Lectures on Inflation

Mehrdad Mirbabayi

Abstract

An introduction to inflation.

Contents

1 FLRW cosmology

Our universe is to a good approximation homogeneous and isotropic at large scales. FLRW model is an idealized version of this in which the symmetries are exact. In this universe there are preferred time-slices, in terms of which, the metric can be written as

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right) \tag{1}$$

where $d\Omega^2$ is the metric of the unit 2-sphere, and there are three inequivalent choices for K:

- $K = 1$, called "closed" cosmology. Constant-t slices (or spatial slices) are positively curved like 3-spheres.
- $K = 0$, called "flat" cosmology. Constant-t slices are flat Euclidean.
- $K = -1$, called "open" cosmology, with hyperbolic spatial slices.

There is a set of distinguished observers on FLRW, called **comoving**. Their 4-velocity is $u^\mu = (1, 0, 0, 0)$. The spatial coordinates are labeling these observers, and are called comoving coordinates.

With these symmetries the energy-momentum tensor, takes the form of a perfect fluid (even though the microscopic origin is often different):

$$T^0_0 = -\rho(t), \qquad T^i_j = \delta^i_j p(t), \tag{2}$$

and 0 for all other components. Energy-momentum conservation, implies

$$\dot{\rho} = -3H(\rho + p), \tag{3}$$

where the **Hubble** parameter is

$$H \equiv \frac{\dot{a}}{a}. \tag{4}$$

In thermodynamics the relation between p and ρ at equilibrium is called the **equation of state**. In cosmology, it is useful to *define* it as

$$w = \frac{p}{\rho}. \tag{5}$$

If $w =$ constant, then (3) and (5) can be solved to give

$$\rho = \rho_0 \left(\frac{a_0}{a} \right)^{3(1+w)}. \tag{6}$$

Some examples are

- *Radiation:* $w = 1/3$, $\rho \propto a^{-4}$.
- Non-relativistic *matter:* $w = 0$, $\rho \propto a^{-3}$.

- *Cosmological Constant (CC):* $T_{\mu\nu} = \Lambda g_{\mu\nu}$, $w = -1$, $\rho =$ constant.

Dynamics of FLRW cosmology is governed by the homogeneous matter equations of motion (often reduced to energy-momentum conservation) and the Friedmann equation that follows from the 00 component of the Einstein equation

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

It is common to think of spatial curvature as another contribution to the ρ:

- *Curvature:* $w = -1/3$, $\rho_K \propto a^{-2}$.

Example: Flat matter dominated universe. Let's denote by index 0, quantities measured today. Since $K = 0$, by a rescaling of spatial coordinates, we can set $a_0 = 1$. Then

$$a = \left(\frac{3H_0 t}{2}\right)^{2/3} \Rightarrow H = \frac{2}{3t}. \tag{8}$$

Therefore the **age of the MD universe** is

$$t_0 = \frac{2}{3H_0}. \tag{9}$$

We can check that the (4d) curvature length is $\sim 1/H$. Hence density and invariants like R^2 diverse as $t \to 0$. This is called the **big bang singularity**.[1]

Critical density: Given the value of the Hubble parameter H, critical density is defined as the density needed for K to vanish:

$$\rho_{\rm cr} = \frac{3H^2}{8\pi G}. \tag{10}$$

The fraction of a particular contribution to the energy density (labeled by i) to the critical density is a good measure of how important that component is

$$\Omega_i = \frac{\rho_i}{\rho_{\rm cr}}. \tag{11}$$

If we also define

$$\Omega_K = -\frac{K}{a^2 H^2} \tag{12}$$

then Friedmann equation implies

$$\sum_i \Omega_i + \Omega_K = 1. \tag{13}$$

Today, in our universe

$$\Omega_\Lambda \approx 0.7, \quad \Omega_m \approx 0.3, \quad \Omega_r \approx 10^{-4}, \quad |\Omega_K| < 10^{-2}. \tag{14}$$

[1] $H \to \infty$ generically corresponds to a singularity, but not always. Milne universe is an exception.

Note that even though the sum is always 1, different components dilute at different rates, so their relative importance changes with time. For instance, when universe was a 1000 times smaller (around recombination), $\Omega_m \sim \Omega_r \sim 1$ while $|\Omega_K| < 10^{-5}$, $\Omega_\Lambda \sim 10^{-9}$. The smallness of the latter two is related to two profound puzzles, the flatness problem and the cosmological constant problem.

Redshift. Cosmological observations are predominantly via photons. In many cases, we know the frequency at which they are emitted (recombination photons with black body spectrum at $T \sim 1$eV, particular atomic lines from stars and interstellar medium, etc.) and we observe them at a lower frequency because of the redshift. The earlier they are emitted, the larger is the redshift. Hence, it is possible (and convenient) to use redshift z, defined as

$$1 + z = \frac{a_0}{a} \tag{15}$$

as a time variable, one that runs backward. The frequency of a photon at the time of emission at redshift z is related to its observed frequency ω_o by

$$\omega_e = (1+z)\omega_o. \tag{16}$$

To derive this relation, it is useful to switch to a new radial coordinate χ

$$ds^2 = -dt^2 + a^2(d\chi^2 + r^2(\chi)d\Omega^2) \tag{17}$$

where

$$r(\chi) = \begin{cases} \sin\chi, & K = 1, \\ \chi, & K = 0, \\ \sinh\chi, & K = -1 \end{cases} \tag{18}$$

Suppose we observe the photon at $\chi = 0$, then by isotropy $k^\mu = (k^t, k^\chi, 0, 0)$, and by χ-independence of the $t - \chi$ part of the metric

$$k^\chi = \frac{p}{a^2}, \qquad p = \text{constant}. \tag{19}$$

Since k^μ is null $k^t = a|k^\chi|$. The frequency of photons measured by comoving observers is $\omega = -u^\mu k_\mu$, where $u^\mu = (1,0,0,0)$. This gives (16).

It is often useful to draw spacetime diagrams to illustrate the causal structure. For this purpose, we switch to conformal time $d\tau = dt/a$, in terms of which the $\tau - \chi$ part of the metric is conformally flat

$$ds^2 = a^2(-d\tau^2 + d\chi^2 + r(\chi)^2 d\Omega^2). \tag{20}$$

1. Find the conformal time between now and the big bang assuming a flat matter dominated universe with expansion rate H_0.

2 Puzzles of hot big bang cosmology

Flatness problem. Suppose the universe is dominated by a single component with $p/\rho = w$, and $|\Omega_{K,0}| < 10^{-2}$. In the past

$$\Omega_K = \Omega_{K,0} a^{1+3w}. \tag{21}$$

If $w > -1/3$, which is the case for most of the cosmic history assuming no phase transition changes the composition (14), then Ω_K becomes extremely small at earlier times. For instance, at the time of a hypothetical GUT phase transition with $T \sim 10^{16}$GeV

$$|\Omega_K(t_{\rm GUT})| < 10^{-56}, \tag{22}$$

or

$$\frac{\sum_i \rho_i}{\rho_{\rm cr}} = 1 + \mathcal{O}(10^{-56}). \tag{23}$$

This looks like an extreme fine-tuning of the initial condition.

Horizon problem. On conformal diagram it is clear that the region at time t_1 that can affect an observer at time $t_2 > t_1$ has the following comoving radius

$$\Delta\chi(t_2, t_1) = \tau_2 - \tau_1 = \int_{t_1}^{t_2} \frac{dt}{a} = \int_{a_1}^{a_2} \frac{da}{Ha^2}. \tag{24}$$

Assuming dominance of a single component, this becomes

$$\Delta\chi(t_2, t_1) = \frac{2}{H_0(1+3w)} (a_2^{(1+3w)/2} - a_1^{(1+3w)/2}). \tag{25}$$

If $w > -1/3$, and at long time-separation, such that $a_2 \gg a_1$, we can neglect the second term. As a result, we find that

$$\Delta\chi(t_{\rm rec}, t_{\rm BB}) \ll \Delta\chi(t_0, t_{\rm rec}), \tag{26}$$

by a factor of about $\sqrt{1/a_{\rm rec}}$. This implies that there are about 1000 patches on the last scattering surface (the section of our past lightcone at recombination time) that didn't have a chance to communicate since the big bang singularity, but they have the same temperature with precision of one part in 10^4.

The above two puzzles (and also the monopole problem) are the standard motivations given for introducing a new phase called inflation. However, I should emphasize that it's a logical possibility that the initial condition after a big bang singularity satisfies the above constraints. It just looks extremely fine-tuned. Inflation is a simple dynamical scenario to produce this initial condition.

3 Inflation

The flatness and horizon problems are both the consequence of the fact that the **comoving Hubble** parameter

$$\mathcal{H} \equiv Ha = \dot{a}, \tag{27}$$

decreases with time. Namely that the universe undergoes a decelerated expansion when $w > -1/3$. A long enough (or fast enough) period of acceleration in the past would solve the problems. $\ddot{a} > 0$ implies

$$H^2 + \dot{H} > 0 \Rightarrow -\dot{H} < H^2. \tag{28}$$

There are two options:

- $\dot{H} < 0$. This is called *inflation*.
- $\dot{H} > 0$. This could happen during a *bouncing cosmology*.

The second option requires violation of the **Null Energy Condition (NEC)**. This is the requirement that

$$T_{\mu\nu} k^\mu k^\nu \geq 0, \qquad \forall k^\mu, k_\mu k^\mu = 0, \tag{29}$$

which can be thought of as the positivity of the energy density measured by ultra-relativistic observers. With symmetries of FLRW, NEC implies

$$\rho + p \geq 0 \to w \geq -1. \tag{30}$$

Taking the derivative of the Friedmann equation with $K = 0$ (which is a good approximation in our universe) gives

$$\dot{H} = -\frac{3}{2} H^2 (1 + w) < 0. \tag{31}$$

So NEC eliminates the second option above. While quantum effects can violate NEC (Casimir energy is an example), there is a version of NEC, called achronal average NEC that is satisfied by all UV complete QFTs that we know. The achronal average NEC still forbids a bounce, but keep in mind that there is no proof of it when gravity is dynamical. Nevertheless, we will assume NEC is satisfied and proceed with inflation, a period of accelerated expansion with

$$-1 < w < -\frac{1}{3} \tag{32}$$

during which the comoving horizon shrinks, $\mathcal{H}^{-1} \propto a^{(1+3w)/2}$.

1. How long need inflation last to solve flatness problem? Suppose $w \approx -1$ during inflation, and transition to radiation domination at $T_{\rm GUT}$.

4 How to drive inflation

A simple way to drive inflation is with a scalar field

$$S = \int d^4x\sqrt{-g}\left[-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right]. \tag{33}$$

The energy-momentum tensor is

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi + V(\phi)\right]. \tag{34}$$

If we take the metric to be FLRW and ϕ to be homogeneous, we find

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi), \qquad p = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{35}$$

Suppose $V(\phi)$ has a minimum at ϕ_0, with $V(\phi_0) > 0$. Then there is a classical solution in which $\phi = \phi_0$ so

$$\dot{\phi} = 0 \Rightarrow w = -1. \tag{36}$$

The resulting geometry is de Sitter[2]

$$ds^2 = -dt^2 + e^{2Ht}dx^2. \tag{37}$$

So far the scalar field plays no role. The minimum of the scalar potential is equivalent to a cosmological constant. However, we want inflation to end and create the hot phase of cosmic evolution. This could happen if the minimum at ϕ_0 is a false minimum and there is a lower minimum at say $\phi = 0$, with $V(0)$ the same as the cosmological constant today. Then bubbles of true vacuum would constantly be formed via quantum mechanical tunneling. This scenario was one of the earliest proposals to realize inflation and is called **false vacuum inflation**. However, the explicit tunneling solution found by Coleman and de Luccia has large positive Ω_K. Therefore, it does not accomplish the goal.

A successful model can be constructed if instead of being stuck at a false minimum, ϕ slowly rolls down a sufficiently flat potential so that

$$\dot{\phi}^2 \ll V \Rightarrow w \approx -1. \tag{38}$$

An elegant way to obtain this for a large number of e-folds is to consider a model that has a **slow-roll** attractor, where

$$H^2 \approx \frac{V}{3M_{\rm pl}^2}, \qquad \dot{\phi} \approx -\frac{V'}{3H}, \tag{39}$$

where we define $M_{\rm pl}^2 \equiv 1/8\pi G$. The validity of this approximation requires $\dot{\phi}^2 \ll V$ and $|\ddot{\phi}| \ll H|\dot{\phi}|$,

[2]This is a particular parametrization of de Sitter. It covers part of the manifold.

which translate into the smallness of the following slow-roll parameters to be small

$$\epsilon = -\frac{\dot{H}}{H^2} \approx \frac{M_{\rm pl}^2 V'^2}{2V^2}, \qquad \eta = \frac{M_{\rm pl}^2 V''}{V}. \tag{40}$$

When $\epsilon, |\eta| \ll 1$, different initial conditions quickly approach the slow-roll solution. This is because if $|\dot{\phi}| \gg |V'|/3H$ then

$$\ddot{\phi} \approx -3H\dot{\phi} \Rightarrow \dot{\phi} \propto a^{-3}, \tag{41}$$

that is, $|\dot{\phi}|$ decays, and if $|\dot{\phi}| \ll |V'|/3H$ then

$$\ddot{\phi} \approx -V' \Rightarrow \dot{\phi} \sim -V't. \tag{42}$$

Hence if V' does not change too fast, $\dot{\phi}$ approaches $-V'/3H$ in a time-scale of order H. The inflaton acts as a clock along the attractor.

4.1 Example: $m^2\phi^2$ inflation

As a concrete example consider

$$V = \frac{1}{2}m^2\phi^2. \tag{43}$$

The slow-roll parameters are

$$\epsilon = \eta = \frac{2M_{\rm pl}^2}{\phi^2}. \tag{44}$$

Therefore, the slow-roll conditions are satisfied as long as $\phi \gg M_{\rm pl}$. You can check the same condition (up to numerical coefficients) apply to any power-law potential. Such models of inflation are called **large field** inflation. Note that this is still within the regime of validity of Einstein theory as long as

$$V \ll M_{\rm pl}^4 \Rightarrow \frac{m^2}{M_{\rm pl}^2} \ll \frac{M_{\rm pl}^2}{\phi^2}. \tag{45}$$

Nevertheless, having a super-Planckian field range is not a necessary condition. For instance, the potential

$$V(\phi) = \mu^4 \tanh^2 \frac{\phi}{f} \tag{46}$$

can realize inflation when $\phi \gg f$. And f could be well below $M_{\rm pl}$. However, large field models are simpler in that one doesn't need to introduce the extra scale f.

4.2 Number of e-folds

Inflation has to last long enough to solve the puzzles of big bang cosmology. This number depends on the efficiency and the energy scale of the transition to the radiation dominated phase. Normally, we need about 60 e-folds. This in turn determines the minimum initial value ϕ_i, given a potential.

We can use the fact that ϕ acts as a clock along the slow-roll solution to write

$$\begin{aligned} N_e &= \int_{t_i}^{t_f} H dt = \int_{\phi_i}^{\phi_f} \frac{H d\phi}{\dot{\phi}} \\ &\approx -\int_{\phi_i}^{\phi_f} \frac{3H^2 d\phi}{V'} \approx -\frac{1}{M_{\rm pl}^2} \int_{\phi_i}^{\phi_f} \frac{V d\phi}{V'} . \end{aligned} \tag{47}$$

In the $m^2\phi^2$ model, we get

$$N_e = \frac{1}{4}\left(\frac{\phi_i^2}{M_{\rm pl}^2} - 1\right), \tag{48}$$

which means $\phi_i \approx 15 M_{\rm pl}$ for $N_e = 60$.

1. Consider the false-vacuum inflation with metric (37) and suppose it continues for an infinite number of e-folds in the past. Compute the affine length λ of a null geodesic between $t = -\infty$ to $t = 0$, assuming $dt/d\lambda = 1$ at $t = 0$. Is λ finite? Can you justify your answer using the global coordinates of dS? How would the result change in the case of $m^2\phi^2$ inflation?

5 Reheating

Inflation ends when the slow-roll conditions are no longer satisfied. At this point the energy in the inflaton has to be transferred to the standard model (or the BSM theory in which it is embedded). This is called **preheating**. Let us focus on the $m^2\phi^2$ model. This is a reasonable assumption because even if $V(\phi)$ is significantly different during the inflation, near the bottom the mass term is generically the most relevant. Once $\phi \ll M_{\rm pl}$, we have

$$H^2 \sim \frac{m^2\phi^2}{M_{\rm pl}^2} \ll m^2. \tag{49}$$

In this regime, the ϕ equation of motion

$$\ddot{\phi} + 3H\dot{\phi} + m^2\phi = 0 \tag{50}$$

can be solved using the WKB method. With a judicious choice of $t = 0$, we can write $\phi(t) = \phi_0(t)\cos(mt)$, with $\dot{\phi}_0/\phi_0 = \mathcal{O}(H)$. Then we find

$$\phi(t) \approx \frac{c}{a^{3/2}}\cos(mt), \tag{51}$$

which results in

$$\rho = \frac{c^2m^2}{2a^3}, \qquad p = \frac{c^2m^2}{2a^3}\cos(2mt). \tag{52}$$

Hence, the energy density decays as non-relativistic matter, while pressure averages to zero. This suggests identifying the oscillating scalar field with a density of ϕ particles that are diluting with the expansion of the universe.

1. Why doesn't the argument around equation (42) apply here?

We need these to decay or annihilate into relativistic particles. Let us denote them by χ (bosonic for simplicity, and a reason which becomes clear shortly), and suppose there are interactions $g\phi\chi^2$ and $\lambda\phi^2\chi^2$ responsible for this. As a simple example suppose the interactions lead to the depletion of ϕ particles at rate Γ

$$\dot{\rho}_\phi + 3H\rho_\phi = -\Gamma\rho_\phi \Rightarrow \rho_\phi = \frac{\rho_i}{a^3}e^{-\Gamma t}. \tag{53}$$

Conservation of energy implies

$$\dot{\rho}_\chi + 4H\rho_\chi = \Gamma\rho_\phi \Rightarrow \rho_\chi = \frac{\Gamma\rho_i}{a^4}\int_0^t dt' a(t')e^{-\Gamma t'}. \tag{54}$$

If $\Gamma \gg H$, then we have *instantaneous preheating* where the radiation energy density reaches a maximum $\rho_{\chi,\rm max} \sim \rho_i$ very fast. If $\Gamma \ll H$, the maximum ρ_χ will be $\mathcal{O}(\Gamma/H)\rho_i$.

It is reasonable to ask if efficient preheating is compatible with successful inflation. The latter requires a sufficiently flat potential. Its radiative stability limits the strength of the couplings we introduced above

$$g < \frac{m^2 \phi_f}{\Lambda^2}, \qquad \lambda < \frac{m^2}{\Lambda^2} \tag{55}$$

where ϕ_f is the field excursion at the end of inflation and Λ is the UV cutoff. It does not have to be as large as $M_{\rm pl}$, but the most minimal and conservative choice would be $\Lambda \sim M_{\rm pl}$. Let's focus on $g\phi\chi^2$ interaction, which at tree-level gives

$$\Gamma_{\rm tree} \sim \frac{g^2}{m} < \frac{m^3}{M_{\rm pl}^2}. \tag{56}$$

This is much smaller than H at the end of inflation. However, as m/H increases with the expansion of the universe, a collective effect called *narrow parametric resonance* greatly enhances the effective decay rate. The momentum modes of χ satisfy

$$\ddot{\chi}_{\boldsymbol{k}} + 3H\dot{\chi}_{\boldsymbol{k}} + \left(\frac{k^2}{a^2} + g\phi_0 \cos(mt)\right)\chi_{\boldsymbol{k}} = 0. \tag{57}$$

In the regime $g\phi_0 \ll m^2$, required for (conservative) radiative stability, and $m \gg H$ this equation has narrow instability bands at

$$k_n \sim a\frac{nm}{2}, \quad n \in \{1, 2, 3, \cdots\}, \qquad \Delta k \sim a\frac{g\phi_0}{m}. \tag{58}$$

These have the interpretation of production of pairs of χ particles via the decay of n inflatons. The decay is Bose-enhanced, and therefore within the bands the modefunctions grow exponentially with a rate[3]

$$\Gamma_{\rm resonance} \sim \frac{g\phi_0}{m}. \tag{59}$$

While radiative stability requires $\Gamma_{\rm resonance} < m$, unlike $\Gamma_{\rm tree}$, the suppression is not parametric. Therefore, instantaneous preheating is conceivable.

Once energy has been transferred to the standard model, its gauge interactions are responsible for thermalization, i.e. a complete **reheating**. To estimate the rate, we use

$$\Gamma_{\rm SM} \sim n\sigma v. \tag{60}$$

Assuming instantaneous production of lots of relativistic particles with energy $\sim m$, we have

$$n \sim \frac{M_{\rm pl}^2 H^2}{m}, \qquad \sigma \sim \frac{\alpha^2}{m^2}, \qquad v \sim 1, \tag{61}$$

where $\alpha = \frac{g_{\rm SM}^2}{4\pi} \sim 0.1$. Requiring $\Gamma_{\rm SM} > H$, and denoting by T the would-be temperature for a successful reheating (namely $T \sim \sqrt{M_{\rm pl} H}$), imply

$$m^3 < \alpha^2 M_{\rm pl} T^2. \tag{62}$$

[3]See Mukhanov's book and 1907.04402 for more details.

On the other hand, we know that at the end of inflation $\rho < M_{\rm pl}^2 m^2$, hence $m > T^2/M_{\rm pl}$. Combined with the above bound, this gives

$$T < \alpha^{1/2} M_{\rm pl}. \tag{63}$$

This condition is safely satisfied for temperatures at which the SM remains in thermal equilibrium, i.e. $T < \alpha^2 M_{\rm pl}$.

6 Perturbations; a first look

As a warm-up, we will consider a massless scalar field on a spatially flat FLRW background. The action is

$$S = -\frac{1}{2}\int d^4x\sqrt{-g}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi. \tag{64}$$

Because of translation invariance, the field can be expanded in momentum space

$$\varphi(\tau, \boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\varphi_{\boldsymbol{k}}(\tau), \tag{65}$$

and different $\boldsymbol{k}$ modes evolve independently in the non-interacting theory:

$$\varphi''_{\boldsymbol{k}} + 2\mathcal{H}\varphi'_{\boldsymbol{k}} + k^2\varphi_{\boldsymbol{k}} = 0. \tag{66}$$

Here prime denotes $d/d\tau$. There are two asymptotic regimes:

1. $k \gg \mathcal{H}$, called **subhorizon**. This is when the physical momentum is much bigger than the Hubble scale (which normally coincides with curvature scale), $k/a \gg H$. In this regime, the equation can be solved using the WKB method. We find two oscillating solutions that are conjugate to each other:
$$\varphi_{\boldsymbol{k}}(\tau) = \frac{c_-}{a}e^{ik\tau} + \frac{c_+}{a}e^{-ik\tau}. \tag{67}$$
2. $k \ll \mathcal{H}$, called **superhorizon**. In this regime the solutions are decaying or growing
$$\varphi_{\boldsymbol{k}} = c_g[1 + \mathcal{O}(k^2/\mathcal{H}^2)] + c_d\int \frac{d\tau}{a^2}[1 + \mathcal{O}(k^2/\mathcal{H}^2)]. \tag{68}$$

For the same reason that inflation solves the puzzles of hot big bang cosmology, namely that the comoving Hubble length shrinks rather than expanding, perturbations during inflation behave in a qualitatively different way. They start their lives in the subhorizon regime and stretch to the superhorizon regime. This is commonly called *exiting* the horizon. In the decelerated phase the superhorizon perturbations *reenter* the horizon and start oscillating. This has a very profound implication: apart from solving the old puzzles, inflation turns out to be a remarkably simple quantum mechanical theory of initial condition for the observed cosmological perturbations.

If slow-roll parameters are small, the metric during inflation is approximately de Sitter (37), for which

$$\mathcal{H} = -\frac{1}{\tau}. \tag{69}$$

In this approximation, the exact solutions to (66) are

$$f_k(\tau) = A_k(1 + ik\tau)e^{-ik\tau}, \tag{70}$$

and its complex conjugate. A_k is a constant, whose role becomes clear shortly.

We now quantize the theory. If this is your first exposure to the subject, this transition from classical to quantum field theory might sound a bit bizarre. After all, observed cosmological perturbations are as classical as anything could be. We are now going to see that quantum zero point oscillations of the subhorizon modes turn into classical fluctuations of the superhorizon modes. So we expand the free field

$$\varphi(\tau, \boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}(a_{\boldsymbol{k}} f_k(\tau) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} + a^\dagger_{\boldsymbol{k}} f^*_k(\tau) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}) \tag{71}$$

where $a_{\boldsymbol{k}}$ and $a^\dagger_{\boldsymbol{k}}$ are the annihilation and creation operators, satisfying

$$[a_{\boldsymbol{k}}, a_{\boldsymbol{k}'}] = (2\pi)^3 \delta^3(\boldsymbol{k} - \boldsymbol{k}'). \tag{72}$$

The normalization A_k is determined by requiring that φ and its conjugate momentum

$$\Pi = \frac{\delta S}{\delta \varphi'} = \frac{1}{H^2\tau^2}\varphi' \tag{73}$$

satisfy the canonical commutation relation:

$$[\varphi(\tau, \boldsymbol{x}), \Pi(\tau, \boldsymbol{x}')] = i(2\pi)^3\delta^3(\boldsymbol{x} - \boldsymbol{x}'). \tag{74}$$

This fixes

$$A_k = \frac{H}{\sqrt{2k^3}}. \tag{75}$$

What is the state of the field? Strictly speaking, this is unknown and can be anything. However, there is a very reasonable choice that high energy modes (i.e. subhorizon modes that are oscillating fast) should not be excited. This is called the **adiabatic vacuum**, which in the case of de Sitter is also known as the **Bunch-Davies** or **Hartle-Hawking** vacuum.

In the adiabatic vacuum the 2-point correlation function of φ is

$$\langle \varphi_{\boldsymbol{k}}(\tau)\varphi_{\boldsymbol{k}'}(\tau)\rangle = (2\pi)^3\delta^3(\boldsymbol{k} + \boldsymbol{k}')|f_k(\tau)|^2 \underset{|k\tau|\ll 1}{\rightarrow} (2\pi)^3\delta^3(\boldsymbol{k} + \boldsymbol{k}')\frac{H^2}{2k^3}(1 + \mathcal{O}(k^2\tau^2)). \tag{76}$$

Recall that $\tau = -1/Ha$. So for a mode that exits the horizon N_e e-folds before the end of inflation, the $k^2\tau^2$ correction at the end of inflation is $\sim e^{-2N_e}$.

On the other hand, the commutator of φ and $\dot{\varphi} = \Pi/a^3$ is given by

$$[\varphi_{\boldsymbol{k}}(t), \dot{\varphi}_{\boldsymbol{k}'}(t)] = (2\pi)^3\delta^3(\boldsymbol{k} + \boldsymbol{k}')\frac{1}{a^3}. \tag{77}$$

The smallness of this compared to $H\langle\varphi_{\boldsymbol{k}}\varphi_{\boldsymbol{k}'}\rangle$ at superhorizon scales is the reason why those perturbations are practically classical. A useful analogy is the position of a heavy particle X, for which

$$[X, \dot{X}] = \frac{i}{m}. \tag{78}$$

For any finite resolution Δx of the measuring apparatus, we can talk about classical trajectories as long as $t \ll m\Delta x^2$, which becomes infinite as $m \to \infty$. Of course, at some point interactions with the environment will also lead to decoherence. However, we do not need that to justify a classical treatment of superhorizon fluctuations.

The slow-roll condition $|\eta| \ll 1$ implies that the inflaton field is an approximately massless field, and hence its fluctuations behave similar to the φ field, studied above. An intuitive way to think about these fluctuations is that during inflation at every time-step of $\sim 1/H$, at every patch of size $L > 1/H$, φ randomly jumps by $\sim H$. Namely, if we define

$$\varphi_L(t) = \frac{1}{V_L} \int_{|x|<L} d^3\boldsymbol{x}\varphi(t, \boldsymbol{x}), \tag{79}$$

then

$$\left\langle \varphi_L^2(t) \right\rangle \sim H^2 \int_0^{a(t)/L} \frac{dk}{k}, \tag{80}$$

which increase by $\mathcal{O}(H^2)$ every Hubble-time.

7 Perturbations; in more detail

A careful analysis of cosmological perturbations has to take into account the metric fluctuations as well as those of other fields, and properly deal with reparametrization invariance. In the context of inflation, the correct analysis was first done by Mukhanov and Chibisov. Here I briefly review a later treatment due to Maldacena (astro-ph/0210603). First, we write the metric in the ADM notation

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt). \tag{81}$$

The Einstein-Hilbert action plus that of the inflaton reads

$$S = \frac{1}{2}\int dt d^3x \sqrt{h}\left[M_{\rm pl}^2 N R^{(3)} + M_{\rm pl}^2 N^{-1}(E_{ij}E^{ij} - E^2) + N^{-1}(\dot\phi - N^i\partial_i\phi)^2 - N h^{ij}\partial_i\phi\partial_j\phi - 2NV(\phi)\right], \tag{82}$$

where E_{ij} is N times the extrinsic curvature of $t =$ constant hypersurfaces

$$E_{ij} = \frac{1}{2}(\dot h_{ij} + \nabla_i N_j + \nabla_j N_i), \tag{83}$$

spatial indices are raised by h^{ij}, the inverse of h_{ij}, and ∇_i is the covariant derivative of this metric. Temporal diffeomorphisms can be fixed by requiring ϕ to be constant on the constant t hypersurfaces:

$$\phi(t, x) = \bar\phi(t). \tag{84}$$

This is possible because on the background solution ϕ is rolling, so $\dot{\bar\phi} \neq 0$.

Next decompose

$$h_{ij} = a^2 e^{2\zeta}(\delta_{ij} + \gamma_{ij} + \mathcal{O}(\gamma^2)), \tag{85}$$

where γ_{ij} is traceless, $\gamma_{ii} = 0$. Spatial diffs with nonzero momentum are all fixed if we impose

$$\partial_i\gamma_{ij} = 0. \tag{86}$$

Note that a full gauge fixing at nonzero $\boldsymbol{k}$ is possible via a Lorentz breaking gauge-fixing (like Coulomb gauge in EM). On a Lorentz invariant background, this often complicates the perturbative calculation. Here, a Lorentz breaking gauge fixing is natural because the background doesn't respect it anyway.

The gauge-fixed action for the fluctuations $S[\zeta, \gamma_{ij}, N, N^i]$ has no time-derivatives acting on N and N^i. They are constraint variables and can be integrated out by solving them (perturbatively in terms of ζ, γ_{ij}) using their equations of motion and substituting back. The result is

$$S = M_{\rm pl}^2 \int dt d^3x a^3\left[\epsilon(\dot\zeta^2 - a^{-2}(\partial_i\zeta)^2) + \frac{1}{8}(\dot\gamma_{ij}^2 - a^{-2}(\partial_i\gamma_{jk})^2)\right] + \cdots \tag{87}$$

where repeated spatial indices are contracted with δ^{ij} (any dependence on ζ and γ_{ij} has to be explicit when doing perturbation theory). We see that the quadratic part of the action has the same form as that of massless scalar fields, except for different normalizations. Dots represent interaction terms. Maldacena shows that not only the quadratic terms but all higher order terms contain two derivatives acting on two ζ or γ. As a result, constant ζ and γ would be a valid super-horizon solution even at nonlinear order.

The transverse-tracelessness of γ_{ij} means that every momentum component $\gamma_{ij}(\boldsymbol{k})$ has two polarizations:

$$\gamma_{ij}(\boldsymbol{k}) = \sum_{r=1,2} \gamma^r_{\boldsymbol{k}} \varepsilon^r_{ij}(\hat{k}), \qquad \varepsilon^r_{ii} = 0 = k_i \varepsilon^r_{ij}. \tag{88}$$

We choose their normalization to be

$$\varepsilon^r_{ij}\varepsilon^s_{ij} = \delta^{r,s}. \tag{89}$$

Let's for now ignore nonlinearities in (87). Then had it not been for the fact that ϵ and H are slowly varying in time, we could have quantized ζ and γ^r as in the previous section, and calculated their 2-point correlation functions. We can still do that, because every k mode is oscillating fast well before horizon-crossing and therefore remains in the adiabatic vacuum regardless of the slow variation of H and ϵ. Well after the horizon crossing, it approaches a constant and again unaffected by the variation of H and ϵ. The transition period between these two phases takes a time $\sim 1/H$, over which ϵ and H do not change appreciably because of the slow-roll conditions. It follows that in the limit $\boldsymbol{k}|\tau| \ll 1$, we can write

$$P_s(k) \equiv \langle \zeta_{\boldsymbol{k}} \zeta_{-\boldsymbol{k}} \rangle' = \frac{H_*^2}{4M_{\rm pl}^2 \epsilon_* k^3}, \tag{90}$$

and

$$P_t(k) \equiv \left\langle \gamma^r_{\boldsymbol{k}} \gamma^s_{-\boldsymbol{k}} \right\rangle' = \delta^{r,s} \frac{2H_*^2}{M_{\rm pl}^2 k^3}, \tag{91}$$

where prime on a correlator means that we have dropped the momentum conserving delta function, and H_* and ϵ_* are the values of these quantities at the horizon crossing time $k\tau_* = -1$.

$P_s(k)$ and $P_t(k)$ are called scalar and tensor **power spectra**. In the minimal inflationary scenario that we considered, they are the origin of all cosmological perturbations that we observe today. Evidently $P_t \ll P_s$, and we have not yet detected any primordial tensor perturbations, while we know $k^3 P_s(k) \simeq 4.1 \times 10^{-8}$. It is common to define **tensor-to-scalar ratio**

$$r \equiv \frac{2P_t(k)}{P_s(k)} = 16\epsilon. \tag{92}$$

The current upper-limit on r is about 10^{-3}.

A power-spectrum that is $\propto 1/k^3$ is often called scale-invariant. This is because in this case every logarithmic interval of k contributes the same to the real space correlation function,

$$\langle \zeta(\boldsymbol{r})\zeta(0) \rangle = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P_s(k) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{93}$$

P_s and P_t are not scale-invariant, because the factors of H_* and ϵ_* introduce nontrivial k-dependence. This deviation from scale-invariance is called tilt:

$$n_s - 1 \equiv \frac{d}{d\log k}\log(k^3 P_s(k)), \qquad n_t \equiv \frac{d}{d\log k}\log(k^3 P_t(k)). \tag{94}$$

Note that these are slow-roll suppressed. We know from observations that $n_s \approx 0.97$.

1. What should be m in $m^2\phi^2$ inflation, to predict the right scalar power? Suppose we observe modes that exit the horizon 60 e-folds before the end of inflation.

2. Calculate r, n_s and n_t in $m^2\phi^2$ inflation. Is it a viable inflationary scenario?

   Hint: Derivatives with respect to k can be related to time-derivatives using $k\tau_* = -1$.

It would be useful to relate the fluctuations of ζ to those of the inflaton $\varphi = \phi - \bar{\phi}$. We gauge-fixed $\varphi = 0$, to define the gauge-invariant ζ. If we instead set $\zeta = 0$, then φ would be a canonically normalized, approximately massless scalar field. Every momentum mode of φ has fluctuations of order H around the horizon crossing time, and then remains approximately constant afterward. We can try to return to the ζ gauge by a time-diff:

$$\xi^0 = -\frac{\varphi}{\dot{\bar{\phi}}}, \tag{95}$$

which would imply

$$\zeta = -\frac{H\varphi}{\dot{\bar{\phi}}}, \tag{96}$$

or

$$\langle \zeta_{\boldsymbol{k}}(\tau)\zeta_{-\boldsymbol{k}}(\tau)\rangle' = \frac{H^2}{\dot{\bar{\phi}}^2}\langle \varphi_{\boldsymbol{k}}(\tau)\varphi_{-\boldsymbol{k}}(\tau)\rangle' . \tag{97}$$

The correlator on the left is constant at superhorizon scales (up to $k^2\tau^2$ corrections), while the one of φ is approximately constant (up to slow-roll effects). φ has a mass of $\mathcal{O}(\sqrt{\epsilon}H)$. Of course, H and $\dot{\bar{\phi}}$ are also time-dependent in such a way that the RHS remains constant. However, we can quickly get the right answer by evaluating it shortly after the horizon-crossing time. Then we can approximate φ 2-point function by (76) evaluated at horizon crossing time. This reproduces (90).

Now I can explain why the perturbations generated during single field inflation are called **adiabatic** perturbations. Consider a radiation dominated universe, where all constituents are in approximate thermal equilibrium and all chemical potentials are zero. Here the state of the universe is fully fixed by T. Superhorizon fluctuations of T, $T(t,x) = \bar{T}(t) + \delta T(t,x)$, are adiabatic fluctuations in the sense that different regions with different T are related by adiabatic expansion or contraction. During cosmic evolution all these regions follow the same thermal history, and since $\dot{\bar{T}} \neq 0$, we can choose time-slices such that in the new coordinates $T(\tilde{t},x) = \bar{T}(t)$:

$$\xi^0 = -\frac{\delta T}{\dot{\bar{T}}}. \tag{98}$$

In this gauge, perturbations are encoded in the nontrivial 3-geometry of the time slices.

The notion of adiabatic perturbations can be generalized beyond radiation dominated cosmology by defining them as perturbations that correspond to the same history for different superhorizon regions but slightly shifted relative to one another. During the slow-roll inflation, ϕ entirely determines the state of the universe and its subsequent evolution. Therefore, superhorizon fluctuations of ϕ correspond to adiabatic fluctuations. We can always absorb them in the perturbations of the 3-geometry by choosing appropriate time-slices. This is indeed what we do by fixing the ζ gauge $\phi = \bar{\phi}$. We have good evidence that cosmological perturbations we observe today are predominantly adiabatic.

8 Non-Gaussianity

A Gaussian distribution is fully specified by the mean and variance

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-x_0)^2/2\sigma^2}. \tag{99}$$

Higher order correlation functions are trivially determined via combinatorics:

$$\left\langle (x-x_0)^{2N} \right\rangle = \frac{(2N)!}{2^N N!} \sigma^{2N}. \tag{100}$$

If we could neglect the interactions, the same would hold for every momentum mode $\zeta_{\boldsymbol{k}}$ and $\gamma^r_{\boldsymbol{k}}$, and their power spectra would contain all information there is. The interactions make the theory non-Gaussian (both in the sense that late time correlators will be non-Gaussian, and more basically because the path integral is no longer Gaussian). Given a model of inflation, there is a prediction for these non-Gaussian correlators and detecting or constraining non-Gaussianity is one of our best tools for learning about the underlying dynamics that drove inflation. Given that $\langle \zeta_{\boldsymbol{k}} \rangle = \langle \gamma^r_{\boldsymbol{k}} \rangle = 0$ by translation invariance, the first place to look for non-Gaussianity is the 3-point correlation function. This would vanish in a Gaussian theory.

A simple model to characterize this effect is the **local non-Gaussianity**,

$$\zeta(\boldsymbol{x}) = g(\boldsymbol{x}) + \frac{1}{2} f^{\text{loc}}_{NL} g(\boldsymbol{x})^2, \tag{101}$$

where $g(\boldsymbol{x})$ is a Gaussian random variable

$$\langle g_{\boldsymbol{k}} g_{-\boldsymbol{k}} \rangle' = P_s(k) = \frac{A_s}{k^{4-n_s}}. \tag{102}$$

It is called local non-Gaussianity because the relation between ζ and g is a local relation. This could arise simply because the curvature perturbations are a nonlinear function of the field that is approximately free. In this model, the leading contribution to the 3-point function is

$$\langle \zeta_{\boldsymbol{k}_1} \zeta_{\boldsymbol{k}_2} \zeta_{\boldsymbol{k}_3} \rangle' = f^{\text{loc}}_{NL} [P_s(k_1)P_s(k_2) + P_s(k_1)P_s(k_3) + P_s(k_2)P_s(k_3)]. \tag{103}$$

Note that by momentum conservation the three momenta form a triangle. By rotation symmetry the orientation of the triangle does not matter. Therefore, the 3-point function only depends on the shape of the triangle, or equivalently on the moduli k_1, k_2, k_3. Cosmologists often call this function the **bispectrum**.

1. Consider an axion field that is light during inflation $H_i \ll f_a$. Sometime after inflation H becomes less than m_a and axion starts oscillating around the minimum of its potential, which is misaligned with respect to $\langle a \rangle$ by an angle θ_0. (a) Show that inflationary fluctuations of a correspond to $\delta\theta \sim H_i/f$. (b) Show that the relation between $\delta\rho/\rho$ and δa is local but nonlinear as in (101). Estimate the corresponding f^{loc}_{NL}.

We are going to see that the **squeezed limit**, where one of the momenta is much less than the others, is of special importance. Since $P(k_1) \gg P(k_2)$ when $k_1 \ll k_2$, in this limit the local shape becomes

$$\langle \zeta_{\boldsymbol{k}_1} \zeta_{\boldsymbol{k}_2} \zeta_{\boldsymbol{k}_3} \rangle' \to 2 f_{NL} P_s(k_1) P_s(k_2). \tag{104}$$

Let us now consider the sources of non-Gaussianity during inflation. Firstly, even if the inflaton is a free field ($m^2\phi^2$ model), gravity is interacting and hence as we have already seen the gauge-fixed action for ζ and γ contains interactions. In particular, there are cubic interactions that lead to a nonzero bispectrum. This was computed by Maldacena. The method used is called in-in perturbation theory. It is nicely explained in the appendix of Weinberg's paper hep-th/0506236. As in the scattering problem one goes to the interaction picture and introduces the Fock states for the free interaction picture field (labeled by I). The states are then evolved by the interaction Hamiltonian written in terms of the interaction picture fields. The difference with scattering is that we are interested in equal-time expectation values in the Bunch-Davies vacuum. Hence, there is a forward time-evolution, insertion of operators and then backward evolution

$$\langle O(t) \rangle = \frac{\left\langle 0_I | \bar{T} \exp(i \int_{-\infty(1-i\epsilon)}^{t} dt' H_I(t')) O_I(t) T \exp(-i \int_{-\infty(1+i\epsilon)}^{t} dt' H_I(t')) | 0_I \right\rangle}{\left\langle 0_I | \bar{T} \exp(i \int_{-\infty(1-i\epsilon)}^{t} dt' H_I(t')) T \exp(-i \int_{-\infty(1+i\epsilon)}^{t} dt' H_I(t')) | 0_I \right\rangle} \tag{105}$$

Here $|0_I\rangle$ is the state that is the interaction picture vacuum (i.e. $a_I|0_I\rangle = 0$). The $i\epsilon$ prescription projects it onto the true vacuum. For us $O_I(t) = \zeta^I_{\boldsymbol{k}_1}(t)\zeta^I_{\boldsymbol{k}_2}(t)\zeta^I_{\boldsymbol{k}_3}(t)$ in the limit $t \to \infty$.

2. This 3-point correlator can be thought of as coming from the interaction of four ϕ via a graviton exchange and with one of the external ϕ fields evaluated on the background $\dot{\phi}$. Show that this gives an estimate $f_{NL} = \mathcal{O}(\epsilon)$.

3. Consider a general slow-roll potential with $V'''(\bar{\phi}) \neq 0$. Estimate the contribution of this interaction to f_{NL}. Is this contribution included in Maldacena's result?

The inflaton field does not have to remain a weakly coupled degree of freedom all the way to the Planck scale. We could imagine a theory with higher derivative self-interactions like

$$\mathcal{L}_{\text{int}} = \frac{1}{\Lambda^4} (\partial\phi)^4 + \cdots \tag{106}$$

Since on the background $\dot{\phi} \neq 0$, for this interaction to be a perturbation we need

$$\Lambda^4 > \dot{\phi}^2. \tag{107}$$

(Sometimes it is consistent to consider larger but slowly varying $\dot{\phi}$ because the dots in (106) have two or more derivatives of ϕ.)

We can estimate the contribution of this interaction to the bispectrum as follows. If we normalize the bispectrum by the variance, it should be of the order of the interaction Lagrangian divided by the free one at scale H

$$\frac{\langle\zeta^3\rangle}{\langle\zeta^2\rangle^{3/2}} \sim \frac{\mathcal{L}_3}{\mathcal{L}_2}. \tag{108}$$

The LHS is an estimate of $f_{NL}\zeta_{\rm rms}$. Decomposing $\phi = \bar{\phi} + \varphi$, the cubic interactions that arise from (106) are

$$4\frac{\dot{\bar{\phi}}}{\Lambda^4}(\dot{\varphi}^3 + a^{-2}\dot{\varphi}(\partial_i\varphi)^2). \tag{109}$$

We can estimate $\varphi \sim H$ and $\partial_t \sim a^{-1}\partial_x \sim H$, to obtain

$$f_{NL} \sim \frac{\dot{\phi}H^2}{\Lambda^4\zeta_{\rm rms}} \sim \frac{\dot{\bar{\phi}}^2}{\Lambda^4}, \tag{110}$$

where we used $\zeta_{\rm rms} \sim H^2/\dot{\bar{\phi}}$. This can be $\mathcal{O}(1)$ if the bound (107) is saturated. Therefore, non-Gaussianity can be much bigger than in the case of minimal slow-roll model. A framework to systematically study non-Gaussianity in single field inflation is the **EFT of Inflation** hep-th/0709.0293.

However, the squeezed limit of the bispectrum is universal in single-field inflation, regardless of the presence of derivative interactions. In this limit, the dominant contribution to the bispectrum comes from the evolution of two short wavelength modes on the background of a long wavelength one. A long-wavelength ζ perturbation changes the metric to

$$ds^2 = -dt^2 + a^2 e^{2\zeta_L} dx^2. \tag{111}$$

Therefore, locally it is identical to a rescaling

$$\tilde{x} = e^{\zeta_L} x. \tag{112}$$

We can now evaluate

$$\begin{aligned}
\langle\zeta_{\boldsymbol{k}}\zeta_{-\boldsymbol{k}}\rangle'_{\zeta_L} &= \int d^3\boldsymbol{x}\,\langle\zeta(\boldsymbol{x})\zeta(0)\rangle_{\zeta_L} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \\
&= \int d^3\boldsymbol{x}\,\langle\zeta(\tilde{\boldsymbol{x}})\zeta(0)\rangle\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \\
&= \int d^3\tilde{\boldsymbol{x}}(1 - 3\zeta_L - i\zeta_L\boldsymbol{k}\cdot\boldsymbol{x}\,\langle\zeta(\tilde{\boldsymbol{x}})\zeta(0)\rangle\, e^{i\boldsymbol{k}\cdot\tilde{\boldsymbol{x}}} + \mathcal{O}(\zeta_L^2) \\
&= (1 - 3\zeta_L - \zeta_L\boldsymbol{k}\cdot\nabla_k)P_s(k) + \mathcal{O}(\zeta_L^2).
\end{aligned} \tag{113}$$

The part linear in ζ_L can be written as $-\zeta_L P_s(k) d\log(k^3 P_s(k))/d\log k = (1-n_s)P_s(k)\zeta_L$. Switching $k \to k_2$ and correlating with $\zeta_{\boldsymbol{k}_1}$ with $k_1 \ll k_2$ gives

$$\langle\zeta_{\boldsymbol{k}_1}\zeta_{\boldsymbol{k}_2}\zeta_{\boldsymbol{k}_3}\rangle'_{k_1\ll k_2} \approx (1-n_s)P_s(k_1)P_s(k_2). \tag{114}$$

This relation is called Maldacena's consistency condition. The smallness of $1 - n_s \approx 0.03$ makes this hard to check. The current limit from CMB observations by Planck collaboration is $|f_{NL}^{\rm loc}| < 5$.

Suppose now that there is another light field during inflation, say σ. And suppose in some way, σ affects the observed cosmological fluctuations. For instance, σ controls the efficiency of reheating. Then the above argument would not apply, because a superhorizon perturbation of σ is a locally observable quantity (unlike the superhorizon ζ, which is equivalent to a rescaling). The power spectrum of the short wavelength modes in the presence of σ_L could respond in different ways depending on the details of the Lagrangian. So the prediction for the squeezed limit of bispectrum in multifield models of inflation is not universal.

4. Check explicitly that the derivative interactions (109) cannot change the squeezed limit behavior.

9 Wavefunction of the Universe

Inflation is a successful theory of initial condition for the cosmological perturbations. However, it is not eternal to past. This is a consequence of the Penrose singularity theorem (see 1312.3956 for a nice review and generalization). It would be interesting to have a theory of initial condition for inflation. Of course, it could emerge from an initial singularity which might be impossible to understand without a full theory of quantum gravity. However, one can wonder about a possible semi-classical theory of initial condition, in the same way that inflation is a semi-classical theory of initial condition for the hot big bang cosmology.

Hartle and Hawking proposed a candidate. It was inspired by the idea that in quantum mechanics Euclidean time evolution can be used to construct the vacuum out of an arbitrary state. Indeed, by expanding in the Hamiltonian basis, we can write

$$|\psi(T)\rangle = \sum_n e^{-E_n T}|n\rangle\langle n|\psi_0\rangle. \tag{115}$$

In the limit $T \to \infty$, all other contributions to the sum will go to zero compared to the ground state. Therefore,

$$|0\rangle \propto \lim_{T\to\infty} e^{-HT}|\psi_0\rangle. \tag{116}$$

This is a trick that we always use in perturbative QFT computations (the $i\epsilon$ prescription) to project the free vacuum onto the interacting vacuum. Let us nevertheless apply it to the concrete problem of finding the ground state of the harmonic oscillator. We can implement time-evolution by the path integral. Euclidean time-evolution would be implemented by the Euclidean path integral

$$\langle x_f|0\rangle \propto \lim_{T\to\infty} \langle x_f|e^{-HT}|x_i = 0\rangle = \lim_{T\to\infty} \int_{x(-T)=0}^{x(0)=x_f} Dx e^{-S_E}, \tag{117}$$

where

$$S_E = \frac{1}{2}\int d\tau (x'^2 + \omega^2 x^2) \tag{118}$$

and we made the arbitrary choice that the initial state is the position eigenstate $|x = 0\rangle$. This is a Gaussian path integral, which can be evaluated by the saddle point approximation in two steps. The leading contribution is the classical action, $S[x_{\rm cl}]$, where taking $T \to \infty$

$$x_{\rm cl} = x_f e^{\omega\tau}. \tag{119}$$

Substituting this in the action, we find

$$\psi_0(x_f) = A_{1-\rm loop} e^{-\omega x_f^2/2} \tag{120}$$

which is the correct x_f dependence of the ground state wavefunction. Keeping T finite but large, the 1-loop factor gives the ground state energy $A_{1-\rm loop} \propto \exp(-\omega T/2)$ (see the appendix of "The uses of instantons" in Coleman's "Aspects of Symmetry").

Similarly, in QFT on Minkowski, we can "prepare" the Minkowski vacuum state using a path integral over Euclidean time in $(-\infty, 0]$. For this reason, we can often calculate vacuum correlators in a Lorentzian QFT, by an appropriate analytic continuation of the Euclidean correlators. Note that in QFT, we have a wavefunctional whose argument is a configuration of the fields (and t in the Schrödinger picture). For instance, for a single scalar field we write $\Psi[\phi(\boldsymbol{x}); t]$.

In quantum gravity, Ψ is still a functional of field configurations on a spatial slice, but there are some differences due to the reparametrization symmetry. Firstly, the Hamiltonian in GR is constrained to vanish:

$$S = \int \sum_a \Pi_a \phi^a - N\mathcal{H} - N^i \mathcal{P}_i \tag{121}$$

where ϕ^a is a collective notation for all dynamical degrees of freedom and Π_a their conjugate momenta, and N and N^i are the lapse and shift parameters in the ADM decomposition of the metric. They act as Lagrange multipliers. The wavefunction has to satisfy this constraint. Therefore,

$$i\partial_t \Psi = (N\mathcal{H} + N^i \mathcal{P}_i)\Psi = 0, \tag{122}$$

which is saying that Ψ cannot depend on t.

This is a general consequence of having time-reparametrization symmetry. As a simple example consider the point-particle action on Minkowski spacetime

$$S = -m \int d\tau \sqrt{-\eta_{\mu\nu} \frac{dX^\mu}{d\tau} \frac{dX^\nu}{d\tau}}. \tag{123}$$

The conjugate momenta are

$$p_\mu = m \frac{\eta_{\mu\nu} dX^\nu / d\tau}{\sqrt{-\eta_{\mu\nu} \frac{dX^\mu}{d\tau} \frac{dX^\nu}{d\tau}}}, \tag{124}$$

and the Hamiltonian

$$H = p_\mu \frac{dX^\mu}{d\tau} - L = 0. \tag{125}$$

At the quantum level, the vanishing of the Hamiltonian is imposed on the Wavefunction(al)

$$\mathcal{H}\Psi = 0. \tag{126}$$

In quantum gravity, this is called the Wheeler-De Witt equation. Suppose we only have the metric and one scalar field. Then the wavefunctional $\Psi[h_{ij}(\boldsymbol{x}), \phi(\boldsymbol{x})]$ depends on the ϕ configuration and the 3-geometry of the slice regardless of its particular parametrization.

WDW equation is a functional PDE, and often too hard to solve. It is usually studied either perturbatively or by looking for highly symmetric saddles of the gravitational path integral. As in quantum mechanics and QFT the final boundary condition of the path integral is fixed by the argument of the wavefunction. We are integrating over 4-geometries and $4d$ field configurations that match this condition. However, we can no longer talk about an infinite amount of Euclidean time evolution. Except for asymptotically flat or asymptotically AdS geometries, there is no absolute notion of time. A candidate wavefunction of the universe was proposed by Hartle and Hawking by imposing the condition that one only sums over smooth 4-geometries with no other boundary than the final one. These geometries are inevitably complex, since there is no non-singular Lorentzian manifold with just one boundary. Hence, there is some resemblance with the usual prescription.

Let us consider one application of this idea. Consider a pure Einstein theory with positive CC Λ. Suppose we ask for the wavefunction $\Psi(a)$ for a homogeneous closed FRW cosmology with scale factor a (which is physical because the universe is closed). Then for large enough a there is a classical solution (hence a saddle of the path integral) that matches this boundary, namely global dS

$$ds^2 = -d\tau^2 + \frac{1}{H^2}\cosh^2(H\tau)d\Omega_3^2. \tag{127}$$

To have a *no-boundary* saddle, we can rotate τ by 90 degrees into the complex plane at $\tau = 0$. The resulting Euclidean geometry is a hemisphere that smoothly ends at $\tau = i\pi/2$. The saddle point approximation to the path integral is

$$\Psi_{\rm HH}(a) \sim e^{iS_{\rm cl}}, \tag{128}$$

where $S_{\rm cl}$ is the action evaluated on the classical solution with the given boundary condition. Note that the action is in general complex (so the factor of i could be absorbed by working with the complexified Euclidean metric). When a is large the phase of the wavefunction changes rapidly; we are in the WKB regime. In this approximation, the amplitude of Ψ comes from the Euclidean part of the solution, namely the hemisphere. We find

$$|\Psi_{\rm HH}|^2 \sim \exp(\mathrm{Vol}_{S^4}\Lambda/H^4) = \exp\left(\frac{24\pi^2 M_{\rm pl}^4}{\Lambda}\right), \tag{129}$$

where we used the Friedmann equation $H^2 = \Lambda/3M_{\rm pl}^2$ and $\mathrm{Vol}_{S^4} = 8\pi^2/3$.

We can imagine that Λ is the value of a scalar potential at a False vacuum. The quantum tunneling from this False vacuum to a true vacuum was studied in the saddle point approximation by Coleman and De Luccia. The above formalism is useful because the tunneling rate can be approximated by the ratio

$$\Gamma \sim H^4 \frac{|\Psi_{\rm CDL}|^2}{|\Psi_{\rm HH}|^2} \sim H^4 e^B \tag{130}$$

where B is the difference between the Euclidean actions with and without the bubble of the true vacuum.

10 Stochastic Method

Consider a light scalar field ϕ with non-derivative interactions, e.g.

$$\mathcal{L} = \frac{1}{4}\lambda\phi^4. \tag{131}$$

As we will see, even for weak interactions, perturbative computation of inflationary correlation functions of ϕ are plagued by infrared divergences. The stochastic method was introduced by Starobinsky to solve this problem. For simplicity suppose the field is massless. At zeroth order in λ, a free massless field averaged over a region with a fixed physical size around a given point performs a random-walk motion:

$$\left\langle(\phi_L(t) - \phi_L(0))^2\right\rangle_{t\gg 1/H} \approx \frac{H^3}{4\pi^2}t. \tag{132}$$

To estimate the leading loop corrections to the 2-point function we isolate two ϕ's in (131) as external modes with momentum k, and let the others run in the loop. The dangerous contribution comes from the contribution of shorter modes that cross the horizon as time passes and lead to the linear growth (132). The time-evolution operator $T\exp\left(-i\int dtH_I\right)$ contains a time-integral over this contribution. Hence, it results in a perturbative expansion

$$\langle\phi_{\boldsymbol{k}}(t)\phi_{-\boldsymbol{k}}(t)\rangle' \approx \frac{H^2}{k^3}\sum_n c_n(\lambda H^2(t-t_k)^2)^n \qquad (t_k \equiv H^{-1}\log(k/H) \ll t) \tag{133}$$

which, even for $\lambda \ll 1$, breaks down if inflation lasts longer than

$$t_\lambda = \frac{1}{H\sqrt{\lambda}}. \tag{134}$$

This break-down (called *secular growth*) has led some to speculate about instability of inflation once the gravitational back-reaction of such a field is taken into account.

We can gain some insight by considering instead of $\lambda\phi^4$, a simple mass term $m^2\phi^2$ with $m \ll H$. This problem can be solved exactly. In particular, when $k \ll aH$

$$\ddot{\phi}_{\boldsymbol{k}} + 3H\dot{\phi}_{\boldsymbol{k}} + m^2\phi_{\boldsymbol{k}} \approx 0 \Rightarrow \phi_{\boldsymbol{k}}(t) \propto e^{-\Lambda H(t-t_k)} \tag{135}$$

where

$$\Lambda = \frac{3}{2} - \sqrt{\frac{9}{4} - \frac{m^2}{H^2}} \approx \frac{m^2}{3H^2} \ll 1. \tag{136}$$

As a result

$$\langle\phi_{\boldsymbol{k}}(t)\phi_{-\boldsymbol{k}}(t)\rangle' \sim \frac{H^2}{k^3}e^{-2\Lambda(t-t_k)}. \tag{137}$$

Imagine we were to include the effect of the mass term perturbatively, as in (133). We would have obtained a series in powers of $m^2(t-t_k)/H$, which breaks down after $t_* = H/m^2$. In this case, we know that the break-down has no dramatic consequence. It is simply saying that the free random-walk will saturate after the field reaches

$$\left\langle \phi_L^2 \right\rangle \sim \frac{H^4}{m^2}, \tag{138}$$

and the potential can no longer be treated as a perturbation.

One way to interpret this is using thermodynamics. Every inflationary observer is surrounded by a cosmological horizon and a temperature associated to it. This is best seen by using the coordinate system that covers the region causally accessible to the observer. In the limit of an exact dS,

$$ds^2 = -(1 - H^2 r^2)dt^2 + \frac{dr^2}{1 - H^2 r^2} + r^2 d\Omega^2. \tag{139}$$

This metric covers only a patch of dS, called the *static patch*. It is bounded by the future and past horizons at $r = 1/H$. Realistic inflationary models are not eternal to the past. A successful theory of initial condition for inflation will presumably removes the past horizon, in the same way that the white holes are removed when a black hole forms from the collapse. As in the case of Schwarzschild, the easiest way to identify the horizon temperature is to analytically continue $t \to -it_E$ and find the unique periodicity of t_E for which the horizon is regular. This gives $T = H/2\pi$. We can now give a thermal interpretation for the field excursion (138). It corresponds to a fluctuation of ϕ in a volume of the order of the static patch, i.e. H^{-3}, that would have a potential energy cost of order T. This argument suggests that a proper treatment of $\lambda\phi^4$ theory should result in the saturation of the variance when

$$\left\langle \phi_L^2 \right\rangle \sim H^3 t_\lambda = \frac{H^2}{\sqrt{\lambda}}. \tag{140}$$

The stochastic method achieves this by identifying $\varphi \equiv \phi_L$ as a slow degree of freedom with time-scale $t_\lambda \gg 1/H$ that interacts with an environment made of fast degrees of freedom with time-scale $1/H$. The latter can be integrated out perturbatively to obtain a diffusion-like (Fokker-Planck) equation for the probability distribution of φ:

$$\partial_t p(\varphi, t) = \frac{H^3}{8\pi^2}\partial_\varphi^2 p(\varphi, t) + \frac{1}{3H}\partial_\varphi(V'(\varphi)p(\varphi, t)). \tag{141}$$

This equation has to be solved *non-perturbatively* in $V(\varphi)$ (and hence in λ) in order to find the late-time behavior of the correlation functions. Without a detailed derivation, let us verify that this equation makes sense.[4]

The first term on the RHS of (141) is called the diffusion term and the second the drift. Diffusion results from the quantum fluctuations, while the drift from the classical rolling of the field on the potential. If $V' = 0$, the diffusion term reproduces (132). If we neglected the diffusion term, we would expect

$$\partial_t p(\varphi, t) = -\dot{\varphi}_{\rm cl}\partial_\varphi p(\varphi, t) \approx \frac{V'}{3H}\partial_\varphi p(\varphi, t), \tag{142}$$

[4] An early derivation can be found in astro-ph/9407016. A more recent and systematic derivation in the static patch can be found in 2010.06604.

up to slow-roll corrections. Once both terms are taken into account, we can find an equilibrium distribution:

$$\partial_t p(\varphi, t) = 0 \Rightarrow p_{\rm eq}(\varphi) = N \exp\left(-\frac{8\pi^2}{3H^4} V(\varphi)\right). \tag{143}$$

This nicely agrees with (138) and (140). This equilibrium distribution easily follows from a saddle point approximation to the thermal partition function. This is computed by performing the path integral on the Euclidean manifold obtained by $t \to -it_E$ in (139). This manifold is a 4-sphere with radius $1/H$. To find the probability distribution for ϕ_L, we divide the computation of the thermal average of ϕ_L into two steps:

$$\text{Tr}\left(e^{-\beta H} f(\phi_L)\right) = \oint D\phi e^{-S_E} f(\phi_L) = \int d\varphi f(\varphi) \oint_{\phi_L(t_E=0)=\varphi} D\phi e^{-S_E}, \tag{144}$$

where $\oint$ means the integration over the thermal circle with the condition $\phi(0, x) = \phi(\beta, x)$. This suggests that $p_{\rm eq}(\varphi) \propto e^{-S_E}$ on the classical solution. Since the potential is not very steep, the saddle point is $\phi = \varphi$ over the entire 4-sphere up to corrections suppressed by the derivatives of the potential. In this approximation $S_E = V(\varphi)\Omega_4/H^4$ where Ω_4 is the volume of unit 4-sphere. This is exactly the exponent in (143).

The behavior of correlation functions at large time or distance is related to how deviations from $p_{\rm eq}$ relax. This relaxation is controlled by the eigenvalues of (141):

$$p(\varphi, t) = \sum_n c_n e^{-\Lambda_n H t} p_n(\varphi), \tag{145}$$

where $p_0 = p_{\rm eq}$ and $\Lambda_0 = 0$. By a field redefinition one can reduce this eigenvalue problem to a Schrodinger problem (see Starobinsky and Yokoyama astro-ph/9407016). For instance, for the ϕ^4 interaction (131), one finds

$$\begin{aligned} \Lambda_1 &\approx 1.37\sqrt{\frac{\lambda}{24\pi^2}}, \\ \Lambda_2 &\approx 4.45\sqrt{\frac{\lambda}{24\pi^2}}. \end{aligned} \tag{146}$$

11 Eternal Inflation

Inflation as a mechanism to explain the initial conditions of the *observable universe* must end within our past light cone and be followed by a phase transition to a radiation dominated universe. However, *globally*, it might enter an eternal phase in which there is always an inflating domain. Also, the present accelerated expansion whose simplest explanation is a nonzero cosmological constant results in eternal inflation. Eternal inflation poses a challenge to a holistic quantum description of cosmology.

11.1 Slow-roll eternal inflation

Consider the evolution of the inflaton field averaged over a spatial region of fixed physical size $L_{\rm ph} \gg 1/H$. We call this $\bar{\phi}(t)$. Classically, $\bar{\phi}(t)$ slowly rolls down the potential, but there are also quantum fluctuations. As argued before, these quantum fluctuations can be treated classically at superhorizon scales, implying that $\bar{\phi}(t)$ is a stochastic variable. Writing $\phi(t,x) = \phi_0(t) + \varphi(t,x)$, and defining the volume average as

$$\bar{\varphi} = \int_0^{k<aH^2/L_{\rm ph}} \frac{d^3\boldsymbol{k}}{(2\pi)^3}\varphi_{\boldsymbol{k}}, \tag{147}$$

we find, at leading order in slow-roll parameters,

$$\left\langle (\bar{\varphi}(t) - \bar{\varphi}(0))^2 \right\rangle = \frac{H^3}{4\pi^2}t. \tag{148}$$

This can be interpreted as saying that $\bar{\varphi}$ performs a random walk motion with step size $\mathcal{O}(H)$ and time steps $\mathcal{O}(1/H)$. These fluctuations win over the classical rolling of $\bar{\phi}$ if

$$\dot{\phi}_0 < H^2. \tag{149}$$

Taking into account the fact that the physical exponentially grows with time, there is a chance that inflaton remains up on the potential somewhere in the universe and inflation never ends.

Let us check this against cosmological observables. Recall that the amplitude of adiabatic fluctuations is given by $\zeta_{\rm rms} \sim H^2/\dot{\phi}_0 \sim 10^{-4}$, safely away from the eternal regime. However, most inflationary potentials have regions that allow eternal inflation. Even if ϕ starts far away from the eternal regime, there is a finite probability that by a rare quantum fluctuation a patch of universe jumps to the eternal region and inflates forever.

Eternal inflation is puzzling because it is an unending process of recreating itself and everything that can follow from that phase. As a result, all possibilities are realized, and they are realized an infinite number of times. This leads to a loss of predictivity since the probability distribution of different cosmological scenarios depends on how the infinities are regulated. This is called the **measure problem**.

1. Does $m^2\phi^2$ inflation allow for eternal inflation? Is Einstein gravity applicable in that regime?

11.1.1 Horizon entropy and eternal inflation

Any observer is surrounded by a cosmological horizon during inflation. Just like black holes, one can associate an entropy to this horizon. Gibbons and Hawking did this:

$$S_{\rm GH} = \frac{A}{4G} \sim \frac{M_{\rm pl}^2}{H^2}. \tag{150}$$

A curious connection between this entropy and transition eternal inflation was pointed out in 0704.1814: As inflaton rolls down the potential H gradually decreases and $S_{\rm GH}$ grows at the rate

$$\frac{dS_{\rm GH}}{dN_e} \sim -\frac{M_{\rm pl}^2 \dot{H}}{H^4} \sim \frac{1}{\zeta_{\rm rms}^2}. \tag{151}$$

The condition for this being < 1 coincides with that of eternal inflation. In other words the maximum length of inflation without being eternal is bounded by the entropy at the end of inflation

$$S_{\rm GH}(t_f) > N_e, \qquad \text{no eternal inflation.} \tag{152}$$

This bound is suggestive of a cosmological analog of Page time in black hole evaporation process.

11.1.2 Phase transition to slow-roll eternal inflation

An interesting work by Creminelli et al (0802.1067) derived a sharp criterion for the phase transition to eternal inflation. This was done at leading order in slow-roll parameters, taking $\dot{\phi}_0$ and H to be constant, and $V(\phi)$ linearly falling until $\phi = \phi_*$ at which potential suddenly drops to zero and inflation ends ("reheating"). The starting point is a finite periodic box of volume L^3, with $\phi = 0$ everywhere. Classically, the field rolls down uniformly and the volume of the reheating surface is

$$\mathrm{Vol}_{\rm RH} = L^3 e^{3H\phi_*/\dot{\phi}_0}. \tag{153}$$

Quantum mechanically, different points arrive to ϕ_* at different times. We can compute the average reheating volume

$$\begin{aligned} \langle \mathrm{Vol}_{\rm RH} \rangle &= \left\langle \int d^3x e^{3Ht_{\rm RH}(x)} \right\rangle = \int d^3x \left\langle e^{3Ht_{\rm RH}(x)} \right\rangle \\ &= L^3 \int_0^\infty dt e^{3Ht} p_{\rm RH}(t) \end{aligned} \tag{154}$$

where in last step we used the fact that by statistical homgeneity averages are x-independent. Hence, we need the distribution function for the reheating time which can be related to the 1-point distribution of ϕ, averaged over a superhorizon region. Separating the background piece, we note that (148) implies that $P(\varphi, t)$ satisfies a diffusion equation (from now on φ means $\bar{\varphi}$, the average defined in (147))

$$\partial_t P(\varphi,t) = D\partial_\varphi^2 P(\varphi,t). \tag{155}$$

With no reheating, the solution with $\varphi(0)=0$ would be

$$P(\varphi,t) = \frac{e^{-\frac{\varphi^2}{4Dt}}}{\sqrt{2\pi Dt}}. \tag{156}$$

However, any trajectory ends once it arrives to $\phi = \phi_*$. This is equivalent to a moving absorbing boundary at $\varphi_* = \phi_* - \dot{\phi}_0 t$ for the above diffusion problem. The solution can be found by the method of images

$$P(\varphi,t) = \frac{1}{\sqrt{2\pi Dt}}\left(e^{-\frac{\varphi^2}{4Dt}} - \text{image}\right), \tag{157}$$

the second term ensuring that $P(\varphi_*(t),t)=0$.

The probability for reheating time being between t and $t+dt$ is the probability that the random walk crosses the absorbing boundary for the first time in this interval:

$$p_{\rm RH}(t) = -\partial_t \int_{-\infty}^{\varphi_*(t)} d\varphi P(\varphi,t) = -D\partial_\varphi P(\varphi,t)|_{\varphi_*(t)}. \tag{158}$$

The key observation is that this quantity has the late time behavior

$$p_{\rm RH}(t) \propto e^{-\frac{\varphi_*^2(t)}{4Dt}} \to e^{-\frac{\dot{\phi}_0^2}{4D}t}. \tag{159}$$

Substituting in (154), one can see that $\langle \mathrm{Vol}_{\rm RH}\rangle = \infty$ when

$$3H > \frac{\dot{\phi}_0^2}{4D} \Rightarrow \frac{3H^4}{2\pi^2\dot{\phi}^2} > 1. \tag{160}$$

In 0802.1067, it was shown that a sharp phase transition happens at this point: there is a finite probability for $\mathrm{Vol}_{\rm RH}$ being infinite after this threshold.

11.2 False-vacuum eternal inflation

What makes eternal inflation more interesting (and more complex) is the possibility of non-perturbative transitions between different vacua, i.e. local minima of the supposedly multidimensional potential for all scalar fields.

11.2.1 Tunneling in quantum mechanics

Consider a potential $V(x)$ that has a local minimum at $x=0$, with $V(0)>0$ separated by a barrier from its asymptotic value $V(x\to\infty)=0$. Classically, a particle trapped in the local minimum will always remain there. Quantum mechanically, it will tunnel out. The standard way to find the tunneling rate is the WKB method. For a given energy E, below the barrier, there are three classical turning points $x_1 < x_2 < x_3$. The leading WKB solution for $x_1 < x < x_2$, is of the form

$$\psi_I(x) \simeq a_1 e^{i\int_{x_1}^{x}\sqrt{2(E-V(x'))}dx'} + a_2 e^{-i\int_{x_1}^{x}\sqrt{2(E-V(x'))}dx'}, \tag{161}$$

for $x_3 < x$

$$\psi_{III}(x) \simeq c e^{i\int_{x_1}^{x}\sqrt{2(E-V(x'))}dx'}, \tag{162}$$

where the relation between c and a_1, a_2 (and hence the tunneling rate) is obtained by matching to the solution in the forbidden region $x_2 < x < x_3$

$$\psi_{II}(x) \simeq b e^{-\int_{x_2}^{x}\sqrt{2(V(x')-E)}dx'}. \tag{163}$$

ψ_I and ψ_{III} are related to the saddles of the path integral $\int Dx e^{iS}$, which are the classical solutions. We can write

$$iS_{cl} = i\int dt L[x_{cl}(t)] = i\int dx p_{cl}(x) - iEt, \tag{164}$$

where the classical momentum is $p_{cl}(x) = \pm\sqrt{2(E-V(x))}$. Coleman made the interesting observation that even the wavefunction in the forbidden region ψ_{II} is related to a classical solution, but a complex one.[5] It can be obtained by analytically continuing $t \to -it$, under which the exponent in the path integral changes to

$$iS \to -S_E = -\int dt\left(\frac{1}{2}\dot{x}^2 + V(x)\right). \tag{165}$$

Note that this is equivalent to a motion in an inverted potential $V_E(x) = -V(x)$, so region II becomes the allowed region and the solution satisfies

$$\frac{1}{2}\dot{x}^2 - V(x) = -E. \tag{166}$$

Substituting in the action gives

$$-S_E = -Et - \int dx\sqrt{2(V(x)-E)}, \tag{167}$$

which reproduces the correct WKB exponent for ψ_{II} after analytically continuing back to Lorentzian time. The exponential suppression of the tunneling rate comes from squaring ψ_{II}, so it is given by the normalized *bounce action*:

$$\Gamma \propto e^{-B}, \qquad B = [S_E - 2Et]_{x_2\to x_3\to x_2}. \tag{168}$$

[5]Most of what is said below are beautifully explained in works of Coleman. In particular "The uses of Instantons" in his book "Aspects of Symmetry".

11.2.2 Decay of false vacuum in QFT

This connection between tunneling and complex saddles of path integral might seem like a cute observation but hardly necessary. In QM, we can instead solve the Schrödinger equation. However, in QFT, wavefunction becomes a wavefunctional and the Schrödinger functional differential equation that it satisfies is not tractable. Looking for complex saddles is one the main analytic tools to study non-perturbative phenomena in QFT.

Let us therefore consider the following potential

$$V(\phi) = \lambda(\phi^2 - v^2)^2 - \epsilon\frac{\phi}{2v}, \tag{169}$$

with two minima $\phi_\pm \simeq \pm v$, with an asymmetry $\Delta V \simeq \epsilon$, the true vacuum being at ϕ_+. What happens if there is a large region where ϕ is initially in the false vacuum ϕ_-? Classically, it stays there. Also perturbation theory around that never shows a sign of instability. However, in the same way that the particle in a metastable minimum eventually tunnels out, the false vacuum should eventually decay into the true vacuum. Indeed, this happens but via the formation of bubbles of true vacuum that subsequently expand and eat the false vacuum, much like what happens when water is boiling. Our task is to find the size of the most probable bubble ("critical bubble") and the formation rate, or at least the exponential that suppresses it.

Typically, most symmetric configurations have the lowest action. Therefore, the critical bubble is expected to be spherically symmetric. Initially, the field profile is interpolating between $\phi(r = 0) \simeq \phi_+$ to $\phi(r \to \infty) = \phi_-$. If the barrier height is large $\lambda v^4 \gg \epsilon$, then the field rapidly transitions between the two minima and a very useful approximation called the *thin-wall approximation* can be made. In this limit, we can reduce the problem to a QM problem of the domain-wall radius r_w.

The domain-wall action is a sum of two contributions: a surface term, having to do with the nonzero surface tension of the wall $\sigma \sim \sqrt{\lambda}v^3$, and a volume term, having to do with the difference in the energy densities ϵ:

$$S_w = -4\pi\sigma \int d\tau r_w^2 + \frac{4\pi}{3}\epsilon \int dt r_w^3, \tag{170}$$

where τ is the proper time of the wall. Using $\dot{t} \equiv \frac{dt}{d\tau} = \sqrt{1 + \dot{r}_w^2}$, we obtain

$$S_w = 4\pi \int d\tau \left[\frac{1}{3}\epsilon r_w^3 \sqrt{1 + \dot{r}_w^2} - \sigma r_w^2\right]. \tag{171}$$

The corresponding Hamiltonian is

$$H = 4\pi\sigma r_w^2 - \frac{4\pi}{3\sqrt{1 + \dot{r}_w^2}}\epsilon r_w^3. \tag{172}$$

There is a family of solutions labeled by the energy E, describing bubbles that start from $r_w = \infty$ at $t = -\infty$, shrink to a minimum radius $r_{\min}(E)$ and expand again. The contracting part is not relevant to tunneling. It is analogous to the left moving solution in region III of the QM tunneling studied above. Instead, there is Euclidean solution that at the turning point matches the onto the

Lorentzian solution when $\dot{r}_w = 0$. The energy has to be set to $E = 0$ because we are interested in the spontaneous nucleation of bubbles. Hence, the critical radius is

$$r_c = \frac{3\sigma}{\epsilon}. \tag{173}$$

The critical bubble always exists: as r_w increases, the positive surface term grows as r_w^2, while the negative volume terms as r_w^3. Hence, for tiny bubbles energy is positive and for very large ones negative, crossing zero at $r_w = r_c$. After nucleation the bubble wall follows a motion with constant acceleration:

$$r_w = r_c \cosh(\tau/r_c). \tag{174}$$

This solution not only has the spherical symmetry but the full $SO(1,3)$ symmetry. The Euclidean solution is $r_w = r_c \cos(\tau/r_c)$: a bubble that starts from $r_w = 0$ at $\tau = -\pi r_c$ expands to $r_w = r_c$ and then bounces back to zero. It is a parametrization of a perfect 3-sphere in $\mathbf{R}_4$. Hence, it has $O(4)$ symmetry. The Euclidean action gives the exponent suppressing the tunneling rate

$$\log \Gamma \sim -\sigma r_c^3. \tag{175}$$

2. Compute the bounce action.

11.2.3 Decay of false vacuum in quantum gravity

Coleman and De Luccia included the gravitational effect. Now the vacuum energy determines the spacetime curvature, so for instance transition rate from a dS vacuum (i.e. one with vacuum energy density $\rho = \Lambda > 0$) to another with $\rho = \Lambda - \epsilon$, would differ compared to a transition from a Minkowski vacuum (i.e. one with $\rho = 0$) to an AdS vacuum with $\rho = -\epsilon$. The latter might even be forbidden because the area and volume of big spheres in AdS spacetime grow at the same rate, hence the balance between the surface tension and bulk energy might never be reached.

We are interested in tunneling out of a dS vacuum. dS can be described as a hyperboloid embedded in a Minkowski space with one higher dimension:

$$-X_0^2 + \sum_{i=1}^{4} X_i^2 = 1/H^2. \tag{176}$$

It is globally covered by the following coordinate system

$$ds^2 = \frac{1}{H^2}\left(-d\theta^2 + \cosh^2\theta d\Omega_3^2\right), \qquad -\infty < \theta < \infty, \tag{177}$$

where $d\Omega_3^2$ is the line element of a 3-sphere. As before, we are looking for Euclidean bounce solutions. Analytic continuation of dS, as can be seen by sending $X_0 \to -iX_0$ in (176), or sending $\theta \to -i\theta$ in (177), gives a 4-sphere.

The leading semi-classical result for tunneling rate is

$$\Gamma \propto e^{-S_E[\Lambda_1,\Lambda_2]+S_E[\Lambda_1]}, \tag{178}$$

where the first term in the exponent is the bounce action for a Euclidean geometry containing a region of the daughter vacuum (Λ_2) separated by a 3-sphere from the parent vacuum (Λ_1). When $\Lambda_1 > 0$, this is topologically a 4-sphere. The second term subtracts the action of the Euclidean geometry in the parent vacuum with no bubble. This is exactly a 4-sphere. I'll conclude with some comments:[6]

- If $r_c \ll 1/H$, the curvature effects become irrelevant and (178) reduces to the QFT answer.
- Tunneling from a dS vacuum to any other vacuum (dS, Minkowski, or AdS) is possible. It is even possible to have $\Lambda_2 > \Lambda_1 > 0$.
- If the tunneling rate is smaller than the expansion rate ($\Gamma < H^4$) the parent vacuum will never be fully consumed by the bubbles. This results in *false vacuum eternal inflation*. It is in sharp contrast with the non-gravitational case, where bubbles of true vacuum expand and eventually occupy the entire volume.
- Eternal inflation lasts for ever, and therefore via CDL tunneling, it populates the entire landscape of minima. This idea is an ingredient for the *anthropic solution to the cosmological constant problem.*
- The late time slices in eternal inflation have a fractal structure consisting of colliding bubbles and bubbles within bubbles in different vacua. The statistics of these bubbles are sensitive to how the time slices are drawn. This is a manifestation of the measure problem.
- When gravity is dynamical, the full action is $S_{\rm EH} + S_{\rm matter}$. It turns out that the Euclidean gravitational action is not bounded below. For a dS vacuum, it is related to the horizon entropy:
$$S_E[\Lambda] = -S_{\rm GH} = -\frac{A}{4G}. \tag{179}$$
Hence (178) seems to support the interpretation of $S_{\rm GH}$ as an entropy. In particular, if we take $\Lambda_2 \gg \Lambda_1$, then $S_E[\Lambda_1,\Lambda_2] \simeq S_E[\Lambda_2]$ and the transition rate proportional to $\exp(\Delta S_{\rm GH})$.
- One could imagine that slow-roll inflation that preceded the radiation domination phase of our universe began via a quantum tunneling. Unfortunately the Euclidean measure does not give an acceptable prediction for the duration of slow-roll inflation. See section 9 and 2403.10510.

[6] See Susskind's lectures at PiTP 2011 for more on eternal inflation.