, this value for the scalar curvature means that the curvature in de Sitter space is everywhere the same. We can use this property to apply what we know about quantum fields in flat space to fix the remaining constant of integration f_k for this field in an expanding space-time.

At very small distances, the fact that this space is curved should not be much apparent. By a small distance, we mean here an interval that is small in comparison with the natural scale defined by the curvature, R . R contains two derivatives, so it scales as a squared inverse-length. Thus the natural lengths where the curvature of de Sitter space becomes very apparent are those that are at least an appreciable factor of $1/H$. If we analyze the field mode by mode, $\varphi_k(\eta)$, the wavelength associated with a particular mode function is

$$\lambda \sim \frac{2\pi}{k}, \quad (49)$$

which means that the small-wavelength modes, compared with radius of curvature of de Sitter space, $\lambda \ll 1/H$, are consequently those for which

$$k \gg H. \quad (50)$$

Actually, it is not enough to look only at small wavelengths. The scale factor affects the time coordinate as well. So for the moment, let us return once more to the ordinary time coordinate t . If we normalize the scale factor so that $a(t) = e^{Ht}$, then we can go between the two coordinate systems by just equating $a(\eta)$ with $a(t)$,

$$\eta = -\frac{e^{-Ht}}{H}. \quad (51)$$

Substituting this expression into $\varphi_k(\eta(t))$ yields,

$$\varphi_k(t) = \frac{H}{\sqrt{1-f_k^* f_k}} \left[\left(i + \frac{k}{H} e^{-Ht} \right) \frac{e^{i\frac{k}{H} e^{-Ht}}}{\sqrt{2k^3}} + f_k \left(i - \frac{k}{H} e^{-Ht} \right) \frac{e^{-i\frac{k}{H} e^{-Ht}}}{\sqrt{2k^3}} \right]; \quad (52)$$

its more complicated form helps to explain why we have been using η rather than t for most of this section.

Just as we should only look at short distances—that is, short wavelengths—to be able to ignore the effects of the curvature, we should similarly only look over short time intervals as well. Let us therefore look at times only near to $t = 0$, so that $Ht \ll 1$ and we can expand the exponential, $e^{-Ht} = 1 - Ht + \dots$. Of course, there is nothing special about $t = 0$; we could just have easily chosen another value $t = t_0$ by translating our coordinates to $t \rightarrow t - t_0$ and then expanding near t_0 instead, $H(t - t_0) \ll 1$. But if we expand near $t = 0$ and use also $k \gg H$, the largest time-dependent terms are

$$\varphi_k(t) = \frac{1}{\sqrt{1 - f_k f_k^*}} \left[e^{ik/H} \frac{e^{-ikt}}{\sqrt{2k}} - f_k e^{-ik/H} \frac{e^{ikt}}{\sqrt{2k}} \right] + \dots \quad (53)$$

We now must make use of one of the properties of quantum field theory in flat space. In flat space, the positive energy eigenmodes for a massless scalar field are

$$\frac{e^{-ikt}}{\sqrt{2k}}. \quad (54)$$

To understand why these are the positive energy modes, recall that in our Fourier expansion for the field, $\varphi_k(t)$ is accompanied by the factor $e^{i\vec{k}\cdot\vec{x}}$, so their product in this limit should be a plane-wave in flat space,

$$\varphi_k(t) e^{i\vec{k}\cdot\vec{x}} \rightarrow e^{-ikt} e^{i\vec{k}\cdot\vec{x}} = e^{-i[k^0 x^0 - \vec{k}\cdot\vec{x}]} = e^{-ik_\mu x^\mu}; \quad (55)$$

remember that in special relativity, the 0th component of the momentum is the energy, E . When a field is massless $m = 0$, as it is here since the potential $V(\phi)$ is constant (and not $\frac{1}{2}m^2\phi^2$), the energy is just equal to the magnitude of its three-momentum,

$$k^0 \equiv E = \sqrt{|\vec{k}|^2 + m^2} = k. \quad (56)$$

If we assume that over small distances and time intervals, where de Sitter space looks much like flat space, the field's behavior should similarly match with that of a field in flat space, then the modes $\varphi_k(t)$ should approach

$$\lim_{\substack{Ht \rightarrow 0 \\ k/H \rightarrow \infty}} \varphi_k(t) \rightarrow \frac{e^{-ikt}}{\sqrt{2k}}, \quad (57)$$

up to an arbitrary phase. This condition fixes the remaining constant of integration to be $f_k = 0$. We now know everything about the quantum part of the

field in this simple setting,

$$\varphi_k(\eta) = \frac{H}{\sqrt{2k^3}}(i - k\eta)e^{-ik\eta}, \quad (58)$$

where we have returned to the conformal time.

One problem, which might have already been apparent, is that while inflation treats the spatial variations of the field quantum mechanically, whenever we measure some real feature of our universe—the patterns in the cosmic microwave background radiation or how galaxies are distributed—we are always measuring entirely classical quantities. We shall therefore need to convert $\varphi(\eta, \vec{x})$ into something classical. One way to obtain classical functions is to take the expectation values of quantum operators. We could take the expectation value of the field in the ground state, but it vanishes

$$\langle 0|\varphi(\eta, \vec{x})|0\rangle \int \frac{d^3\vec{k}}{(2\pi)^3} \left[\varphi_k(\eta)e^{i\vec{k}\cdot\vec{x}}\langle 0|a_{\vec{k}}|0\rangle + \varphi_k^*(\eta)e^{-i\vec{k}\cdot\vec{x}}\langle 0|a_{\vec{k}}^\dagger|0\rangle \right] = 0, \quad (59)$$

since $a_{\vec{k}}|0\rangle = 0$ and $\langle 0|a_{\vec{k}}^\dagger = 0$. Actually, this result ought not to be too surprising; $\varphi(\eta, \vec{x})$ describes the fluctuations about the spatially constant part $\phi_0(\eta)$; so, on average, these fluctuations should be zero.

Let us go to the next simplest expectation value, that of two fields at different places,

$$\langle 0|\varphi(\eta, \vec{x})\varphi(\eta, \vec{y})|0\rangle. \quad (60)$$

If $\langle 0|\varphi|0\rangle$ is the functional analogue of the mean of a distribution, $\langle 0|\varphi\varphi|0\rangle$ is like its variance. It is often called the *two-point correlation function* and we shall see that it does not vanish. It provides the simplest description of the spatial fluctuations produced by inflation.

If we substitute the Fourier-expanded form for the fields into this expectation value, and use the properties of the creation and annihilation operators when they act on the ground state, we find that this two-point function is

$$\begin{aligned} \langle 0|\varphi(\eta, \vec{x})\varphi(\eta, \vec{y})|0\rangle &= \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{d^3\vec{p}}{(2\pi)^3} \varphi_k(\eta)e^{i\vec{k}\cdot\vec{x}} \varphi_p^*(\eta)e^{-i\vec{p}\cdot\vec{y}} \langle 0|a_{\vec{k}}^\dagger a_{\vec{p}}|0\rangle \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{d^3\vec{p}}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x} - i\vec{p}\cdot\vec{y}} \varphi_k(\eta)\varphi_p^*(\eta)(2\pi)^3 \delta^3(\vec{k} - \vec{p}) \\ &= \int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x} - \vec{y})} \varphi_k(\eta)\varphi_k^*(\eta). \end{aligned} \quad (61)$$

When experiments measure the properties of how galaxies are distributed in the universe or the pattern of temperature fluctuations in the microwave

background radiation, the results are often described in terms of the Fourier transform (or in terms of their spherical analogue, spherical harmonics). The Fourier transform of the two-point function is sometimes normalized in a convenient form, called the *power spectrum* of the distribution that we are measuring. Here we shall define the power spectrum with some of the factors of the momentum extracted,

$$\langle 0|\varphi(\eta, \vec{x})\varphi(\eta, \vec{y})|0\rangle = \int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})} \frac{2\pi^2}{k^3} P_k^\varphi(\eta), \quad (62)$$

but other conventions are often used as well. From our derivation of the two-point function, we can immediately see what inflation predicts for the power spectrum of the primordial perturbations,

$$P_k^\varphi(\eta) = \frac{k^3}{2\pi^2} \varphi_k(\eta) \varphi_k^*(\eta) = \frac{k^3}{2\pi^2} \frac{H^2}{2k^3} (1 + k^2 \eta^2) = \frac{H^2}{4\pi^2} (1 + k^2 \eta^2) \quad (63)$$

using what we derived for $\varphi_k(\eta)$.

We have one final fact to include if we are to select which values of $k\eta$ are the important ones for cosmological measurements. Remember from the very beginning, inflation is essentially a mechanism for hiding information behind a horizon. Or equivalently, inflation makes the size of a causally connected part of the universe much larger than that part of it that can be observed. In de Sitter space, the distance from an observer to this horizon is $1/H$, so anything that inflation has stretched larger than this size remains hidden until inflation ends. Thereafter, we return to a phase where radiation and then matter govern how the universe evolves. In this more conventional phase, the range of what we can see begins to grow faster than the rate at which the causally connected region grows. All the perturbations that were hidden behind the horizon during inflation are revealed as soon as the size of the horizon catches up to the physical wavelength of a particular fluctuation. So smaller wavelength fluctuations re-emerge first, while larger ones only enter later.

The *physical* wavelength of a fluctuation is stretched along with the expansion of the universe. So the physical size of a particular Fourier mode $\varphi_k(\eta)$ should include a factor of the scale factor, $a(\eta)\lambda$, since $a(\eta)$ tells by how much space stretches in any of its directions over time. Similarly, the physical momentum scales as $k/a(\eta)$. If inflation is meant to hide fluctuations beyond the horizon, the relevant wavelengths for cosmological experiments are those whose wavelengths are much larger than the horizon by the end of inflation, $a(\eta)\lambda \gg 1/H$, or

$$\frac{k}{a(\eta)} \ll H. \quad (64)$$

Using $a(\eta) = -1/H\eta$, we learn that the momenta appropriate for observations are those for which

$$\frac{k}{a(\eta)} \ll H \quad \Rightarrow \quad -k\eta \ll 1. \quad (65)$$

The power spectrum in this regime is very nearly constant,

$$P_k^\varphi(\eta) \approx \frac{H^2}{4\pi^2}. \quad (66)$$

Although we mentioned from the early stages of this section that this example is only an idealization of what is meant to happen in an inflationary model, we have learned at least one important property of the initial spatial fluctuations produced by such a model. The power spectrum, at the cosmologically important scales k , is entirely featureless; that is, since k corresponds to roughly the inverse size of a particular fluctuation, primordial fluctuations should exist on all scales that we can observe, and with the same amplitude.

Unfortunately, to obtain the full predictions for these primordial perturbations, we shall need to analyze the response of the gravitational background to the fluctuations in the field. What we have found here are only the quantum fluctuations of some scalar field, but not the distortions of space-time itself. This is an important distinction; everything feels the influence of gravity, so fluctuations of space-time will directly influence how matter and radiation will behave. The calculation of the true primordial perturbations is far more complicated than the example here, although the results for both have much in common. But if we wish to know the amplitude of these perturbations, we shall need to leave a purely de Sitter universe and consider one where the potential for the field ϕ is no longer perfectly flat.

III. A short sketch of the true primordial perturbations

This final section summarizes, in an extremely abbreviated form, the true calculation for the power spectrum of the primordial perturbations predicted by inflation. The reason that we have not included the full details is that the general set of *scalar perturbations*—those that eventually lead to the density perturbations in matter—can be represented in the metric as four functions, and expanding the full action out to quadratic order in these perturbations becomes quite tedious. Moreover, some of these functions are redundant, corresponding to particular ways for choosing the perturbed coordinates about the simple background that we have been considering. One of the tenets of general relativity is that how we choose our coordinates does not contain any genuine physical information. In a complete calculation, this property becomes apparent in the fact that some of these functions have no dynamical effect—they

appear in the action only as total derivatives. To reach this point, however, requires quite a bit of labor.

The reason that we do mention the true calculation at all—and did not simply conclude with the last section—is that without a proper treatment, we would miss some of the important predictions of inflation. One of these is for the *amplitude* of the power spectrum. The experimental result, as we mentioned earlier, is that there should be fluctuations in an otherwise uniform background at the level of one part in one hundred thousand. So far we have been calculating the fluctuations of a scalar field in an expanding background, but we have not yet connected these fluctuations with the actual fluctuations of the space-time background itself. We shall do so now.

Let us start by including some quantum fluctuations in the metric. We shall not write them in the most general form; but instead, we shall choose our coordinates so that the fluctuating part of the metric is still diagonal,

$$\begin{aligned} g_{\eta\eta}(\eta, \vec{x}) &= a^2(\eta)[1 + 2\Phi(\eta, \vec{x})] \\ g_{\eta i}(\eta, \vec{x}) &= 0 \\ g_{ij}(\eta, \vec{x}) &= -a^2(\eta)\delta_{ij}[1 - 2\Psi(\eta, \vec{x})]. \end{aligned} \quad (67)$$

As before, we shall also include a small quantum part of the field, except that this time we shall call it $\delta\phi$ rather than φ (which we shall reserve for something closely related, and which becomes $\delta\phi$ in the de Sitter limit),

$$\phi(\eta, \vec{x}) = \phi_0(\eta) + \delta\phi(\eta, \vec{x}). \quad (68)$$

All that we need to do, then, is to substitute these expressions for the metric and the field into the complete action for the theory,

$$S = \int d^4x \sqrt{-g} \left\{ \frac{1}{16\pi G} R + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right\}. \quad (69)$$

Since the fluctuations are small, terms that contain more powers of the fluctuations—either Φ , Ψ or $\delta\phi$ —will have less of an influence on predictions than terms with fewer powers. The terms with no factors of these fields obviously will not tell us anything about how they behave; instead, such terms govern the behavior of the background— $a(\eta)$ and $\phi_0(\eta)$. If we look at the terms with just one factor of these fields, these too will not be of much use since all these terms vanish upon substituting in the equations of motion of the background. Therefore, to be able to understand the fluctuations, and thus to be able to refine our prediction from the last section, we shall need to expand the action S to quadratic order in the perturbations. We shall neglect terms beyond this order, not because they are ultimately unimportant, but rather because

their influence on the two-point function—which is, after all, itself a quadratic quantity in the fields—is negligible compared with the current experimental precisions.

Despite many complicated intermediate steps, the final expression for the quadratic terms is remarkably simple. After applying a constraint that emerges in the course of the calculation (to eliminate Φ) and neglecting terms that can be written as total spatial or time derivatives, all of the dynamically influential terms in the action can be written succinctly in terms of a linear combination of two of the fields,

$$\varphi = \delta\phi + \frac{\phi'_0}{aH}\Psi, \quad (70)$$

as

$$S^{(2)}[\varphi] = \int d^4x a^2 \left\{ \frac{1}{2}\varphi'^2 - \frac{1}{2}\vec{\nabla}\varphi \cdot \vec{\nabla}\varphi - \frac{1}{2}a^2 m^2 \varphi^2 \right\} + \dots \quad (71)$$

This action essentially describes a free, massive scalar field in an expanding background, although the mass m is not a constant. In terms of the slow-roll parameters, the exact expression for this mass is

$$m^2(\eta) = -H^2(3 + \delta)(\varepsilon + \delta) - aH^3(\varepsilon' + \delta'); \quad (72)$$

for the rest of this section, however, we shall assume that the universe is expanding in a slowly rolling phase of inflation. There, the parameters ε and δ are small and approximately constant, so it is sufficient to retain only

$$m^2(\eta) \approx -3H^2(\varepsilon + \delta) \quad (73)$$

in the expression for the mass.

The reason that de Sitter space is only an unrealistic idealization now becomes apparent when we look at the relative amounts of the two fields that compose φ . In the de Sitter limit, the spatially constant part of the field, ϕ_0 , is also constant in time; so $\phi'_0 = 0$. Thus, in de Sitter space $\varphi = \delta\phi$, and our connection with the fluctuations with the space-time are lost. As a consequence, it would have done us no good at all to have included fluctuations of the metric in the last section. It is important always to have the field roll at least a little bit, $\phi'_0 \neq 0$.

We have chosen φ so that—from the perspective of the quantum theory—it is a “canonically normalized” field, meaning that its kinetic term in the action has the standard normalization. This choice lets us immediately apply all the results of quantum field theory. For example, we can use the canonical commutation relation between φ and its conjugate momentum, as before, to fix the normalization of the allowed solutions for $\varphi_k(\eta)$. However, this normalization

is not the appropriate one when comparing with experiments. Instead we shall rescale the field so it is the term Ψ that has no prefactor,

$$\zeta = \Psi + \frac{aH}{\phi_0'} \delta\phi = \frac{aH}{\phi_0'} \varphi. \quad (74)$$

The reason for doing so is that Ψ , being a fluctuation of space-time, is something whose influence we can directly measure. Ψ is related to the fluctuations in the curvature of a three-dimensional slice of the universe at a particular time. For example, the three-dimensional metric along such a slice would be

$$g_{ij}^{(3)}(\eta, \vec{x}) = a^2(\eta) (1 - 2\Psi(\eta, \vec{x})) \delta_{ij}, \quad (75)$$

and its three-dimensional scalar curvature (which is completely different from the full space-time curvature R) is

$$R^{(3)} = \frac{4}{a^2} \vec{\nabla} \cdot \vec{\nabla} \Psi + \dots, \quad (76)$$

where we have kept only the leading term in powers of Ψ . In Fourier space, the proportionality is even more explicit since the right side of this equation becomes

$$-\frac{4}{a^2} k^2 \Psi_k(\eta) + \dots. \quad (77)$$

Thus, the correct way to calculate the power spectrum is first to find the appropriate solution to the equations of motion for φ as before, but in the end to evaluate the power spectrum of the field ζ ,

$$\langle 0 | \zeta(\eta, \vec{x}) \zeta(\eta, \vec{y}) | 0 \rangle = \int \frac{d^3 \vec{k}}{(2\pi)^3} e^{i\vec{k} \cdot (\vec{x} - \vec{y})} \frac{2\pi^2}{k^3} P_k^\zeta(\eta). \quad (78)$$

Since φ and ζ are so closely related, all that is needed is to calculate the power spectrum of φ as before and then rescale by the necessary factor as the final step,

$$P_k^\zeta(\eta) = \left(\frac{aH}{\phi_0'} \right)^2 P_k^\varphi(\eta). \quad (79)$$

This rescaling factor perhaps does not convey very much, so let us rewrite it in a more illuminating form. If we differentiate both sides of the time (tt) component of the background Einstein equation,

$$3H^2 = 8\pi G \left[\frac{1}{2} \dot{\phi}_0^2 + V \right] \Rightarrow 6H\dot{H} = 8\pi G \left[\ddot{\phi}_0 + \frac{\delta V}{\delta \phi_0} \right] \dot{\phi}_0 \quad (80)$$

and then use the definition for ε on the left side and the equation of motion for ϕ_0 on the right, we find

$$-6H^3 \varepsilon = 8\pi G[-3H\dot{\phi}_0]\dot{\phi}_0 \quad \Rightarrow \quad \frac{4\pi G}{\varepsilon} = \frac{H^2}{\dot{\phi}_0^2} = \left(\frac{aH}{\dot{\phi}_0'}\right)^2, \quad (81)$$

and thus the power spectrum of $\zeta(\eta, \vec{x})$ can also be written as

$$P_k^\zeta(\eta) = \frac{4\pi G}{\varepsilon} P_k^\varphi(\eta). \quad (82)$$

Here G is just the Newton constant and the slow-roll parameter ε is a small number.

The calculation of the power spectrum of φ now follows exactly the same steps as in the de Sitter example of the previous section. The equation of motion implied by the action $S^{(2)}[\varphi]$ for the Fourier modes $\varphi_k(\eta)$ is

$$\varphi_k'' + 2aH\varphi_k' + (k^2 + a^2m^2)\varphi_k = 0, \quad (83)$$

where the modes $\varphi_k(\eta)$ are defined by the same operator expansion as before,

$$\varphi(\eta, \vec{x}) = \int \frac{d^3\vec{k}}{(2\pi)^3} [\varphi_k(\eta)e^{i\vec{k}\cdot\vec{x}} a_{\vec{k}} + \varphi_k^*(\eta)e^{-i\vec{k}\cdot\vec{x}} a_{\vec{k}}^\dagger]. \quad (84)$$

In the slowly rolling limit, $aH = -(1 + \varepsilon)/\eta$, so the equation of motion becomes approximately

$$\varphi_k'' - \frac{2(1 + \varepsilon)}{\eta}\varphi_k' + k^2\varphi_k - \frac{3(\varepsilon + \delta)}{\eta^2}\varphi_k = 0. \quad (85)$$

To put this differential equation into one of the standard forms, let us define a dimensionless variable $z = -k\eta$ and rescale the field by $\varphi_k(\eta) = z^{\frac{3}{2} + \varepsilon} Z_\nu(z)$. Then the equation of motion becomes

$$\frac{d^2 Z_\nu}{dz^2} + \frac{1}{z} \frac{dZ_\nu}{dz} + \left[1 - \frac{\nu^2}{z^2}\right] Z_\nu = 0, \quad (86)$$

where the index ν in the slowly rolling limit is

$$\nu = \left(\frac{9}{4} + 6\varepsilon + 3\delta\right)^{1/2} \approx \frac{3}{2} + 2\varepsilon + \delta. \quad (87)$$

Again, these details might appear to be merely technical; however, the fact that this index is not exactly equal to $\frac{3}{2}$ is directly responsible for the fact that—for a slowly rolling model—the power spectrum *does* slightly depend on the scale, which seems to agree with what can be inferred from observations.

The differential equation for $Z_\nu(z)$ is, in fact, Bessel’s equation, so $Z_\nu(z)$ must be a linear combination of Bessel functions. Applying the canonical commutation relation and choosing the modes to match the flat space ground state modes over short intervals as before provides two constraints to fix the two constants of integration. Applying both of these conditions yields

$$\varphi_k(\eta) = \frac{\sqrt{\pi} H(-\eta)^{3/2}}{2(1+\varepsilon)} H_\nu^{(1)}(-k\eta). \quad (88)$$

$H_\nu^{(1)}(-k\eta)$ corresponds to a particular linear combination of Bessel functions called a *Hankel function*. Thus, the power spectrum for $\varphi(\eta, \vec{x})$ is

$$P_k^\varphi(\eta) = \frac{k^3}{2\pi^2} \varphi_k(\eta) \varphi_k^*(\eta) = \frac{H^2 (-k\eta)^3}{8\pi (1+\varepsilon)^2} |H_\nu^{(1)}(-k\eta)|^2. \quad (89)$$

A power spectrum expressed in terms of Hankel functions might not seem at first to be especially illuminating. However, as we explained in the last section, the physically important scales k for cosmological observations are those for which $-k\eta \rightarrow 0$, which, combined with the smallness of the parameters ε and δ hidden in the index of the Hankel function, allows us to write the expression for the power spectrum of these modes in a far simpler form,

$$P_k^\varphi(\eta) = \frac{H^2}{4\pi^2} (-k\eta)^{-4\varepsilon-2\delta} + \dots \quad (90)$$

Before converting this result into an inflationary prediction for the power spectrum of the rescaled field $\zeta(\eta, \vec{x})$, let us first define an *amplitude*, $\Delta_\zeta^2(k_0)$, and a *tilt*, n_s , by writing the power spectrum in the following “standard” form,

$$P_k^\zeta(\eta) = \Delta_\zeta^2(k_0) \left(\frac{k}{k_0}\right)^{n_s-1}, \quad (91)$$

which is obviously strongly influenced by the expectations of inflation. Here we have defined the amplitude with respect to a particular reference scale k_0 .¹ From our calculations, we find

$$P_k^\zeta(\eta) = \frac{4\pi G}{\varepsilon} \frac{H^2}{4\pi^2} (-k\eta)^{-4\varepsilon-2\delta} + \dots, \quad (92)$$

¹For the WMAP experiment, for example, k_0 was chosen to be 2 Gpc^{-1} .

so we learn that the amplitude and the tilt are

$$\Delta_{\zeta}^2(k_0) \approx \frac{4\pi G}{\varepsilon} \frac{H^2}{4\pi^2} \quad \text{and} \quad n_s \approx 1 - 4\varepsilon - 2\delta, \quad (93)$$

if we neglect the higher-order corrections of the slowly rolling limit.

The case where $n_s = 1$ would be a power spectrum like the one in the last section which has equal power at all scales. Since ε and δ are small and *positive*, the spectrum is still nearly independent of the scale at which it is evaluated, but it should show a slight tendency for the fluctuations at large scales to have larger amplitudes than those at smaller scales.

To conclude, let us rewrite the amplitude once more by introducing a natural mass scale for gravity. In the system of units that we have been implicitly using, where c and \hbar are neglected, the Newton constant can be expressed as a mass scale,

$$M_{\text{pl}}^2 = \frac{1}{8\pi G} \quad (94)$$

called the (reduced) *Planck mass*. In terms of this scale, the inflationary model that we have been studying predicts the following amplitude for the primordial fluctuations,

$$\Delta_{\zeta}^2(k_0) = \frac{1}{8\pi^2} \frac{1}{\varepsilon} \frac{H^2}{M_{\text{pl}}^2}. \quad (95)$$

This form is useful since it expresses the amplitude in terms of a dimensionless ratio between the expansion rate during the inflationary era, H , and the natural interaction strength of gravity, M_{pl} .