Early Universe: Inflation and the generation of fluctuations

I. PROBLEMS OF THE STANDARD MODEL

A. The horizon problem

We know from CMB observations that the universe at recombination was the same temperature in all directions to better than 10^{-4}. Unless the universe has been able to smooth itself out dynamically, this means that the universe must have started in a very special initial state, being essentially uniform over the entire observable universe. Furthermore if we look at the small fluctuations in the temperature we see correlations on degree scales: there are 10^{-4} hot and cold spots of larger than degree size on the sky. Since random processes cannot generate correlations on scales larger than the distance light can travel, these fluctuations must have been in causal contact at some time in the past. This is telling us something very interesting about the evolution of the early universe, as we shall now see.

Let's investigate the distance travelled by a photon from the big bang to recombination and then from recombination to us. A photon follows a null trajectory, $ds^2 = 0$, and travels a proper distance cdt in interval dt. Setting $c = 1$ the comoving distance travelled is $dr = dt/a \equiv d\eta$, where we have defined η, the *conformal time*. The comoving distance light can travel from the big bang to a time t, the *particle horizon*, is therefore

$$\int dr = \eta = \int_0^t \frac{dt'}{a(t')}. \tag{1}$$

So starting at a given point, light can only reach inside a sphere of comoving radius η_* by the time of recombination t_*. *If* we assume the early universe contained only matter and radiation we can solve for η_*, about 300Mpc. So we only expect to see correlations over patches with comoving radius smaller than about 300Mpc at recombination.

What angular size would we expect to see this area on the sky? Today the comoving distance from recombination is $\sim \eta_0 \sim 14000$Mpc, so we would only expect correlations over an angular radius[1] $\sim 300/14000 \sim 1^\circ$. No causal process could have generated the correlations seen on larger scales!

We must have made an incorrect assumption. The most likely suspect is the assumption that the universe only contained matter and radiation in the early universe: perhaps some different content could change the evolution at early times so that $\eta_* \gg 300$Mpc, as required to generate correlations on large scales? For a constant equation of state

$$\eta \propto \int t^{-2/(3+3w)} dt, \tag{2}$$

so large conformal times can be obtained if $2/(3+3w) \geq 1$, or $w < -1/3$. If this is the case very large distances could have been in causal contact. If $w \sim -1$ ($a \sim e^{Ht}$, $H \sim$ const) for some early-universe epoch so that $\ddot{a} > 0$, this is termed *inflation*, the most popular solution to the horizon problem. We use the term *hot big bang* to describe the time after inflation (usually taken to be $t = 0$) at which the normal radiation-dominated universe started.

A useful rule of thumb is that the universe approximately doubles in size in a 'Hubble time' $\sim 1/H$. The Hubble distance $1/H$ sets approximately the distance light can travel in this time. A scale larger than $\gtrsim 1/H$ is therefore expanded faster by the expansion than the extra distance light can travel in an expansion time, and so becomes out of causal contact with itself. Physical scales grow with the expansion, $\lambda_{\rm phys} \propto a \sim e^{Ht}$, so in inflation where $H \sim$ const every scale eventually becomes larger

---

[1] Why are we allowed to calculate angles using comoving lengths and conformal time as though it were flat space? Basically because for light $d\chi = d\eta$, so on a spacetime diagram with comoving distance against conformal time, light rays are straight lines at 45°. More generally, for light (null geodesics) $ds^2 = g_{\mu\nu} dx^\mu dx^\nu = 0$, hence under a change of metric $g_{\mu\nu} \to \Omega g_{\mu\nu}$ we still have $ds^2 = 0$; so the flat (conformal) FRW metric $ds^2 = a^2[d\eta^2 - d\mathbf{x}^2]$ gives the same results as the Minkowski metric $ds^2 = d\eta^2 - d\mathbf{x}^2$.

than $1/H$ and goes out of causal contact with itself, when $\lambda_{\rm phys} \gtrsim H^{-1}$. This is what is required to solve the horizon problem: going back in time from the hot big bang, a physical scale $\lambda_{\rm phys}$ shrunk rapidly, and was inside the horizon (in causal contact with itself) at some earlier point in inflation.

Equivalently in terms of comoving length, a comoving scale $\lambda = \lambda_{\rm phys}/a$ goes out of causal contact with itself when $\lambda \sim (aH)^{-1}$, the comoving Hubble length. During inflation $H \sim$ const but $a \sim e^{Ht}$ grows rapidly, so $(aH)^{-1}$ is rapidly shrinking. In one Hubble time $\Delta t = H^{-1}$ light travels the comoving distance

$$\Delta\eta = \frac{\Delta t}{a} = \frac{1}{aH}, \tag{3}$$

so a comoving scale λ is in causal contact with itself when $\lambda \lesssim (aH)^{-1}$ (termed *inside the horizon*), and out of causal contact with itself when $\lambda \gtrsim (aH)^{-1}$ (termed *outside the horizon*).

Any solution to the horizon problem must have some time in the past (prior to recombination) at which all relevant scales have $\lambda < (aH)^{-1}$ so they are in causal contact. An alternative to inflation would be to have some contracting phase before the hot big bang, so that there is a time in the contracting phase when the observable universe was all in causal contact. Inflation solves the horizon problem because a grows rapidly and H is nearly constant, so that early in inflation $(aH)^{-1}$ was very large and then shrank rapidly. In a pre-big bang contracting universe light can travel a large comoving distance during the contracting phase, and the horizon problem is also solved. In the pre-big bang a can be decreasing, but if $|H|$ shrinks rapidly enough it's still possible to have $\lambda < |aH|^{-1}$ early in the contracting phase and also $\lambda > |aH|^{-1}$ as the big bang is approached, making it in some ways remarkably similar to inflation; such models have been called *ekpyrotic* models (see e.g. astro-ph/0404480) and can be motivated as resulting from the collision of 4-D 'branes' colliding in a higher-dimensional space. However the contracting universe has the disadvantage of being much harder to study, since it requires knowing how to go through $a = 0$ without singularities. We shall not discuss it further here, and focus on the more popular model of inflation.

Ex: It is often useful to use conformal time η rather than t as a time variable. Defining the conformal Hubble parameter $\mathcal{H} \equiv a'/a$, where a dash denotes conformal time derivative, show that

$$\mathcal{H} = aH \tag{4}$$

$$\mathcal{H}^2 + K = \frac{\kappa}{3} a^2 \rho \tag{5}$$

$$\mathcal{H}^2 - \mathcal{H}' + K = \frac{\kappa}{2} a^2 (\rho + P) \tag{6}$$

where $\kappa \equiv 8\pi G$ ($= 1/M_P^2$ where M_P is the reduced Planck mass). Also show that in radiation domination $a \propto \eta$ and in matter domination $a \propto \eta^2$.

B. The flatness problem

Recall the Friedmann equation

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho, \tag{7}$$

and the critical density ρ_c defined so that

$$H^2 = \frac{8\pi G}{3}\rho_c. \tag{8}$$

In general $\rho = \Omega\rho_c$, with $\Omega \neq 1$ if the universe is not flat, and Ω is a function of time.

The curvature radius of the FRW universe $R_K^2 \equiv a^2/|K|$ is given by

$$R_K^{-2} = \frac{|K|}{a^2} = \left|\frac{8\pi G\rho}{3} - H^2\right| = H^2|\Omega - 1|, \tag{9}$$

so the ratio of the Hubble radius to the curvature scale is determined by Ω:

$$\frac{H^{-1}}{R_K} = |\Omega - 1|^{1/2}. \tag{10}$$

Today, the energy density of the universe is close to the critical density, $\Omega \sim 1$ (observations measure this to about 1%). Equivalently, the curvature radius is significantly larger than the current Hubble radius. Is this surprising?

Let's consider how the relative energy density $\Omega(t)$ evolves as a function of time when the background equation of state is constant. We shall assume $\Omega(t) \sim 1$, and see if that is stable, so we can use the flat-space result for $a(t)$ (Eq. (**??**)), which gives $H \propto 1/t \propto a^{-(3+3w)/2}$ so that

$$\Omega(t) - 1 = \frac{K}{a^2H^2} \propto Ka^{1+3w}. \tag{11}$$

During matter and radiation domination $|\Omega(t)-1|$ grows, and hence had to be much smaller in the past to give a small value today. If for example $\frac{H^{-1}}{R_K} \sim 1$ at some early epoch, it would be enormous today, not small as observed. Why was the universe initially so flat? We have seen that $|\Omega(t) - 1| \propto a^{1+3w}$. So if the universe was at some point in the past dominated by a fluid with $w < -1/3$, then $|\Omega(t) - 1|$ was growing smaller instead of larger: the $\Omega(t) = 1$ is a stable fixed point. If this phase persisted for long enough it would explain the observed flatness: if inflation started in a universe that was slightly non-flat, the exponential expansion expands the curvature scale to be much larger than the current horizon, so the universe we see is very flat.

C. The monopole problem

Phase transitions are often associated with symmetry breaking. In a supersymmetric model, all running coupling constants seem to reach the same magnitude at about 10^{16}GeV, suggesting a unification of the gauge groups $U(1) \times SU(2) \times SU(3) \to G$. If this is indeed the case, and if G is a simple group, then a general theorem states that monopoles are formed when G is broken into subgroups containing $U(1)$. This is what happens in GUT (grand unification) theories when the universe cools below the critical temperature T_c at which the unification happens.

The characteristic mass of the monopoles is the critical temperature of the phase transition, $m_M \approx T_c \sim 10^{16}$GeV. We expect generically to form about one monopole per Hubble volume, so their number density is

$$n_M(T_c) \sim H(T_c)^3 \sim g_*^{3/2}\frac{T_c^6}{m_P^3}. \tag{12}$$

Using the entropy density $s \sim g_*T_c^3$ we find

$$\frac{n_M}{s} \sim \left(\frac{T_c}{m_P}\right)^3. \tag{13}$$

As the monopoles are "hidden" from each other by their associated gauge field, they freeze out very rapidly and no monopole-antimonopole annihilation occurs. Their contribution to the total energy density in the universe is about

$$\rho_M(t_0) = \frac{m_M}{m_b}\frac{n_M}{s}\frac{s}{n_b}m_b n_b \tag{14}$$

and therefore $\rho_M/\rho_b \sim 10^{16} \cdot 10^{-12} \cdot 10^{10} \sim 10^{14}$ – far too much mass in monopoles compared to baryons to be consistent with observation.

This problem can be avoided by postulating that there is no unification of the gauge groups to a simple group. Or there could be a rapid expansion: in this case all the matter is diluted as $1/a^3$, so a sufficiently large increase in the scale factor can get rid of all the monopoles. All the other matter and radiation is also redshifted away, but the energy in the expansion is released at the end of the "inflationary" phase (a process called reheating), generating the radiation and baryons that we see. Afterwards, the thermal history can proceed as in the standard model. But if $T_c > T_{\rm reheat} \gg m_x$ where x is the normal kind of matter, then the monopoles are no longer a problem.

II. INFLATION WITH A SCALAR FIELD

We have seen that a period of inflationary expansion with $w < -1/3$ (negative pressure, $P < -\rho/3$) can solve the monopole, flatness and horizon problem. A cosmological constant ($w = -1$) would solve the problems too, but then inflation never ends and the "normal" evolution of the standard model never starts. We need a dynamical mechanism to create the negative pressure.

Fortunately there is a very simple possibility, namely a scalar field. Scalar fields represent spin-0 particles in field theories, for example the Higgs field in the standard model. Mathematically a scalar field is just a scalar function of space and time, usually written $\phi(x,t)$. We will continue to assume our universe is approximately homogeneous, so that the background field depends only on time, $\phi(t)$. If the universe expands very rapidly, this is self-consistent if inflation starts with a small Hubble-radius sized patch which is homogeneous and grows to contain the observable universe. The existence of such a smooth initial patch is an assumption of this simple model. Unfortunately the Higgs field itself cannot drive observationally-consistent inflation in the simplest models, so the 'inflaton' field is usually assumed to be some other as-yet-unidentified scalar field.

A. Field theory

The action in a field theory is given by the integral of the Lagrangian[2] L, for example in 4-dimensions by

$$S = \int d^4x L. \tag{15}$$

The Lagrangian is in general a function of all the relevant fields (and their derivatives). The classical field equations minimize the action. So if $L = L(\psi, \partial_\mu \psi)$ and ψ_c is a classical solution we expect $\delta S[\psi]|_{\psi=\psi_c} = 0$. If we expand about the classical solution $\psi = \psi_c + \delta\psi$ we have

$$\begin{aligned} S &= \int d^4x L[\psi_c + \delta\psi, \partial_\mu(\psi_c + \delta\psi)] && (16)\\ &= \int d^4x \left\{ L_c + \delta\psi \frac{\partial}{\partial\psi} L + \partial_\mu(\delta\psi) \frac{\partial}{\partial(\partial_\mu\psi)} L \right\} + \mathcal{O}(\delta\psi^2) && (17)\\ &= \int d^4x \left\{ L_c + \delta\psi \frac{\partial L}{\partial\psi} - (\delta\psi)\partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu\psi)} \right) \right\} + \mathcal{O}(\delta\psi^2) && (18)\\ &= S_c + \int d^4x \delta\psi \left\{ \frac{\partial L}{\partial\psi} - \partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu\psi)} \right) \right\} + \mathcal{O}(\delta\psi^2). && (19) \end{aligned}$$

where we integrated by parts assuming boundary terms vanish. The action is extremised when the $\delta\psi$ term is zero for all $\delta\psi$, in other words when the Euler-Lagrange equations are satisfied

$$\partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu\psi)} \right) = \frac{\partial L}{\partial\psi}. \tag{20}$$

---

[2] Perhaps more properly called the *Lagrangian density*

This is the equation that tells us the ψ that minimizes the action, which is the solution to the classical field equation.

1. 1D example

This is all analogous to what happens in 1D, where for example for $L = T - V = \frac{1}{2}\dot{q}^2 - V(q)$ (kinetic minus potential energy) we have

$$\partial_t \left(\frac{\partial L}{\partial \dot{q}} \right) = \frac{\partial L}{\partial q} \tag{21}$$

$$\implies \ddot{t} = -\frac{dV(q)}{dq}, \tag{22}$$

i.e. the acceleration is given by the force (the gradient of the potential).

Ex: A ball is falling directly downwards near the bottom of a cylinder with parabolic cross-section, so the height is $y = \frac{1}{2}\alpha^2 x^2$. It has kinetic energy $\frac{m}{2}(\dot{x}^2 + \dot{y}^2) \approx \frac{m}{2}\dot{x}^2$ for small x, and the potential is $V = my$ for some parameter α. Write down the Lagrangian and solve the Euler-Lagrange equations for the equation of motion. What would a friction term look like in the Lagrangian?

2. Action in GR

For completeness I'll give the results for GR here, though we only use them once or twice. In general relativity the measure becomes $d^4x\sqrt{-g}$ where g is the determinant of the metric tensor; this combination is invariant under coordinate transforms. The action[3] is

$$S = \int d^4x \sqrt{-g} \left\{ -\frac{R}{16\pi G} + \mathcal{L}_m \right\} \tag{23}$$

where $\mathcal{L}_m$ is the lagrangian for any matter fields and R is the Ricci scalar. This is consistent because the Euler-Lagrange equations for the elements of the metric in this action give the Einstein equations (see e.g. *Carroll: Spacetime and Geometry*)

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G\, T_{\mu\nu}, \tag{24}$$

where the stress-energy tensor is given in terms of the Lagrangian by

$$T^{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\partial \left(\mathcal{L}_m \sqrt{-g} \right)}{\partial g_{\mu\nu}} = -2\frac{\partial \mathcal{L}_m}{\partial g_{\mu\nu}} - g^{\mu\nu}\mathcal{L}_m. \tag{25}$$

The full system of equations of motion for this action contains also the corresponding Euler-Lagrange equations for every field that appears in the matter lagrangian $\mathcal{L}_m$. For the last step above we used the derivative of the determinant, which can be calculated for a matrix A using

$$\ln|A| \quad = \quad \mathrm{Tr}(\ln A) \implies |A|^{-1} \frac{d|A|}{dA_{ij}} = \mathrm{Tr}\left(A^{-1} \frac{dA}{dA_{ij}} \right) = A^{-1}_{pq} \frac{dA_{qp}}{dA_{ij}} = A^{-1}_{ji} \tag{26}$$

$$\implies \frac{d|A|}{dA_{ij}} = |A| A^{-1}_{ji}. \tag{27}$$

The result $\ln|A| = \mathrm{Tr}(\ln \mathrm{A})$ follows trivially if $A_{ij} = \lambda_i \delta_{ij}$ is diagonal, since then $\ln|A| = \ln\prod_i \lambda_i = \sum_i \ln\lambda_i = \mathrm{Tr}(\ln \mathrm{A})$; since determinant and trace are rotationally invariant, it also holds more generally.

---

[3] Beware that in this and the next section signs can move around in different signature conventions.

3. Scalar field equations

The Lagrangian of a minimally-coupled[4] scalar field is given by

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi), \tag{28}$$

where $V(\phi)$ is a potential. This has the usual 'kinetic energy' - 'potential energy' form in a homogeneous universe, and generalizes the 'kinetic' term to include gradient energy when there are perturbations. We can calculate the stress-energy tensor from Eq. (25) by writing Eq. (28) as

$$\mathcal{L}_\phi = \frac{1}{2}g^{\mu'\nu'}\partial_{\mu'}\phi\partial_{\nu'}\phi - V(\phi). \tag{29}$$

We need to take the derivative of $g^{\mu'\nu'}$. Remembering that $g_{\mu\nu'}g^{\nu'\nu} = \delta_\mu{}^\nu$ we have

$$(\delta g_{\mu\nu'})g^{\nu'\nu} + g_{\mu\nu'}\delta g^{\nu'\nu} = 0 \quad \Longrightarrow \quad \delta g^{\mu\nu} = -g^{\nu\nu'}g^{\mu\mu'}\delta g_{\mu'\nu'}, \tag{30}$$

and hence

$$\frac{\partial g^{\mu'\nu'}}{\partial g_{\mu\nu}} = -g^{\nu\nu'}g^{\mu\mu'}. \tag{31}$$

The stress-energy tensor is therefore

$$T^{\mu\nu} = -2\frac{\partial\mathcal{L}_m}{\partial g_{\mu\nu}} - g^{\mu\nu}\mathcal{L}_m = \partial^\mu\phi\partial^\nu\phi - g^{\mu\nu}\left(\frac{1}{2}\partial_{\nu'}\phi\partial^{\nu'}\phi - V\right), \tag{32}$$

or lowering an index

$$T^\mu{}_\nu = g^{\mu\rho}\partial_\rho\phi\partial_\nu\phi - \delta^\mu{}_\nu\left(\frac{1}{2}\partial_\rho\phi\partial^\rho\phi - V\right). \tag{33}$$

In an isotropic and homogenous universe the density and pressure can then be determined from $\rho = T^0{}_0$, $T^i{}_j = -P\delta^i{}_j$, for $i, j = 1..3$. In general $T^i{}_j$ can have off-diagonal components, corresponding to anisotropic stresses; $T^0{}_i$ measures the heat flux, in the fluid context proportional to the fluid velocity.

B. Equations for Inflation

For the moment we only consider the case of an FRW universe filled with a scalar field. For a homogeneous field (i.e. $\partial_i\phi = 0,\ i = 1, 2, 3$)

$$\mathcal{L}_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{34}$$

and the energy density and pressure are given from Eq. (33) by

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{35}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{36}$$

---

[4] That is precisely the definition of *minimal coupling to gravity*: that the only place in which the elements of the metric appear in the matter Lagrangian $\mathcal{L}_m$ is a *kinetic term*, and that there are no metric elements within $V(\phi)$.

Why is the potential contribution to the pressure negative? If $\dot{\phi} = 0$, the scalar field is just a cosmological constant, with $P = -\rho$. Expanding a box of cosmological constant generates more volume of constant energy density, which must be put in by working to expand the box - i.e. the fluid has a negative pressure (opposite to what one expects from a gas, where the energy density dilutes as the gas expands).

As we don't know which particle is responsible for inflation we do not know the form of the potential. A few simple examples would be

$$V(\phi) = \frac{1}{2}m^2\phi^2 \qquad \text{massive scalar field} \tag{37}$$

$$V(\phi) = \frac{1}{2}\lambda(\phi^2 - \sigma^2)^2 \quad \text{Higgs-type potential} \tag{38}$$

$$V(\phi) = \frac{1}{2}\lambda\phi^4 \qquad \text{self-interacting scalar field} \tag{39}$$

The joint evolution of the metric and the scalar field can be read from the Euler-Lagrange equations (20) of the action for GR (23) and the scalar field Lagrangian (34) as the matter Lagrangian. The resulting equations of motion for the metric and the scalar field are, respectively, the Friedman equations and the couple Klein-Gordon equation for the scalar field (using M_P for the reduced Planck mass $M_P = m_P/\sqrt{8\pi} = 1/\sqrt{8\pi G}$):

$$H^2 = \frac{1}{3M_P^2}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right), \qquad \dot{H} = \frac{1}{M_P}\left(-\frac{1}{2}\dot{\phi}^2\right), \tag{40}$$

$$\ddot{\phi} + 3H\dot{\phi} = -\frac{dV(\phi)}{d\phi}. \tag{41}$$

Ex: Show that Eq. (41) also follows from using the energy-momentum conservation equation, $\dot{\rho} = -3H(\rho + P)$. Change variables to conformal time to show

$$\phi'' + 2\mathcal{H}\phi' + a^2 V_{,\phi} = 0, \tag{42}$$

where a prime denotes $d/d\eta$.

Let's take a look at the scalar field equation (41). We can immediately identify an acceleration, a *damping term* proportional to H and due to the expansion of the universe, and the force given by $-V_{,\phi}$ (from now on for brevity we will use the notation $V_{,\phi} \equiv dV/d\phi$, $V_{,\phi\phi} \equiv d^2V/d\phi^2$). Now let's imagine the scalar field as a ball rolling on a hill with some friction – the shape of the hill is determined by the potential and the friction is determined by the expansion rate H. Generically, if the field (ball) starts somewhere away from the minimum of the potential (up the hill) it will roll down towards the minimum.

Now let's impose a very simple condition on top of that situation: that the field is rolling down *slowly*, i.e. its kinetic energy is much smaller than its potential energy: $\frac{1}{2}\dot{\phi}^2 \ll V(\phi)$. In addition, in order for this condition to remain fulfilled for long enough, we will impose that the second derivative of the field, its *acceleration*, is also small, in particular with respect to the rest of the terms in eq. (41). These approximations, called the *slow-roll approximations*, are not strictly required for inflation to take place, but they are consistent and natural. The resulting *approximate* slow-roll Friedman and scalar field evolution equations are

$$H^2 \approx \frac{1}{3M_P^2}V(\phi), \qquad \dot{H} = \frac{1}{M_P}\left(-\frac{1}{2}\dot{\phi}^2\right), \tag{43}$$

$$3H\dot{\phi} \approx -\frac{dV(\phi)}{d\phi}. \tag{44}$$

and the equation of state for the scalar field is, approximately

$$P_\phi \approx -\rho_\phi, \tag{45}$$

which fulfills the desired inflationary condition $P < -\rho/3$. [NB: don't use this approximate equation of state when solving a problem: get the exact one from eqs. (35) and (36) instead, using eqs. (43) and (44), or use eq. (50) below.]

The universe inflates as the field is rolling down the hill, due to H being *approximately* constant:

$$H \approx \text{const} \quad \Longrightarrow \quad a(t) \sim e^{Ht}. \tag{46}$$

Notice that this is an approximate solution to the slow-roll Friedman equations above. Realistic potentials will produce different, but approximately exponential, solutions for $a(t)$.

Inflation will last for as long as the slow-roll conditions are fulfilled. In order to quantify that condition, and to relate it to the inflaton potential, let's parameterize the condition by expanding the potential in a Taylor series, i.e. in higher and higher derivatives of V. The first two so called *slow-roll parameters* are

$$\epsilon_V(\phi) \equiv \frac{M_P^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2, \tag{47}$$

$$\eta_V(\phi) \equiv M_P^2 \frac{V_{,\phi\phi}}{V}. \tag{48}$$

The first one measures the slope of the potential and the second one the curvature, both with respect to the value of the potential during inflation.

It's possible to express the equation of state with the help of ϵ_V. For that we notice that to first-order in the slow-roll approximation

$$\dot\phi^2 \approx \frac{V_{,\phi}^2}{9H^2} \approx M_P^2 \frac{V_{,\phi}^2}{3V} \approx \frac{2}{3}\epsilon_V V, \tag{49}$$

and therefore, to first order in ϵ_V,

$$P = \left(\frac{2}{3}\epsilon_V - 1\right)\rho. \tag{50}$$

Our old condition $w < -1/3$ translates into $\epsilon_V < 1$. So $\epsilon_V < 1$ is required for inflation, and inflation will end when $\epsilon_V = 1$. For $w \sim -1$, the slow-roll limit, we need $\epsilon_V \ll 1$. This is equivalent to $|\dot H|/H^2 \ll 1$, so the Hubble expansion is approximately constant.

For $3H\dot\phi \approx -V_{,\phi}$ to be consistent over time we need

$$3\dot H\dot\phi + 3H\ddot\phi \approx -V_{,\phi\phi}\dot\phi. \tag{51}$$

$$\Longrightarrow\ 3\frac{\ddot\phi}{H\dot\phi} \approx -\frac{V_{,\phi\phi}}{H^2} - 3\frac{\dot H}{H^2}. \tag{52}$$

Since we want $|\ddot\phi| \ll |V_{,\phi}| \approx |3H\dot\phi|$ and $|\dot H|/H^2 \ll 1$ this implies $|V_{,\phi\phi}/H^2| \ll 1$, or $|\eta_V| \ll 1$. In order for slow-roll to be self-consistent we therefore need that

$$\epsilon_V \ll 1 \quad \text{and} \quad |\eta_V| \ll 1. \tag{53}$$

In other words the inflation potential needs to be very flat with respect to its value during inflation. This can be realised in two different ways: either $V_{,\phi}$ is very small while V is sizeable, so the potential is actually very flat and the terms at both sides of eq. (44) are small (remember that according to eq. (43) $H^2 \propto V$); or the terms at both sides of eq. (44) are large, i.e. the slope $V_{,\phi}$ is large but inflation occurs with a largely damped field evolution, or, equivalently, at high values of V.[5] Getting a potential that is flat enough can be difficult in particle physics models.

---

[5] This generically corresponds to high values of ϕ. We call those models *large-field* inflation.

The very fact that in this slow-roll setting there is a natural way for inflation to end is quite remarkable: as we stated at the beginning of this section the lack of such stopping mechanism in the case of a cosmological constant is the main reason why we discarded it as an explanation, despite its dynamic behaviour being inflationary ($a(t) \propto e^{Ht}$, $H = \text{const}$).

C. The amount of inflation

The amount by which the universe inflates is measured as the number of *e-foldings* N, so the scale factor before and after inflation are related by

$$\frac{a(t_f)}{a(t_i)} = e^N. \tag{54}$$

By differentiating the equation above, we see that the number of e-foldings is just given by the integral of the expansion rate

$$N = \int_{a(t_i)}^{a(t_f)} \ln a = \int_{t_i}^{t_f} H dt. \tag{55}$$

We can rewrite this result with the help of the slow-roll equations (43) and (44),

$$H dt = H \frac{dt}{d\phi} d\phi = H \frac{d\phi}{\dot{\phi}} = -\frac{3H^2 d\phi}{V_{,\phi}} = -\frac{1}{M_P^2} \frac{V}{V_{,\phi}} d\phi. \tag{56}$$

so we finally find

$$N = -\frac{1}{M_P^2} \int_{\phi_i}^{\phi_f} \frac{V}{V_{,\phi}} d\phi. \tag{57}$$

Typically $N \gtrsim 40$–60 is large enough to solve the horizon and flatness problems, depending on the reheating temperature. If N is larger than this, on the one hand all the CMB sky would belong to a single causally connected patch, and on the other hand the universe is expected to be flat to within observational accuracy even if the curvature scale at the beginning of inflation was comparable to the Hubble radius.

D. Example: chaotic inflation with a quadratic potential

Let us now work out a particular example for an inflationary model. Given a particular potential parametrized by a few numbers (e.g. the mass of the scalar field), we should be able to express in terms of (1) those few numbers and (2) the initial conditions just before inflation happens (e.g. ϕ_i, $\dot{\phi}_i$, etc.), the theoretical predictions for the relevant cosmological observables (e.g. the *power spectrum*, that we will describe later), which will ultimately tell us about the likelihood that said model is the true one driving inflation.

The probably simplest example of inflation involves just a non-interacting massive scalar field with a potential $V(\phi) = m^2\phi^2/2$. In this case, the slow roll equations are

$$3H\dot{\phi} + m^2\phi = 0 \tag{58}$$

$$H^2 = \frac{m^2}{6M_P^2}\phi^2. \tag{59}$$

The slow-roll parameters, eq. (53), are

$$\epsilon_V = \eta_V = \frac{2M_P^2}{\phi^2}, \tag{60}$$

so inflation breaks down at $|\phi_f| \approx \sqrt{2} M_P$. If the field starts out initially at ϕ_i then the total amount of inflation will be

$$N \approx \frac{\phi_i^2}{4M_P^2} - \frac{1}{2}. \tag{61}$$

We therefore require $\phi_i \gg M_P$. On the other hand, we can ensure that $V \ll M_P^4$ by choosing the mass small enough, so we may not need to know worry too much about complications from new physics at Planck-scale energies (quantum gravity).

Notice that in eq. (61) we can see that the amount of e-folds depends on the value of the field, in particular on its initial value ϕ_i.

The slow-roll equations can also be solved, resulting in

$$\phi(t) = \phi_i - m M_P \sqrt{\frac{2}{3}} t \tag{62}$$

$$a(t) = a_i \exp\left\{ \sqrt{\frac{1}{6}} \frac{m}{M_P} \left(\phi_i t - \frac{m M_P}{\sqrt{6}} t^2 \right) \right\}. \tag{63}$$

Notice how the expansion is approximately exponential, as follows from the approximate slow-roll condition $H \approx$ const.

E. Reheating after inflation

During the inflationary period, any radiation and matter are redshifted away and the temperature drops rapidly to very low values. All the energy density is now stored in the inflaton field. We need to convert this energy back into particles and radiation if the universe is to continue with the normal hot big-bang evolution at the end of inflation.

Once slow-roll inflation breaks down, the inflaton accelerates, reaches the minimum of the potential and starts to oscillate around it.[6] At this point, the coupling to other fields becomes important. Originally, this decay was modeled by an additional dissipative term of the form $\Gamma \dot{\phi}$ in the evolution equation. As long as $\Gamma < H$ (i.e. during inflation) no particles are produced. As inflation ends, H decreases and the coupling leads to a relatively slow decay of the inflaton and to a relatively low temperature after inflation (less than maybe 10^{10} GeV).

However the process may be more complicated. When looking at a the coupling between a matter field χ and the inflaton, there can be a resonance for certain wavenumbers of χ, a phenomenon called *parametric resonance*. This can lead to much faster and more efficient transfer of energy from the inflaton field. An explosive reheating ensues, and the final energy density in matter and radiation will be similar to the one in the inflaton field at the end of inflation. This can lead to significantly higher reheating temperatures and to non-thermal distribution of the energy (since the resonant modes receive most of the energy). This version of reheating is commonly called *preheating*, and there seems to be still some controversy about the relative efficiency of the preheating and the normal (slower) reheating stage.

At any rate, the particles created during reheating are then supposed to decay, the particles and radiation start to interact and thermalize. The evolution of the universe re-enters the one of the standard hot big-bang model which we have discussed before.

---

[6] Notice that the Taylor expansion of a function around a minimum lacks the derivative term, since it is a critical point. In particular, for the potential around its minumum at ϕ_0: $V(\phi) \approx V(\phi_0) + V_{,\phi\phi}(\phi - \phi_0)^2$, so $-V_{,\phi} \propto -\phi$, and the evolution equation for the scalar field (41) turns into that of a damped harmonic oscillator. Notice also that the oscillating inflaton stops behaving as an *inflaton* as $P < -\rho/3$ is not fulfilled any more. Instead, the oscillating behaviour continuously transforms kinetic into potential energy and back, so from eq. (36) one sees that the pressure averages to 0 for fast enough oscillations, making the reheating inflaton effectively behave like a *matter field*, whilst it decays into relativistic species (radiation).

F. Inflation models

There are a very large number of inflation models in the literature, none of which is uniquely compelling. We just briefly mention some of the important classes:

1. Historical inflation models

In the old inflationary scenarios (the original inflation, proposed by Guth in 1981), the scalar field starts trapped in a "false vacuum" state, that is in a local minimum of the potential, with $V(\phi) > 0$. In this case bubbles of true vacuum (the ground state) nucleate by tunneling, ending inflation. Old inflation has been shown not to work, since it has what is called a "graceful exit" problem: The tunneling process creates bubbles of true vacuum, but they expand too slowly to ever fill the whole volume and the scenario does not lead to suitable reheating.

The graceful exit problem can be cured by postulating a second order phase transition (e.g. in a double-well potential). In this case, the field within a fluctuation domain is very homogeneous ($\nabla\phi \sim 0$), and if the "top" of the potential is sufficiently flat, the field will start to roll slowly ($\dot{\phi} \sim 0$). During this "slow-roll" phase, the universe inflates. Later on, the field accelerates, leading to a breakdown of inflation. It finally reaches the true vacuum and starts to oscillate. The coupling to other fields then leads to the creation of particles and to reheating.

2. Eternal inflation

There are various situations in which inflation may be eternal, in that parts of the universe continue to expand for ever, with an indefinite number of 'pocket' universes forming in which inflation ends and leading to an observationally-consistent universe. Since these pocket universes are mostly out of causal contact eternal inflation can be viewed as a 'multiverse' scenario, with lots of similar but mutually unobservable universes like our own.

New inflation can lead to eternal inflation, as different regions of space move away from $\phi = 0$ at different times.

Eternal 'chaotic' inflation happens in simple high-energy potentials (e.g. $V(\phi) = m^2\phi^2/2$) where quantum fluctuations in the field as it rolls down the potential lead to some regions of the universe actually moving *up* the potential, by $\Delta\phi \sim H$ (see later section on perturbation generation). For large values of H (so high V) the volume in regions moving up can be growing, leading to eternal inflation with an infinite number of separate regions ending inflation in different places, while other regions continuing to expand.

Eternal inflation can also happen when there are two or more false vacua, via nucleation of bubbles near the different vacuum that then undergo slow-roll inflation, leading to independent bubble universes. In this case the bubble universes are in fact open universes, though as long as there is enough inflation once they form the radius of curvature can be very large so that they look locally flat. This generalizes to 'string landscape' multiverse models, where each false vacuum corresponds to slightly different physics.

Learn more: astro-ph/0702178.

In almost all eternal inflation scenarios, the observable effects are determined by what happened when our particular pocket universe expanded and then ended inflation. In what follows we therefore focus on the inflation in our past rather than worrying about different things that may be happening in other unobservable regions.

3. Multi-field inflation

We have been assuming single-field inflation. But if there is one such field, why not more than one – or perhaps many? Some motivation for this might be the large number of effective fields that can appear in string theory models, and some of these models may be better theoretically motivated that simply writing down a single-field model out of thin air. However many of the strongest predictions

of single-field inflation that we will see later can break down if we allow more than one field, so these models can be rather less predictive; conversely any conflict of the simplest predictions found in the data might (none yet!) be an indication that we need to look at more complicated models.

III. COSMOLOGICAL PERTURBATION THEORY

We know from CMB observations that fluctuations must have been small in the early universe. We therefore use a linear approximation to lowest order in the perturbations: we keep only linear terms like $\delta\rho$ and consistently drop any terms of higher order, $(\delta\rho)^2$, etc. In general perturbations are a function of position and time so in the perturbed universe we might have $X(\mathbf{x},t) = X(t) + \delta X(\mathbf{x},t)$, where we split into a background (zero-order) part $X(t)$ that is homogeneous and so only depends on time, and $\delta X(\mathbf{x},t)$ that describes the perturbation. Since the background universe is assumed to be isotropic, all 3-space vectors must be first order or higher.

A. Mode decomposition

The rather accurate linear approximation leads to several big simplifications. The first is that any linear equation involving spatial derivatives can be Fourier-transformed, so for example an equation of the form

$$A\ddot{X} + B\boldsymbol{\nabla}^2 X = 0. \tag{64}$$

If X is first order, e.g. $X = \delta\rho$, then to first order A and B can be taken to be zeroth order ($\delta\rho\delta A \approx 0$), and hence are a function only of time t and not position $\mathbf{x}$. Hence the equation can be transformed using

$$X(\mathbf{x},t) = \frac{1}{(2\pi)^{3/2}} \int d^3\mathbf{k}\, X(\mathbf{k},t) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{65}$$

to becomes a set of evolution equations

$$A(t)\ddot{X}(\mathbf{k},t) - B(t)k^2 X(\mathbf{k},t) = 0 \tag{66}$$

for each wavevector $\mathbf{k}$ (where $k = |\mathbf{k}|$). These mode equations can easily be integrated separately for each k, and then at a later time Fourier-transformed back to get the solution to the original equation. Usually we parameterize position in comoving coordinates, so $\mathbf{k}$ is a *comoving* wavevector: a plane wave with fixed $\mathbf{k}$ expands with the universe, having a fixed comoving wavelength.

This further simplifies because in the above example the coefficients in the equation are only a function of $k = |\mathbf{k}|$, not $\mathbf{k}$, because of isotropy: i.e. all modes with the same k will evolve in exactly the same way. So we can write

$$X(\mathbf{k},t) = X(\mathbf{k},t_0)Y(k,t) \tag{67}$$

where $Y(k,t_0) = 1$ at some initial time, and we just need to solve

$$A(t)\ddot{Y}(k,t) - k^2 B(t) Y(k,t) = 0 \tag{68}$$

with that initial condition for each k. The solution $Y(k,t)$ is called the *transfer function*: it is the scaling that relates $X(\mathbf{k},t_0)$ at an initial time to $X(\mathbf{k},t)$ at a later time.

A further useful simplification due to linearity comes if we consider equations involving vectors or tensors. For example a vector equation like

$$A(t)\dot{\mathbf{v}}(\mathbf{x},t) + B(t)\mathbf{v}(\mathbf{x},t) + \boldsymbol{\nabla}\Psi(\mathbf{x},t) = 0 \tag{69}$$

could be Fourier transformed directly, but then the time evolution equation would have full $\mathbf{k}$ dependence:

$$A(t)\dot{\mathbf{v}}(\mathbf{k},t) + B(t)\mathbf{v}(\mathbf{k},t) + i\mathbf{k}\Psi(\mathbf{k},t) = 0. \tag{70}$$

Instead we could dot the equation with $\hat{\mathbf{k}} \equiv \mathbf{k}/k$ to project out the component along the direction of $\mathbf{k}$

$$A(t)\dot{v}(\mathbf{k},t) + B(t)v(\mathbf{k},t) - k\Psi(\mathbf{k},t) = 0, \tag{71}$$

where $v(\mathbf{k},t) = i\hat{\mathbf{k}}\cdot\mathbf{v}(\mathbf{k},t)$. Equivalently $k\,v(\mathbf{k},t)$ is the Fourier transform of $\boldsymbol{\nabla}\cdot\mathbf{v}$, where the factor of k keeps dimensions consistent. The scalar function v (or in real space $\boldsymbol{\nabla}\cdot\mathbf{v}$) is called the scalar mode component of $\mathbf{v}$.

If the velocity is determined in terms of a velocity potential $\mathbf{v} = \boldsymbol{\nabla}\phi_c$, then the scalar mode solution gives the full solution to the original equation (69). However $\mathbf{v}$ could also have components that cannot be written as $\boldsymbol{\nabla}\phi$, corresponding to the components of Eq. (70) perpendicular to $\hat{\mathbf{k}}$. If $\mathbf{e}_1$ and $\mathbf{e}_2$ are a basis orthogonal to $\mathbf{k}$ then taking either perpendicular component

$$A(t)\dot{v}^{(1)}(\mathbf{k},t) + B(t)v^{(1)}(\mathbf{k},t) = 0, \tag{72}$$

where $v^{(1)}(\mathbf{k},t)$ is either $\mathbf{e}_1\cdot\mathbf{v}(\mathbf{k},t)$ or $\mathbf{e}_2\cdot\mathbf{v}(\mathbf{k},t)$. The two modes $v^{(1)}$ are called the *vector mode* solutions of the equation. So we see that the original equation (69) can be split up into uncoupled scalar and vector mode equations, each of which gives simple transfer function solutions. Physically it is scalar mode velocities that are induced by linear gravitational forces, since the acceleration $\boldsymbol{\nabla}\Psi$ is purely scalar mode. The vector mode solution correspond to vortical solutions, which decay away in an expanding universe and are not generated by linear density perturbations.

This decomposition generalizes to tensor equations, for example involving a tensor h_{ij}. A general tensor can be decomposed into

$$h_{ij} = \frac{1}{3}\delta_{ij}h + h_{[ij]} + h_{\langle ij\rangle} \tag{73}$$

where

$$h_{\langle ij\rangle} \equiv h_{(ij)} - \frac{h}{3}\delta_{ij} \tag{74}$$

is the symmetric trace-free part and $h \equiv h^i_i$. By taking the trace, antisymmetric or symmetric trace-free (STF) part of a tensor equation we can generate three separate uncoupled equations. Since h is a scalar, it generates just scalar modes. The antisymmetric part $h_{[ij]}$ has three independent components, and so can be written $h_{[ij]} = \epsilon_{ijk}v^k$ for some vector $\mathbf{v}$, and so only has vector and scalar modes. The remaining intrinsically tensor part $h_{\langle ij\rangle}$ in general has scalar, vector and tensor mode components. For example the scalar mode component could be written in terms of a scalar as $h_{\langle ij\rangle} = \boldsymbol{\nabla}_{\langle i}\boldsymbol{\nabla}_{j\rangle}\phi$. Again the different modes decouple in linear theory.

Since the modes decouple at linear order, and density perturbations do not source vector or tensor modes, we shall focus mostly on scalar modes. However later we shall see that tensor modes of the metric are generated by inflation, which describe gravitational waves.

Ex: Show that

$$X^{\pm}_{ij} = \frac{1}{(2\pi)^{3/2}}\int d^3\mathbf{x}\, X(\mathbf{x},t)e^{\pm}_{\langle i}e^{\pm}_{j\rangle}e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{75}$$

is purely tensor mode (is symmetric, trace free, and orthogonal to $\hat{\mathbf{k}}$ on both indices), where $\mathbf{e}^{\pm} \equiv \mathbf{e}_1 \pm i\mathbf{e}_2$. Also that

$$X^{\pm}_{ij} = \frac{1}{(2\pi)^{3/2}}\int d^3\mathbf{x}\, X(\mathbf{x},t)e^{\pm}_{\langle i}\hat{\mathbf{k}}_{j\rangle}e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{76}$$

is vector mode and

$$X_{ij} = \frac{1}{(2\pi)^{3/2}}\int d^3\mathbf{x}\, X(\mathbf{x},t)\hat{\mathbf{k}}_{\langle i}\hat{\mathbf{k}}_{j\rangle}e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{77}$$

scalar mode.

B. Frames and gauges

In a perturbed universe, the decomposition into a background and perturbed part is not unique: we are in a relativistic theory so there are different possible ways to slice spacetime into a set of slices

with different constant t. This freedom is called a choice of *gauge*, basically the freedom to choose coordinates.

Alternatively we could describe the universe in terms of quantities that would be observed by a particular set of observers, for example observers at rest with respect to the matter or some other component. The world lines of these observers define a *threading*. The four-velocity of the observers $u^\mu = dx^\mu/d\tau$ define a *frame*, so the gauge choice is equivalent in most cases to a choice of frame: when discussing 3-vectors and energy density perturbations, which depend on the choice of frame (as in special relativity), we must specify which frame we are using in order not to be ambiguous. However we are free to choose any frame we like; popular choices are a frame in which the dark matter is at rest ($\mathbf{v}_c = 0$, called the 'synchronous gauge'), or the frame comoving with (at rest with respect to) the total energy density (the 'comoving gauge'). There are many other choices, for example choosing the frame so that the curvature or densities remains unperturbed. We shall use the terms *frame* and *gauge* interchangeably, since they are essentially the same thing as far as physical observables are concerned.

1. Gauge transformation example

Let's say we have a Lorentz scalar X, and defined δX_1 in one gauge. Now in another gauge with $t_2 = t_1 + \delta t$, where $\delta t(\mathbf{x}, t)$ is first order, we have to $\mathcal{O}(\delta^2)$

$$X = X(t_1) + \delta X_1(\mathbf{x}, t_1) \tag{78}$$

$$= X(t_2) + \delta X_2(\mathbf{x}, t_2) = X(t_1 + \delta t) + \delta X_2(\mathbf{x}, t_1). \tag{79}$$

Note that $\delta X(\mathbf{x}, t_1) = \delta X(\mathbf{x}, t_2)$ to linear order, since the difference is $\mathcal{O}(\delta t \delta X) = \mathcal{O}(2)$. Taylor expanding, to lowest order we see that

$$\delta X_2(\mathbf{x}, t) = \delta X_1(\mathbf{x}, t) - \dot{X}(t)\delta t(\mathbf{x}, t), \tag{80}$$

explicitly demonstrating that the perturbation changes when the coordinates are changed.

Consider a linearized Lorentz transformation by a linear velocity $\delta\mathbf{v}$ so that

$$\mathbf{x}_2 = \mathbf{x}_1 - \delta\mathbf{v}t_1 \qquad t_2 = t_1 - \delta\mathbf{v}\cdot\mathbf{x}_1, \tag{81}$$

and hence $\delta t \sim -\delta\mathbf{v}\cdot\mathbf{x}_1$. Since we are interested in small linear velocities, we have neglected terms of $\mathcal{O}(\delta\mathbf{v}^2)$, so $\gamma = (1 - \delta\mathbf{v}^2)^{-1/2} \approx 1$ (i.e. velocities are non-relativistic). Total energies and volumes change by γ-factors under the transformation, so they are unchanged under a first order Lorentz transformation. However density *perturbations* can change, since what we mean by the background density is different in different frames. If we take $X = \rho$, the density, we see that under the Lorentz transformation $\delta\rho$ changes by[7]

$$\dot{\rho}(t)\delta t(\mathbf{x}, t) \sim |3H\rho\,\delta\mathbf{v}\cdot\mathbf{x}| \sim |\rho\,\delta\mathbf{v}\cdot(H\mathbf{x})|. \tag{82}$$

For linear perturbations we expect $\mathcal{O}(|\delta\mathbf{v}|) = \mathcal{O}(\delta\rho/\rho)$, so this is much less than $\delta\rho$ if $|H\mathbf{x}| \ll 1$, i.e. $x \ll H^{-1}$: we are considering distances much smaller than a Hubble length. However for distances of the order or larger than a Hubble length, $|H\mathbf{x}| \gtrsim 1$, the change can be significant. In Fourier space this means that density perturbation modes with comoving wavenumber $k \gg (aH)$ will be nearly the same in any gauge related by a small linear velocity $\delta\mathbf{v}$. The choice of gauge only becomes important on super-horizon scales. However even in Newtonian theory velocities change with a change of frame, so on all scales it is still important to specify the frame used to describe velocities.

---

[7] We shouldn't in general do non-local Lorentz transformations like this in GR, but it gives the right idea for an order of magnitude argument. We give a more careful local argument in the next subsection.

2. Perturbation frame-dependence and frame-invariant variables

As above, consider a linear Lorentz transformation so that

$$\mathbf{x}_2 = \mathbf{x}_1 - \delta\mathbf{v}t_1 \qquad t_2 = t_1 - \delta\mathbf{v}\cdot\mathbf{x}_1, \tag{83}$$

corresponding to changing to a frame moving with velocity $\delta\mathbf{v}$. In general relativity, a Lorentz transformation can be applied locally to relate things that would be observed by observers with different relative velocities at the same point. An observer in frame 1 would calculate spatial gradients of $\rho = \rho(t_1) + \delta\rho_1(\mathbf{x}_1, t_1)$

$$\begin{aligned}
\frac{\partial\rho}{\partial x_1^i} &= \frac{\partial(\delta\rho_1)}{\partial x_1^i} && (84)\\
&= \frac{\partial x_2^\nu}{\partial x_1^i}\frac{\partial\rho}{\partial x_2^\nu} && (85)\\
&= \frac{\partial x_2^j}{\partial x_1^i}\frac{\partial\rho}{\partial x_2^j} + \frac{\partial t_2}{\partial x_1^i}\frac{\partial\rho}{\partial t_2} && (86)\\
&= \frac{\partial(\delta\rho_2)}{\partial x_2^i} - \delta v^i\dot{\rho}, && (87)
\end{aligned}$$

where in the last line we dropped second order terms. Taking the spatial divergence to get back to a scalar equation, this becomes

$$\begin{aligned}
\nabla^2\delta\rho_2 &= \nabla^2\delta\rho_1 + \dot{\rho}\nabla\cdot\delta\mathbf{v} && (88)\\
\implies \delta\rho_2 &= \delta\rho_1 - 3H(\rho+P)\nabla^{-2}\nabla\cdot\delta\mathbf{v}, && (89)
\end{aligned}$$

where we used $\dot{\rho} = -3H(\rho + P)$. This equation describes how density perturbations change when evaluated in different frames related by a relative 3-velocity $\delta\mathbf{v}$. The ∇^{-2} operator may look odd, but in harmonic space it is simply[8] $-a^2/k^2$.

To describe physics unambiguously on all scales we must be clear to specify which gauge (frame) any quantities are evaluated in, or use combinations of quantities that are the same in any frame ('gauge-invariant variables'). In what follows we shall mostly calculate things in some particular convenient frame. Beware that the equations may take different forms in different gauges.

Under the same local linear Lorentz transformation as above, velocities transform as $\mathbf{v}_2 = \mathbf{v}_1 - \delta\mathbf{v}$, and hence the combination

$$\bar{\delta\rho} = \delta\rho_1 - 3H(\rho+P)\nabla^{-2}(\nabla\cdot\mathbf{v}_1) = \delta\rho_2 - 3H(\rho+P)\nabla^{-2}(\nabla\cdot\mathbf{v}_2) \tag{90}$$

is the same in either frame: it is a frame-invariant quantity. If we work in a frame with $\mathbf{v} = 0$, then $\bar{\delta\rho} = \delta\rho$, so the frame-invariant combination has the interpretation as the comoving density perturbation: the density perturbation evaluated in the frame in which the fluid has no velocity. In harmonic space we have

$$\bar{\delta\rho} = \delta\rho + 3(\rho+P)\frac{Hav}{k} = \delta\rho + 3(\rho+P)\frac{\mathcal{H}v}{k}, \tag{91}$$

where $\mathcal{H} = aH$ is the comoving Hubble parameter and k is the usual comoving wavenumber.

3. Conformal Newtonian gauge

We shall mainly be interested in scalar mode perturbations, since these are the perturbations responsible for density perturbations and potential flows. In general there are two independent physical

---

[8] Note that in this and the previous section x is a local non-comoving coordinate, so ∇ is non-comoving.

scalar perturbations to the metric, which we can parameterize by two functions Ψ and Φ, describing the change to the time and space parts:

$$ds^2 = a(\eta)^2 \left[(1+2\Psi)d\eta^2 - (1-2\Phi)d\mathbf{x}^2\right]. \tag{92}$$

Here we are switching to using conformal time η rather than t. This particular parameterization of the perturbations is called the Conformal Newtonian Gauge (CNG)[9]. As we shall see later, for matter density and inflation perturbations $\Psi = \Phi$, and Φ is the Newtonian potential that obeys the Poisson equation. To study the perturbations we shall first look at how they are produced during inflation, and then calculate the transfer functions that tell how the inflationary fluctuations change after inflation to give the structure we see in the universe today.

Ex: Show that $d^4x\sqrt{-g}$ is invariant under the change of coordinates $d\eta \to dt/a$, where g is the determinant of the metric.

C. Conserved scalar perturbation

To solve the horizon problem we need to explain the origin of perturbations larger than the particle horizon in the hot big bang, which means propagating perturbations generated during inflation (or some other theory) though various unknown process that happened between their generation and the start of the hot big bang (reheating). Fortunately this is possible on super-horizon scales, even if we know almost nothing about physics in some intermediate regime.

Recall what an FRW universe is: it is an isotropic solution for a homogeneous universe, with some constant K that measures the spatial curvature. As the FRW universe evolves, K remains a constant. Now if we are in a perturbed universe, but all the perturbations of interest are on scales much larger than the horizon, locally we expect it to look a lot like an FRW universe, since everything is locally smooth and homogeneous: this is called the *separate universe assumption*. In fact we can choose a gauge so that $\delta\rho = 0$ everywhere, so that the density is exactly homogenous. If we then also have $\delta P = 0$ this corresponds to slicing the universe into what looks exactly like a set of FRW universes, each of which will have their own curvature constant K but the same density and pressure. But K is a constant, so at some later time, as long as the perturbations are well outside the horizon, K should remain the same, independent of what happens to the densities and pressures inside each FRW region. This is precisely what we need to relate super-horizon perturbations at the end of inflation or perturbations at the beginning of the hot big bang.

How do we calculate the relevant K as a function of position in the perturbed universe? A local curvature on a 3-surface can be defined in just the same way as the curvature of spacetime, so we can define a scalar $R^{(3)}$, the Ricci scalar on the comoving spatial metric. For the FRW universe this is

$$ds_3^2 = a^2\left[\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right], \tag{93}$$

which gives $R^{(3)} = 6K/a^2$. The FRW solution holds on super-horizon scales if $\delta\rho = 0$, so to calculate the conserved quantity analogous to K in a perturbed spacetime we just need to evaluate something proportional to $a^2R^{(3)}$ in the $\delta\rho = 0$ gauge [10]. In the FRW limit there are no velocities, so the $\delta\rho = 0$ gauge is also the same as the zero-velocity gauge on large scales; i.e. we can equally well calculate $a^2R^{(3)}$ in the comoving frame, the *comoving curvature perturbation*. Since $R^{(3)}$ contains two spatial

---

[9] Note that all possible permutations of $\Psi \leftrightarrow \pm\Phi$ and signs are used in equivalent definitions in the literature

[10] This works if setting $\delta\rho = 0$ enforces $\delta P = 0$ (or example if $P = P(\rho)$) so that both pressure and density are homogeneous in the $\delta\rho = 0$ gauge. For single-field inflation, there is only one degree of freedom, and this is the case. If there are more degrees of freedom, say two fields, then in general you can't choose a frame where $\delta\rho = \sum_i \delta\rho_i = 0$ and $\delta P = \sum_i \delta P_i = 0$. What you can still do is define a set of curvatures K_i, each evaluated on different slices with $\delta\rho_i = 0$; these have similar gauge-invariant form to Eq. (98), and if the fluids are uncoupled having $\rho_i' = -3\mathcal{H}(\rho_i + P_i)$, and $P_i = P_i(\rho_i)$ so that $\delta P = (P'/\rho')\delta\rho$, then these are separately conserved.

derivatives we define $\mathcal{R}$ so that $4\boldsymbol{\nabla}^2\mathcal{R} = -a^2\bar{R}^{(3)}$, or equivalently in harmonic space we define

$$\mathcal{R} \equiv \frac{a^2\bar{R}^{(3)}}{4k^2}, \tag{94}$$

where the bar denotes evaluation in the comoving gauge (or equivalently on super-horizon scales, the $\delta\rho = 0$ gauge). This should be a constant on super-horizon scales.

The 3-Ricci scalar can be calculated from the metric. Let's parameterize the spatial perturbation by ζ, and take it to be isotropic so the 3-metric is

$$ds_3^2 = a(t)^2 e^{2\zeta(\mathbf{x},t)} d\mathbf{x}^2. \tag{95}$$

This is like the CNG, but we now haven't imposed any specific choice of the time coordinate. There are many possible slicings of spacetime into different spatial slices with different ζ; indeed we could choose to slice so that $\zeta = 0$ — the *flat slicing*. In a general gauge the parameter ζ has the natural interpretation as the fractional perturbation to the scale factor $\zeta = \delta \ln a = (\delta a)/a$, so $a(\mathbf{x}, t) = a(t)e^{\zeta(\mathbf{x},t)} = a(t)(1 + \zeta(\mathbf{x}, t))$ to first order. Calculating the Ricci scalar to first order we get $a^2R^{(3)} = -4\boldsymbol{\nabla}^2\zeta$: since the metric is isotropic, and the 3-Ricci scalar is a scalar, unsurprisingly the two scalar measures of the perturbation are related to each other. The conserved $\mathcal{R}$ we defined before is just ζ evaluated in the $\delta\rho = 0$ slicing.

Recall from Eq. (80) how we expect scalar perturbations to change under a change δt in the choice of time coordinate t. Using this under a change of slicing we expect

$$\delta\rho \;\to\; \delta\rho - \dot{\rho}\delta t \tag{96}$$

$$\frac{\delta a}{a} \;\to\; \frac{\delta a}{a} - \frac{\dot{a}}{a}\delta t \implies \zeta \to \zeta - H\delta t. \tag{97}$$

It follows that we can construct the gauge-invariant quantity

$$\zeta - \frac{H\delta\rho}{\dot{\rho}} = \zeta + \frac{\delta\rho}{3(\rho + P)},$$

and this is equal to $\mathcal{R}$ because in the $\delta\rho = 0$ gauge it is just ζ: in a general gauge we can write the conserved super-horizon curvature perturbation as

$$\mathcal{R} = \frac{a^2R^{(3)}}{4k^2} + \frac{\delta\rho}{3(\rho + P)} = \zeta + \frac{\delta\rho}{3(\rho + P)}. \tag{98}$$

On super-horizon scales the $\delta\rho = 0$ slices have no peculiar velocity and one can also construct $\mathcal{R}$ using

$$\mathcal{R} = \frac{a^2R^{(3)}}{4k^2} - \frac{\mathcal{H}v}{k} \tag{99}$$

where v is the fluid velocity (see later), and equivalence of definitions holds for $k \ll \mathcal{H}$ (see the relativistic Poisson equation later). In the conformal Newtonian gauge $\zeta = -\Phi$ and hence

$$\mathcal{R} = -\Phi + \frac{\delta\rho}{3(\rho + P)}. \tag{100}$$

D. Power spectrum and transfer functions

Perturbations are random and have zero mean (they are equally likely to be positive or negative), but they have non-zero variance. In cosmology the variance is quantified by what is called the *power spectrum*: this measures the variance of the fluctuations as a function of scale (wavenumber). So if the power spectrum is larger at a particular k, the variance of modes with wavenumber k is larger, and so the positive and negative perturbations typically have larger amplitude on a scale $\lambda \sim 2\pi/k$.

Statistical homogeneity and isotropy restrict the form of the power spectrum: consider the variance of any field $\chi(\mathbf{x},t)$ at some point $\mathbf{x}$

$$\langle[\chi(\mathbf{x},t)]^2\rangle = \langle \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \frac{d^3\mathbf{k}'}{(2\pi)^{3/2}} \chi(\mathbf{k},t)\chi(\mathbf{k}',t) e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}} \rangle. \tag{101}$$

However we think there should be no special place in the universe on average. In other words, $\langle[\chi(\mathbf{x},t)]^2\rangle$ should be independent of $\mathbf{x}$, so that every point is statistically equivalent: we assume statistical homogeneity. For this to be true we need $\langle\chi(\mathbf{k},t)\chi(\mathbf{k}',t)\rangle \propto \delta(\mathbf{k}+\mathbf{k}')$ so the $\mathbf{x}$-dependence cancels. For statistical isotropy to hold (no preferred direction on average), the proportionality constant can only be a function of $k = |\mathbf{k}|$. We choose the normalization so that

$$\langle\chi(\mathbf{k},t)\chi(\mathbf{k}',t)\rangle = \frac{2\pi^2}{k^3}\mathcal{P}_\chi(k,t)\delta(\mathbf{k}+\mathbf{k}'), \tag{102}$$

which gives a definition for the power spectrum $\mathcal{P}(k,t)$ to quantify the variance of the fluctuations. With this definition

$$\langle[\chi(\mathbf{x},t)]^2\rangle = \int \frac{dk}{k}\mathcal{P}_\chi(k,t), \tag{103}$$

and $\mathcal{P}_\chi$ has the same dimensions as χ^2. $\mathcal{P} = $ const is called a scale-invariant spectrum, and corresponds to equal amplitude of fluctuations of all possible sizes (so any random realization is statistically indistinguishable from an enlarged version of itself). Be aware that different authors sometimes use different definitions.

If the fluctuations are *Gaussian*, the power spectrum completely defines the statistics of the random perturbations. Even if the perturbations are non-Gaussian, the power spectrum is a useful way to quantify the size of the typical perturbations as a function of scale. Any theory that solves the problems of the hot big bang must also produce a power spectrum of fluctuations that can then evolve to give the large-scale structure seen today. Image we are interested in a density perturbation at late times, Δ. This will have come originally from some random super-horizon curvature perturbation $\mathcal{R}$ at the beginning of the hot big bang. But the *evolution* of a given initial perturbation is determined by the laws of physics (gravitational collapse, pressure, etc, as we show later). In linear theory we can define a transfer function T_Δ that describes the linear relationship between the original perturbation at the start of the hot big bang, and the later density perturbation $\Delta(\mathbf{k},t) = T_\Delta(k,t)\mathcal{R}(\mathbf{k},0)$. Remember that because of statistical isotropy $T_\Delta(k,t)$ is only a function of $k = |\mathbf{k}|$. The initial $\mathcal{R}$ is a random variable determining the initial conditions, but the evolution given by T is entirely deterministic. So the power spectrum of Δ can easily be calculated in terms of the transfer function

$$\begin{aligned}
\langle\Delta(\mathbf{k},t)\Delta(\mathbf{k}',t)\rangle = \frac{2\pi^2}{k^3}\mathcal{P}_\Delta(k,t)\delta(\mathbf{k}+\mathbf{k}') &= \langle T_\Delta(k,t)\mathcal{R}(\mathbf{k},0)T_\Delta(k',t)\mathcal{R}(\mathbf{k}',0) && (104)\\
&= T_\Delta(k,t)T_\Delta(k',t)\langle\mathcal{R}(\mathbf{k},0)\mathcal{R}(\mathbf{k}',0)\rangle && (105)\\
&= [T_\Delta(k,t)]^2\frac{2\pi^2}{k^3}\mathcal{P}_\mathcal{R}(k)\delta(\mathbf{k}+\mathbf{k}'). && (106)
\end{aligned}$$

So the late-time density power spectrum is given simply by

$$\mathcal{P}_\Delta(k,t) = [T_\Delta(k,t)]^2\mathcal{P}_\mathcal{R}(k). \tag{107}$$

In the last section of the course we will see how to calculate the transfer function; first we will look at how inflation can generate the primordial perturbations with power spectrum $\mathcal{P}_\mathcal{R}(k)$ (which is constant outside the horizon for adiabatic perturbations, because $\mathcal{R}$ is constant).

IV. THE CREATION OF FLUCTUATIONS BY INFLATION

Inflation can do more for us than solve the horizon, flatness and monopole problem. It can also explain the primordial fluctuations that grew to create the large scale structure (galaxies, clusters, etc) visible in the universe, and lead to the observed anisotropies in the CMB.

The basic picture is that quantum fluctuations during inflation are stretched beyond the horizon by inflation, where we see them as classical fluctuations on scales larger than the horizon at the end of inflation. So we need to study the inflaton field at the quantum level to calculate the fluctuations produced and their power spectrum, and then the subsequent evolution.

A. Perturbations of the inflaton field

The Lagrangian for the scalar field is

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) \tag{108}$$

with action

$$S = \int d^4x\sqrt{-g}\mathcal{L}_\phi. \tag{109}$$

A classical solution by definition minimizes the action. We are interested in fluctuations about the classical background evolution, so split the field up into the homogeneous part $\phi(\eta)$ and a small perturbation $\delta\phi$:

$$\phi(\mathbf{x},\eta) = \phi(\eta) + \delta\phi(\mathbf{x},\eta). \tag{110}$$

In general we also need to consider perturbations to the metric. These enter the perturbed field equation multiplied by derivatives of $\phi(\eta)$, which for slow-roll inflation are expected to be small. The Einstein equations relate perturbations in the metric to perturbations in the matter, so in fact there is only one physical degree of freedom, since anisotropic stress is zero during inflation. We quantify this degree of freedom by $\delta\phi$ and approximate the metric as being unperturbed; this gives results that are substantially correct at horizon crossing if $\delta\phi$ is interpreted as the perturbation in the spatially flat (zero curvature) gauge — the gauge in which the 3-curvature is zero and the spatial part of the metric is unperturbed.

The action is then given by $S = S_c[\phi] + S_2[\phi,\delta\phi] + \ldots$, where $S_c[\phi]$ is the classical background action, and S_2 is the leading quadratic part due to the perturbation (there is no term linear in $\delta\phi$ since it must vanish for $\phi(t)$ to minimize the action). Taylor expanding the potential we have

$$S_2 = \int d^4x\sqrt{-g}\frac{1}{2}\left\{\partial_\mu(\delta\phi)\partial^\mu(\delta\phi) - V_{,\phi\phi}(\phi)(\delta\phi)^2\right\}. \tag{111}$$

For an unperturbed conformal time FRW metric $\sqrt{-g} = a^4$. The corresponding field equation is

$$(\delta\phi)'' + 2\mathcal{H}(\delta\phi)' - \boldsymbol{\nabla}^2\delta\phi + a^2V_{,\phi\phi}\delta\phi = 0. \tag{112}$$

Here $\boldsymbol{\nabla}$ is the comoving spatial Laplacian. As mentioned we are neglecting perturbations in g, which is not strictly valid. In fact to the same level of accuracy we can also drop the $V_{,\phi\phi}$ term, and the result will still be approximately accurate at horizon crossing in standard slow-roll models. Although the analysis is approximate here, the method applies to the full result, it just takes more work to find the exact form of S_2.

Neglecting $V_{,\phi\phi}$, in Fourier space the field equation then becomes

$$(\delta\phi)'' + 2\mathcal{H}(\delta\phi)' + k^2\delta\phi = 0, \tag{113}$$

where $\mathbf{k}$ is the comoving wave vector. To see what happens for the quantum fluctuations, we start by considering the case then $k \gg \mathcal{H}$, so that the perturbations are well inside the horizon and the Hubble damping term can be neglected; we just have the equation for a free scalar field in flat space. The inflationary result then follows from a small generalization.

Ex: If H is a constant show that $\mathcal{H} = -1/\eta$ (for $-\infty < \eta < 0$).

B. The quantisation of the free scalar field

The equation of motion of a massless free scalar field $\phi(\mathbf{x}, t)$ in flat space-time is

$$\ddot{\phi} - \nabla^2\phi = 0. \tag{114}$$

This equation is called the *Klein-Gordon equation*. To solve it we change into Fourier space giving

$$\ddot{\phi}(\mathbf{k}, t) + k^2\phi(\mathbf{k}, t) = 0. \tag{115}$$

This is just the equation of a harmonic oscillator, and its solutions are $\phi(\mathbf{k}, t) \propto \exp(\pm ikt)$.

To quantise we need to calculate the momentum conjugate to ϕ. We can obtain this from the Lagrangian

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi \tag{116}$$

since

$$\pi_\phi \equiv \frac{\partial\mathcal{L}_\phi}{\partial\dot{\phi}} = \dot{\phi}. \tag{117}$$

The commutation relations imposed by quantum mechanics on the corresponding operators are then

$$[\hat{\phi}(\mathbf{x}, t), \hat{\pi}_\phi(\mathbf{x}', t)] = i\delta(\mathbf{x} - \mathbf{x}') \tag{118}$$

$$[\hat{\phi}(\mathbf{x}, t), \hat{\phi}(\mathbf{x}', t)] = 0 \tag{119}$$

$$[\hat{\pi}_\phi(\mathbf{x}, t), \hat{\pi}_\phi(\mathbf{x}', t)] = 0 \tag{120}$$

where our units have $\hbar = c = 1$. We use the Heisenburg picture where states are constant, but operators evolve with the classical equations of motion. In harmonic space we can write

$$\hat{\phi}(\mathbf{k}, t) = w(k, t)\hat{a}(\mathbf{k}) + w(k, t)^*\hat{a}^\dagger(-\mathbf{k}), \tag{121}$$

where

$$w(k, t) = \frac{1}{\sqrt{2k}}e^{-ikt}, \tag{122}$$

where the factor of $\sqrt{2k}$ is inserted so that consistency with (118) is obtained by the operators $\hat{a}$ and $\hat{a}^\dagger$ satisfying the usual harmonic oscillator canonical commutation relations

$$[\hat{a}(\mathbf{k}), \hat{a}^\dagger(\mathbf{k}')] = \delta(\mathbf{k} - \mathbf{k}'), \qquad [\hat{a}(\mathbf{k}), \hat{a}(\mathbf{k}')] = 0 \qquad [\hat{a}^\dagger(\mathbf{k}), \hat{a}^\dagger(\mathbf{k}')] = 0. \tag{123}$$

Ex: show these relations are consistent with Eq. (118).

Postulating a space of state vectors $|N_1, N_2, \ldots >$ where the states N_i have a defined momentum $\mathbf{k}_i$ and the number operator $N_i \equiv \hat{a}(\mathbf{k}_i)^\dagger\hat{a}(\mathbf{k}_i)$ we can then show with the commutation relations that

$$\hat{a}_1^\dagger\,|N_1, N_2, \ldots > = \sqrt{N_1 + 1}\,|N_1 + 1, N_2, \ldots > \tag{124}$$

$$\hat{a}_1\,|N_1, N_2, \ldots > = \sqrt{N_1}\,|N_1 - 1, N_2, \ldots > \tag{125}$$

Hence $\hat{a}^\dagger(\mathbf{k})$ plays the role of a *creation operator* which creates new particles with a given momentum $\mathbf{k}$, while $\hat{a}(\mathbf{k})$ is an *annihilation operator* which destroys particles of a given momentum $\mathbf{k}$. Finally, we demand that the vacuum is normalized, $< 0|0 >= 1$ and that there are no states containing a negative number of particles, $\hat{a}(\mathbf{k})|0 >= 0$.

C. Inflationary fluctuations and their spectrum

With the un-perturbed metric approximation the second-order part of the Lagrangian for the fluctuations in the expanding spacetime is now

$$L_2 = \frac{1}{2}\sqrt{-g}\partial_\mu\delta\phi\partial^\mu\delta\phi = \frac{a^4}{2}\partial_\mu\delta\phi\partial^\mu\delta\phi = \frac{a^2}{2}\left[(\delta\phi')^2 - (\boldsymbol{\nabla}\delta\phi)^2\right]. \tag{126}$$

So fluctuations that are produced will evolve with the field equation (113):

$$(\delta\phi)'' + 2\mathcal{H}(\delta\phi)' + k^2\delta\phi = 0, \tag{127}$$

and the conjugate momentum is

$$\pi_{\delta\phi} \equiv \frac{\partial L_2}{\partial(\delta\phi)'} = a^2\delta\phi'. \tag{128}$$

The harmonic operator expansion in creation and annihilation operators is the same as before

$$\hat{\delta\phi}(\eta) = w(k,\eta)\hat{a}(\mathbf{k}) + w^*(k,\eta)\hat{a}^\dagger(-\mathbf{k}), \tag{129}$$

but now the time evolution $w(k,\eta)$ is a solution of the field equation

$$w'' + 2\mathcal{H}w' + k^2 w = 0. \tag{130}$$

Again we need to normalize w so that the creation and annihilation operators satisfy the canonical commutation relations, which is achieved if

$$a^2(ww^{*\prime} - w'w^*) = i. \tag{131}$$

The term in brackets is the Wronskian[11] which for Eq. (127) is $\propto 1/a^2$, so the equation is satisfied by choosing the correct constant normalization for w.

Finding a general analytic solution to Eq. (130) is difficult, but taking H to be approximately constant over the range of interest we have $\mathcal{H} = -1/\eta$ and the solution is $w \propto (k\eta - i)e^{-ik\eta}$. Using $a = -1/(H\eta)$ the normalized solution (satisfying Eq. (131)) is then

$$w(k,\eta) = \frac{H(k\eta - i)}{\sqrt{2k^3}}e^{-ik\eta}. \tag{132}$$

At late times the fluctuations on scales well outside the horizon, $k \ll \mathcal{H}$ (or $|k\eta| \ll 1$), are then determined by

$$|w(k,\eta)|^2 = \frac{H^2}{2k^3}, \tag{133}$$

1. *Scalar power spectrum from inflation*

The non-zero variance due to the quantum fluctuations in the scalar field is given from Eq. (121) and the commutation relations by

$$< 0|\,\delta\phi(\mathbf{k},\eta)\delta\phi(\mathbf{k}',\eta)\,|0> = w(k,\eta)w^*(k',\eta) < 0|\,\hat{a}(\mathbf{k})\hat{a}^\dagger(-\mathbf{k}')\,|0> = |w(k,\eta)|^2\delta(\mathbf{k}+\mathbf{k}'), \tag{134}$$

so the power spectrum of the fluctuations is

$$\mathcal{P}_{\delta\phi}(k,\eta) = \frac{k^3}{2\pi^2}|w(k,\eta)|^2. \tag{135}$$

---

[11] Reminder: For solutions A and B of a 2nd order equation, the Wronksian is $W = AB' - BA'$, and so $W' = AB'' - BA''$ and substituting for A'' and B'' from the 2nd order equation gives an equation for W that can be solved

Substituting the result for $w(k,\eta)$ from Eq. (133) the spectrum is then

$$\mathcal{P}_{\delta\phi}(k) \approx \mathcal{P}_{\delta\phi}(k,\eta) \approx \left(\frac{H}{2\pi}\right)^2, \tag{136}$$

where in general this expression is a reasonable approximation if H is evaluated at horizon crossing. If H is constant the spectrum is scale invariant. Once well outside the horizon $\delta\phi \sim$ const since that is a solution to Eq. (127) for $k \ll \mathcal{H}$, and hence the power spectrum is also constant.

Correlations of three, four or more perturbations can be calculated in a similar way to Eq. (134), using the commutation relations to produce a series of products of $|w|^2$ and delta-functions. In other words, the higher-point functions can be written in terms of the two-point function (the power spectrum), which is precisely the property of a Gaussian random field. All the information on the statistics of the fluctuations is contained in the power spectrum: simple inflationary models predict a Gaussian spectrum of perturbations.

D. The spectrum of the primordial comoving curvature perturbation

In general, we do not use the perturbation spectrum of $\delta\phi$ directly, but a gauge-invariant measure of the fluctuations induced in the geometry, or more specifically the comoving curvature perturbation. This remains constant outside the horizon, even if the background equation of state changes, and remains well defined after the scalar field has decayed. It can therefore be used to relate the super-horizon fluctuations during inflation to the super-horizon fluctuations in the early radiation dominated universe, without having to worry about the unknown details of reheating. Our calculation of the quantum fluctuations is most accurate in the zero-spatial curvature gauge (flat slicing — recall we neglected metric perturbations). The comoving curvature perturbation during slow-roll inflation is given in this flat slicing (zero-curvature gauge[12]) from Eq. (98) by

$$\mathcal{R} = \frac{\delta\rho}{3(\rho+P)} = -\frac{H}{\dot\phi}\delta\phi. \tag{137}$$

Recall from Sec. III C that $\mathcal{R}$ also has the interpretation of a fractional scale-factor perturbation in the $\delta\rho = 0$ gauge, $\mathcal{R} = \delta\ln a = \delta N$, so the curvature perturbation measures the relative perturbation to the inflationary expansion (number of e-foldings N) on equal density hypersurfaces. This can also be thought of as due to time differences δt for different patches to reach a given density, since $\mathcal{R} = \delta\ln a = H\delta t$.

In single-field inflation, on slices with $\delta\rho = 0$ local physics at the end of inflation is the same everywhere. There is one scalar degree of freedom that measures the slightly different times of reheating, but the physics is the same so the end products should be the same. In other words, there should be some unique relation between the densities of the different species at the beginning of the hot big bang, so if $\delta\rho = 0$ then $\delta\rho_i = 0$. If we change from $\delta\rho = \delta\rho_i = 0$ gauge to any other gauge, the δt change in slicing gives $\delta\rho_i = -\dot\rho_i\delta t$. It follows that $\delta\rho_i/\dot\rho_i$ should be the same for all species. In fact in the zero curvature gauge the initial conditions are[13]

$$H\delta t = -\frac{H\delta\rho_i}{\dot\rho_i} = \frac{\delta\rho_i}{3(\rho_i+P_i)} = \mathcal{R}. \tag{138}$$

Perturbations where all the species are perturbed this way are called adiabatic. Single-field inflation predicts *adiabatic perturbations*: where there is an overdensity in dark matter we also expect a corresponding overdensity in the photons, baryons and neutrinos. If there were more than one

---

[12] This might sound confusing. Remember both the 3-curvature and $\delta\phi$ are gauge-dependent, only one linear combination is gauge invariant. In the comoving frame $\delta\phi = 0$, so the comoving curvature is just the curvature, but in the zero-curvature frame, it is proportional to $\delta\phi$.

[13] Note that $\delta\rho_i$ here is not the comoving or synchronous gauge density that satisfies the Poisson equation and is usually discussed for the matter power spectrum. On large scales they differ by a factor of k^2.

degree of freedom, potentially there could be other types of perturbations where this is not the case: *isocurvature modes* — orthogonal combinations of compensating densities that do not contribute to $\mathcal{R}$. Observationally there is no evidence for isocurvature modes.

For single-field inflation the spectrum of $\mathcal{R}$ therefore uniquely determines the statistics of the adiabatic perturbations at the start of the hot big bang, and from Eq. (137) we have

$$\mathcal{P}_{\mathcal{R}}(k) \approx \left(\frac{H}{\dot{\phi}}\right)^2 \left(\frac{H}{2\pi}\right)^2 \Bigg|_{aH=k} \tag{139}$$

where the right hand side is evaluated at/after horizon crossing (from which point $\mathcal{R}(k,t)$ is constant). Using the slow-roll equations $3H\dot{\phi} = -V_{,\phi}$ and $3H^2M_P^2 = V$ we can write

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{1}{3(2\pi)^2 M_P^6}\frac{V^3}{V_{,\phi}^2} = \frac{1}{24\pi^2 M_P^4}\frac{V}{\epsilon_V} = \frac{1}{2M_P^2\epsilon_V}\left(\frac{H}{2\pi}\right)^2 \Bigg|_{aH=k} \tag{140}$$

Since $\mathcal{R}$ is conserved on super-horizon scales, this power spectrum gives the predicted spectrum of curvature perturbations at the beginning of the hot big bang after reheating.

Since H and $\dot{\phi}$ only vary slowly during inflation, $\mathcal{P}_{\mathcal{R}}$ is close to scale invariant, with a weak k dependence coming from the change in the background quantities when different k leave the horizon. The spectrum is often parameterized as

$$\mathcal{P}_{\mathcal{R}}(k) = A_s\left(\frac{k}{k_s}\right)^{n_s-1} \tag{141}$$

where $n_s \sim 1$ is the conventional definition of the scalar spectral index and k_s is some chosen scale of observational interest (e.g. $k_s = 0.05\text{Mpc}^{-1}$). From the Planck satellite we know $A_s \sim 2\times10^{-9}$ and $n_s \approx 0.968 \pm 0.005$, slightly but significantly less than unity. This implies a numerical value for the potential at k_s,

$$V^{1/4} \sim 6\times10^{16}\text{GeV}\epsilon_V^{1/4}. \tag{142}$$

To calculate the spectral index n_s we define it as

$$n_s(k) - 1 \equiv \frac{d\ln\mathcal{P}_{\mathcal{R}}}{d\ln k}. \tag{143}$$

Since we evaluate all quantities at $k = aH$ and since H is essentially constant, we have $d\ln k = d\ln a = (H/\dot{\phi})d\phi$, so using the slow-roll equations

$$\frac{d}{d\ln k} = -M_P^2\frac{V_{,\phi}}{V}\frac{d}{d\phi}. \tag{144}$$

We find therefore that

$$n_s - 1 = -M_P^2\frac{V_{,\phi}}{V}\frac{d\ln\mathcal{P}_{\mathcal{R}}}{d\phi} = -M_P^2\frac{V_{,\phi}}{V}\frac{V_{,\phi}^2}{V^3}\frac{d}{d\phi}\left(\frac{V^3}{V_{,\phi}^2}\right) = -6\epsilon_V + 2\eta_V. \tag{145}$$

In slow-roll inflation, we see that the deviation from $n_s = 1$ is expected to be small, consistent with observations. The case $n_s > 1$ is often referred to as a *blue* spectrum, and the other case as a *red* spectrum; most simple inflationary models predict a slightly red spectrum.

Exact result

We used the approximation of no metric fluctuations, however the full result including metric perturbations is very analogous, where the Lagrangian can be reduced to

$$L = \frac{1}{2}\left(\frac{a\phi'}{\mathcal{H}}\right)^2\left[(\mathcal{R}')^2 - (\boldsymbol{\nabla}\mathcal{R})^2\right] \tag{146}$$

and in the Newtonian gauge

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}}{\phi'}\delta\phi. \tag{147}$$

This can be compared with the approximate form (126) that we used.

E. Gravitational waves

General relativity predicts the existence of gravitational waves, distortions of the geometry that can be excited by quantum fluctuations in a similar way to fluctuations in the scalar field. An FRW universe perturbed by gravitational waves has a metric

$$ds^2 = a(\eta)^2\left[d\eta^2 - (\delta_{ij} + 2H_{ij})\,dx^i dx^j\right]. \tag{148}$$

For gravitational waves H_{ij} is restricted to be trace-free $H^i_i = 0$ and be transverse, $\partial_i H_{ij} = 0$ so that there are only two independent components, H_+ and H_-. The perturbation is distinct from scalar perturbations since H_{ij} with the above properties cannot be written in terms of spatial derivatives of a scalar field. In Fourier space and a coordinate system where the wave vector $\mathbf{k}$ is pointing into the 3-direction we find $H_{11} = -H_{22} = H_+$ and $H_{12} = H_{21} = H_-$, while all other components vanish. The action and equations of motion can be obtained with some work by inserting the perturbed metric into the action

$$S = \int d^4x\,\sqrt{-g}\left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) - \frac{M_P^2}{2}R\right) \tag{149}$$

and keeping the terms quadratic in H_{ij}, where R is the Ricci scalar. The inflationary part has no tensor component itself and so is just the background value, but $\sqrt{-g}$ does contain important H_{ij} dependence. Calculating the quadratic part and using the background equations gives

$$S_2 = \int d^4x\,M_P^2a^2\left[(H_-{}')^2 - (\boldsymbol{\nabla}H_-)^2\right] + \int d^4x\,M_P^2a^2\left[(H_+{}')^2 - (\boldsymbol{\nabla}H_+)^2\right]. \tag{150}$$

The field equations for H_+ and H_- give equations of motion for each polarization of the gravitational waves:

$$H_\pm{}'' + 2\mathcal{H}H_\pm{}' + k^2H_\pm = 0, \tag{151}$$

which is the equation of motion of a massless scalar field, just like in the scalar case. Note that for $k \ll \mathcal{H}$ a solution is $H_\pm = $ const, so the tensor perturbations are also conserved on super-horizon scales (as for the curvature perturbation in the scalar case). This equation also follows from using the Einstein equations if the stress-energy tensor has no transverse symmetric trace-free part (no tensor anisotropic stress); as such it also describes the evolution of gravitational waves at any epoch in the universe after inflation in the approximation that anisotropic stress can be neglected.

During inflation the vacuum fluctuations of the $H_\pm$ field will hence lead to a background of gravitational waves in precisely the same way as the vacuum fluctuations of the inflaton field. We just need to be careful about the normalization, since quadratic part of the perturbed action (150) is normalized like the scalar case for the variable combination

$$h_\pm \equiv \sqrt{2M_P^2}H_\pm. \tag{152}$$

The variable $h_\pm$ is given quantum fluctuations $H/(2\pi)$ in each polarization as before. The tensor spectrum is defined similarly to the scalar case (Eq. (103)) so that the power in the metric is

$$\langle 2H_{ij}(\mathbf{x})2H^{ij}(\mathbf{x})\rangle = 8\langle H_+(\mathbf{x})H_+(\mathbf{x})\rangle + 8\langle H_-(\mathbf{x})H_-(\mathbf{x})\rangle = \int\frac{dk}{k}\mathcal{P}_T(k), \tag{153}$$

which includes contributions from both polarizations. It is constant for scales that have left the horizon. Hence the tensor power spectrum is

$$\mathcal{P}_T(k) = 2 \times 8 \times \frac{1}{2M_P^2} \times \left(\frac{H}{2\pi}\right)^2\Bigg|_{k=aH} = \frac{8}{M_P^2}\left(\frac{H}{2\pi}\right)^2\Bigg|_{k=aH} = \frac{2}{3\pi^2}\frac{V_*}{M_P^4}. \tag{154}$$

So the amplitude of the tensor modes depends directly on the value of the potential when the modes left the horizon; if the amplitude could be measured it would tell us directly about the energy scale of inflation. For the amplitude to be significant (observable for $\mathcal{P}_T \gtrsim 10^{-12}$) this requires inflation at an energy scale $V_*^{1/4} \gtrsim 5 \times 10^{15}$GeV. Low-energy inflation produces negligible amounts of gravitational waves.

As in the scalar case we can define the spectral index (note conventional -1 difference in the definition)

$$n_T = \frac{d\ln \mathcal{P}_T(k)}{d\ln k} = -2\epsilon_V. \tag{155}$$

Since $\epsilon_V \geq 0$ for slow-roll down a potential, $n_T \leq 0$ corresponding to the decreasing amplitude of fluctuations $\propto H$ as the field rolls to lower values of the V at later times; the spectrum is red.

The relative size of the primordial super-horizon tensor and scalar mode perturbation can be quantified using the definition (at some scale)

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_\mathcal{R}} = 16\epsilon_V. \tag{156}$$

Using this expression, we can directly relate r and n_T:

$$n_T = -\frac{r}{8}. \tag{157}$$

This is called the *consistency relation*. It applies only to single-field inflationary models, and is different in more general cases. The existence of this relation is not surprising, since the creation of both the inflaton density perturbations and the gravitational waves are connected to the potential $V(\phi)$. The experimental confirmation of the consistency relation would be a major success of the single-field inflation model. Unfortunately it is hard to test at high accuracy since the tensor modes are only observable over a relatively small range of scales (modes that were horizon-sized at recombination and larger since gravitational waves decay on sub-horizon scales), so measuring n_T is difficult.

Ex: Show that $u \equiv aH_\pm$ satisfies the equation

$$u'' + \left(k^2 - \frac{a''}{a}\right)u = 0. \tag{158}$$

Hence show that for $k \gg \mathcal{H}$ (specifically $a''/a \ll k^2$) that the amplitude of the sub-horizon tensor modes decays as $1/a$. [Hence: Since the modes decay on sub-horizon scales, only modes entering the horizon between us and last scattering have a chance to appreciably distort the CMB background, leading to an observational signal only in the large-scale CMB temperature anisotropy due to horizon-scale gravitational waves.]

F. Example: quadratic potential

Returning to the example of $V(\phi) = m^2\phi^2/2$ let's compute the expected perturbations. First, we remember that the slow-roll parameters were

$$\epsilon_V(\phi) = \eta_V(\phi) = \frac{2M_P^2}{\phi^2} \tag{159}$$

and the number of e-foldings until the end of inflation

$$N = \frac{\phi_i^2}{4M_P^2} - \frac{1}{2} \tag{160}$$

(where we will neglect the 1/2 from now on).

The scalar spectral index is then

$$n_s - 1 = -6\epsilon_V + 2\eta_V = -4\epsilon_V \tag{161}$$

$$= -\frac{8M_P^2}{\phi^2} = -\frac{2}{N} \approx -0.04 \tag{162}$$

where we assumed in the last step that the scales of interest left the horizon 50 e-foldings before the end of inflation. We expect therefore $n_s \approx 0.96$, very compatible with the Planck results of $n_s = 0.96 \pm 0.01$ at 1σ.

The spectral index of the gravitational waves, or tensor spectral index n_T is found to be

$$n_T = -2\epsilon_V = \frac{n_s - 1}{2}, \tag{163}$$

and the ratio of scalar to tensor contributions is

$$r = 16\epsilon_V = 0.16. \tag{164}$$

This model leads to an observable contribution from gravity waves. However current observations indicate that r is likely to be less than this (the detection by BICEP in 2014 of $r \sim 0.2$ is now known to be contaminated significantly by galactic dust, with current dust-corrected limits suggesting $r \lesssim 0.1$ at 2σ.).

G. Summary

Inflation solves the horizon problem by exponentially inflating the early universe so that scales out of causal contact at the beginning of the hot big bang were in causal contact earlier on during inflation. Inflation also generates perturbations by quantum fluctuations, which for slow-roll models are expected to give close to scale-invariant gaussian curvature perturbations, leading to adiabatic initial conditions for modes outside the horizon at the beginning of the hot big bang. It is therefore an attractive solution to some of the problems with a pure hot big bang. Inflation can also generate gravitational waves, with an amplitude that depends on the energy scale of inflation, which may give an observational handle on the energy scale involved (physics at much high energies than can be reached in accelerators). Since the fluctuations are expected to be very nearly Gaussian in single-field models, a detection of significant non-Gaussianity is a potentially powerful way to rule out simple single-field inflationary models.

There are however some potential problems and unanswered question with inflation

- Why did the scalar field begin 'up the hill'? (rather than always being at the minimum, so no inflation)
- Why did the universe start smooth over a size $1/H$ - much larger than the Planck length?
- What *is* the scalar field? - is it anything that fits into particle physics or string theory models?
- Why was the potential so flat, flat enough to allow slow-roll inflation? Does the near-flatness indicate a weakly broken underlying shift symmetry in the field, or is it another fine tuning?
- How did inflation end and how did reheating take place?
- Without knowing what the scalar field is, we have no idea what the potential is, so the theory is not very predictive - just a few consistency conditions. With multiple fields these consistency conditions weaken further - is the theory really falsifiable? Aren't the basic predictions the sort of thing you'd naturally expect in *any* simple theory that solved the horizon problem?

- If the universe is infinite, do quantum fluctuations randomly push the field further *up* the hill in some places, giving eternal/chaotic inflation? And how do you calculate predictions when the universe contains infinite regions, ending inflation in different ways at different times, and being observed by infinite numbers of causally disconnected observers?