Inflation

Marieke Postma[1]

[1] *NIKHEF, Science Park 105 1098 XG Amsterdam, The Netherlands.*

Abstract: Notes on inflation

Contents

1. Big Bang Cosmology

1.1 FRW universe

Modern cosmology is grounded on the "cosmological principle": nobody is at the center of the universe, and the cosmos viewed from any point looks the same as from any other point. It is the Copernican principle, that we are not at the center of the solar system, taken to the extreme. It implies that the universe (on large scales) is isotropic and homogeneous (as seen by a freely falling observer), i.e. it is invariant under spatial translations and rotations. Cosmological "principle" instead of "law" because at the time it was introduced, in the 1920's, it was done mainly for mathematical simplicity, not based on any data. But in the last two decades observational evidence from the cosmic microwave background (CMB) and large scale structure surveys confirm the homogeneity and isotropy of the universe on *large scales* > 100 Mpc.[1]

An isotropic and homogenous universe is described by the Friedman-Robertson-Walker (FRW) metric:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2 \left[\frac{\mathrm{d}r^2}{(1-kr^2)} + r^2 \left(d\theta^2 + \sin^2\theta \mathrm{d}\varphi^2 \right) \right] = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu \tag{1.1}$$

with $a(t)$ the time-dependent cosmic scale factor. $\mathrm{d}s$ measures the proper distance between two points in spacetime separated by $\mathrm{d}x^\mu$. The constant $k = -1, 0, 1$ for an open, flat, or closed universe respectively, corresponding to the 3-dimensional spatial slices being hyperbolic surfaces with negative curvatore, flat Euclidean surfaces with zero curvature, or 3-spheres with positive curvature. To write the metric in the above form, the freedom to redefine $r \to \lambda r$ has been used to absorb the radius of curvature in the scale factor $R = a/\sqrt{k}$ and normalize $|k| = 1$ for curved universes.

$\{r, \theta, \varphi\}$ are called comoving coordinates, a particle initially at rest in these coordinates remains at rest, i.e. $\{r, \theta, \varphi\}$ remains constant. The physical separation between freely moving particles at $(t, 0)$ and (t, r) is

$$d(r,t) = \int \mathrm{d}s = a(t) \int_0^r \frac{\mathrm{d}r}{\sqrt{1-kr^2}} = a(t) \times \begin{cases} \sinh^{-1} r, & k = -1, \\ r, & k = 0, \\ \sin^{-1} r, & k = 1. \end{cases} \tag{1.2}$$

Thus physical distances and wavelengths scale $\lambda \propto a$, and momenta $p \propto a^{-1}$. The distance increases with time in an expanding universe ($\dot{a} > 0$):

$$\dot{d} = \frac{\dot{a}}{a} d \equiv H d, \tag{1.3}$$

with $H(t)$ the Hubble parameter or constant (to indicate it is independent of spacial coordinates). The above is nothing but Hubble's law: galaxies recede from each other with a velocity that is proportional to the distance. Hubble's law is borne out by observations; the present day measured Hubble parameter is $H_0 \sim 72 \pm 8$ km/sec/Mpc [2].

[1] The observable patch of our universe is ~ 3000Mpc, and 1Mpc $\approx 3 \times 10^{19}$km $\approx 3.3 \times 10^6$ light years.

[2] A subscript 0 will always denote the present day value of the corresponding quantity.

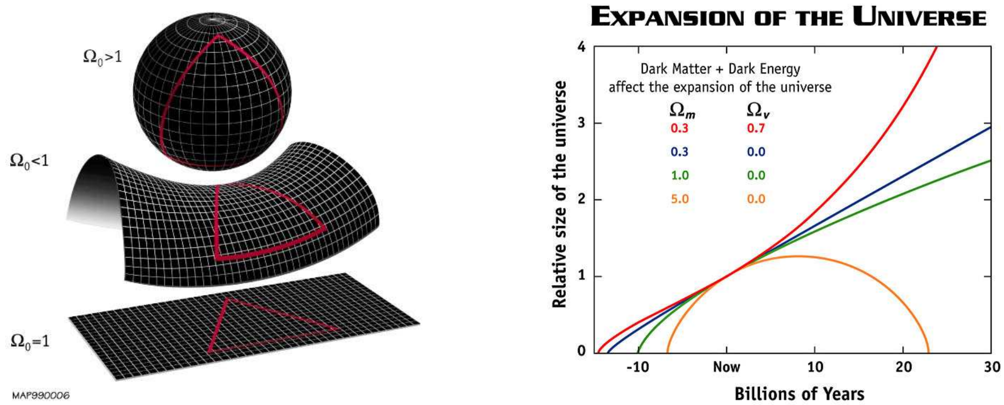

Figure 1: Curvature and expansion history of the universe.

A freely moving particle will eventually come at rest in comoving coordinates as its momentum is red shifted $p \propto a^{-1}$ to zero. The expansion of the universe creates a kind of dynamical friction for everything moving in it. It will be useful to define comoving distance and momenta, with the expansion factored out, via

$$\lambda_{\rm com} = \lambda_{\rm phys}/a(t), \qquad k_{\rm com} = a(t) k_{\rm phys}. \tag{1.4}$$

Motion w.r.t. comoving coordinates is called peculiar motion, it probes the local mass density.

A photon emitted with wavelength $\lambda_{\rm em}$ from a distant galaxy is red shifted, and observed at present with a longer wavelength λ_0, given by

$$(1+z) \equiv \frac{\lambda_{\rm em}}{\lambda_0} = \frac{a(t_0)}{a(t_{\rm em})}, \tag{1.5}$$

that is light with red shift $(1+z)$ was emitted when the universe was a factor $(1+z)^{-1}$ smaller. Another way to look at the effects is that the wave length of a photon traveling through sapcetime increases because the uncerlying spacetime is expanding.

Friedman equation In general relativity the metric is a dynamical object. The time evolution of the scale factor in (1.1) is governed by Einstein's equations

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G_N T_{\mu\nu} \tag{1.6}$$

with R and $R_{\mu\nu}$ the scalar curvature and Ricci curvature tensor respectively, which are both functions of the metric with up to two metric derivatives. I will use units in which $m_{\rm p}^2 = (8\pi G_N)^{-1} = 1$ (to restore units add appropiate powers of the reduced planck mass with $m_{\rm p} = 2.4 \times 10^{18}$GeV). The gravitational field, that is the metric of spacetime, is sourced and curved by matter/energy. The energy-momentum tensor is dictated by isotropy and homogeneity to be of the perfect fluid form $T_\mu^\nu = {\rm diag}(-\rho, p, p, p)$ (see Appendix for definition). Then Einstein's equations reduce to two independent equations

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{\rho}{3} - \frac{k}{a^2} \qquad \text{(Friedmann eq.)}, \tag{1.7}$$

$$\frac{\ddot{a}}{a} = -\frac{1}{6}(\rho + 3p) \qquad \text{(Raychaudhuri eq.)}. \tag{1.8}$$

Eq. (1.8) can also be traded for the continuity equation

$$\dot{\rho} + 3H(\rho + p) = 0 \qquad \text{(continuity eq.)} \tag{1.9}$$

which follows from (1.7, 1.8), and encodes energy conservation; it can also be derived from $\nabla_\nu T^{\mu\nu} = 0$. Heuristically, (1.9) is just the 1st law of thermodynamics:

$$\mathrm{d}U = -p\mathrm{d}V \quad \Rightarrow \quad \mathrm{d}(\rho a^3) = -p\mathrm{d}(a^3). \tag{1.10}$$

Introduce the equation of state parameter $p \equiv \omega\rho$. Then the continuity equation can be integrated to give

$$\frac{\mathrm{d}\rho}{\rho} = -3(1+\omega)\frac{\mathrm{d}a}{a} \quad \Rightarrow \quad \rho \propto a^{-3(1+\omega)} \tag{1.11}$$

From (1.7), neglecting the curvature term, it then follows

$$a \propto \begin{cases} t^{2/(3(1+\omega))} & \omega \neq 1 \\ \mathrm{e}^{Ht} & \omega = 1 \end{cases} \tag{1.12}$$

[This can be derived substituting $a = t^n$ and (1.11) in (1.7), to give $(n/t)^2 = 1/3t^{-3n(1+\omega)}$. This has the solution $n = 2/(3(1+\omega))$ provided $\omega \neq 1$. For $\omega = -1$ then $\rho =$ const. and (1.7) has an exponential solution.]

The matter in the universe consists of several fluids $T^\nu_\mu = \sum_i T^{(i)}{}^\nu{}_\mu$, with $i = \{\mathrm{rad}, \mathrm{mat}, \Lambda\}$ for radiation, non-relativistic matter and vacuum respectively. If the energy exchange between them is negligible, it follows that all fluids separately satisfy the continuity equation. We can define an equation of state parameter for each fluid separately $p_i = \omega_i \rho_i$.

- Radiation includes all relativistic species, at present only photons (generically, species are relativistic when $m \ll T$). For radiation $\omega_{\mathrm{rad}} = 1/3$ and thus (1.11) gives $\rho_{\mathrm{rad}} \propto a^{-4}$ [3]. If the universe is *dominated* by radiation, it follows from (1.12) that the scale factor grows $a \propto t^{1/2}$.

- Matter includes all non-relativistic or cold matter, at present baryons, dark matter and neutrinos. For matter $\omega_{\mathrm{mat}} = 0$ and thus $\rho_{\mathrm{mat}} \propto a^{-3}$. If the universe is *dominated* by matter, the scale factor grows $a \propto t^{2/3}$.

[3] The energy-density in radiation is the number density of relativistic particles times the momenta $\rho_{\mathrm{rad}} = pN_{\mathrm{rad}}/V$ with N_{rad} the number of relativistic particles. The volume factor $V \propto a^3$ and $p \propto a^{-1}$, giving rise to the $\rho_{\mathrm{rad}} \propto a^{-4}$. This agrees with the assignment $w = 1/3$ for radiation. For non-relativistic particles with mass m instead $\rho_{\mathrm{mat}} = mN_{\mathrm{mat}}/V \propto a^{-3}$, the redshift now only coming from the volume factor.

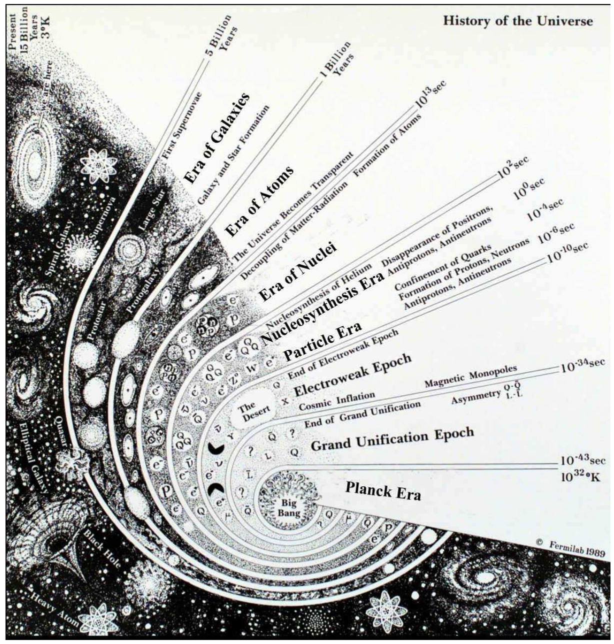

Figure 2: Timeline of the universe.

- Vacuum energy (a cosmological constant) ρ_Λ with $\omega_\Lambda = -1$ remains constant in time. If it *dominates* the universe $a \propto \mathrm{e}^{Ht}$.

Define $\Omega_i = \rho_i/\rho_c$ with $\rho_c = 3H^2$ the critical density. Then the Friedmann equation (1.7) becomes

$$\Omega = \sum_i \Omega_i = 1 + \frac{k}{(aH)^2} \tag{1.13}$$

Thus Ω is larger, equal, or smaller than unity for an open, flat or closed universe respectively. From observations (CMB data, supernovae, large scale structure, lensing, big bang nucleosynthesis (BBN)) we find for the present values

$$\Omega - 1 \approxeq 0, \quad \Omega_{\mathrm{B}} \approxeq 0.04, \quad \Omega_{\mathrm{DM}} \approxeq 0.23, \quad \Omega_\gamma \approxeq 8 \times 10^{-5}, \quad \Omega_\Lambda \approxeq 0.072 \tag{1.14}$$

with B and DM denoting baryons and dark matter. Visible matter only makes up a very small part.

Thermal history of the universe Hubble's law and other observations indicate the universe is expanding. The temperature of the radiation bath in the universe $\rho_{\mathrm{rad}} \propto T^4 \propto a^{-4}$, where for the first expression we used Stefan-Boltzman's law. It follows that the temperature decreases as $T \propto a^{-1}$ with the expansion. Initially the universe is hot and dense (because small), and it coolse as it expands with time. Key events in history of the universe are summarized by fig 2.

- Big bang nucleosythesis (BBN) at $t \sim 10^2$s. As the temperature drops below MeV isotopes of light nuclei (hydrogen, helium, lithium,...) are formed from protons and neutrons. Theoretical predictions and observations are in excellent agreement about the primordial abundances (75% hydrogen, 25% helium, and trace amounts of heavier elements). BBN constitutes the earliest direct evidence for big bang fireball picture of a universe in thermal equilibrium, and that coolse with the expansion of the universe $T \propto a^{-1}$. Everything that happened before BBN is speculation, in the sense that there is no direct observational evidence.

- Matter - radiation equality at 10^4yr or $T \sim 1$eV. Intially, at high temperatures the universe is radiation dominated. Since radiation red shifts faster than cold matter, at some point the latter comes to dominate the energy density. Small primordial density perturbations start growing in the radiation dominated universe.

- Recombination at 10^5yr or $T \sim 0.1\,$eV. As the temperature drops below the binding energy nuclei and electrons combine to form neutral atoms. The free path length of the photons (set by electromagnetic interactions) suddenly increases enormously, and photons from this period reach us unscattered. These are the CMB photons, they come from the "surface of last scattering". Due to the red shift we observe them today in the microwave range with $T = 2.73K$.

- Formation of gravitational bound states/galaxies at $\sim 10^9$yr or $T \sim 10^{-3}$eV. This happened only rather recently at red shifts $z \lesssim 10$.

- Present with $T \approx 2.73\,\mathrm{K} \sim 10^{-4}$eV.

1.2 Shortcomings of the big bang cosmology

The big bang, the picture of the universe as an exploding fireball, is confirmed at low red shifts by the presence of the CMB, BBN, and Hubble's law. However, there are some problems as well.

Horizon problem Despite the fact that the early universe was vanishingly small, the rapid expansion precluded causal contact from being established throughout. The CMB has a perfect black body spectrum. Two photons coming from opposite direction have nearly equal temperatures. Yet these photons come from different regions, that at the time of last scattering were not in causal contact with each other.

Photons travel on null geodesics with $\mathrm{d}s^2 = 0 \;\Rightarrow\; \mathrm{d}r = \mathrm{d}t/a(t)$ for a radial path. The particle horizon is the maximum distance a light ray can travel between $t = 0$ and t (and thus gives the size of a causal region)

$$\begin{aligned} R_p(t) &= a(t)\int_0^t \frac{\mathrm{d}t'}{a(t')} = a(t)\int_0^a \frac{\mathrm{d}(\ln a)}{(aH)} \\ &= \frac{t}{(1-n)} \sim H^{-1} \propto \begin{cases} a^{3/2} & \text{(MD)} \\ a^2 & \text{(RD)} \end{cases} \end{aligned} \tag{1.15}$$

In the second line we used $a \propto t^n$ with $n < 1$ as apropiate for normal matter such as matter and radiation. Particles separated by a distance $d_{\rm phys} > R_H$ never could have talked to each other. Particles separated by a distance $d_{\rm phys} > H^{-1}$ can have no causal contact *now*, as they are flying away from each other with a velocity greater than the speed of light. This can be seen from Hubbles law (1.3). This defines the Hubble horizon R_H; in comoving coordinates $d_{\rm com} > (aH)^{-1}$ with the rhs comoving Hubble radius. In big bang cosmology the particle and Hubble horizon are or the same order of magnitude $R_p \sim R_H \sim H^{-1}$, and they are often used (sloppily) interchangebly. But caution should be taken, as in an inflationary spacetime, discussed in the next section, the two concepts do differ.

Note that the particle horizon (1.15) is set by the comoving Hubble radius $(aH)^{-1}$. Physical lengths are stretched by the expansion $\lambda \propto a$. Since $(aH)^{-1}$ grows w/ time, so does the ratio R_p/λ: scales that are inside the horizon at present were outside at earlier times, see 4. Concretely, consider two CMB photons emitted, which were emitted at the time of last scattering. Nowadays we see the on the sky a separated by a distance $\lambda(t_0) < R_p(t_0)$ (because we cannot see beyond the hubble horizon). Extrapolating back in time the the time of last scattering, it follows that $\lambda(t_{\rm ls}) > R_p(t_{\rm ls})$ was larger than the horizon. No causal physics could have acted at such large scales. Yet, although these photons come from caussally disconnected regions, to a very good precision they have nearly the same temperature. How is this possible? This is the horizon problem.

The volume that is now within our horizon consist of $\sim 10^6$ causally disconnected regions at the time of last scattering. The computation goes as follows. The length scale corresponding to our present horizon $R_H(t_0) \sim H_0^{-1}$ (= observable part of our universe) at the time of last scattering was:

$$\lambda_H(t_{ls}) = R_p(t_0)\left(\frac{a_{ls}}{a_0}\right) = R_p(t_0)\left(\frac{T_0}{T_{ls}}\right). \tag{1.16}$$

The particle horizon at last scattering is

$$R_p(t_{ls}) \sim H_{ls}^{-1} = H_0^{-1}\left(\frac{H_0}{H_{ls}}\right) \sim R_p(t_0)\left(\frac{T_{ls}}{T_0}\right)^{-3/2} \tag{1.17}$$

where we used that during matter domination $\rho_m \sim H^2 \propto a^{-3} \propto T^3$. Indeed $R_p(t_{ls}) \ll \lambda_H(t_{ls})$. Comparing the volumes of these two scales

$$\frac{\lambda_H^3}{R_H^3} \sim \left(\frac{T_0}{T_{ls}}\right)^{-3/2} \sim 10^6 \tag{1.18}$$

Flatness problem Consider the Friedmann equation in the form

$$\Omega - 1 = \frac{k}{(aH)^2}. \tag{1.19}$$

The comoving Hubble radius $(aH)^{-1}$ grows with time, and thus $\Omega = 1$ is an unstable fixed points. Indeed

$$\frac{|\Omega - 1|_{\rm pl}}{|\Omega - 1|_0} \sim \left(\frac{a_{\rm pl}}{a_0}\right)^2 \sim \left(\frac{T_0}{T_{\rm pl}}\right)^2 \sim \mathcal{O}(10^{-64}) \tag{1.20}$$

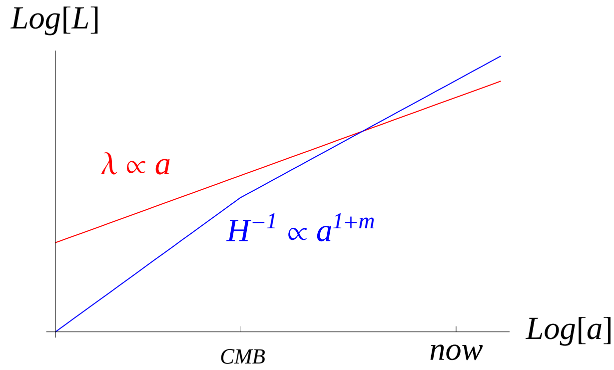

Figure 3: Timeline of the universe.

Figure 4: Evolution of a physical scale λ and Hubble horizon H^{-1} with the expansion of the universe parametrized by $\ln a$. The power $m = 1\,(1/2) \leq 1$ during radiation (matter) domination. Subhorizon physical scales today ($\lambda < R_H$) were superhorizon at early times ($\lambda > R_H \sim R_p$) .

where in the 1st step we took a radiation dominated universe (valid up till recombination), and in the last step we set $|\Omega - 1|_0 = \mathcal{O}(1)$. To have a flat universe at present, the value of Ω at earlier times need to be extremely fine-tuned.

When the strong energy condition $(1 + 3\omega > 1)$ is satisfied, $\Omega = 1$ is an unstable fixed point. Calculation:

$$\begin{aligned} \dot{\Omega} &= -\frac{2k}{(aH)^3}(H\dot{a} + \dot{H}a) = -\frac{2k}{(aH)^3}\left(\frac{a\rho}{3} - \frac{k}{a} - \frac{a}{2}(\rho + p) + \frac{k}{a}\right) \\ &= -\frac{2k}{(aH)^3}\frac{a}{6}(\rho + 3p) = H(\Omega - 1)\Omega(1 + 3\omega) \end{aligned} \tag{1.21}$$

In the 2nd step we used (1.7, 1.8), in particular

$$\dot{H} = \frac{\ddot{a}}{a} - H^2 = -\frac{1}{2}(\rho + p) + \frac{k}{a^2}. \tag{1.22}$$

In the last step we used $\Omega = \rho/\rho_c$ with $\rho_c = 3H$. We can rewrite the above equation as

$$\frac{\mathrm{d}|\Omega - 1|}{\mathrm{d}\ln a} = \Omega|\Omega - 1|(1 + 3\omega). \tag{1.23}$$

It follows that in an expanding universe with a growing with time, $|\Omega - 1|$ grows if $(1+3\omega) > 0$, or $\omega > -1/3$. The strong energy condition is satisfied for normal matter and radiation.

Monopole problem If the universe can be extrapolated back it time to high temperatures (remember, we only have direct evidence for the big bang picture for low temperatures $T < T_{\rm BBN} \sim$MeV), it is likely the universe went through a series of phase transitions during its evolution. There are the electroweak and QCD phase transition, and possibly other ones at (much) higher scales, such as grand unified theory (GUT) phase transition(s). Depending on the symmetry broken in the phase transition topological defects — domain walls, cosmic strings, monopoles or textures — may form. If a semi-simple GUT group is broken down to the SM, either directly or via some intermediate steps, monopoles form.

Monopoles are heavy pointlike objects, which behave as cold matter $\rho_{\rm mp} \propto a^{-3}$. If produced in the early universe the energy density in monopoles decreases slower than the radiation background, and comes to dominate the energy density in the universe early on (it "overcloses" the universe), in conflict with observations.

Whether monopoles (or other defects) are a problem is a model dependent issue.

1.3 Problems

P1.1 Metric on the 2-sphere The 2-sphere can be embedded in 3D Euclidean space $\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$ via the embedding equation $x^2 + y^2 + z^2 = R^2$. Show that the meteric of the sphere can be brought in the form

$$\mathrm{d}l^2 = R^2 \left(\frac{\mathrm{d}r^2}{1 - r^2} + r^2 \mathrm{d}\theta^2 \right)$$

Notice the similarity between the FRW metric with $k = 1$.

P1.2 Friedmann equation The Friedman equation is one of the most important equations in cosmology. A proper derivation is done using the Einstein equation. But it can also be derived using classical mechanics. This gives some intuition about its meaning.

a) Consider a large mass M at rest, and a probe mass m a distance a away, moving radially away from M with velocity v. What is the total energy of the system (in the limit $m \ll M$)? Calculate the escape velocity for the small mass. How does $a(t)$ change with time for m moving with the escape velocity?

b) Consider now a spherically expanding universe, with mass $M = \frac{4}{3}\pi a^3 \rho$. A probe particle with mass $m \ll M$ is located at the edge of the sphere. The energy of the system is the same as in part a). Use Hubble's law to express the velocity of m in terms of H to derive the Friedmann equation. What is the relation between k and E?

c) What is the faith of the universe for $k = 0$ (note that since we considered a sphere of matter, this corresponds to a universe with $\Omega_{\rm mat} = 1$)? And for $k < 0$ and $k > 0$?

d) Consider the Friedmann equation with both cold matter and a cosmological constant. What is now the faith of the universe for a closed universe with $k = -1$?

P1.3 Continuity equation The continuity equation also has a simple interpretation. Show that it follows from the first law of thermodynamics $U = -p\mathrm{d}V$.

2. Inflation

2.1 Big bang problems revisited

The flatness and horizon problem are initial value problems. In principle it is possible to tune initial conditions in the big bang cosmology so that our current universe emerges. But the amount of tuning is enormous.

Inflation — the idea that the early universe went through a period of superluminal expansion — solves all problems in one go. The flatness and isotropy emerges dynamically, and not as a result of special initial conditions, and the monopole problem is addressed as well. Inflation explains why our observable universe is so large, flat, and (on large scales) homogenous. As a bonus, quantum fluctuations during inflation can provide the seeds for structure formation.

Inflation is a period of accelerated expansion, defined by:

$$\ddot{a} > 0 \quad \Leftrightarrow \quad (\rho + 3p) < 0 \quad \Leftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}(aH)^{-1} < 0 \tag{2.1}$$

The accelerated expansion reverses the behavior of the comoving Hubble radius, it will decrease instead of increase. The growing comoving Hubble was the root of both the flatness and horizon problem. To get inflation the strong energy condition has to be broken. The equivalences above follow from (1.7, 1.8, 1.9). Indeed the 2nd equivalence comes from (1.8), whereas the 3rd follows from a calculation analogous to (1.21, 1.22): $\partial_t(aH)^{-1} = -(aH)^{-2}(a\dot{H} + \dot{a}H) = (aH)^{-2}(a/6)(\rho + 3p)$.

We will now show how inflation overcomes the shortcomings of the big bang. Consider a fluid with $\rho \approx -p$ (postpone for the moment the discussion how to get something like that). Then ρ_I, $H_I \approx$ constant and $a \propto \mathrm{e}^{H_I t}$ with H_I the Hubble constant during inflation. The universe during inflation is close to a deSitter (dS) geometry.

Horizon problem As we have seen the problem in big bang cosmology is that the particle horizon $R_p \sim H^{-1}$ increases faster than a physical length scale $\lambda \propto a$ with time, and thus extrapolating back a scale that now is inside the horizon $\lambda < R_H$, it was outside at earlier times. This problem can be solved if in the early universe there is a (long enough) phase in which λ decreases faster than the horizon, which is the case during inflation during which the comoving Hubble distance $(aH)^{-1}$, which enters the particle horizon in (1.15), decreases. Indeed, the ratio

$$\frac{R_p}{\lambda} \propto \int_0^a \frac{\mathrm{d}\ln a}{(aH)} \tag{2.2}$$

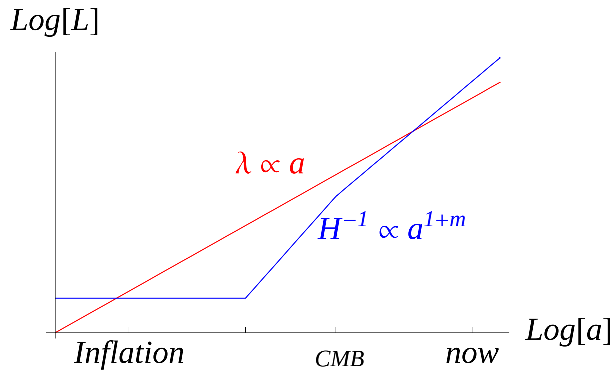

Figure 5: Evolution of physical scale λ and Hubble horizon H^{-1} with the expansion of the universe parametrized by $\ln a$. Subhorizon physical scales were superhorizon at intermediate times, and subhorizon again during inflation provided $N \gtrsim 70$.

decreases with time during inflation, since $(aH)^{-1}$ decreases.

Another way to look at is is that if we want the homogenity and isotropy of the CMB photons to be explained by causal physics, the lenght scale we observe today should have been inside the Hubble horizon at some point before the surface of last scattering to homogenize the temperature/initial conditions. In big bang cosmology, the largest CMB scales are superhorizon at all times before last scattering, as explained in the previous section. But with inflation in the picture things change. Since the hubble constant during inflation H_I nearly constant, the behavior of λ/H^{-1} changes: whereas it is decreasing during the big bang phase following inflation (and giving rise to the horizon problem of big bang cosmology), it increases during inflation. With this reversed behavior it is possible that two CMB photons we observe today, were at some early time before last scattering separated by a distance smaller than the Hubble horizon and thus causal physics could homogenize them, resulting in their nearly equal temperature we measure today. This is illustrated in fig. 5.

To solve the horizon problem the largest scales observed today should be within the horizon at the beginning of inflation. Take $\lambda(t_0) = H_0^{-1}$, then this means

$$\lambda(t_i) \sim H_0^{-1}\left(\frac{a_f}{a_0}\right)\left(\frac{a_i}{a_f}\right) = H_0^{-1}\left(\frac{T_0}{T_f}\right)\mathrm{e}^{-N} < H_I^{-1} \tag{2.3}$$

Here t_i, t_f are the time inflation begins and ends, and $N \equiv \ln(H_I(t_f - t_i))$ the number of efolds of inflation. The bound on the number of e-folds is

$$N > \ln\left(\frac{T_0}{H_0}\right) - \ln\left(\frac{T_f}{H_I}\right) \approx 67 - \ln\left(\frac{T_f}{H_I}\right) \sim 60 + \log\left(\frac{T_f}{10^{15}\mathrm{GeV}}\right) \tag{2.4}$$

where in the last step used that inflationary energy is transferred to radiation instantaneously with $T_f^4 \sim \rho_I$.

Flatness problem The freedman equation (1.19) during inflation

$$\Omega - 1 = \frac{k^2}{(aH)^2} \propto \mathrm{e}^{-2N} \to 0 \tag{2.5}$$

goes closer to zero the longer inflation takes. From (1.20) it follows one needs $|\Omega - 1|_{t_f} \lesssim 10^{-60}$ to avoid tuning of the initial conditions, with t_f the time inflation ends. Since

$$\frac{|\Omega - 1|_{t_f}}{|\Omega - 1|_{t_i}} = \left(\frac{a_i}{a_f}\right)^2 = \mathrm{e}^{-2N} \tag{2.6}$$

this requires $N \gtrsim 60 - 70$.

Inflation predicts $\Omega_0 = 1$.

Heuristically, if you consider space-time as the surface of a 2-sphere (lower dimensional analogue), and expand the sphere by an enormous amount, than for a local observer living on the surface the geometry is indistinguishable from flat.

Monopole problem If inflation takes place after the phase transition during which monopoles form, the monopole density is diluted by inflation to harmless size

$$n_{\mathrm{mp}} \propto \frac{N_{\mathrm{mp}}}{a^3} \to 0. \tag{2.7}$$

2.2 Scalar field dynamics

The vacuum like period that drives inflation must be dynamic, it cannot be a true cosmological constant as inflation must end. How to violate the strong energy condition and get a system with $\rho \approx -p$ (matter with negative pressure)? The answer is scalar fields. Note that scalar fields have the same quantum numbers as the vacuum, and thus can mimic a vacuum like state. Expectation values of scalar fields can be non-zero without breaking Lorentz invariance.

Consider the action for a scalar field φ, which we will call the inflaton field,

$$S = \int \mathrm{d}^4 x \sqrt{-g} \left[\frac{1}{2} R + \mathcal{L}_\varphi\right] \tag{2.8}$$

with

$$\mathcal{L}_\varphi = -\frac{1}{2} g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - V(\varphi) \tag{2.9}$$

For a FRW metric $\sqrt{-g} \equiv \sqrt{-\det(g_{\mu\nu})} = a^3$. The Euler-Lagrange equations are

$$\partial^\mu \left(\frac{\delta(\sqrt{-g}\mathcal{L})}{\delta(\partial^\mu \varphi)}\right) - \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta\varphi} = 0 \quad \Rightarrow \quad -\partial^\mu (a^3 \partial_\mu \varphi) - V_\varphi a^3 = 0 \tag{2.10}$$

Using the Leibniz rule for the 1st term, the fact that $a(t)$ is a function of time only, and dividing by a factor a^3, then gives

$$\ddot{\varphi} + 3H\dot{\varphi} - \frac{\nabla^2}{a^2}\varphi + V_\varphi = 0. \tag{2.11}$$

The expansion of the universe provides a friction term for the scalar field proportional to the expansion rate $H = \dot{a}/a$.

The energy momentum tensor is

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_\varphi)}{\delta g_{\mu\nu}} = \partial_\mu\varphi\partial_\nu\varphi + g_{\mu\nu}\mathcal{L}_\varphi \tag{2.12}$$

from which the energy and momentum follow:

$$\begin{aligned} \rho = T_{00} &= \frac{1}{2}\dot{\varphi}^2 + V + \frac{1}{2}\frac{(\nabla\varphi)^2}{a^2} \\ p = T_{ii} &= \frac{1}{2}\dot{\varphi}^2 - V - \frac{1}{6}\frac{(\nabla\varphi)^2}{a^2} \end{aligned} \tag{2.13}$$

Consider the homogeneous mode and set $\nabla\varphi/a \to 0$ (inflation rapidly smooths out spatial variation, so the approximation is certainly valid after a couple of e-folds of inflation). Then $\omega_\varphi = p/\rho = (\frac{1}{2}\dot{\varphi}^2 - V)/(\frac{1}{2}\dot{\varphi}^2 + V) \approx -1$ for $\frac{1}{2}\dot{\varphi}^2 \ll V$. If in addition the inflaton dominates energy density in the universe $\rho_\varphi \gg \rho_{\rm rad}$, ρ_m, then $\omega \approx \omega_\varphi$. From the Friedmann equation it then follows that we get a nearly exponentially growing scale factor.

The scalar potential is non-zero $V \neq 0$ if φ is displaced from the minimum of the potential (assuming $V \approx 0$ at present to get a nearly vanishing cosmological constant in agreement with observations).

Slow roll inflation. The slow roll approximation consists of

$$\dot{\varphi}^2/2 \ll V \qquad \Rightarrow \qquad H^2 = \rho/3 \approx V/3 \tag{2.14}$$

$$\ddot{\varphi} \ll 3H\dot{\varphi} \qquad \Rightarrow \qquad 3H\dot{\varphi} \approx -V_\varphi \tag{2.15}$$

The first approximations assures that H is nearly constant $\dot{H} \ll H^2$, leading to quasi-exponential expansion, i.e. inflation with $a \sim \mathrm{e}^{Ht}$. The second approximation allows to drop the double derivative in the inflaton equation of motion (2.11), and assures inflation is prolonged. The the slow roll conditions are equivalent to requiring the slow roll parameters

$$\epsilon \equiv \frac{1}{2}\left(\frac{V_\varphi}{V}\right)^2, \qquad \eta \equiv \frac{V_{\varphi\varphi}}{V} \tag{2.16}$$

to be small: ϵ, $|\eta| \ll 1$.

Proof: The 1st condition (2.15) can be rewritten

$$\frac{\frac{1}{2}\dot{\varphi}^2}{V} = \frac{V_\varphi^2}{9H^2V} = \frac{1}{6}\left(\frac{V_\varphi}{V}\right)^2 \equiv \frac{\epsilon}{3} \tag{2.17}$$

where we used (2.15) in the 1st step and (2.14) in the 2nd. The 2nd condition (2.15) can be rewritten as follows. First note that differentiating (2.14) gives $2H\dot{H} = 1/3\dot{V} = 1/3V_\varphi\dot{\varphi}$, and thus

$$\frac{\dot{H}}{H^2} = \frac{V_\varphi\dot{\varphi}}{6H^3} = \frac{V_\varphi^2}{18H^4} = -\frac{1}{2}\left(\frac{V_\varphi}{V}\right)^2 = -\epsilon \tag{2.18}$$

Differentiating (2.15) gives

$$\ddot{\varphi} = -\frac{V_{\varphi\varphi}\dot{\varphi}}{3H} + \frac{V_\varphi \dot{H}}{3H^2} = -\frac{V_{\varphi\varphi}\dot{\varphi}}{3H} + \frac{V_\varphi}{3}\epsilon \tag{2.19}$$

which allows to write the 2nd condition as

$$\frac{\ddot{\varphi}}{3H\dot{\varphi}} = \frac{1}{2}\left(\frac{V_\varphi}{V}\right)^2 - \frac{V_{\varphi\varphi}}{V} \equiv -\frac{1}{3}(\epsilon + \eta) \tag{2.20}$$

Define N as the number of e-folds left to the end of inflation.

$$N(\varphi) = \ln\frac{a_f}{a} = \int_t^{t_f} H\mathrm{d}t = H\int_\varphi^{\varphi_f}\frac{\mathrm{d}\varphi}{\dot{\varphi}} \approx \int_{\varphi_f}^{\varphi}\frac{V}{V_\varphi}\mathrm{d}\varphi \tag{2.21}$$

where in the last step we used (2.14, 2.15). Here $\varphi_i = \varphi(t_i)$. Inflation ends when slow roll is violated $\epsilon(\varphi_f) \approx 1$ (or $\eta(\varphi_f) \approx 1$).

A length scale λ crosses the horizon when $\lambda \approx H^{-1}$. CMB scales, which correspond to the largest scales observed today (i.e. are of the present horizon scale) leave the horizon $N_* \sim 60$ before the end of inflation, see (2.4) and Fig. 5 [4].

Example: chaotic inflation with a quadratic potential $V = \frac{1}{2}m^2\varphi^2$. The slow roll parameters $\epsilon = \eta = 2/(\varphi^2)$ are small for superplanckian field values $\varphi \gg 1$. (Note that $V \ll 1$ otherwise quantum gravity is needed). Inflation proceeds if the inflaton has initially very large field values, and slowly rolls down its potential until $\varphi_f \sim 1$ when $\epsilon \sim 1$ and inflation ends. Observable scales leave the horizon when

$$N_* = \int_{\varphi_f}^{\varphi_*}\frac{\varphi}{2}\mathrm{d}\varphi \approx \frac{1}{4}\varphi_*^2 \qquad \Rightarrow \qquad \varphi_* \approx \sqrt{4\times 60} \approx 15. \tag{2.22}$$

2.3 Reheating

The universe should be reheated after the end of inflation, to start the successful big bang cosmology. During inflation particle densities $n \propto a^{-3} \to 0$ are diluted to basically zero; the universe is empty and cold. At end of inflation the vacuum energy density stored in the inflaton field is to be transferred to radiation, via inflaton decay $\varphi \to$ radiation.

2.4 Problems

P2.1 Slow roll parameters Show that (2.16) is equivalent to the slow roll approximation in (2.14, 2.15).

[4] The subscript $*$ will be used to denote the corresponding quantity at the time observable scales leave the horizon

3. Cosmological perturbations

Primeval density inhomogeneities are amplified by gravity, and grow into the structures we see today: galaxy clusters, galaxies, planets, everything.

A fluid of self-gravitating particles is unstable to growth of small inhomogeneities; this is called the Jeans instability. The essence of gravitational instability can already be seen at the level of Newtonian perturbation theory. Consider Minkowski space time filled with incompressible matter $p = 0$. If there is some inhomogeneity $\delta\rho$ in some particular point in space, this inhomogeneity starts to attract nearby matter towards the point, and according to Newton's law the attracting force is proportional to $\delta\rho$. Therefore, $\ddot{\delta\rho} \propto \delta\rho$, and it follows an exponential instability develops. A proper analysis should include background expansion of the FRW universe. This will tame the instability (expansion leads to friction), making the instability grow as power law instead of exponential, but does not remove it.

Quantum fluctuations during inflation provide the initial inhomogeneities (the seeds) for structure formation. Heuristically this can be understood as follows. The quantum vacuum is never empty, particle and anti-particle pairs constantly pop out of the vacuum and annihilate again. During inflation, due to the enormous expansion, the particle and antiparticle are ripped apart, and they may get separated by a distance larger than the causal horizon H^{-1}, and cannot find each other again to annihilate. They remain as perturbations on the background.

Fluctuations are stretched to superhorizon size $\lambda > H^{-1}$ by the expansion, and get "frozen in". The amplitude remains approximately constant on superhorizon size, the wavelength grows $\propto a$. The result is the appearance of a classical field $\delta\varphi$ that does not vanish after averaging over a macroscopic time interval.

The quantum fluctuations of the inflaton source perturbations in the metric and vice versa.

3.1 Quantum fluctuations of generic massless scalar field during inflation

Consider a scalar field other than the inflaton. Its fluctuations will give a subdominant contribution to the total energy density and consequently the back reaction on the metric is negligible small. To determine the quantum fluctuations produced during inflation, we can then study the perturbed KG equation for this scalar in a fixed spacetime background. For the inflaton field the equations derived below are still valid, but only in a special gauge (spatially flat gauge — more on this in the next subsection) for which the metric and inflaton perturbations decouple.

Split the scalar field in a homogeneous background field plus a fluctuation $\varphi(x,t) = \varphi(t) + \delta\varphi(x,t)$. The fluctuation can be expanded in Fourier modes

$$\delta\varphi(x,t) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^{3/2}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \delta\varphi_k(t) \tag{3.1}$$

with $\mathbf{k}, \mathbf{x}$ comoving momenta and distance (physical momenta and distance scale $x_{\text{phys}} = a(t)x$ and $k_{\text{phys}} = k/a(t)$). Because of isotropy of the background φ_k only depends on $k = |\mathbf{k}|$.

Perturbing the KG equation (2.11) gives

$$\delta\ddot{\varphi}_k + 3H\delta\dot{\varphi}_k + \frac{k^2}{a^2}\delta\varphi_k + V_{\varphi\varphi}\delta\varphi_k = 0 \tag{3.2}$$

Using the slow roll condition $V_{\varphi\varphi} \ll H^2$ (2.16) for the inflaton (or taking the massless limit if φ is not the inflaton) allows to neglect the last term.

Now switch to conformal time

$$\mathrm{d}\eta \equiv \frac{\mathrm{d}t}{a(t)} \tag{3.3}$$

which brings the FRW metric (1.1) in the form

$$\mathrm{d}s^2 = a^2(\eta)\left[-\mathrm{d}\eta^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right] \tag{3.4}$$

conformal to the Minkowski line element. During a dS phase $a = \mathrm{e}^{Ht}$ and thus

$$\eta = \int \mathrm{d}t\mathrm{e}^{-Ht} = -\frac{1}{Ha} \qquad (\eta < 0) \tag{3.5}$$

We fixed the integrations constant such that the beginning of inflation corresponds to some initial time $\eta_i \ll 1$, whereas $\eta \to 0$ as $a,\, t \to \infty$. Further introduce the variable

$$v_k = a\delta\varphi_k \tag{3.6}$$

With these definitions (3.2) becomes

$$v_k'' + \left(k^2 - \frac{a''}{a}\right)v_k = 0 \tag{3.7}$$

with primes indicating derivatives w.r.t. conformal time. [Algebra:

$$\begin{aligned}
\dot{\varphi}_k &= \frac{1}{a}\partial_\eta\left(\frac{v}{a}\right) = \frac{v'}{a^2} - \frac{a'v}{a^3} \\
\ddot{\varphi}_k &= \frac{1}{a}\partial_\eta\left(\frac{v'}{a^2} - \frac{a'v}{a^3}\right) = -\frac{2a'v'}{a^4} + \frac{v''}{a^3} - \frac{a''v}{a^4} + \frac{3{a'}^2v}{a^5} - \frac{a'v'}{a^3} \\
H &= \frac{\dot{a}}{a} = \frac{a'}{a^2}
\end{aligned} \tag{3.8}$$

where it was used that $\partial_t = a^{-1}\partial_\eta$. Plugging in (3.2) gives

$$\frac{1}{a^3}\left[-3\frac{a'}{a}v' + v'' - \frac{a''}{a}v + 3\frac{a'^2}{a^2}v + 3\frac{a'}{a}v' - 3\frac{a'^2}{a^2} + k^2v\right] = 0 \tag{3.9}$$

and thus (3.7)].

The equation (3.7) is that of a collection of decoupled harmonic oscillators with time dependent frequency, one for each k. Eq. (3.7) can be obtained from an action

$$S^{(2)} = \int \mathrm{d}\eta\mathrm{d}\mathbf{x}^3\frac{1}{2}\left[v'^2 - (\nabla v)^2 + \frac{a''}{a}v^2\right] = \int \mathrm{d}\eta\mathrm{d}\mathbf{x}^3\mathcal{L}^{(2)} \tag{3.10}$$

which could also have been found perturbing the original action S to second order in perturbations (this is needed to get the correct normalization). It is the action for a canonically normalized (standard kinetic terms) free field with an effective time-dependent mass. One can canonically quantize the action in the usual way. Define the canonical momentum

$$\pi = \frac{\partial \mathcal{L}^{(2)}}{\partial v'} = v'. \tag{3.11}$$

Promote the classical fields $\{v, \pi\}$ to quantum operators $\{\hat{v}, \hat{\pi}\}$. Impose equal time canonical commutation relations $[\hat{v}(\eta, \mathbf{x}), \hat{\pi}(\eta, \tilde{\mathbf{x}})] = i\hbar\delta^{(3)}(\mathbf{x} - \tilde{\mathbf{x}})$. Define the mode decomposition

$$\hat{v}(\eta, \mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^{3/2}} \left[v_k(\eta)\hat{a}_k \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} + v_k^*(\eta)\hat{a}_k^\dagger \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} \right] \tag{3.12}$$

The $\hat{a}$ and $\hat{a}^\dagger$ are annihilation and creation operators with the familiar commutation relation $[\hat{a}_k, \hat{a}_{\tilde{k}}^\dagger] = \delta^{(3)}(\mathbf{k} - \tilde{\mathbf{k}})$. Finally, the mode functions are normalized $(v_k^* v_k' - (v_k^*)' v_k) = i/\hbar$; it follows from substituting (3.12) in the commutation relation for $\{\hat{v}, \hat{\pi}\}$, and demanding the usual normalization of the commutation relation for the creation and annihilation operators. The vacuum is defined via $\hat{a}_k \,|0\rangle = 0$ for all k.

The mode functions $v_k(\eta)$ satisfy the classical equation of motion (3.7). Now determine its solutions. Consider first the subhorizon limit with the physical wavelenght of the perturbation smaller than the causal horizon $\lambda a < H^{-1}$. This is equivalent to $k > (aH) = -1/\eta$ where the last equation applies to dS space; here λ, k are comoving wavelenth and momenta . In this limit $k^2 \gg a''/a = 2/\eta^2$ and the mode equation reduces to a harmonic oscillator with time-independent frequency $v_k'' + k^2 v_k = 0$. Spacetime is locally Minkowski, on scales much smaller than the curvature radius of dS. In Minkowski space there is a unique solution (requiring the vacuum to be the minimum energy eigenstate), which is the positive frequency solution

$$\lim_{k\eta \gg -1} v_k = \frac{\mathrm{e}^{-ik\eta}}{\sqrt{2k}} \tag{3.13}$$

which is properly normalized, and satisfies $v_k' = -i\omega_k v$ with frequency $\omega_k^2 = k^2$. This choiche of vacuum is called the Bunch-Davies vacuum. Modes oscillate inside the horizon.

Now the superhorizon limit $a\lambda > H^{-1}$, or $k < (aH) = -1/\eta$. The equation of motion reduces to $v_k'' - (a''/a)v_k = 0$. The growing mode solution is $v_k = B_k a$ with B_k an integration constant (in dS the other solution is $v_k \propto a^{-2}$). The scalar field perturbation is constant on superhorizon scales $\delta\varphi_k = v_k/a$ =const, from (3.6). Modes are frozen outside the horizon. Matching the solution for sub- and superhorizon scale at horizon exit $k = aH$ $(-k\eta = 1)$ fixes the integration constant $a|B_k| = k/H|B_k| = 1/\sqrt{2k}$. On superhorizon scales the amplitude is $|\delta\varphi_k| \approx H/\sqrt{2k^3}$.

The matching procedure can generically be used. For the present case the mode equation can also be solved exactly

$$v_k = \alpha \frac{\mathrm{e}^{-ik\eta}}{\sqrt{2k}} \left(1 - \frac{i}{k\eta}\right) + \beta \frac{\mathrm{e}^{ik\eta}}{\sqrt{2k}} \left(1 + \frac{i}{k\eta}\right) \tag{3.14}$$

The integration constants α, β can be fixed by the boundary condition, which is the requirement that the solution reduces to the Minkowski result in the subhorizon limit. This sets $\alpha = 0$, $\beta = 1$.

Quantum to classical transition Quantum fluctuations can be regarded as classical when their corresponding wavelengths cross the horizon, which motivates the usual description of cosmological perturbations in terms of classical random fields. The transition occurs roughly at horizon exit. Indeed on super horizon scales $v_k \sim \hbar^{1/2}/(\eta k^{3/2})$, where we reinserted factors of $\hbar$. And thus both terms in the Wronskian are $\sim v_k^* v_k' \sim v_k^2/\eta \sim \hbar/(\eta k)^3 \gg 1$, and the non-commutativity can be neglected. Much more refinement is needed to make things precise.

Power spectrum Consider a generic quantity $g(\mathbf{x}, t)$ which can be expanded in Fourier modes as

$$g(t, \mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^{3/2}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} g_k(t) \tag{3.15}$$

The power spectrum $\mathcal{P}_g(k)$ is defined via [5]

$$\langle g_{k_1}^* g_{k_2} \rangle = \delta^{(3)}(\mathbf{k}_1 - \mathbf{k}_2) \frac{2\pi^2}{k^3} \mathcal{P}_g(k) \tag{3.16}$$

The power spectrum gives the power per logarithmic momentum interval:

$$\langle g^2(t, \mathbf{x}) \rangle = \int \frac{\mathrm{d}k}{k} \mathcal{P}_g(k) \tag{3.17}$$

Consider the inflaton perturbation on super-horizon scales. The corresponding power spectrum is

$$\lim_{k\eta \ll -1} \langle |\delta\varphi_k|^2 \rangle = \lim_{k\eta \ll -1} \frac{|v_k|^2}{a^2} = \frac{H^2}{2k^3} = \frac{2\pi^2}{k^3} \mathcal{P}_{\delta\varphi}(k) \tag{3.18}$$

and thus

$$\mathcal{P}_{\delta\varphi}(k) = \left(\frac{H}{2\pi} \right)^2 \tag{3.19}$$

It is scale invariant in the dS and massless limit (limit $\epsilon, \eta = 0$).

3.2 Curvature perturbation

Scalar perturbations give rise to perturbations in the energy-momentum tensor, which source the Einstein equations and thus lead to metric perturbations. Vice versa, metric perturbations back react through the perturbed Laplace operator in the KG equations, giving rise to matter fluctuations. To calculate the perturbation spectrum one needs to consider the coupled system of equations.

[5]For the quantum theory with $g = \hat{g}$ an quantum operate, $\langle = \langle 0|$. This is the formalism used to compute the quantum fluctuations during inflation. When on superhorizon scales, treating the fluctuations as a classsical random field, $\langle ... \rangle$ is the ensemble average.

Since the measured CMB perturbations are small a linearized analysis of the KG and Einstein equations suffices, and in particular we do not need a theory of quantum gravity to describe the fluctuations. We quantize the perturbations, but keep the background classical.

As we will see there is only one scalar degree of freedom corresponding to the curvature perturbation. We will construct a gauge invariant definition of this, which then can be related directly to the observed temperature fluctuations of the CMB. While the inflaton decays at the end of inflation (and $\delta\varphi$ loses its meaning), the curvature perturbation persists. For single field inflation (adiabatic perturbations) the curvature perturbation is constant on superhorizon scales, and it thus suffices to calculate its value at horizon exit. Let's discuss this all in more detail.

The metric tensor can be split in a homogeneous background piece plus small perturbations on top

$$g_{\mu\nu}(t,\mathbf{x}) = g_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\mathbf{x}) \tag{3.20}$$

The most general perturbation of the FRW metric has the form

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\phi)\mathrm{d}\eta^2 + 2\bar{B}_i \mathrm{d}x^i + (\delta_{ij} + \bar{h}_{ij})\mathrm{d}x^i \mathrm{d}x^j\right] \tag{3.21}$$

The metric perturbations can be decomposed according to their spin w.r.t. a local rotation of the spatial coordinates on hypersurfaces of constant time. This leads to scalar, vector and tensor perturbations. For example a spatial vector field $\bar{B}^i$ can be decomposed uniquely into a longitudinal part and a transverse part

$$\bar{B}_i = B_{,i} + B_i \tag{3.22}$$

with a comma denoting partial derivative $\partial_i x = x_{,i}$. The longitudinal part is curl-free and can thus be expressed as a gradient; the transverse part is divergenceless $\partial_i B^i = 0$. $\bar{B}_i$ contains one scalar and two vector d.o.f. Similarly we can decompose

$$\bar{h}_{ij} = -2\psi\delta_{ij} + 2E_{,ij} + 2E_{(i,j)} + h_{ij} \tag{3.23}$$

with h_{ij} transverse and traceless, and E_i transverse. The round brackets on the indices denote symmetrization. Thus $\bar{h}_{ij}$ contains 2 scalar, 2 vector and 2 tensor degrees of freedom.

At linear order scalar, vector and tensor perturbations decouple, and one can follow their evolution separately. No vector perturbations are produced during scalar field inflation as there are no rotational velocity fields; moreover vector perturbations decay in an expanding universe — thus we can neglect them. The most general scalar metric perturbation to 1st order can be written (in conformal time)

$$\mathrm{d}s^2 = a^2\left\{-(1+2\phi)\mathrm{d}\eta^2 + 2B_{,i}\mathrm{d}\eta\mathrm{d}x^i + [(1-2\psi)\delta_{ij} + 2E_{,ij} + h_{ij}]\right\} \tag{3.24}$$

Gauge invariance General relativity is a gauge theory where the gauge transformations are the generic coordinate transformations from a local reference frame to another. The coordinates $t, \mathbf{x}$ carry no independent physical meaning. By performing a coordinate/gauge transformation we can create "fictitious" fluctuations in a homogeneous and isotropic universe, which are just gauge artifacts. For a FRW universe there is a special gauge choice in which the metric is homogeneous and isotropic, which singles out a preferred coordinate choice. But the situation is more complicated in a perturbed universe, and we have to be careful with this.

Consider first a field/scalar perturbation in a fixed spacetime. It can be defined via $\delta\varphi(p) = \varphi(p) - \varphi_0(p)$ with φ_0 the unperturbed field and p is any point of the spacetime. Generalizing this to GR where spacetime is not a fixed background, but is perturbed if matter is perturbed, the above definition is ill defined. Indeed φ "lives" in the perturbed real spacetime $\mathcal{M}$ whereas φ_0 lives in *another* spacetime, the unperturbed reference spacetime $\mathcal{M}_0$. To define a perturbation requires an identification ι that maps points in $\mathcal{M}_0$ to points in $\mathcal{M}$. The perturbation can then be defined via

$$\delta\varphi = \varphi(\iota(p_0)) - \varphi_0(p_0) \tag{3.25}$$

However, the identification ι is not uniquely defined, and therefore the definition of the perturbation depends on the choice of map. This freedom of choosing a map is the freedom of choosing coordinates. The choice of map is a gauge choice, changing the map is a gauge transformation.

Thus fixing a gauge in GR implies choosing a coordinate system, a threading of spacetime into lines (corresponding to fixed $\mathbf{x}$) and a slicing into hypersurfaces of fixed time. A general gauge transformation reads

$$\tilde{\eta} = \eta + \xi^0, \qquad \tilde{x}^i = x^i + \xi_,{}^i + \zeta^i \tag{3.26}$$

with ξ^0, ξ arbitrary scalars and ζ^i a divergence-free 3-vector $\zeta^i{}_{,j} = 0$ Calculate how the metric functions transform under the gauge transformation (3.26) by directly perturbing the line-element. We can neglect ζ^i in the following, as only vector perturbations will depend on this. Then

$$\begin{aligned} \mathrm{d}\eta &= \mathrm{d}\tilde{\eta} - \mathrm{d}\xi^0 = \mathrm{d}\tilde{\eta} - \xi^{0'}\mathrm{d}\tilde{\eta} - \xi^0_{,i} d\tilde{x}^i \\ \mathrm{d}x^i &= \mathrm{d}\tilde{x}^i - \mathrm{d}\xi_,{}^i = \mathrm{d}\tilde{x}^i - \xi'{}_,{}^i d\tilde{\eta} - \xi_{,j}{}^i \mathrm{d}\tilde{x}^j \end{aligned} \tag{3.27}$$

where we used $\xi^0(\eta, x) = \xi^0(\tilde{\eta}, \tilde{x})$ and similar for ξ. Further expanding $a(\eta) = a\eta(\tilde{\eta}) - \xi^0 a'(\tilde{\eta})$, and plugging all in the line element (3.24) gives to first order in both the metric perturbations and the coordinate transformations

$$\begin{aligned} \mathrm{d}s^2 &= a^2(\tilde{\eta})\Big\{ -(1 + 2\phi - 2\xi^{0'} - 2\xi^0\mathcal{H})\mathrm{d}\tilde{\eta}^2 + (2\xi^0{}_{,i} + 2B_{,i} - 2\xi'{}_,{}^i)\mathrm{d}\tilde{\eta}\mathrm{d}\tilde{x}^i \\ &\qquad + [(1 - 2\psi - 2\xi^0\mathcal{H})\gamma_{ij} - 2\xi_{,ij} + 2E_{,ij}]\mathrm{d}\tilde{x}^i\mathrm{d}\tilde{x}^j \Big\} \\ &= \tilde{a}^2(\tilde{\eta}) \left\{ -(1 + 2\tilde{\phi})\mathrm{d}\tilde{\eta}^2 + 2\tilde{B}_{,i}\mathrm{d}\tilde{\eta}\mathrm{d}\tilde{x}^i + (1 - 2\tilde{\psi}\delta_{ij} + 2\tilde{E}_{,ij}) \right\} \end{aligned} \tag{3.28}$$

with the tilde quantities in the second line all a function of $\tilde{x}^\mu$; further we used that to lowest order $\tilde{a}(\tilde{\eta}) = a(\tilde{\eta})$ (corrections to this affect the results only at 2nd order). The metric functions transform as

$$\begin{aligned} \tilde{\phi} &= \phi - \xi^{0'} - \mathcal{H}\xi^0 \\ \tilde{B} &= B - \xi' + \xi^0 \\ \tilde{\psi} &= \psi + \mathcal{H}\xi^0 \\ \tilde{E} &= E - \xi \end{aligned} \tag{3.29}$$

The hubble constant in comoving time is defined

$$\mathcal{H} \equiv a'/a. \tag{3.30}$$

Any scalar ρ that is homogeneous on the background FRW model can be written as $\rho(\eta, x) = \rho_0(\eta) + \delta\rho(\eta, x)$. The perturbation in the scalar quantity transforms as

$$\begin{aligned} \tilde{\delta}\rho(\tilde{\eta}, \tilde{x}) &= \tilde{\rho}(\tilde{\eta}, \tilde{x}) - \tilde{\rho}^0(\tilde{\eta}) = \rho(\tilde{x}, \tilde{\eta}) - \rho^0(\tilde{\eta}) = \rho(x, \eta) - \rho^0(\eta) - \rho^{0'}\mathrm{d}\eta \\ &= \delta\rho(\eta, x) - \rho^{0'}\xi^0 \end{aligned} \tag{3.31}$$

where we used $\rho(\tilde{x}^\mu) = \rho(x^\mu)$. Physical scalars only depend on the choice of time slicing (temporal gauge, determined by the choice of ξ^0), but are independent of the coordinates within the constant time hypersurfaces determined by ξ.

There are two ways to proceed, and remove the gauge artifacts. Do the computation 1) in terms of gauge invariant quantities, or 2) in a fixed gauge.

Gauge fixing removes 2 of the 4 scalar degrees of freedom. There are 3 propagating d.o.f. left, 2 metric degrees of freedom and $\delta\varphi$. In the absence of anisotropic stress (no off diagonal perturbations of the energy momentum tensor $\delta T^i_j = 0$), which holds for scalar field inflation, one metric d.o.f. is removed by the off-diagonal Einstein equations which acts as a constraint. For example, in the gauge $E = B = 0$, this implies $\phi = \psi$. The remaining metric and field perturbation are related by the other Einstein equations, and are not independent. This can be understood, as without a field perturbation there is no metric perturbation, and vice versa. It follows that there is one independent scalar degree of freedom, which is directly related to the temperature anisotropies of the CMB.

The 2 d.o.f. in h_{ij} are gauge invariant, and correspond to the two (independent) polarizations.

Comoving curvature perturbation The intrinsic spatial curvature on hypersurfaces of constant conformal time η is given by (for a flat universe with $k = 0$)

$${}^{(3)}R = \frac{4}{a^2}\nabla^2\psi \tag{3.32}$$

The quantity ψ is usually referred to as the curvature perturbation. It is not gauge invariant but defined only on a given time slicing. Indeed under $\eta \to \eta + \delta\eta$ the curvature transforms $\psi \to \psi + \mathcal{H}\delta\eta$.

We now consider the comoving slicing which is defined to be the slicing orthogonal to the world lines of a comoving observer. Comoving observers are freely falling, and the expansion defined by them is isotropic. This means that there is no flux of energy measured by these observers. During inflation $0 = T_{0i} \propto \partial_i \delta\varphi(\eta, x)\varphi'(\eta)$ and thus $\delta\varphi_{\rm com} = 0$. From (3.31) we see that to go from a time slicing with arbitrary $\delta\varphi$ to go to comoving slicing requires a time translation

$$\delta\varphi \to \delta\varphi_{\rm com} = \delta\varphi - \varphi'\delta\eta = 0 \quad \Rightarrow \quad \delta\eta = \frac{\delta\varphi}{\varphi'} \tag{3.33}$$

At the same time the curvature perturbations transforms

$$\psi \to \psi_{\rm com} = \psi + \mathcal{H}\delta\eta = \psi + \mathcal{H}\frac{\delta\varphi}{\varphi'} \tag{3.34}$$

By construction the quantity

$$\mathcal{R} = \psi + \mathcal{H}\frac{\delta\varphi}{\varphi'} = \psi + H\frac{\delta\varphi}{\dot\varphi} \tag{3.35}$$

is gauge invariant. It is called the comoving curvature perturbation. The interpretation of $\mathcal{R}$ is that it gives the curvature perturbation on comoving time slices $\mathcal{R}|_{\delta\varphi=0} = \psi$.

Another commonly defined invariant quantity is the curvature on slices of constant density $\zeta|_{\delta\rho=0} = \psi$. It can be found analogously to the comoving curvature perturbation, by finding the time translation to go from an arbitrary time slice to a slice with $\delta\rho = 0$ (which gives $\delta\eta = \delta\rho/\rho'$); and then subsequently see how ψ transforms under the same time shift. The result is

$$\zeta = \psi + H\frac{\dot{\delta\rho}}{\rho} \tag{3.36}$$

On superhorizon scales and using slow roll so that $\rho' \approx V'$ (i.e. kinetic and gradient energy negligible), it is easy to show that $R \approx \zeta$. Using the continuity equation one rewrite $\zeta = \psi - \delta\rho/(3(\rho+p))$. During inflation $\rho + p = \dot\varphi^2$. Furhter $\delta\rho = \dot\varphi\delta\dot\varphi + V'\delta\varphi \approx V'\delta\varphi = -3H\dot\varphi$, where we used that on superhorizon scales the perturbation is frozen $\delta\dot\varphi \ll \varphi$ (actual calculation shows it is suprressed by the smallness of the slow roll parameters). Putting it all together gives the announced result $\mathcal{R} \approx \zeta$.

The usefulness of $\mathcal{R}$ and ζ is that they are constant on superhorizon scales $\dot\zeta = 0$. This means that all we have to do is calculate its value at horizon exit; then it remains constant until it enters the horizon again, and can be directly related to the CMB temperature fluctuations. The constancy of ζ can be derived from the conservation of energy density $T_{\mu\nu;\sigma}$ =0 (see appendix), which gives (for $\sigma = 0$)

$$\delta\dot\rho = -3H(\delta\rho + \delta p) + (\rho + p)\left[3\dot\psi - \nabla^2(\dot E + v)\right] \tag{3.37}$$

with $v_{,i}$ the perturbed 3-velocity of the fluid. On superhorizon scale the ∇^2-term is negligible. In the uniform-density gauge $\delta\rho = 0$ and $\zeta|_{\delta\rho=0} = \psi$, the energy conservation equation then gives

$$\dot\zeta = -\frac{H}{\rho + p}\delta p_{\rm non-ad} \tag{3.38}$$

Here we defined the non-adiabatic pressure perturbation via $\delta p = c_s^2 \delta\rho + \delta p_{\rm non-ad}$ with $c_s^2 = \delta p_{\rm ad}/\delta\rho$ the sound speed. For an adiabatic perturbation $\delta p \propto \delta\rho = 0$ in the uniform-density gauge, and $\dot\zeta = 0$. This is the required result: for an adiabatic perturbation the curvature perturbation ζ, which is equal to $\mathcal{R}$ on superhorizon scales, is constant.

Perturbation equations in general relativity Perturb the action (2.8) to 2nd order in the perturbations to get

$$S[g_{\mu\nu},\varphi] = S^{(0)}[g^{(0)}_{\mu\nu},\varphi^{(0)}] + S^{(2)}[\delta g_{\mu\nu},\delta\varphi; g^{(0)}_{\mu\nu},\varphi^{(0)}] \tag{3.39}$$

where $S^{(0)}$ contains only the homogeneous part, $S^{(1)} = 0$ if the background fields are extremized, and $S^{(2)}$ contains the terms quadratic in the linear perturbations with coefficients depending on the homogeneous variables. The action $S^{(2)}$ gives the equations for the perturbations, and enables us to quantize the linear perturbations and to find the correct normalization. As argued before, there is only one true scalar degree of freedom. This is reflected in the fact that $S^{(2)}$ can be written (after some cumbersome manipulations...) in terms of a single variable (normalized to have a canonical kinetic term)

$$v = a\left(\delta\varphi + \frac{\varphi'}{\mathcal{H}}\psi\right) \tag{3.40}$$

Note that v is proportional to the comoving curvature perturbation $v = a(\varphi'/\mathcal{H})\mathcal{R} = z\mathcal{R}$,with $z = a\delta\varphi'/\mathcal{H}$. The quadratic action takes the simple form

$$S^{(2)} = \frac{1}{2}\int {\rm d}\eta {\rm d}^3x \left[v'^2 - (\nabla v)^2 + \frac{z''}{z}v^2\right]. \tag{3.41}$$

Except $a \leftrightarrow z$ this is the same action as for the perturbations of a scalar in a fixed FRW background (3.10). But in the slow roll approximation the evolution of φ and H is much slower than that of the scale factor a, and thus $z''/z \approx a''/a$, and all the results of our previous calculation apply [6]. And thus, using (3.18),

$$P_{\mathcal{R}} = \frac{k^3}{2\pi^2}\frac{|v_k|^2}{z^2} = \left(\frac{H^2}{\dot\varphi^2}\right)\left(\frac{H}{2\pi}\right)^2_{k=aH} \tag{3.42}$$

where the subscript $k = aH$ is a reminder that the corresponding quantity is to be evaluated at horizon exit.

Time delay formalism An intuitive way to understand the perturbations is via the time delay or δN formalism. The physical picture for adiabatic perturbations is that in flat gauge, the universe goes everywhere through the same history, but at slightly different times. Consider for example a chaotic inflation model, where φ slowly rolls down the potential during inflation until it drops below some critical value at which point slow roll breaks down and

[6] In the spatially flat gauge v is just the scalar perturbation, which confirms our earlier statement that in this case the gravitational and scalar field perturbations decouple

inflation ends. In regions where the fluctuations take the field up the potential $\delta\varphi > 0$ inflation will end later than in regions in which $\delta\varphi < 0$. During inflation $a \sim \mathrm{e}^{Ht}$ whereas after inflation $a \sim t^n$. Thus the regions in which inflation lasts longer has expanded more than the regions in which inflation ends early, and as a result the energy density $\propto a^{-3}$ is less. Hence, density perturbations are created. Schematically

$$\mathcal{R}|_{\psi=0} = H\frac{\delta\varphi}{\dot{\varphi}} = H\delta t = \frac{\delta a}{a} = \delta N \tag{3.43}$$

3.3 Gravitational waves

From (3.23) the tensor perturbations decompose

$$\delta g_{\mu\nu} = a^2 \begin{pmatrix} 0 & 0 \\ 0 & h_{ij} \end{pmatrix} \tag{3.44}$$

with h_{ij} traceless and divergenceless. The 2 independent d.o.f. h_{ij} are gauge invariant and correspond to physical degrees of freedom, they correspond to the two polarizations of gravitational waves. For a diagonal stress-energy tensor the tensor modes do not have any source in their equation of motion, and their action is that of 2 independent massless fields.

$$S^{(2)} = \frac{1}{2}\int \mathrm{d}x^4 \sqrt{-g}\frac{1}{2}\partial_\sigma h_{ij}\partial^\sigma h_{ij} \tag{3.45}$$

The tensor structure can be put in a polarization tensor via

$$a h_{ij} = \frac{1}{\sqrt{2}} \sum_{\lambda=+,\times} \int \frac{\mathrm{d}^3 k}{(2\pi)^3} v_{k,\lambda}\epsilon_{ij}(\mathbf{k};\lambda)\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}. \tag{3.46}$$

Then $v_{k,\lambda}$ satisfy the same equation as the massless scalar perturbation (3.7). The factor $1/\sqrt{2}$ in the definition is needed to make $v_{k,\lambda}$ canonically normalized; there is a factor $1/2$ difference in $S^{(2)}$ w.r.t. the perturbed action for scalar perturbations (3.10). The power spectrum is

$$\mathcal{P}_T = 2 \times 4 \times \left(\frac{H}{2\pi}\right)^2 \tag{3.47}$$

where the factor 2 comes from the two polarizations, and the 4 is a renormalization factor that can be traced back to the factor $1/\sqrt{2}$ mentioned above.

3.4 Problems

P3.1 Gauge transformations Show that (3.29) follows from (3.26).

P3.2 Curvature perturbation on superhorizon scales The constancy of ζ on superhorizon scales follows from energy momentum conservation

$$0 = \nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\kappa} T^\kappa_\nu - \Gamma^\kappa_{\nu\mu} T^\mu_\kappa \tag{3.48}$$

To calculate the covariant derivative we first need to calculate the perturbed affine connections. We also need the perturbed energy momentum tensor.

a) The energy momentum tensor for a perfect fluid is of the form

$$T^\mu_\nu = (\rho + p) u^\mu u_\nu + p \delta^\mu_\nu \tag{3.49}$$

with $u^\mu = \mathrm{d}x^\mu / \mathrm{d}\tau$, with τ the proper time, the fluid 4-velocity normalized to $u^\mu u_\mu = -1$. Because of isotropy, the 0th order fluid 3-velocity is zero and $u^\mu = a^{-1}(1, 0)$. The 3-velocity only enters at 1st order and can be defined via $\delta u^i = v^i$. As before the 3-vector can be decomposed into a scalar and vector part $v^i = v_,{}^i + \bar{v}^i$, of which we are only interested in the scalar. What are the components of the perturbed energy momentum tensor?

b) The 1st order metric in conformal time is

$$g_{\mu\nu} = a^2 \begin{pmatrix} -(1+2\phi) & B_{,i} \\ B_{,i} & (1-2\psi)\delta_{ij} + 2E_{,ij} \end{pmatrix}, \tag{3.50}$$

Find its inverse to 1st order in the perturbations.

c) The affine connections are:

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2} g^{\alpha\rho} \left(\frac{\partial g_{\rho\gamma}}{\partial x^\beta} + \frac{\partial g_{\beta\rho}}{\partial x^\gamma} - \frac{\partial g_{\beta\gamma}}{\partial x^\rho} \right) \tag{3.51}$$

Which connections enter the equation $0 = \nabla_\mu T^\mu_\nu$ when evaluated to 1st order in the perturbations? Calculate the perturbed connections needed, and show that energy conservations gives (3.37).

4. CMB data

Superhorizon modes are frozen in. Once the perturbations re-enter the Hubble horizon, which sets the scale for regions in which causal physics can happen, they start to evolve again. Matter tends to collapse due to gravity onto regions where the density is higher than average, and baryons falls into overdense regions. Before recombination, the baryons and photons are still strongly coupled, the photon pressure tends to resist collapse and push the baryon-photon plasma outward. The result are oscillatory modes of compression and rarefraction in

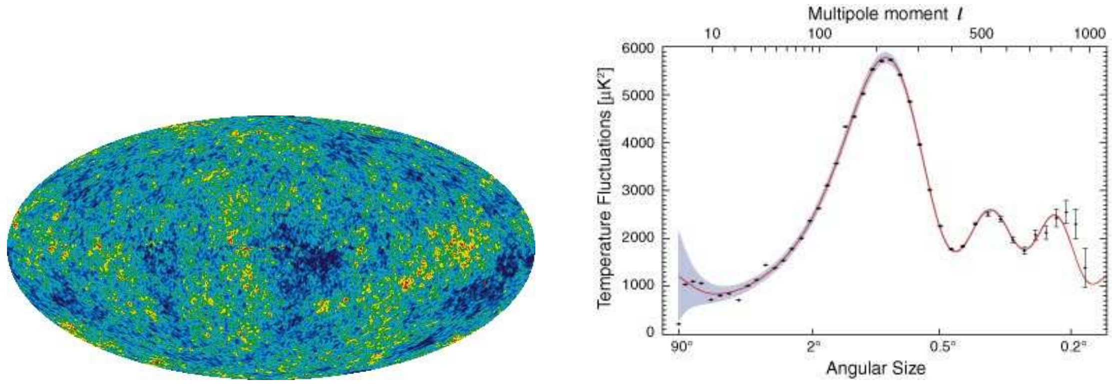

Figure 6: WMAP data.

the baryon-photon fluid, which are called acoustic oscillations. “Acoustic” because the waves move with the sound speed. The plasma heats as it compresses and cools expands, giving rise to the CMB temperature fluctuations observed.

At last scattering the photons decouple, and fly unimpeded towards us, where we observe them as microwave radiation. The CMB provides a snapshot of the universe at the time of last scattering, a baby photo of the universe. The observed oscillatory behavior in the power spectrum plotted as function of (angular) scale is due to the acoustic oscillations. The first peak corresponds to the mode that at recombination just had time to do a partial oscillation and compress, and is overdense. The second peak corresponds to the mode that just had time to compress and decompress again, and is underdense. Etc. Note that the figure shows the power spectrum, which is $\propto$ perturbation squared, and thus both over- and underdensities show up as peaks. Note that different scales enter the horizon at different times, and are thus at different stages in their oscillatory cycle.

The CMB power spectrum is created by complicated but well understood physics, depending not only on the perturbation spectrum, but also on the matter composition of the universe. It allows us to extract information on the primordial spectrum of density fluctuations. The main sources of error are parameter degeneracies, cosmic noise (on small scales), and cosmic variance (on large scales).

On the largest scales, matter has not yet collapsed, and the primordial fluctuations are directly related to the measured temperature fluctuations via $\delta T/T = -(1/5)\mathcal{R}$. Overdense regions are cold, and underdense regions are hot. The fluctuations are a result of two competing effects: in overdense regions the temperature is higher ($\rho_\gamma \propto T^4$ by Stefan-Boltzmann’s law), but for the photons to reach us they have to climb out of a gravitational potential well and are red shifted. The latter is known as the Sachs-Wolfe effect, and it dominates the temperature fluctuations on large scales.

The CMB temperature anisotropies are commonly expanded in spherical harmonics (which form a complete set on the 2D spherical surface of last scattering).

$$\frac{\delta T(\mathbf{n})}{T} = \sum_{lm} a_{lm} Y_{lm}(\mathbf{n}) \tag{4.1}$$

with l the multipoles number and $-l < m < l$ integer. If the fluctuations are due to a Gaussian random process with no distinct direction in the sky, the complete information about the fluctuations is given by the ensemble average

$$\langle a_{lm} a^*_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l \tag{4.2}$$

The C_l's (the power spectrum) are independent of m and our position as a consequence of homogeneity and isotropy. The 2-point correlator is then

$$\langle \frac{\delta T(\mathbf{n})}{T} \frac{\delta T(\mathbf{n}')}{T} \rangle = \frac{1}{4\pi} \sum_l (2l+1) C_l P_l(\mathbf{n} \cdot \mathbf{n}') \tag{4.3}$$

with P_l the Legendre polynomial (which enters via the orthogonality relation for the spherical harmonics: $\sum_m YY \propto P_l$).

The polarization of light can also be measured. It can be uniquely decomposed in an E-mode and B-mode polarization. The E-mode is produced by Thompson scattering at last scattering. This signal is (anti) correlated with the scalar perturbation spectrum. WMAP has measured the $T-E$ cross correlation. This helps reduce the degeneracy between parameters. Gravitational waves give rise to B-mode polarization. However, this signal is expected to be small. Moreover, it will be hard to measure because of background noise (e.g. cosmic lensing transfers E-modes into B-modes). Detecting a gravitational wave background is one of the main goals of the new generation of CMB experiments.

Gaussian distribution and cosmic variance The (Gaussian) statistical nature of the inflaton perturbations are inherited by other perturbations which linearly depend on them. To a very good degree the curvature perturbation and the temperature fluctuations are Gaussian for single field inflation models. Let's discuss a generic Gaussian perturbation

$$g(\mathbf{x}) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} g_{\mathbf{k}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \tag{4.4}$$

where the fourier coefficients $g_{\mathbf{k}}$ are the result of a Gaussian random process, ie modes with different wavevector k are uncorrelated and the phases are random. For a real perturbation $g_{-\mathbf{k}} = g^*_{\mathbf{k}}$. The random process is encoded in the probability distribution $P(g_k)$, defined via

$$\langle f(g_{\mathbf{k}}) \rangle = \int f(g_{\mathbf{k}}) P(g_{\mathbf{k}}) \mathrm{d} g_{\mathbf{k}}. \tag{4.5}$$

For a gaussian distribution the probability distribution is a gaussian:

$$P(g_{\mathbf{k}}) = \frac{1}{2\pi\sigma^2_{\mathbf{k}}} \exp\left(-\frac{|g_{\mathbf{k}}|^2}{2\sigma^2_{\mathbf{k}}} \right) \tag{4.6}$$

with mean $\langle g_k \rangle = 0$ and variance $\langle |g_k|^2 \rangle = 2\sigma_k^2 \delta^3(k-k')$. The probabilities of different Fourier modes are independent. For an isotropic process $\sigma_{\mathbf{k}} = \sigma_k$ independent of the direction of the Fourier mode. The distribution has one free parameter σ_k which gives the width. In terms of real space perturbations $\langle g(x) \rangle = 0$ and $\langle g(x)^2 \rangle = \int 2\sigma_k^2 \mathrm{d}^3k/(2\pi)^3$.

The ensemble average in (4.3) can be regarded as an average over the possible observer positions. On large scales there will always be a big systematic uncertainty due to the fact that we have only one sky to observe (and we cannot average over a whole ensemble of universes). The usual hypothesis is that we observe a typical realization of the ensemble. This means we expect the difference between measured $|a_{lm}|^2$ and the theoretical ensemble average C_l to be of the order of the mean square deviation 2σ. Since we are dealing with a Gaussian distribution $\sigma = C_l$ for each multipole l. For a single l averaging over the $(2l+1)$ values of m, which give independent measurements, reduces $\sigma \to \sigma/(2l+1)$. Cosmic variance, the difference between the observed correlator and the theoretical one, is a serious problem especially for small multipoles. The statistical error due to cosmic variance is indicated by the grey band in the CMB plot of fig. (6).

Position of the first peak The first peak corresponds to the acoustic oscillations that just had time to compress once before last scattering. The scale of this oscillation is the horizon at last scattering. Since the plasma waves propagate with the sound velocity $c_s \approx 1/\sqrt{3}$, it is the sound horizon we are interested in.

The comoving distance to the last scattering surface is $\int_{t_{\rm ls}}^{t_0} a^{-1}\mathrm{d}t = \eta_0 - \eta_{\rm ls}$. A given comoving scale λ is projected on the last scattering-surface sky on an angular scale $\theta \approx \lambda/(\eta_0 - \eta_{\rm ls})$. Consider that $\lambda \sim c_s \eta_{\rm ls}$ is the comoving sound horizon at last scattering. Using $\eta_0 \gg \eta_{\rm ls}$ it follows

$$\theta_{\rm hor} \approx c_s \eta_{\rm ls}/\eta_0 = c_s \left(\frac{T_0}{T_{\rm ls}}\right)^{1/2} \sim 1\,\mathrm{deg} \tag{4.7}$$

where in the 2nd step we used that during matter domination $a \propto T^{-1} \propto t^{2/3} \propto \tau^2$, and the last step we inserted $T_{\rm ls} \sim 0.3\mathrm{eV}$ and $T_0 \sim 10^{-13}\mathrm{GeV}$. We thus expect the first peak at

$$l_{\rm hor} \approx \frac{\pi}{\theta_{\rm hor}} \sim 200 \tag{4.8}$$

4.1 Inflationary prediction for the CMB

Power spectra The scalar power spectrum (3.42) can be rewritten as

$$P_{\mathcal{R}} = \left(\frac{V}{24\pi^2\epsilon}\right)_{k=aH} \tag{4.9}$$

with the subscript denoting that the expression is to be evaluated at horizon exit. Observable scales leave the horizon $N_* \sim 60$ e-folds before the end of inflation, with N_* given by (2.21) (thus V, ϵ are to be evaluated at $\varphi = \varphi(t_*)$).

In a pure dS and for a massless scalar the power spectrum is scale independent. Scale dependence enters because $\dot{H} \propto \epsilon$ is non-zero during inflation, and because the inflaton mass is non-zero $\eta \neq 0$. Different scales leave the horizon at different times (the large scales we observe today leave the horizon before the smallest scales). The Hubble rate (and thus the horizon size) changes with time, and also ϕ as the non-zero mass leads to a classical (slow roll) evolution of the inflaton field. The scale dependence is parametrized by the spectral index via $P_R \propto k^{n_s-1}$, or

$$n_s - 1 \equiv \frac{\mathrm{d}\ln P_R}{\mathrm{d}\ln k} \approx 2\eta - 6\epsilon \tag{4.10}$$

To derive the above formula we used that at horizon exit $\mathrm{d}\ln k = \mathrm{d}\ln(aH) \approx \mathrm{d}\ln a = -\mathrm{d}N = (H/\dot{\varphi})\mathrm{d}\varphi = -(V/V_\varphi)\mathrm{d}\varphi$. Then

$$n_s - 1 \approx -\frac{V_\varphi}{V}\frac{\mathrm{d}\ln P_R}{\mathrm{d}\varphi} = -\frac{V_\varphi}{V}\left(3\frac{V_\varphi}{V} - 2\frac{V_{\varphi\varphi}}{V}\right) = 2\eta - 6\epsilon \tag{4.11}$$

Deviation from scale invariance is small since $\epsilon, \eta \ll 1$ during inflation, but non-zero. The same result is obtained by calculating the power spectrum, but now taking the inflaton mass, and deviation from dS into account. More specifically, this amounts to keeping the $V_{\varphi\varphi}$ term in the equation of motion (which is proportional to η), and taking for the scale factor during inflation $a(\eta) = -1/(H\eta)1/(1-\epsilon)$ with ϵ paramterizing deviations from a perfect dS expansion. The result of the calculation is a power spectrum with specral index as given in (4.10).

There may be deviations from the parametrization of the spectrum by a power law. This is parametrized by the running of the spectral index

$$\frac{\mathrm{d}n_s}{\mathrm{d}\ln k} = -16\epsilon\eta + 24\epsilon^2 + 2\xi^2 \tag{4.12}$$

with $\xi^2 = V_\varphi V_{\varphi\varphi\varphi}/V^2$. The running is small, 2nd order in slow roll parameters.

The tensor-to-scalar ration, from (3.42, 3.47), is

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_R} = 16\epsilon \tag{4.13}$$

The tensor spectral index is $n_T = -2\epsilon$.

The power spectrum was first measured by the COBE satellite. WMAP gives for the amplitude $\mathcal{P}_{\mathcal{R}} \approx (5 \times 10^{-5})^2$. Combining with (4.9) gives $(V/\epsilon)^{1/4} \approx 0.027 = 7 \times 10^{16}\mathrm{GeV}$. WMAP has first measured a deviation from scale invariance $n_s = 0.963^{+0.014}_{-0.015}$. No tensor perturbations have been measured which gives an upper bound $r < 0.2$ at the 95% confidence level (analysis including baryon acoustic oscillations and supernova data). Combining this bound with (4.9) gives an upper bound on the scale during inflation $V^{1/4} < 9.3 \times 10^{-3} = 2 \times 10^{16}\mathrm{GeV}$ or $H < 5.0 \times 10^{-5} = 1.2 \times 10^{14}\mathrm{GeV}$.

The upcoming Planck satellite will be sensitive to $r \sim 10^{-2} - 10^{-3}$. The observables $r = -8n_T$ are not independent, and this consistency relation for single field slow roll inflation can be tested by experiments.

Adiabatic vs. isotropy perturbations Adiabatic perturbations are perturbations along the same trajectory in phase space as the background solution. Perturbations in any scalar X can be described by a unique perturbation in the expansion via

$$H\delta t = H\frac{\delta X}{\dot{X}} \qquad \forall X \tag{4.14}$$

In particular $\delta\rho/\dot{\rho} = \delta p/\dot{p}$ which implies that pressure is a unique function of energy density $p = p(\rho) = c_s^2\rho$. We used this before to proof that the curvature perturbation $\mathcal{R}$ is constant on superhorizon scales for adiabatic perturbations. Another way to characterize adiabatic perturbations is to note that it implies that all fluids in the universe (radiation, DM, neutrinos, baryons) have the same density perturbation, ie $\delta\rho_i/\dot{\rho}_i = \delta\rho_j/\dot{\rho}_j$.

In single field inflation there is only one scalar degree of freedom, and the only perturbation is adiabatic. The inflaton dominates the energy density, and thus perturbations in the inflaton give rise to energy-density perturbations and consequently metric perturbations. After inflaton decay all fluids in the universe will inherit the same perturbation. To get isocurvature perturbations requires an extra light field during inflation (an extra scalar field d.o.f.), so that the perturbations can be split in perturbations along the inflaton direction (adiabatic with $\delta\mathcal{R} \neq 0$) and those orthogonal (isocurvature with $\delta\mathcal{R} = 0$). Isocurvature perturbations are characterized by

$$\frac{\delta X}{\dot{X}} \neq \frac{\delta Y}{\dot{Y}} \qquad \text{for some } X, Y \tag{4.15}$$

One simple example is the baryon to photon ratio $\delta(n_b/n_\gamma) = (\delta n_b/n_b) - (\delta n_\gamma/n_\gamma)$. A gauge invariant parametrization is

$$S_{ij} = 3H\left(\frac{\delta\rho_i}{\dot{\rho}_i} - \frac{\delta\rho_j}{\dot{\rho}_j}\right) \tag{4.16}$$

Isocurvature perturbations correspond to a fluctuation in the local pressure, i.e. spatial variations in the equation of state since now $p = p(\rho, S_{ij}(\mathbf{x}))$.

Adiabatic and isocurvature lead to a very different peak structure for the CMB power spectrum. Heuristically, this can be understood as they provide very different initial conditions. For an adiabatic perturbation the initial condition is a maximum amplitude (corresponding to the density contrast) and zero velocity, for an isocurvature perturbation it is zero amplitude but maximum velocity. Hence one is a cosine oscillation, and the other is a sine. The CMB data is consistent with a fully adiabatic spectrum. It puts upper bounds on the level of isocurvature perturbations.

Non-gaussianity Studying the perturbation equations at linear order, the action at 2nd order in the fluctuations is that of a free field. The corresponding perturbation is Gaussian: it's Fourier components are uncorrelated and have random phases. Deviations from pure Gaussian statistics arise from cubic and higher order terms in the action, and give rise to primordial non-Gaussianity. Non-linearity's in the evolution after horizon exit give a secondary contribution to non-Gaussianity; since gravity is a non-linear theory this contribution always enters at some level.

A gaussian temperature distribution give rise to as many cold as hot spots. Measuring the temperature at N points, the distribution is a Gaussian centered around δT=0, and with width parameterized by $\langle \delta T^2 \rangle$.

For a Gaussian spectrum all information is contained in the 2-pnt function. Higher order correlators

$$\langle \mathcal{R}(\mathbf{n}_1)\mathcal{R}(\mathbf{n}_2)...\mathcal{R}(\mathbf{n}_n)\rangle \tag{4.17}$$

are zero for n odd, and can be expressed in terms of 2-pnt functions for n even. The 3-pnt function is the lowest order statistics to distinguish a Gaussian from a non-Gaussian spectrum of fluctuations (going to Fourier space it is called the bispectrum)

$$\langle \mathcal{R}(t)^3 \rangle = \langle U^{-1}(t_0,t)\mathcal{R}(t)^3 U(t_0,t)\rangle = -i\int_{t_0}^{t} \mathrm{d}t' \langle [\mathcal{R}^3(t'), H_{\mathrm{int}}]\rangle \tag{4.18}$$

with $U = T\exp(-i\int_{t_0}^t H_{\mathrm{int}}(t')\mathrm{d}t')$ the time evolution operator inserted to take into account that the vacuum is not that of the free theory, and $H_{\mathrm{int}} = -\mathcal{L}_{\mathrm{int}}$ the Hamiltonian in the interaction picture. The interactions are switched on at some early time t_0. In the last step we expanded the exponent to get the 1st order result. Note we are computing an expectation value, and not a transition amplitude for which the expression is $\langle T\mathcal{R}^3 \exp(\int_{-\infty}^{\infty} H_{\mathrm{int}}(t')\mathrm{d}t')\rangle$.

The quantity $\mathcal{R}$ is not Gaussian but contains a non-linear correction usually parametrized as

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f_{\mathrm{NL}}\left(\mathcal{R}_g(\mathbf{x})^2 - \langle \mathcal{R}_g(\mathbf{x})^2\rangle\right) \tag{4.19}$$

with $\mathcal{R}_g(\mathbf{x})$ Gaussian. The factor (3/5) is convention, introduced so that f_{NL} parametrizes non-linearities of a matter-era gravitational potential on large scales ($\Phi = 3/5\mathcal{R}$ in the matter dominated era). In principle f_{NL} is a momentum-dependent function. For f_{NL} constant, the 5yr WMAP data gives

$$-9 < f_{\mathrm{NL}} < 111 \tag{4.20}$$

The upcoming Planck satellite will be sensitive to $f_{\mathrm{NL}} \sim 5$. Non-linear GR effects give a lower bound $f_{\mathrm{NL}} \sim 1$. In slow roll single field inflation the level of non-Gaussianity $f_{\mathrm{NL}} \sim \mathcal{O}(\epsilon)$ is unmeasurably small. The slow roll conditions force the inflaton potential to be flat, with interaction terms suppressed.

Using rotational and translational invariance the bispectrum depends only on k_i, with i = 1,2,3, and is independent of directions. Furthermore if follows from the (nearly) scale invariance of the perturbation spectrum that it is invariant under rescalings of the triangle (note that the bisprectum is $\propto \delta^3(\sum k_i)$, i.e the k_i-vectors form a triangle). Hence one of the k_i can be scaled to unity, and the bi-spectrum depends on the two others. If detected, non-Gaussianity can give a lot of information on the underlying inflation model, as it gives both size and shape information.

The parametrization (4.19) has only one d.o.f., which describes at leading order the most generic form of non-Gaussianity which is local in real space, i.e. NG depends only on the local value of the potential. It is therefore expected for models where non-linearity's develop outside the horizon. This can happen in multifield models of inflation, where isocurvature perturbation source the adiabatic curvature perturbation on superhorizon scale (3.38) (e.g. in the curvaton model). Local NG is peaked in the squeezed triangle limit $k_3 \ll k_1, k_2$.

If $\mathcal{R}$ is constant on superhorizon scales, non-linearity's only have limited time to grow before horizon exit. In this case maximal application is achieved when all wave vectors have magnitude comparable to the Hubble radius, leading to NG that is peaked in the equitriangle limit. This can occur e.g. in single field models with non-canonical kinetic terms such as DBI inflation.

4.2 Problems

P4.1 Sachse-Wolfe effect The observed CMB temperature is

$$\left(\frac{\Delta T}{T}\right)_{\rm obs} = \left(\frac{\Delta T}{T}\right)_{\rm emit} - \Phi_{\rm emit} \tag{4.21}$$

with Φ the gravitational potential.

$$\mathrm{d}s = -\sqrt{1-2\Phi}\mathrm{d}t + ... \approx (1-\Phi)\mathrm{d}t + ... \tag{4.22}$$

The gravitational potential term on the rhs of (4.21) represents the red shift as the photons climb out of the potential well towards us. The 1st term is the intrinsic temperature fluctuation at the time of emission. For adiabatic fluctuations the number of photons inside the well is expected to be larger than average, and one expects $(\Delta T/T)_{\rm emit} \propto \Phi_{\rm emit}$. Note that the two terms on the rhs of (4.21) are not gauge invariant by themselves, but their sum is; when evaluated in the same gauge we can add the two contributions. Here we are in the comoving gauge $B = E = 0$.

a) Calculate the constant of proportionality using that the universe is matter dominated at the time of last scattering. Further use that (aT) =const.

b) In (4.21) we took $\dot{\Phi} = 0$, which is appropiate for a matter dominated universe. The correct generalization is

$$\left(\frac{\Delta T}{T}\right)_{\rm obs} = \left(\frac{\Delta T}{T}\right)_{\rm emit} - \Phi_{\rm emit} + 2\int_{t_{\rm dec}}^{t_0} \dot{\Phi}\mathrm{d}t \tag{4.23}$$

The time integral is from the time of last scattering to the present along a null geodesic. At present the universe is dominated by a cosmological constant. What is its effect on the measured temperature anisotropies on the largest scales?

P4.2 Curvaton scenario In the curvaton scenario not the inflaton field but some other field that is light during inflation (called the "curvaton" field σ) gives the dominant contribution to the observed density perturbations. The energy density in the curvaton field is negligible during inflation, and the perturbations in the curvaton field give $\mathcal{R} = 0$; they are isocurvature perturbations. To be in agreement with observations these isocurvature perturbations have to be converted to adiabatic perturbations after inflation.

a) Consider a potential $V = V_I(\varphi) + m_\sigma^2 \sigma^2$, ie is an inflationary potential plus a light curvaton field with $m_\sigma^2 \ll H_I^2$. Write down the KG equation for the curvaton field in a FRW universe. How does the curvaton evolve in the limit $m_\sigma^2 \ll H^2$ (at early times), and in the opposite limit $m_\sigma^2 \ll H^2$ (late times)?

b) In the limit $m_\sigma^2 \ll H^2$ show that the pressure is $\bar{p} = 0$ when averaged over one oscillations. It follows that the equation of state for the curvaton is $w_\sigma = 0$, and it behaves as cold matter with $\rho_\sigma \propto a^{-3}$. Assume that the inflaton promptly decays into radiation at the end of inflation which red shifts $\rho_\gamma \propto a^{-4}$. The ratio ρ_σ/ρ_γ grows with time until the curvaton comes to dominate the energy density. Use the δt formalism to argue this leads to the formation of adiabatic density perturbations.

c) Go to flat gauge $\psi = 0$. Define $\zeta_i = \psi + H\dot{\rho}_i/rho$ for $i = \sigma, \gamma$. Show that

$$\sigma = \zeta_\gamma(1 - f) + \zeta_\sigma f, \qquad f = \frac{3\rho_\sigma}{4\rho_\gamma + 3\rho_\sigma} \tag{4.24}$$

d) Calculate the perturbation spectrum in the limit of constant decay, and with the curvaton dominating the energy density at the time of decay $f = 1$. Use that

$$\left.\frac{\delta\rho}{\rho}\right|_{\rm dec} \approx 2\frac{\delta\sigma}{\sigma}_{\rm dec} \equiv 2q\frac{\delta\sigma}{\sigma}_I \tag{4.25}$$

What is q for a quadratic potential?

e) If the curvaton energy density does not dominate the energy density at the time of decay, it can be shown that

$$\zeta \approx r\zeta|_{f=1}, \qquad r = \frac{\rho_\sigma}{\rho} \tag{4.26}$$

The relation between $\delta\rho$ and $\delta\sigma$ used (4.25) is non-linear, which gives rise to non-gaussianities in the spectrum. Include the second order term in the expansion, and calculate f_{NL}. Can you understand why non-gaussianity is larger for smaller r? What is the bound on r from CMB observations?

5. Particle physics models of inflation

Inflation was first proposed in January, 1980 by Alan Guth as a mechanism for resolving the big bang problems. His model was based on a 1st order phase transition. Inflation is driven by the false vacuum energy, and ends via tunneling to the true vacuum. However, for inflation to last long enough to solve the horizon and flatness problem, the tunneling rate has to be sufficiently slow. The bubbles of true vacuum are to sparsely produced and will never coalesce to reheat the universe. The model is phenomenologically not viable. It has a "graceful exit" problem.

In 1982 the second generation of models appeared, called "new inflation", by Linde and Albrecht & Steinhardt. These models were based on a 2nd order phase transition. The fields are initially in thermal equilibrium, although it is questionably whether at such high temperatures the thermal interactions can confine the field in the potential minimum, as assumed.

Around the same time it was shown that inflation produces tiny density fluctuations. The perturbations were first calculated by Mukhanov and Chibisov. Independently of that work it was calculated at the three-week 1982 Nuffield Workshop on the Very Early Universe at Cambridge University by four groups working separately over the course of the workshop: Hawking; Starobinsky; Guth and Pi; and Bardeen, Steinhardt and Turner.

Newer models of inflation, starting with chaotic inflation invented by Linde in 1983, gave up on the assumption of an initial thermal state.

We will discuss chaotic inflation, new/hilltop inflation, and hybrid inflation in turn.

5.1 Models

Chaotic inflation There is no reason the universe should be in a thermal state before inflation, it might as well be cold, with "chaotic" field values as initial conditions. This is the the starting point for chaotic inflation. The idea is that pieces of classical spacetime emerge continually from the "foam". In a small (but non-zero) fraction of them conditions are such – gradient and kinetic energy subdominant — that inflation commences.

Consider as in section 2.2 the example of a quadratic potential. The slow roll parameters are $\epsilon = \eta = 2/\varphi^2 = 1/(2N)$. Normalizing the power spectrum (4.9) to the WMAP data gives $m^2N^2/(6\pi^2) = (5 \times 10^{-5})^2$, and thus $m \approx 10^{13}$GeV. The prediction for the spectral index (4.10) is $n_s = 1 - 2/N \approx 0.97$ for $N = 60$. Finally the tensor-to-scalar ratio (4.13) is $r = 8/N \approx 0.13$, which is close to the current WMAP bound.

Chaotic inflation is an example of large field inflation, with the displacement of the inflaton field during inflation large on Planckian scales $\delta\varphi \gtrsim 1$. Lyth showed that only large field models can give a large tensor perturbation, measurable by experiment. The proof of the "Lyth bound" is as follows. Use the slow roll approximation to write

$$\epsilon = \frac{1}{2}(V_\varphi/V)^2 = \frac{1}{2}(3H\dot{\varphi}/V)^2 = \frac{1}{2}(\dot{\varphi}/H)^2 \tag{5.1}$$

Now $\dot{\varphi}/H = \mathrm{d}\varphi/(H\mathrm{d}t) = \mathrm{d}\varphi/\mathrm{d}N$. For single field slow roll inflation $r = 16\epsilon$, and thus we can rewrite the above relation as

$$r = 8(\mathrm{d}\varphi/\mathrm{d}N)^2. \tag{5.2}$$

Integrating gives

$$\Delta\varphi = \frac{1}{\sqrt{8}}\int_0^{N_*} \mathrm{d}N\sqrt{r} \sim \sqrt{\frac{r}{8}}N_* \tag{5.3}$$

where in the 2nd step we took r constant during the last N_* efolds of inflation to get a rough estimate. Now $N_* \sim 60$, and using that $r \gtrsim 10^{-3}$ will be measurable, this implies $\Delta\varphi \gtrsim 1$. The inflaton field range during the last N_* efolds of inflation needs to be of the Planck scale. Gravity couples to energy density, hence the higher the scale of inflation (large $H^2 = V/3$) the larger the amplitude of gravitational waves. Lyth's result is that a large inflaton scale implies large field ranges.

Examples of chaotic inflation are polynomial potentials (such as quadratic chaotic inflation discussed in the example above), natural inflation with a pseudo Nambu-Goto boson as the inflaton and a cosine-potential (need Planckian period). Monodromy inflation is a rare example of large field in string theory, with the inflaton one of the many string axions.

Advantages: no initial value problem, very simple potential, predictive. Disadvantages: against effective field theory approach, tuned parameters.

New/Hilltop inflation New/Hilltop models are small field models of inflation $\Delta\varphi \ll 1$, with inflation taking place near a maximum or saddle of the potential. Consider the potential

$$V = V_0 - \frac{1}{2}m^2\varphi^2 - \frac{1}{4}\lambda\varphi^4 + ... = V_0\left(1 - \frac{1}{2}|\eta_0|\varphi^2 - \frac{1}{4}C\varphi^4 + ...\right) \tag{5.4}$$

Here $\eta_0 = \eta|_{\varphi=0} = m^2/V_0 = m^2/(3H^2)$, and $C = \lambda/V_0$; the quartic term needs to be negative for the model to work. Higher order terms are neglected, but they are essential to stabilize the inflaton with some finite vev after inflation. Inflation takes place for the field close to the maximum.

For inflation to work $\eta = -|\eta_0| - 3C\varphi^2 \ll 1$; one needs $|\eta_0| \ll 1$, but C can be larger than unity. Although it seems we can make the term proportional to C arbitrarily small by going close enough to the maximum $\varphi \to 0$, this is not quite correct, as there is a lower bound on the variation of φ during (observable) inflation. Indeed, from (2.21) $\Delta\varphi \sim \sqrt{2\epsilon}\Delta N$, and using the WMAP normalization $(V/\epsilon)^{1/4} \approx 0.027$ this gives $\Delta\varphi \approx \sqrt{2V_0}\Delta N/(0.027)^2$. Using this the constraint $\eta < 1$ translates to $\lambda < 10^{-9}$, where we took conservatively ΔN corresponding to the 10-efolds that observable scales leave the horizon.

Let's calculate the inflationary predictions for this model. Integrating (2.21) gives

$$N(\varphi) = \left[\frac{1}{2|\eta_0|}\log\left(\frac{|\eta_0| + C\varphi^2}{\varphi^2}\right)\right]_{\varphi_{\mathrm{end}}}^{\varphi} \tag{5.5}$$

Inflation ends when $\epsilon, \eta \approx 1$, for $C \gtrsim 1$ this gives $\varphi \sim 1/C^{1/3}$ (although higher order terms in the potential may come important before). The results are rather insensitive to the exact field value at the end of inflation. Now invert the above equation to get

$$\varphi(N) \approx \sqrt{\frac{|\eta_0|}{C(\mathrm{e}^{2|\eta_0|N} - 1)}} \tag{5.6}$$

where we used the approximation that the quadratic term, although being dominant when observable scales leave the horizon, is negligible small towards the end of inflation $|\eta_0| \ll C\varphi_{\rm end}^2$. Plugging the above in the expression for the power spectrum (4.9), and using the limit $|\eta_0| \ll 1$, gives

$$P_\zeta \approx \frac{2V_0 C N_*}{3\pi^2} \tag{5.7}$$

Fitting to the WMAP data with $N_* = 50$ fixes the scale of inflation $V_0 = 3 \times 10^{-13}/C$. Note that even though $\lambda \ll 1$ it is indeed consistent to take $C = V_0/\lambda \gtrsim 1$ during inflation. Inflation is low scale, and tensor perturbations are immeasurable small. Finally, the spectral index is

$$n_s \approx 1 + 2\eta \approx 1 - 2|\eta_0| - 2\frac{3|\eta_0|}{\mathrm{e}^{2\eta_0 N_*} - 1} \lesssim 0.945 \tag{5.8}$$

in agreement with recent WMAP data. (In the limit $|\eta_0| \to 0$ the spectral index is $n_s = (1 - 3/N_*) = 0.94$).

The moduli potential in the string landscape is a complicated multi-field potential, which may have saddle points suitable for inflation. An explicit realization of this idea is racetrack inflation. In this model the volume modulus has a non-perturbatively generated superpotential of the form $W = W_0 + \sum A_i \mathrm{e}^{-a_i T}$. This potential stabilizes the modulus at one of the minima, but it may in addition give rise to inflation (at least two non-perturbative terms are needed). The scalar potential is a bunch of cosines in $\mathrm{Im}(T)$, with maxima/minima/saddles. The η-parameter at a generic saddle is of order unity, and must be tuned to get inflation. In this model $C \gg 1$, and our analytical approximation applies.

Advantages: easy to implement in particle physics models, less sensitive to UV physics, simple and predictive. Disadvantages: initial conditions for low scale inflation, tuning.

Hybrid inflation Hybrid inflation is a multi-field model of inflation. During inflation the inflaton field rolls slowly down its potential, while all other fields are heavy and frozen, and inflation is effectively single field. Inflation ends as a consequence of the extra fields, when one of the so-called waterfall fields become tachyonic, setting off a phase transition. Hybrid inflation is somewhere in between small and large field models.

Hybrid inflation is naturally embedded into SUSY. Consider for example F-term hybrid inflation, which has a superpotential of the form

$$W = \lambda S(\chi\bar{\chi} - v^2) \tag{5.9}$$

with S a singlet field, and $\chi, \bar{\chi}$ the waterfall fields which are oppositely charged fields under some $U(1)$. Without loss of generality we can take λ, v real. The scalar potential is

$$V = \sum_i |W_i|^2 + V_D = \lambda^2|\chi\bar{\chi} - v^2|^2 + \lambda^2|S|^2(|\chi|^2 + |\bar{\chi}|^2) + V_D \tag{5.10}$$

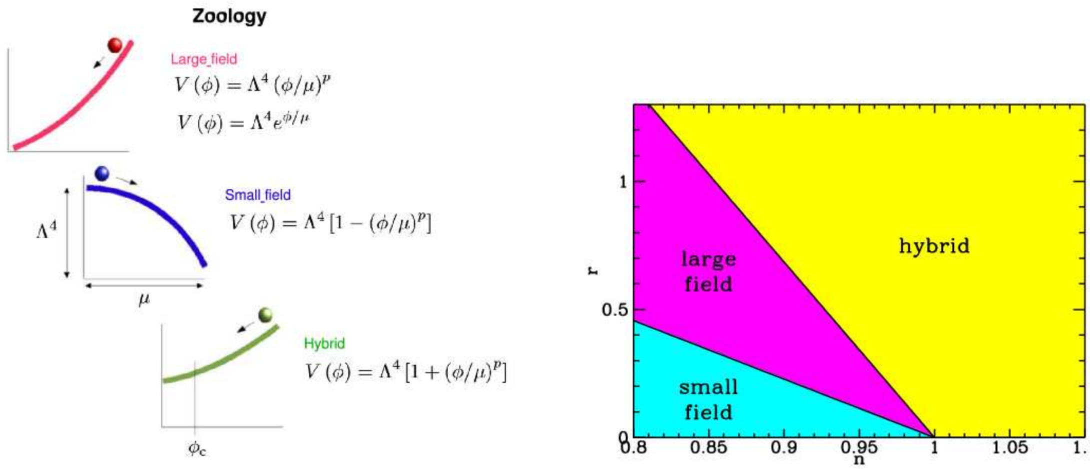

Figure 7: Inflationary models.

Vanishing of the D-term potential enforces $|\chi| = |\bar{\chi}|$. The real canonically normalized inflaton $s = |S|/\sqrt{2}$, and the two waterfall field mass eigen states $\chi_\pm = (\chi \pm \bar{\chi})/\sqrt{2}$ have masses

$$m_s^2 = \lambda^2(|\chi|^2 + |\bar{\chi}|^2), \qquad m_\pm^2 = \lambda^2(|S|^2 \pm v^2) \tag{5.11}$$

Inflation takes place for large initial values $|S| > v$. From (5.11) it follows that then the waterfall field masses are both positive definite $m_\pm^2 > 0$, and consequently the waterfall fields are minimized at the origin $\chi = \bar{\chi} = 0$. The inflaton is massless at tree level, and thus has an exactly flat potential. As discussed below the flatness is lifted by loop corrections, providing a small slope for the inflaton field which slowly rolls down the potential during inflation. The potential during inflation is $V \approx \lambda^2 v^4$.

Inflation ends as $|S| < v$ drops below the critical value, and the waterfall field χ_- becomes tachyonic. Inflation ends in a Higgs-like phase transition. The minimum after inflation is $|\chi_-| = v$, $\chi_+ = S = 0$, which sets $V = 0$ (zero cosmological constant).

Susy is broken spontaneously during inflation by the non-zero energy density in the universe. Remember $\langle 0|H|0\rangle \quad \Leftrightarrow \quad Q|0\rangle \neq 0$ or $\bar{Q}|0\rangle \neq 0$ by virtue of the susy algebra $\{Q_\alpha, \bar{Q}_\beta\} = 2\sigma^\mu_{\alpha\beta} P^\mu$. Another way to see this is by noting that the bosonic mass eigenstates of the waterfall fields and their fermionic susy partners are split in mass; the fermion masses are

$$\tilde{m}_\pm^2 = \lambda^2 |S|^2 \tag{5.12}$$

Loops involving the waterfall fields no longer cancel because of this mass split, and give a one-loop corrections to the potential. Since the waterfall field masses depend on the inflaton vev this correction provides a potential for the inflaton field. The 1-loop potential is given by the Coleman-Weinberg formula

$$V_{\rm CW} = \frac{1}{64\pi^2} \sum_i {\rm Str} M_i^4 \log \frac{M_i^2}{\Lambda^2} \tag{5.13}$$

where the sum is over all mass eigenstates (NB sum over the helicity states), $\mathrm{Str} f(M) = f(M_{(\mathrm{boson})}) - f(M_{(\mathrm{fermion})})$, and Λ is the cutoff scale. Plugging in the masses gives

$$\begin{aligned} V_{\mathrm{CW}} &= \frac{\lambda^4 v^4}{32\pi^2}\left[2\ln\left(\frac{v^2\lambda^2 z}{\Lambda^2}\right) + (z+1)^2\ln(1+z^{-1}) + (z-1)^2\ln(1-z^{-1})\right] \\ &\xrightarrow{z\gg 1} \frac{\lambda^4 v^4}{8\pi^2}\log\frac{\lambda s}{\sqrt{2}\Lambda} \end{aligned} \tag{5.14}$$

with $z = (|S|/v)^2$. In the second line we took the limit $z \gg 1$ (inflation takes place for field values much larger than the critical value), which can be shown to be a good approximation for large couplings $\lambda^2 \gtrsim 10^{-5}$. Note that $V_{\mathrm{CW}} \ll V_0 \approx \lambda^2 v^4$ during inflation. It does not contribute significantly to the energy density, which sets the scale of inflation $H_I^2 \approx V_0/3$. However, the inflaton slope and curvature are generated solely by the 1-loop potential.

Using the slow roll expressions it follows that N efolds before the end of inflation the inflaton vev is $s \approx \lambda\sqrt{N}/(2\pi)$. The power spectrum is $P_\zeta \approx 16N_* v^4/75$; matching to the WMAP result gives the scale of symmetry breaking $v^2 \approx 6\times 10^{-6}$. The spectral index is $n_s \approx 1 + 2\eta \approx 1 - 1/N_* \approx 0.98$ where we took $N_* = 55$. Tensor perturbations are negligibly small.

The scale of inflation $v \sim 10^{15} - 10^{16}$GeV is of the GUT scale. This has prompted much effort to embed hybrid inflation in GUT models. For example, one can identify the waterfall fields with GUT Higgs fields that brake some GUT symmetry, such as $B-L$. In this set-up the inflaton is still a singlet, and not directly associated with the GUT symmetry. (Another approach would be to use the right-handed neutrino as the inflaton, but this is more involved than the simple superpotential that we discussed). Hybrid inflation also readily emerges in string theory, as the effective field theory (EFT) description of brane inflation. In such a set up the inflaton is played by the interbrane distance. The inflaton potential is the attracting coulomb interaction between the branes, plus possible additional corrections. Inflation ends when the branes get close to each other and one of the strings stretching between them becomes tachyonic, setting in the process of brane annihilation. The set-up is more involved than the simple example above; extra ingredients are non-trivial geometry (warping), moduli stabilization, and susy breaking. This is discussed in more detail in the next section.

In all viable GUT hybrid inflation models, as well in the string inspired models, the symmetry broken in the phase transition ending inflation is a gauged U(1) symmetry. This means cosmic strings form (in the string case, there are both D- and F-strings) at the end of inflation. It can be shown that the network of cosmic strings enters a scaling regime with the energy in strings a constant fraction of total energy density in the universe; this is possible as the network can lose energy via the decay of string loops. Strings are heavy objects, which mean they can deflect CMB photons. This leaves an imprint in CMB, whose magnitude is set by the string tension. The string signal is highly non-Gaussian, very different from the peak structure seen. Because of this the data constrains the string contribution to the power spectrum to less than 10%. This bounds the string tension, which is related to parameters in the model. In the simple example above, this rules out the model for the large couplings discussed.

5.2 Problems

Despite the simplicity of the idea, as a phenomenon in quantum field theory coupled to general relativity, inflation does not appear to be natural. The set of Lagrangians suitable for inflation is a minute subset of all possible Lagrangians. Moreover, in a wide class of models inflation only emerges for rather special initial conditions.

The problems are associated with our ignorance about the UV physics.

Initial conditions Having an inflationary solution is not enough. Why should the inflaton field start in the particular slow roll domain? Inflation only commences if potential energy dominates over kinetic and gradient energies.

For chaotic inflation the issue of initial conditions is resolved easily. The idea is that continually pieces of classical spacetime emerge from the "foam". The natural energy scale of a Planck scale region of spacetime is $V \sim 1$, which implies that for $m^2 \ll 1$ it is natural to have the large $\varphi \gg 1$, as needed for inflation. If in one of the many disconnected regions gradient and kinetic terms are small, inflation starts.

Chaotic inflation is eternal. The evolution of the inflaton field is dominated by quantum fluctuations (random walk) rather than the classical evolution (down the potential). There will always be some regions in which the field moves up the potential by means of quantum diffusion, and inflation continues. In a typical Hubble time $\delta t \sim H^{-1}$ by random walk the quantum displacement is $\delta\varphi_{\mathrm{q}} \sim H$. In the same time interval the classical motion, given by the slow roll equation, is $\delta\varphi_{\mathrm{cl}} \sim \dot{\varphi}\delta t$. Inflation is eternal inflation for $\delta\varphi_{\mathrm{q}} \gtrsim \delta\varphi_{\mathrm{cl}}$, which is satisfied for $\phi \gg 1$ (not during observable inflation, but for energies well below m_{p}^4 where gravity is classical). The total volume of space emerging from inflation by this process is infinitely greater than the total volume of all non-inflationary domains.

The issue of initial conditions is more complicated for low scale inflation, which starts at energies $V \ll 1$. If the universe is closed $\Omega > 1$, then the universe collapses before the onset of inflation unless it is initially homogeneous over many causally disconnected Planck regions. Even for open or flat universes, inflation only begins if the patch of space-time is homogeneous on scales larger than the horizon $\Delta l \gtrsim H^{-1}$. Note in this respect that the horizon is large compared to typical length scales over which particle interactions can homogenize the universe.

One may argue (anthropically) that in models where inflation is eternal, the issue of initial conditions is largely irrelevant. Even if the probability of proper initial conditions for inflation is extremely small, the parts of the universe where they are satisfied enter the regime of eternal inflation, producing an infinite amount of homogeneous space where life of our type is possible. Thus, even if the fraction of patches of universes with inflationary initial conditions is exponentially suppressed, one may argue that eventually most of the observers will live in the parts of the universe produced by eternal topological inflation.

This argument applies to hilltop inflation, which is eternal for the inflaton field close enough to the maximum. This may offset the unlikely initial conditions needed for inflation, with the inflaton sitting initially at the metastable maximum. Another way to circumvent the initial value problem, is to make the maximum part of a symmetry breaking potential. As the universe expands it goes through a phase transition in which topological defects — domain walls, cosmic strings, monopoles — form. The field inside the defect interpolates between the false and true vacuum. In the core there is always a region where the field goes through $\varphi = 0$. This is where initial conditions are suitable for hilltop inflation, and topological inflation begins.

Hybrid inflation is not eternal. Moreover, starting with random initially conditions at the Planck scale, only a very small part of parameter space leads to inflation. In this sense hybrid inflation is not natural.

The problems with low scale inflation can be ameliorated/solved if initial conditions are set by a previous inflationary scale, starting at the Planck scale. This inflationary phase leaves no observable imprints; only the last 60 efolds of inflation are testable.

η-problem For slow roll inflation the potential needs to be flat, with both the slope and curvature (parametrized by the slow roll parameters) small in Planck units. Quantum and gravity correction will lift the flatness of the potential, quite generically leading to large, order one, contributions to the η-parameter. This is the η-problem.

The slow roll condition, and thus inflation, is sensitive to Planck-scale physics. New physics (quantum gravity) at the Planck scale is required in order to render graviton-graviton scattering sensible, just as unitarity of W-W scattering requires new physics at the TeV scale. In effective field theories (EFTs) the effects of high-scale physics above some cutoff Λ are described by non-renormalizable (NR) operators which can be thought of as originating from integrating out all particles with mass $m > \Lambda$. Low energy physics is sensitive to these operators (eg FCNC, proton life-time), but in all known examples this involves operators with a cutoff well below the Planck scale and of dimension 6 or lower. However, inflation is sensitive to Planck-suppressed operators, and an understanding of these is needed to address the smallness of the η-parameter.

In a generic EFT theory the mass of a scalar field runs to the cutoff scale unless protected by some symmetry. Since the cutoff in an EFT of inflation is at least the Hubble scale, this implies that a small inflaton mass is radiatively unstable. The difficulty here is analogous to the Higgs hierarchy problem. In the absence of any specific symmetries protecting the inflaton potential, contributions to the Lagrangian of the general form

$$\frac{\mathcal{O}_6}{m_{\rm p}^2} = \frac{\mathcal{O}_4}{m_{\rm p}^2}\varphi^2 \tag{5.15}$$

are allowed. If $\langle \mathcal{O}_4 \rangle \sim R\varphi^2$ with R the Ricci scalar, or $\langle \mathcal{O}_4 \rangle \sim V$ comparable to the inflationary energy density, then this term corrects the inflaton mass by order H and $\eta \sim 1$.

The problem is especially severe for large field models with superplanckian inflaton vev's $\varphi > m_{\rm p}$. Indeed we can write down a whole series of NR operators $\sum \lambda_p \phi^p / m_{\rm p}^{p-4}$. Naturalness suggests $\lambda_p \sim 1$. Then it follows that every term in the series kills inflation. One needs a mechanism (symmetry) to keep the NR terms under control for chaotic inflation to be consistent with the usual EFT lore.

The hybrid inflation model was formulated in terms of global susy. To incorporate gravity we have to extend it to sugra. It can be readily shown from the generic form of the sugra potential that $\eta \sim 1$ (see below). But this can already be seen at the susy level. Following the rules of EFT we have to write down all operators that are not forbidden by symmetries; one of these is

$$\delta K = \int \mathrm{d}^4\theta \frac{1}{m_\mathrm{p}^2} \varphi^\dagger \varphi \chi^\dagger \chi \tag{5.16}$$

(sloppily denoting the superfield and the scalar potential by the same symbol). Such terms are actually needed, as they are counterterms for operators generated by 1-loop diagrams. If φ dominates the energy density, then $\langle \int \mathrm{d}^4\theta \varphi^\dagger \varphi \rangle = |F_\varphi|^2 \approx V$, and

$$\delta \mathcal{L} = \frac{V}{m_\mathrm{p}^2} |\chi|^2 \approx 3H^2 |\chi|^2. \tag{5.17}$$

A soft mass for all scalars χ is generated, including the inflaton, which magnitude is set by the susy breaking scale.

For completenes we also give the sugra argument. The F-term scalar potential in sugra is

$$V_F = \mathrm{e}^{K/m_\mathrm{p}^2} \left[D_I W K^{I\bar{J}} \overline{D_J W} - \frac{3}{m_\mathrm{p}^2} |W|^2 \right] \tag{5.18}$$

where $D_I W = m_\mathrm{p}^{-2} K_I W + W_I$. The kahler K determines the inflaton kinetic term, while the superpotential W determines the interactions. For simplicity just consider the inflaton potential and no other light fields, i.e. $I = J = \varphi$. To find the inflaton mass, expand K around some chosen origin, which we can take $\varphi = 0$ without loss of generality, $K = K_0 + K_{\varphi\bar{\varphi}}|_0 \varphi\bar{\varphi} + \ldots$. Then

$$\mathcal{L} \approx -K_{\varphi\bar{\varphi}} \partial\varphi\partial\bar{\varphi} - V_0 \left(1 + K_{\varphi\bar{\varphi}}|_0 \frac{\varphi\bar{\varphi}}{m_\mathrm{p}^2} + \ldots \right) \equiv -\partial\phi\partial\bar{\phi} - V_0 \left(1 + \frac{\phi\bar{\phi}}{m_\mathrm{p}^2} + \ldots \right) \tag{5.19}$$

where we defined the canonical normalized inflaton field $\phi\bar{\phi} \approx K_{\varphi\bar{\varphi}}|_0 \varphi\bar{\varphi}$, and $V_0 = V_F|_{\varphi=0}$. We retained the leading order correction of the expansion of $\mathrm{e}^{K/m_\mathrm{p}^2}$, which is a "universal" correction in sugra theories. The omitted terms, some of which can be of the same order, arise from expanding the term between square brackets in (5.18), and are model dependent. The result is an operator $\mathcal{O}_6 = V_0 \phi\bar{\phi}$ and $\delta\eta \sim 1$.

How to solve the η-problem? (1) Fine-tuning. The procedure is clear in the sugra discussion: tune the model independent and dependent contributions to the inflaton mass, so that they cancel for a large part, and $\eta \ll 1$. This is done in racetrack models, for which a fine tuning of order 1 promile is needed, as η naturally of order $1 - 10$. (2) Invoke symmetries to protect the inflaton mass against corrections, as in natural inflation. Note that the symmetry needs to be softly broken, as an unbroken symmetry would forbid an inflaton mass altogether, and there is no gracefull ending to inflation. On also has to make sure that other parts of the scalar potential (e.g. moduli stabilization physics that also breaks susy) do not badly break the symmetries. (3) Go beyond slow roll inflation. Introduce non-trivial kinetic terms (K-inflation, DBI inflation), chain inflation, or even ekpyrotic type of scenarios.

6. Brane inflation

There are several motivations for considering inflation in string theory. First of all, as argued in the last subsection, inflation is sensitive to UV physics. String theory provides an UV completion, making it possible to explicitly calculate observables. Secondly, string theory contains new ingredients not (readily) present in field theories, such as branes, moduli, extra dimensions and warp factors. Maybe these can be used for inflation. It is not impossible that stringy physics leaves an imprints on the CMB (although it will be hard if not impossible to disentangle these effects). Then inflaton provides a way to probe string theory.

Most models constructed sofar have the Hubble constant during inflation H_I much smaller than the Kaluza-Klein and the string scale. In such a set-up only string zero modes enter, and one can study the low effective sugra approximation instead of the full string theory. The effects of the extra dimensions can be integrated out, and the inflationary solutions are intrinsically 4D (instead of that we have to look for solutions of the full 10D Einstein equations).

We discuss here the KKLMMT brane inflation model, which was the first inflation model constructed in an explicit string compactification (of IIB string theory). In the end inflation does not come out naturally. It is quite intstructive to see why, as the problems that arise are quite generic for stringy models of inflation.

In brane inflation models, the inflaton is identified with the distance between a pair of branes. The interaction potential is generated by graviton, dilaton and RR-fields exchange between the branes, and is of the form of a Coulomb potential. Moreover, the branes respresent a source of energy density, which provides nearly constant energy density that drives inflation. As the branes approach each other, strings stretching between them (remember: branes are higher dimensional hypersurfaces in string theory on which strings can end), become light. If the distance drops between some critical distance, one of these strings becomes actually tachyonic, and the two branes annihilate; in the effective field theory this is described by a phase transition which ends inflation. The effective potential is that of hybrid inflation.

We will only outline the main ingredients that go into the model. More details can be found in the original references.

6.1 Brane potential

We will consider a model with 3+1 dimensional $D3$ and anti-$D3$ brane that span our observable universe. Both objects are pointlike in the compactified 6D manifold $\mathcal{M}$. Consider the following 10D metric, which is a product space of our 4D minkowski space and the extra 6D manifold $\mathcal{M}$ (with $\mathcal{M}$ Ricci flat)

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + (\mathrm{d}r^2 + r^2 \mathrm{d}s^2_{X_5}) \tag{6.1}$$

with x^μ parametrizing 4D space, and r the direction along the D3-$\overline{D3}$ separation, and X_5 some angular space whose details are not important.

In addition to being a delta-function source for stress energy, the $D3/\overline{D3}$ branes also source the $F_{(5)}$ field strength, which comes form the RR gauge potential $C_{(4)}$ (we are working in IIB string theory). The full action is the Dirac-Born-Infeld (DBI) action plus a topological Chern-Simons term [7]

For the simple case that the background fields $B_{ab} = F_{ab} = 0$ and the dilaton $\Phi = 0$ are set to zero, the action is

$$S_{D3/\overline{D3}} = -T_3 \int \mathrm{d}^4x \sqrt{|G|} \pm T_3 \int \mathrm{d}^4x (C_4)_{0123} \tag{6.2}$$

with T_3 the D3 brane tension. The 1st term is the DBI action with the langrangian just the worldvolume of the brane, and the 2nd term is the Chern-Simons term. The plus and minus sign is for the $D3$ brane and $\overline{D3}$ anti-brane respecively, which have opposite C_4 charges. The equation of motion for the RR C_4 field has the solution $(C_4)_{0123} = 1$, and the CS term is trivial. Now evaluate the DBI action. The induced metric on the brane is given by

$$G_{\mu\nu} = G_{AB}\frac{\mathrm{d}x^A}{\mathrm{d}x^\nu}\frac{\mathrm{d}x^B}{\mathrm{d}x^\nu} = g_{\mu\nu} + \partial_\mu r \partial_\nu r \tag{6.3}$$

For inflation we are inteterested in homogoneous and isotropic backgrounds with $r = r(t)$. Then

$$G = \mathrm{diag}(g_{00} + \dot{r}^2, g_{11}, g_{22}, g_{33}) \quad \Rightarrow \quad \sqrt{|\det(G_{AB})|} = \sqrt{-g}\sqrt{1 - g^{00}\dot{r}^2} \tag{6.4}$$

here $\mu, \nu = 1, 4$ run over our (3+1)-dimensions, whereas $A, B = 1, .., 10$ run over all dimension. In the limit of small velocities we can Taylor expand the square root to get

$$S_{D3/\overline{D3}} = T_3 \int \mathrm{d}^4x \sqrt{-g}\left(-1 \pm 1 + \frac{1}{2}\dot{r}^2 + \mathcal{O}(\dot{r}^4)\right) \tag{6.5}$$

For the D3 the two constant terms cancel and the action is that of a free field. There is no net force on the brane. For the $\overline{D}3$ brane the terms instead add, and this provides a non-zero constant energy density $V_0 = 2T_3$ that can drive inflation (the $\overline{D3}$ breaks susy). From the kinetic term, it follows that the canonically normalized interbrane distance, to be identified with the inflaton field, is $\phi = T_3 \delta r$ with $\delta r = r_{D3} - r_{\overline{D3}}$.

The action (6.5) is for a $D3$ or a $\bar{D}3$ brane. But if both are added, there is in addition an interaction term. This comes about as a $D3$ is added somewhere a r, this perturbs

[7] This is just the higher dimensional analogue for the action of a charged point particle, which consist of the gravitational action of the worldline (the DBI part), plus a Wilson line coupling to the gauge field $e\int A_\mu \mathrm{d}x^\mu$ where the integral is along the worldlin of the particle, and e is the charge (the CS part)

the background metric. Pluggin in the perturbed metric in the action for the antibrane, and expanding, then gives the interaction potential $\propto 1/r^4$. Alternatively, this term can be understood as due to closed string exchange between the branes. It is just the higher dimensional analogue of the Coulomb interaction between two charged particles in electrodynamics, and follows from the analogue of Gauss law to be $F \propto 1/r^5$. In terms of the canonically normalized field ϕ the potential is (for $r \gg l_s$)

$$V = 2T_3 \left(1 - \frac{1}{2\pi^3} \frac{T_3^3}{m_{\rm p,10}^8 \phi^4} \right) \tag{6.6}$$

with $m_{\rm p,10}$ the 10D planck mass. The Coulomb interaction is attractive. The relation between the 10D and 4D planck mass can be found by integrating out the extra dimensions from the 10D Einstein action

$$S_{\rm bulk} = \frac{1}{16\pi G_{10}} \int {\rm d}^{10}x \sqrt{-g_{10}} R_{10} \quad \Rightarrow \quad \frac{1}{16\pi G_N} \int {\rm d}^4 x \sqrt{-g} R + ... \tag{6.7}$$

with $8\pi G_{10} = m_{\rm p10}^{-8}$ and $8\pi G_N = m_{\rm p}{}^{-2}$. Thus $m_{\rm p}^2 = m_{\rm p10}^8 L^6$ with L^6 the volume of $\mathcal{M}$.

The η parameter for (6.6) is

$$\eta = -\frac{10}{\pi^3} (L/r)^6 \sim -0.3 (L/r)^6 \tag{6.8}$$

Hence $\eta \ll 1$ is possible only for $r > L$ — but two branes cannot be separated by a distance greater than L in a manifold $\mathcal{M}$ of size L! The potential is not flat enough for inflation. [8]

6.2 Inflation in a warped background

Consider the warped metric

$${\rm d}s^2 = h^{-1/2} g_{\mu\nu} {\rm d}x^\mu {\rm d}x^\nu + h^{1/2} ({\rm d}r^2 + r^2 {\rm d}s_{X_5}^2) \tag{6.9}$$

with X_5 some 5D compact angular space whose details will not be important (for the Klebanov-Strassler solution it is the Einstein-Sasaki manifold, and the extra dimensions are conformal CY). In IIB string theory this is a solution of 10D sugra equations in the presence of fluxes, and as such a good background. It is a stringy realization of the Randall-Sundrum scenario. The AdS region is truncated at some minimum $r = r_0$, and at large $r/R \sim 1$ the throat is glued into a CY space. The metric above describes the geometry in the "throat" away from the tip. Extradimensional space is compact, and Newton's constant G_N is finite.

The warp factor is $h(r) = (R/r)^4 +$ log corrections. Neglecting the log corrections the geometry is $AdS_5 \times X_5$ for 4D Minkowski times a radial direction (times the angular space), and R is the curvature scale of anti deSitter (AdS). R is a constant, depending on the fluxes. The region of small r is the bottom of a gravitational well. Energies along the x^μ coordinates

[8]It can be shown that taking asymmetric extra dimensions, with some compact dimensions much larger than others, does not improve the situation.

therefore get increasingly redshifted as r decreases. Thus the energy of an object, such as a brane, depends on its position along the radial direction.

Reconsider brane inflation in the presence of warping. One can redo the small velocity expansion of the DBI action (6.5) to give

$$S_{D3} = T_3 \int \mathrm{d}^4x\sqrt{-g}\left(-h^{-1}(1\pm 1) + \frac{1}{2}\dot{r}^2 + ...\right) \tag{6.10}$$

The canonical normalized field is as before. But $V_0 = 2T_3 h^{-1}$ is redshifted. The net force on a $D3$ is still zero, but for the anti-brane $V_0 = 2T_3 h^{-1}$ is now dependent on r which enters via the warp factor. The $\overline{D3}$ brane energy is minimized for $r \to r_0$. Inflation takes place with the $\overline{D3}$ localized at the bottom of the throat, and the the $D3$ slowly moving towards the bottom, attracted by the coulomb force. The coulomb force also picks up warp factors

$$V = 2T_3 h_0^{-1}\left(1 - h_0^{-1}\frac{R^4T_3^3}{N\phi^4}\right) \tag{6.11}$$

with $N = MK$ all integers (corresponding to flux quanta). Further $h_0^{-1} = h(r_0^{-1}) = (r_0/R)^4 = \mathrm{e}^{-8\pi K/(3g_sM)}$ is the warp factor at the bottom of the throat. The η parameter is

$$\eta \sim \eta_{\text{unwarped}}\left(\frac{r_0}{R}\right)^4 . \tag{6.12}$$

Moderate warping is needed to make the potential flat enough for inflation.

The warping dependence of the coulomb term can be understood as follows. In the KS throat the warp factor is given in terms of a function $h = \mathrm{e}^{-4A}$ which obeys Laplace equation, with fluxes and branes as sources. In particular, a single D3 located at $r = r_1$ will correct the background according to $h_{\text{new}} = h(r) + \Delta h(r, r_1)$. Here $h(r)$ is the background warp factor appearing in the metric, and $\Delta h(r, r_1)$ the correction due to the brane. In a region where the original warp factor is very small $h(r_0) \gg 1$, so that the total warp factor can be expanded

$$V \propto h(r_0)^{-1}\left(1 - \frac{\Delta h(r, r_1)}{h(r_0)}\right) \tag{6.13}$$

If $h(r_0) \gg 1$ this typically gives a very flat potential.

6.3 Moduli stabilization

So far we have discussed brane inflation in a fixed background. However, in string theory spacetime is dynamic. For an internal manifold of size L, the main contribution to the inflationary energy comes from the $\overline{D3}$ tension, which with length scales factored out explicitly is (derivation below)

$$V \sim \frac{2T_3\tilde{h}(r_0)^{-1}}{L^{12}} \tag{6.14}$$

which is a runaway potential for L. The extra-dimensional space needs to be stabilized. Only then does it make sense to discuss inflation.

The parameters required to describe a geometry of compactified space are known as moduli. They arise as there are a whole class of solutions to the 10D einstein equations. As the solution are degenerate in energy and curvature, each modulus corresponds to a massless 4D scalar in the low energy EFT. This can be seen from the KK reduction of the metric (see appendix for more details).

Kachru, Kallosh, Linde & Trivedi (KKLT) came with the first explicit scheme to stabilize all moduli. They did so in a three step way. First they introduced fluxes, non-trivial configurations of various antisymmetric tensor fields. In particular

$$\frac{1}{2\pi\alpha'^2}\int_A F = M, \qquad \frac{1}{2\pi\alpha'^2}\int_B H = K, \tag{6.15}$$

with $F = \mathrm{d}C$ and $H = \mathrm{d}B$ are 3-form fluxes, M, K are integers, and A, B are 3-surfaces. Flux quantization — flux of these fields through topolgically nontrinvial surfaces in the extra dimensions is quantized — implies the value for fields like C_{MN} must grow as the areas of these surfaces shrink, ensuring such changes come with an energy. Fluxes can stabilize the dilaton and the complex structure moduli which parametrize the shapes of the extra dimensions.

Another important consequence of non-zero fluxes is that the backreaction on the geometry gives rise to a warped throat. Hence flux compactifications give as a byproduct the setting for warped brane inflation.

The resulting 4D EFT is $N = 1$ SUSY with the complex moduli stabilized at some high string scale. The Kahler moduli, in particular the volume modulus, are not stabilized by fluxes. They in turn are stabilized by non-perturbative effecs from gaugino condensation on $D7$ branes, or from instanton effects. Consider the case that the only Kahler modulus is the volume modulus. The superpotential is

$$W = W_0 + A\mathrm{e}^{-aT} \tag{6.16}$$

with T a 4D complex scalar whose real part parametrizes the volume of compactified space. W_0 is a constant from integrating out the complex structure moduli. The exponential term is from non-perturbative physics (note that $g^{-2} = \mathrm{Re}(f) \propto \mathrm{Re}T$ with f the gauge kinetic function). This does the job of stabilizing the modulus, but the minimum is susy and AdS.

The last step is to add an uplifting term to get a Minkowski vacuum (or dS with small cosmological constant). KKLT suggested doing so by adding an anti-D3 brane to the system. The problem is that a $\overline{D3}$ breaks $N = 1$ SUSY explicitly (and its effects are not captured by the $N = 1$ sugra action). Although this gives much less control over the corrections to the calculation, the damage can be kept small if the contribution of the antibrane to the low-energy action can be made parametrically weak. This can plausibly be done in the case that there is a strongly warped throat, because in this case the antibrane can minimize its energy by moving to the throat's tip. Then the contribution of the antibrane to the low-energy action can be computed perturbatively in T_3/h_0, which to leading order means simply adding the sugra potential and the uplifting term.

6.4 The η problem rears its head again

The volume modulus needs to be included in the low energy EFT. Let's see how brane inflation is affected. The metric depencence on the breathing mode can be made explicit by introducing $\mathrm{Re}(T) = \mathrm{e}^{4u}$, and

$$\mathrm{d}s^2 = \frac{\mathrm{e}^{-6u}}{\sqrt{\tilde{h}}}\tilde{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b + \mathrm{e}^{2u}\sqrt{\tilde{h}}\tilde{g}_{mn}\mathrm{d}y^m\mathrm{d}y^n \tag{6.17}$$

where $\tilde{g}_{ab}$ the 4D metric in the Einstein frame, and $\tilde{g}_{mn}$ is the metric with some ficucial volume on the CY. Note that the warp factor depends on the extradimensional coordinates, whereas T is a 4D scalar field, and consequently e^{2u} only depends on the 4D coordinates. The warp factor in (6.17) is related to the one defined in the previous subsection (6.9) by $\mathrm{e}^{4u}\tilde{h} = h$, as we factored out the length scale explicitly. The factor e^{2u} in the 6D part of the metric is understandable with $\mathrm{e}^u \sim L$, as it parametrizes the volume of compactified space. The factor e^{-6u} in the 4D part is factored out to put the metric in the Einstein frame, where there is no mixing of the Kahler modulus with the 4D graviton. Indeed, dimensional reduction of the the 10D action gives

$$S_{\mathrm{bulk}} = \frac{1}{16\pi G_{10}}\int \mathrm{d}^{10}x\sqrt{-g_{10}}R_{10} \supset \frac{1}{16\pi G_N}\int \mathrm{d}^4x\sqrt{-g}\left[R + K_{T\bar{T}}\partial T\partial\bar{T}\right] \tag{6.18}$$

where the 4D curvature is constructed from the $\tilde{g}_{ab}$ metric. The Kahler potential is $K = -3\log[T+\bar{T}] = -2\log(V_6)$, amd

$$2\sigma = \mathrm{e}^{4u} = V_6^{2/3} \tag{6.19}$$

In the absence of the $D3$ one has $2\sigma = T + \bar{T}$.

With the metric in hand we can calculate explicitly how the energy density driving inflation depends on the volume modulus. Indeed, as before consider the DBI + CS action for $\overline{D3}$ located at the tip of the throat, but now in the background (6.17)

$$\mathcal{L}_{\overline{D3}} = -2T_3\int \mathrm{d}^4x\sqrt{-G} = -T_3\int \mathrm{d}^4x\frac{\sqrt{-\tilde{g}}}{\tilde{h}_0 L^{12}} \tag{6.20}$$

Using that $\tilde{h}_0 = h_0\mathrm{e}^{4u}$, and (6.19), then gives

$$V_{\overline{D3}} = \frac{E}{(2\sigma)^2} \tag{6.21}$$

with $E = T_3/h_0$. Withouth the volume T stabilized, this potential presents a runaway direction for σ, as was mentioned above.

Up to now we described the background: a flux compactification with warped throat, volume modulus in low energy EFT, and a $\overline{D3}$ for uplifting and (another one) to provide inflationary energy density (can have several throats with anti-branes, or other sources of susy breaking). To get inflation we now need to add a $D3$ to this set-up. The presence of the brane perturbs the background. Let's see how.

Consider the DBI + CS action for the $D3$ in the slow velocity limit, using the metric (6.17). Now

$$G_{\mu\nu} = \frac{\mathrm{e}^{-6u}}{\sqrt{h}} - \mathrm{e}^{2u}\sqrt{h}\tilde{g}_{mn}\frac{\partial X^m}{\partial x^\mu}\frac{\partial X^n}{\partial x^\nu} \tag{6.22}$$

As before, aligning the brane fluctuations, ie the inflaton direction, with the radial coordinate, and taking the low velocity limit gives

$$S_{D3} = \int \mathrm{d}^4x\sqrt{-\tilde{g}}\frac{3T_3}{4\sigma}(\partial_\mu r)^2 + ... = \int \mathrm{d}^4x\sqrt{-\tilde{g}}\frac{3}{2\sigma}(\partial_\mu\phi)^2 + ... \tag{6.23}$$

where we used (6.19), and as before we absorbed the brane tension in the ϕ-field. Note that ϕ is not the canonically normalized inflaton field. The form of kinetic term suggest that the kahler potential in the presence of the $D3$ should be generalized to

$$K = -3\log(2\sigma) = -3\log\left(T + \bar{T} - \kappa k(\phi,\bar{\phi})\right) \tag{6.24}$$

with $k(\phi,\bar{\phi})$ the kahler potential of the CY metric $\tilde{g}_{mn}$, in the sense that $g_{i\bar{j}} = \partial_i\partial_{\bar{j}}k$. It can be shown that near the bottom of the throat $k = \sum\phi^i\bar{\phi}^i + ...$ Calculating the kinetic term from the above kahler, $\mathcal{L}_{\rm kin} = K_{i\bar{j}}\partial\phi_i\overline{\partial\phi_j}$, indeed agrees with the DBI result

$$K_{i\bar{j}} = \frac{3k_{ij}}{2\sigma} + \frac{3k_ik_{\bar{j}}}{2\sigma} \approx \frac{3\delta_{ij}}{2\sigma} \tag{6.25}$$

where in the last line we used the explicit form of k.

Now that we know how the modulus enters the EFT, we can see how it affects the η-problem of inflation. Using the sugra expression for the F-term part of the potential (5.18), and adding the effects of the anti $D3$ (6.21) gives

$$\begin{aligned} V &= \frac{1}{(T-\phi\bar{\phi}/2)^2}\left(W_T^2T - 3WW_T + \frac{E}{4}\right) \\ &= \frac{V(T)}{(1-\phi\bar{\phi}/(2\sigma))^2} \approx V_0(T)\left(1 + 2\frac{\phi\bar{\phi}}{2\sigma} + ...\right) \end{aligned} \tag{6.26}$$

which generates a mass for the inflaton field. Noting that the canonically normalized field is $\varphi = \phi\sqrt{3/(2\sigma)}$, and assuming $\sigma = \sigma_0$ stabilized by the non-perturbative potential, it follows

$$\eta = \frac{2}{3} \tag{6.27}$$

which is too large for inflation. The problem is that the moduli stabilization mechanism, the non-perturbative potential, stabilizes T and not σ. This implies that as the brane moves and ϕ changes, there is a change in the potential.

The ϕ dependence in the Kahler potential, and its effect on the potential, describs a force on the $D3$ brane once the modulus get stabilized because V acquires non-trival dependence on ϕ. Physically, the absence of ϕ in V expresses the absence of a net static force at tree level between the $D3$ and $\overline{D3}$, with the fluctuations in the brane position just a free field, which is the situation we discussed in the previous subsection. However, with moduli effects included, what happens is that if the $D3$ is moved within the extra dimensions the distibution of forcess acting on the branes adjusts, as they try to maintain their cancellation at the new position of the D3. This adjustment in turn causes the volume modulus to change, as the internal geometry responds to the new distirbution of forces. The change of the extra-dimensional volume costs no energy so long as the breathing mode is an unstabilized modulus. But once this modulus has been stabilized the energy cost associated with this adjustment induces a force (expressed by the interactions of T and ϕ in K) which tends to localize the $D3$ at a specific position within the extra dimensions.

Inflation is ruined unless there is a compensating contribution to the mass term from some other source. One possibility is a dependence of the superpotential on ϕ. If $V_0 = V_0(T, \phi)$ (instead of $V_0(T)$), there is an additional contribution to the mass term. Writing $V_0 = X(T, \phi)/T^2$, and expanding aroun (T_0, ϕ_0) this gives

$$V = \frac{X(T,\phi)}{T^2}\left(1 + 2\frac{\phi\bar{\phi}}{2\sigma} + ...\right)\Big|_0 + \frac{X_{\phi\bar{\phi}}}{\sigma^2}\Big|_0 \phi\bar{\phi} \tag{6.28}$$

In principle the second contribution to the mass term might substaialy cancel the first, alleviating the problem of the inflaton mass. This would certainly require fine-tuning at the level of one percent.

Modifications that introduce ϕ dependence directly into W describe a scond kind of force experienced by the $D3$. This force arises due to the back reaction of the $D3$ onto the background extra-demensional geometry, since this changes the volume of the cycle wrapped by any D7 branes, and thereby changes the gauge couplings of the interactions on these branes (such as those that generate $W_{\rm np}$). This force is calculable for a KS throat, it changes $A = A(\phi)$ in the non-perturbative superpotential. Fine-tuning the various contributions to the potential an inflation model has been constructed (with inflation at an inflection point).

7. Literature

We list here mainly some review papers/lectures. All references to the original literature can be found in those. Ref. [1, 5] are review papers, explaining the basics of inflation and the generation of density perturbations. A good starting point for non-gaussianities, is [6, 7]; the curvaton scenario is described in [8]. Review papers more geared towards string inflation are [9, 11]. Original papers on KKLMMT are [12, 15].

A. Curvature perturbation constant on superhorizon scales

The constancy of $\dot{\zeta}$ follows from energy momentum conservation

$$0 = \nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\kappa} T^\kappa_\nu - \Gamma^\kappa_{\nu\mu} T^\mu_\kappa \tag{A.1}$$

To calculate the covariant derivative we first need to calculate the perturbed affine connections. We also need the perturbed energy momentum tensor. The energy momentum tensor for a perfect fluid is of the form

$$T^{\mu\nu} = (\rho + p)u^\mu u_\nu + p\delta^{\mu\nu} \tag{A.2}$$

with $u^\mu = \mathrm{d}x^\mu/\mathrm{d}\tau$, with τ the proper time, the fluid 4-velocity normalized to $u^\mu u_\mu = -1$. Because of isotropy, the 0th order fluid 3-velocity is zero and $u^\mu = a^{-1}(1, 0)$. The 3-velocity only enters at 1st order and can be defined via $\delta u^i = v^i$. As before the 3-vector can be decomposed into a scalar and vector part $v^i = v_,{}^i + \bar{v}^i$, of which we are only interested in the scalar. Thus at 1st order $u^\mu = a^{1-}(1 - \phi, v_,{}^i)$. The perturbed energy momentum tensor is

$$T^0_0 = -(\rho + \delta\rho), \qquad T^i_0 = (\rho + p)v_,{}^i, \qquad T^i_j = (p + \delta p)\delta^i_j \tag{A.3}$$

and there is no anisotropic stress T^i_j for $i \neq j$ is zero.

The 1st order metric is and its inverse is

$$g_{\mu\nu} = a^2 \begin{pmatrix} -(1+2\phi) & B_{,i} \\ B_{,i} & (1-2\psi)\delta_{ij} + 2E_{,ij} \end{pmatrix}, \qquad g^{\mu\nu} = \frac{1}{a^2}\begin{pmatrix} -(1-2\phi) & B_,{}^i \\ B_,{}^i & (1+2\psi)\delta^{ij} - 2E_,{}^{ij} \end{pmatrix} \tag{A.4}$$

To find the inverse we used that in general we can write $g^{00} = -a^{-2}(1-X)$, $g^{0i} = a^{-2}Y_,{}^i$ and $g^{ij} = a^{-2}((1+2Z)\delta^{ij} + K_,{}^{ij})$. Then from the relation $g_{\mu\rho}g^{\rho\nu} = \delta^\nu_\mu$ the unknowns X, Y, Z, K can be determined to 1st order. The expression for the affine connections in terms of the metric is

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2}g^{\alpha\rho}\left(\frac{\partial g_{\rho\gamma}}{\partial x^\beta} + \frac{\partial g_{\beta\rho}}{\partial x^\gamma} - \frac{\partial g_{\beta\gamma}}{\partial x^\rho}\right) \tag{A.5}$$

The perturbed corrections are

$$\begin{aligned} &\Gamma^0_{00} = \mathcal{H} + \phi', && \Gamma^i_{0j} = (\mathcal{H} - \psi')\delta^i_j + E'{}_{,ij}, \\ &\Gamma^0_{ij} = \mathcal{H}(1 - 2(\phi+\psi)\delta_{ij} - 2E_{,ij}) - \psi_i\delta_{ij} + (E' - B)_{,ij} && \Gamma^0_{0i} = \phi_{,i} + \mathcal{H}B_{,i}, \\ &\Gamma^i_{jk} = \psi_{,j}\delta^i_k - \psi_{,k}\delta^i_j + \psi_,{}^i\delta_{jk} - \mathcal{H}B_,{}^i\delta_{jk} + E_{ijk} && \Gamma^i_{00} = \mathcal{H}B_,{}^i + B'^i_, + \phi_,{}^i, \end{aligned} \tag{A.6}$$

Only Γ^0_{00}, Γ^i_{0j}, Γ^0_{ij} are non-zero at 0th order.

With this in hand we can evaluate A.1 with $\nu = 0$ to 1st order

$$\begin{aligned} 0 &= \partial_0 T^0_0 + \partial_i T^i_0 + \Gamma^\mu_{\mu 0}T^0_0 + \Gamma^\mu_{\mu i}T^i_0 - \Gamma^\kappa_{0\mu}T^\mu_\kappa \\ &= \partial_0 T^0_0 + \partial_i T^i_0 + (\Gamma^0_{00}T^0_0 + \Gamma^i_{i0}T^0_0) - (\Gamma^0_{00}T^0_0 + \Gamma^i_{0j}T^i_j + \Gamma^i_{0j}T^i_j) + \mathcal{O}(2^{\mathrm{nd}}) \\ &= (\rho + \delta\rho)' + (\rho + p)\partial_i v_,{}^i + (3\mathcal{H} - 3\psi' + E'_{,ij})(\rho + \delta\rho - p - \delta p) \end{aligned} \tag{A.7}$$

Using that the background fields satisfy energy momentum conservation at 0th order then gives the required result (3.37).

B. Moduli are massless modes in the 4D EFT

[From [11]]. To see how this works, first recall how to compactify a fluctuation in a 10D scalar field, $\delta\phi(x,y)$, whose 10D field equation is $\Box_{10}\delta\phi = g^{MN}D_M D_N \delta\phi = 0$. Evaluated for a product metric like eq. (6.1), this becomes $(\Box_4 + \Box_6)\delta\phi = 0$, where $\Box_6 = g^{mn}D_m D_n$ and $\Box_4 = \eta^{\mu\nu}\partial_\mu\partial_\nu$. If we decompose $\delta\phi(x,y)$ in terms of eigenfunctions, $u_k(y)$, of $\Box_6$ — *i.e.* where $\Box_6 u_k = -\mu_k^2\, u_k$ — we have

$$\delta\phi(x,y) = \sum_k \varphi_k(x)\, u_k(y)\,, \tag{B.1}$$

and the equations of motion for ϕ imply $(\Box_4 - \mu_k^2)\varphi_k = 0$. The 10D field decomposes as an infinite number of 4D Kaluza-Klein fields, each of whose 4D mass is given by the corresponding eigenvalue, μ_k. In particular a massless mode in 4D corresponds to a zero eigenvalue: $\Box_6 u_k = 0$.

A similar analysis also applies for the fluctuations, $\delta g_{MN}(x,y)$, in the 10D metric about a specific background geometry. Focussing on metric components in the extra dimensions, $\delta g_{mn}(x,y)$, allows an expansion similar to eq. (B.1)

$$\delta g_{mn}(x,y) = \sum_k \varphi_k(x)\, h^k_{mn}(y)\,, \tag{B.2}$$

where $h_{mn}(y)$ are tensor eigenfunctions for a particular 6D differential operator (the *Lichnerowitz operator*) obtained by linearizing the Einstein equations, $\Delta_6 h^k_{mn} = -\mu_k^2\, h^k_{mn}$. Again the 10D equation of motion, $\Delta_{10}\delta g_{mn} = 0$, implies each 4D mode, $\varphi_k(x)$, satisfies $(\Box_4 - \mu_k^2)\varphi_k = 0$, and so has mass μ_k.

The significance of moduli is that they provide zero eigenfunctions for Δ_6, and so identify massless 4D scalar fields within the KK reduction of the extra-dimensional metric. The zero eigenfunction is given by the variation of the background metric in the direction of the moduli. Schematically, if ω_a are the moduli of the background metric, $g_{mn}(y;\omega)$, and if $h^a_{mn} = \partial g_{mn}/\partial\omega_a$, then $\Delta_6 h^a_{mn} = 0$. Physically, these are zero eigenfunctions because varying a modulus in a given solution to the Einstein equations gives (by definition) a new solution to the same equations, and so in particular an infinitesimal variation in this direction is a zero mode of the linearized equations.

Because the 4D moduli fields, $\varphi_a(x)$, are massless they necessarily appear in the low-energy 4D effective action which governs the dynamics at scales below the KK scale, M_c. If we focus purely on the moduli and the 4D metric (and ignore other fields), then the low-energy part of this action must have a potential, V, which is independent of the moduli, $\varphi^a(x)$.

References

[1] A. Riotto, arXiv:hep-ph/0210162.

[2] D. H. Lyth and A. Riotto, Phys. Rept. **314** (1999) 1 [arXiv:hep-ph/9807278].

[3] D. Baumann, arXiv:0710.3187 [hep-th].

[4] D. Langlois, arXiv:hep-th/0405053.

[5] W. H. Kinney, arXiv:0902.1529 [astro-ph.CO].

[6] J. M. Maldacena, JHEP **0305** (2003) 013 [arXiv:astro-ph/0210603].

[7] D. Babich, P. Creminelli and M. Zaldarriaga, JCAP **0408** (2004) 009 [arXiv:astro-ph/0405356].

[8] D. H. Lyth, C. Ungarelli and D. Wands, Phys. Rev. D **67**, 023503 (2003) [arXiv:astro-ph/0208055].

[9] J. M. Cline, arXiv:hep-th/0612129.

[10] D. Baumann and L. McAllister, arXiv:0901.0265 [hep-th].

[11] C. P. Burgess, PoS **P2GC** (2006) 008 [Class. Quant. Grav. **24** (2007) S795] [arXiv:0708.2865 [hep-th]].

[12] S. Kachru, R. Kallosh, A. Linde and S. P. Trivedi, Phys. Rev. D **68**, 046005 (2003) [arXiv:hep-th/0301240].

[13] S. Kachru, R. Kallosh, A. Linde, J. M. Maldacena, L. P. McAllister and S. P. Trivedi, JCAP **0310** (2003) 013 [arXiv:hep-th/0308055].

[14] D. Baumann, A. Dymarsky, S. Kachru, I. R. Klebanov and L. McAllister, arXiv:0808.2811 [hep-th].

[15] D. Baumann, A. Dymarsky, I. R. Klebanov and L. McAllister, JCAP **0801**, 024 (2008) [arXiv:0706.0360 [hep-th]].