Primordial non-Gaussianity in the cosmological perturbations

Antonio Riotto

Département de Physique Théorique, Université de Genève,

24 quai Ansermet, CH-1211, Genève, Switzerland

Abstract

This set of notes have been written down as supplementary material for the course on primordial non-Gaussianity in the cosmological perturbations at the II Jayme Tiomno School of Cosmology held at Brazilian Center for Research in Physics in Rio de Janeiro from 6 -10 August, 2012. Hopefully they are self-contained, but by no means they are intended to substitute any of the reviews on the subject. The notes contain some extended introductory material and a set of exercises, whose goal is to familiarize the students with the basic notions necessary to deal with the issue of non-Gaussianity in the cosmological perturbations.

email: `Antonio.Riotto@unige.ch`, phone: +41 22 379 6310 August 7, 2012

Literature

During the preparation of this set of notes, we have been found useful consulting the following reviews and textbooks:

N. Bartolo, E. Komatsu, S. Matarrese and A. Riotto, "Non-Gaussianity from inflation: Theory and observations," Phys. Rept. **402**, 103 (2004). [astro-ph/0406398];

X. Chen, "Primordial Non-Gaussianities from Inflation Models," Adv. Astron. **2010**, 638979 (2010) [arXiv:1002.1416 [astro-ph.CO]];

V. Desjacques and U. Seljak, "Primordial non-Gaussianity in the large scale structure of the Universe," Adv. Astron. **2010** (2010) 908640 [arXiv:1006.4763 [astro-ph.CO]];

S. Dodelson, "Modern Cosmology", Academic Press, 2003;

J.A. Peacock, "Cosmological Physics", Cambridge University Press, 1999.

Units

We will adopt natural, or high energy physics, units. There is only one fundamental dimension, energy, after setting $\hbar = c = k_{\rm b} = 1$,

$$[\text{Energy}] = [\text{Mass}] = [\text{Temperature}] = [\text{Length}]^{-1} = [\text{Time}]^{-1}\,.$$

The most common conversion factors and quantities we will make use of are

1 $\text{GeV}^{-1} = 1.97 \times 10^{-14}$ cm$=6.59 \times 10^{-25}$ sec,

1 Mpc$=$ $3.08{\times}10^{34}$ cm$=1.56{\times}10^{33}$ GeV^{-1},

$M_{\rm Pl} = 1.22 \times 10^{19}$ GeV,

H_0= 100 h Km sec^{-1} Mpc^{-1}=2.1 h $\times 10^{-42}$ GeV^{-1},

$\rho_{\rm c} = 1.87h^2 \cdot 10^{-29}{\rm g\,cm^{-3}} = 1.05h^2 \cdot 10^4\,{\rm eV\,cm^{-3}} = 8.1h^2 \times 10^{-47}$ GeV^4,

$T_0 = 2.75$ K$=2.3{\times}10^{-13}$ GeV,

$T_{\rm eq} = 5.5(\Omega_0 h^2)$ eV,

$T_{\rm ls} = 0.26\,(T_0/2.75\,{\rm K})$ eV.

Contents

Part I

Introduction

Our current understanding of the evolution of the universe is based upon the Friedmann-Robertson-Walker (FRW) cosmological model, or the hot big bang model as it is usually called. The model is so successful that it has become known as the standard cosmology. times. The FRW cosmology is so robust that it is possible to make sensible speculations about the universe at times as early as 10^{-43} sec after the Big Bang.

The most important feature of our universe is its large scale homogeneity and isotropy. This feature ensures that observations made from our single point are representative of the universe as a whole and can therefore be legitimately used to test cosmological models. For most of the twentieth century, the homogeneity and isotropy of the universe had to be taken as an assumption, known as the Cosmological Principle. The assumption of isotropy and homogeneity dates back to the earliest work of Einstein, who made the assumption not based upon observations, but as theorists often do, to simplify the mathematical analysis. The Cosmological Principle remained an intelligent guess until firm empirical data, confirming large scale homogeneity and isotropy, were finally obtained at the end of the twentieth century. The best evidence for the isotropy of the observed universe

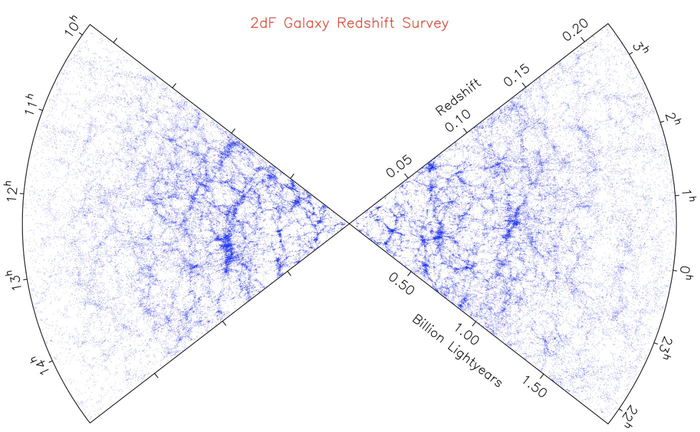

Figure 1: The large-scale structure from the 2dF Galaxy Survey

is the uniformity of the temperature of the cosmic microwave background (CMB) radiation: aside from the observed dipole anisotropy, the temperature difference between two antennas separated by angles ranging from about 10 arc seconds to 180° is smaller than about one part in 10^5. The simplest interpretation of the dipole anisotropy is that it is the result of our motion relative to the cosmic rest frame. If the expansion of the universe were not isotropic, the expansion anisotiopy would lead to a temperature anisotiopy in the CMBR of similar magnitude. Likewise, inhomogeneities in the density of the universe on the last scattering surface would lead to temperature anisotropies. In this iegard,

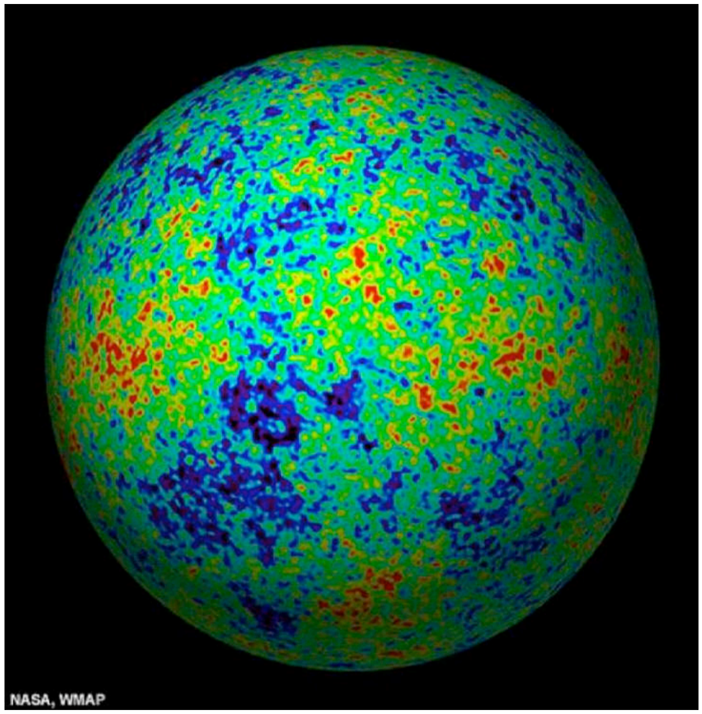

Figure 2: The CMB radiation projected onto a sphere

the CMBR is a very powerful probe: It is even sensitive to density inhomogeneities on scales larger than our present Hubble volume. The remarkable uniformity of the CMB radiation indicates that at the epoch of last scattering for the CMB radiation (about 200,000 yr after the bang) the universe was to a high degree of precision (order of 10^{-5} or so) isotropic and homogeneous. Homogeneity and isotropy is of course true if the universe is observed at sufficiently large scales. The observable patch of the universe is of order 3000 Mpc. Redshift surveys suggest that the universe is homogeneous and isotropic only when coarse grained on scales of the order of 100 Mpc. On smaller scales there exist large inhomogeneities, such as galaxies, clusters and superclusters. Hence, the Cosmological Principle is only valid within a limited range of scales, spanning a few orders of magnitude. The inflationary theory, as we shall see, suggests that the universe continues to be homogeneous and isotropic over distances larger than 3000 Mpc.

It is firmly established by observations that our universe therefore

- is homogeneous and isotropic on scales larger than 100 Mpc and has well developed inhomogeneous structure on smaller scales;
- expands according to the Hubble law for which the recession velocity of, say, galaxies is proportional to their distances.

Concerning the matter composition of the universe, we know that

- it is pervaded by thermal microwave background radiation with temperature $T_0 \simeq 2.73$ K;

- there is baryonic matter, roughly one baryon per 10^9 photons, but no substantial amount of antimatter;
- the chemical composition of baryonic matter is about 75% hydrogen, 25% helium, plus trace amounts of heavier elements;
- baryons contribute only a small percentage of the total energy density; the rest is a dark component, which appears to be composed of cold dark matter with negligible pressure (25%) and dark energy with negative pressure (70%).

Observations of the fluctuations in the cosmic microwave background radiation suggest that

- there were only small fluctuations of order 10^{-5} in the energy density distribution when the universe was a thousand times smaller than now.

Any cosmological model worthy of consideration must be consistent with established facts. While the standard big bang model accommodates most known facts, a physical theory is also judged by its predictive power. At present, inflationary theory, naturally incorporating the success of the standard big bang, has no competitor in this regard. Therefore, we will build upon the standard big bang model, which will be our starting point, until we reach contemporary ideas of inflation. We will then show how its prediction influences the CMB physics as well as the physics of the matter perturbations. This set of notes also offers two Appendices about the generation of the baryon asymmetry of the universe and the phase transitions in the early universe. These subjects will not be covered during the lectures.

1 The Friedmann-Robertson-Walker metric

As discussed in the previous section, the distribution of matter and radiation in the observable universe is homogeneous and isotropic. While this by no means guarantees that the entire universe is smooth, it does imply that a region at least as large as our present Hubble volume is smooth. So long as the universe is spatially homogeneous and isotropic on scales as large as the Hubble volume, for purposes of description of our local Hubble volume we may assume the entire universe is homogeneous and isotropic. While a homogeneous and isotropic region within an otherwise inhomogeneous and anisotropic universe will not remain so forever, causality implies that such a region will remain smooth for a time comparable to its light-crossing time. This time corresponds to the Hubble time, about 10 Gyr. We have determined during the last part of the GR course that the metric of a maximally symmetric space satisfying the Cosmological Principle is the the metric for a space with homogeneous and isotropic spatial sections, that is the Friedmann-Robertson-Walker metric, which can be written in the form

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\right], \tag{1}$$

where (t, r, θ, ϕ) are are coordinates (referred to as comoving coordinates), $a(t)$ is the cosmic scale factor, and with an appropriate reseating of the coordinates, k can be chosen to be $+1, -1$, or 0 for spaces of constant positive, negative, or zero spatial curvature, respectively. The coordinate r is dimensionless, *i.e.* $a(t)$ has dimensions of length, and r ranges from 0 to 1 for $k = +1$. The time coordinate is just the proper (or clock) time measured by an observer at rest in the comoving frame, *i.e.*, (r, θ, ϕ)=constant. As we shall discover shortly, the term comoving is well chosen: observers at rest in the comoving frame remain at rest, *i.e.*, (r, θ, ϕ) remain unchanged, and observers initially moving with respect to this frame will eventually come to rest in it. Thus, if one introduces a homogeneous, isotropic fluid initially at rest in this frame, the t =constant hypersurfaces will always be orthogonal to the fluid flow, and will always coincide with the hypersurfaces of both spatial homogeneity and constant fluid density.

We already know the curvature tensor of the maximally symmetric metric entering the FRW metric (and its contractions), this is not difficult.

1. First of all, we write the FRW metric as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\widetilde{g}_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{2}$$

   From now on, all objects with a tilde will refer to three-dimensional quantities calculated with the metric $\widetilde{g}_{ij}$.

2. One can then calculate the Christoffel symbols in terms of $a(t)$ and $\widetilde{\Gamma}^i{}_{jk}$. The nonvanishing components are (we had already established that $\Gamma^\mu{}_{00} = 0$)

$$\begin{aligned} \Gamma^i{}_{jk} &= \widetilde{\Gamma}^i{}_{jk}, \\ \Gamma^i{}_{j0} &= \frac{\dot{a}}{a}\delta^i_j, \\ \Gamma^0{}_{ij} &= \frac{\dot{a}}{a}\widetilde{g}_{ij}. \end{aligned} \tag{3}$$

3. The relevant components of the Riemann tensor are

$$\begin{aligned} R^i{}_{0j0} &= -\frac{\ddot{a}}{a}\delta^i_j, \\ R^0{}_{i0j} &= \ddot{a}a\widetilde{g}_{ij}, \\ R^k{}_{ikj} &= \widetilde{R}_{ij} + 2\dot{a}^2\widetilde{g}_{ij}. \end{aligned} \tag{4}$$

4. Now we can use $\widetilde{R}_{ij} = 2k\widetilde{g}_{ij}$ (as a consequence of the maximally symmetry of $\widetilde{g}_{ij}$) to calculate $R_{\mu\nu}$. The nonzero components are

$$\begin{aligned} R_{00} &= -3\frac{\ddot{a}}{a}, \\ R_{ij} &= \left(a\ddot{a} + 2\dot{a}^2 + 2k\right)\widetilde{g}_{ij}, \\ &= \left(\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{k}{a^2}\right) g_{ij}. \end{aligned} \tag{5}$$

5. The Ricci scalar is

$$R = \frac{6}{a^2}\left(a\ddot{a} + \dot{a}^2 + k\right), \tag{6}$$

and

6. the Einstein tensor has the components

$$\begin{aligned} G_{00} &= 3\left(\frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right), \\ G_{0i} &= 0, \\ G_{ij} &= -\left(2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right) g_{ij}. \end{aligned} \tag{7}$$

1.1 Open, closed and flat spatial models

In order to illustrate the construction of the metric, consider the simpler case of a two spatial dimensions. dimensions. Examples of two-dimensional spaces that are homogeneous and isotropic are the flat plane $\mathbb{R}^2$, the positively-curved closed two sphere $\mathbb{S}^2$ and the negatively-curved hyperbolic plane $\mathbb{H}^2$.

Consider first a two sphere $\mathbb{S}^2$ of radius a and embedded in a three-dimensional space $\mathbb{R}^3$. The equation of the sphere of radius a is

$$x_1^2 + x_2^2 + x_3^2 = a^2. \tag{8}$$

The element of length is the three-dimensional Euclidean space is

$$\mathrm{d}\mathbf{x}^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2. \tag{9}$$

Since x_3 is a fictitious coordinate, we can eliminate it in favour of the other two and write

$$\mathrm{d}\mathbf{x}^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \frac{(x_1\mathrm{d}x_1 + x_2\mathrm{d}x_2)^2}{a^2 - x_1^2 - x_2^2}. \tag{10}$$

Now, let us introduce the polar coordinates

$$x_1 = r' \cos\theta, \;\; x_2 = r' \sin\theta, \tag{11}$$

in terms of which the infinitesimal line length becomes

$$\mathrm{d}\mathbf{x}^2 = \frac{a^2 \mathrm{d}r'^2}{a^2 - r'^2} + r'^2 \mathrm{d}\theta^2. \tag{12}$$

Finally, with the definition of a dimensionless coordinate $r = r'/a$ $(0 \leq r \leq 1)$, the spatial metric becomes

$$\mathrm{d}\mathbf{x}^2 = a^2 \left[\frac{\mathrm{d}r^2}{1 - r^2} + r^2 \mathrm{d}\theta^2 \right]. \tag{13}$$

Note the similarity between this metric and the $k = 1$ FRW metric. It should also now be clear that $a(t)$ is the radius of the space. The poles of the two-sphere are at $r = 0$, the equator is at $r = 1$. The locus of points of constant r sweep out the latitudes of the sphere, while the locus of points of constant θ sweep out the longitudes of the sphere. Another convenient coordinate system for the two sphere is that specified by the usual polar and azimuthal angles (θ, ϕ) of spherical coordinates, related to the x_i by

$$x_1 = a \sin\theta \cos\phi, \;\; x_2 = a \sin\theta \sin\phi, \;\; x_3 = a \cos\theta. \tag{14}$$

In terms of these coordinates, the spatial line element becomes

$$\mathrm{d}\mathbf{x}^2 = a^2 \left[\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2 \right]. \tag{15}$$

This form makes manifest the fact that the space is the two sphere of radius a. The volume of the two sphere is easily calculated

$$V_{\mathbb{S}^2} = \int \mathrm{d}^2 \sqrt{\tilde{g}} = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \, a^2 \sin^2\theta = 4\pi a^2, \tag{16}$$

as expected. The two sphere is homogeneous and isotropic. Every point in the space is equivalent to every other point, and there is no preferred direction. In other words, the space embodies the Cosmological Principle, *i.e.*, no observer (especially us) occupies a preferred position in the universe. Note that this space is unbounded; there are no edges on the two sphere. It is possible to circumnavigate the two sphere, but it is impossible to fall off. Although the space is unbounded, the volume is finite. The expansion (or contraction) of this two-dimensional universe equivalent to an increase (or decrease) in the radius of the two sphere a. Since the universe is apatiatly homogeneous and isotropic, the scale factor can only be a function of time. As the two sphere expands or contracts, the coordinates (r and θ in the case of the two sphere) remain unchanged; they are "comoving."

Also note that the physical distance between any two comoving points in the space scales with a (hence the name scale factor).

The equivalent formulas for a space of constant negative curvature can be obtained with the replacement $a \to ia$. In this case the embedding is in a three-dimensional Minkowski space. The metric corresponding to the form for the negative curvature case is

$$\mathrm{d}\mathbf{x}^2 = a^2 \left[\frac{\mathrm{d}r^2}{1-r^2} + r^2 \mathrm{d}\theta^2 \right] = a^2 \left[\mathrm{d}\theta^2 + \sinh^2\theta \mathrm{d}\phi^2 \right]. \tag{17}$$

The hyperbolic plane $\mathbb{H}^2$ is unbounded with infinite volume since $0 \le \theta < \infty$. The embedding of $\mathbb{H}^2$ in a Euclidean space requires three fictitious extra dimensions, and such an embedding is of little use in visualizing the geometry. The spatially-flat model can be obtained from either of the above examples by taking the radius a to infinity. The flat model is unbounded with infinite volume. For the flat model the scale factor does not represent any physical radius as in the closed case, or an imaginary radius as in the open case, but merely represents how the physical distance between comoving points scales as the space expands or contracts.

The generalization of the two-dimensional models discussed above to three spatial dimensions is trivial. For the three sphere a fictitious fourth spatial dimension is introduced, and in cartesian coordinates the three sphere is denned by: $a^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$. The spatial metric is $\mathrm{d}\mathbf{x}^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 + \mathrm{d}x_4^2$. The fictitious coordinate can be removed to give

$$\mathrm{d}\mathbf{x}^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 \frac{(x_1\mathrm{d}x_1 + x_2\mathrm{d}x_2 + x_3\mathrm{d}x_3)^2}{a^2 - x_1^2 - x_2^2 - x_3^2}. \tag{18}$$

In terms of the coordinates $x_1 = r' \sin\theta\cos\phi$, $x_2 = r'\sin\theta\sin\phi$ and $x_3 = r'\cos\theta$, this metric becomes equal to the FRW metric with $k = +1$ and $r = r'/a$. In the coordinate system that employes the three angular coordinates (χ, θ, ϕ) of a four-dimensional spherical coordinates system, $x_1 = a\sin\chi\sin\theta\cos\phi$, $x_2 = a\sin\chi\sin\theta\sin\phi$, $x_3 = a\sin\chi\cos\theta$ and $x_4 = a\cos\chi$, the metric is given by

$$\mathrm{d}\mathbf{x}^2 = a^2 \left[\mathrm{d}\chi^2 + \sin^2\chi(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2) \right]. \tag{19}$$

The volume of the three-sphere is

$$V_{\mathbb{S}^3} = \int \mathrm{d}^3x \sqrt{g} = 2\pi a^3. \tag{20}$$

As for the two-dimensional case, the three-dimensional open model is obtained by the replacement $a \to ia$, which gives the FRW metric with $k = -1$. Again the Again the space is unbounded with infinite volume and $a(t)$ sets the curvature scale. Embedding $\mathbb{H}^3$ in a Euclidean space requires four fictitious extra dimensions.

It should be noted that the assumption of local homogeneity and isotropy only implies that the spatial metric is locally $\mathbb{S}^3$, $\mathbb{H}^3$ or $\mathbb{R}^3$ and the space can have different global properties. For instance, for the spatially flat case the global properties of the space might be that of the three torus, $\mathbb{T}^3$, rather than $\mathbb{R}^3$; this is accomplished by identifying the opposite sides of a fundamental spatial volume element. Such non-trivial topologies may be relevant in light of recent work on theories with extra dimensions. In many such theories the internal space (of the extra dimensions) is compact, but with non-trivial topology, *e.g.*, containing topological defects such as holes, handles, and so on. If the internal space is not simply connected, it suggests that the external space may also be non-trivial, and the global properties of our three-space might be much richer than the simple $\mathbb{S}^3$, $\mathbb{H}^3$ or $\mathbb{R}^3$ topologies.

1.2 The particle horizon and the Hubble radius

A fundamental question in cosmology that one might ask is: what fraction of the universe is in causal contact? More precisely, for a comoving observer with coordinates (r_0, θ_0, ϕ_0), for what values of (r, θ, ϕ) would a light signal emitted at $t = 0$ reach the observer at, or before, time t? This can be calculated directly in terms of the FRW metric. A light signal satisfies the geodesic equation $\mathrm{d}s^2 = 0$. Because of the homogeneity of space, without loss of generality we may choose $r_0 = 0$. Geodesies passing through r=0 are lines of constant θ and ϕ, just as great circles eminating from the poles of a two sphere are lines of constant θ (*i.e.*, constant longitude), so $\mathrm{d}\theta = \mathrm{d}\phi = 0$. Of course, the isotropy of space makes the choice of direction θ_0, ϕ_0) irrelevant. Thus, a light signal emitted from coordinate position $(r_\mathrm{H}, \theta_0, \phi_0)$ at time $t = 0$ will reach $r_0 = 0$ in a time t determined by

$$\int_0^t \frac{\mathrm{d}t'}{a(t')} = \int_0^{r_\mathrm{H}} \frac{\mathrm{d}r'}{\sqrt{1 - kr'^2}}. \tag{21}$$

The proper distance to the horizon measured at time t is

$$\boxed{R_\mathrm{H}(t) = a(t) \int_0^t \frac{\mathrm{d}t'}{a(t')} = a(t) \int_0^a \frac{\mathrm{d}a'}{a'} \frac{1}{a' H(a')} = a(t) \int_0^{r_\mathrm{H}} \frac{\mathrm{d}r'}{\sqrt{1 - kr'^2}} \ \text{(PARTICLE HORIZON).}} \tag{22}$$

If $R_\mathrm{H}(t)$ is finite, then our past light cone is limited by this particle horizon, which is the boundary between the visible universe and the part of the universe from which light signals have not reached us. The behavior of $a(t)$ near the singularity will determine whether or not the particle horizon is finite. We will see that in the standard cosmology $R_\mathrm{H}(t) \sim t$, that is the particle horizon is finite. The particle horizon should not be confused with the notion of Hubble radius

$$\boxed{\frac{1}{H} = \frac{a}{\dot{a}} \ \text{(HUBBLE RADIUS).}} \tag{23}$$

Let us emphasize a subtle distinction between the particle horizon and the Hubble: if particles are separated by distances greater than $R_H(t)$ they never could have communicated with one another; if they are separated by distances greater than the Hubble radius H^{-1}, they cannot talk to each other at the time t.

We shall see that the standard cosmology the distance to the horizon is finite, and up to numerical factors, equal to the Hubble radius, H^{-1}, but during inflation, for instance, they are drastically different. One can also define a comoving particle horizon distance

$$\boxed{\tau_{\rm H} = \int_0^t \frac{{\rm d}t'}{a(t')} = \int_0^a \frac{{\rm d}a'}{H(a')a'^2} = \int_0^a {\rm d}\ln a' \left(\frac{1}{Ha'}\right) \text{ (COMOVING PARTICLE HORIZON).}} \tag{24}$$

Here, we have expressed the comoving horizon as an integral of the comoving Hubble radius,

$$\boxed{\frac{1}{aH} \text{ (COMOVING HUBBLE RADIUS),}} \tag{25}$$

which will play a crucial role in inflation. We see that the comoving horizon then is the logarithmic integral of the comoving Hubble radius $(aH)^{-1}$. The Hubble radius is the distance over which particles can travel in the course of one expansion time, *i.e.* roughly the time in which the scale factor doubles. So the Hubble radius is another way of measuring whether particles are causally connected with each other: if they are separated by distances larger than the Hubble radius, then they cannot communicate at a given time t (or τ). Let us reiterate that there is a subtle distinction between the comoving horizon $\tau_{\rm H}$ and the comoving Hubble radius $(aH)^{-1}$. If particles are separated by comoving distances greater than $\tau_{\rm H}$, they never could have communicated with one another; if they are separated by distances greater than $(aH)^{-1}$, they cannot talk to each other at some time τ. It is therefore possible that $\tau_{\rm H}$ could be much larger than $(aH)^{-1}$ now, so that particles cannot communicate today but were in causal contact early on. As we shall see, this might happen if the comoving Hubble radius early on was much larger than it is now so that $\tau_{\rm H}$ got most of its contribution from early times. We will see that this could happen, but it does not happen during matter-dominated or radiation-dominatd epochs. In those cases, the comoving Hubble radius increases with time, so typically we expect the largest contribution to $\tau_{\rm H}$ to come from the most recent times.

A similar concept is the event horizon, or the maximum distance we can probe in the infinite future

$$R_{\rm e}(t) = a(t)\int_t^\infty \frac{{\rm d}t'}{a(t')}, \tag{26}$$

which is clearly infinite if the universe expands a $a \sim t^n$ $(n < 1)$.

1.3 Particle kinematics of a particle

Consider the geodesic motion of a particle that is not necessarily massless. The four-velocity u^μ of a particle with respect to the comoving frame is referred to as the peculiar velocity. The geodesic equation of motion in terms of the affine parameter chosen to be the proper length is

$$\frac{\mathrm{d}u^\mu}{\mathrm{d}s} + \Gamma^\mu_{\ \nu\sigma} u^\nu u^\sigma = 0,\ u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}s}. \tag{27}$$

The $\mu = 0$ component of the geodesic equation is

$$\frac{\mathrm{d}u^0}{\mathrm{d}s} + \Gamma^0_{\ \nu\sigma} u^\nu u^\sigma = \frac{\mathrm{d}u^0}{\mathrm{d}s} + \Gamma^0_{\ ij} u^i u^j = \frac{\mathrm{d}u^0}{\mathrm{d}s} + \frac{\dot{a}}{a}\widetilde{g}_{ij} u^i u^j = 0. \tag{28}$$

Denoting by $|\mathbf{u}|^2 = \widetilde{g}_{ij} u^i u^j$ and recalling that $-(u^0)^2 + |\mathbf{u}|^2 = -1$, it follows that $u^0 \mathrm{d}u^0 = |\mathbf{u}|\mathrm{d}|\mathbf{u}|$, the geodesic equation becomes

$$\frac{\mathrm{d}u^0}{\mathrm{d}s} + \frac{\dot{a}}{a}|\mathbf{u}|^2 = \frac{1}{u^0}\frac{\mathrm{d}|\mathbf{u}|}{\mathrm{d}s} + \frac{\dot{a}}{a}|\mathbf{u}| = 0. \tag{29}$$

Finally, since $u^0 = \mathrm{d}t/\mathrm{d}s$, this equation reduces to

$$\frac{1}{|\mathbf{u}|}\frac{\mathrm{d}|\mathbf{u}|}{\mathrm{d}t} = -\frac{\dot{a}}{a}. \tag{30}$$

It implies that $|\mathbf{u}| \sim a^{-1}$. recalling that the four-momentum $p^\mu = m u^\mu$, we see that the magnitude of the three-momentum of a freely propagating particle red-shifts away as a^{-1}. Note that in eq. (29) the factors of $\mathrm{d}s$ cancel. That implies that the above discussion also applies to massless particles, where $\mathrm{d}s = 0$ (formally by the choice of a different affine parameter). In terms of the ordinary three-velocity v^i for which $u^\mu = (u^0, u^i) = (\gamma, \gamma v^i)$ and $\gamma = 1/\sqrt{1 - |\mathbf{v}|^2}$, we have

$$|\mathbf{u}| = \frac{|\mathbf{v}|}{\sqrt{1 - |\mathbf{v}|^2}} \sim \frac{1}{a}. \tag{31}$$

We again see why the comoving frame is the natural frame. Consider an observer initially ($t = t_1$) moving non-relativistically with respect to the comoving frame with physical three velocity of magnitude $|\mathbf{v}|_1$. At a later time t_2 the magnitude of the observer's physical three-velocity $|\mathbf{v}|_2$ will be

$$|\mathbf{v}|_2 = |\mathbf{v}|_1 \frac{a(t_1)}{a(t_2)}. \tag{32}$$

In an expanding universe, the freely-falling observer is destined to come to rest in the comoving frame even if he has some initial velocity with respect to it.

1.4 The cosmological redshift

Without explicitly solving Einstein's equations for the dynamics of the expansion, it is still possible to understand many of the kinematic effects of the expansion upon light from distant galaxies. The light emitted by a distant object can be viewed quantum mechanically as freely-propagating photons, or classically as propagating plane waves. In the quantum mechanical description, the wavelength of light is in- inversely proportional to the photon momentum $\lambda = h/p$. If the momentum changes, the wavelength of the light must change. It was shown in the previous section that the momentum of a photon changes in proportion to a^{-1}. Since the wavelength of a photon is inversely proportional to its momentum, the wavelength at time t_0, denoted as λ_0, will differ from that at time t_1, denoted as λ_1, by

$$\boxed{\frac{\lambda_1}{\lambda_0} = \frac{a(t_1)}{a(t_0)}.} \tag{33}$$

As the universe expands, the wavelength of a freely-propagating photon increases, just as all physical distances increase with the expansion. This means that the red shift of the wavelength of a photon is due to the fact that the universe was smaller when the photon was emitted.

It is also possible to derive the same result by considering the propagation of light from a distant galaxy as a classical wave phenomenon. Let us again place ourselves at the origin $r = 0$. We consider a radially travelling electro-magnetic wave (a light ray) and consider the equation $\mathrm{d}s^2 = 0$ or

$$\mathrm{d}t^2 = a^2(t)\frac{\mathrm{d}r^2}{1 - kr^2}\,. \tag{34}$$

Let us assume that the wave leaves a galaxy located at r at time t. Then it will reach us at time t_0 given by

$$\int_t^{t_0} \frac{\mathrm{d}t}{a(t)} = f(r) = \int_0^r \frac{\mathrm{d}r}{\sqrt{1 - kr^2}} = \begin{cases} \sin^{-1} r = r + r^3/6 + \cdots & (k = +1), \\ r & (k = 0), \\ \sinh^{-1} r = r - r^3/6 + \cdots & (k = -1). \end{cases} \tag{35}$$

As typical galaxies will have constant coordinates, $f(r)$ (which can of course be given explicitly, but this is not needed for the present analysis) is time-independent. If the next wave crest leaves the galaxy at r at time $(t + \delta t)$, it will arrive at time $(t_0 + \delta t_0)$ given by

$$f(r) = \int_0^r \frac{\mathrm{d}r}{\sqrt{1 - kr^2}} = \int_{t+\delta t}^{t_0 + \delta t_0} \frac{\mathrm{d}t}{a(t)}\,. \tag{36}$$

Subtracting these two equations and making the (eminently reasonable) assumption that the cosmic scale factor $a(t)$ does not vary significantly over the period δt given by the frequency of light, we obtain

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t}{a(t)}. \tag{37}$$

Therefore the observed frequency ν_0 is related to the emitted frequency ν by

$$\frac{\nu_0}{\nu} = \frac{a(t)}{a(t_0)}. \tag{38}$$

Astronomers like to express this in terms of the red-shift parameter

$$\boxed{1 + z = \frac{\lambda_0}{\lambda},} \tag{39}$$

which implies

$$\boxed{z = \frac{a(t_0)}{a(t)} - 1}. \tag{40}$$

Thus if the universe expands one has $z > 0$ and ther is a red-shift while in a contracting universe with $a(t_0) < a(t)$ the light of distant glaxies would be blue-shifted. A few remarks:

1. This cosmological red-shift has nothing to do with the stars own gravitational field - that contribution to the red-shift is completely negligible compared to the effect of the cosmological red-shift.

2. Unlike the gravitational red-shift i GR, this cosmological red-shift is symmetric between receiver and emitter, *.e.* light sent from the earth to the distant galaxy would likewise be red-shifted if we observe a red-shift of the distant galaxy.

3. This red-shift is a combined effect of gravitational and Doppler red-shifts and it is not very meaningful to interpret this only in terms of, say, a Doppler shift. Nevertheless, as mentioned before, astronomers like to do just that, calling $v = zc$ the recessional velocity.

4. Nowadays, astronomers tend to express the distance of a galaxy not in terms of light-years or megaparsecs, but directly in terms of the observed red-shift factor z, the conversion to distance then following from some version of Hubbles law.

5. The largest observed redshift of a galaxy is currently $z \sim 10$, corresponding to a distance of the order of 13 billion light-years, while the cosmic microwave background radiation, which originated just a couple of 10^5 years after the Big Bang, has $z \sim 10^3$.

2 Standard cosmology

All of the discussions in the previous section concerned the kinematics of a universe described by a FRW. The dynamics of the expanding universe only appeared implicitly in the time dependence of the scale factor $a(t)$. To make this time dependence explicit, one must solve for the evolution of the scale factor using the Einstein equations

$$\boxed{R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G_{\rm N} T_{\mu\nu} + \Lambda g_{\mu\nu},} \tag{41}$$

where $T_{\mu\nu}$ is the stress-energy tensor for all the fields present (matter, radiation, and so on) and we have also included the presence of a cosmological constant. With very minimal assumptions about the right-hand side of the Einstein equations, it is possible to proceed without detailed knowledge of the properties of the fundamental fields that contribute to the stress tensor $T_{\mu\nu}$.

2.1 The stress-energy momentum tensor

To be consistent with the symmetries of the metric, the total stress-energy tensor tensor must be diagonal, and by isotropy the spatial components must be equal. The simplest realization of such a stress-energy tensor is that of a perfect fluid characterized by a time-dependent energy density $\rho(t)$ and pressure $P(t)$

$$\boxed{T^{\mu}_{\nu} = (\rho + P) u^{\mu} u_{\nu} + P \delta^{\mu}_{\nu} = \mathrm{diag}(-\rho, P, P, P),} \tag{42}$$

where $u^{\mu} = (1, 0, 0, 0)$ in a comoving coordinate system. This is precisely the energ-ymomentum tensor of a perfect fluid. The four-vector u_{μ} is known as the velocity field of the fluid, and the comoving coordinates are those with respect to which the fluid is at rest. In general, this matter content has to be supplemented by an equation of state. This is usually assumed to be that of a barytropic fluid, *i.e.* one whose pressure depends only on its density, $P = P(\rho)$. The most useful toy-models of cosmological fluids arise from considering a linear relationship between P and ρ of the type

$$\boxed{P = w\rho}, \tag{43}$$

where w is known as the equation of state parameter. Occasionally also more exotic equations of state are considered. For non-relativistic particles (NR) particles, there is no pressure, $p_{\rm NR} = 0$, *i.e.* $w_{\rm NR} = 0$, and such matter is usually referred to as dust. The trace of the energy-momentum tensor is

$$T^{\mu}_{\mu} = -\rho + 3P. \tag{44}$$

For relativistic particles, radiation for example, the energy-momentum tensor is (like that of Maxwell theory) traceless, and hence relativistic particles hve the equation of state

$$P_{\rm r} = \frac{1}{3}\rho_{\rm r}, \tag{45}$$

and thus $w_{\rm r} = 1/3$. For physical (gravitating instead of anti-gravitating) matter one usually requires $\rho > 0$ (positive energy) and either $P > 0$, corresponding to $w > 0$ or, at least, $(\rho + 3P) > 0$, corresponding to the weaker condition $w > 1/3$. A cosmological constant, on the other hand, corresponds, as we will see, to a matter contribution with $w_\Lambda = -1$ and thus violates either $\rho > 0$ or $(\rho + 3P) > 0$.

Let us now turn to the conservation laws associated with the energy-momentum tensor,

$$\nabla_\mu T^{\mu\nu} = 0. \tag{46}$$

The spatial components of this conservation law,

$$\nabla_\mu T^{\mu i} = 0, \tag{47}$$

turn out to be identically satisfied, by virtue of the fact that the u^μ are geodesic and that the functions ρ and P are only functions of time. This could hardly be otherwise because $\nabla_\mu T^{\mu i}$ would have to be an invariant vector, and we know that there are none. Nevertheless it is instructive to check this explicitly

$$\nabla_\mu T_{\mu i} = \nabla_0 T^{0i} + \nabla_j T^{ji} = 0 + \nabla_j T^{ji} = P\nabla_j g^{ij} = 0. \tag{48}$$

The only interesting conservation law is thus the zero-component

$$\nabla_\mu T^{\mu 0} = \partial_\mu T^{\mu 0} + \Gamma^\mu{}_{\mu\nu} T^{\nu 0} + \Gamma^0{}_{\mu\nu} T^{\mu\nu} = 0, \tag{49}$$

which for a perfect fluid becomes

$$\dot\rho + \Gamma^\mu{}_{\mu 0}\rho + \Gamma^0{}_{00}\rho + \Gamma^0{}_{ij} T^{ij} = 0. \tag{50}$$

Using the Christoffel symbols previously computed, see Eq. (3), we get

$$\boxed{\dot\rho + 3H(\rho + P) = 0}. \tag{51}$$

For instance, when the pressure of the cosmic matter is negligible, like in the universe today, and we can treat the galaxies (without disrespect) as dust, then one has

$$\boxed{\rho_{\rm NR}\, a^3 = \text{constant} \quad \text{(MATTER)}}. \tag{52}$$

The energy (number) density scales like the inverse of the volume whose size is $\sim a^3$ On the other hand, if the universe is dominated by, say, radiation, then one has the equation of state $P = \rho/3$, then

$$\boxed{\rho_{\rm r}\, a^4 = \text{constant} \quad (\text{RADIATION}).} \tag{53}$$

The energy density scales the like the inverse of the volume (whose size is $\sim a^3$) and the energy which scales like $1/a$ because of the red-shift: photon energies scale like the inverse of their wavelenghts which in turn scale like $1/a$. More generally, for matter with equation of state parameter w, one finds

$$\boxed{\rho\, a^{3(1+w)} = \text{constant}.} \tag{54}$$

In particular, for $w = -1$, ρ is constant and corresponds, as we will see more explicitly below, to a cosmological constant vacuum energy

$$\boxed{\rho_\Lambda = \text{constant} \quad (\text{VACUUM ENERGY}).} \tag{55}$$

The early universe was radiation dominated, the adolescent universe was matter dominated and the adult universe is dominated, as we shall see, by the cosmological constant. If the universe underwent inflation, there was again a very early period when the stress-energy was dominated by vacuum energy. As we shall see next, once we know the evolution of ρ and P in terms of the scale factor $a(t)$, it is straightforward to solve for $a(t)$. Before going on, we want to emphasize the utility of describing the stress energy in the universe by the simple equation of state $P = w\rho$. This is the most general form for the stress energy in a FRW space-time and the observational evidence indicates that on large scales the RW metric is quite a good approximation to the space-time within our Hubble volume. This simple, but often very accurate, approximation will allow us to explore many early universe phenomena with a single parameter.

2.2 The Friedmann equations

After these preliminaries, we are now prepared to tackle the Einstein equations. We allow for the presence of a cosmological constant and thus consider the equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_{\rm N} T_{\mu\nu}. \tag{56}$$

It will be convenient to rewrite these equations in the form

$$R_{\mu\nu} = 8\pi G_{\rm N} \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T^\lambda_\lambda \right) + \Lambda g_{\mu\nu}. \tag{57}$$

Because of isotropy, there are only two independent equations, namely the 00-component and any one of the non-zero ij-components. Using Eqs. (5) we find

$$\begin{aligned} -3\frac{\ddot{a}}{a} &= 4\pi G_{\rm N}(\rho + 3P) - \Lambda\,, \\ \frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{k}{a^2} &= 4\pi G_{\rm N}(\rho - P) + \Lambda. \end{aligned} \tag{58}$$

Using the first equation to eliminate $\ddot{a}$ from the second, one obtains the set of equations for the Hubble rate

$$\boxed{H^2 + \frac{k}{a^2} = \frac{8\pi G_{\rm N}}{3}\rho + \frac{\Lambda}{3}}, \tag{59}$$

for the acceleration

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G_{\rm N}}{3}(\rho + 3P) - \frac{\Lambda}{3}}. \tag{60}$$

Together, this set of equation is known as the Friedman equations. They are supplemented this by the conservation equation (51). Note that because of the Bianchi identities, the Einstein equations and the conservation equations should not be independent, and indeed they are not. It is easy to see that (59) and (51) imply the second order equation (60) so that, a pleasant simplification, in practice one only has to deal with the two first order equations (59) and (51). Sometimes, however, (60) is easier to solve than (59), because it is linear in $a(t)$, and then (59) is just used to fix one constant of integration.

Notice that Eqs. (59) and (60) can be obtained, in the non-relativistic limit $P = 0$ from Newtonian physics. Imagine that the distribution of matter is uniform and its matter density is ρ. Put a test particle with mass m on a surface of a sphere of radius a and let gravity act. The total energy is constant and therefore

$$E_{\rm kin} + E_{\rm pot} = \frac{1}{2}m\dot{a}^2 - G_{\rm N}\frac{mM}{a} = \kappa = \text{constant}. \tag{61}$$

Since the mass M contained in a sphere of radius a is $M = (4\pi\rho a^3/3)$, we obtain

$$\frac{1}{2}m\dot{a}^2 - \frac{4\pi G_{\rm N}}{3}ma^2 = \kappa = \text{constant}. \tag{62}$$

By dividing everything by $(ma^2/2)$ we obtain Eq. (59) with of course no cosmological constant and after setting $k = 2\kappa/m$. Eq. (60) can be analogously obtained from Newton's law relating the gravitational force and the acceleration (but still with $P = 0$).

The expansion rate of the universe is determined by the Hubble rate H which is not a constant and generically scales like t^{-1}. The Friedmann equation (59) can be recast as

$$\boxed{\Omega - 1 = \frac{\rho}{3H^2/8\pi G_{\rm N}} = \frac{k}{a^2H^2}}, \tag{63}$$

where we have defined the parameter Ω as the ratio between the energy density ρ and the critical energy density $\rho_{\rm c}$

$$\boxed{\Omega = \frac{\rho}{\rho_{\rm c}}, \quad \rho_{\rm c} = \frac{3H^2}{8\pi G_{\rm N}}.} \tag{64}$$

Since $a^2H^2 > 0$, there is a correspondence between the sign of k and the sign of $(\Omega - 1)$

$$\begin{aligned} k = +1 \quad &\Rightarrow \quad \Omega > 1 \quad \text{CLOSED}, \\ k = 0 \quad &\Rightarrow \quad \Omega = 1 \quad \text{FLAT}, \\ k = -1 \quad &\Rightarrow \quad \Omega < 1 \quad \text{OPEN}. \end{aligned} \tag{65}$$

Eq. (63) is valid at all times, note also that both Ω and $\rho_{\rm c}$ are not constant in time. At early times once has a radiation-dominated (RD) phase radiation and $H^2 \sim a^{-4}$ with $(\Omega - 1) \sim a^2$; during the matter-dominated phase (MD) one finds $H^2 \sim a^{-3}$ with $(\Omega - 1) \sim a$. These relations will be crucial when we will study the inflationary universe. The present day value of the critical energy density is

$$\rho_{\rm c} = 1.87\, h^2 \cdot 10^{-29}\, {\rm gr\, cm^{-3}} = 1.05 h^2 \cdot 10^4\, {\rm eV\, cm^{-3}} = 8.1 h^2 \times 10^{-47}\, {\rm GeV}^4. \tag{66}$$

It is also common practice to define the Ω parameters for all the components of the universe

$$\Omega_i = \frac{\rho_i}{\rho_{\rm c}}, \quad (i = \text{MATTER}, \text{RADIATION}, \text{VACUUM}, \text{ENERGY}, \cdots). \tag{67}$$

If we define

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{\rm c}} = \frac{\Lambda}{8\pi G_{\rm N}}\frac{1}{\rho_{\rm c}} = \frac{\Lambda}{3H^2}, \tag{68}$$

and a curvature density parameter

$$\Omega_k = -\frac{k}{H^2a^2}, \tag{69}$$

we can obtain the so-called golden rule of cosmology

$$\boxed{\Omega_{\rm m} + \Omega_\gamma + \Omega_{\rm b} + \Omega_\Lambda + \Omega_k + \cdots = 1.} \tag{70}$$

We have indicated here with the subscript $_{\rm m}$ the dark matter (DM) (see below) and $_{\rm b}$ the baryons (ordinary matter). Present day values do no carry the index $_0$ unless needed for the clarity of the presentation. From each discussion it should be clear when we intend that the Ω parameters are at the present epoch or at a generic instant of time. In particular, the Fridmann equation (59) can be wriiten as

$$\boxed{H^2 = H_0^2(\Omega_{\rm m} a^{-3} + \Omega_{\rm r} a^{-4} + \Omega_\Lambda + \Omega_k a^{-2} + \cdots),} \tag{71}$$

where we have set $a_0 = 1$. In the previous section we have also introduced the deceleration parameter, see Eq. (**??**). By combining Eqs. (59) and (60) and using the definition of Ω_0, that is the value of the parameter Ω today, it follows that

$$\boxed{q_0 = -\frac{4\pi G_{\rm N}}{3H_0^2}\sum_i \rho_i(1+3w_i) \simeq \frac{1}{2}\left(\Omega_{\rm m} + 2\Omega_{\rm r} - 2\Omega_\Lambda\right).} \tag{72}$$

For a MD universe we have $q_0 = \Omega_{\rm m}/2$, for a RD we have $q_0 = \Omega_{\rm r}$, both positive. Nevertheless, for a vacuum-dominated universe, we obtain $q_0 = -\Omega_\Lambda$ and the universe is indeed accelerating.

Recall also that from (6) the curvature of the three-dimensioanl spatial slices is ${}^3\mathcal{R} = 6k/a^2$. Using the definition of Ω we obtain

$${}^3\mathcal{R} = \frac{6H^2}{\Omega - 1}. \tag{73}$$

From the FRW metric, it is clear that the effect of the curvature becomes important only at a comoving radius $r \sim |k|^{-1/2}$. So we define the physical radius of curvature of the universe $R_{\rm curv} = a(t)|k|^{-1/2} = (6/|{}^3\mathcal{R}|)^{1/2}$, related to the Hibble radius H^{-1} by

$$R_{\rm curv} = \frac{H^{-1}}{|\Omega - 1|^{1/2}}. \tag{74}$$

When $|\Omega - 1| \ll 1$, such a curvature radius turns out to be much larger than the Hubble radius and we can safely neglect the effect of curvature in the universe. Note also that for closed universes, $k = +1$, $R_{\rm curv}$ is just the physical radius of the three-sphere.

2.3 Exact solutions of the Friedman-Robertson-Walker Cosmology

To solve the Friedman equations we have to account for the presence of several species of matter, characterised by different equations of state or different equation of state parameters w_i will coexist. If we assume that these do not interact, then one can just add up their contributions in the Friedman equations.

In order to make the dependence of the Friedman equation (59)

$$\dot{a}^2 = \frac{8\pi G_{\rm N}}{3}\rho a^2 - k + \frac{\Lambda}{3}a^2 \tag{75}$$

on the equation of state parameters w_i more manifest, it is useful to use the conservation law

$$\frac{8\pi G_{\rm N}}{3}\rho_i a^2 = c_i a^{-(1+3w_i)}, \tag{76}$$

for some constant c_i. Then the Friedman equation takes the more explicit form

$$\dot{a}^2 = \sum_i c_i a^{-(1+3w_i)} - k + \frac{\Lambda}{3}a^2. \tag{77}$$

In addition to the vacuum energy (and pressure), there are typically two other kinds of matter which are relevant in our approximation, namely matter in the form of dust and radiation. Denoting the corresponding constants by $c_{\rm m}$ and $c_{\rm r}$ respectively, the Friedman equation that we will be dealing with takes the form

$$\dot{a}^2 = \frac{c_{\rm m}}{a} + \frac{c_{\rm r}}{a^2} - k + \frac{\Lambda}{3}a^2, \tag{78}$$

illustrating the qualitatively different conntributions to the time-evolution. One can then characterise the different eras in the evolution of the universe by which of the above terms dominates, i.e. gives the leading contribution to the equation of motion for a. This already gives some insight into the physics of the situation. We will call a universe matter-diminated if the piece $c_{\rm m}/a$ dominates; radiation-dominated if the piece $c_{\rm r}/a^2$ dominates; curvature-dominated if the piece k dominates and vacuum-dominated if the piece Λa^2 dominates. As mentioned before, for a long time it was believed that our present universe is purely matter dominated while recent observations appear to indicate that contributions from both matter and the cosmological constant are non-negligible.

Here are some immediate consequences of the Friedman equation (78):

1. No matter how small $c_{\rm r}$ is, provided that it is non-zero, for sufficiently small values of a that term will dominate and one is in the radiation dominated era. In that case, one finds the characteristic behaviour

$$\dot{a}^2 = \frac{c_{\rm r}}{a^2} \Rightarrow a(t) \sim t^{1/2} \quad \text{(RD)}. \tag{79}$$

2. If matter dominates, one finds the characteristic behaviour

$$\dot{a}^2 = \frac{c_{\rm m}}{a} \Rightarrow a(t) \sim t^{2/3} \quad \text{(MD)}. \tag{80}$$

3. For a general equation of state $w \neq -1$, one finds

$$\boxed{a(t) \sim t^{\frac{2}{3(1+w)}}}. \tag{81}$$

4. For sufficiently large a, a nonzero cosmological constant will always dominate, no matter how small the cosmological constant may be, as all the other energy-content of the universe gets more and more diluted.

5. Only for $\Lambda = 0$ does k dominates for large a.

6. Finally, for $\Lambda = 0$ the Friedman equation can be integrated in terms of elementary functions whereas for $\Lambda \neq 0$ one typically encounters elliptic integrals.

7. If we extrapolate at $t = 0$, we see that the scale factor vanishes there and the energy density becomes infinite. This is a mathematical, rather than physical singularity and goes under the name of Big Bang. In practice, if the inflationary cosmology is correct, we are not really interested in such a epoch as there is no observation which could test it, simply inflation erased any information about it. Let us also stress that, because of the Cosmological Principle, such singularity should have happened everywhere uniformly.

Let us study, for hystorical reasons, the so-called Einstein static solution. Einstein was looking for a static cosmological solution and for this he was forced to introduce the cosmological constant. Static means $\dot{a} = 0$. Energy conservation then tells us that $\dot{\rho} = 0$ and Eq. (60) tells us that $(\rho + 3P) = \Lambda$, therefore also P must be a constant. We see that with $\Lambda = 0$ we would already not be able to satisfy this equation for physical matter content $(\rho + 3P) > 0$. Furthermore Eq. (78) indicates that k must be positive. Finally, going back to Eq. (59) we find

$$a^2 = (8\pi G_{\rm N}\rho/3 + 8\pi G_{\rm N}(\rho + 3P)/3)^{-1} = (4\pi G_{\rm N}(\rho + P))^{-1}\,. \tag{82}$$

This is thus a static universe, with topology $\mathbb{R} \times S^3$ in which the gravitational attraction is precisely balanced by the cosmological constant. Note that even though a positive cosmological constant has a positive energy density, it has a negative pressure, and the net effect of a positive cosmological constant is that of gravitational repulsion rather than attraction.

In the matter-dominated universe we have to solve

$$\dot{a}^2 = \frac{c_{\rm m}}{a} - k. \tag{83}$$

For $k = 0$ this is the equation (79) we already discussed and goes under the name of Einstein-de Sitter universe. For $k = +1$, the equation is

$$\dot{a}^2 = \frac{c_{\rm m}}{a} - 1. \tag{84}$$

We recall that in this case we will have a recollapsing universe with $a_{\rm max} = c_{\rm m}$, which is attained for $\dot{a} = 0$. This can be solved in closed form for t as a function of a, and the solution to

$$\frac{\mathrm{d}t}{\mathrm{d}a} = \left(\frac{a}{a_{\max} - a}\right)^{-1/2} \tag{85}$$

is

$$t(a) = \frac{a_{\max}}{2} \arccos\left(1 - 2\frac{a}{a_{\max}}\right) - \left(a a_{\max} - a^2\right)^{1/2}, \tag{86}$$

as can be easily verified. The universe starts at $t = 0$ with $a(0) = 0$, reaches its maximum at $a = a_{\max}$ at

$$t_{\max} = a_{\max} \arccos(-1)/2 = \frac{\pi}{2}\, a_{\max}, \tag{87}$$

and ends in a Big Crunch at $t = 2t_{\max}$. 2tmax. The curve $a(t)$ is a cycloid, as is most readily seen by writing the solution in parametrised form. For this it is convenient to introduce the time-coordinate τ via

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{1}{a(t)}. \tag{88}$$

As an aside, not that this time-coordinate renders the FRW metric conformal to Minkowski $\mathrm{d}s^2 = a^2(\tau)(-\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2)$. This coordinate system is very convenient for discussing the causal structure of the FRW universes. In terms of the parameter τ, the solution to the Friedman equation for $k = +1$ can be written as

$$\begin{aligned} a(\tau) &= \frac{a_{\max}}{2}(1 - \cos\tau)\,, \\ t(\tau) &= \frac{a_{\max}}{2}(\tau - \sin\tau)\,, \end{aligned} \tag{89}$$

which makes it transparent that the curve is indeed a cycloid. The maximal radius is reached at

$$t_{\max} = t(a = a_{\max}) = t(\tau = \pi) = \frac{\pi}{2} a_{\max}, \tag{90}$$

as before. Analogously, for $k == -1$, the Friedman equation can be solved in parametrised form, with the trignometric functions replaced by hyperbolic functions

$$\begin{aligned} a(\tau) &= \frac{a_{\max}}{2}(\cosh\tau - 1)\,, \\ t(\tau) &= \frac{a_{\max}}{2}(\sinh\tau - \tau)\,, \end{aligned} \tag{91}$$

In the case in which radiation is dominating the equation to solve is

$$a^2 \dot{a}^2 = c_{\mathrm{r}} - k a^2. \tag{92}$$

It is convenient to make the change of variable $b = a^2$ to obtain

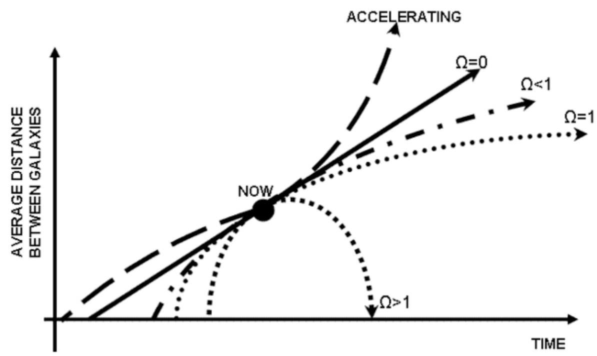

Figure 3: Qualitative behaviour of the Friedman-Robertson-Walker models for $\Lambda = 0$.

$$\frac{\dot{b}^2}{4} + kb = c_{\mathrm{r}}. \tag{93}$$

For $k = 0$ we already saw that the solution is $a(t) \sim t^{1/2}$. For $k = \pm 1$, one necessarily has $b(t) = b_0 + b_1 t + b_2 t^2$. Fixing $b(0) = 0$ we find the solution

$$a(t) = \left(2c_{\mathrm{r}}^{1/2} t - kt\right)^{1/2}, \tag{94}$$

so, for $k = +1$

$$a(0) = a\left(2c_{\mathrm{r}}^{1/2}\right) = 0. \tag{95}$$

Thus already electro-magnetic radiation is sufficient to shrink the universe again and make it recollapse. For $k = -1$ on the other hand, the universe expands forever. All this is of course in agreement with the results of the qualitative discussion given earlier.

Finally and for future applications, let us see what happens when the universe is dominated by a cosmological constant. The equation to solve is

$$\dot{a}^2 = -k + \frac{\Lambda}{3} a^2. \tag{96}$$

We see that Λ has to be positive for $k = +1$ or $k = 0$, whereas for $k = -1$ both positive and negative λ are possible. This is one instance where the solution to the second order equation (60)

$$\ddot{a} = \frac{\Lambda}{3} a, \tag{97}$$

is more immediate, namely trigonometric functions for $\Lambda < 0$ (only possible for $k =?1$) and hyperbolic functions for $\Lambda > 0$. The first order equation then fixes the constants of integration according to the value of k. For $k = 0$, the solution is

$$a_\pm(t) = \sqrt{\frac{3}{\Lambda}}\, e^{\pm\sqrt{\Lambda/3}\,t}, \tag{98}$$

and for $k = +1$, thus $\Lambda > 0$, one has

$$a_\pm(t) = \sqrt{\frac{3}{\Lambda}}\,\cosh\,\sqrt{\Lambda/3}\,t. \tag{99}$$

This is also known as the de Sitter universe. It is a maximally symmetric (in space-time) solution of the Einstein equations with a cosmological constant and thus has a metric of constant curvature. But we know that such a metric is unique. Hence the three solutions with $\lambda > 0$, for $k = 0, \pm 1$ must all represent the same space-time metric, only in different coordinate systems. This is interesting because it shows that de Sitter space is so symmetric that it has space-like slicings by three-spheres, by three-hyperboloids and by three-planes. The solution for $k = -1$ involves $\sin\sqrt{\Lambda/3}t$ for $\Lambda < 0$ and $\sinh\sqrt{\Lambda/3}t$ for $\Lambda > 0$. The former is known as the anti de Sitter universe.

Part II

Equilibrium thermodynamics

Today the radiation, or relativistic particles, in the universe is comprised of the 2.75 K microwave photons and the three cosmic seas of about 1.96 K relic neutrinos. because the early universe was to a good approximation in thermal equilibrium, there should have been other relativistic particles present, with comparable abundances. Before going on to discuss the early RD phase, we will quickly review some basic thermodynamics.

The number density n, energy density ρ and pressure P of a dilute, weakly interacting gas of particles with g internal degrees of freedom is given in terms of its phase space distribution function $f(\mathbf{p})$

$$\begin{aligned} n &= \frac{g}{(2\pi)^3}\int \mathrm{d}^3p\, f(\mathbf{p}), \\ \rho &= \frac{g}{(2\pi)^3}\int \mathrm{d}^3p\, E(\mathbf{p})\, f(\mathbf{p}), \\ P &= \frac{g}{(2\pi)^3}\int \mathrm{d}^3p\, \frac{|\mathbf{p}|^2}{3E(\mathbf{p})} f(\mathbf{p}), \end{aligned} \tag{100}$$

where $E^2 = |\mathbf{p}|^2 + m^2$. For a species in kinetic equilibrium, the phase occupancy f is given by the familiar Fermi-Dirac or Bose-Einstein distributions

$$\boxed{f(\mathbf{p}) = \left[\exp(E-\mu)/T \pm 1\right]^{-1}}\,, \tag{101}$$

where μ is the chemical potential and $+1$ refers to Fermi-Dirac species and -1 to Bose-Einstein species. Moreover, if the species is in chemical equilibrium, then its chemical potential μ is related to the chemical potentials of other species with which it interacts. For example, if the species i interacts with the species j, k and l

$$i + j \leftrightarrow k + l, \tag{102}$$

then we have

$$\mu_i + \mu_j = \mu_k + \mu_l. \tag{103}$$

From the equilibrium distributions, it follows that the number density n, energy density ρ and pressure P of a species of mass m, chemical potential μ and temperature T are

$$\begin{aligned} n &= \frac{g}{2\pi^2}\int_m^\infty \mathrm{d}E\, E\, \frac{(E^2-m^2)^{1/2}}{\exp\left[(E-\mu)/T\right]\pm 1}, \\ \rho &= \frac{g}{2\pi^2}\int_m^\infty \mathrm{d}E\, E^2\, \frac{(E^2-m^2)^{1/2}}{\exp\left[(E-\mu)/T\right]\pm 1}, \\ P &= \frac{g}{6\pi^2}\int_m^\infty \mathrm{d}E \frac{(E^2-m^2)^{3/2}}{\exp\left[(E-\mu)/T\right]\pm 1}. \end{aligned} \tag{104}$$

In the relativistic limit $T \gg m$ and $T \gg \mu$ we obtain

$$\begin{aligned} \rho &= \begin{cases} (\pi^2/30)gT^4 & \text{(BOSE)} \\ (7/8)(\pi^2/30)gT^4 & \text{(FERMI)} \end{cases} \\ n &= \begin{cases} (\zeta(3)/\pi^2)gT^3 & \text{(BOSE)} \\ (3/4)(\zeta(3)/\pi^2)gT^3 & \text{(FERMI)} \end{cases} \\ P &= \rho/3. \end{aligned} \tag{105}$$

For a degenerate gas for which $\mu \gg T$ we have

$$\begin{aligned} \rho &= (1/8\pi^2)g\mu^4, \\ n &= (1/6\pi^2)g\mu^3, \\ P &= (1/24\pi^2)g\mu^4. \end{aligned} \tag{106}$$

Here $\zeta(3) \simeq 1.2$ is the Riemann zeta function of three. For a Bose-Einstein species $\mu > 0$ indicates the presence of a Bose condensate, which should be treated separately from the other modes. For relativistic bosons or fermions with $\mu < 0$ and $|\mu| < T$, it follows that

$$
\begin{aligned}
n &= \exp(\mu/T)(g/\pi^2)T^3, \\
\rho &= \exp(\mu/T)(3g/\pi^2)T^4, \\
P &= \exp(\mu/T)(g/\pi^2)T^4.
\end{aligned} \tag{107}
$$

In the non-relativistic limit, $m \gg T$, the number density and pressure are the same for the Bose and Fermi species

$$
\begin{aligned}
n &= g\left(\frac{mT}{2\pi}\right)^{3/2} \exp[-(m-\mu)/T], \\
\rho &= m\,n, \\
P &= n\,T \ll \rho.
\end{aligned} \tag{108}
$$

For a nondegenerate relativistic species, its average energy density per particle is

$$
\begin{aligned}
\langle E\rangle &= \rho/n = [\pi^4/30\zeta(3)] \simeq 2.7\,T \text{ (BOSE)}, \\
\langle E\rangle &= \rho/n = [7\pi^4/180\zeta(3)] \simeq 3.15\,T \text{ (FERMI)}.
\end{aligned} \tag{109}
$$

For a degenerate, relativistic species

$$
\langle E\rangle = \rho/n = (3\mu/4). \tag{110}
$$

Finally, for a non-relativistic particle

$$
\langle E\rangle = m + (3/2)\,T. \tag{111}
$$

The excess of a fermionic species over its antiparticle is often of interest and can be computed in the relativistic and non-relativistic regimes. Assuming that $\mu_+ = -\mu_-$ (true if the reactions like particle + antiparticles $\leftrightarrow \gamma + \gamma$ occur rapidly), then the net fermion number is

$$
\begin{aligned}
n_+ - n_- &= \frac{g}{2\pi^2}\int_m^\infty \mathrm{d}E\,E(E^2-m^2)^{1/2} \\
&\times \left[\frac{1}{\exp[(E-\mu)/T]+1} - \frac{1}{\exp[(E+\mu)/T]+1}\right] \\
&= \frac{gT^3}{6\pi^2}\left[\pi^2\left(\frac{\mu}{T}\right) + \left(\frac{\mu}{T}\right)^3\right] \;\; (T \gg m), \\
&= 2g(mT/2\pi)^{3/2}\sinh(\mu/T)\,\exp(-m/T) \;\; (T \ll m).
\end{aligned} \tag{112}
$$

The total energy density and pressure of all species in equilibrium can be expressed in terms of the photon temperature T

$$\begin{aligned}
\rho_{\rm r} &= T^4 \sum_{\rm all} {\rm species} \left(\frac{T_i}{T}\right)^4 \frac{g_i}{2\pi^2} \int_{x_i}^{\infty} {\rm d}u\, \frac{(u^2 - x_i^2)^{1/2} u^2}{\exp(u - y_i) \pm 1}, \\
P_{\rm r} &= T^4 \sum_{\rm all} {\rm species} \left(\frac{T_i}{T}\right)^4 \frac{g_i}{6\pi^2} \int_{x_i}^{\infty} {\rm d}u\, \frac{(u^2 - x_i^2)^{3/2}}{\exp(u - y_i) \pm 1},
\end{aligned} \tag{113}$$

where $x_i = m_i/T$ and $y_i = /\mu_i/T$ and we taken into account the possibility that the species have a different temperature than the photons.

Since the energy density and pressure of non-relativistic species is exponentially smaller than that of relativistic species, it is a very good approximation to include only the relativistic species in the sums and we obtain

$$\boxed{\rho_{\rm r} = 3P_{\rm r} = \frac{\pi^2}{30} g_*(T)\, T^4,} \tag{114}$$

where

$$\boxed{g_*(T) = \sum_{\rm bosons} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{\rm fermions} g_i \left(\frac{T_i}{T}\right)^4,} \tag{115}$$

counts the effective total number of relativistic degrees of freedom in the plasma. For instance, for $T \ll$ MeV, the only relativistic species are the three neutrinos with $T_\nu = (4/11)^{1/3} T_\gamma$ (see below) and $g_*(\ll \text{ MeV}) \simeq 3.36$. For 100 MeV $\gtrsim T \gtrsim 1$ MeV, the electron and positron are additional relativistic degrees of freeedom and $T_\nu = T_\gamma$ and $g_* \simeq 10.75$. For $T \gtrsim 300$ GeV, all the species of the Standard Model (SM) are in equilibrium: 8 gluons, $W^\pm$, Z, three generations of quarks and leptons and one complex Higgs field and $g_* \simeq 106.75$.

During early RD phase when $\rho \simeq \rho_{\rm r}$ and supposing that $g_* \simeq$ constant, we have that the Hubble rate is

$$\boxed{H \simeq 1.66\, g_*^{1/2} \frac{T^2}{M_{\rm Pl}}} \tag{116}$$

and the corresponding time is

$$\boxed{t \simeq 0.3\, g_*^{-1/2} \frac{M_{\rm Pl}}{T^2} \simeq \left(\frac{T}{\rm MeV}\right)^{-2} \text{ sec.}} \tag{117}$$

3 Entropy

Throughout most of the history of the universe the reaction rates of particles in the thermal bath were much greater than the expansion rate of the universe and local thermal equilibrium (LTE) was attained. In this case the entropy per comoving volume element remains constant. The entropy in a comoving volume provides a very useful quantity during the expansion of the universe. The second law of theormodynamics as applied to a comoving volume elementof unit coordinate volume and physical volume $V = a^3$, implies that (we assume small chemical potentials)

$$T\mathrm{d}S = \mathrm{d}(\rho V) + P\mathrm{d}V = \mathrm{d}[(\rho + P)V] - V\mathrm{d}P. \tag{118}$$

The integrability condition

$$\frac{\partial^2 S}{\partial T \partial V} = \frac{\partial^2 S}{\partial V \partial T} \tag{119}$$

relates the energy density and pressure

$$T\frac{\mathrm{d}P}{\mathrm{d}T} = \rho + P, \tag{120}$$

or, equivalently,

$$\mathrm{d}P = \frac{\rho + P}{T}\mathrm{d}T. \tag{121}$$

We therefore obtain from Eq. (118) that

$$\mathrm{d}S = \frac{1}{T}\mathrm{d}[(\rho + P)V] - (\rho + P)V\frac{\mathrm{d}T}{T^2} = \mathrm{d}\left[\frac{(\rho + P)V}{T} + \mathrm{const.}\right]. \tag{122}$$

That is, up to an additional constant, the entropy per comoving volume is

$$S = a^3\frac{(\rho + P)}{T}. \tag{123}$$

Reacll that the first law (energy conservation) can be written as

$$\mathrm{d}[(\rho + P)V] = V\mathrm{d}P. \tag{124}$$

Thus substituting (121) in Eq. (124), we obtain

$$\mathrm{d}\left[\frac{(\rho + P)V}{T}\right] = 0. \tag{125}$$

This implies that in thermal equilibrium the entropy per comoving volume V, S, is conserved. It is useful to define the entropy density s as

$$s = \frac{S}{V} = \frac{\rho + P}{T}. \tag{126}$$

It is dominated by the relativistic degrees of freedom and to a very good approximation

$$\boxed{s = \frac{2\pi^2}{45} g_{*S} T^3,} \tag{127}$$

where

$$\boxed{g_{*S} = g_*(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T}\right)^3,} \tag{128}$$

For most of the history of the universe, all particles had the same temperature and one replace therefore g_{*S} with g_*. Note also that s is proportional to the number density of relativistic degrees of freedom and in particular it can be related to the photon number density n_γ

$$\boxed{s \simeq 1.8 g_{*S} n_\gamma.} \tag{129}$$

Today $s_0 \simeq 7.04 n_{\gamma,0}$. The conservation of S implies that $s \sim a^{-3}$ and therefore

$$\boxed{g_{*S} T^3 a^3 = \text{constant}} \tag{130}$$

during the evolution of the universe and that the number density of a given species $Y = a^3 n$ can be written as

$$\boxed{Y = \frac{n}{s}.} \tag{131}$$

For a species in thermal equilbrium

$$\begin{aligned} Y &= \frac{45\zeta(3)g}{2\pi^4 g_{*S}} \; (T \gg m), \\ Y &= \frac{45g}{4\sqrt{2}\pi^5 g_{*S}} (m/T)^{3/2} \exp(-m/T + \mu/T) \; (T \ll m). \end{aligned} \tag{132}$$

If the number of a given species in a comoving volume is not changing, *i.e.* particles are neither created nor destroyed, then Y remains constant at a given temperature. For instance, as long as the baryon number processes are out-of-equilibrium, then $n_{\rm b}/s$ is conserved. Although $\eta = n_{\rm b}/\nu_\gamma = 1.8 g_{*S}(n_{\rm b}/s)$, the baryon number-to-phton ratio does not remain constant with time because g_{*S} chages. During the era of $e^\pm$ annihilations, the number density of photons per comoving volume increases by a factor 11/4, so that η decreases by the same factor. After the time of $e^\pm$ annihilations, however, g_{*S} is constant and $\eta \simeq 7 n_{\rm b}/s$ and $n_{\rm b}/s$ can be used interchangeably.

The second fact, that $S = g_{*S} T^3 a^3$ = constant, implies that the temperature of the universe evolves as

$$\boxed{T \sim g_{*S}^{-1/3} a^{-1}.} \tag{133}$$

When g_{*S} is constant one gets the familiar result $T \sim a^{-1}$. The factor $g_{*S}^{-1/3}$ enters because whenever a particle species becomes non-relativistic and disappears from the plasma, its entropy is transferred to the other relativistic particles in the thermal plasma causing T to decrease slightly less slowly (sometimes it is said, but in a wrong way, that the universe slightly reheats up).

Part III

The inflationary cosmology

In this chapter we will discuss the inflationary universe. As we will come out along the way, inflation is responsible not only for the observed homogeneity and isotropy of the universe, but also for its inhomogeneities. Furthermore, inflation links the quantum mechanical microphysics to the the macrophysics of the universe as a whole. It is a beautiful example of connection between high energy physics and cosmology.

Before launching ourselves into the description of inflation, we would like to go back to the concept of conformal time which will be useful in the next sections. The conformal time τ is defined through the following relation

$$\mathrm{d}\tau = \frac{\mathrm{d}t}{a}. \tag{134}$$

The metric $\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\mathbf{x}^2$ then becomes

$$\mathrm{d}s^2 = a^2(\tau)\left[-\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2\right]. \tag{135}$$

The reason why τ is called conformal is manifest from Eq. (135): the corresponding FRW line element is conformal to the Minkowski line element describing a static four dimensional hypersurface.

Any function $f(t)$ satisfies the rule

$$\dot{f}(t) = \frac{f'(\tau)}{a(\tau)}, \tag{136}$$

$$\dot{f}(t) = \frac{f''(\tau)}{a^2(\tau)} - \mathcal{H}\frac{f'(\tau)}{a^2(\tau)}, \tag{137}$$

where a prime now indicates differentation wrt to the conformal time τ and

$$\boxed{\mathcal{H} = \frac{a'}{a}}. \tag{138}$$

In particular we can set the following rules

$$
\begin{aligned}
H &= \frac{\dot{a}}{a} = \frac{a'}{a^2} = \frac{\mathcal{H}}{a}, \\
\dot{a} &= \frac{a''}{a^2} - \frac{\mathcal{H}^2}{a}, \\
\dot{H} &= \frac{\mathcal{H}'}{a^2} - \frac{\mathcal{H}^2}{a^2}, \\
H^2 &= \frac{8\pi G\rho}{3} - \frac{k}{a^2} \Longrightarrow \mathcal{H}^2 = \frac{8\pi G\rho a^2}{3} - k \\
\dot{H} &= -4\pi G\,(\rho + P) \Longrightarrow \mathcal{H}' = -\frac{4\pi G}{3}\,(\rho + 3P)\,a^2, \\
\dot{\rho} &+ 3H(\rho + P) = 0 \Longrightarrow \rho' + 3\,\mathcal{H}(\rho + P) = 0
\end{aligned}
$$

Finally, if the scale factor $a(t)$ scales like $a \sim t^n$, solving the relation (134) we find

$$a \sim t^n \Longrightarrow a(\tau) \sim \tau^{\frac{n}{1-n}}. \tag{139}$$

Therefore, for a RD era $a(t) \sim t^{1/2}$ one has $a(\tau) \sim \tau$ and for a MD era $a(t) \sim t^{2/3}$, that is $a(\tau) \sim \tau^2$.

4 Again on the concept of particle horizon

We have already encountered the concept of the particle horizon. Let us see how it behaves in an expanding universe and what this implies. In spite of the fact that the universe was vanishingly small at early times, the rapid expansion precluded causal contact from being established throughout. Photons travel on null paths characterized by $\mathrm{d}s^2 = 0$ or (along straight lines in polar coordinates) $\mathrm{d}r = \mathrm{d}t/a(t)$; the physical distance that a photon could have traveled since the bang until time t, the distance to the particle horizon, is

$$
\begin{aligned}
R_H(t) &= a(t)\int_0^t \frac{\mathrm{d}t'}{a(t')} = a(\tau)\int_{\tau_0}^{\tau} \mathrm{d}\tau' \\
&= \frac{t}{(1-n)} = n\,\frac{H^{-1}}{(1-n)} \sim H^{-1} \qquad \text{for } a(t) \propto t^n, \quad n < 1.
\end{aligned} \tag{140}
$$

Recall that in a universe dominated by a fluid with equation of state $P = w/\rho$ we have $n = 2/3(1+w)$. The comoving Hubble radius goes like

$$\text{COMOVING HUBBLE RADIUS} = \frac{1}{aH} \sim \frac{t}{t^n} = t^{1-n} \tag{141}$$

In particular, for a MD universe $w = 0$ and $n = 2/3$, while for a RD universe $w = 1/3$ and $1/2$. In bot cases the comoving Huble radius increases with time. We see that in the standard cosmology the distance to the horizon is finite, and up to numerical factors, equal to the Hubble radius, H^{-1}.

For this reason, one can use the words horizon and Hubble radius interchangeably for standard cosmology. As we shall see, in inflationary models the horizon and Hubble radius are drastically different as the horizon distance grows exponentially relative to the Hubble radius; in fact, at the end of inflation they differ by e^N, where N is the number of e-folds of inflation. The horizon sets the length scale for which two points separated by a distance larger than $R_H(t)$ they could never communicate, while the Hubble radius sets the scale at which these two points could not comunicate at the time t.

Note also that a physical length scale λ is within the Hubble radius if $\lambda < H^{-1}$. Since we can identify the length scale λ with its wavenumber k, $\lambda = 2\pi a/k$, we will have the following rule

$$
\boxed{
\begin{aligned}
\frac{k}{aH} &\ll 1 \Longrightarrow \text{SCALE } \lambda \text{ OUTSIDE THE HORIZON} \\
\frac{k}{aH} &\gg 1 \Longrightarrow \text{SCALE } \lambda \text{ WITHIN THE HORIZON}
\end{aligned}
}
$$

Notice that in standard cosmology

$$
\frac{\lambda}{\text{PARTICLE HORIZON}} = \frac{\lambda}{R_H} = \lambda\, H \sim \frac{aH}{k}. \tag{142}
$$

This shows once more that Hubble radius and particle horizon can be used interchangeably in standard cosmology.

5 The shortcomings of the Standard Big-Bang Theory

By now the shortcomings of the standard cosmology are well appreciated: the horizon or large-scale smoothness problem; the small-scale inhomogeneity problem (origin of density perturbations); and the flatness or oldness problem. We will only briefly review them here. They do not indicate any logical inconsistencies of the standard cosmology; rather, that very special initial data seem to be required for evolution to a universe that is qualitatively similar to ours today. Nor is inflation the first attempt to address these shortcomings: over the past two decades cosmologists have pondered this question and proposed alternative solutions. Inflation is a solution based upon well-defined, albeit speculative, early universe microphysics describing the post-Planck epoch.

5.1 The Flatness Problem

Let us make a tremendous extrapolation and assume that Einstein equations are valid until the Plank era, when the temperature of the universe is $T_{\rm Pl} \sim\sim 10^{19}$ GeV. From the equation for the curvature

$$\Omega - 1 = \frac{k}{H^2 a^2}, \tag{143}$$

we read that if the universe is perfectly flat, then ($\Omega = 1$) at all times. On the other hand, if there is even a small curvature term, the time dependence of ($\Omega - 1$) is quite different.

During a RD period, we have that $H^2 \propto \rho_{\rm r} \propto a^{-4}$ and

$$\Omega - 1 \propto \frac{1}{a^2 a^{-4}} \propto a^2. \tag{144}$$

During MD, $\rho_{\rm NR} \propto a^{-3}$ and

$$\Omega - 1 \propto \frac{1}{a^2 a^{-3}} \propto a. \tag{145}$$

In both cases ($\Omega - 1$) decreases going backwards with time. Since we know that today ($\Omega_0 - 1$) is of order unity at present, we can deduce its value at $t_{\rm Pl}$ (the time at which the temperature of the universe is $T_{\rm Pl} \sim 10^{19}$ GeV)

$$\frac{\mid \Omega - 1 \mid_{T=T_{\rm Pl}}}{\mid \Omega - 1 \mid_{T=T_0}} \approx \left(\frac{a_{\rm Pl}^2}{a_0^2}\right) \approx \left(\frac{T_0^2}{T_{\rm Pl}^2}\right) \approx \mathcal{O}(10^{-64}). \tag{146}$$

where 0 stands for the present epoch, and $T_0 \sim 10^{-13}$ GeV is the present-day temperature of the CMB radiation. If we are not so brave and go back simply to the epoch of nucleosynthesis when light elements abundances were formed, at $T_N \sim 1$ MeV, we get

$$\frac{\mid \Omega - 1 \mid_{T=T_N}}{\mid \Omega - 1 \mid_{T=T_0}} \approx \left(\frac{a_N^2}{a_0^2}\right) \approx \left(\frac{T_0^2}{T_N^2}\right) \approx \mathcal{O}(10^{-16}). \tag{147}$$

In order to get the correct value of $(\Omega_0 - 1) \sim 1$ at present, the value of ($\Omega - 1$) at early times have to be fine-tuned to values amazingly close to zero, but without being exactly zero. This is the reason why the flatness problem is also dubbed the 'fine-tuning problem'.

5.2 The Entropy Problem

Let us now see how the hypothesis of adiabatic expansion of the universe is connected with the flatness problem. From the Friedman equations we know that during a RD period

$$H^2 \simeq \rho_R \simeq \frac{T^4}{M_{\rm Pl}^2}, \tag{148}$$

from which we deduce

$$\Omega - 1 = \frac{k M_{\rm Pl}^2}{a^4 T^4} = \frac{k M_{\rm Pl}^2}{S^{\frac{2}{3}} T^2}. \tag{149}$$

Under the hypothesis of adiabaticity, S is constant over the evolution of the universe and therefore

$$|\Omega - 1|_{t=t_{\rm Pl}} = \frac{M_{\rm Pl}^2}{T_{\rm Pl}^2} \frac{1}{S_{\rm U}^{2/3}} = \frac{1}{S_{\rm U}^{2/3}} \approx 10^{-60}, \tag{150}$$

where we have used the fact that the present horizon contains a total entropy

$$S_{\rm U} = \frac{4\pi}{3} H_0^{-3} s = \frac{4\pi}{3} H_0^{-3} \frac{2\pi^2 g_* T^3}{45} \simeq 10^{90}. \tag{151}$$

We have discovered that $(\Omega - 1)$ is so close to zero at early epochs because the total entropy of our universe is so incredibly large. The flatness problem is therefore a problem of understanding why the (classical) initial conditions corresponded to a universe that was so close to spatial flatness. In a sense, the problem is one of fine–tuning and although such a balance is possible in principle, one nevertheless feels that it is unlikely. On the other hand, the flatness problem arises because the entropy in a comoving volume is conserved. It is possible, therefore, that the problem could be resolved if the cosmic expansion was non–adiabatic for some finite time interval during the early history of the universe.

5.3 The horizon problem

According to the standard cosmology, photons decoupled from the rest of the components (electrons and baryons) at a temperature of the order of 0.3 eV. This corresponds to the so-called surface of 'last-scattering' at a red shift of about 1100 and an age of about $180,000\,(\Omega_0 h^2)^{-1/2}$ yrs. From the

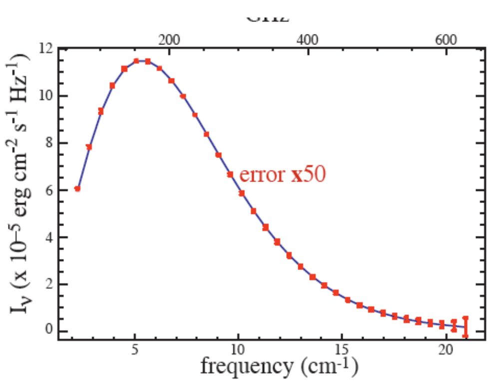

Figure 4: The black body spectrum of the cosmic background radiation.

epoch of last-scattering onwards, photons free-stream and reach us basically untouched. Detecting primordial photons is therefore equivalent to take a picture of the universe when the latter was about 300,000 yrs old. The spectrum of the cosmic background radiation is consistent that of a black body at temperature 2.73 K over more than three decades in wavelength; see Fig. 4. The length corresponding to our present Hubble radius (which is approximately the radius of our observable universe) at the time of last-scattering was

$$\lambda_{\rm H}(t_{\rm ls}) = R_{\rm H}(t_0)\left(\frac{a_{\rm ls}}{a_0}\right) = R_{\rm H}(t_0)\left(\frac{T_0}{T_{\rm ls}}\right).$$

On the other hand, during the MD period, the Hubble length has decreased with a different law

$$H^2 \propto \rho_{\rm NR} \propto a^{-3} \propto T^3.$$

At last-scattering

$$H_{\rm ls}^{-1} = R_{\rm H}(t_0)\left(\frac{T_{\rm ls}}{T_0}\right)^{-3/2} \ll R_{\rm H}(t_0).$$

The length corresponding to our present Hubble radius was much larger that the horizon at that time. This can be shown comparing the volumes corresponding to these two scales

$$\frac{\lambda_H^3(T_{\rm ls})}{H_{\rm ls}^{-3}} = \left(\frac{T_0}{T_{\rm ls}}\right)^{-\frac{3}{2}} \approx 10^6. \tag{152}$$

There were $\sim 10^6$ casually disconnected regions within the volume that now corresponds to our horizon! It is difficult to come up with a process other than an early hot and dense phase in the history of the universe that would lead to a precise black body for a bath of photons which were causally disconnected the last time they interacted with the surrounding plasma.

The horizon problem is well represented by Fig. 5 where the green line indicates the horizon scale and the red line any generic physical length scale λ. Suppose, indeed that λ indicates the distance between two photons we detect today. From Eq. (152) we discover that at the time of emission (last-scattering) the two photons could not talk to each other, the red line is above the green line.

There is another aspect of the horizon problem which is related to the problem of initial conditions for the cosmological perturbations. We have every indication that the universe at early times, say $t \ll 300,000\,$yrs, was very homogeneous; however, today inhomogeneity (or structure) is ubiquitous: stars ($\delta\rho/\rho \sim 10^{30}$), galaxies ($\delta\rho/\rho \sim 10^5$), clusters of galaxies ($\delta\rho/\rho \sim 10 - 10^3$), superclusters, or "clusters of clusters" ($\delta\rho/\rho \sim 1$), voids ($\delta\rho/\rho \sim -1$), great walls, and so on. For some twenty-five years the standard cosmology has provided a general framework for understanding this picture. Once the universe becomes matter dominated (around 1000 yrs after the bang) primeval density inhomogeneities ($\delta\rho/\rho \sim 10^{-5}$) are amplified by gravity and grow into the structure we see today. The existence of density inhomogeneities has another important consequence: fluctuations in the temperature of the CMB radiation of a similar amplitude. The temperature difference measured between two points separated by a large angle ($\gtrsim 1^\circ$) arises due to a very simple physical effect: the difference in the gravitational potential between the two points on the last-scattering surface, which in turn is related to the density perturbation, determines the temperature anisotropy on the angular scale subtended by that length scale,

$$\left(\frac{\delta T}{T}\right)_\theta \approx \left(\frac{\delta\rho}{\rho}\right)_\lambda, \tag{153}$$

where the scale $\lambda \sim 100 h^{-1}\,{\rm Mpc}(\theta/{\rm deg})$ subtends an angle θ on the last-scattering surface. This is known as the Sachs-Wolfe effect. The CMB experiments looking for the tiny anisotropies are of three

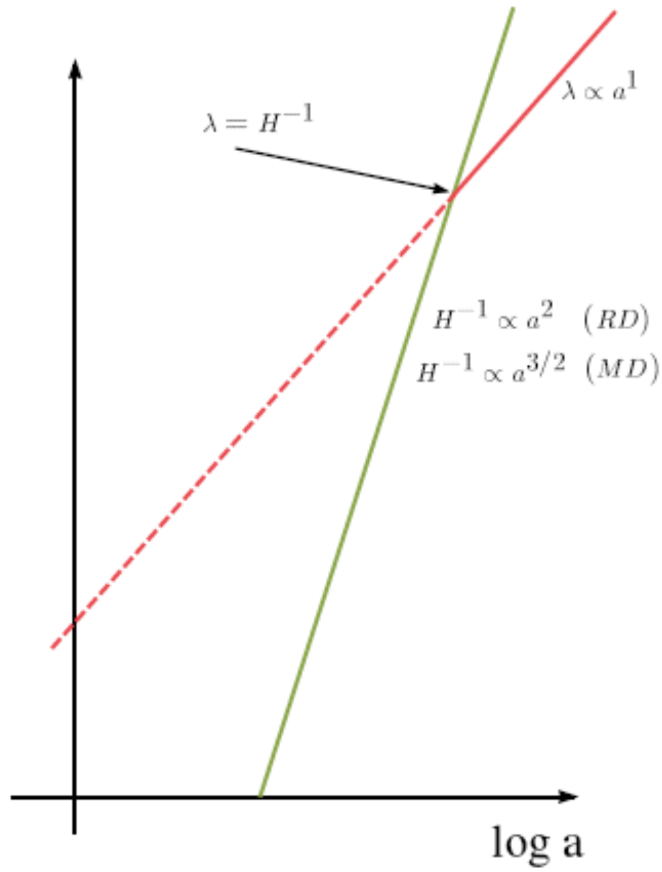

Figure 5: The horizon scale (green line) and a physical scale λ (red line) as function of the scale factor a.

kinds: satellite experiments, balloon experiments, and ground based experiments. The technical and economical advantages of ground based experiments are evident, but their main problem is atmospheric fluctuations. The temperature anisotropy is commonly expanded in spherical harmonics

$$\frac{\Delta T}{T}(x_0, \tau_0, \mathbf{n}) = \sum_{\ell m} a_{\ell,m}(x_0) Y_{\ell m}(\mathbf{n}), \tag{154}$$

where x_0 and τ_0 are our position and the preset time, respectively, $\mathbf{n}$ is the direction of observation, ℓ's are the different multipoles and[1]

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell,\ell'} \delta_{m,m'} C_\ell, \tag{155}$$

[1]An alternative definition is $C_\ell = \langle |a_{\ell m}|^2 \rangle = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2$.

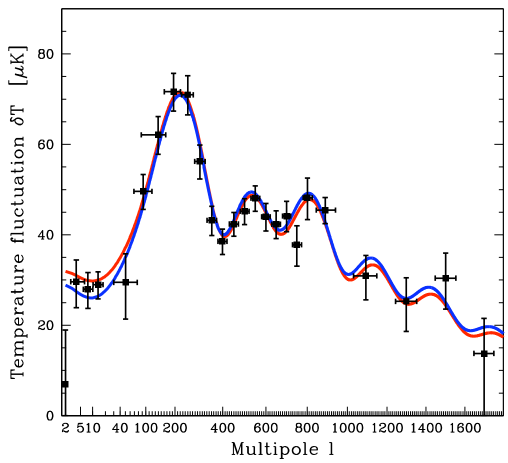

Figure 6: The CMBR anisotropy as function of ℓ.

where the deltas are due to the fact that the process that created the anisotropy is statistically isotropic. The C_ℓ are the so-called CMB power spectrum. For homogeneity and isotropy, the C_ℓ's are neither a function of x_0, nor of m. The two-point-correlation function is related to the C_ℓ's in the following way

$$\begin{aligned} \left\langle \frac{\delta T(\mathbf{n})}{T} \frac{\delta T(\mathbf{n}')}{T} \right\rangle &= \sum_{\ell\ell' mm'} \langle a_{\ell m} a^*_{\ell' m'} \rangle Y_{\ell m}(\mathbf{n}) Y^*_{\ell' m'}(\mathbf{n}') \\ &= \sum_\ell C_\ell \sum_m Y_{\ell m}(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}') = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\mu = \mathbf{n}\cdot\mathbf{n}') \end{aligned} \tag{156}$$

where we have used the addition theorem for the spherical harmonics, and P_ℓ is the Legendre polynom of order ℓ. In expression (156) the expectation value is an ensamble average. It can be regarded as an average over the possible observer positions, but not in general as an average over the single sky we observe, because of the cosmic variance[2].

Let us now consider the last-scattering surface. In comoving coordinates the latter is 'far' from us a distance equal to

$$\int_{t_{\rm ls}}^{t_0} \frac{\mathrm{d}t}{a} = \int_{\tau_{\rm ls}}^{\tau_0} \mathrm{d}\tau = (\tau_0 - \tau_{\rm ls})\,. \tag{157}$$

[2]The usual hypothesis is that we observe a typical realization of the ensamble. This means that we expect the difference between the observed values $|a_{\ell m}|^2$ and the ensamble averages C_ℓ to be of the order of the mean-square deviation of $|a_{\ell m}|^2$ from C_ℓ. The latter is called cosmic variance and, because we are dealing with a Gaussian distribution, it is equal to $2C_\ell$ for each multipole ℓ. For a single ℓ, averaging over the $(2\ell+1)$ values of m reduces the cosmic variance by a factor $(2\ell+1)$, but it remains a serious limitation for low multipoles.

A given comoving scale λ is therefore projected on the last-scattering surface sky on an angular scale

$$\theta \simeq \frac{\lambda}{(\tau_0 - \tau_{\rm ls})}, \tag{158}$$

where we have neglected tiny curvature effects. Consider now that the scale λ is of the order of the comoving sound horizon at the time of last-scattering, $\lambda \sim c_s \tau_{\rm ls}$, where $c_s \simeq 1/\sqrt{3}$ is the sound velocity at which photons propagate in the plasma at the last-scattering. This corresponds to an angle

$$\theta \simeq c_s \frac{\tau_{\rm ls}}{(\tau_0 - \tau_{\rm ls})} \simeq c_s \frac{\tau_{\rm ls}}{\tau_0}, \tag{159}$$

where the last passage has been performed knowing that $\tau_0 \gg \tau_{\rm ls}$. Since the universe is MD from the time of last-scattering onwards, the scale factor has the following behaviour: $a \sim T^{-1} \sim t^{2/3} \sim \tau^2$, where we have made use of the relation (139). The angle $\theta_{\rm HOR}$ subtended by the sound horizon on the last-scattering surface then becomes

$$\theta_{\rm HOR} \simeq c_s \left(\frac{T_0}{T_{\rm ls}}\right)^{1/2} \sim 1^\circ, \tag{160}$$

where we have used $T_{\rm ls} \simeq 0.3$ eV and $T_0 \sim 10^{-13}$ GeV. This corresponds to a multipole $\ell_{\rm HOR}$

$$\ell_{\rm HOR} = \frac{\pi}{\theta_{\rm HOR}} \simeq 200. \tag{161}$$

From these estimates we conclude that two photons which on the last-scattering surface were separated by an angle larger than $\theta_{\rm HOR}$, corresponding to multipoles smaller than $\ell_{\rm HOR} \sim 200$ were not in causal contact. On the other hand, from Fig. 6 it is clear that small anisotropies, of the *same* order of magnitude $\delta T/T \sim 10^{-5}$ are present at $\ell \ll 200$. We conclude that one of the striking features of the CMB fluctuations is that they appear to be noncausal. Photons at the last-scattering surface which were causally disconnected have the same small anisotropies! The existence of particle horizons in the standard cosmology precludes explaining the smoothness as a result of microphysical events: the horizon at decoupling, the last time one could imagine temperature fluctuations being smoothed by particle interactions, corresponds to an angular scale on the sky of about 1°, which precludes temperature variations on larger scales from being erased.

To account for the small-scale lumpiness of the universe today, density perturbations with horizon-crossing amplitudes of 10^{-5} on scales of $1\,{\rm Mpc}$ to $10^4\,{\rm Mpc}$ or so are required. As can be seen in Fig. 5, in the standard cosmology the physical size of a perturbation, which grows as the scale factor, begins larger than the horizon and relatively late in the history of the universe crosses inside the horizon. This precludes a causal microphysical explanation for the origin of the required density perturbations.

From the considerations made so far, it appears that solving the shortcomings of the standard Big Bang theory requires two basic modifications of the assumptions made so far:

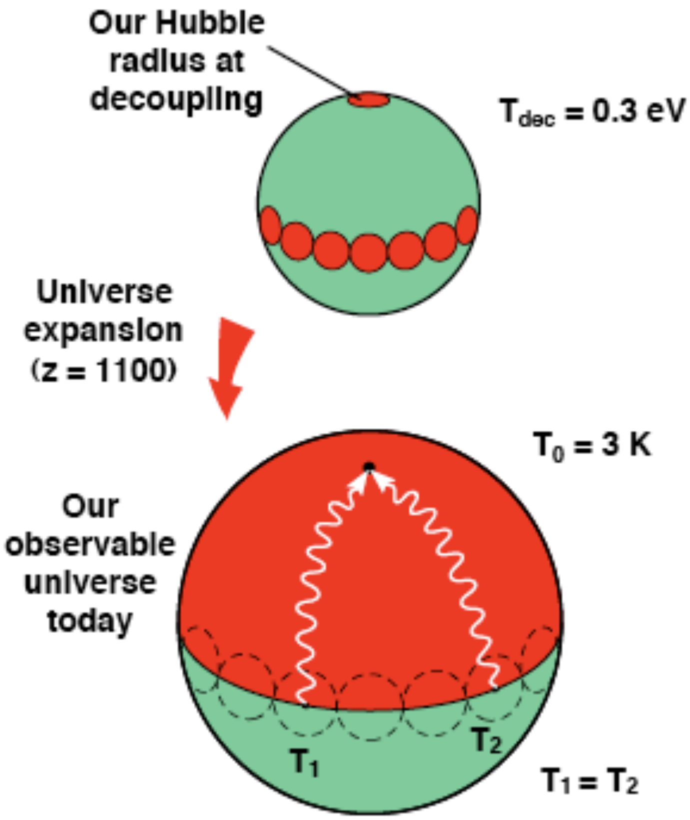

Figure 7: An illustration of the horizon problem stemming from the CMB anisotropy.

- The universe has to go through a non-adiabatic period. This is necessary to solve the entropy and the flatness problem. A non-adiabatic phase may give rise to the large entropy $S_{\rm U}$ we observe today.

- The universe has to go through a primordial period during which the physical scales λ evolve faster than the Hubble radius H^{-1}.

The second condition is obvious from Fig. 8. If there is period during which physical length scales grow faster than the Hubble radius H^{-1}, length scales λ which are within the horizon today, $\lambda < H^{-1}$ (such as the distance between two detected photons) and were outside the Hubble radius at some period, $\lambda > H^{-1}$ (for istance at the time of last-scattering when the two photons were emitted), had a chance to be within the Hubble radius at some primordial epoch, $\lambda < H^{-1}$ again. If this happens, the homogeneity and the isotropy of the CMB can be easily explained: photons that we receive today and were emitted from the last-scattering surface from causally disconnected regions have the same temperature because they had a chance to talk to each other at some primordial stage of the evolution of the universe. The distinction between the (comoving) particle horizon and

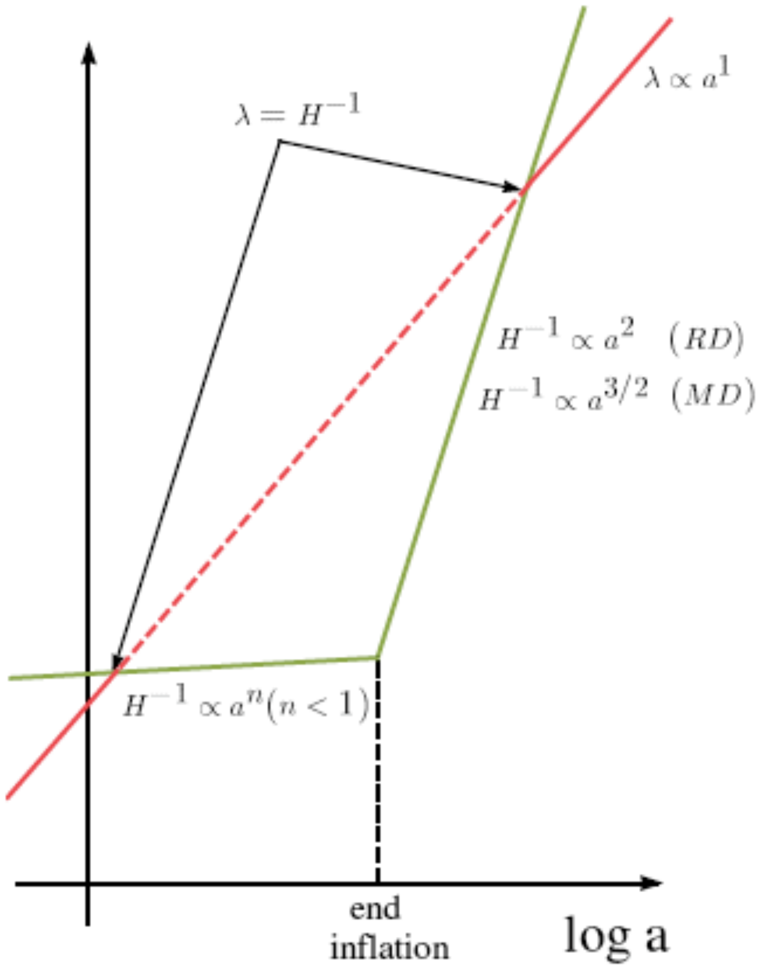

Figure 8: The behaviour of a generic scale λ and the Hubble radius H^{-1} in the standard inflationary model.

the (comoving) Hubble radius is crucial now for the solution to the horizon problem which relies on the following: It is possible that R_H is much larger than the Hubble radius now, so that particles cannot communicate today but were in causal contact early on.

The second condition can be easily expressed as a condition on the scale factor a. Since a given scale λ scales like $\lambda \sim a$ and the Hubble radius $H^{-1} = a/\dot{a}$, we need to impose that there is a period during which

$$\left(\frac{\lambda}{H^{-1}}\right)^{\cdot} > 0 \Rightarrow \ddot{a} > 0. \tag{162}$$

Notice that is equivalent to require that the ratio between the comoving length scales λ/a the comoving Hubble radius during inflation

$$\left(\frac{\lambda}{H^{-1}}\right)^{\cdot} = \left(\frac{\lambda/a}{H^{-1}/a}\right)^{\cdot} = \left(\frac{\lambda/a}{1/aH}\right)^{\cdot} > 0. \tag{163}$$

increases with time. We can therefore introduced the following rigorous definition: an inflationary stage is a period of the universe during which the latter accelerates

$$\text{INFLATION} \quad \Longleftrightarrow \quad \ddot{a} > 0.$$

<u>*Comment*</u>: Let us stress that during such a accelerating phase the universe expands *adiabatically*. This means that during inflation one can exploit the usual FRW equations. It must be clear therefore that the non-adiabaticity condition is satisfied not during inflation, but during the phase transition between the end of inflation and the beginning of the RD phase. At this transition phase a large entropy is generated under the form of relativistic degrees of freedom: the Big Bange has taken place.

6 The standard inflationary universe

From the previous section we have learned that an accelerating stage during the primordial phases of the evolution of the universe might be able to solve the horizon problem. Therefore we obtain we learn that

$$\dot{a} > 0 \Longleftrightarrow (\rho + 3P) < 0.$$

An accelerating period is obtainable only if the overall pressure p of the universe is negative: $P < -\rho/3$. Neither a RD phase nor a MD phase (for which $P = \rho/3$ and $P = 0$, respectively) satisfy such a condition. Let us postpone for the time being the problem of finding a 'candidate' able to provide the condition $P < -\rho/3$. For sure, inflation is a phase of the history of the universe occurring before the era of nucleosynthesis ($t \approx 1$ sec, $T \approx 1$ MeV) during which the light elements abundances were formed. This is because nucleosynthesis is the earliest epoch we have experimental data from and they are in agreement with the predictions of the standard Big-Bang theory. However, the thermal history of the universe before the epoch of nucleosynthesis is unknown.

In order to study the properties of the period of inflation, we assume the extreme condition $P = -\rho$ which considerably simplifies the analysis. A period of the universe during which $P = -\rho$ is called *de Sitter* stage. By inspecting the FRW equations and the energy conservation equation, we learn that during the de Sitter phase

$$\begin{aligned}\rho &= \text{constant},\\ H_{\rm I} &= \text{constant},\end{aligned}$$

where we have indicated by $H_{\rm I}$ the value of the Hubble rate during inflation. Correspondingly, we obtain

$$a = a_{\rm I}\, e^{H_{\rm I}(t-t_{\rm I})}, \tag{164}$$

where $t_{\rm I}$ denotes the time at which inflation starts. Let us now see how such a period of exponential expansion takes care of the shortcomings of the standard Big Bang Theory.[3]

6.1 Inflation and the horizon Problem

During the inflationary (de Sitter) epoch the horizon scale $H_{\rm I}^{-1}$ is constant. If inflation lasts long enough, all the physical scales that have left the Hubble radius during the RD or MD phase can re-enter the Hubble radius in the past: this is because such scales are exponentially reduced. Indeed, while during inflation the particle horizon grow exponential

$$R_H(t) = a(t)\int_{t_{\rm I}}^{t}\frac{{\rm d}t'}{a(t')} = a_{\rm I}\, e^{H_{\rm I}(t-t_{\rm I})}\left(-\frac{1}{H_{\rm I}}\right)\left[e^{-H_{\rm I}(t-t_{\rm I})}\right]_{t_{\rm I}}^{t} \simeq \frac{a(t)}{H_{\rm I}}, \tag{165}$$

while the Hubble radius remains constant

$$\text{HUBBLE RADIUS} = \frac{a}{\dot{a}} = H_{\rm I}^{-1}, \tag{166}$$

and points that our causally disconnected today could have been in contact during inflation. Notice that in comoving coordinates the comoving Hubble radius shrink exponentially

$$\text{COMOVING HUBBLE RADIUS} = H_{\rm I}^{-1} e^{-H_{\rm I}(t-t_{\rm I})}, \tag{167}$$

while comoving length scales remain constant. An illustration of the solution to the horizon problem can therefore be visualized as in Fig. 9. As we have seen in the previous section, this explains both the problem of the homogeneity of CMB and the initial condition problem of small cosmological perturbations. Once the physical length is within the horizon, microphysics can act, the universe can be made approximately homogeneous and the primaeval inhomogeneities can be created.

[3]Despite the fact that the growth of the scale factor is exponential and the expansion is *superluminal*, this is not in contradiction with what dictated by relativity. Indeed, it is the space-time itself which is progating so fast and not a light signal in it.

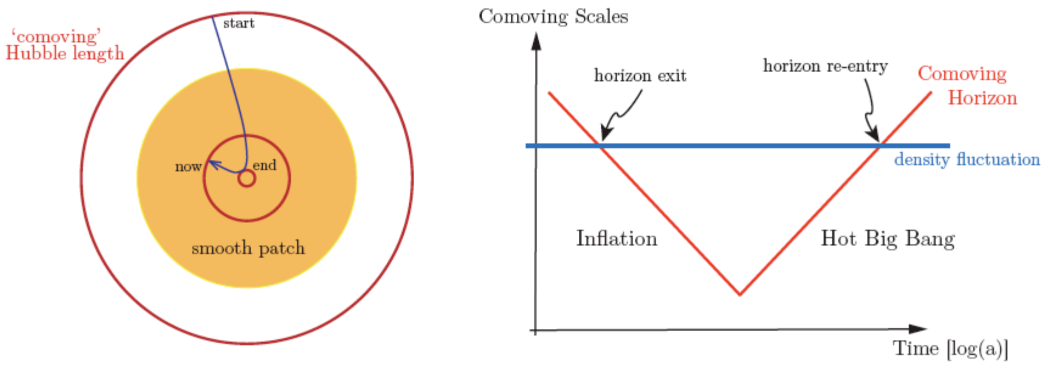

Figure 9: The solution of the horizon problem by inflation in comoving coordinates

Let us see how long inflation must be sustained in order to solve the horizon problem. Let $t_{\rm I}$ and $t_{\rm f}$ be, respectively, the time of beginning and end of inflation. We can define the corresponding number of e-foldings N

$$N = \ln\left[H_{\rm I}(t_{\rm e} - t_{\rm I})\right]. \tag{168}$$

A necessary condition to solve the horizon problem is that the largest scale we observe today, the present horizon H_0^{-1}, was reduced during inflation to a value $\lambda_{H_0}(t_{\rm I})$ smaller than the value of horizon length $H_{\rm I}^{-1}$ during inflation. This gives

$$\lambda_{H_0}(t_{\rm I}) = H_0^{-1}\left(\frac{a_{t_{\rm f}}}{a_{t_0}}\right)\left(\frac{a_{t_{\rm I}}}{a_{t_{\rm f}}}\right) = H_0^{-1}\left(\frac{T_0}{T_{\rm f}}\right)e^{-N} \lesssim H_{\rm I}^{-1},$$

where we have neglected for simplicity the short period of MD and we have called $T_{\rm f}$ the temperature at the end of inflation (to be indentified with the reheating temperature $T_{\rm RH}$ at the beginning of the RD phase after inflation, see later). We get

$$N \gtrsim \ln\left(\frac{T_0}{H_0}\right) - \ln\left(\frac{T_{\rm f}}{H_{\rm I}}\right) \approx 67 + \ln\left(\frac{T_{\rm f}}{H_{\rm I}}\right).$$

Apart from the logarithmic dependence, we obtain $N \gtrsim 70$.

6.2 Inflation and the flateness problem

Inflation solves elegantly the flatness problem. Since during inflation the Hubble rate is constant

$$\Omega - 1 = \frac{k}{a^2H^2} \propto \frac{1}{a^2}.$$

On the other end the condition (150) tells us that to reproduce a value of $(\Omega_0 - 1)$ of order of unity today the initial value of $(\Omega - 1)$ at the beginning of the RD phase must be $|\Omega - 1| \sim 10^{-60}$. Since we identify the beginning of the RD phase with the beginning of inflation, we require

$$|\Omega - 1|_{t=t_{\rm f}} \sim 10^{-60}.$$

During inflation

$$\frac{|\Omega - 1|_{t=t_{\rm f}}}{|\Omega - 1|_{t=t_{\rm I}}} = \left(\frac{a_{\rm I}}{a_{\rm f}}\right)^2 = e^{-2N}. \tag{169}$$

Taking $|\Omega - 1|_{t=t_{\rm I}}$ of order unity, it is enough to require that $N \approx 70$ to solve the flatness problem.

1. Comment: In the previous section we have written that the flateness problem can be also seen as a fine-tuning problem of one part over 10^{60}. Inflation ameliorates this fine-tuning problem, by explaining a tiny number $\sim 10^{-60}$ with a number N of the order 70.

2. Comment: The number $N \simeq 70$ has been obtained requiring that the present-day value of $(\Omega_0 - 1)$ is of order unity. For the expression (169), it is clear that –if the period of inflation lasts longer than 70 e-foldings the present-day value of Ω_0 will be equal to unity with a great precision. One can say that a generic prediction of inflation is that

$$\text{INFLATION} \quad \Longrightarrow \quad \Omega_0 = 1.$$

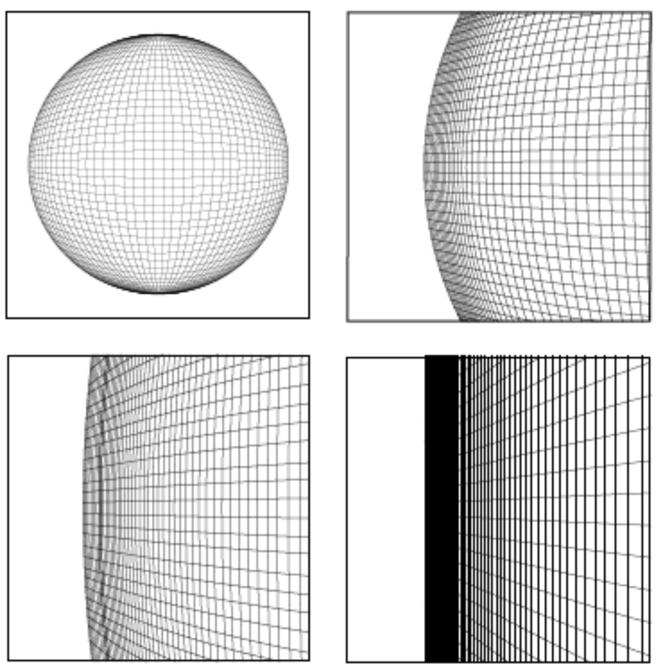

Figure 10: Inflation predicts a local flat universe.

This statement, however, must be taken *cum grano salis* and properly specified. Inflation does not change the global geometric properties of the space-time. If the universe is open or closed, it will always remain flat or closed, independently from inflation. What inflation does is to magnify the radius of curvature $R_{\rm curv}$ so that locally the universe is flat with a great precision. As we shall see, the current data on the CMB anisotropies confirm this prediction!

6.3 Inflation and the entropy problem

In the previous section, we have seen that the flatness problem arises because the entropy in a comoving volume is conserved. It is possible, therefore, that the problem could be resolved if the cosmic expansion was non-adiabatic for some finite time interval during the early history of the universe. We need to produce a large amount of entropy $S_{\rm U} \sim 10^{90}$. Let us postulate that the entropy changed by an amount

$$S_{\rm f} = Z^3 \, S_{\rm m,i} \tag{170}$$

from the beginning to the end of the inflationary period, where Z is a numerical factor. It is very natural to assume that the total entropy of the universe at the beginning of inflation was of order unity, one particle per horizon. Since, from the end of inflation onwards, the universe expands adiabatically, we have $S_{\rm f} = S_{\rm U}$. This gives $Z \sim 10^{30}$. On the other hand, since $S_{\rm f} \sim (a_{\rm f} T_{\rm f})^3$ and $S_{\rm m,i} \sim (a_{\rm I} t_{\rm I})^3$, where $T_{\rm f}$ and $t_{\rm I}$ are the temperatures of the universe at the end and at the beginning of inflation, we get

$$\left(\frac{a_{\rm f}}{a_{\rm I}}\right) = e^N \approx 10^{30} \left(\frac{t_{\rm I}}{T_{\rm f}}\right), \tag{171}$$

which gives again $N \sim 70$ up to the logarithmic factor $\ln\left(\frac{t_{\rm I}}{T_{\rm f}}\right)$. We stress again that such a large amount of entopy is not produced during inflation, but during the non-adiabatic phase transition which gives rise to the usual RD phase.

6.4 Inflation and the inflaton

In the previous subsections we have described the various adavantages of having a period of accelerating phase. The latter required $P < -\rho/3$. Now, we would like to show that this condition can be attained by means of a simple scalar field. We shall call this field the *inflaton* ϕ.

The action of the inflaton field reads

$$S = \int {\rm d}^4 x \, \sqrt{-g} \, \mathcal{L} = \int d^4 x \, \sqrt{-g} \, \left[-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right], \tag{172}$$

where $\sqrt{-g} = a^3$ for the FRW metric. From the Eulero-Lagrange equations

$$\partial^\mu \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta\, \partial^\mu\phi} - \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta\phi} = 0, \tag{173}$$

we obtain

$$\boxed{\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a^2} + V'(\phi) = 0}, \tag{174}$$

where $V'(\phi) = (\mathrm{d}V(\phi)/\mathrm{d}\phi)$. Note, in particular, the appearance of the friction term $3H\dot{\phi}$: a scalar field rolling down its potential suffers a friction due to the expansion of the universe.

We can write the energy-momentum tensor of the scalar field

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\,\mathcal{L}.$$

The corresponding energy density ρ_ϕ and pressure density P_ϕ are

$$T_{00} = \rho_\phi = \frac{\dot{\phi}^2}{2} + V(\phi) + \frac{(\nabla\phi)^2}{2a^2}, \tag{175}$$

$$T_{ii} = P_\phi = \frac{\dot{\phi}^2}{2} - V(\phi) - \frac{(\nabla\phi)^2}{6a^2}. \tag{176}$$

Notice that, if the gradient term were dominant, we would obtain $P_\phi = -\rho_\phi/3$, not enough to drive inflation. We can now split the inflaton field in

$$\phi(t) = \phi_0(t) + \delta\phi(\mathbf{x}, t),$$

where ϕ_0 is the 'classical' (infinite wavelength) field, that is the expectation value of the inflaton field on the initial isotropic and homogeneous state, while $\delta\phi(\mathbf{x}, t)$ represents the quantum fluctuations around ϕ_0. as for now, we will be only concerned with the evolution of the classical field ϕ_0. This separation is justified by the fact that quantum fluctuations are much smaller than the classical value and therefore negligible when looking at the classical evolution. To not be overwhelmed by the notation, we will keep indicating from now on the classical value of the inflaton field by ϕ. The energy-momentum tensor becomes

$$T_{00} = \rho_\phi = \frac{\dot{\phi_0}^2}{2} + V(\phi_0) \tag{177}$$

$$T_{ii} = P_\phi = \frac{\dot{\phi}_0^2}{2} - V(\phi_0). \tag{178}$$

If

$$V(\phi_0) \gg \dot{\phi}_0^2$$

we obtain the following condition

$$P_\phi \simeq -\rho_\phi$$

From this simple calculation, we realize that a scalar field whose energy is dominant in the universe and whose potential energy dominates over the kinetic term gives inflation! Inflation is driven by the vacuum energy of the inflaton field.

6.5 Slow-roll conditions

Let us now quantify better under which circumstances a scalar field may give rise to a period of inflation. The equation of motion of the field is

$$\dot{\phi}_0 + 3H\dot{\phi}_0 + V'(\phi_0) = 0 \tag{179}$$

If we require that $\dot{\phi}_0^2 \ll V(\phi_0)$, the scalar field is slowly rolling down its potential. This is the reason why such a period is called *slow-roll*. We may also expect that – being the potential flat – $\ddot{\phi}$ is negligible as well. We will assume that this is true and we will quantify this condition soon. The FRW equation becomes

$$H^2 \simeq \frac{8\pi G_{\rm N}}{3} V(\phi_0), \tag{180}$$

where we have assumed that the inflaton field dominates the energy density of the universe. The new equation of motion becomes

$$3H\dot{\phi}_0 = -V'(\phi_0), \tag{181}$$

which gives $\dot{\phi}_0$ as a function of $V'(\phi_0)$. Using Eq. (181) slow-roll conditions then require

$$\boxed{\dot{\phi}_0^2 \ll V(\phi_0) \Longrightarrow \frac{(V')^2}{V} \ll H^2}$$

and

$$\boxed{\ddot{\phi}_0 \ll 3H\dot{\phi}_0 \Longrightarrow V'' \ll H^2}.$$

It is now useful to define the slow-roll parameters, ϵ and η in the following way

$$\boxed{\begin{aligned} \epsilon &= -\frac{\dot{H}}{H^2} = 4\pi G_{\rm N}\frac{\dot{\phi}_0^2}{H^2} = \frac{1}{16\pi G_{\rm N}}\left(\frac{V'}{V}\right)^2, \\ \eta &= \frac{1}{8\pi G_{\rm N}}\left(\frac{V''}{V}\right) = \frac{1}{3}\frac{V''}{H^2}, \\ \delta &= \eta - \epsilon = -\frac{\ddot{\phi}_0}{H\dot{\phi}_0}. \end{aligned}}$$

It might be useful to have the same parameters expressed in terms of conformal time

$$\boxed{\begin{aligned} \epsilon &= 1 - \frac{\mathcal{H}'}{\mathcal{H}^2} = 4\pi G_{\rm N}\frac{{\phi_0'}^2}{\mathcal{H}^2} \\ \delta &= \eta - \epsilon = 1 - \frac{\phi_0''}{\mathcal{H}\phi'}. \end{aligned}}$$

The parameter ϵ quantifies how much the Hubble rate H changes with time during inflation. Notice that, since

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = (1-\epsilon)\, H^2,$$

inflation can be attained only if $\epsilon < 1$:

$$\text{INFLATION} \quad \Longleftrightarrow \quad \epsilon < 1.$$

As soon as this condition fails, inflation ends. In general, slow-roll inflation is attained if $\epsilon \ll 1$ and $|\eta| \ll 1$. During inflation the slow-roll parameters ϵ and η can be considered to be approximately constant since the potential $V(\phi)$ is very flat.

Comment: In the following, we will work at *first-order* perturbation in the slow-roll parameters, that is we will take only the first power of them. Since, using their definition, it is easy to see that $\dot{\epsilon}, \dot{\eta} = \mathcal{O}\left(\epsilon^2, \eta^2\right)$, this amounts to saying that we will trat the slow-roll parameters as constant in time.

Within these approximations, it is easy to compute the number of e-foldings between the beginning and the end of inflation. If we indicate by $\phi_{\rm m,i}$ and $\phi_{\rm f}$ the values of the inflaton field at the beginning and at the end of inflation, respectively, we have that the *total* number of e-foldings is

$$\begin{aligned} N &\equiv \int_{t_{\rm I}}^{t_{\rm f}} H\,{\rm d}t \\ &\simeq H \int_{\phi_{\rm m,i}}^{\phi_{\rm f}} \frac{{\rm d}\phi_0}{\dot{\phi}_0} \\ &\simeq -3H^2 \int_{\phi_{\rm m,i}}^{\phi_{\rm f}} \frac{{\rm d}\phi_0}{V'} \\ &\simeq -8\pi G_{\rm N} \int_{\phi_{\rm m,i}}^{\phi_{\rm f}} \frac{V}{V'}\,{\rm d}\phi_0. \end{aligned} \tag{182}$$

We may also compute the number of e-foldings ΔN which are left to go to the end of inflation

$$\Delta N \simeq 8\pi G_{\rm N} \int_{\phi_{\rm f}}^{\phi_{\Delta N}} \frac{V}{V'}\,{\rm d}\phi_0, \tag{183}$$

where $\phi_{\Delta N}$ is the value of the inflaton field when there are ΔN e-foldings to the end of inflation.

1. Comment: A given scale length $\lambda = a/k$ leaves the horizon when $k = aH_k$ where H_k is the the value of the Hubble rate at that time. One can compute easily the rate of change of H_k^2 as a function of k

$$\frac{{\rm dln}\,H_k^2}{{\rm dln}\,k} = \left(\frac{{\rm dln}\,H_k^2}{{\rm d}t}\right)\left(\frac{{\rm d}t}{{\rm dln}\,a}\right)\left(\frac{{\rm dln}\,a}{{\rm dln}\,k}\right) = 2\frac{\dot{H}}{H} \times \frac{1}{H} \times 1 = 2\frac{\dot{H}}{H^2} = -2\epsilon. \tag{184}$$

2. Comment: Take a given physical scale λ today which crossed the horizon scale during inflation. This happened when

$$\lambda \left(\frac{a_{\rm f}}{a_0}\right) e^{-\Delta N_\lambda} = \lambda \left(\frac{T_0}{T_{\rm f}}\right) e^{-\Delta N_\lambda} = H_{\rm I}^{-1}$$

where ΔN_λ indicates the number of e-foldings from the time the scale crossed the horizon during inflation and the end of inflation. This relation gives a way to determine the number of e-foldings to the end of inflation corresponding to a given scale

$$\Delta N_\lambda \simeq 65 + \ln\left(\frac{\lambda}{3000\ \mathrm{Mpc}}\right) + 2\ln\left(\frac{V^{1/4}}{10^{14}\ \mathrm{GeV}}\right) - \ln\left(\frac{T_{\rm f}}{10^{10}\ \mathrm{GeV}}\right).$$

Scales relevant for the CMB anisotropies correspond to $\Delta N \sim$60.

6.6 The last stage of inflation and reheating

Inflation ended when the potential energy associated with the inflaton field became smaller than the kinetic energy of the field. By that time, any pre-inflation entropy in the universe had been inflated away, and the energy of the universe was entirely in the form of coherent oscillations of the inflaton condensate around the minimum of its potential. The universe may be said to be frozen after the end of inflation. We know that somehow the low-entropy cold universe dominated by the energy of coherent motion of the ϕ field must be transformed into a high-entropy hot universe dominated by radiation. The process by which the energy of the inflaton field is transferred from the inflaton field to radiation has been dubbed *reheating*. In the old theory of reheating, the simplest way to envision this process is if the comoving energy density in the zero mode of the inflaton decays into normal particles, which then scatter and thermalize to form a thermal background. It is usually assumed that the decay width of this process is the same as the decay width of a free inflaton field.

Of particular interest is a quantity known usually as the reheat temperature, denoted as $T_{\rm RH}$ (so far, we have indicated it with $T_{\rm f}$). The reheat temperature is calculated by assuming an instantaneous conversion of the energy density in the inflaton field into radiation when the decay width of the inflaton energy, Γ_ϕ, is equal to H, the expansion rate of the universe.

The reheat temperature is calculated quite easily. After inflation the inflaton field executes coherent oscillations about the minimum of the potential at some $\phi_0 \simeq \phi_{\rm m}$

$$V(\phi_0) \simeq \frac{1}{2}V''(\phi_{\rm m})(\phi_0-\phi_{\rm m})^2 \equiv \frac{1}{2}m^2(\phi_0-\phi_{\rm m})^2 \tag{185}$$

Indeed, the equation of motion for ϕ_0 is

$$\dot{\phi}_0 + 3H\dot{\phi}_0 + m^2(\phi_0-\phi_{\rm m}) = 0, \tag{186}$$

whose solution is

$$\phi_0(t) = \phi_{\rm m,i}\left(\frac{a_{\rm I}}{a}\right)^3 \cos\left[m(t-t_{\rm I})\right], \tag{187}$$

where $t_{\rm I}$ denotes here the beginning of the oscillations. Since the period of the oscillation is much shorter than the Hubble time, $H \gg m$, we can compute over many oscillations the the equation satisfied by average energy density stored in the oscillating field

$$
\begin{aligned}
\langle\dot{\rho}_\phi\rangle &= \left\langle \frac{\rm d}{{\rm d}t}\left(\frac{1}{2}\dot{\phi}_0^2 + V(\phi_0)\right)\right\rangle_{\rm many\, oscillations} \\
&= \left\langle \dot{\phi}_0\left(\dot{\phi}_0 + V'(\phi_0)\right)\right\rangle_{\rm many\, oscillations} \\
&= \left\langle \dot{\phi}_0\left(-3H\dot{\phi}_0\right)\right\rangle_{\rm many\, oscillations} \\
&- 3H\left\langle\dot{\phi}_0^2\right\rangle_{\rm many\, oscillations} \\
&= -3H\left\langle\rho_\phi\right\rangle_{\rm many\, oscillations},
\end{aligned}
\tag{188}
$$

where we have used the equipartition property of the energy density during the oscillations $\langle\dot{\phi}_0^2/2\rangle = \langle V(\phi_0)\rangle = \langle\rho_\phi/2\rangle$ and Eq. (174). The solution of Eq. (188) is (removing the symbol of averaging)

$$
\rho_\phi = (\rho_\phi)_{\rm m,i}\left(\frac{a_{\rm I}}{a}\right)^3. \tag{189}
$$

The Hubble expansion rate as a function of a is

$$
H^2(a) = \frac{8\pi}{3}\frac{(\rho_\phi)_{\rm m,i}}{M_{\rm Pl}^2}\left(\frac{a_{\rm I}}{a}\right)^3. \tag{190}
$$

Equating $H(a)$ and Γ_ϕ leads to an expression for a_0/a. Now if we assume that all available coherent energy density is instantaneously converted into radiation at this value of a_0/a, we can find the reheat temperature by setting the coherent energy density, $\rho_\phi = (\rho_\phi)_0(a_0/a)^3$, equal to the radiation energy density, $\rho_R = (\pi^2/30)g_* T_{\rm RH}^4$, where g_* is the effective number of relativistic degrees of freedom at temperature $T_{\rm RH}$. The result is

$$
\boxed{T_{\rm RH} = \left(\frac{90}{8\pi^3 g_*}\right)^{1/4}\sqrt{\Gamma_\phi M_{\rm Pl}} = 0.2\left(\frac{200}{g_*}\right)^{1/4}\sqrt{\Gamma_\phi M_{\rm Pl}}.} \tag{191}
$$

In some models of inflation reheating can be anticipated by a period of preheating when the the classical inflaton field very rapidly (explosively) decays into ϕ-particles or into other bosons due to broad parametric resonance. This stage cannot be described by the standard elementary approach to reheating based on perturbation theory. The bosons produced at this stage further decay into other particles, which eventually become thermalized.

The presence of a preheating stage at the beginning of the reheating process is based on the fact that, for some parameter ranges, there is a new decay channel that is non-perturbative: due to the coherent oscillations of the inflaton field stimulated emissions of bosonic particles into energy bands with large occupancy numbers are induced. The modes in these bands can be understood as Bose condensates, and they behave like classical waves. The back-reaction of these modes on the homogeneous inflaton field and the rescattering among themselves produce a state that is far from thermal equilibrium and may induce very interesting phenomena, such as non-thermal phase transitions with production of a stochastic background of gravitational waves and of heavy particles in a state far from equilibrium, which may constitute today the dark matter in our universe.

The idea of preheating is relatively simple, the oscillations of the inflaton field induce mixing of positive and negative frequencies in the quantum state of the field it couples to because of the *time-dependent* mass of the quantum field. Let us focus – for sake of simplicity – to the case of a massive inflaton ϕ with quadratic potential $V(\phi) = \frac{1}{2}m^2\phi^2$ and coupled to a massless scalar field χ via the quartic coupling $g^2\phi^2\chi^2$.

The evolution equation for the Fourier modes of the χ field with momentum $\mathbf{k}$ is

$$\dot{X}_{\mathbf{k}} + \omega_k^2 X_{\mathbf{k}} = 0, \tag{192}$$

with

$$\begin{aligned} X_{\mathbf{k}} &= a^{3/2}(t)\chi_{\mathbf{k}}, \\ \omega_k^2 &= k^2/a^2(t) + g^2\phi^2(t). \end{aligned} \tag{193}$$

This Klein-Gordon equation may be cast in the form of a Mathieu equation

$$X''_{\mathbf{k}} + [A(k) - 2q\cos 2z]X_{\mathbf{k}} = 0, \tag{194}$$

where $z = mt$ and

$$\begin{aligned} A(k) &= \frac{k^2}{a^2m^2} + 2q, \\ q &= g^2\frac{\Phi^2}{4m^2}, \end{aligned} \tag{195}$$

where Φ is the amplitude and m is the frequency of inflaton oscillations, $\phi(t) = \Phi(t)\sin(mt)$. Notice that, at least initially, if $\Phi \gg M_{\rm Pl}$

$$g^2\frac{\Phi^2}{4m^2} \gg g^2\frac{M_{\rm Pl}^2}{m^2} \tag{196}$$

can be extremely large. If so, the resonance is broad. For certain values of the parameters (A, q) there are exact solutions $X_{\mathbf{k}}$ and the corresponding number density $n_{\mathbf{k}}$ that grow exponentially with time because they belong to an instability band of the Mathieu equation

$$X_{\mathbf{k}} \propto e^{\mu_{\mathbf{k}}mt} \Rightarrow n_{\mathbf{k}} \propto e^{2\mu_{\mathbf{k}}mt}, \tag{197}$$

where the parameter $\mu_{\mathbf{k}}$ depends upon the instability band and, in the broad resonance case, $q \gg 1$, it is ~ 0.2.

These instabilities can be interpreted as coherent "particle" production with large occupancy numbers. One way of understanding this phenomenon is to consider the energy of these modes as that of a harmonic oscillator, $E_{\mathbf{k}} = |\dot{X}_{\mathbf{k}}|^2/2 + \omega_{\mathbf{k}}^2 |X_{\mathbf{k}}|^2/2 = \omega_{\mathbf{k}}(n_{\mathbf{k}} + 1/2)$. The occupancy number of level $\mathbf{k}$ can grow exponentially fast, $n_{\mathbf{k}} \sim \exp(2\mu_{\mathbf{k}} mt) \gg 1$, and these modes soon behave like classical waves. The parameter q during preheating determines the strength of the resonance. It is possible that the model parameters are such that parametric resonance does *not* occur, and then the usual perturbative approach would follow, with decay rate Γ_ϕ. In fact, as the universe expands, the growth of the scale factor and the decrease of the amplitude of inflaton oscillations shifts the values of (A, q) along the stability/instability chart of the Mathieu equation, going from broad resonance, for $q \gg 1$, to narrow resonance, $q \ll 1$, and finally to the perturbative decay of the inflaton.

It is important to notice that, after the short period of preheating, the universe is likely to enter a long period of matter domination where the biggest contribution to the energy density of the universe is provided by the residual small amplitude oscillations of the classical inflaton field and/or by the inflaton quanta produced during the back-reaction processes. This period will end when the age of the universe becomes of the order of the perturbative lifetime of the inflaton field, $t \sim \Gamma_\phi^{-1}$. At this point, the universe will be reheated up to a temperature $T_{\rm RH}$ obtained applying the old theory of reheating described previously.

6.7 A brief survey of inflationary models

Even restricting ourselves to a simple single-field inflation scenario, the number of models available to choose from is large. It is convenient to define a general classification scheme, or "zoology" for models of inflation. We divide models into three general types: *large-field*, *small-field*, and *hybrid*, with a fourth classification. A generic single-field potential can be characterized by two independent mass scales: a "height" Λ^4, corresponding to the vacuum energy density during inflation, and a "width" μ, corresponding to the change in the field value $\Delta\phi$ during inflation:

$$V(\phi) = \Lambda^4 f\left(\frac{\phi}{\mu}\right). \tag{198}$$

Different models have different forms for the function f. Let us now briefly describe the different class of models.

6.7.1 Large-field models

Large-field models are potentials typical of the "chaotic" inflation scenario, in which the scalar field is displaced from the minimum of the potential by an amount usually of order the Planck mass. Such models are characterized by $V''(\phi) > 0$, and $-\epsilon < \delta \leq \epsilon$. The generic large-field potentials we consider are polynomial potentials $V(\phi) = \Lambda^4 (\phi/\mu)^p$, and exponential potentials,

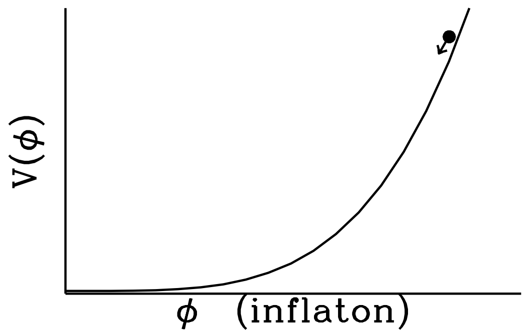

Figure 11: Large field models of inflation.

$V(\phi) = \Lambda^4 \exp(\phi/\mu)$. In the chaotic inflation scenario, it is assumed that the universe emerged from a quantum gravitational state with an energy density comparable to that of the Planck density. This implies that $V(\phi) \approx M_{\rm Pl}^4$ and results in a large friction term in the Friedmann equation. Consequently, the inflaton will slowly roll down its potential. The condition for inflation is therefore satisfied and the scale factor grows as

$$a(t) = a_{\rm I} e^{\left(\int_{t_{\rm I}}^{t} {\rm d}t' H(t')\right)}. \tag{199}$$

The simplest chaotic inflation model is that of a free field with a quadratic potential, $V(\phi) = m^2\phi^2/2$, where m represents the mass of the inflaton. During inflation the scale factor grows as

$$a(t) = a_{\rm I} e^{2\pi(\phi_{\rm m,i}^2 - \phi^2(t))} \tag{200}$$

and inflation ends when $\phi = \mathcal{O}\ M_{\rm Pl}$. If inflation begins when $V(\phi_{\rm m,i}) \approx M_{\rm Pl}^4$, the scale factor grows by a factor $\exp(4\pi M_{\rm Pl}^2/m^2)$ before the inflaton reaches the minimum of its potential. We will later show that the mass of the field should be $m \approx 10^{-6} M_{\rm Pl}$ if the microwave background constraints are to be satisfied. This implies that the volume of the universe will increase by a factor of $Z^3 \approx 10^{3\times 10^{12}}$ and this is more than enough inflation to solve the problems of the hot big bang model.

In the chaotic inflationary scenarios, the present-day universe is only a small portion of the universe which suffered inflation! Notice also that the typical values of the inflaton field during inflation are of the order of $M_{\rm Pl}$, giving rise to the possibility of testing planckian physics.

6.7.2 Small-field models

Small-field models are the type of potentials that arise naturally from spontaneous symmetry breaking (such as the original models of "new" inflation and from pseudo Nambu-Goldstone modes (natural inflation). The field starts from near an unstable equilibrium (taken to be at the origin) and rolls down the potential to a stable minimum. Small-field models are characterized by $V''(\phi) < 0$ and

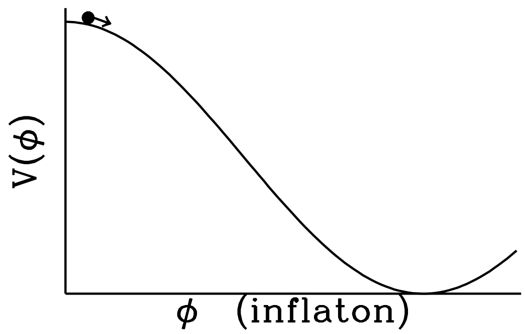

Figure 12: Small field models of inflation.

$\eta < -\epsilon$. Typically ϵ is close to zero. The generic small-field potentials we consider are of the form $V(\phi) = \Lambda^4 \left[1 - (\phi/\mu)^p\right]$, which can be viewed as a lowest-order Taylor expansion of an arbitrary potential about the origin.

6.7.3 Hybrid models

The hybrid scenario frequently appears in models which incorporate inflation into supersymmetry and supergravity. In a typical hybrid inflation model, the scalar field responsible for inflation evolves toward a minimum with nonzero vacuum energy. The end of inflation arises as a result of instability in a second field. Such models are characterized by $V''(\phi) > 0$ and $0 < \epsilon < \delta$. We consider generic potentials for hybrid inflation of the form $V(\phi) = \Lambda^4 \left[1 + (\phi/\mu)^p\right]$. The field value at the end of inflation is determined by some other physics, so there is a second free parameter characterizing the models. This enumeration of models is certainly not exhaustive. There are a number of single-

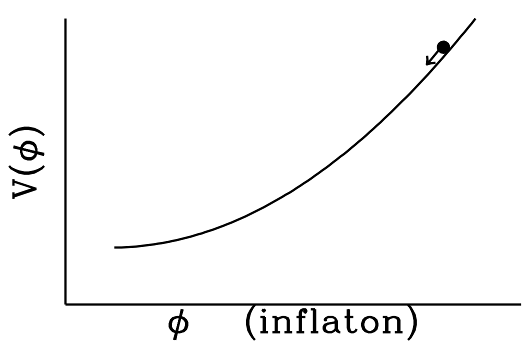

Figure 13: Hybrid field models of inflation.

field models that do not fit well into this scheme, for example logarithmic potentials $V(\phi) \propto \ln(\phi)$ typical of supersymmetry. Another example is potentials with negative powers of the scalar field

$V(\phi) \propto \phi^{-p}$ used in intermediate inflation and dynamical supersymmetric inflation. Both of these cases require and auxilliary field to end inflation and are more properly categorized as hybrid models, but fall into the small-field class. However, the three classes categorized by the relationship between the slow-roll parameters as $-\epsilon < \delta \leq \epsilon$ (large-field), $\delta \leq -\epsilon$ (small-field) and $0 < \epsilon < \delta$ (hybrid) seems to be good enough for comparing theoretical expectations with experimental data.

Part IV

Inflation and the cosmological perturbations

As we have seen in the previous section, the early universe was made very nearly uniform by a primordial inflationary stage. However, the important caveat in that statement is the word 'nearly'. As we shall see, our current understanding of the origin of structure in the universe is that it originated from small 'seed' perturbations, which over time grew to become all of the structure we observe. Once the universe becomes matter dominated (around 1000 yrs after the bang) primeval density inhomogeneities ($\delta\rho/\rho \sim 10^{-5}$) are amplified by gravity and grow into the structure we see today. The fact that a fluid of self-gravitating particles is unstable to the growth of small inhomogeneities was first pointed out by Jeans and is known as the Jeans instability. Furthermore, the existence of these inhomogeneities is confirmed by detailed measurements of the CMB anisotropies; the temperature anisotropies detected almost certainly owe their existence to primeval density inhomogeneities, since, as we have seen, causality precludes microphysical processes from producing anisotropies on angular scales larger than about 1°, the angular size of the horizon at last-scattering.

Let us just anticipate for the sake of the argument that the growth of small matter inhomogeneities of wavelength smaller than the Hubble scale ($\lambda \lesssim H^{-1}$) is governed by a Newtonian equation:

$$\dot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + v_s^2 \frac{k^2}{a^2}\delta_{\mathbf{k}} = 4\pi G_{\rm N}\rho_{\rm NR}\delta_{\mathbf{k}}, \tag{201}$$

where $v_s^2 = \partial p/\partial\rho_{\rm NR}$ is the square of the sound speed and we have expanded the perturbation to the matter density in plane waves

$$\frac{\delta\rho_{\rm NR}(\mathbf{x},t)}{\rho_{\rm NR}} = \frac{1}{(2\pi)^3}\int {\rm d}^3k\, \delta_{\mathbf{k}}(t) e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{202}$$

Competition between the pressure term and the gravity term on the rhs of Eq. (201) determines whether or not pressure can counteract gravity: perturbations with wavenumber larger than the

Jeans wavenumber, $k_J^2 = 4\pi G_N a^2 \rho_{NR}/v_s^2$, are Jeans stable and just oscillate; perturbations with smaller wavenumber are Jeans unstable and can grow.

Let us discuss solutions to this equation under different circumstances. First, consider the Jeans problem, evolution of perturbations in a static fluid, *i.e.*, $H = 0$. In this case Jeans unstable perturbations grow exponentially, $\delta_{\mathbf{k}} \propto \exp(t/\tau)$ where $\tau = 1/\sqrt{4G_N\pi\rho_{NR}}$. Next, consider the growth of Jeans unstable perturbations in a MD universe, *i.e.*, $H^2 = 8\pi G_N\rho_{NR}/3$ and $a \propto t^{2/3}$. Because the expansion tends to "pull particles away from one another," the growth is only power law, $\delta_{\mathbf{k}} \propto t^{2/3}$; *i.e.*, at the same rate as the scale factor. Finally, consider a RD universe. In this case, the expansion is so rapid that matter perturbations grow very slowly, as $\ln a$ in RD epoch. Therefore, perturbations may grow only in a MD period. Once a perturbation reaches an overdensity of order unity or larger it "separates" from the expansion –*i.e.*, becomes its own self-gravitating system and ceases to expand any further. In the process of virial relaxation, its size decreases by a factor of two—density increases by a factor of 8; thereafter, its density contrast grows as a^3 since the average matter density is decreasing as a^{-3}, though smaller scales could become Jeans unstable and collapse further to form smaller objects of higher density.

In order for structure formation to occur via gravitational instability, there must have been small preexisting fluctuations on physical length scales when they crossed the Hubble radius in the RD and MD eras. In the standard Big-Bang model these small perturbations have to be put in by hand, because it is impossible to produce fluctuations on any length scale while it is larger than the horizon. Since the goal of cosmology is to understand the universe on the basis of physical laws, this appeal to initial conditions is unsatisfactory. The challenge is therefore to give an explanation to the small seed perturbations which allow the gravitational growth of the matter perturbations.

Our best guess for the origin of these perturbations is quantum fluctuations during an inflationary era in the early universe. Although originally introduced as a possible solution to the cosmological conundrums such as the horizon, flatness and entopy problems, by far the most useful property of inflation is that it generates spectra of both density perturbations and gravitational waves. These perturbations extend from extremely short scales to scales considerably in excess of the size of the observable universe.

During inflation the scale factor grows quasi-exponentially, while the Hubble radius remains almost constant. Consequently the wavelength of a quantum fluctuation – either in the scalar field whose potential energy drives inflation or in the graviton field – soon exceeds the Hubble radius. The amplitude of the fluctuation therefore becomes 'frozen in'. This is quantum mechanics in action at macroscopic scales!

According to quantum field theory, empty space is not entirely empty. It is filled with quantum fluctuations of all types of physical fields. The fluctuations can be regarded as waves of physical fields with all possible wavelenghts, moving in all possible directions. If the values of these fields, averaged over some macroscopically large time, vanish then the space filled with these fields seems to us empty and can be called the vacuum.

In the exponentially expanding universe the vacuum structure is much more complicated. The wavelenghts of all vacuum fluctuations of the inflaton field ϕ grow exponentially in the expnading universe. When the wavelength of any particular fluctuation becomes greater than H^{-1}, this fluctuation stops propagating, and its amplitude freezes at some nonzero value $\delta\phi$ because of the large friction term $3H\delta\dot{\phi}$ in the equation of motion of the field $\delta\phi$. The amplitude of this fluctuation then remains almost unchanged for a very long time, whereas its wavelength grows exponentially. Therefore, the appearance of such frozen fluctuation is equivalent to the appearance of a classical field $\delta\phi$ that does not vanish after having averaged over some macroscopic interval of time. Because the vacuum contains fluctuations of all possible wavelength, inflation leads to the creation of more and more new perturbations of the classical field with wavelength larger than the horizon scale.

Once inflation has ended, however, the Hubble radius increases faster than the scale factor, so the fluctuations eventually reenter the Hubble radius during the radiation- or MD eras. The fluctuations that exit around 60 e-foldings or so before reheating reenter with physical wavelengths in the range accessible to cosmological observations. These spectra provide a distinctive signature of inflation. They can be measured in a variety of different ways including the analysis of microwave background anisotropies.

The physical processes which give rise to the structures we observe today are well-explained in Fig. 14.

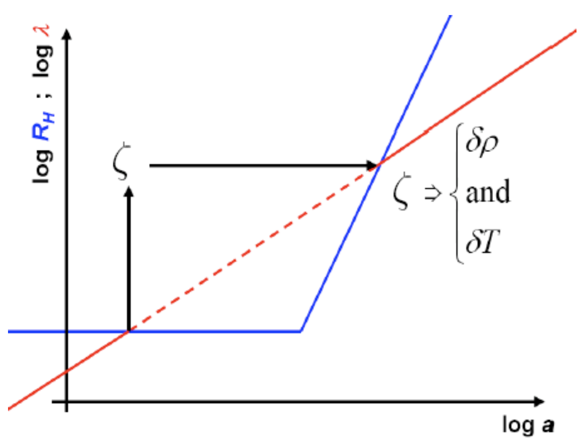

Figure 14: A schematic representation of the generation of quantum fluctuations during inflation.

Since gravity talks to any component of the universe, small fluctuations of the inflaton field are intimately related to fluctuations of the space-time metric, giving rise to perturbations of the curvature $\mathcal{R}$ (which will be defined in the following; the reader may loosely think of it as a gravitational potential). The wavelenghts λ of these perturbations grow exponentially and leave soon the horizon when $\lambda > R_{\rm H}$. On superHubble scales, curvature fluctuations are frozen in and may be considered as classical. Finally, when the wavelength of these fluctuations reenters the horizon, at some radiation- or MD epoch, the curvature (gravitational potential) perturbations of the space-time give rise to matter (and temperature) perturbations $\delta\rho$ via the Poisson equation. These fluctuations will then start growing giving rise to the structures we observe today.

In summary, two are the key ingredients for understanding the observed structures in the universe within the inflationary scenario:

- Quantum fluctuations of the inflaton field are excited during inflation and stretched to cosmological scales. At the same time, being the inflaton fluctuations connected to the metric perturbations through Einstein's equations, ripples on the metric are also excited and stretched to cosmological scales.
- Gravity acts a messanger since it communicates to baryons and photons the small seed perturbations once a given wavelength becomes smaller than the horizon scale after inflation.

Let us know see how quantum fluctuations are generated during inflation. We will proceed by steps. First, we will consider the simplest problem of studying the quantum fluctuations of a generic scalar field during inflation: we will learn how perturbations evolve as a function of time and compute their spectrum. Then – since a satisfactory description of the generation of quantum fluctuations have to take both the inflaton and the metric perturbations into account – we will study the system composed by quantum fluctuations of the inflaton field and quantum fluctuations of the metric.

7 Quantum fluctuations of a generic massless scalar field during inflation

Let us first see how the fluctuations of a generic scalar field χ, which is *not* the inflaton field, behave during inflation. To warm up we first consider a de Sitter epoch during which the Hubble rate is constant.

7.1 Quantum fluctuations of a generic massless scalar field during a de Sitter stage

We assume this field to be massless. The massive case will be analyzed in the next subsection. Expanding the scalar field χ in Fourier modes

$$\delta\chi(\mathbf{x},t) = \int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^{3/2}}\, e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta\chi_{\mathbf{k}}(t),$$

we can write the equation for the fluctuations as

$$\delta\dot{\chi}_{\mathbf{k}} + 3H\,\delta\dot{\chi}_{\mathbf{k}} + \frac{k^2}{a^2}\,\delta\chi_{\mathbf{k}} = 0. \tag{203}$$

Let us study the qualitative behaviour of the solution to Eq. (203).

- For wavelengths within the Hubble radius, $\lambda \ll H^{-1}$, the corresponding wavenumber satisfies the relation $k \gg a\,H$. In this regime, we can neglect the friction term $3H\,\delta\dot{\chi}_{\mathbf{k}}$ and Eq. (203) reduces to

  $$\delta\dot{\chi}_{\mathbf{k}} + \frac{k^2}{a^2}\,\delta\chi_{\mathbf{k}} = 0, \tag{204}$$

  which is – basically – the equation of motion of an harmonic oscillator. Of course, the frequency term k^2/a^2 depends upon time because the scale factor a grows exponentially. On the qualitative level, however, one expects that when the wavelength of the fluctuation is within the horizon, the fluctuation oscillates.

- For wavelengths above the Hubble radius, $\lambda \gg H^{-1}$, the corresponding wavenumber satisfies the relation $k \ll aH$ and the term k^2/a^2 can be safely neglected. Eq. (203) reduces to

  $$\delta\dot{\chi}_{\mathbf{k}} + 3H\,\delta\dot{\chi}_{\mathbf{k}} = 0, \tag{205}$$

  which tells us that on superHubble scales $\delta\chi_{\mathbf{k}}$ remains constant.

We have therefore the following picture: take a given fluctuation whose initial wavelength $\lambda \sim a/k$ is within the Hubble radius. The fluctuations oscillates till the wavelength becomes of the order of the horizon scale. When the wavelength crosses the Hubble radius, the fluctuation ceases to oscillate and gets frozen in.

Let us know study the evolution of the fluctuation is a more quantitative way. To do so, we perform the following redefinition

$$\delta\chi_{\mathbf{k}} = \frac{\delta\sigma_{\mathbf{k}}}{a}$$

and we work in conformal time $\mathrm{d}\tau = \mathrm{d}t/a$. For the time being, we solve the problem for a pure de Sitter expansion and we take the scale factor exponentially growing as $a \sim e^{Ht}$; the corresponding conformal factor reads (after choosing properly the integration constants)

$$a(\tau) = -\frac{1}{H\tau} \quad (\tau < 0).$$

In the following we will also solve the problem in the case of quasi de Sitter expansion. The beginning of inflation coincides with some initial time $\tau_{\mathrm{m,i}} \ll 0$. Using the set of rules (139), we find that Eq. (203) becomes

$$\delta\sigma''_{\mathbf{k}} + \left(k^2 - \frac{a''}{a}\right)\delta\sigma_{\mathbf{k}} = 0. \tag{206}$$

We obtain an equation which is very 'close' to the equation for a Klein-Gordon scalar field in flat space-time, the only difference being a negative time-dependent mass term $-a''/a = -2/\tau^2$. Eq. (206) can be obtained from an action of the type

$$\delta S_{\mathbf{k}} = \int \mathrm{d}\tau \left[\frac{1}{2}\delta\sigma_{\mathbf{k}}'^2 - \frac{1}{2}\left(k^2 - \frac{a''}{a}\right)\delta\sigma_{\mathbf{k}}^2\right], \tag{207}$$

which is the canonical action for a simple harmonic oscillator with canonical commutation relations

$$\delta\sigma_{\mathbf{k}}^*\delta\sigma_{\mathbf{k}}' - \delta\sigma_{\mathbf{k}}\delta\sigma_{\mathbf{k}}^{*\prime} = -i. \tag{208}$$

Let us study the behaviour of this equation on subHubble and superHubble scales. Since

$$\frac{k}{aH} = -k\,\tau,$$

on subHubble scales $k^2 \gg a''/a$ Eq. (206) reduces to

$$\delta\sigma_{\mathbf{k}}'' + k^2\,\delta\sigma_{\mathbf{k}} = 0,$$

whose solution is a plane wave

$$\delta\sigma_{\mathbf{k}} = \frac{e^{-ik\tau}}{\sqrt{2k}} \quad (k \gg aH). \tag{209}$$

We find again that fluctuations with wavelength within the horizon oscillate exactly like in flat space-time. This does not come as a surprise. In the ultraviolet regime, that is for wavelengths much smaller than the Hubble radius scale, one expects that approximating the space-time as flat is a good approximation.

On superHubble scales, $k^2 \ll a''/a$ Eq. (206) reduces to

$$\delta\sigma_{\mathbf{k}}'' - \frac{a''}{a}\delta\sigma_{\mathbf{k}} = 0,$$

which is satisfied by

$$\delta\sigma_{\mathbf{k}} = B(k)\,a \quad (k \ll aH). \tag{210}$$

where $B(k)$ is a constant of integration. Roughly matching the (absolute values of the) solutions (209) and (210) at $k = aH$ ($-k\tau = 1$), we can determine the (absolute value of the) constant $B(k)$

$$|B(k)|\,a = \frac{1}{\sqrt{2k}} \Longrightarrow |B(k)| = \frac{1}{a\sqrt{2k}} = \frac{H}{\sqrt{2k^3}}.$$

Going back to the original variable $\delta\chi_{\mathbf{k}}$, we obtain that the quantum fluctuation of the χ field on superHubble scales is constant and approximately equal to

$$|\delta\chi_{\mathbf{k}}| \simeq \frac{H}{\sqrt{2k^3}} \quad \text{(ON superHubble SCALES)}$$

In fact we can do much better, since Eq. (206) has an *exact* solution:

$$\delta\sigma_{\mathbf{k}} = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right). \tag{211}$$

This solution reproduces all what we have found by qualitative arguments in the two extreme regimes $k \ll aH$ and $k \gg aH$. The reason why we have performed the matching procedure is to show that the latter can be very useful to determine the behaviour of the solution on superHubble scales when the exact solution is not known.

7.2 Quantum fluctuations of a generic massive scalar field during a de Sitter stage

So far, we have solved the equation for the quantum perturbations of a generic massless field, that is neglecting the mass squared term m_χ^2. Let us know discuss the solution when such a mass term is present. Eq. (206) becomes

$$\delta\sigma_{\mathbf{k}}'' + \left[k^2 + M^2(\tau)\right]\delta\sigma_{\mathbf{k}} = 0, \tag{212}$$

where

$$M^2(\tau) = \left(m_\chi^2 - 2H^2\right)a^2(\tau) = \frac{1}{\tau^2}\left(\frac{m^2}{H^2} - 2\right).$$

Eq. (212) can be recast in the form

$$\delta\sigma_{\mathbf{k}}'' + \left[k^2 - \frac{1}{\tau^2}\left(\nu_\chi^2 - \frac{1}{4}\right)\right]\delta\sigma_{\mathbf{k}} = 0, \tag{213}$$

where

$$\nu_\chi^2 = \left(\frac{9}{4} - \frac{m_\chi^2}{H^2}\right). \tag{214}$$

The generic solution to Eq. (212) for ν_χ *real* is

$$\delta\sigma_{\mathbf{k}} = \sqrt{-\tau}\left[c_1(k)\, H_{\nu_\chi}(-k\tau) + c_2(k)\, H^{(2)}_{\nu_\chi}(-k\tau)\right],$$

where H_{ν_χ} and $H^{(2)}_{\nu_\chi}$ are the Hankel's functions of the first and second kind, respectively. If we impose that in the ultraviolet regime $k \gg aH$ $(-k\tau \gg 1)$ the solution matches the plane-wave solution $e^{-ik\tau}/\sqrt{2k}$ that we expect in flat space-time and knowing that

$$H_{\nu_\chi}(x \gg 1) \sim \sqrt{\frac{2}{\pi x}}\, e^{i\left(x - \frac{\pi}{2}\nu_\chi - \frac{\pi}{4}\right)} \quad, H^{(2)}_{\nu_\chi}(x \gg 1) \sim \sqrt{\frac{2}{\pi x}}\, e^{-i\left(x - \frac{\pi}{2}\nu_\chi - \frac{\pi}{4}\right)},$$

we set $c_2(k) = 0$ and $c_1(k) = \frac{\sqrt{\pi}}{2}\, e^{i\left(\nu_\chi + \frac{1}{2}\right)\frac{\pi}{2}}$. The exact solution becomes

$$\delta\sigma_{\mathbf{k}} = \frac{\sqrt{\pi}}{2}\, e^{i\left(\nu_\chi + \frac{1}{2}\right)\frac{\pi}{2}}\, \sqrt{-\tau}\, H_{\nu_\chi}(-k\tau). \tag{215}$$

On superHubble scales, since $H_{\nu_\chi}(x \ll 1) \sim \sqrt{2/\pi}\, e^{-i\frac{\pi}{2}}\, 2^{\nu_\chi - \frac{3}{2}}\, (\Gamma(\nu_\chi)/\Gamma(3/2))\, x^{-\nu_\chi}$, the fluctuation (215) becomes

$$\delta\sigma_{\mathbf{k}} = e^{i\left(\nu_\chi - \frac{1}{2}\right)\frac{\pi}{2}} 2^{\left(\nu_\chi - \frac{3}{2}\right)} \frac{\Gamma(\nu_\chi)}{\Gamma(3/2)} \frac{1}{\sqrt{2k}} \left(-k\tau\right)^{\frac{1}{2}-\nu_\chi}.$$

Going back to the old variable $\delta\chi_{\mathbf{k}}$, we find that on superHubble scales, the fluctuation with nonvanishing mass is not exactly constant, but it acquires a tiny dependence upon the time

$$|\delta\chi_{\mathbf{k}}| \simeq \frac{H}{\sqrt{2k^3}} \left(\frac{k}{aH}\right)^{\frac{3}{2}-\nu_\chi} \quad \text{(ON superHubble SCALES)}$$

If we now define, in analogy with the definition of the slow roll parameters η and ϵ for the inflaton field, the parameter $\eta_\chi = (m_\chi^2/3H^2) \ll 1$, one finds

$$\frac{3}{2} - \nu_\chi \simeq \eta_\chi. \tag{216}$$

7.3 Quantum to classical transition

We have previously said that the quantum flactuations can be regarded as classical when their corresponding wavelengths cross the horizon. To better motivate this statement, we should compute the number of particles $n_{\mathbf{k}}$ per wavenumber $\mathbf{k}$ on superHubble scales and check that it is indeed much larger than unity, $n_{\mathbf{k}} \gg 1$ (in this limit one can neglect the "quantum" factor 1/2 in the Hamiltonian $H_{\mathbf{k}} = \omega_{\mathbf{k}}\left(n_{\mathbf{k}} + \frac{1}{2}\right)$ where $\omega_{\mathbf{k}}$ is the energy eigenvalue). If so, the fluctuation can be regarded as classical. The number of particles $n_{\mathbf{k}}$ can be estimated to be of the order of $H_{\mathbf{k}}/\omega_{\mathbf{k}}$, where $H_{\mathbf{k}}$ is the Hamiltonian corresponding to the action

$$\delta S_{\mathbf{k}} = \int \mathrm{d}\tau \left[\frac{1}{2}\delta\sigma_{\mathbf{k}}^{\prime 2} + \frac{1}{2}\left(k^2 - M^2(\tau)\right)\delta\sigma_{\mathbf{k}}^2\right]. \tag{217}$$

One obtains on superHubble scales

$$n_{\mathbf{k}} \simeq \frac{M^2(\tau)\,|\delta\chi_{\mathbf{k}}|^2}{\omega_{\mathbf{k}}} \sim \left(\frac{k}{aH}\right)^{-3} \gg 1,$$

which confirms that fluctuations on superHubble scales may be indeed considered as classical.

7.4 The power spectrum

Let us define now the power spectrum, a useful quantity to characterize the properties of the perturbations. For a generic quantity $g(\mathbf{x}, t)$, which can expanded in Fourier space as

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

the power spectrum can be defined as

$$\langle 0|g^*_{\mathbf{k}_1} g_{\mathbf{k}_2}|0\rangle \equiv \delta^{(3)}\left(\mathbf{k}_1+\mathbf{k}_2\right)\frac{2\pi^2}{k^3}\mathcal{P}_g(k), \tag{218}$$

where $|0\rangle$ is the vacuum quantum state of the system. This definition leads to the relation

$$\boxed{\langle 0|g^2(\mathbf{x},t)|0\rangle = \int \frac{\mathrm{d}k}{k}\mathcal{P}_g(k).} \tag{219}$$

7.5 Quantum fluctuations of a generic scalar field in a quasi de Sitter stage

So far, we have computed the time evolution and the spectrum of the quantum flutuations of a generic scalar field χ supposing that the scale factor evolves like in a pure de Sitter expansion, $a(\tau) = -1/(H\tau)$. However, during inflation the Hubble rate is not exactly constant, but changes with time as $\dot{H} = -\epsilon H^2$ (quasi de Sitter expansion), In this subsection, we will solve for the perturbations in a quasi de Sitter expansion. Using the definition of the conformal time, one can show that the scale factor for small values of ϵ becomes

$$a(\tau) = -\frac{1}{H}\frac{1}{\tau(1-\epsilon)}.$$

Eq. (212) has now a squared mass term

$$M^2(\tau) = m_\chi^2 a^2 - \frac{a''}{a},$$

where

$$\begin{aligned}\frac{a''}{a} &= a^2\left(\frac{\dot{a}}{a}+H^2\right) = a^2\left(\dot{H}+2\,H^2\right)\\ &= a^2\left(2-\epsilon\right)H^2 = \frac{(2-\epsilon)}{\tau^2\left(1-\epsilon\right)^2}\\ &\simeq \frac{1}{\tau^2}\left(2+3\epsilon\right).\end{aligned} \tag{220}$$

Taking $m_\chi^2/H^2 = 3\eta_\chi$ and expanding for small values of ϵ and η we get Eq. (213) with

$$\nu_\chi \simeq \frac{3}{2}+\epsilon-\eta_\chi. \tag{221}$$

Armed with these results, we may compute the variance of the perturbations of the generic χ field

$$\begin{aligned}\langle 0|\left(\delta\chi(\mathbf{x},t)\right)^2|0\rangle &= \int\frac{\mathrm{d}^3k}{(2\pi)^3}\,|\delta\chi_{\mathbf{k}}|^2\\ &= \int\frac{\mathrm{d}k}{k}\frac{k^3}{2\pi^2}\,|\delta\chi_{\mathbf{k}}|^2\\ &= \int\frac{\mathrm{d}k}{k}\mathcal{P}_{\delta\chi}(k),\end{aligned} \tag{222}$$

which defines the power spectrum of the fluctuations of the scalar field χ

$$\mathcal{P}_{\delta\chi}(k) \equiv \frac{k^3}{2\pi^2} \, |\delta\chi_{\mathbf{k}}|^2 \,. \tag{223}$$

Since we have seen that fluctuations are (nearly) frozen in on superHubble scales, a way of characterizing the perturbations is to compute the spectrum on scales larger than the horizon. For a massive scalar field, we obtain

$$\mathcal{P}_{\delta\chi}(k) = \left(\frac{H}{2\pi}\right)^2 \left(\frac{k}{aH}\right)^{3-2\nu_\chi} . \tag{224}$$

We may also define the *spectral index* $n_{\delta\chi}$ of the fluctuations as

$$n_{\delta\chi} - 1 = \frac{d\ln \mathcal{P}_{\delta\phi}}{d\ln k} = 3 - 2\nu_\chi = 2\eta_\chi - 2\epsilon.$$

The power spectrum of fluctuations of the scalar field χ is therefore *nearly flat*, that is is nearly independent from the wavelength $\lambda = \pi/k$: the amplitude of the fluctuation on superHubble scales does not (almost) depend upon the time at which the fluctuations crosses the horizon and becomes frozen in. The small tilt of the power spectrum arises from the fact that the scalar field χ is massive and because during inflation the Hubble rate is not exactly constant, but nearly constant, where 'nearly' is quantified by the slow-roll parameters ϵ. Adopting the traditional terminology, we may say that the spectrum of perturbations is blue if $n_{\delta\chi} > 1$ (more power in the ultraviolet) and red if $n_{\delta\chi} < 1$ (more power in the infrared). The power spectrum of the perturbations of a generic scalar field χ generated during a period of slow roll inflation may be either blue or red. This depends upon the relative magnitude between η_χ and ϵ. For instance, in chaotic inflation with a quadric potential $V(\phi) = m^2\phi^2/2$, one can easily compute

$$n_{\delta\chi} - 1 = 2\eta_\chi - 2\epsilon = \frac{2}{3H^2}\left(m_\chi^2 - m^2\right),$$

which tells us that the spectrum is blue (red) if $m_\chi^2 > m_\phi^2$ ($m_\chi^2 > m^2$).

Comment: We might have computed the spectral index of the spectrum $\mathcal{P}_{\delta\chi}(k)$ by first solving the equation for the perturbations of the field χ in a di Sitter stage, with $H =$ constant and therefore $\epsilon = 0$, and then taking into account the time-evolution of the Hubble rate introducing the subscript in H_k whose time variation is determined by Eq. (184). Correspondingly, H_k is the value of the Hubble rate when a given wavelength $\sim k^{-1}$ crosses the horizon (from that point on the fluctuations remains frozen in). The power spectrum in such an approach would read

$$\mathcal{P}_{\delta\chi}(k) = \left(\frac{H_k}{2\pi}\right)^2 \left(\frac{k}{aH}\right)^{3-2\nu_\chi} \tag{225}$$

with $3-2\nu_\chi \simeq \eta_\chi$. Using Eq. (184), one finds

$$n_{\delta\chi}-1=\frac{\mathrm{d}\ln \mathcal{P}_{\delta\phi}}{\mathrm{d}\ln k}=\frac{\mathrm{d}\ln H_k^2}{\mathrm{d}\ln k}+3-2\nu_\chi=2\eta_\chi-2\epsilon$$

which reproduces our previous findings.

Comment: Since on superHubble scales

$$\delta\chi_{\mathbf{k}}\simeq\frac{H}{\sqrt{2k^3}}\left(\frac{k}{aH}\right)^{\eta_\chi-\epsilon}\simeq\frac{H}{\sqrt{2k^3}}\left[1+(\eta_\chi-\epsilon)\ln\left(\frac{k}{aH}\right)\right],$$

we discover that

$$|\delta\dot{\chi}_{\mathbf{k}}|\simeq|H\,(\eta_\chi-\epsilon)\,\delta\chi_{\mathbf{k}}|\ll|H\,\delta\chi_{\mathbf{k}}|\,, \tag{226}$$

that is on superHubble scales the time variation of the perturbations can be safely neglected.

8 Quantum fluctuations during inflation

As we have mentioned in the previous section, the linear theory of the cosmological perturbations represent a cornerstone of modern cosmology and is used to describe the formation and evolution of structures in the universe as well as the anisotrpies of the CMB. The seeds for these inhomogeneities were generated during inflation and stretched over astronomical scales because of the rapid superluminal expansion of the universe during the (quasi) de Sitter epoch.

In the previous section we have already seen that pertubations of a generic scalar field χ are generated during a (quasi) de Sitter expansion. The inflaton field is a scalar field and, as such, we conclude that inflaton fluctuations will be generated as well. However, the inflaton is special from the point of view of perturbations. The reason is very simple. By assumption, the inflaton field dominates the energy density of the universe during inflation. Any perturbation in the inflaton field means a perturbation of the stress energy-momentum tensor

$$\delta\phi\Longrightarrow\delta T_{\mu\nu}.$$

A perturbation in the stress energy-momentum tensor implies, through Einstein's equations of motion, a perturbation of the metric

$$\delta T_{\mu\nu}\Longrightarrow\left[\delta R_{\mu\nu}-\frac{1}{2}\delta\left(g_{\mu\nu}R\right)\right]=8\pi G\delta T_{\mu\nu}\Longrightarrow\delta g_{\mu\nu}.$$

On the other hand, a pertubation of the metric induces a backreaction on the evolution of the inflaton perturbation through the perturbed Klein-Gordon equation of the inflaton field

$$\delta g_{\mu\nu} \Longrightarrow \delta\left(\partial_\mu \partial^\mu \phi + \frac{\partial V}{\partial \phi}\right) = 0 \Longrightarrow \delta\phi.$$

This logic chain makes us conclude that the perturbations of the inflaton field and of the metric are tightly coupled to each other and have to be studied together

$$\boxed{\delta\phi \Longleftrightarrow \delta g_{\mu\nu}}$$

As we will see shortly, this relation is stronger than one might thought because of the issue of gauge invariance.

Before launching ourselves into the problem of finding the evolution of the quantum perturbations of the inflaton field when they are coupled to gravity, let us give a heuristic explanation of why we expect that during inflation such fluctuations are indeed present.

If we take Eq. (174) and split the inflaton field as its classical value ϕ_0 plus the quantum flucutation $\delta\phi$, $\phi(\mathbf{x}, t) = \phi_0(t) + \delta\phi(\mathbf{x}, t)$, the quantum perturbation $\delta\phi$ satisfies the equation of motion

$$\delta\ddot{\phi} + 3H\,\delta\dot{\phi} - \frac{\boldsymbol{\nabla}^2 \delta\phi}{a^2} + V''\,\delta\phi = 0. \tag{227}$$

Differentiating Eq. (179) with respect to time and taking H constant (de Sitter expansion) we find

$$(\phi_0)^{\cdots} + 3H\ddot{\phi}_0 + V''\,\dot{\phi}_0 = 0. \tag{228}$$

Let us consider for simplicity the limit $k/a \ll 1$ and let us disregard the gradient term. Under this condition we see that $\dot{\phi}_0$ and $\delta\phi$ solve the same equation. The solutions have therefore to be related to each other by a constant of proportionality which depends upon time, that is

$$\delta\phi = -\dot{\phi}_0\,\delta t(\mathbf{x}). \tag{229}$$

This tells us that $\phi(\mathbf{x}, t)$ will have the form

$$\boxed{\phi(\mathbf{x}, t) = \phi_0\left(\mathbf{x}, t - \delta t(\mathbf{x})\right).}$$

This equation indicates that the inflaton field does not acquire the same value at a given time t in all the space. On the contrary, when the inflaton field is rolling down its potential, it acquires different values from one spatial point $\mathbf{x}$ to the other. The inflaton field is not homogeneous and fluctuations are present. These fluctuations, in turn, will induce fluctuations in the metric.

8.1 The metric fluctuations

The mathematical tool do describe the linear evolution of the cosmological perturbations is obtained by perturbing at the first-order the FRW metric $g^{(0)}_{\mu\nu}$,

$$g_{\mu\nu} \quad = \quad g^{(0)}_{\mu\nu}(t) \, + \, g_{\mu\nu}(\mathbf{x}, t)\,; \qquad g_{\mu\nu} \ll g^{(0)}_{\mu\nu}\,. \tag{230}$$

The metric perturbations can be decomposed according to their spin with respect to a local rotation of the spatial coordinates on hypersurfaces of constant time. This leads to

- *scalar perturbations*
- *vector perturbations*
- *tensor perturbations*

Tensor perturbations or gravitational waves have spin 2 and are the "true" degrees of freedom of the gravitational fields in the sense that they can exist even in the vacuum. Vector perturbations are spin 1 modes arising from rotational velocity fields and are also called vorticity modes. Finally, scalar perturbations have spin 0.

Let us make a simple exercise to count how many scalar degrees of freedom are present. Take a space-time of dimensions $D = n + 1$, of which n coordinates are spatial coordinates. The symmetric metric tensor $g_{\mu\nu}$ has $\frac{1}{2}(n + 2)(n + 1)$ degrees of freedom. We can perform $(n + 1)$ coordinate transformations in order to eliminate $(n+1)$ degrees of freedom, this leaves us with $\frac{1}{2}n(n+1)$ degrees of freedom. These $\frac{1}{2}n(n+1)$ degrees of freedom contain scalar, vector and tensor modes. According to Helmholtz's theorem we can always decompose a vector U_i $(i = 1, \cdots, n)$ as $U_i = \partial_i v + v_i$, where v is a scalar (usually called potential flow) which is curl-free, $v_{[i,j]} = 0$, and v_i is a real vector (usually called vorticity) which is divergence-free, $\nabla \cdot v = 0$. This means that the real vector (vorticity) modes are $(n-1)$. Furthermore, a generic traceless tensor Π_{ij} can always be decomposed as $\Pi_{ij} = \Pi^S_{ij} + \Pi^V_{ij} + \Pi^T_{ij}$, where $\Pi^S_{ij} = \left(-\frac{k_i k_j}{k^2} + \frac{1}{3}\delta_{ij}\right)\Pi$, $\Pi^V_{ij} = (-i/2k)\,(k_i\Pi_j + k_j\Pi_i)$ $(K_i\Pi_i = 0)$ and $K_i\Pi^T_{ij} = 0$. This means that the true symmetric, traceless and transverse tensor degreees of freedom are $\frac{1}{2}(n-2)(n+1)$.

The number of scalar degrees of freedom are therefore

$$\frac{1}{2}n(n+1) - (n-1) - \frac{1}{2}(n-2)(n+1) = 2,$$

while the degrees of freedom of true vector modes are $(n-1)$ and the number of degrees of freedom of true tensor modes (gravitational waves) are $\frac{1}{2}(n-2)(n+1)$. In four dimensions $n = 3$, meaning that one expects 2 scalar degrees of freedom, 2 vector degrees of freedom and 2 tensor degrees of freedom. As we shall see, to the 2 scalar degrees of freedom from the metric, one has to add an another one, the inflaton field perturbation $\delta\phi$. However, since Einstein's equations will tell us that the two scalar degrees of freedom from the metric are equal during inflation, we expect a total number of scalar degrees of freedom equal to 2.

At the linear order, the scalar, vector and tensor perturbations evolve independently (they decouple) and it is therefore possible to analyze them separately. Vector perturbations are not excited during inflation because there are no rotational velocity fields during the inflationary stage. We will analyze the generation of tensor modes (gravitational waves) in the following. For the time being we want to focus on the scalar degrees of freedom of the metric.

Considering only the scalar degrees of freedom of the perturbed metric, the most generic perturbed metric reads

$$g_{\mu\nu} = a^2 \begin{pmatrix} -1 - 2\,\Phi & \partial_i B \\ \partial_i B & (1 - 2\,\Psi)\,\delta_{ij} + D_{ij}E \end{pmatrix}, \tag{231}$$

while the line-element can be written as

$$\mathrm{d}s^2 = a^2\big((-1-2\,\Phi)\mathrm{d}\tau^2 + 2\,\partial_i B\,\mathrm{d}\tau\,\mathrm{d}x^i + ((1-2\,\Psi)\delta_{ij} + D_{ij}E)\,\mathrm{d}x^i\,\mathrm{d}x^j\big). \tag{232}$$

Here $D_{ij} = \left(\partial_i\partial_j - \frac{1}{3}\,\delta_{ij}\,\boldsymbol{\nabla}^2\right)$.

We now want to determine the inverse $g^{\mu\nu}$ of the metric at the linear order

$$g^{\mu\alpha}\,g_{\alpha\nu} = \delta^\mu_\nu. \tag{233}$$

We have therefore to solve the equations

$$\left(g^{\mu\alpha}_{(0)} + g^{\mu\alpha}\right)\left(g^{(0)}_{\alpha\nu} + g_{\alpha\nu}\right) = \delta^\mu_\nu\,, \tag{234}$$

where $g^{\mu\alpha}_{(0)}$ is simply the unperturbed FRW metric. Since

$$g^{\mu\nu}_{(0)} = \frac{1}{a^2}\begin{pmatrix} -1 & 0 \\ 0 & \delta^{ij} \end{pmatrix}, \tag{235}$$

we can write in general

$$\begin{aligned} g^{00} &= \frac{1}{a^2}\,(-1 + X)\,; \\ g^{0i} &= \frac{1}{a^2}\,\partial^i Y\,; \\ g^{ij} &= \frac{1}{a^2}\,\big((1 + 2\,Z)\,\delta^{ij} + D^{ij}K\big)\,. \end{aligned} \tag{236}$$

Plugging these expressions into Eq. (234) we find for $\mu = \nu = 0$

$$(-1 + X)(-1 - 2\,\Phi) + \partial^i Y\,\partial_i B = 1. \tag{237}$$

Neglecting the terms $-2\,\Phi\cdot X$ e $\partial^i Y\cdot\partial_i B$ because they are second-order in the perturbations, we find

$$1\,-\,X\,+\,2\,\Phi\,=\,1\qquad\Rightarrow\qquad X\,=\,2\,\Phi\,. \tag{238}$$

Analogously, the components $\mu=0$, $\nu=i$ of Eq. (234) give

$$(-1\,+\,2\,\Phi)(\partial_i B)\,+\,\partial^j Y\,[(1\,-\,2\,\Psi)\delta_{ji}\,+\,D_{ji}E]\,=\,0. \tag{239}$$

At the first-order, we obtain

$$-\partial_i B\,+\,\partial_i Y\,=\,0\qquad\Rightarrow\qquad Y=B\,. \tag{240}$$

Finally, the components $\mu=i,\quad\nu=j$ give

$$\partial^i B\,\partial_j B\,+\,\Big((1\,+\,2\,Z)\delta^{ik}\,+\,D^{ik}K\Big)\,((1-2\,\Psi)\delta_{kj}\,+\,D_{kj}E)\,=\,\delta^i_j. \tag{241}$$

Neglecting the second-order terms, we obtain

$$(1\,-\,2\,\Psi\,+\,2\,Z)\delta^i_j\,+\,D^i_jE\,+\,D^i_jK\,=\,\delta^i_j\Rightarrow Z\,=\,\Psi\,;\qquad K\,=\,-\,E\,. \tag{242}$$

The metric $g^{\mu\nu}$ finally reads

$$g^{\mu\nu}\,=\,\frac{1}{a^2}\begin{pmatrix}-1\,+\,2\,\Phi & \partial^i B\\ \partial^i B & (1\,+\,2\,\Psi)\delta^{ij}\,-\,D^{ij}E\end{pmatrix}. \tag{243}$$

8.2 Perturbed affine connections and Einstein's tensor

In this subsection we provide the reader with the perturbed affine connections and Einstein's tensor. First, let us list the unperturbed affine connections

$$\Gamma^0_{00}\,=\,\frac{a'}{a}\,;\qquad\Gamma^i_{0j}\,=\,\frac{a'}{a}\,\delta^i_j\,;\qquad\Gamma^0_{ij}\,=\,\frac{a'}{a}\,\delta_{ij}\,; \tag{244}$$

$$\Gamma^i_{00}\quad=\quad\Gamma^0_{0i}\quad=\quad\Gamma^i_{jk}\quad=\quad 0\,. \tag{245}$$

The expression for the affine connections in terms of the metric is

$$\Gamma^\alpha_{\beta\gamma}\,=\,\frac{1}{2}\,g^{\alpha\rho}\left(\frac{\partial g_{\rho\gamma}}{\partial x^\beta}\,+\,\frac{\partial g_{\beta\rho}}{\partial x^\gamma}\,-\,\frac{\partial g_{\beta\gamma}}{\partial x^\rho}\right) \tag{246}$$

which implies

$$\begin{aligned}\delta\Gamma^\alpha_{\beta\gamma}\quad&=\quad\frac{1}{2}\,\delta g^{\alpha\rho}\left(\frac{\partial g_{\rho\gamma}}{\partial x^\beta}\,+\,\frac{\partial g_{\beta\rho}}{\partial x^\gamma}\,-\,\frac{\partial g_{\beta\gamma}}{\partial x^\rho}\right)\\ &+\quad\frac{1}{2}\,g^{\alpha\rho}\left(\frac{\partial\delta g_{\rho\gamma}}{\partial x^\beta}\,+\,\frac{\partial\delta g_{\beta\rho}}{\partial x^\gamma}\,-\,\frac{\partial\delta g_{\beta\gamma}}{\partial x^\rho}\right),\end{aligned} \tag{247}$$

or in components

$$\delta\Gamma^0_{00}\,=\,\Phi'\,; \tag{248}$$

$$\delta\Gamma^0_{0i} \, = \, \partial_i \, \Phi \, + \, \frac{a'}{a} \partial_i \, B \, ; \tag{249}$$

$$\delta\Gamma^i_{00} \, = \, \frac{a'}{a} \, \partial^i B \, + \, \partial^i B' \, + \, \partial^i \Phi \, ; \tag{250}$$

$$\delta\Gamma^0_{ij}, = - \, 2 \, \frac{a'}{a} \, \Phi \, \delta_{ij} \, - \, \partial_i \partial_j B \, - \, 2 \, \frac{a'}{a} \, \psi \, \delta_{ij} - \, \Psi' \, \delta_{ij} \, - \, \frac{a'}{a} \, D_{ij} E \, + \, \frac{1}{2} \, D_{ij} E' \, ;$$

$$\delta\Gamma^i_{0j} \, = \, - \, \Psi' \delta_{ij} \, + \, \frac{1}{2} \, D_{ij} E' \; ; \tag{251}$$

$$\delta\Gamma^i_{jk} = \partial_j \Psi \, \delta^i_k \, - \, \partial_k \Psi \, \delta^i_j \, + \, \partial^i \Psi \, \delta_{jk} \, - \, \frac{a'}{a} \, \partial^i B \, \delta_{jk} + \, \frac{1}{2} \, \partial_j D^i_k E \, + \, \frac{1}{2} \, \partial_k D^i_j E \, - \, \frac{1}{2} \, \partial^i D_{jk} E \, . \tag{252}$$

We may now compute the Ricci scalar defines as

$$R_{\mu\nu} \, = \, \partial_\alpha \, \Gamma^\alpha_{\mu\nu} \, - \, \partial_\mu \, \Gamma^\alpha_{\nu\alpha} \, + \, \Gamma^\alpha_{\sigma\alpha} \, \Gamma^\sigma_{\mu\nu} \, - \, \Gamma^\alpha_{\sigma\nu} \, \Gamma^\sigma_{\mu\alpha} \, . \tag{253}$$

Its variation at the first-order reads

$$\begin{aligned} \delta R_{\mu\nu} \quad &= \quad \partial_\alpha \, \delta\Gamma^\alpha_{\mu\nu} \, - \, \partial_\mu \, \delta\Gamma^\alpha_{\nu\alpha} \, + \, \delta\Gamma^\alpha_{\sigma\alpha} \, \Gamma^\sigma_{\mu\nu} \, + \, \Gamma^\alpha_{\sigma\alpha} \, \delta\Gamma^\sigma_{\mu\nu} \\ &- \quad \delta\Gamma^\alpha_{\sigma\nu} \, \Gamma^\sigma_{\mu\alpha} \, - \, \Gamma^\alpha_{\sigma\nu} \, \delta\Gamma^\sigma_{\mu\alpha} \, . \end{aligned} \tag{254}$$

The background values are given by

$$R_{00} \, = \, - \, 3 \, \frac{a''}{a} \, + \, 3 \left(\frac{a'}{a} \right)^2 ; \qquad R_{0i} \, = \, 0 \, ; \tag{255}$$

$$R_{ij} \, = \, \left(\frac{a''}{a} \, + \, \left(\frac{a'}{a} \right)^2 \right) \delta_{ij} \tag{256}$$

which give

$$\delta R_{00} = \frac{a'}{a} \partial_i \partial^i B + \partial_i \partial^i B' + \partial_i \partial^i \Phi + 3\Psi'' + 3\frac{a'}{a}\Psi' + 3\frac{a'}{a}\Phi' \; ; \tag{257}$$

$$\delta R_{0i} = \frac{a''}{a} \partial_i B + \left(\frac{a'}{a} \right)^2 \partial_i B + 2\partial_i \Psi' + 2\frac{a'}{a} \partial_i \Phi + \frac{1}{2} \partial_k D^K_i E' \; ; \tag{258}$$

$$\begin{aligned} \delta R_{ij} \quad &= \quad \Bigg(- \frac{a'}{a} \Phi' - 5 \frac{a'}{a} \psi' - 2 \frac{a''}{a} \Phi - 2 \left(\frac{a'}{a} \right)^2 \Phi \\ &- \quad 2 \frac{a''}{a} \Psi - 2 \left(\frac{a'}{a} \right)^2 \Psi - \Psi'' + \partial_k \partial^k \Psi - \frac{a'}{a} \partial_k \partial^k B \Bigg) \delta_{ij} \\ &- \quad \partial_i \partial_j B' + \frac{a'}{a} D_{ij} E' + \frac{a''}{a} D_{ij} E + \left(\frac{a'}{a} \right)^2 D_{ij} E \\ &+ \quad \frac{1}{2} D_{ij} E'' + \partial_i \partial_j \Psi - \partial_i \partial_j \Phi - 2 \frac{a'}{a} \partial_i \partial_j B \\ &+ \quad \frac{1}{2} \partial_k \partial_i D^k_j E + \frac{1}{2} \partial_k \partial_j D^k_i E - \frac{1}{2} \partial_k \partial^k D_{ij} E \; ; \end{aligned} \tag{259}$$

The perturbation of the scalar curvature

$$R \, = \, g^{\mu\alpha} \, R_{\alpha\mu} \, , \tag{260}$$

for which the first-order perturbation is

$$\delta R \,=\, \delta g^{\mu\alpha}\, R_{\alpha\mu} \,+\, g^{\mu\alpha}\, \delta R_{\alpha\mu}\,. \tag{261}$$

The background value is

$$R \,=\, \frac{6}{a^2}\,\frac{a''}{a} \tag{262}$$

while from Eq. (261) one finds

$$\begin{aligned}\delta R \;&=\; \frac{1}{a^2}\Big(-6\frac{a'}{a}\partial_i\partial^i B - 2\partial_i\partial^i B' - 2\partial_i\partial^i\Phi - 6\Psi'' \\ &-\; 6\frac{a'}{a}\Phi' - 18\frac{a'}{a}\Psi' - 12\frac{a''}{a}\Phi + 4\partial_i\partial^i\Psi + \partial_k\partial^i D_i^K E\Big).\end{aligned} \tag{263}$$

Finally, we may compute the perturbations of the Einstein tensor

$$G_{\mu\nu} \,=\, R_{\mu\nu} \,-\, \frac{1}{2}\, g_{\mu\nu}\, R\,, \tag{264}$$

whose background components are

$$G_{00} \,=\, 3\left(\frac{a'}{a}\right)^2; \qquad G_{0i} \,=\, 0\,; \qquad G_{ij} \,=\, \left(-\, 2\,\frac{a''}{a} \,+\, \left(\frac{a'}{a}\right)^2\right) \delta_{ij}\,. \tag{265}$$

At first-order, one finds

$$\delta G_{\mu\nu} \,=\, \delta R_{\mu\nu} \,-\, \frac{1}{2}\,\delta g_{\mu\nu}\, R \,-\, \frac{1}{2}\, g_{\mu\nu}\, \delta R\,, \tag{266}$$

or in components

$$\delta G_{00} = -2\frac{a'}{a}\,\partial_i\partial^i\, B - 6\frac{a'}{a}\,\Psi' + 2\,\partial_i\partial^i\,\Psi + \frac{1}{2}\,\partial_k\partial^i D_i^K E\,; \tag{267}$$

$$\delta G_{0i} = -2\frac{a''}{a}\,\partial_i B + \left(\frac{a'}{a}\right)^2 \partial_i B + 2\partial_i\,\Psi' + \frac{1}{2}\partial_k\, D_i^K E' + 2\,\frac{a'}{a}\,\partial_i\Phi\,; \tag{268}$$

$$\begin{aligned}\delta G_{ij} \;&=\; \Bigg(2\frac{a'}{a}\,\Phi' + 4\frac{a'}{a}\,\Psi' + 4\frac{a''}{a}\,\Phi - 2\left(\frac{a'}{a}\right)^2\Phi \\ &+\; 4\frac{a''}{a}\,\Psi - 2\left(\frac{a'}{a}\right)^2\Psi + 2\Psi'' - \partial_k\partial^k\,\Psi \\ &+\; 2\frac{a'}{a}\,\partial_k\partial^k B + \partial_k\partial^k B' + \partial_k\partial^k\Phi + \frac{1}{2}\partial_k\partial^m\, D_m^k\, E\Bigg)\delta_{ij} \\ &-\; \partial_i\partial_j B' + \partial_i\partial_j\Psi - \partial_i\partial_j A + \frac{a'}{a}\, D_{ij}E' - 2\,\frac{a''}{a}\, D_{ij}E \\ &+\; \left(\frac{a'}{a}\right)^2 D_{ij}E + \frac{1}{2}D_{ij}E'' + \frac{1}{2}\partial_k\partial_i D_j^k E \\ &+\; \frac{1}{2}\partial^k\partial_j D_{ik}E - \frac{1}{2}\partial_k\partial^k D_{ij}E - 2\,\frac{a'}{a}\partial_i\partial_j B\,.\end{aligned} \tag{269}$$

For convenience, we also give the expressions for the pertubations with one index up and one index down

$$\begin{aligned} \delta G^\mu_\nu &= \delta(g^{\mu\alpha}\, G_{\alpha\nu}) \\ &= \delta g^{\mu\alpha}\, G_{\alpha\nu} \,+\, g^{\mu\alpha}\, \delta G_{\alpha\nu}\,, \end{aligned} \tag{270}$$

or in components

$$\delta G^0_0 \,=\frac{1}{a^2}\left[6\left(\frac{a'}{a}\right)^2\Phi \,+\, 6\,\frac{a'}{a}\Psi' \,+\, 2\,\frac{a'}{a}\partial_i\,\partial^i B \,-\, 2\,\partial_i\,\partial^i\Psi \,-\, \frac{1}{2}\,\partial_k\partial^i\, D^K_i E\right]. \tag{271}$$

$$\delta G^0_i \,=\, \frac{1}{a^2}\left[-2\,\frac{a'}{a}\partial_i\Phi \,-\, 2\,\partial_i\Psi' \,-\, \frac{1}{2}\,\partial_k D^K_i E'\right]. \tag{272}$$

$$\begin{aligned} \delta G^i_j \;&=\; \frac{1}{a^2}\Bigg[\Bigg(2\,\frac{a'}{a}\Phi' \,+\, 4\,\frac{a''}{a}\Phi \,-\, 2\left(\frac{a'}{a}\right)^2\Phi \,+\, \partial_i\,\partial^i\Phi \,+\, 4\,\frac{a'}{a}\Psi' \,+\, 2\,\Psi'' \\ &-\; \partial_i\,\partial^i\Psi \,+\, 2\,\frac{a'}{a}\partial_i\,\partial^i B \,+\, \partial_i\,\partial^i B' \,+\, \frac{1}{2}\partial_k\partial^m D^k_m E\Bigg)\delta^i_j \\ &-\; \partial^i\partial_j\Phi \,+\, \partial^i\partial_j\Psi \,-\, 2\,\frac{a'}{a}\partial^i\partial_j B \,-\, \partial^i\partial_j B' \,+\, \frac{a'}{a}D^i_j E' \,+\, \frac{1}{2}\,D^i_j E'' \\ &+\; \frac{1}{2}\,\partial_k\partial^i\, D^k_j E \,+\, \frac{1}{2}\,\partial_k\partial_j\, D^{ik} E \,-\, \frac{1}{2}\,\partial_k\partial^k\, D^i_j E\Bigg]. \end{aligned} \tag{273}$$

8.3 Perturbed stress energy-momentum tensor

As we have seen previously, the perturbations of the metric are induced by the perturbations of the stress energy-momentum tensor of the inflaton field

$$T_{\mu\nu} \,=\, \partial_\mu\phi\,\partial_\nu\phi \,-\, g_{\mu\nu}\left(\frac{1}{2}\,g^{\alpha\beta}\,\partial_\alpha\phi\,\partial_\beta\phi \,+\, V(\phi)\right), \tag{274}$$

whose background values are

$$\begin{aligned} T_{00} &= \frac{1}{2}\,\phi'^2 \,+\, V(\phi)\,a^2\,; \\ T_{0i} &= 0\,; \\ T_{ij} &= \left(\frac{1}{2}\,\phi'^2 \,-\, V(\phi)\,a^2\right)\,\delta_{ij}\,. \end{aligned} \tag{275}$$

The perturbed stress energy-momentum tensor reads

$$\begin{aligned} \delta T_{\mu\nu} \;&=\; \partial_\mu\delta\phi\,\partial_\nu\phi \,+\, \partial_\mu\phi\,\partial_\nu\delta\phi \,-\, \delta g_{\mu\nu}\left(\frac{1}{2}\,g^{\alpha\beta}\,\partial_\alpha\phi\,\partial_\beta\phi \,+\, V(\phi)\right) \\ &-\; g_{\mu\nu}\left(\frac{1}{2}\delta g^{\alpha\beta}\,\partial_\alpha\phi\,\partial_\beta\phi \,+\, g^{\alpha\beta}\,\partial_\alpha\delta\phi\,\partial_\beta\phi \,+\, \frac{\partial V}{\partial\phi}\,\delta\phi \,+\, \frac{\partial V}{\partial\phi}\delta\phi\right). \end{aligned} \tag{276}$$

In components we have

$$\delta T_{00} \,=\, \delta\phi'\,\phi' \,+\, 2\,\Phi\,V(\phi)\,a^2 \,+\, a^2\,\frac{\partial V}{\partial \phi}\,\delta\phi\,; \tag{277}$$

$$\delta T_{0i} \,=\, \partial_i\,\delta\phi\,\phi' \,+\, \frac{1}{2}\,\partial_i B\,\phi'^2 \,-\, \partial_i B\,V(\phi)\,a^2\,; \tag{278}$$

$$\begin{aligned} \delta T_{ij} \;&=\; \left(\delta\phi'\,\phi' \,-\, \Phi\,\phi'^2 \,-\, a^2\,\frac{\partial V}{\partial\phi}\,\delta\phi \,-\, \Psi\,\phi'^2 \,+\, 2\,\Psi\,V(\phi)\,a^2\right)\delta_{ij} \\ &+\; \frac{1}{2}\,D_{ij}E\,\phi'^2 \,-\, D_{ij}E\,V(\phi)\,a^2\,. \end{aligned} \tag{279}$$

For covenience, we list the mixed components

$$\begin{aligned} \delta T^\mu_\nu \;&=\; \delta(g^{\mu\alpha}\,T_{\alpha\nu}) \\ &=\; \delta g^{\mu\alpha}\,T_{\alpha\nu} \,+\, g^{\mu\alpha}\,\delta T_{\alpha\nu} \end{aligned} \tag{280}$$

or

$$\begin{aligned} \delta T^0_0 \;&=\; \Phi\,\phi'^2 \,-\, \delta\phi'\,\phi' \,-\, \delta\phi\,\frac{\partial V}{\partial\phi}\,a^2\,; \\ \delta T^i_0 \;&=\; \partial^i B\,\phi'^2 \,+\, \partial^i\delta\phi\,\phi'\,; \\ \delta T^0_i \;&=\; -\,\partial^i\delta\phi\,\phi'\,; \\ \delta T^i_j \;&=\; \left(-\,\Phi\,\phi'^2 \,+\, \delta\phi'\,\phi' \,-\, \delta\phi\,\frac{\partial V}{\partial\phi}\,a^2\right)\delta^i_j\,. \end{aligned} \tag{281}$$

8.4 Perturbed Klein-Gordon equation

The inflaton equation of motion is the Klein-Gordon equation of a scalar field under the action of its potential $V(\phi)$. The equation to perturb is therefore

$$\begin{aligned} \partial^\mu\partial_\mu\,\phi \;&=\; \frac{\partial V}{\partial\phi}\,; \\ \partial_\mu\partial^\mu\phi \;&=\; \frac{1}{\sqrt{-g}}\,\partial_\nu\left(\sqrt{-g}\,g^{\mu\nu}\,\partial_\nu\phi\right)\,; \end{aligned} \tag{282}$$

which at the zero-th order gives the inflaton equation of motion

$$\phi'' \,+\, 2\,\frac{a'}{a}\,\phi' \,=\, -\,\frac{\partial V}{\partial\phi}\,a^2\,. \tag{283}$$

The variation of Eq. (282) is the sum of four different contributions corresponding to the variations of $\frac{1}{\sqrt{-g}}$, $\sqrt{-g}$, $g^{\mu\nu}$ and ϕ. For the variation of g we have

$$\delta g \,=\, g\,g^{\mu\nu}\delta g_{\nu\mu} \tag{284}$$

which give at the linear order

$$\begin{aligned} \delta\sqrt{-g} \;&=\; -\,\frac{\delta g}{2\,\sqrt{-g}}\,; \\ \delta\frac{1}{\sqrt{-g}} \;&=\; \frac{\delta\sqrt{-g}}{g}\,. \end{aligned} \tag{285}$$

Plugging these results into the expression for the variation of Eq. (283)

$$\begin{aligned}\delta\partial_\mu\partial^\mu\,\phi &= -\,\delta\phi'' - 2\,\frac{a'}{a}\delta\phi' + \partial_i\,\partial^i\delta\phi + 2\,\Phi\,\phi'' + 4\,\frac{a'}{a}\Phi\,\phi' + \Phi'\phi' \\ &+ 3\,\Psi'\phi' + \partial_i\,\partial^i B\,\phi' \\ &= \delta\phi\,\frac{\partial^2 V}{\partial\phi^2}\,a^2\,. \end{aligned} \tag{286}$$

Using Eq. (283) to write

$$2\,\Phi\,\phi'' + 4\,\frac{a'}{a}\phi' = 2\,\Phi\,\frac{\partial V}{\partial\phi}\,, \tag{287}$$

Eq. (286) becomes

$$\begin{aligned}\delta\phi'' + 2\,\frac{a'}{a}\delta\phi' &- \partial_i\,\partial^i\delta\phi - \Phi'\phi' - 3\,\Psi'\phi' - \partial_i\,\partial^i B\,\phi' \\ &= -\delta\phi\,\frac{\partial^2 V}{\partial\phi^2}\,a^2 - 2\,\Phi\,\frac{\partial V}{\partial\phi}. \end{aligned} \tag{288}$$

After having computed the perturbations at the linear order of the Einstein's tensor and of the stress energy-momentum tensor, we are ready to solve the perturbed Einstein's equations in order to quantify the inflaton and the metric fluctuations. We pause, however, for a moment in order to deal with the problem of gauge invariance.

8.5 The issue of gauge invariance

When studying the cosmological density perturbations, what we are interested in is following the evolution of a space-time which is neither homogeneous nor isotropic. This is done by following the evolution of the differences between the actual space-time and a well understood reference space-time. So we will consider small perturbations away from the homogeneous, isotropic space-time (see Fig. 15). The reference system in our case is the spatially flat Friedmann–Robertson–Walker space-time, with line element $\mathrm{d}s^2 = a^2(\tau)\left(-\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2\right)$. Now, the key issue is that general relativity is a gauge theory where the gauge transformations are the generic coordinate transformations from a local reference frame to another.

When we compute the perturbation of a given quantity, this is defined to be the difference between the value that this quantity assumes on the real physical space-time and the value it assumes on the unperturbed background. Nonetheless, to perform a comparison between these two values, it is necessary to compute the at the same space-time point. Since the two values "live" on two different geometries, it is necessary to specify a map which allows to link univocally the same point on the two different space-times. This correspondance is called a gauge choice and changing the map means performing a gauge transformation.

Fixing a gauge in general relativity implies choosing a coordinate system. A choice of coordinates defines a *threading* of space-time into lines (corresponding to fixed spatial coordinates $\mathbf{x}$) and a *slicing*

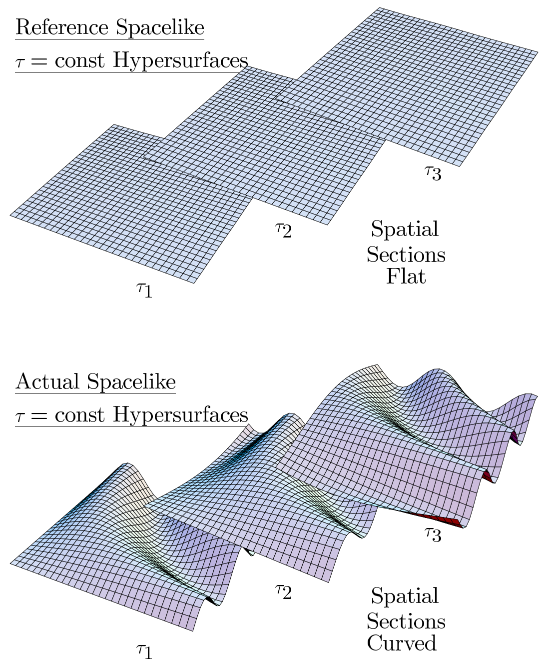

Figure 15: In the reference unperturbed universe, constant-time surfaces have constant spatial curvature (zero for a flat FRW model). In the actual perturbed universe, constant-time surfaces have spatially varying spatial curvature

into hypersurfaces (corresponding to fixed time τ). A choice of coordinates is is called a *gauge* and there is no unique preferred gauge

GAUGE CHOICE $\Longleftrightarrow$ SLICING AND THREADING

Similarly, we can look at the change of coordinates either as an active transformation, in which we slightly alter the manifold or as a passive transformation, in which we do not alter the manifold, all the points remain fixed, and we just change the coordinate system. So this is tantamount to a relabelling of the points. From the passive point of view, in which a coordinate transformation represents a relabelling of the points of the space, one then compares a quantity, say the metric (or its perturbations), at a point P (with coordinates x^μ) with the new metric at the point P' which has the same values of the new coordinates as the point P had in the old coordinate system, $\widetilde{x^\mu}(P') = x^\mu(P)$. This is by the way an efficient way to detect symmetries (isometries if one is concerned with the metric), we only need to consider infinitesimal coordinate transformations.

From a more formal point of view, operating an infinitesimal gauge tranformation on the coordinates

$$\widetilde{x^\mu} = x^\mu + \delta x^\mu \tag{289}$$

implies on a generic quantity Q a tranformation on its perturbation

$$\widetilde{\delta Q} = \delta Q + \pounds_{\delta x} Q_0 \tag{290}$$

where Q_0 is the value assumed by the quantity Q on the background and $\pounds_{\delta x}$ is the Lie-derivative of Q along the vector δx^μ. Notice that for a scalar, the Lie derivative is just the ordinary directional derivative (and this is as it should be since saying that a function has a certain symmetry amounts to the assertion that its derivative in a particular direction vanishes).

Decomposing in the usual manner the vector δx^μ

$$\begin{aligned}
\delta x^0 &= \xi^0(x^\mu)\,; \\
\delta x^i &= \partial^i\beta(x^\mu) + v^i(x^\mu)\,; \qquad \partial_i v^i = 0\,,
\end{aligned} \tag{291}$$

we can easily deduce the transformation law of a scalar quantity f (like the inflaton scalar field ϕ and energy density ρ). Instead of applying the formal definition (290), we find the transformation law in an alternative (and more pedagogical) way. We first write $\delta f(x) = f(x) - f_0(x)$, where $f_0(x)$ is the background value. Under a gauge transformation we have $\widetilde{\delta f}(\widetilde{x^\mu}) = \widetilde{f}(\widetilde{x^\mu}) - \widetilde{f}_0(\widetilde{x^\mu})$. Since f is a scalar we can write $f(\widetilde{x^\mu}) = f(x^\mu)$ (the value of the scalar function in a given physical point is the same in all the coordinate system). On the other side, on the unperturbed background hypersurface $\widetilde{f}_0 = f_0$. We have therefore

$$\begin{aligned}
\widetilde{\delta f}(\widetilde{x^\mu}) &= \widetilde{f}(\widetilde{x^\mu}) - \widetilde{f}_0(\widetilde{x^\mu}) \\
&= f(x^\mu) - f_0(\widetilde{x^\mu}) \\
&= f(x^\mu) - \delta x^\mu \frac{\partial f_0}{\partial x^\mu}(x) - f_0(x^\mu),
\end{aligned} \tag{292}$$

from which we finally deduce, being $f_0 = f_0(x^0)$,

$$\boxed{\widetilde{\delta f} = \delta f - f_0'\,\xi^0}$$

For the spin zero perturbations of the metric, we can proceed analogously. We use the following trick. Upon a coordinate transformation $x^\mu \to \widetilde{x^\mu} = x^\mu + \delta x^\mu$, the line element is left invariant, $\mathrm{d}s^2 = \widetilde{\mathrm{d}s^2}$. This implies, for instance, that $a^2(\widetilde{x^0})\left(1 + 2\widetilde{\Phi}\right)\left(\widetilde{\mathrm{d}x^0}\right)^2 = a^2(x^0)\,(1 + 2\Phi)\,(\mathrm{d}x^0)^2$. Since $a^2(\widetilde{x^0}) \simeq a^2(x^0) + 2a\,a'\,\xi^0$ and $\widetilde{\mathrm{d}x^0} = \left(1 + \xi^{0\prime}\right)\mathrm{d}x^0 + \frac{\partial x^0}{\partial x^i}\,\mathrm{d}x^i$, we obtain $1 + 2\Phi = 1 + 2\widetilde{\Phi} + 2\,\mathcal{H}\xi^0 + 2\xi^{0\prime}$. A similar procedure leads to the following transformation laws

$$
\begin{aligned}
\widetilde{\Phi} &= \Phi - \xi^{0\prime} - \frac{a'}{a}\xi^0\,; \\
\widetilde{B} &= B + \xi^0 + \beta' \\
\widetilde{\Psi} &= \Psi - \frac{1}{3}\nabla^2\beta + \frac{a'}{a}\xi^0\,; \\
\widetilde{E} &= E + 2\,\beta\,.
\end{aligned}
$$

The gauge problem stems from the fact that a change of the map (a change of the coordinate system) implies the variation of the perturbation of a given quantity which may therefore assume different values (all of them on a equal footing!) according to the gauge choice. To eliminate this ambiguity, one has therefore a double choice:

- Indentify those combinations representing gauge invariant quantities;
- choose a given gauge and perform the calculations in that gauge.

Both options have advantages and drawbacks. Choosing a gauge may render the computation technically simpler with the danger, however, of including gauge artifacts, *i.e.* gauge freedoms which are not physical. Performing a gauge-invariant computation may be technically more involved, but has the advantage of treating only physical quantities.

Let us first indicate some gauge-invariant quantities. They are the so-called gauge invariant potentials or Bardeen's potentials

$$
\Phi_{\rm GI} = -\Phi + \frac{1}{a}\left[\left(-B + \frac{E'}{2}\right)a\right]'\,, \tag{293}
$$

$$
\Psi_{\rm GI} = -\Psi - \frac{1}{6}\nabla^2 E + \frac{a'}{a}\left(B - \frac{E'}{2}\right)\,. \tag{294}
$$

Analogously, one can define a gauge invariant quantity for the perturbation of the inflaton field. Since ϕ is a scalar field $\widetilde{\delta\phi} = \left(\delta\phi - \phi'\,\xi^0\right)$ and therefore

$$
\delta\phi_{\rm GI} = -\delta\phi + \phi'\left(\frac{E'}{2} - B\right).
$$

is gauge-invariant. Analogously, one can define a gauge-invariant energy-density perturbation

$$
\delta\rho_{\rm GI} = -\delta\rho + \rho'\left(\frac{E'}{2} - B\right).
$$

We now want to pause to introduce in details some gauge-invariant quantities which play a major role in the computation of the density perturbations. In the following we will be interested only in the coordinate transformations on constant time hypersurfaces and therefore gauge invariance will be equivalent to independent of the slicing.

8.6 The comoving curvature perturbation

The intrinsic spatial curvature on hypersurfaces on constant conformal time τ and for a flat universe is given by

$$^{(3)}R = \frac{4}{a^2}\boldsymbol{\nabla}^2\,\Psi.$$

The quantity Ψ is usually referred to as the *curvature perturbation*. We have seen, however, that the the curvature potential Ψ is *not* gauge invariant, but is defined only on a given slicing. Under a transformation on constant time hypersurfaces $t \to t + \delta\tau$ (change of the slicing)

$$\Psi \to \Psi + \mathcal{H}\,\delta\tau.$$

We now consider the *comoving slicing* which is defined to be the slicing orthogonal to the worldlines of comoving observers. The latter are are free-falling and the expansion defined by them is isotropic. In practice, what this means is that there is no flux of energy measured by these observers, that is $T_{0i} = 0$. During inflation this means that these observers measure $\delta\phi_{\rm com} = 0$ since T_{0i} goes like $\partial_i\delta\phi(\mathbf{x},\tau)\phi'(\tau)$.

Since $\delta\phi \to \delta\phi - \phi'\delta\tau$ for a transformation on constant time hypersurfaces, this means that

$$\delta\phi \to \delta\phi_{\rm com} = \delta\phi - \phi'\,\delta\tau = 0 \Longrightarrow \delta\tau = \frac{\delta\phi}{\phi'},$$

that is $\delta\tau = \frac{\delta\phi}{\phi'}$ is the time-displacement needed to go from a generic slicing with generic $\delta\phi$ to the comoving slicing where $\delta\phi_{\rm com} = 0$. At the same time the curvature pertubation ψ transforms into

$$\Psi \to \Psi_{\rm com} = \Psi + \mathcal{H}\,\delta\tau = \Psi + \mathcal{H}\frac{\delta\phi}{\phi'}.$$

The quantity

$$\mathcal{R} = \Psi + \mathcal{H}\frac{\delta\phi}{\phi'} = \Psi + H\frac{\delta\phi}{\dot{\phi}}$$

is the *comoving curvature perturbation.* This quantity is gauge invariant by construction and is related to the gauge-dependent curvature perturbation ψ on a generic slicing to the inflaton perturbation $\delta\phi$ in that gauge. By construction, the meaning of $\mathcal{R}$ is that it represents the gravitational potential on comoving hypersurfaces where $\delta\phi = 0$

$$\mathcal{R} = \Psi|_{\delta\phi=0}\,.$$

8.7 The curvature perturbation on spatial slices of uniform energy density

We now consider the *slicing of uniform energy density* which is defined to be the the slicing where there is no perturbation in the energy density, $\delta\rho = 0$.

Since $\delta\rho \to \delta\rho - \rho'\,\delta\tau$ for a transformation on constant time hypersurfaces, this means that

$$\delta\rho \to \delta\rho_{\rm unif} = \delta\rho - \rho'\,\delta\tau = 0 \Longrightarrow \delta\tau = \frac{\delta\rho}{\rho'},$$

that is $\delta\tau = \frac{\delta\rho}{\rho'}$ is the time-displacement needed to go from a generic slicing with generic $\delta\rho$ to the slicing of uniform energy density where $\delta\rho_{\rm unif} = 0$. At the same time the curvature pertubation ψ transforms into

$$\Psi \to \Psi_{\rm unif} = \psi + \mathcal{H}\,\delta\tau = \Psi + \mathcal{H}\frac{\delta\rho}{\rho'}.$$

The quantity

$$\boxed{\zeta = \Psi + \mathcal{H}\frac{\delta\rho}{\rho'} = \Psi + H\frac{\delta\rho}{\dot{\rho}}}$$

is the *curvature perturbation on slices of uniform energy density.* This quantity is gauge invariant by construction and is related to the gauge-dependent curvature perturbation Ψ on a generic slicing and to the energy density perturbation $\delta\rho$ in that gauge. By construction, the meaning of ζ is that it represents the gravitational potential on slices of uniform energy density

$$\zeta = \Psi|_{\delta\rho=0}\,.$$

Notice that, using the energy-conservation equation $\rho' + 3\,\mathcal{H}(\rho + P) = 0$, the curvature perturbation on slices of uniform energy density can be also written as

$$\zeta = \Psi - \frac{\delta\rho}{3(\rho + P)}.$$

During inflation $\rho + P = \dot{\phi}^2$. Furthermore, on superHubble scales from what we have learned in the previous section (and will be rigously shown in the following) the inflaton fluctuation $\delta\phi$ is frozen in and $\delta\dot{\phi} =$ (slow roll parameters) $\times\, H\,\delta\phi$. This implies that $\delta\rho = \dot{\phi}\delta\dot{\phi} + V'\delta\phi \simeq V'\delta\phi \simeq -3H\dot{\phi}$, leading to

$$\zeta \simeq \Psi + \frac{3H\dot{\phi}}{3\dot{\phi}^2} = \Psi + H\frac{\delta\phi}{\dot{\phi}}\mathcal{R} \qquad \text{(ON superHubble SCALES)}$$

The comoving curvature pertubation and the curvature perturbation on uniform energy density slices are equal on superHubble scales.

8.8 Scalar field perturbations in the spatially flat gauge

We now consider the *spatially flat gauge* which is defined to be the the slicing where there is no curvature $\Psi_{\rm flat} = 0$.

Since $\psi \to \Psi + \mathcal{H}\,\delta\tau$ for a transformation on constant time hypersurfaces, this means that

$$\Psi \to \Psi_{\rm flat} = \Psi + \mathcal{H}\,\delta\tau = 0 \Longrightarrow \delta\tau = -\frac{\Psi}{\mathcal{H}},$$

that is $\delta\tau = -\Psi/\mathcal{H}$ is the time-displacement needed to go from a generic slicing with generic ψ to the spatially flat gauge where $\Psi_{\rm flat} = 0$. At the same time the fluctuation of the inflaton field transforms a

$$\delta\phi \to \delta\phi - \phi'\,\delta\tau = \delta\phi + \frac{\phi'}{\mathcal{H}}\,\Psi.$$

The quantity

$$\boxed{Q = \delta\phi + \frac{\phi'}{\mathcal{H}}\,\Psi = \delta\phi + \frac{\dot{\phi}}{H}\Psi \equiv \frac{\dot{\phi}}{H}\,\mathcal{R}}$$

is the inflaton perturbation on spatially flat gauges. This quantity is gauge invariant by construction and is related to the inflaton perturbation $\delta\phi$ on a generic slicing and to to the curvature perturbation Ψ in that gauge. By construction, the meaning of Q is that it represents the inflaton potential on spatially flat slices

$$Q = \left.\delta\phi\right|_{\delta\Psi=0}.$$

Notice that $\delta\phi = -\phi'\delta\tau = -\dot{\phi}\delta t$ on flat slices, where δt is the time displacement going from flat to comoving slices. This relation makes somehow rigorous the expression (229). Analogously, going from flat to comoving slices one has $\mathcal{R} = H\,\delta t$.

8.9 Comments about gauge invariance

While comparing the theoretical predictions (*e.g.* the CMB power spectrum) with obervations does not represent a problem on sub-horizon scales where the matter perturbations computed in the different gauges all coincide, it is a delicate operation on scales comparable with the horizon where different gauges provide different results even at the linear level. Truly gauge-independent perturbations must be exactly constant in the background space-time. This apparently limits ones ability to make a gauge-invariant study of quantities that evolve in the background space-time, *e.g.* density perturbations in an expanding cosmology. In practice one can construct gauge-invariant definitions of unambiguous, that is physically defined, perturbations. These are not unique gauge-independent perturbations, but are gauge-invariant in the sense commonly used by cosmologists to define a physical perturbation. There is a distinction between quantities that are automatically gauge-independent, *i.e.*, those that have no gauge dependence (such as perturbations about a constant scalar field), and quantities that are in general gauge-dependent (such as the curvature perturbation) but can have a gauge-invariant definition once their gauge-dependence is fixed (such as the curvature perturbation on uniform-density hypersurfaces). In other words, one can define gauge-invariant quantities which are simply a coordinate independent definition of the perturbations in the given gauge. This can be often achieved by defining unambiguously a specific slicing into spatial hypersurfaces. In this sense it should be clear that one may define an infinite number of, *e.g.*, gauge-invariant density contrasts. Which one to use is a matter that can be decided only considering how the determination of a given observable is performed.

8.10 Adiabatic and isocurvature perturbations

Arbitrary cosmological perturbations can be decomposed into:

- *adiabatic or curvature perturbations* which perturb the solution along the same trajectory in phase-space as the as the background solution. The perturbations in any scalar quantity X can be described by a unique perturbation in expansion with respect to the background

  $$H\,\delta t = H\,\frac{\delta X}{\dot{X}} \qquad \text{FOR EVERY } X.$$

  In particular, this holds for the energy density and the pressure

  $$\frac{\delta\rho}{\dot{\rho}} = \frac{\delta P}{\dot{P}}$$

  which implies that $P = P(\rho)$. This explains why they are called adiabatic. They are called curvature perturbations because a given time displacement δt causes the same relative change

$\delta X/\dot{X}$ for all quantities. In other words the perturbations is democratically shared by all components of the universe.

- *isocurvature perturbations* which perturb the solution off the background solution

$$\frac{\delta X}{\dot{X}} \neq \frac{\delta Y}{\dot{Y}} \quad \text{FOR SOME } X \text{ AND } Y.$$

One way of specifying a generic isocurvature perturbation δX is to give its value on uniform-density slices, related to its value on a different slicing by the gauge-invariant definition

$$H \left. \frac{\delta X}{\dot{X}} \right|_{\delta\rho=0} = H \left(\frac{\delta X}{\dot{X}} - \frac{\delta \rho}{\dot{\rho}} \right).$$

For a set of fluids with energy density ρ_i, the isocurvature perturbations are conventionally defined by the gauge invariant quantities

$$\boxed{S_{ij} = 3H \left(\frac{\delta \rho_i}{\dot{\rho}_i} - \frac{\delta \rho_j}{\dot{\rho}_j} \right) = 3\left(\zeta_i - \zeta_j\right).}$$

One simple example of isocurvature perturbations is the baryon-to-photon ratio

$$S = \delta(n_{\rm b}/n_\gamma) = \frac{\delta n_{\rm b}}{n_{\rm b}} - \frac{\delta n_\gamma}{n_\gamma}. \tag{295}$$

1. Comment:

From the definitions above, it follows that the cosmological perturbations generated during inflation are of the adiabatic type *if* the inflaton field is the only fiels driving inflation. However, if inflation is driven by more than one field, isocurvature perturbations are expected to be generated (and they might even be cross-correlated to the adiabatic ones. In the following we will give one example of the utility of generating isocurvature perturbations.

2. Comment: The perturbations generated during inflation are *Gaussian*, *i.e.* the two-point correlation functions (like the power spectrum) suffice to define all the higher-order even correlation fucntions, while the odd correlation functions (such as the three-point correlation function) vanish. This conclusion is drawn by the very same fact that cosmological perturbations are studied *linearizing* Einstein's and Klein-Gordon equations. This turns out to be a good approximation because we know that the inflaton potential needs to be very flat in order to drive inflation and the interaction terms in the inflaton potential might be present, but they are small. Non-Gaussian features are therefore suppressed since the non-linearities of the inflaton potential are suppressed too. The same argument applies to the metric perturbations; non-linearities appear only at the second-order in deviations from the homogeneous background solution and are therefore small. This expectation is confirmed by a direct computation of the cosmological perturbations generated during inflation up to second-order in deviations from the homogeneous background solution which fully account for the inflaton self-interactions as well as for the second-order fluctuations of the background metric. While the subject of non-Gaussianity is extremely interesting both theoretically and observationally, it goes beyond the scope of these lectures. The interested reader can ask more details during and/or after the lectures.

8.11 The next steps

After all these technicalities, it is useful to rest for a moment and to go back to physics. Up to now we have learned that during inflation quantum fluctuations of the inflaton field are generated and their wavelengths are stretched on large scales by the rapid expansion of the universe. We have also seen that the quantum fluctuations of the inflaton field are in fact impossible to disantagle from the metric perturbations. This happens not only because they are tightly coupled to each other through Einstein's equations, but also because of the issue of gauge invariance. Take, for instance, the gauge invariant quantity $Q = \delta\phi + \frac{\phi'}{\mathcal{H}}\,\Psi$. We can always go to a gauge where the fluctuation is entirely in the curvature potential Ψ, $Q = \frac{\phi'}{\mathcal{H}}\,\Psi$, or entirely in the inflaton field, $Q = \delta\phi$. However, as we have stressed at the end of the previous section, once ripples in the curvature are frozen in on superHubble scales during inflation, it is in fact gravity that acts as a messanger communicating to baryons and photons the small seeds of perturbations once a given scale reenters the horizon after inflation. This happens thanks to Newtonian physics; a small perturbation in the gravitational potential Ψ induces a small perturbation of the energy density ρ through Poisson's equiation $\boldsymbol{\nabla}^2\Psi = 4\pi G_{\rm N}\delta\rho$. Similarly, if perturbations are adiabatic/curvature perturbations and, as such, treat democratically all the components, a ripple in the curvature is communicated to photons as well, giving rise to a nonvanishing $\delta T/T$.

These considerations make it clear that the next steps of these lectures will be

- Compute the curvature perturbation generated during inflation on superHubble scales. As we have seen we can either compute the comoving curvature perturbation $\mathcal{R}$ or the curvature on uniform energy density hypersurfaces ζ. They will tell us about the fluctuations of the gravitational potential.
- See how the fluctuations of the gravitational potential are transmitted to photons, baryons and matter in general.

We now intend to address the first point. As stressed previously, we are free to follow two alternative roads: either pick up a gauge and compute the gauge-invariant curvature in that gauge or perform a gauge-invariant calculation. We take both options.

8.12 Computation of the curvature perturbation using the longitudinal gauge

The longitudinal (or conformal newtonian) gauge is a convenient gauge to compute the cosmological perturbations. It is defined by performing a coordinate transformation such that $B = E = 0$. This leaves behind two degrees of freedom in the scalar perturbations, Φ and Ψ. As we have previously seen, these two degrees of freedom fully account for the scalar perturbations in the metric.

First of all, we take the non-diagonal part ($i \neq j$) of Eq. (273). Since the stress energy-momentum tensor does not have any non-diagonal component (no stress), we have

$$\partial_i \partial_j (\Psi - \Phi) = 0 \Longrightarrow \Psi = \Phi$$

and we can now work only with one variable, let it be Ψ.

Eq. (272) gives (in cosmic time)

$$\dot{\Psi} + H\,\Psi = 4\pi G_{\rm N} \dot{\phi}\,\delta\phi = \epsilon\, H \frac{\delta\phi}{\dot{\phi}}, \tag{296}$$

while Eq. (271) and the diagonal part of (273) ($i = j$) give respectively

$$-3H\left(\dot{\Psi} + H\Psi\right) + \frac{\boldsymbol{\nabla}^2\Psi}{a^2} = 4\pi G_{\rm N}\left(\dot{\phi}\delta\dot{\phi} - \dot{\phi}^2\Psi + V'\delta\phi\right), \tag{297}$$

$$-\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right)\Psi - 3H\dot{\Psi} - \ddot{\Psi} = -\left(\dot{\phi}\delta\dot{\phi} - \dot{\phi}^2\Psi - V'\delta\phi\right), \tag{298}$$

If we now use the fact that $\dot{H} = 4\pi G_{\rm N}\dot{\phi}^2$, sum Eqs. (297) and 298) and use the background Klein-Gordon equation to eliminate V', we arrive at the equation for the gravitational potential

$$\ddot{\Psi}_{\mathbf{k}} + \left(H - 2\frac{\ddot{\phi}}{\dot{\phi}}\right)\dot{\Psi}_{\mathbf{k}} + 2\left(\dot{H} - H\frac{\ddot{\phi}}{\dot{\phi}}\right)\Psi_{\mathbf{k}} + \frac{k^2}{a^2}\Psi_{\mathbf{k}} = 0. \tag{299}$$

We may rewrite it in conformal time

$$\Psi''_{\mathbf{k}} + 2\left(\mathcal{H} - \frac{\phi''}{\phi'}\right)\Psi'_{\mathbf{k}} + 2\left(\mathcal{H}' - \mathcal{H}\frac{\phi''}{\phi'}\right)\Psi_{\mathbf{k}} + k^2\,\Psi_{\mathbf{k}} = 0 \tag{300}$$

and in terms of the slow-roll parameters ϵ and η

$$\Psi''_{\mathbf{k}} + 2\,\mathcal{H}\,(\eta - \epsilon)\,\Psi'_{\mathbf{k}} + 2\,\mathcal{H}^2\,(\eta - 2\epsilon)\,\Psi_{\mathbf{k}} + k^2\,\Psi_{\mathbf{k}} = 0. \tag{301}$$

Using the same logic leading to Eq. (226), from Eq. (299) we can infer that on superHubble scales the gravitational potential Ψ is nearly constant (up to a mild logarithmic time-dependence proportional to slow-roll parameters), that is $\dot{\Psi}_{\mathbf{k}} \sim$ (slow-roll parameters)$\times\Psi_{\mathbf{k}}$. This is hardly surprising, we know that fluctuations are frozen in on superHubble scales.

Using Eq. (296), we can therefore relate the fluctuation of the gravitational potential Ψ to the fluctuation of the inflaton field $\delta\phi$ on superHubble scales

$$\Psi_{\mathbf{k}} \simeq \epsilon\, H \frac{\delta\phi_{\mathbf{k}}}{\dot{\phi}} \quad \text{(ON superHubble SCALES)} \tag{302}$$

This gives us the chance to compute the gauge-invariant comoving curvature perturbation $\mathcal{R}_{\mathbf{k}}$

$$\mathcal{R}_{\mathbf{k}} = \Psi_{\mathbf{k}} + H\,\frac{\delta\phi_{\mathbf{k}}}{\dot{\phi}} = (1+\epsilon)\,\frac{\delta\phi_{\mathbf{k}}}{\dot{\phi}} \simeq \frac{\delta\phi_{\mathbf{k}}}{\dot{\phi}}. \tag{303}$$

The power spectrum of the the comoving curvature perturbation $\mathcal{R}_{\mathbf{k}}$ then reads on superHubble scales

$$\mathcal{P}_{\mathcal{R}} = \frac{k^3}{2\pi^2}\frac{H^2}{\dot{\phi}^2}\left|\delta\phi_{\mathbf{k}}\right|^2 = \frac{k^3}{4M_{\rm Pl}^2\epsilon\,\pi^2}\left|\delta\phi_{\mathbf{k}}\right|^2.$$

What is left to evaluate is the time evolution of $\delta\phi_{\mathbf{k}}$. To do so, we consider the perturbed Klein-Gordon equation (288) in the longitudinal gauge (in cosmic time)

$$\delta\ddot{\phi}_{\mathbf{k}} + 3H\delta\dot{\phi}_{\mathbf{k}} + \frac{k^2}{a^2}\delta\phi_{\mathbf{k}} + V''\delta\phi_{\mathbf{k}} = -2\Psi_{\mathbf{k}}V' + 4\dot{\Psi}_{\mathbf{k}}\dot{\phi}.$$

Since on superHubble scales $\left|4\dot{\Psi}_{\mathbf{k}}\dot{\phi}\right| \ll |\Psi_{\mathbf{k}}V'|$, using Eq. (302) and the relation $V' \simeq -3H\dot{\phi}$, we can rewrite the perturbed Klein-Gordon equation on superHubble scales as

$$\delta\ddot{\phi}_{\mathbf{k}} + 3H\delta\dot{\phi}_{\mathbf{k}} + \left(V'' + 6\epsilon H^2\right)\delta\phi_{\mathbf{k}} = 0.$$

We now introduce as usual the field $\delta\chi_{\mathbf{k}} = \delta\phi_{\mathbf{k}}/a$ and go to conformal time τ. The perturbed Klein-Gordon equation on superHubble scales becomes, using Eq. (220),

$$\begin{aligned} \delta\chi_{\mathbf{k}}'' &- \frac{1}{\tau^2}\left(\nu^2 - \frac{1}{4}\right)\delta\chi_{\mathbf{k}} = 0, \\ \nu^2 &= \frac{9}{4} + 9\epsilon - 3\eta. \end{aligned} \tag{304}$$

Using what we have learned in the previous section, we conclude that

$$\left|\delta\phi_{\mathbf{k}}\right| \simeq \frac{H}{\sqrt{2k^3}}\left(\frac{k}{aH}\right)^{\frac{3}{2}-\nu} \quad \text{(ON superHubble SCALES)}$$

which justifies our initial assumption that both the inflaton perturbation and the gravitational potential are nearly constant on superHubble scale.

We may now compute the power spectrum of the comoving curvature perturbation on superHubble scales

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{1}{2M_{\rm Pl}^2\epsilon}\left(\frac{H}{2\pi}\right)^2\left(\frac{k}{aH}\right)^{n_{\mathcal{R}}-1} \equiv A_{\mathcal{R}}^2\left(\frac{k}{aH}\right)^{n_{\mathcal{R}}-1}$$

where we have defined the *spectral index* $n_{\mathcal{R}}$ of the comoving curvature perturbation as

$$n_{\mathcal{R}} - 1 = \frac{{\rm dln}\,\mathcal{P}_{\mathcal{R}}}{{\rm dln}\,k} = 3 - 2\nu = 2\eta - 6\epsilon.$$

We conclude that inflation is responsible for the generation of adiabatic/curvature perturbations with an almost scale-independent spectrum.

From the curvature perturbation we can easily deduce the behaviour of the gravitational potential $\Psi_{\mathbf{k}}$ from Eq. (296). The latter is solved by

$$\Psi_{\mathbf{k}} = \frac{A(k)}{a} + \frac{4\pi G_{\rm N}}{a}\int^t {\rm d}t'\,a(t')\,\dot{\phi}(t')\,\delta\phi_{\mathbf{k}}(t') \simeq \frac{A(k)}{a} + \epsilon\,\mathcal{R}_{\mathbf{k}}.$$

We find that during inflation and on superHubble scales the gravitational potential is the sum of a decreasing function plus a nearly constant in time piece proportional to the curvature perturbation. Notice in particular that in an exact de Sitter stage, that is $\epsilon = 0$, the gravitational potential is not sourced and any initial condition in the gravitational potential is washed out as a^{-1} during the inflationary stage.

Comment: We might have computed the spectral index of the spectrum $\mathcal{P}_{\mathcal{R}}(k)$ by first solving the equation for the perturbation $\delta\phi_{\mathbf{k}}$ in a di Sitter stage, with $H =$ constant ($\epsilon = \eta = 0$), whose solution is Eq. (211) and then taking into account the time-evolution of the Hubble rate and of ϕ introducing the subscript in H_k and $\dot{\phi}_k$. The time variation of the latter is determined by

$$\frac{{\rm dln}\,\dot{\phi}_{\mathbf{k}}}{{\rm dln}\,k} = \left(\frac{{\rm dln}\,\dot{\phi}_{\mathbf{k}}}{{\rm d}t}\right)\left(\frac{{\rm d}t}{{\rm dln}\,a}\right)\left(\frac{{\rm dln}\,a}{{\rm dln}\,k}\right) = \frac{\ddot{\phi}_{\mathbf{k}}}{\dot{\phi}_{\mathbf{k}}} \times \frac{1}{H} \times 1 = -\delta = \epsilon - \eta. \tag{305}$$

Correspondingly, $\dot{\phi}_{\mathbf{k}}$ is the value of the time derivative of the inflaton field when a given wavelength $\sim k^{-1}$ crosses the horizon (from that point on the fluctuations remains frozen in). The curvature perturbation in such an approach would read

$$\mathcal{R}_{\mathbf{k}} \simeq \frac{H_k}{\dot{\phi}_k}\,\delta\phi_{\mathbf{k}} \simeq \frac{1}{2\pi}\left(\frac{H_k^2}{\dot{\phi}_{\mathbf{k}}}\right).$$

Correspondigly

$$n_{\mathcal{R}} - 1 = \frac{{\rm dln}\,\mathcal{P}_{\mathcal{R}}}{{\rm dln}\,k} = \frac{{\rm dln}\,H_k^4}{{\rm dln}\,k} - \frac{{\rm dln}\,\dot{\phi}_{\mathbf{k}}^2}{{\rm dln}\,k} = -4\epsilon + (2\eta - 2\epsilon) = 2\eta - 6\epsilon$$

which reproduces our previous findings.

During inflation the curvature perturbation is generated on superHubble scales with a spectrum which is nearly scale invariant, that is is nearly independent from the wavelength $\lambda = \pi/k$: the amplitude of the fluctuation on superHubble scales does not (almost) depend upon the time at which the fluctuations crosses the horizon and becomes frozen in. The small tilt of the power spectrum arises from the fact that the inflaton field is massive, giving rise to a nonvanishing η and because during inflation the Hubble rate is not exactly constant, but nearly constant, where 'nearly' is quantified by the slow-roll parameters ϵ.

Comment: From what found so far, we may conclude that on superHubble scales the comoving curvature perturbation $\mathcal{R}$ and the uniform-density gauge curvature ζ satisfy on superHubble scales the relation

$$\dot{\mathcal{R}}_{\mathbf{k}} \simeq \dot{\zeta}_{\mathbf{k}} \simeq 0.$$

An independent argument of the fact that they are nearly constant on superHubble scales is given in the next subsection.

8.13 A proof of time-independence of the comoving curvature perturbation for adiabatic modes: linear level

We give here a general argument following from energy-momentum conservation to show that the curvature perturbation on constant-time hypersurfaces Ψ is constant on superHubble scales if perturbations are adiabatic. Let us consider a generic fluid with energy-momentum tensor $T^{\mu\nu} = (\rho + P)u^\mu u^\nu + g^{\mu\nu}P$. The four-velocity u^μ is subject to the constraint $u^\mu u_\nu = -1$. Since it can be decomposed as

$$u^\mu = \frac{1}{a}\left(\delta^\mu_0 + v^\mu\right), \tag{306}$$

we get

$$v^0 = -\Psi. \tag{307}$$

Similarly, we obtain

$$\begin{aligned} u_0 &= a(-1 - \Phi), \\ U_i &= a v_i. \end{aligned} \tag{308}$$

Notice that, since we will work on superHubble scales we have only taken the gravitational potentials in the metric. The associated perturbation of the energy-momentum tensor is

$$
\begin{aligned}
\delta T^0_0 &= -(\delta\rho + \delta P) + (\bar\rho + \bar P)(1-\Psi)(-1-\Phi) + \delta P \simeq -\delta\rho, \\
\delta T^i_0 &\simeq 0, \\
\delta T^i_j &= \delta P \delta^i_j,
\end{aligned}
\tag{309}
$$

The associated continuity equation

$$
\boldsymbol{\nabla}_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu{}_{\mu\lambda} T^\lambda_\nu - \Gamma^\lambda{}_{\mu\nu} T^\mu_\lambda \tag{310}
$$

gives

$$
\begin{aligned}
&\partial_0 T^0_0 + \partial_i T^i_0 + \Gamma^\mu{}_{\mu\lambda} T^\lambda_0 - \Gamma^\lambda{}_{\mu 0} T^\mu_\lambda \\
&= \partial_0 T^0_0 + \Gamma^\mu{}_{\mu\lambda} T^\lambda_0 - \Gamma^\lambda{}_{\mu 0} T^\mu_\lambda \\
&= \partial_0 T^0_0 + \Gamma^\mu{}_{\mu 0} T^0_0 - \Gamma^\lambda{}_{00} T^0_\lambda - \partial_0 T^0_0 - \Gamma^\lambda{}_{i0} T^i_\lambda \\
&= \partial_0 T^0_0 + \Gamma^0{}_{00} T^0_0 + \Gamma^i{}_{i0} T^0_0 - \Gamma^0{}_{00} T^0_0 - \Gamma^j{}_{i0} T^i_j.
\end{aligned}
\tag{311}
$$

This expression, using the Christoffel symbols (246) gives

$$
\delta\dot\rho = -3H\left(\delta\rho + \delta P\right) - 3\dot\Psi\left(\bar\rho + \bar P\right).
$$

We write $\delta P = \delta P_{\rm nad} + c_s^2 \delta\rho$, where $\delta P_{\rm nad}$ is the non-adiabatic component of the pressure perturbation and $c_s^2 = \delta P_{\rm ad}/\delta\rho$ is the adiabatic one. In the uniform-density gauge $\Psi = \zeta$ and $\delta\rho = 0$ and therefore $\delta p_{\rm ad} = 0$. The energy conservation equation implies

$$
\boxed{\dot\zeta = -\frac{H}{\bar P + \bar\rho}\,\delta P_{\rm nad}.}
$$

If perturbations are adiabatic, the curvature on uniform-density gauge is constant on superHubble scales. The same holds for the comoving curvature $\mathcal{R}$ as the latter and ζ are equal on superHubble scales.

8.14 A proof of time-independence of the comoving curvature perturbation for adiabatic modes: linear level

We give here a general argument following from energy-momentum conservation to show that the curvature perturbation on constant-time hypersurfaces Ψ is constant on superHubble scales if perturbations are adiabatic. Let us consider a generic fluid with energy-momentum tensor $T^{\mu\nu} = (\rho + P)u^\mu u^\nu + g^{\mu\nu}P$. The four-velocity u^μ is subject to the constraint $u^\mu u_\nu = -1$. Since it can be decomposed as

$$u^\mu = \frac{1}{a}\left(\delta^\mu_0 + v^\mu\right), \tag{312}$$

we get

$$v^0 = -\Psi. \tag{313}$$

Similarly, we obtain

$$\begin{aligned} u_0 &= a(-1-\Phi), \\ U_i &= a v_i. \end{aligned} \tag{314}$$

Notice that, since we will work on superHubble scales we have only taken the gravitational potentials in the metric. The associated perturbation of the energy-momentum tensor is

$$\begin{aligned} \delta T^0_0 &= -(\delta\rho + \delta P) + (\bar\rho + \bar P)(1-\Psi)(-1-\Phi) + \delta P \simeq -\delta\rho, \\ \delta T^i_0 &\simeq 0, \\ \delta T^i_j &= \delta P \delta^i_j, \end{aligned} \tag{315}$$

The associated continuity equation

$$\boldsymbol{\nabla}_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\ \mu\lambda} T^\lambda_\nu - \Gamma^\lambda_{\ \mu\nu} T^\mu_\lambda \tag{316}$$

gives

$$\begin{aligned} &\partial_0 T^0_0 + \partial_i T^i_0 + \Gamma^\mu_{\ \mu\lambda} T^\lambda_0 - \Gamma^\lambda_{\ \mu 0} T^\mu_\lambda \\ &= \partial_0 T^0_0 + \Gamma^\mu_{\ \mu\lambda} T^\lambda_0 - \Gamma^\lambda_{\ \mu 0} T^\mu_\lambda \\ &= \partial_0 T^0_0 + \Gamma^\mu_{\ \mu 0} T^0_0 - \Gamma^\lambda_{\ 00} T^0_\lambda - \partial_0 T^0_0 - \Gamma^\lambda_{\ i0} T^i_\lambda \\ &= \partial_0 T^0_0 + \Gamma^0_{\ 00} T^0_0 + \Gamma^i_{\ i0} T^0_0 - \Gamma^0_{\ 00} T^0_0 - \Gamma^j_{\ i0} T^i_j. \end{aligned} \tag{317}$$

This expression, using the Christoffel symbols (246) gives

$$\delta\dot\rho = -3H\left(\delta\rho + \delta P\right) - 3\dot\Psi\left(\bar\rho + \bar P\right).$$

We write $\delta P = \delta P_{\rm nad} + c_s^2 \delta\rho$, where $\delta P_{\rm nad}$ is the non-adiabatic component of the pressure perturbation and $c_s^2 = \delta P_{\rm ad}/\delta\rho$ is the adiabatic one. In the uniform-density gauge $\Psi = \zeta$ and $\delta\rho = 0$ and therefore $\delta p_{\rm ad} = 0$. The energy conservation equation implies

$$\boxed{\dot\zeta = -\frac{H}{\bar P + \bar\rho}\,\delta P_{\rm nad}.}$$

If perturbations are adiabatic, the curvature on uniform-density gauge is constant on superHubble scales. The same holds for the comoving curvature $\mathcal{R}$ as the latter and ζ are equal on superHubble scales.

8.15 A proof of time-independence of the comoving curvature perturbation for adiabatic modes: all orders

We prove now that the comoving curvature perturbation is conserved at all orders in perturbation theory for adiabatic models on scales larger than the horizon. To do so, at momenta $k \ll Ha$ the universe looks like a collection of separate almost homogeneous universes. We choose a threading of spatial coordinates comoving with the fluid

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}t},\; v^i = \frac{u^i}{u^0} = \frac{\mathrm{d}x^i}{\mathrm{d}t} = 0. \tag{318}$$

The rate of the expansion is

$$\Theta = \boldsymbol{\nabla}_\mu u^\mu = \frac{1}{\mathcal{N}}\partial_0\, e^{3\alpha}, \tag{319}$$

where $g_{00} = \mathcal{N}^2$, $g_{ij} = e^{2\alpha}\gamma_{ij}$, with $\det\gamma_{ij} = 1$. The energy conservation equation

$$u_\nu \boldsymbol{\nabla}_\mu T^{\mu\nu} = 0 \Rightarrow \frac{\mathrm{d}}{\mathrm{d}\tau}\rho + (\rho + P)\Theta = 0, \tag{320}$$

where $\mathrm{d}t/\mathrm{d}\tau = u^0 = 1/\mathcal{N}$. Therefore, we obtain

$$\dot\rho + 3(\rho + P)\dot\alpha = 0. \tag{321}$$

Upon defining

$$a(t)e^{-\Psi} = e^\alpha, \tag{322}$$

we obtain

$$3\left(\frac{\dot a}{a} - \dot\Psi\right) = 3\dot\alpha = -\frac{\dot\rho}{\rho + P}. \tag{323}$$

This implies that the number of e-folds of expansion along an integral curve of the four-velocity comoving with the fluid is

$$N(t_2, t_1, x^i) = \frac{1}{3}\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \Theta = \frac{1}{3}\int_{t_1}^{t_2} \mathrm{d}t\, \mathcal{N}\,\Theta = -\frac{1}{3}\int_{t_1}^{t_2} \mathrm{d}t\; \left.\frac{\dot\rho}{\rho + P}\right|_{x^i}. \tag{324}$$

This implies that

$$\Psi(t_2, x^i) - \Psi(t_1, x^i) = -N(t_2, t_1, x^i) + \ln\frac{a(t_2)}{a(t_1)}, \tag{325}$$

that is the change in Ψ from one slice to another equals the difference of the actual number of e-folds and the background. In particular, in a flat slice

$$N(t_2, t_1, x^i) = \ln\frac{a(t_2)}{a(t_1)}, \tag{326}$$

From (325) we find therefore

$$-\Psi(t_2,x^i)+\Psi(t_1,x^i)=-\frac{1}{3}\int_{\rho(t_1,x^i)}^{\rho(t_2,x^i)}\frac{{\rm d}\rho}{\rho+P}-\ln\frac{a(t_2)}{a(t_1)}. \tag{327}$$

If the perturbation are adiabatic, that is if $P=P(\rho)$, then we conclude that

$$\boxed{\zeta(x^i)=\Psi(t,x^i)-\frac{1}{3}\int_{\rho(t)}^{\rho(t,x^i)}\frac{{\rm d}\rho}{\rho+P}} \tag{328}$$

is constant and this holds at any order in perturbation theory. This is the non-linear generalization of the comoving curvature perturbation.

Consider now two different slices A and B which coincide at $t=t_1$. From (325) we have that

$$-N_A(t_2,t_1,x^i)+N_B(t_2,t_1,x^i)=\Psi_A(t_2,x^i)-\Psi_B(t_2,x^i). \tag{329}$$

Now, choose the slice A such that it is flat at $t=t_1$ and ends on a uniform energy slice at $t=t_2$ and B to be flat both at t_1 and t_2

$$\boxed{-\Psi_A(t_2,x^i)=N_A(t_2,t_1,x^i)-N_0(t_2,t_1)\equiv\delta N}, \tag{330}$$

since B is flat. This means that $-\Psi_A(t_2,x^i)$ is the difference in the number of e-folds (from $t=t_1$ to $t=t_2$) between the uniform-density slicing and the flat slicing. Therefore, by choosing the initial slice at the t_1 to be the flat slice and the slice at generic time t to have uniform energy density, the curvature perturbation on that slice is the difference in the number of e-folds between the uniform energy density slice and the flat slice from t_1 to t

$$\boxed{-\zeta=\delta N=\delta N(\phi({\bf x},t))\Rightarrow\zeta=\frac{\partial N}{\partial\phi}\delta\phi=\frac{\partial N}{\partial t}\frac{\delta\phi}{\dot\phi}=H\frac{\delta\phi}{\dot\phi}.} \tag{331}$$

This is indeed the easiest way of computing the comoving curvature perturbation and is dubbed the δN formalism. In general

$$\boxed{\zeta(x^i)=-\delta N-\frac{1}{3}\int_{\rho(t)}^{\rho(t,x^i)}\frac{{\rm d}\rho}{\rho+P}} \tag{332}$$

where δN must be interpreted ad the amount of expansion along the worldline of a comoving observer from a spatially flat $\Psi=0$ slice at time t_1 to a generic slice at time t.

9 Comoving curvature perturbation from isocurvature perturbation

Let us give one example of how the fact that the comoving curvature perturbation is not constant when there are isocurvature perturbation can be useful. The paradigm we will describe goes under the name of the curvaton mechanism.

Suppose that during inflation there is another field σ, the curvaton, which is supposed to give a negligible contribution to the energy density and to be an almost free scalar field, with a small effective mass $m_\sigma^2 = |\partial^2 V/\partial\sigma^2| \ll H^2$.

The unperturbed curvaton field satisfies the equation of motion

$$\sigma'' + 2\,\mathcal{H}\sigma' + a^2\frac{\partial V}{\partial\sigma} = 0\,. \tag{333}$$

It is also usually assumed that the curvaton field is very weakly coupled to the scalar fields driving inflation and that the curvature perturbation from the inflaton fluctuations is negligible. Thus, if we expand the curvaton field up to first-order in the perturbations around the homogeneous background as $\sigma(\tau,\mathbf{x}) = \sigma_0(\tau) + \delta\sigma$, the linear perturbations satisfy on large scales

$$\delta\sigma'' + 2\,\mathcal{H}\delta\sigma' + a^2\frac{\partial^2 V}{\partial\sigma^2}\,\delta\sigma = 0\,. \tag{334}$$

As a result on superHubble scales its fluctuations $\delta\sigma$ will be Gaussian distributed and with a nearly scale-invariant spectrum given by

$$\mathcal{P}^{\frac{1}{2}}_{\delta\sigma}(k) \approx \frac{H_*}{2\pi}\,, \tag{335}$$

where the subscript $*$ denotes the epoch of horizon exit $k = aH$. Once inflation is over the inflaton energy density will be converted to radiation (γ) and the curvaton field will remain approximately constant until $H^2 \sim m_\sigma^2$. At this epoch the curvaton field begins to oscillate around the minimum of its potential which can be safely approximated to be quadratic $V \approx \frac{1}{2}m_\sigma^2\sigma^2$. During this stage the energy density of the curvaton field just scales as non-relativistic matter $\rho_\sigma \propto a^{-3}$. The energy density in the oscillating field is

$$\rho_\sigma(\tau,\mathbf{x}) \approx m_\sigma^2\sigma^2(\tau,\mathbf{x})\,, \tag{336}$$

and it can be expanded into a homogeneous background $\rho_\sigma(\tau)$ and a first-order perturbation $\delta\rho_\sigma$ as

$$\rho_\sigma(\tau,\mathbf{x}) = \rho_\sigma(\tau) + \delta\rho_\sigma(\tau,\mathbf{x}) = m_\sigma^2\sigma + 2m_\sigma^2\,\sigma\,\delta\sigma\,. \tag{337}$$

As it follows from Eqs. (333) and (334) for a quadratic potential the ratio $\delta\sigma/\sigma$ remains constant and the resulting relative energy density perturbation is

$$\frac{\delta\rho_\sigma}{\rho_\sigma} = 2\left(\frac{\delta\sigma}{\sigma}\right)_*\,, \tag{338}$$

where the $*$ stands for the value at horizon crossing. Such perturbations in the energy density of the curvaton field produce in fact a primordial density perturbation well after the end of inflation. The primordial adiabatic density perturbation is associated with a perturbation in the spatial curvature Ψ and it is, as we have shown, characterized in a gauge-invariant manner by the curvature perturbation ζ on hypersurfaces of uniform total density ρ. We recall that at linear order the quantity ζ is given by the gauge-invariant formula

$$\zeta = \Psi + \mathcal{H}\frac{\delta\rho}{\rho'}\,, \tag{339}$$

and on large scales it obeys the equation of motion

$$\zeta' = -\frac{\mathcal{H}}{\rho + P}\,\delta P_{\rm nad}\,, \tag{340}$$

In the curvaton scenario the curvature perturbation is generated well after the end of inflation during the oscillations of the curvaton field because the pressure of the mixture of matter (curvaton) and radiation produced by the inflaton decay is not adiabatic. A convenient way to study this mechanism is to consider the curvature perturbations ζ_i associated with each individual energy density components, which to linear order are defined as

$$\zeta_i \equiv \Psi + \mathcal{H}\left(\frac{\delta\rho_i}{\rho_i'}\right)\,. \tag{341}$$

Therefore, during the oscillations of the curvaton field, the total curvature perturbation in Eq. (339) can be written as a weighted sum of the single curvature perturbations

$$\zeta = (1-f)\zeta_\gamma + f\zeta_\sigma\,, \tag{342}$$

where the quantity

$$f = \frac{3\rho_\sigma}{4\rho_\gamma + 3\rho_\sigma} \tag{343}$$

defines the relative contribution of the curvaton field to the total curvature perturbation. From now on we shall work under the approximation of sudden decay of the curvaton field. Under this approximation the curvaton and the radiation components ρ_σ and ρ_γ satisfy separately the energy conservation equations

$$\rho_\gamma' = -4\,\mathcal{H}\rho_\gamma\,,$$
$$\rho_\sigma' = -3\,\mathcal{H}\rho_\sigma\,, \tag{344}$$

and the curvature perturbations ζ_i remains constant on superHubble scales until the decay of the curvaton. Therefore from Eq. (342) it follows that the first-oder curvature pertubation evolves on large scales as

$$\zeta' = f'(\zeta_\sigma - \zeta_\gamma) = \mathcal{H}f(1-f)(\zeta_\sigma - \zeta_\gamma)\,, \tag{345}$$

and by comparison with Eq. (340) one obtains the expression for the non-adiabatic pressure perturbation at first order

$$\delta P_{\rm nad} = \rho_\sigma (1-f)(\zeta_\gamma - \zeta_\sigma)\,. \tag{346}$$

Since in the curvaton scenario it is supposed that the curvature perturbation in the radiation produced at the end of inflation is negligible

$$\zeta_\gamma = \Psi - \frac{1}{4}\frac{\delta\rho_\gamma}{\rho_\gamma} = 0\,. \tag{347}$$

Similarly the value of ζ_σ is fixed by the fluctuations of the curvaton during inflation

$$\zeta_\sigma = \Psi - \frac{1}{3}\frac{\delta\rho_\sigma}{\rho_\sigma} = \zeta_{\sigma \rm I}\,, \tag{348}$$

where I stands for the value of the fluctuations during inflation. From Eq. (342) the total curvature perturbation during the curvaton oscillations is given by

$$\zeta = f\zeta_\sigma\,. \tag{349}$$

As it is clear from Eq. (349) initially, when the curvaton energy density is subdominant, the density perturbation in the curvaton field ζ_σ gives a negligible contribution to the total curvature perturbation, thus corresponding to an isocurvature (or entropy) perturbation. On the other hand during the oscillations $\rho_\sigma \propto a^{-3}$ increases with respect to the energy density of radiation $\rho_\gamma \propto a^{-4}$, and the perturbations in the curvaton field are then converted into the curvature perturbation. Well after the decay of the curvaton, during the conventional radiation and matter dominated eras, the total curvature perturbation will remain constant on superHubble scales at a value which, in the sudden decay approximation, is fixed by Eq. (349) at the epoch of curvaton decay

$$\zeta = f_{\rm D}\,\zeta_\sigma\,, \tag{350}$$

where D stands for the epoch of the curvaton decay.

Going beyond the sudden decay approximation it is possible to introduce a transfer parameter r defined as

$$\zeta = r\zeta_\sigma, \tag{351}$$

where ζ is evaluated well after the epoch of the curvaton decay and ζ_σ is evaluated well before this epoch. Numerical studies of the coupled perturbation equations has been performed show that the sudden decay approximation is exact when the curvaton dominates the energy density before it decays ($r=1$), while in the opposite case

$$r \approx \left(\frac{\rho_\sigma}{\rho}\right)_{\rm D}\,. \tag{352}$$

9.1 Gauge-invariant computation of the curvature perturbation

In this subsection we would like to show how the computation of the curvature perturbation can be performed in a gauge-invariant way. We first rewrite Einstein's equations in terms of Bardeen's potentials (293) and (294)

$$\delta G^0_0 = \frac{2}{a^2}\left(-3\,\mathcal{H}\left(\mathcal{H}\,\Phi_{\rm GI} + \Psi'_{\rm GI}\right) + \boldsymbol{\nabla}^2\Psi_{\rm GI} + 3\,\mathcal{H}\left(-\mathcal{H}' + \mathcal{H}^2\right)\left(\frac{E'}{2} - B\right)\right), \quad (353)$$

$$\delta G^0_i = \frac{2}{a^2}\,\partial_i\left(\mathcal{H}\,\Phi_{\rm GI} + \Psi'_{\rm GI} + \left(\mathcal{H}' - \mathcal{H}^2\right)\left(\frac{E'}{2} - B\right)\right), \quad (354)$$

$$\delta G^i_j = -\frac{2}{a^2}\left(\left(\left(2\,\mathcal{H}' + 2\,\mathcal{H}^2\right)\Phi_{\rm GI} + \mathcal{H}\,\Phi'_{\rm GI} + \Psi''_{\rm GI} + 2\,\mathcal{H}\,\Psi'_{\rm GI} + \frac{1}{2}\,\boldsymbol{\nabla}^2 D_{\rm GI}\right)\delta^i_j \right.$$
$$\left. + \left(\mathcal{H}'' - \mathcal{H}\,\mathcal{H}' - \mathcal{H}^3\right)\left(\frac{E'}{2} - B\right)\delta^i_j - \frac{1}{2}\,\partial^i\partial_j D_{\rm GI}\right), \quad (355)$$

with $D_{\rm GI} = \Phi_{\rm GI} - \Psi_{\rm GI}$. These quantities are not gauge-invariant, but using the gauge transformations described previously, we can easily generalize them to gauge-invariant quantities

$$\delta G^{({\rm GI})0}_0 = \delta G^0_0 + (G^0_0)'\left(\frac{E'}{2} - B\right), \quad (356)$$

$$\delta G^{({\rm GI})0}_i = \delta G^0_i + \left(G^0_i - \frac{1}{3}\,T^k_k\right)\partial_i\left(\frac{E'}{2} - B\right), \quad (357)$$

$$\delta G^{({\rm GI})i}_j = \delta G^i_j + (G^i_j)'\left(\frac{E'}{2} - B\right) \quad (358)$$

and

$$\delta T^{({\rm GI})0}_0 = \delta T^0_0 + (T^0_0)'\left(\frac{E'}{2} - B\right) = -\delta\rho^{({\rm GI})}, \quad (359)$$

$$\delta T^{({\rm GI})0}_i = \delta T^0_i + \left(T^0_i - \frac{1}{3}\,T^k_k\right)\partial_i\left(\frac{E'}{2} - B\right) = \left(\bar{\rho} + \bar{P}\right)a^{-1}v^{({\rm GI})}_i, \quad (360)$$

$$\delta T^{({\rm GI})i}_j = \delta T^i_j + (T^i_j)'\left(\frac{E'}{2} - B\right) = \delta P^{({\rm GI})} \quad (361)$$

where we have written the stress energy-momentum tensor as $T^{\mu\nu} = (\rho + P)\,u^\mu u^\nu + P\eta^{\mu\nu}$ with $u^\mu = (1, v^i)$. Barred quantities are to be intended as background quantities. Einstein's equations can now be written in a gauge-invariant way

$$
\begin{aligned}
& - \quad 3\,\mathcal{H}\left(\mathcal{H}\,\Phi_{\rm GI} \,+\, \Psi'_{\rm GI}\right) \,+\, \boldsymbol{\nabla}^2\,\Psi_{\rm GI} \qquad (362)\\
& = \quad 4\,\pi\,G_{\rm N}\left(-\Phi_{\rm GI}\,\phi'^2 \,+\, \delta\phi^{\rm (GI)}\,\phi' \,+\, \delta\phi^{\rm (GI)}\,\frac{\partial V}{\partial \phi}\,a^2\right)\,,\\
& \qquad \partial_i\,\left(\mathcal{H}\,\Phi \,+\, \Psi'_{\rm GI}\right) \,=\, 4\,\pi\,G_{\rm N}\left(\partial_i\,\delta\phi^{\rm (GI)}\,\phi'\right)\,,\\
& \qquad \left(\left(2\,\mathcal{H}' \,+\, \mathcal{H}^2\right)\Phi_{\rm GI} \,+\, \mathcal{H}\,\Phi'_{\rm GI} \,+\, \Psi''_{\rm GI} \,+\, 2\,\mathcal{H}\,\Psi'_{\rm GI} \,+\, \frac{1}{2}\,\boldsymbol{\nabla}^2 D_{\rm GI}\right)\delta^i_j \,-\, \frac{1}{2}\partial^i\partial_j D_{\rm GI},\\
& = \quad -\,4\,\pi\,G_{\rm N}\left(\Phi_{\rm GI}\,\phi'^2 \,-\, \delta\phi^{\rm (GI)}\,\phi' \,+\, \delta\phi^{\rm (GI)}\,\frac{\partial V}{\partial \phi}\,a^2\right)\delta^i_j\,. \qquad (363)
\end{aligned}
$$

Taking $i \neq j$ from the third equation, we find $D_{\rm GI} = 0$, that is $\Psi_{\rm GI} = \Phi_{\rm GI}$ and from now on we can work with only the variable $\Phi_{\rm GI}$. Using the background relation

$$
2\Big(\frac{a'}{a}\Big)^2 \,-\, \frac{a''}{a} \,=\, 4\,\pi\,G_{\rm N}\,\phi'^2 \qquad (364)
$$

we can rewrite the system of Eqs. (363) in the form

$$
\begin{aligned}
\boldsymbol{\nabla}^2\,\Phi_{\rm GI} \,-\, 3\,\mathcal{H}\,\Phi'_{\rm GI} \,-\, \left(\mathcal{H}' \,+\, 2\,\mathcal{H}^2\right)\Phi_{\rm GI} \quad &= \quad 4\,\pi\,G_{\rm N}\left(\delta\phi^{\rm (GI)}\,\phi' \,+\, \delta\phi^{\rm (GI)}\,\frac{\partial V}{\partial \phi}\,a^2\right)\,;\\
\Phi'_{\rm GI} \,+\, \mathcal{H}\,\Phi_{\rm GI} \quad &= \quad 4\,\pi\,G_{\rm N}\left(\delta\phi^{\rm (GI)}\,\phi'\right)\,;\\
\Phi''_{\rm GI} \,+\, 3\,\mathcal{H}\,\Phi'_{\rm GI} \,+\, \left(\mathcal{H}' \,+\, 2\,\mathcal{H}^2\right)\Phi_{\rm GI} \quad &= \quad 4\,\pi\,G_{\rm N}\left(\delta\phi^{\rm (GI)}\,\phi' \,-\, \delta\phi^{\rm (GI)}\,\frac{\partial V}{\partial \phi}\,a^2\right)\,. \qquad (365)
\end{aligned}
$$

Substracting the first equation from the third, using the second equation to express $\delta\phi^{\rm (GI)}$ as a function of $\Phi_{\rm GI}$ and $\Phi'_{\rm GI}$ and using the Klein-Gordon equation one finally finds the

$$
\Phi''_{\rm GI} \,+\, 2\left(\mathcal{H} \,-\, \frac{\phi''}{\phi'}\right)\Phi'_{\rm GI} \,-\, \boldsymbol{\nabla}^2\,\Phi_{\rm GI} \,+\, 2\left(\mathcal{H}' \,-\, \mathcal{H}\,\frac{\phi''}{\phi'}\right)\Phi_{\rm GI} \,=\, 0\,, \qquad (366)
$$

for the gauge-invariant potential $\Phi_{\rm GI}$. We now introduce the gauge-invariant quantity

$$
\begin{aligned}
u \quad &\equiv \quad a\,\delta\phi^{\rm (GI)} \,+\, z\,\Psi_{\rm GI}\,, \qquad (367)\\
z \quad &\equiv \quad a\,\frac{\phi'}{\mathcal{H}} = a\,\frac{\dot\phi}{H}. \qquad (368)
\end{aligned}
$$

Notice that the variable u is equal to $-a\,Q$, the gauge-invariant inflaton perturbation on spatially flat gauges. Eq. (366) becomes

$$
u'' \,-\, \boldsymbol{\nabla}^2\,u \,-\, \frac{z''}{z}\,u \,=\, 0\,, \qquad (369)
$$

while the two remaining equations of the system (365) can be written as

$$\nabla^2 \Phi_{\rm GI} = 4\,\pi\,G_{\rm N}\,\frac{\mathcal{H}}{a^2}\left(z\,u' - z' u\right)\,, \tag{370}$$

$$\left(\frac{a^2\,\Phi_{\rm GI}}{\mathcal{H}}\right)' = 4\,\pi\,G_{\rm N}\,z\,u\,, \tag{371}$$

which allow to determine the variables Φ and $\delta\phi^{(\rm GI)}$.

We have now to solve Eq. (369). First, we have to evaluate z''/z in terms of the slow-roll parameters

$$\frac{z'}{\mathcal{H}z} = \frac{a'}{\mathcal{H}a} + \frac{\phi''}{\mathcal{H}\phi'} - \frac{\mathcal{H}'}{\mathcal{H}^2} = \epsilon + \frac{\phi''}{\mathcal{H}\phi'}.$$

We then deduce that

$$\delta \equiv 1 - \frac{\phi''}{\mathcal{H}\phi'} = 1 + \epsilon - \frac{z'}{\mathcal{H}z}.$$

Keeping the slow-roll parameters constat in time (as we have mentioned, this corresponds to expand all quantities to first-order in the slow-roll parameters), we find

$$0 \simeq \delta' = \epsilon'(\simeq 0) - \frac{z''}{\mathcal{H}z} + \frac{z'\,\mathcal{H}'}{z\,\mathcal{H}^2} + \frac{(z')^2}{\mathcal{H}z^2},$$

from which we deduce

$$\frac{z''}{z} \simeq \frac{z'\,\mathcal{H}'}{z\,\mathcal{H}} + \frac{(z')^2}{z^2}.$$

Expanding in slow-roll parameters we find

$$\frac{z''}{z} \simeq (1+\epsilon-\delta)\,(1-\epsilon)\;\mathcal{H}^2 + (1+\epsilon-\delta)^2\;\mathcal{H}^2 \simeq\; \mathcal{H}^2\,(2+2\epsilon-3\delta)\,.$$

If we set

$$\frac{z''}{z} = \frac{1}{\tau^2}\left(\nu^2 - \frac{1}{4}\right),$$

this corresponds to

$$\nu \simeq \frac{1}{2}\left[1 + 4\frac{(1+\epsilon-\delta)\,(2-\delta)}{(1-\epsilon)^2}\right]^{1/2} \simeq \frac{3}{2} + (2\epsilon-\delta) \simeq \frac{3}{2} + 3\epsilon - \eta.$$

On subHubble scales ($k \gg aH$), the solution of equation (369) is obviously $u_{\mathbf{k}} \simeq e^{-ik\tau}/\sqrt{2k}$. Rewriting Eq. (371) as

$$\Phi^{\rm GI}_{\mathbf{k}} = -\frac{4\pi G a^2}{k^2}\frac{\dot\phi^2}{H}\left(\frac{H}{a\dot\phi}u_{\mathbf{k}}\right)^{\cdot},$$

we infer that on subHubble scales

$$\Phi_{\mathbf{k}}^{\rm GI} \simeq i\,\frac{4\pi G\dot{\phi}}{\sqrt{2k^3}}\,e^{-i\frac{k}{a}}.$$

On superHubble scales ($k \ll aH$), one obvious solution to Eq. (369) is $u_{\mathbf{k}} \propto z$. To find the other solution, we may set $u_{\mathbf{k}} = z\,\widetilde{u}_{\mathbf{k}}$, which satisfies the equation

$$\frac{\widetilde{u}''_{\mathbf{k}}}{\widetilde{u}'_{\mathbf{k}}} = -2\frac{z'}{z},$$

which gives

$$\widetilde{u}_{\mathbf{k}} = \int^{\tau} \frac{{\rm d}\tau'}{z^2(\tau')}.$$

On superHubble scales therefore we find

$$u_{\mathbf{k}} = c_1(k)\frac{a\dot{\phi}}{H} + c_2(k)\frac{a\dot{\phi}}{H}\int^t {\rm d}t'\,\frac{H^2}{a^3\dot{\phi}^2}. \simeq c_1(k)\frac{a\dot{\phi}}{H} - c_2(k)\frac{1}{3a^2\dot{\phi}},$$

where the last passage has been performed supposing a de Sitter epoch, H = constant. The first piece is the constant mode $c_1(k)z$, while the second is the decreasing mode. To find the constant $c_1(k)$, we apply what we have learned previously. We know that on superHubble scales the exact solution of equation (369) is

$$u_{\mathbf{k}} = \frac{\sqrt{\pi}}{2}\,e^{i\left(\nu+\frac{1}{2}\right)\frac{\pi}{2}}\,\sqrt{-\tau}\,H_\nu(-k\tau). \tag{372}$$

On superHubble scales, since $H_\nu(x \ll 1) \sim \sqrt{2/\pi}\,e^{-i\frac{\pi}{2}}\,2^{\nu-\frac{3}{2}}\,(\Gamma(\nu_\chi)/\Gamma(3/2))\,x^{-\nu}$, the fluctuation (372) becomes

$$u_{\mathbf{k}} = e^{i\left(\nu-\frac{1}{2}\right)\frac{\pi}{2}}2^{\left(\nu-\frac{3}{2}\right)}\frac{\Gamma(\nu)}{\Gamma(3/2)}\frac{1}{\sqrt{2k}}\,(-k\tau)^{\frac{1}{2}-\nu}.$$

Therefore

$$c_1(k) = {\rm lim}_{k\to 0}\left|\frac{u_{\mathbf{k}}}{z}\right| = \frac{H}{a\dot{\phi}}\frac{1}{\sqrt{2k}}\left(\frac{k}{aH}\right)^{\frac{1}{2}-\nu} = \frac{H}{\dot{\phi}}\frac{1}{\sqrt{2k^3}}\left(\frac{k}{aH}\right)^{\eta-3\epsilon} \tag{373}$$

The last steps consist in relating the variable u to the comoving curvature $\mathcal{R}$ and to the gravitational potential $\Phi_{\rm GI}$. The comoving curvature takes the form

$$\mathcal{R} \equiv -\Psi_{\rm GI} - \frac{H}{\phi'}\,\delta\phi^{\rm (GI)} = -\frac{u}{z}. \tag{374}$$

Since $z = a\dot{\phi}/H = a\sqrt{2\epsilon}M_{\rm Pl}$, the power spectrum of the comoving curvature can be expressed on superHubble scales as

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\,\pi^2}\left|\frac{u_{\mathbf{k}}}{z}\right|^2 = \frac{1}{2M_{\rm Pl}^2\epsilon}\left(\frac{H}{2\pi}\right)^2\left(\frac{k}{aH}\right)^{n_{\mathcal{R}}-1} \equiv A_{\mathcal{R}}^2\left(\frac{k}{aH}\right)^{n_{\mathcal{R}}-1} \tag{375}$$

with

$$n_{\mathcal{R}} - 1 = 3 - 2\nu = 2\eta - 6\epsilon. \tag{376}$$

These results reproduce those found in the previous subsection. The last step is to find the behaviour of the gauge-invariant potential $\Phi_{\rm GI}$ on superHubble scales. If we recast equation (371) in the form

$$u_{\mathbf{k}} = \frac{1}{4\pi G_{\rm N}} \frac{H}{\dot{\phi}} \left(\frac{a}{H} \Phi^{\rm GI}_{\mathbf{k}} \right)^{\cdot}, \tag{377}$$

we can infer that on superHubble scales the nearly constant mode of the gravitational potential during inflation reads

$$\Phi^{\rm GI}_{\mathbf{k}} = c_1(k) \left[1 - \frac{H}{a} \int^t {\rm d}t'\, a\left(t'\right) \right] \simeq -c_1(k) \frac{\dot{H}}{H^2} = \epsilon\, c_1(k) \simeq \epsilon \frac{u_{\mathbf{k}}}{z} \simeq -\epsilon\, \mathcal{R}_{\mathbf{k}}. \tag{378}$$

Indeed, plugging this solution into Eq. (377), one reproduces $u_{\mathbf{k}} = c_1(k) \frac{a\dot{\phi}}{H}$.

10 Transferring the perturbation to radiation during reheating

When the inflaton decays, the comoving curvature perturbation associated to the inflaton field are transferred to radiation. Let us see how this works.

Let us consider the system composed by the oscillating scalar field ϕ and the radiation fluid. Each component has energy-momentum tensor $T^{\mu\nu}_{(\phi)}$ and $T^{\mu\nu}_{(\gamma)}$. The total energy momentum $T^{\mu\nu} = T^{\mu\nu}_{(\phi)} + T^{\mu\nu}_{(\gamma)}$ is covariantly conserved, but allowing for an interaction between the two fluids

$$\begin{aligned} \boldsymbol{\nabla}_\mu T^{\mu\nu}_{(\phi)} &= Q^\nu_{(\phi)}\,, \\ \boldsymbol{\nabla}_\mu T^{\mu\nu}_{(\gamma)} &= Q^\nu_{(\gamma)}\,, \end{aligned} \tag{379}$$

where $Q^\nu_{(\phi)}$ and $Q^\nu_{(\gamma)}$ are the generic energy-momentum transfer to the scalar field and radiation sector respectively and are subject to the constraint

$$Q^\nu_{(\phi)} + Q^\nu_{(\gamma)} = 0\,. \tag{380}$$

The energy-momentum transfer $Q^\nu_{(\phi)}$ and $Q^\nu_{(\gamma)}$ can be decomposed for convenience as

$$\begin{aligned} Q^\nu_{(\phi)} &= \hat{Q}_\phi u^\nu + f^\nu_{(\phi)}\,, \\ Q^\nu_{(\gamma)} &= \hat{Q}_\gamma u^\nu + f^\nu_{(\gamma)}\,, \end{aligned} \tag{381}$$

where the f^{ν}'s are required to be orthogonal to the the total velocity of the fluid u^{ν}. The energy continuity equations for the scalar field and radiation can be obtained from $u_{\nu}\nabla_{\mu}T^{\mu\nu}_{(\phi)} = u_{\nu}Q^{\nu}_{(\phi)}$ and $u_{\nu}\nabla_{\mu}T^{\mu\nu}_{(\gamma)} = u_{\nu}Q^{\nu}_{(\gamma)}$ and hence from Eq. (381)

$$
\begin{aligned}
u_{\nu}\nabla_{\mu}T^{\mu\nu}_{(\phi)} &= \hat{Q}_{\phi}\,, \\
u_{\nu}\nabla_{\mu}T^{\mu\nu}_{(\gamma)} &= \hat{Q}_{\gamma}\,.
\end{aligned} \tag{382}
$$

In the case of an oscillating scalar field decaying into radiation the energy transfer coefficient $\hat{Q}_{\phi}$ is given by

$$
\begin{aligned}
\hat{Q}_{\phi} &= -\Gamma\rho_{\phi}, \\
\hat{Q}_{\gamma} &= \Gamma\rho_{\phi},
\end{aligned} \tag{383}
$$

where Γ is the decay rate of the scalar field into radiation.

The equations of motion for the curvature perturbations ζ_{ϕ} and ζ_{γ} can be obtained perturbing at first order the continuity energy equations (382) for the scalar field and radiation energy densities, including the energy transfer. Expanding the transfer coefficients $\hat{Q}_{\phi}$ and $\hat{Q}_{\gamma}$ up to first order in the perturbations around the homogeneous background as

$$\hat{Q}_{\phi} = Q_{\phi} + \delta Q_{\phi}\,, \tag{384}$$

$$\hat{Q}_{\gamma} = Q_{\gamma} + \delta Q_{\gamma}\,, \tag{385}$$

Eqs. (382) give on wavelengths larger than the horizon scale

$$
\begin{aligned}
&\delta\rho'_{\phi} + 3\,\mathcal{H}\,(\delta\rho_{\phi} + \delta P_{\phi}) - 3\,(\rho_{\phi} + P_{\phi})\,\Psi' \\
&= a\,Q_{\phi}\phi + a\,\delta Q_{\phi}\,,
\end{aligned} \tag{386}
$$

$$
\begin{aligned}
&\delta\rho'_{\gamma} + 3\,\mathcal{H}\,(\delta\rho_{\gamma} + \delta P_{\gamma}) - 3\,(\rho_{\gamma} + P_{\gamma})\,\Psi' \\
&= a\,Q_{\gamma}\phi + a\,\delta Q_{\gamma}\,.
\end{aligned} \tag{387}
$$

Notice that the oscillating scalar field and radiation have fixed equations of state with $\delta P_{\phi} = 0$ and $\delta P_{\gamma} = \delta\rho_{\gamma}/3$ (which correspond to vanishing intrinsic non-adiabatic pressure perturbations). Using the perturbed $(0-0)$-component of Einstein's equations for super-horizon wavelengths $\Psi' + \mathcal{H}\Phi = -\,\mathcal{H}(\delta\rho/\rho)/2$, we can rewrite Eqs. (386) and (387) in terms of the gauge-invariant curvature perturbations ζ_ϕ and ζ_γ

$$\begin{aligned}
\zeta_\phi^{'} &= \frac{a\,\mathcal{H}}{\rho_\phi'}\Bigg[\delta Q_\phi - \frac{Q_\phi'}{\rho_\phi'}\delta\rho_\phi \\
&+ Q_\phi\frac{\rho'}{2\rho}\left(\frac{\delta\rho_\phi}{\rho_\phi'} - \frac{\delta\rho}{\rho'}\right)\Bigg], \qquad (388)\\
\zeta_\gamma^{'} &= \frac{a\,\mathcal{H}}{\rho_\gamma'}\Bigg[\delta Q_\gamma - \frac{Q_\gamma'}{\rho_\gamma'}\delta\rho_\gamma \\
&+ Q_\gamma\frac{\rho'}{2\rho}\left(\frac{\delta\rho_\gamma}{\rho_\gamma'} - \frac{\delta\rho}{\rho'}\right)\Bigg], \qquad (389)
\end{aligned}$$

where $\delta Q_\gamma = -\delta Q_\phi$ from the constraint in Eq (380). If the energy transfer coefficients $\hat{Q}_\phi$ and $\hat{Q}_\gamma$ are given in terms of the decay rate Γ as in Eq. (383), the first order perturbation are respectively

$$\begin{aligned}
\delta Q_\phi &= -\Gamma\delta\rho_\phi, \qquad (390)\\
\delta Q_\gamma &= \Gamma\delta\rho_\phi. \qquad (391)
\end{aligned}$$

Plugging the expressions (390-391) into Eqs. (388-389), the first order curvature perturbations for the scalar field and radiation obey on large scales

$$\begin{aligned}
\zeta_\phi^{'} &= \frac{a\Gamma}{2}\frac{\rho_\phi}{\rho_\phi'}\frac{\rho'}{\rho}\left(\zeta - \zeta_\phi\right), \qquad (392)\\
\zeta_\gamma^{'} &= -\frac{a}{\rho_\gamma'}\left[\Gamma\rho'\frac{\rho_\phi'}{\rho_\gamma'}\left(1 - \frac{\rho_\phi}{2\rho}\right)\left(\zeta - \zeta_\phi\right)\right]. \qquad (393)
\end{aligned}$$

From the total comoving curvature perturbation

$$\zeta = \frac{\dot\rho_\phi}{\dot\rho}\zeta_\phi + \frac{\dot\rho_\gamma}{\dot\rho}\zeta_\gamma,\ \ \rho = \rho_\phi + \rho_\gamma. \qquad (394)$$

it is thus possible to find the equation of motion for the total curvature perturbation ζ using the evolution of the individual curvature perturbations in Eqs. (392) and (393)

$$\begin{aligned}
\zeta^{'} &= f'\left(\zeta_\phi - \zeta_\gamma\right) + f\zeta_\phi^{'} + (1-f)\zeta_\gamma^{'} \\
&= \mathcal{H}f(1-f)\left(\zeta_\phi - \zeta_\gamma\right) = -\,\mathcal{H}f\left(\zeta - \zeta_\phi\right)\,, \qquad (395)
\end{aligned}$$

where $f = (\dot\rho_\phi/\dot\rho)$. Notice that during the decay of the scalar field into the radiation fluid, ρ_γ' in Eq. (393) may vanish. So it is convenient to close the system of equations by using the two equations (392) and (395) for the evolution of ζ_ϕ and ζ. These equations say that $\zeta = \zeta_\phi$ is a fixed point: during the reheating phase the comoving curvature pertubation stored in the inflaton field is transferred to radiation smoothly.

11 The initial conditions provided by inflation

Inflation provides the initial conditions for all perturbations once the latter re-enter the horizon. Let us turn again to the longitudinal gauge. On superHubble scales, from Eq. (271) we have

$$6\mathcal{H}^2\Phi_{\mathbf{k}} = -4\pi G_{\rm N} a^2 \frac{\delta\rho_{\mathbf{k}}}{\bar\rho} \Rightarrow \frac{\delta\rho_{\mathbf{k}}}{\bar\rho} = -2\Phi, \tag{396}$$

on superHubble scales, where $\mathcal{H}^2 = (8\pi G_{\rm N}/3)\bar\rho a^2$ defines the average energy density. Recalling now that $\Psi = \Phi$ and that $\zeta = \Psi + \mathcal{H}\delta\rho/\rho'$ and $\rho' = -3\,\mathcal{H}(\rho + P) = -3\,\mathcal{H}(1+w)\rho$, where we have defined $P/\rho = w$, we find that on superHubble scales

$$\zeta_{\mathbf{k}} = \Phi_{\mathbf{k}} - \frac{\delta\rho_{\mathbf{k}}}{3(1+w)\bar\rho} = \left(1 + \frac{2}{3(1+w)}\right)\Phi_{\mathbf{k}} = \frac{5+3w}{3(1+w)}\Phi_{\mathbf{k}}. \tag{397}$$

This means that during the RD phase one has ($w = 1/3$)

$$\boxed{\Phi_{\mathbf{k}}^{\rm RD} = \frac{2}{3}\zeta_{\mathbf{k}}\ ({\rm RD})}, \tag{398}$$

and during the MD phase

$$\boxed{\Phi_{\mathbf{k}}^{\rm MD} = \frac{3}{5}\zeta_{\mathbf{k}}\ ({\rm MD})}, \tag{399}$$

In particular, notice that

$$\boxed{\Phi_{\mathbf{k}}^{\rm MD} = \frac{9}{10}\Phi_{\mathbf{k}}^{\rm RD}}. \tag{400}$$

One of the last steps we wish to take is now fixing the amplitude of the density perturbation in the CMB through inflation. The CMB anisotropy has an oscillating structure (the famous Doppler peaks) because the baryon-phton fluid oscillates, with 1) a boost for those modes which enter the horizon at last scattering and 2) a damping due to photon diffusion. The overall amplitude of the CMB anisotropy can be fixed at large angular scales (superHubble modes) where there is no evolution and therefore one can match the amplitude with the theoretical prediction from inflation.

As on large scales and during matter-domination (recall that $\tau_{\rm ls} > \tau_{\rm eq}$) we have at last scattering

$$\frac{\delta\rho_{\rm m}}{\bar\rho_{\rm m}} = -2\Phi^{\rm MD}(\tau_{\rm ls}), \tag{401}$$

and, if the adiabatic condition holds,

$$\frac{1}{3}\frac{\delta\rho_{\rm m}}{\bar\rho_{\rm m}} = \frac{1}{4}\frac{\delta\rho_\gamma}{\bar\rho_\gamma} \equiv \Delta_0(\tau_{\rm ls}), \tag{402}$$

we obtain that the observed CMB anisotropy on large scales at the last scattering epoch should be the Sachs-Wolfe term (see later for an explanation of how this expression comes about)

$$\boxed{\left(\frac{\Delta}{4}+\Phi^{\rm MD}\right)_{\rm SW}(\tau_{\rm ls})=\left(-\frac{2}{3}+1\right)\Phi^{\rm MD}=\frac{1}{3}\Phi^{\rm MD}(\tau_{\rm ls}).} \tag{403}$$

We have seen previously that the temperature anisotropy is commonly expanded in spherical harmonics

$$\frac{\Delta T}{T}(\mathbf{x}_0,\tau_0,\mathbf{n})=\sum_{\ell m}a_{\ell,m}(\mathbf{x}_0)Y_{\ell m}(\mathbf{n}), \tag{404}$$

where $\mathbf{x}_0$ and τ_0 are our position and the preset time, respectively, $\mathbf{n}$ is the direction of observation, ℓ's are the different multipoles and

$$\langle a_{\ell m}a^*_{\ell' m'}\rangle=\delta_{\ell,\ell'}\delta_{m,m'}C_\ell, \tag{405}$$

where the deltas are due to the fact that the process that created the anisotropy is statistically isotropic. The C_ℓ are the so-called CMB power spectrum. For homogeneity and isotropy, the C_ℓ's are neither a function of $\mathbf{x}_0$, nor of m. The two-point-correlation function is related to the $C_{\rm L}$'s according to Eq. (156). WE get therefore that

$$a_{\ell m}(\mathbf{x}_0,\tau_0)=\int\frac{{\rm d}^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{x}_0}\int{\rm d}\Omega\,Y^*_{\ell m}(\mathbf{n})\Theta(\mathbf{k},\mathbf{n},\tau_0), \tag{406}$$

where we have made use the orthonormality property of the spherical harmonics

$$\int{\rm d}\Omega\,Y^*_{\ell m}(\mathbf{n})\,Y_{\ell' m'}(\mathbf{n})=\delta_{\ell\ell'}\delta_{mm'}. \tag{407}$$

The C_ℓ

$$\begin{aligned}C_\ell &= \int\frac{{\rm d}^3k}{(2\pi)^3}\int\frac{{\rm d}^3p}{(2\pi)^3}e^{i(\mathbf{k}+\mathbf{p})\cdot\mathbf{x}_0}\int{\rm d}\Omega\,Y^*_{\ell m}(\mathbf{n})\int{\rm d}\Omega'\,Y^*_{\ell m}(\mathbf{n}')\Big\langle\Theta(\mathbf{k},\mathbf{n},\tau_0)\Theta(\mathbf{p},\mathbf{n}',\tau_0)\Big\rangle\\ &= \sum_{\ell'\ell''}(-i)^{\ell'+\ell''}(2\ell'+1)(2\ell''+1)\int\frac{{\rm d}^3k}{(2\pi)^3}\int\frac{{\rm d}^3p}{(2\pi)^3}e^{i(\mathbf{k}+\mathbf{p})\cdot\mathbf{x}_0}\\ &\times \int{\rm d}\Omega\,Y^*_{\ell m}(\mathbf{n})P_{\ell'}(\mathbf{k}\cdot\mathbf{n})\int{\rm d}\Omega'\,Y^*_{\ell m}(\mathbf{n}')P_{\ell''}(\mathbf{p}\cdot\mathbf{n}')\Big\langle\Theta_{\ell'}(\mathbf{k})\Theta_{\ell''}(\mathbf{p})\Big\rangle.\end{aligned} \tag{408}$$

where we have decomposed the temperature anisotropy in multipoles as usual

$$\Theta(\mathbf{k},\mathbf{k},\tau_0)=\sum_\ell(-i)^\ell(2\ell+1)P_\ell(\mathbf{k}\cdot\mathbf{n})\Theta_\ell(k). \tag{409}$$

In the SW limit we have

$$\Theta^{\rm SW}_\ell(\mathbf{k})\simeq\frac{1}{3}\Phi^{\rm MD}(\mathbf{k},\tau_{\rm ls})j_\ell(k\tau_0), \tag{410}$$

with the spectrum of the gravitational potential defined as

$$\left\langle \Phi^{\rm MD}(\mathbf{k},\tau_{\rm ls})\Phi^{\rm MD}(\mathbf{p},\tau_{\rm ls})\right\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}+\mathbf{p})P_{\Phi^{\rm MD}}(k). \tag{411}$$

Therefore we obtain

$$\begin{aligned}
C_\ell^{\rm SW} &= \int \frac{{\rm d}^3k}{(2\pi)^3}P_{\Phi^{\rm MD}}(k)j_\ell^2(k\tau_0)\\
&\times \sum_{\ell'\ell''}(-i)^{\ell'+\ell''}(2\ell'+1)(2\ell''+1)\int {\rm d}\Omega\, Y_{\ell m}^*(\mathbf{n})P_{\ell'}(\mathbf{k}\cdot\mathbf{n})\int {\rm d}\Omega'\, Y_{\ell m}^*(\mathbf{n}')P_{\ell''}(\mathbf{p}\cdot\mathbf{n}')\\
&= \int \frac{{\rm d}^3k}{(2\pi)^3}P_{\Phi^{\rm MD}}(k)j_\ell^2(k\tau_0)\\
&\times \sum_{\ell'\ell''}(-i)^{\ell'+\ell''}(2\ell'+1)(2\ell''+1)\frac{4\pi}{(2\ell+1)}\delta_{\ell\ell'}Y_{\ell m}(\mathbf{n})\frac{4\pi}{(2\ell+1)}\delta_{\ell\ell''}Y_{\ell m}^*(\mathbf{n})\\
&= \frac{2}{\pi}\int {\rm d}k\, k^2 P_{\Phi^{\rm MD}}(k)j_\ell^2(k\tau_0)\int {\rm d}\Omega\, \left|Y_{\ell m}(\mathbf{n})\right|^2\\
&= \frac{2}{\pi}\int \; dk\, k^2 P_{\Phi^{\rm MD}}(k)j_\ell^2(k\tau_0).
\end{aligned} \tag{412}$$

If we generically indicate by

$$\left\langle \left|\Phi_{\mathbf{k}}^{\rm MD}\right|^2\right\rangle k^3 = A^2\,(k\tau_0)^{n-1}, \tag{413}$$

we can perform the integration and get

$$\frac{\ell(\ell+1)C_\ell^{\rm SW}}{2\pi} = \left[\frac{\sqrt{\pi}}{2}\ell(\ell+1)\frac{\Gamma(\frac{3-n}{2})\Gamma(\ell+\frac{n-)}{2})}{\Gamma\left(\frac{4-n}{2}\right)\Gamma\left(\ell+\frac{5-n}{2}\right)}\right]\frac{A^2}{9}\left(\frac{H_0}{2}\right)^{n-1} \tag{414}$$

For $n \simeq 1$ and $100 \gg \ell \gg 1$, we can approximate this expression to

$$\frac{\ell(\ell+1)C_\ell^{\rm SW}}{2\pi} = \frac{A^2}{9}. \tag{415}$$

This result shows that inflation predicts a very flat spectrum for low ℓ. This prediction has been confirmed by CMB anisotropy measurements. Furthermore, since inflation predicts $\Phi_{\mathbf{k}}^{\rm MD} = \frac{3}{5}\zeta_{\mathbf{k}}$, we find that

$$\pi\,\ell(\ell+1)C_{\rm L}^{\rm SW} = \frac{A_\zeta^2}{25} = \frac{1}{25}\frac{1}{2\,M_{\rm Pl}^2\,\epsilon}\left(\frac{H}{2\pi}\right)^2. \tag{416}$$

Assuming that

$$\frac{\ell(\ell+1)C_\ell^{\rm SW}}{2\pi} \simeq 10^{-10}, \tag{417}$$

we find

$$\boxed{\left(\frac{V}{\epsilon}\right)^{1/4} \simeq 6.7\times 10^{16}\,{\rm GeV}.}$$

Take for instance a model of chaotic inflation with quadratic potential $V(\phi) = \frac{1}{2}m_\phi^2\phi^2$. Using Eq. (183) one easily computes that when there are ΔN e-foldings to go, the value of the inflaton field is $\phi^2_{\Delta N} = (\Delta N/2\pi G)$ and the corresponding value of ϵ is $1/(2\Delta N)$. Taking $\Delta N \simeq 60$ (corresponding to large-angle CMB anisotropies), one finds that COBE normalization imposes $m_\phi \simeq 10^{13}$ GeV.

12 Symmetries of the de Sitter geometry

Before launching ourselves into the topic of non-Gaussianity of the cosmological perturbations, we wish to summarize the symmetries of the de Sitter geometry to understand better the properties of the inflationary perturbations.

The four-dimensional de Sitter space-time of radius H^{-1} is described by the hyperboloid

$$\eta_{AB}X^AX^B = -X_0^2 + X_i^2 + X_5^2 = \frac{1}{H^2} \quad (i = 1, 2, 3), \tag{418}$$

embedded in five-dimensional Minkowski space-time $\mathbb{M}^{1,4}$ with coordinates X^A and flat metric $\eta_{AB} = \mathrm{diag}(-1, 1, 1, 1, 1)$. A particular parametrization of the de Sitter hyperboloid is provided by

$$\begin{aligned} X^0 &= \frac{1}{2H}\left(H\eta - \frac{1}{H\eta}\right) - \frac{1}{2}\frac{x^2}{\eta}, \\ X^i &= \frac{x^i}{H\eta}, \\ X^5 &= -\frac{1}{2H}\left(H\eta + \frac{1}{H\eta}\right) + \frac{1}{2}\frac{x^2}{\eta}, \end{aligned} \tag{419}$$

which may easily be checked that satisfies Eq. (418). The de Sitter metric is the induced metric on the hyperboloid from the five-dimensional ambient Minkowski space-time

$$\mathrm{d}s_5^2 = \eta_{AB}dX^AdX^B. \tag{420}$$

For the particular parametrization (419), for example, we find

$$\mathrm{d}s^2 = \frac{1}{H^2\eta^2}\left(-\mathrm{d}\eta^2 + \mathrm{d}\mathbf{x}^2\right). \tag{421}$$

The group SO(1,4) acts linearly on $\mathbb{M}^{1,4}$. Its generators are

$$J_{AB} = X_A\frac{\partial}{\partial X^B} - X_B\frac{\partial}{\partial X^A} \quad A, B = (0, 1, 2, 3, 5) \tag{422}$$

and satisfy the $SO(1,4)$ algebra

$$[J_{AB}, J_{CD}] = \eta_{AD}J_{BC} - \eta_{AC}J_{BD} + \eta_{BC}J_{AD} - \eta_{BD}J_{AC}. \tag{423}$$

We may split these generators as

$$J_{ij}, \quad P_0 = J_{05}\,, \quad \Pi_i^+ = J_{i5} + J_{0i}\,, \quad \Pi_i^- = J_{i5} - J_{0i}, \tag{424}$$

which act on the de Sitter hyperboloid as

$$
\begin{aligned}
J_{ij} &= x_i\frac{\partial}{\partial x_j} - x_j\frac{\partial}{\partial x_i},\\
P_0 &= \eta\frac{\partial}{\partial \eta} + x^i\frac{\partial}{\partial x^i},\\
\Pi_i^- &= -2H\eta x^i\frac{\partial}{\partial \eta} + H\left(x^2\delta_{ij} - 2x_ix_j\right)\frac{\partial}{\partial x_j} - H\eta^2\frac{\partial}{\partial x_i},\\
\Pi_i^+ &= \frac{1}{H}\frac{\partial}{\partial x_i}
\end{aligned}
\tag{425}
$$

and satisfy the commutator relations

$$
\begin{aligned}
[J_{ij}, J_{kl}] &= \delta_{il}J_{jk} - \delta_{ik}J_{jl} + \delta_{jk}J_{il} - \delta_{jl}J_{ik},\\
[J_{ij}, \Pi_k^\pm] &= \delta_{ik}\Pi_j^\pm - \delta_{jk}\Pi_i^\pm,\\
[\Pi_k^\pm, P_0] &= \mp\Pi_k^\pm,\\
[\Pi_i^-, \Pi_j^+] &= 2J_{ij} + 2\delta_{ij}P_0.
\end{aligned}
\tag{426}
$$

This is nothing else that the conformal algebra. Indeed, by defining

$$
L_{ij} = iJ_{ij}\,, \quad D = -iP_0\,, \quad P_i = -i\Pi_i^+\,, \quad K_i = i\Pi_i^-\,,
\tag{427}
$$

we get

$$
\begin{aligned}
P_i &= -\frac{i}{H}\partial_i,\\
D &= -i\left(\eta\frac{\partial}{\partial\eta} + x^i\partial_i\right),\\
K_i &= -2iHx_i\left(\eta\frac{\partial}{\partial\eta} + x^i\partial_i\right) - iH(-\eta^2 + x^2)\partial_i,\\
L_{ij} &= i\left(x_i\frac{\partial}{\partial x_j} - x_j\frac{\partial}{\partial x_i}\right).
\end{aligned}
\tag{428}
$$

These are also the Killing vectors of de Sitter space-time corresponding to symmetries under space translations (P_i), dilitations (D), special conformal transformations (K_i) and space rotations (L_{ij}). They satisfy the conformal algebra in its standard form

$$[D, P_i] = iP_i, \tag{429}$$

$$[D, K_i] = -iK_i, \tag{430}$$

$$[K_i, P_j] = 2i\Big(\delta_{ij}D - L_{ij}\Big) \tag{431}$$

$$[L_{ij}, P_k] = i\Big(\delta_{jk}P_i - \delta_{ik}P_j\Big), \tag{432}$$

$$[L_{ij}, K_k] = i\Big(\delta_{jk}K_i - \delta_{ik}K_j\Big), \tag{433}$$

$$[L_{ij}, D] = 0, \tag{434}$$

$$[L_{ij}, L_{kl}] = i\Big(\delta_{il}L_{jk} - \delta_{ik}L_{jl} + \delta_{jk}L_{il} - \delta_{jl}L_{ik}\Big). \tag{435}$$

The de Sitter algebra SO(1,4) has two Casimir invariants

$$\mathcal{C}_1 = -\frac{1}{2} J_{AB} J^{AB}\,, \tag{436}$$

$$\mathcal{C}_2 = W_A W^A\,, \qquad W^A = \epsilon^{ABCDE} J_{BC} J_{DE}. \tag{437}$$

Using Eqs. (424) and (427), we find that

$$\mathcal{C}_1 = D^2 + \frac{1}{2}\{P_i, K_i\} + \frac{1}{2} L_{ij} L^{ij}, \tag{438}$$

which turns out to be, in the explicit representation Eq. (428),

$$H^{-2}\mathcal{C}_1 = -\frac{\partial^2}{\partial \eta^2} - \frac{2}{\eta}\frac{\partial}{\partial \eta} + \boldsymbol{\nabla}^2. \tag{439}$$

As a result, $\mathcal{C}_1$ is the Laplace operator on the de Sitter hyperboloid and for a scalar field $\phi(x)$ we have

$$\mathcal{C}_1 \phi(x) = \frac{m^2}{H^2}\phi(x). \tag{440}$$

Let us now consider the case $H\eta \ll 1$. The parametrization (419) turns out then to be

$$\begin{aligned} X^0 &= -\frac{1}{2H^2\eta} - \frac{1}{2}\frac{x^2}{\eta}, \\ X^i &= \frac{x^i}{H\eta}, \\ X^5 &= -\frac{1}{2H^2\eta} + \frac{1}{2}\frac{x^2}{\eta} \end{aligned} \tag{441}$$

and we may easily check that the hyperboloid has been degenerated to the hypercone

$$-X_0^2 + X_i^2 + X_5^2 = 0. \tag{442}$$

We identify points $X^A \equiv \lambda X^A$ (which turns the cone (442) into a projective space). As a result, η in the denominator of the X^A can be ignored due to projectivity condition. Then, on the cone, the conformal group acts linearly, whereas induces the (non-linear) conformal transformations $x_i \to x_i'$ with

$$x_i' = a_i + M_i^j x_j, \tag{443}$$

$$x_i' = \lambda x_i, \tag{444}$$

$$x_i' = \frac{x_i + b_i x^2}{1 + 2b_i x_i + b^2 x^2}. \tag{445}$$

on Euclidean $\mathbb{R}^3$ with coordinates x^i. These transformations correspond to translations and rotations (generated by P_i, L_{ij}), dilations (generated by D) and special conformal transformations (generated by K_i), respectively, acting now on the constant time hypersurfaces of de Sitter space-time. It should be noted that special conformal transformations can be written in terms of inversion

$$x_i \to x_i' = \frac{x_i}{x^2} \tag{446}$$

as inversion×translation×inversion.

12.1 Killing vectors of the de Sitter space

We have seen that the essential kinematical feature of a vacuum dominated de Sitter universe is that the conformal group of certain embeddings of three dimensional hypersurfaces in de Sitter space-time may be mapped (either one-to-one or multiple-to-one) to the geometric isometry group of the full four dimensional space-time into which the hypersurfaces are embedded. The first example of such an embedding of three dimensional hypersurfaces is that of flat Euclidean $\mathbb{R}^3$ in de Sitter space-time in coordinates. The conformal group of the three dimensional spatial $\mathbb{R}^3$ sections is in fact identical (isomorphic) to the isometry group SO(4,1) of the four dimensional de Sitter space-time, as we now review.

Since (eternal) de Sitter space is maximally symmetric, it posseses the maximum number of isometries for a space-time in $n = 4$ dimensions, namely $\frac{n(n+1)}{2} = 10$, corresponding to the 10 solutions of the Killing equation,

$$\nabla_\mu \epsilon^{(\alpha)}_\nu + \nabla_\nu \epsilon^{(\alpha)}_\mu = 0\,, \qquad \mu,\nu = 0,1,2,3\,; \quad \alpha = 1,\dots,10\,. \tag{447}$$

Each of the 10 linearly independent solutions to this equation (labelled by α) is a vector field in de Sitter space corresponding to an infinitesimal coordinate transformation, $x^\mu \to x^\mu + \epsilon^\mu(x)$ that leaves the de Sitter geometry and line element invariant. These are the 10 generators of the de Sitter isometry group, the non-compact Lie group SO(4,1).

The isomorphism with conformal transformations of $\mathbb{R}^3$ is that each of these 10 solutions of (447) may be placed in one-to-one correspondence with the 10 solutions of the conformal Killing equation of three dimensional flat space $\mathbb{R}^3$, *i.e.*

$$\partial_i \xi^{(\alpha)}_j + \partial_j \xi^{(\alpha)}_i = \frac{2}{3}\,\delta_{ij}\,\partial_k \xi^{(\alpha)}_k\,, \qquad i,j,k = 1,2,3; \quad \alpha = 1,\dots 10\,. \tag{448}$$

In (447) the space-time indices μ,ν range over 4 values and ∇_ν is the covariant derivative with respect to the full four dimensional metric of de Sitter space-time, whereas in (448), i,j are three dimensional spatial indices of the three Cartesian coordinates x^i of Euclidean $\mathbb{R}^3$ of one dimension lower with flat metric δ_{ij}. Solutions to the conformal Killing Eq. (448) are transformations of $x^i \to x^i + \xi^i(\vec{x})$ which preserve all angles in $\mathbb{R}^3$. This isomorphism between geometric isometries of $(3+1)$ dimensional de Sitter space-time and conformal transformations of 3 dimensional flat space embedded in it is the origin of conformal invariance of correlation functions generated in a de Sitter phase of the universe.

The 10 solutions of (448) for vector fields in flat $\mathbb{R}^3$ are easily found. They are of two kinds. First there are 6 solutions of (448) with $\partial_k \xi_k = 0$, corresponding to the strict isometries of $\mathbb{R}^3$, namely 3 translations and 3 rotations. Second, there are also 4 solutions of (448) with $\partial_k \xi_k \neq 0$. These are the 4 conformal transformations of flat space that are not strict isometries but preserve all angles. They consist of one global dilation and three special conformal transformations. The Killing Eq. (447) can be rewritten as

$$g_{\nu\lambda}\partial_\mu \epsilon^\lambda + g_{\mu\lambda}\partial_\nu \epsilon^\lambda + \partial_\sigma g_{\mu\nu}\epsilon^\sigma = 0\,, \tag{449}$$

which, for de Sitter space, they provide

$$\partial_t \epsilon_t = 0\,, \tag{450}$$

$$\partial_t \epsilon_i + \partial_i \epsilon_t = 2H\epsilon_i\,, \tag{451}$$

$$\partial_i \epsilon_j + \partial_j \epsilon_i = 2Ha^2 \delta_{ij}\epsilon_t\,. \tag{452}$$

Its solutions of can be catalogued as follows. For $\epsilon_t = 0$ we have the three translations,

$$\epsilon_t^{(\mathrm{T}j)} = 0\,, \qquad \epsilon_i^{(\mathrm{T}j)} = a^2 \delta_i{}^j\,, \qquad j = 1,2,3\,, \tag{453}$$

and the three rotations,

$$\epsilon_\tau^{(\mathrm{R}\ell)} = 0\,, \qquad \epsilon_i^{(\mathrm{R}\ell)} = a^2 \epsilon_{i\ell m} x^m\,, \qquad \ell = 1,2,3\,. \tag{454}$$

The spatial $\mathbb{R}^3$ sections also have four conformal Killing vectors which satisfy the Killing vector equations with $\epsilon_t \neq 0$. They are the three special conformal transformations of $\mathbb{R}^3$,

$$\epsilon_t^{(\mathrm{C})} = -2Hx^n\,, \qquad \epsilon_i^{(\mathrm{C})} = H^2 a^2 (\delta_i{}^n \delta_{jk} x^j x^k - 2\delta_{ij} x^j x^n) - \delta_i^n\,, \qquad n = 1,2,3\,, \tag{455}$$

and the dilation,

$$\epsilon_t^{(\mathrm{D})} = 1\,, \qquad \epsilon_i^{(\mathrm{D})} = Ha^2\, \delta_{ij} x^j\,. \tag{456}$$

This last dilational Killing vector is the infinitesimal form of the finite dilational symmetry,

$$\mathbf{x} \to \lambda \mathbf{x}\,, \tag{457}$$

$$a(\tau) \to \lambda^{-1} a(\tau)\,, \tag{458}$$

$$t \to t - H^{-1} \ln \lambda \tag{459}$$

of de Sitter space. Since the maximum number of Killing isometries in 4 dimensions is 10, there are no other solutions and de Sitter space, being a fully symmetric space, possesses the maximum number of symmetries.

We can understand the issue of scale-invariance rather easily looking at the symmetries of de Sitter. In conformal time the metric during inflation reads approximately de Sitter

$$\mathrm{d}s^2 = \frac{1}{H^2 \tau^2}(-\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2)\,, \tag{460}$$

whose isometry group is SO(4,1). The time-evolving inflaton background is homogeneous and rotationally invariant, so that translations and rotations are good symmetries of the whole system. The dilation isometry

$$\tau \to \lambda\tau, \quad \mathbf{x} \to \lambda\mathbf{x}\,, \tag{461}$$

is also an approximate symmetry of the inflaton background in the limit in which its dynamics varies slowly in time. It is this isometry which guarantees a scale invariant spectrum, independently of the inflaton dynamics. In Fourier space dilatations act on a scalar field $\phi(\mathbf{x}, \tau)$ on large scales as

$$\phi_{\mathbf{k}} \to \lambda^{-3}\phi_{\mathbf{k}/\lambda}\,. \tag{462}$$

Indeed, consider a transformation $\mathbf{x} \to \lambda\mathbf{x}$. Then, in real space $\phi(\mathbf{x}) \to \phi_\lambda(\mathbf{x}) = \phi(\lambda\mathbf{x})$. Expressing this in terms of the Fourier transform of $\phi(\mathbf{x})$ gives how the rescaling acts in Fourier space

$$\phi(\lambda\mathbf{x}) = \int \mathrm{d}^3k\, e^{-i\mathbf{k}\cdot\lambda\mathbf{x}}\phi(\mathbf{k}) = \lambda^{-3}\int \mathrm{d}^3p\, e^{-i\mathbf{p}\cdot\mathbf{x}}\phi(\mathbf{p}/\lambda)\,, \tag{463}$$

where, in the last step, we have made a change in the variable of integration with $\mathbf{p} = \lambda\mathbf{k}$. Therefore, the two-point function is constrained to have the form

$$\langle\phi_{\mathbf{k}_1}\phi_{\mathbf{k}_2}\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2)\frac{F(k_1\tau)}{k_1^3}. \tag{464}$$

If perturbations become time independent when out of the Hubble radius, the function F must be a constant in this limit and this gives a scale invariant spectrum.

13 Non-Gaussianity of the cosmological perturbations

Non-Gaussianity (NG), *i.e.* the study of non-Gaussian contributions to the correlations of cosmological fluctuations, is emerging as an important probe of the early universe. Being a direct measure of inflaton interactions, constraints on primordial NG's will teach us a great deal about the inflationary dynamics and on the mechanism giving rise to the primordial cosmological perturbations. Over the last decade we have accumulated a good deal of observational evidence from CMB and LSS power spectra that the observed structures originated from seed fluctuations in the very early universe. As we have seen, the leading theory explaining the primordial origin of cosmological fluctuations is cosmic inflation, a period of accelerated expansion at very early times. During inflation, microscopic quantum fluctuations were stretched to macroscopic scales to provide the seed fluctuations for the formation of large-scale structures like our own Galaxy. Despite the simplicity of the inflationary paradigm, the mechanism by which cosmological perturbations are generated is not yet established. In the standard slow-roll inflationary scenario associated to one-single field, the inflaton, density perturbations are due to fluctuations of the inflaton itself when it slowly rolls down along its potential. In the curvaton mechanism the final curvature perturbation $\mathcal{R}$ is produced from an initial isocurvature mode associated with the quantum fluctuations of a light scalar (other than the inflaton), the curvaton, whose energy density is negligible during inflation. Recently, other mechanisms for the generation of cosmological perturbations have been proposed: the inhomogeneous reheating scenario, ghost-inflation, the DBI scenario, and from broken symmetries to mention a few.

A precise measurement of the spectral index $n_{\mathcal{R}}$ of comoving curvature perturbations will provide a powerful constraint to slow-roll inflation models and the standard scenario for the generation of cosmological perturbations which predicts $n_{\mathcal{R}}$ close to unity. However, alternative mechanisms generically also predict a value of $n_{\mathcal{R}}$ very close to unity. Thus, even a precise measurement of the spectral index will not allow us to efficiently discriminate among them. On the other hand, the lack gravity-wave signals in CMB anisotropies will not give us any information about the perturbation generation mechanism, since alternative mechanisms predict an amplitude of gravity waves far too small to be detectable by future experiments aimed at observing the B-mode of the CMB polarization.

There is, however, a third observable which will prove fundamental in providing information about the mechanism chosen by Nature to produce the structures we see today. It is the deviation from a Gaussian statistics, *i.e.*, the presence of higher-order connected correlation functions of the perturbations. Indeed, a possible source of NG could be primordial in origin, being specific to a particular mechanism for the generation of the cosmological perturbations. This is what makes a positive detection of NG so relevant: it might help in discriminating among competing scenarios which otherwise might be undistinguishable. While, as we shall see, single-field models of inflation with canonical kinetic terms generically predict a tiny level of NG (of the order of the slow-roll parameters), other models for the generation of the curvature perturbation, such as the curvaton models, may predict a high level of NG. While detection of large primordial NG would not rule out inflation, it would rule out in a single shot the large class of slow-roll models where inflation is driven by a single scalar field with canonical kinetic energy.

NG can measured by various methods. A standard approach is to measure non-Gaussian correlations, *i.e.*, the correlations that vanish for a Gaussian distribution, in the CMB and in high-redshift galaxy surveys. The three-point function, or its Fourier transform, the bispectrum, and the four-point function, the trispectrum, are examples of such correlations. The dimensionless quantities $f_{\rm NL}$ and $g_{\rm NL}$ set the amplitude of the the bispectrum $B_\Phi(k_1, k_2, k_3)$ and trispectrum $T_\Phi(k_1, k_2, k_3, k_4)$ of the (gauge-invariant) gravitational potential Φ, respectively.

A large, detectable amount of NG can be produced when any of the following conditions is violated: single field, canonical kinetic energy, slow-roll and initial adiabatic (the Bunch-Davies) vacuum; an important theoretical discovery made toward the end of the last decade is that violation of each of the above conditions results in unique signals with specific triangular shapes: multi-field models, non-canonical kinetic term models, non-adiabatic-vacuum models (*e.g.*, initially excited states), and non-slow-roll models can generate a NG of the local type where the amplitude of the bispectrum is maximized for squeezed triangles ($k_3 \ll k_2 \simeq k_1$); in such a case $f_{\rm NL}^{\rm loc}$ enters in the second-order gravitational potential Φ expressed in terms of the linear Gaussian field $\Phi_{\rm g}$ (on super-horizon scales)

$$\Phi(\mathbf{x}) = \Phi_{\rm g}(\mathbf{x}) + f_{\rm NL}^{\rm loc}\, \Phi_{\rm g}^2(\mathbf{x})\,. \tag{465}$$

Notice that the parameter of the expansion is $f_{\rm NL}^{\rm loc}\, \Phi_{\rm g}(\mathbf{x}$ which is much smaller than unity, for sure perturbation theory holds. Alternatively, the NG can be for instance of the equilateral type ($k_1 = k_2 = k_3$) or of the flattened/folded type ($k_3 \simeq k_2 \simeq 2k_1$), or even strongly scale-dependent with a sharp cut-off so that NG is very suppressed on large cosmological scales, but sizeable on small scales. The latest constraint on $f_{\rm NL}$ come from the WMAP 7-year data; for instance, $f_{\rm NL}^{\rm loc} = 32 \pm 21$ (68% CL) and $f_{\rm NL}^{\rm eq} = 26 \pm 140$ (68% CL). While the statistical significance of the signal is still low, future experiments, as we shall see, such as the Planck CMB satellite might lead to a detection of local NG as small as $f_{\rm NL}^{\rm loc} \sim 3$ by combining the temperature and polarization bispectra. Bispectra measured from high-redshift galaxy surveys at redshifts $z > 2$ should yield constraints on $f_{\rm NL}^{\rm loc}$ and $f_{\rm NL}^{\rm eq}$ that are comparable to, or even better than, those from CMB experiments.

Non-Gaussianities are also particularly relevant in the high-mass end of the power spectrum of perturbations, *i.e.* on the scale of galaxy clusters, since the effect of NG fluctuations becomes especially visible on the tail of the probability distribution. As a result, both the abundance and the clustering properties of very massive halos are sensitive probes of primordial NG. The dark matter (DM) mass function $\mathrm{d}n(M,z)/\mathrm{d}M$ of halos of mass M at redshift z has been computed in the presence of NG adopting various different techniques: via N-body simulations and analytically through the Press-Schechter (PS) approach to mildly NG fields, the Edgeworth expansion and the excursion set formalism. Deviations from Gaussianity in the DM halo mass function could be detected or significantly constrained by the various planned large-scale galaxy surveys, both ground based (such as DES, PanSTARRS and LSST) and on satellite (such as EUCLID).

The local primordial NG also alters the clustering of DM halos inducing a scale-dependent bias on large scales. Indeed, in the local biasing model the galaxy density field at a given position is described as a local function of the DM density field at the same position. As the primordial NG generates a cross-talk between short and long wavelengths, it alters significantly the local bias and introduces a strong scale dependence. The corresponding limit is $-29 < f_{\rm NL}^{\rm loc} < +70$ at 95% CL.

It is clear that measuring the primordial component of NG correlations offers a new window into the details of the fundamental physics of the primordial universe that are not accessible by Gaussian correlations. To some extent, understanding NG does for inflation what direct detection experiments do for dark matter, or the Large Hadron Collider for the Higgs particle. It probes the interactions of the field sourcing inflation, revealing the fundamental aspects of the physics at very high energies that are not accessible to any collider experiments.

13.1 The generation of non-Gaussianity in the primordial cosmological perturbations: generic considerations

In this subsection we wish to give a generic description of how large NG's can be generated during the primordial expansion of the universe. Suppose that there is a period of inflation, that is (quasi) de Sitter expansion and that there are a number of light fields σ^I which are quantum mechanically excited. A we have previously seen, by the δN formalism, the comoving curvature perturbation ζ on a uniform energy density hypersurface at time $t_{\rm f}$ is, on sufficiently large scales, equal to the perturbation in the time integral of the local expansion from an initial flat hypersurface ($t = t_*$) to the final uniform energy density hypersurface. On sufficiently large scales, the local expansion can be approximated quite well by the expansion of the unperturbed Friedmann universe. Hence the curvature perturbation at time $t_{\rm f}$ can be expressed in terms of the values of the relevant scalar fields $\sigma^I(t_*, \vec{x})$ at t_* (notice the change of an irrelevant sign with respect to the previous definition of ζ (332))

$$\zeta(t_{\rm f}, \vec{x}) = N_I \sigma^I + \frac{1}{2} N_{IJ} \sigma^I \sigma^J + \cdots , \tag{466}$$

where N_I and N_{IJ} are the first and second derivative, respectively, of the number of e-folds

$$N(t_{\rm f}, t_*, \vec{x}) = \int_{t_*}^{t_{\rm f}} {\rm d}t \, H(t, \vec{x}) \, . \tag{467}$$

with respect to the field σ^I. From the expansion (466) one can read off the n-point correlators. For instance, the three- and four-point correlators of the comoving curvature perturbation, the so-called bispectrum and trispectrum respectively, is given by

$$B_\zeta(k_1, k_2, k_3) = N_I N_J N_K B^{IJK}_{k_1 k_2 k_3} + N_I N_{JK} N_L \left(P^{IK}_{k_1} P^{JL}_{k_2} + 2 \text{ permutations} \right) \tag{468}$$

and

$$\begin{aligned} T_\zeta(k_1, k_2, k_3, k_4) &= N_I N_J N_K N_L T^{IJKL}_{k_1 k_2 k_3 k_4} \\ &+ N_{IJ} N_K N_L N_M \left(P^{IK}_{k_1} B^{JLM}_{k_{12} k_3 k_4} + 11 \text{ permutations} \right) \\ &+ N_{IJ} N_{KL} N_M N_N \left(P^{JL}_{k_{12}} P^{IM}_{k_1} P^{KN}_{k_3} + 11 \text{ permutations} \right) \\ &+ N_{IJK} N_L N_M N_N \left(P^{IL}_{k_1} P^{JM}_{k_2} P^{KN}_{k_3} + 3 \text{ permutations} \right) , \end{aligned} \tag{469}$$

where

$$\begin{aligned}
\langle\sigma^I_{\mathbf{k}_1}\sigma^J_{\mathbf{k}_2}\rangle &= (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2)P^{IJ}_{k_1}\,, &(470)\\
\langle\sigma^I_{\mathbf{k}_1}\sigma^J_{\mathbf{k}_2}\sigma^K_{\mathbf{k}_3}\rangle &= (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B^{IJK}_{k_1k_2k_3}\,, &(471)\\
\langle\sigma^I_{\mathbf{k}_1}\sigma^J_{\mathbf{k}_2}\sigma^J_{\mathbf{k}_3}\sigma^L_{\mathbf{k}_4}\rangle &= (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4)T^{IJKL}_{k_1k_2k_3k_4}\,, &(472)
\end{aligned}$$

and $\mathbf{k}_{ij} = (\mathbf{k}_i + \mathbf{k}_j)$. We see that the three-point correlator (and similarly for the four-point one) of the comoving curvature perturbation is the sum of two pieces. One, proportional to the three-point correlator of the σ^I fields, is model-dependent and present when the fields σ^I are intrinsically NG. The second one is universal and is generated when the modes of the fluctuations are superHubble and is present even if the σ^I fields are Gaussian. Therefore, we learn immediately that NG can be induced even if the light fields are purely Gaussian at horizon-crossing, this is, for instance, the case of the curvaton. Nevertheless, in general the NG gets both contributions. Therefore, to compute the three-point function for a specific inflationary model requires a careful treatment of the time-evolution of the vacuum in the presence of interactions (while for the two-point function this effect is higher-order). In practice, computing three-point functions can be algebraically very cumbersome, so we restrict us to citing the final results. The details on how to compute these three-point functions deserves a review of its own. Let us just sketch them.

13.2 A brief Review of the in-in formalism

The problem of computing correlation functions in cosmology differs in important ways from the corresponding analysis of quantum field theory applied to particle physics. In particle physics the central object is the S-matrix describing the transition probability for a state in the far past $|\psi\rangle$ to become some state $|\psi'\rangle$ in the far future, $\langle\psi'|S|\psi\rangle = \langle\psi'(+\infty)|\psi(-\infty)\rangle$. Imposing asymptotic conditions at very early *and* very late times makes sense in this case, since in Minkowski space, states are assumed to non-interacting in the far past and the far future, *i.e.* the asymptotic state are taken to be vacuum state of the free Hamiltonian H_0.

In cosmology, however, we evaluate the expectation values of products of fields *at a fixed time*. Conditions are *not* imposed on the fields at both very early and very late times, but only at very early times, when the wavelength is deep inside the horizon. In this limit (according to the equivalence principle) the interaction picture fields should have the same firm as in Minkowski space. This lead us to the definition of the Bunch-Davies vacuum (the free vacuum in Minkowski space).

To describe the time evolution of cosmological perturbations we split the Hamiltonian into a free part and an interacting part

$$H = H_0 + H_{\rm int}\,. \tag{473}$$

The free-field Hamiltonian H_0 is quadratic in perturbations. Quadratic order was sufficient to compute the two-point correlations. However, the higher-order correlations that concerned us in our study of NG require going beyond quadratic order and defining the interaction Hamiltonian $H_{\rm int}$. The interaction Hamiltonian defines the evolution of states via the well-known time-evolution operator

$$U(\tau_2,\tau_1) = T\exp\left(-i\int_{\tau_1}^{\tau_2} {\rm d}\tau' H_{\rm int}(\tau')\right)\,, \tag{474}$$

where T denotes the time-ordering operator. The time-evolution operator U may be used to relate the interacting vacuum at arbitrary time $|\Omega(\tau)\rangle$ to the free (Bunch-Davies) vacuum $|0\rangle$. We first expand $\Omega(\tau)$ in eigenstates of the free Hamiltonian,

$$|\Omega\rangle = \sum_n |n\rangle\langle n|\Omega(\tau)\rangle\,. \tag{475}$$

Then we evolve $|\Omega(\tau)\rangle$ as

$$|\Omega(\tau_2)\rangle = U(\tau_2,\tau_1)|\Omega(\tau_1)\rangle = |0\rangle\langle 0|\Omega\rangle + \sum_{n\geq 1} e^{+iE_n(\tau_2-\tau_1)}|n\rangle\langle n|\Omega(\tau_1)\rangle\,. \tag{476}$$

From Eq. (476) we see that the choice $\tau_2 = -\infty(1-i\epsilon)$ projects out all excited states. Hence, we have the following relation between the interacting vacuum at $\tau = -\infty(1-i\epsilon)$ and the free vacuum $|0\rangle$

$$|\Omega(-\infty(1-i\epsilon))\rangle = |0\rangle\langle 0|\Omega\rangle\,. \tag{477}$$

Finally, the interacting vacuum at an arbitrary time τ is

$$|{\rm in}\rangle \;\equiv\; |\Omega(\tau)\rangle = U(\tau,-\infty(1-i\epsilon))|\Omega(-\infty(1-i\epsilon))\rangle \tag{478}$$

$$\;=\; T\exp\left(-i\int_{-\infty(1-i\epsilon)}^{\tau} {\rm d}\tau' H_{\rm int}(\tau')\right)|0\rangle\langle 0|\Omega\rangle\,. \tag{479}$$

In the "in-in" formalism, the expectation value $\langle W(\tau)\rangle$, of a product of operators $W(\tau)$ at time τ, is evaluated as

$$\langle W(\tau)\rangle \;\equiv\; \frac{\langle {\rm in}|W(\tau)|{\rm in}\rangle}{\langle {\rm in}|{\rm in}\rangle} \tag{480}$$

$$\;=\; \left\langle 0\right|\left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau'){\rm d}\tau'}\right)^{\dagger} W(\tau)\left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau''){\rm d}\tau''}\right)\left|0\right\rangle\,, \tag{481}$$

or

$$\boxed{\langle W(\tau)\rangle = \left\langle 0\right|\left(\bar{T}e^{-i\int_{-\infty^-}^{\tau} H_{\rm int}(\tau'){\rm d}\tau'}\right) W(\tau)\left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau''){\rm d}\tau''}\right)\left|0\right\rangle}\,, \tag{482}$$

where we defined the anti-time-ordering operator $\bar{T}$ and the notation $-\infty^{\pm} \equiv -\infty(1\mp i\epsilon)$. This definition of the correlation functions $\langle W(\tau)\rangle$ in terms of the interaction Hamiltonian $H_{\rm int}$ is the main result of the "in-in" formalism. The interaction Hamiltonian is computed in the ADM approach to General Relativity and $\langle W(\tau)\rangle$ is then evaluated perturbatively.

For instance, this formalism can be used to compute the three-point function of the curvature perturbation ζ for various inflationary models,

$$\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle(\tau) = \left\langle 0\right| \left(\bar{T}e^{-i\int_{-\infty^-}^{\tau} H_{\rm int}(\tau')\mathrm{d}\tau'}\right) \zeta_{\mathbf{k}_1}(\tau)\zeta_{\mathbf{k}_2}(\tau)\zeta_{\mathbf{k}_3}(\tau) \left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau'')\mathrm{d}\tau''}\right) \left|0\right\rangle . \quad (483)$$

Let us sketch how the interaction Hamiltonian is computed The inflationary action is expanded perturbatively

$$S = S_0[\bar{\phi}, \bar{g}_{\mu\nu}] + S_2[\zeta^2] + S_3[\zeta^3] + \cdots \, . \quad (484)$$

Here, we have defined a background part S_0, a quadratic free-field part S_2 and a non-linear interaction term S_3. The background action S_0 defines the Hubble parameter H and the slow-roll parameters ϵ and η. The free-field action S_2 defines the time-evolution of the mode functions $\zeta(\tau)$ in the interaction picture (often denoted by $\zeta_{\rm I}(\tau)$). The non-linear part of the action defines the interaction Hamiltonian, *e.g.* at cubic order $S_3 = -\int \mathrm{d}\tau H_{\rm int}(\zeta_{\rm I})$. Schematically, the interaction Hamiltonian takes the following form

$$H_{\rm int} = \sum_i f_i(\epsilon, \eta, \dots)\zeta_{\rm I}^3(\tau) \, . \quad (485)$$

If we define the expansion of the operator corresponding to the Mukhanov variable, $v = 2a^2\epsilon\mathcal{R}$, in terms of creation and annihilation operators

$$\hat{v}_{\mathbf{k}}(\tau) = v_k(\tau)\hat{a}_{\mathbf{k}} + v_k^*(\tau)\hat{a}^\dagger_{-\mathbf{k}} \, . \quad (486)$$

The mode functions $v_k(\tau)$ were defined uniquely by initial state boundary conditions when all modes were deep inside the horizon

$$v_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right) . \quad (487)$$

The free two-point correlation function is

$$\langle 0|\hat{v}_{\mathbf{k}_1}(\tau_1)\hat{v}_{\mathbf{k}_2}(\tau_2)|0\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2)G_{k_1}(\tau_1, \tau_2) \, , \quad (488)$$

with

$$G_{k_1}(\tau_1, \tau_2) \equiv v_k(\tau_1)v_k^*(\tau_2) \, . \quad (489)$$

Expansion of Eqn. (483) in powers of $H_{\rm int}$ gives:

- at zeroth order

$$\langle W(\tau)\rangle^{(0)} = \langle 0|W(\tau)|0\rangle \, , \quad (490)$$

  where $W(\tau) \equiv \zeta_{\mathbf{k}_1}(\tau)\zeta_{\mathbf{k}_2}(\tau)\zeta_{\mathbf{k}_3}(\tau)$.

- at first order

$$\langle W(\tau)\rangle^{(1)} = 2\,\mathrm{Re}\left[-i\int_{-\infty^+}^{\tau} \mathrm{d}\tau' \langle 0|W(\tau)H_{\rm int}(\tau')|0\rangle\right] . \quad (491)$$

- at second order

$$\langle W(\tau)\rangle^{(2)} = -2\,\mathrm{Re}\left[\int_{-\infty^+}^{\tau} \mathrm{d}\tau' \int_{-\infty^+}^{\tau'} \mathrm{d}\tau'' \langle 0|W(\tau)H_{\rm int}(\tau')H_{\rm int}(\tau'')|0\rangle\right] + \int_{-\infty^-}^{\tau} \mathrm{d}\tau' \int_{-\infty^+}^{\tau} \mathrm{d}\tau'' \langle 0|H_{\rm int}(\tau')W(\tau)H_{\rm int}(\tau'')|0\rangle\,. \qquad (492)$$

In the bispectrum calculations the zeroth-order term (490) vanishes for Gaussian initial conditions. The leading result therefore comes from Eq. (491). Evaluating Eq. (491) makes use of Wick's theorem to expresses the result as products of two-point functions (489).

13.3 The shapes of non-Gaussianity

Let us discuss the various shapes of the NG. One of the first ways to parameterize non-Gaussianity phenomenologically was via a non-linear correction to a Gaussian perturbation $\mathcal{R}_{\rm g}$,

$$\boxed{\zeta(\mathbf{x}) = \zeta_{\rm g}(\mathbf{x}) + \frac{3}{5} f_{\rm NL}^{\rm local}\left[\zeta_{\rm g}(\mathbf{x})^2 - \langle \zeta_{\rm g}(\mathbf{x})^2\rangle\right]}\,. \qquad (493)$$

This definition is local in real space and therefore called *local* NG. Experimental constraints on non-Gaussianity are often set on the parameter $f_{\rm NL}^{\rm local}$ defined via Eq. (493). The factor of 3/5 in Eq. (493) is conventional since non-Gaussianity was first defined in terms of the Newtonian potential, $\Phi(\mathbf{x}) = \Phi_{\rm g}(\mathbf{x}) + f_{\rm NL}^{\rm local}\left[\Phi_{\rm g}(\mathbf{x})^2 - \langle \Phi_{\rm g}(\mathbf{x})^2\rangle\right]$, which during the matter era is related to ζ by a factor of 3/5. Using Eq. (493) the bispectrum of local non-Gaussianity may be derived

$$B_\zeta(k_1,k_2,k_3) = \frac{6}{5} f_{\rm NL}^{\rm local} \times \left[P_\zeta(k_1)P_\zeta(k_2) + P_\zeta(k_2)P_\zeta(k_3) + P_\zeta(k_3)P_\zeta(k_1)\right]\,. \qquad (494)$$

For a scale-invariant spectrum, $P_\zeta(k) = Ak^{-3}$, this is

$$B_\zeta(k_1,k_2,k_3) = \frac{6}{5} f_{\rm NL}^{\rm local} \times A^2 \left[\frac{1}{(k_1k_2)^3} + \frac{1}{(k_2k_3)^3} + \frac{1}{(k_3k_1)^3}\right]\,. \qquad (495)$$

Without loss of generality, let us order the momenta such that $k_3 \le k_2 \le k_1$. The bispectrum for local non-Gaussianity is then largest when the smallest k (*i.e.* k_3) is very small, $k_3 \ll k_1 \sim k_2$. The other two momenta are then nearly equal. In this *squeezed* limit, the bispectrum for local non-Gaussianity becomes

$$\lim_{k_3\ll k_1\sim k_2} B_{\mathcal{R}}(k_1,k_2,k_3) = \frac{12}{5} f_{\rm NL}^{\rm local} \times P_\zeta(k_1)P_\zeta(k_3)\,. \qquad (496)$$

The delta function in the definition of the bispectrum enforces that the three Fourier modes of the bispectrum form a closed triangle. Different inflationary models predict maximal signal for different triangle configurations. This *shape of non-Gaussianity* is potentially a powerful probe of the mechanism that laid down the primordial perturbations.

It will be convenient to define the shape function

$$\mathcal{S}(k_1, k_2, k_3) \equiv N(k_1 k_2 k_3)^2 B_\zeta(k_1, k_2, k_3)\,, \tag{497}$$

where N is an appropriate normalization factor. Two commonly discussed shapes are the *local* model, *cf.* Eq. (495),

$$\mathcal{S}^{\rm local}(k_1, k_2, k_3) \propto \frac{K_3}{K_{111}}\,, \tag{498}$$

and the *equilateral* model,

$$\mathcal{S}^{\rm equil}(k_1, k_2, k_3) \propto \frac{\tilde{k}_1 \tilde{k}_2 \tilde{k}_3}{K_{111}}\,. \tag{499}$$

Here, we have introduced a notation

$$K_p = \sum_i (k_i)^p \quad \text{with} \quad K = K_1 \tag{500}$$

$$K_{pq} = \frac{1}{\Delta_{pq}} \sum_{i \neq j} (k_i)^p (k_j)^q \tag{501}$$

$$K_{pqr} = \frac{1}{\Delta_{pqr}} \sum_{i \neq j \neq l} (k_i)^p (k_j)^q (k_l)^q \tag{502}$$

$$\tilde{k}_{ip} = K_p - 2(k_i)^p \quad \text{with} \quad \tilde{k}_i = \tilde{k}_{i1}\,, \tag{503}$$

where $\Delta_{pq} = 1 + \delta_{pq}$ and $\Delta_{pqr} = \Delta_{pq}(\Delta_{qr} + \delta_{pr})$ (no summation). This notation significantly compresses the increasingly complex expressions for the bispectra discussed in the literature.

We have argued above that for scale-invariant fluctuations the bispectrum is only a function of the two ratios k_2/k_1 and k_3/k_1. We hence define the rescaled momenta

$$x_i \equiv \frac{k_i}{k_1}\,. \tag{504}$$

We have ordered the momenta such that $x_3 \leq x_2 \leq 1$. The triangle inequality implies $x_2 + x_3 > 1$. In the following we plot $\mathcal{S}(1, x_2, x_3)$ (see Figs. 16, 18, and 19). We use the normalization, $\mathcal{S}(1,1,1) \equiv 1$. To avoid showing equivalent configurations twice $\mathcal{S}(1, x_2, x_3)$ is set to zero outside the triangular region $1 - x_2 \leq x_3 \leq x_2$. We see in Fig. 16 that the signal for the local shape is concentrated at $x_3 \approx 0$, $x_2 \approx 1$, while the equilateral shape peaks at $x_2 \approx x_3 \approx 1$. Fig. 17 illustrates how the different triangle shapes are distributed in the x_2-x_3 plane.

Physically motivated models for producing non-Gaussian perturbations often produce signals that peak at special triangle configurations. Three important special cases are:

i) *squeezed triangle* ($k_1 \approx k_2 \gg k_3$)

This is the dominant mode of models with multiple light fields during inflation.

ii) *equilateral triangle* ($k_1 = k_2 = k_3$)

Signals that peak at equilateral triangles arise in models with higher-derivative interactions and non-trivial speeds of sound.

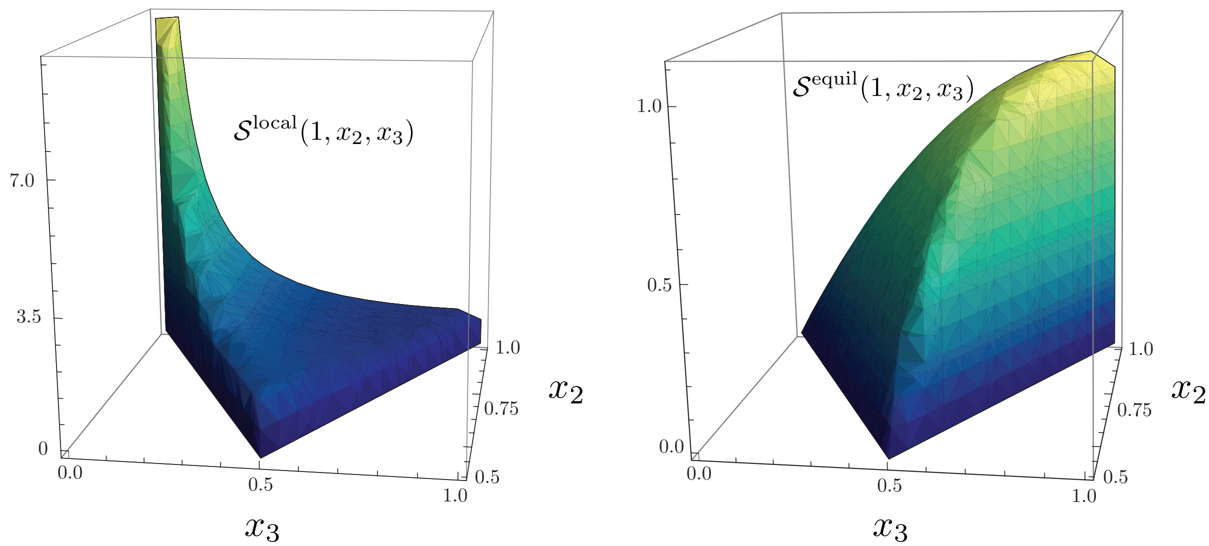

Figure 16: 3D plots of the *local* and *equilateral* bispectra. The coordinates x_2 and x_3 are the rescaled momenta k_2/k_1 and k_3/k_1, respectively. Momenta are order such that $x_3 < x_2 < 1$ and satsify the triangle inequality $x_2 + x_3 > 1$.

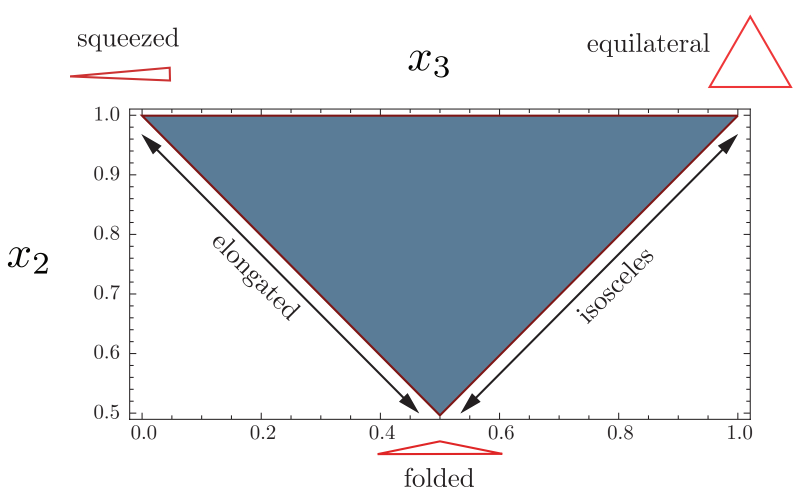

Figure 17: Shapes of Non-Gaussianity. The coordinates x_2 and x_3 are the rescaled momenta k_2/k_1 and k_3/k_1, respectively. Momenta are order such that $x_3 < x_2 < 1$ and satsify the triangle inequality $x_2 + x_3 > 1$.

iii) *folded triangle* ($k_1 = 2k_2 = 2k_3$)

Folded triangles arise in models with non-standard initial states.

In addition, there are the intermediate cases: *elongated triangles* ($k_1 = k_2 + k_3$) and *isosceles triangles* ($k_1 > k_2 = k_3$). For arbitrary shape functions we measure the magnitude of NG by defining the

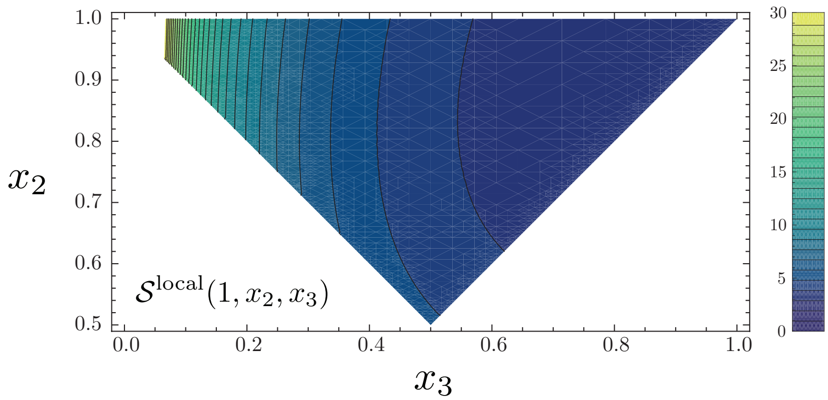

Figure 18: Contour plot of the *local* bispectrum.

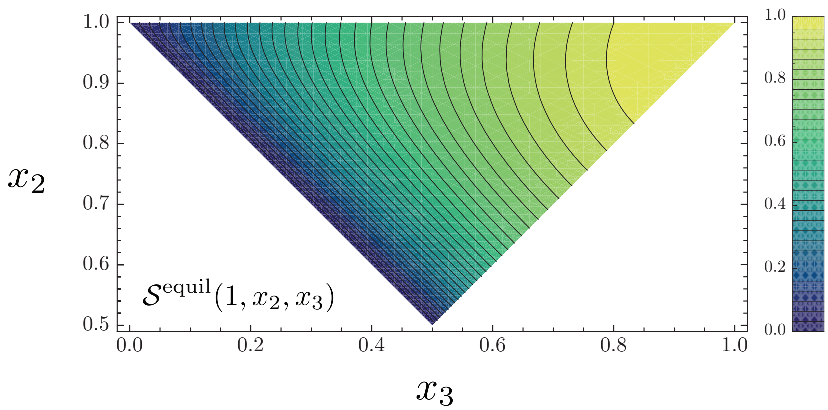

Figure 19: Contour plot of the *equilateral* bispectrum.

generalized f_{NL} parameter

$$\boxed{f_{\mathrm{NL}} \equiv \frac{5}{18} \frac{B_\zeta(k,k,k)}{P_\zeta(k)^2}} . \tag{505}$$

In this definition the amplitude of non-Gaussianity is normalized in the equilateral configuration.

13.4 Theoretical Expectations

Let us analyze what are the theoretical expectations from the various classes of models.

13.4.1 Single-Field Slow-Roll Inflation

Successful slow-roll inflation demands that the interactions of the inflaton field are weak. Since the wave function of free fields in the ground state is Gaussian, the fluctuations created during slow-roll inflation are expected to be Gaussian. A lengthy computation gives

$$\mathcal{S}^{\rm SR}(k_1,k_2,k_3) \;\propto\; (\epsilon-2\eta)\frac{K_3}{K_{111}}+\epsilon\left(K_{12}+8\frac{K_{22}}{K}\right) \tag{506}$$

$$\approx\; (4\epsilon-2\eta)\,\mathcal{S}^{\rm local}(k_1,k_2,k_3)+\frac{5}{3}\epsilon\,\mathcal{S}^{\rm equil}(k_1,k_2,k_3)\,, \tag{507}$$

where $\mathcal{S}^{\rm local}$ and $\mathcal{S}^{\rm equil}$ are normalized so that $\mathcal{S}^{\rm local}(k,k,k)=\mathcal{S}^{\rm equil}(k,k,k)$. The bispectrum for slow-roll inflation peaks at squeezed triangles and has an amplitude that is suppressed by slow-roll parameters

$$f_{\rm NL}^{\rm SR}=\mathcal{O}(\epsilon,\eta)\,. \tag{508}$$

To get convinced about this result one can use the δN formalism applied to a single field model. One finds that

$$f_{\rm NL}\sim\frac{N_{\phi\phi}}{N_\phi^2}\,. \tag{509}$$

Using the fact that $N_\phi=H/\dot\phi$, one gets that

$$\frac{N_{\phi\phi}}{N_\phi^2}=\frac{1}{N_\phi^2}\frac{\rm d}{\dot\phi\,{\rm d}t}N_\phi=\left(\frac{\dot H}{\dot\phi}-\frac{H\ddot\phi}{\dot\phi^2}\right)\times\frac{1}{\dot\phi}\times\frac{\dot\phi^2}{H^2}=\left(\frac{\dot H}{H^2}-\frac{\ddot\phi}{H\dot\phi}\right)=(-\epsilon+\eta-\epsilon)=\eta-2\epsilon\,. \tag{510}$$

This incomplete result makes intuitive sense since the slow-roll parameters characterize deviations of the inflaton from a free field. To get the full behaviour, let us consider Eq. (468) restricting ourselves to the one-single field case. Then

$$B_\zeta(k_1,k_2,k_3)=N_\phi^3B^\phi_{k_1k_2k_3}+N_\phi^2N_{\phi\phi}\left(P^\phi(k_1)P^\phi(k_2)+2\text{ permutations}\right)\,. \tag{511}$$

At first-order we have $\delta\phi^{(1)}_{\mathbf{k}}\simeq(H/2\pi)$. However at second-order there is a local correction to the amplitude of vacuum fluctuations at Hubble exit due to first-order perturbations in the local Hubble rate $\widetilde{H}(\phi)$. This is determined by the local scalar field value due to longer wavelength modes that have already left the horizon

$$\widetilde{H}(\phi)=H(\phi)+H'(\phi)\int_0^{k_{\rm c}}\frac{{\rm d}^3k}{(2\pi)^3}\delta\phi_{\mathbf{k}}\,, \tag{512}$$

where $k_{\rm c}$ is the cut-off wavenumber which selects only long wavelength perturbation at horizon crossing. Thus for a mode $k_1\simeq k_2\gg k_3$ one can write at second-order

$$\delta\phi^{(2)}_{\mathbf{k}_1}\simeq\frac{H'}{H}\int_0^{k_{\rm c}}\frac{{\rm d}^3k'}{(2\pi)^3}\delta\phi^{(1)}_{\mathbf{k}'}\delta\phi^{(1)}_{\mathbf{k}_1-\mathbf{k}'}\,, \tag{513}$$

where $k_1\simeq k_2\gg k_{\rm c}$. The bispectrum for the inflation field therefore reads in the squeezed limit

$$\begin{aligned}
B^{\phi}_{k_1k_2k_3} &\simeq \langle \delta\phi^{(2)}_{\mathbf{k}_1}\delta\phi^{(1)}_{\mathbf{k}_2}\delta\phi^{(1)}_{\mathbf{k}_3}\rangle + \langle \delta\phi^{(1)}_{\mathbf{k}_1}\delta\phi^{(2)}_{\mathbf{k}_2}\delta\phi^{(1)}_{\mathbf{k}_3}\rangle \simeq (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)2\frac{H'}{H}P^{\phi}(k_3)P^{\phi}(k_1) \\
&\simeq -2\epsilon\left(\frac{H}{\dot\phi}\right)(2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)P^{\phi}(k_3)P^{\phi}(k_1) \\
&= -2\epsilon N_{\phi}(2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)P^{\phi}(k_3)P^{\phi}(k_1)\,. \qquad (514)
\end{aligned}$$

Using Eq. (510) we then get

$$\begin{aligned}
B_{\zeta}(k_1,k_2,k_3) &= (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\left[-2\epsilon N^4_{\phi}P^{\phi}(k_3)P^{\phi}(k_1)+2(\eta-2\epsilon)P^{\zeta}(k_3)P^{\zeta}(k_1)\right] \\
&= (2\eta-6\epsilon)P^{\zeta}(k_3)P^{\zeta}(k_1) \\
&= (n_{\zeta}-1)P^{\zeta}(k_3)P^{\zeta}(k_1) \qquad (515)
\end{aligned}$$

and we have obtained a $(n_\zeta-1)$ suppression. In fact, this result goes beyond the slow-roll assumption: under the assumption of single-field inflation, but *no* other assumptions about the inflationary action, one is able to prove a powerful theorem

$$\boxed{\lim_{k_3\to 0}\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,(n_\zeta-1)\,P_\zeta(k_1)P_\zeta(k_3)}\,. \qquad (516)$$

Eq. (516) states that for single-field inflation, the squeezed limit of the three-point function is suppressed by $(1-n_\zeta)$ and vanishes for perfectly scale-invariant perturbations. The same happens for higher-order correlators. A detection of non-Gaussianity in the squeezed limit can therefore rule out single-field inflation. In particular, this statement is independent of: the form of the potential, the form of the kinetic term (or sound speed) and the initial vacuum state.

The proof is the following. The squeezed triangle correlates one long-wavelength mode, $k_{\rm L}=k_3$ to two short-wavelength modes, $k_{\rm S}=k_1\approx k_2$,

$$\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle \approx \langle(\zeta_{\mathbf{k}_{\rm S}})^2\zeta_{\mathbf{k}_{\rm L}}\rangle\,. \qquad (517)$$

Modes with longer wavelengths freeze earlier. Therefore, $k_{\rm L}$ will be already frozen outside the horizon when the two smaller modes freeze and acts as a background field for the two short-wavelength modes.

Why should $(\zeta_{\mathbf{k}_{\rm S}})^2$ be correlated with $\zeta_{\mathbf{k}_{\rm L}}$? The theorem says that "it is not correlated if $\zeta_{\rm k}$ is precisely scale-invariant". The proof is simplest in real-space. The long-wavelength curvature perturbation $\zeta_{\mathbf{k}_{\rm L}}$ rescales the spatial coordinates (or changes the effective scale factor) within a given Hubble patch

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)e^{-2\zeta}\mathrm{d}\mathbf{x}^2\,. \qquad (518)$$

The two-point function $\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\rangle$ will depend on the value of the background fluctuations $\zeta_{\mathbf{k}_{\rm L}}$ already frozen outside the horizon. In position space the variation of the two-point function given by the long-wavelength fluctuations $\zeta_{\rm L}$ is at linear order

$$\frac{\partial}{\partial \zeta_{\rm L}} \langle \zeta(x)\zeta(0)\rangle \cdot \zeta_{\rm L} = -x \frac{\mathrm{d}}{\mathrm{d}x} \langle \zeta(x)\zeta(0)\rangle \cdot \zeta_{\rm L} \, . \tag{519}$$

To get the three-point function one multiplies Eq. (519) by $\zeta_{\rm L}$ and average over it. Going to Fourier space gives Eq. (516).

13.4.2 Models with Large Non-Gaussianity

Although for a single-field *slow-roll* inflation non-Gaussianity is always small, single-field models can still give large non-Gaussianity if higher-derivative terms are important during inflation (as opposed to assuming a canonical kinetic term and no higher-derivative corrections as in slow-roll inflation). Consider the following action

$$S = \frac{1}{2} \int \mathrm{d}^4 x \sqrt{-g} \, [R - P(X, \phi)] \, , \quad \text{where} \quad X \equiv (\partial_\mu \phi)^2 \, . \tag{520}$$

Here, $P(X, \phi)$ is an arbitrary function of the kinetic term $X = (\partial_\mu \phi)^2$ and hence can contain higher-derivative interactions. These models in general have a non-trivial sound speed for the propagation of fluctuations

$$c_s^2 \equiv \frac{P_{,X}}{P_{,X} + 2X P_{,XX}} \, . \tag{521}$$

The second-order action for ζ (giving $P_{\mathcal{R}}$) is

$$S_{(2)} = \int \mathrm{d}^4 x \, \epsilon \left[a^3 (\dot{\zeta})^2 / c_s^2 - a (\partial_i \zeta)^2 \right] + \mathcal{O}(\epsilon^2) \tag{522}$$

The third-order action for ζ is

$$S_{(3)} = \int \mathrm{d}^4 x \, \epsilon^2 \left[\dots a^3 (\dot{\zeta})^2 \zeta / c_s^2 + \dots a (\partial_i \zeta)^2 \zeta + \dots a^3 (\dot{\zeta})^3 / c_s^2 \right] + \mathcal{O}(\epsilon^3) \, . \tag{523}$$

We notice that the third-order action is surpressed by an extra factor of ϵ relative to the second-order action. This is a reflection of the fact that non-Gaussianity is small in the slow-roll limit: $P(X, \phi) = X - V(\phi)$, $c_s^2 = 1$. However, away from the slow-roll limit, for small sound speeds, $c_s^2 \ll 1$, a few interaction terms in Eq. (523) get boosted and non-Gaussianity can become significant. The signal is peaked at equilateral triangles, with

$$f_{\rm NL}^{\rm equil} = -\frac{35}{108} \left(\frac{1}{c_s^2} - 1 \right) + \frac{5}{81} \left(\frac{1}{c_s^2} - 1 - 2\Lambda \right) , \tag{524}$$

where

$$\Lambda \equiv \frac{X^2 P_{,XX} + \frac{2}{3} X^3 P_{,XXX}}{X P_{,X} + 2 X^2 P_{,XX}} \, . \tag{525}$$

Whether actions with arbitrary $P(X, \phi)$ exist in consistent high-energy theories is an important challenge for these models. It is encouraging that one of the most interesting models of inflation in string theory,

13.4.3 Multiple Fields

In single-field slow-roll inflation interactions of the inflaton are constrained by the requirement that inflation should occur. However, if more than one field was relevant during inflation this constraint may be circumvented. Models like the curvaton mechanism or inhomogeneous reheating exploit this to create non-Gaussian fluctuations via fluctuations is a second field that is not the inflaton. The signal is peaked at squeezed triangles. Let us describe in some detail the curvaton case. We expand the curvaton field up to first-order in the perturbations around the homogeneous background as $\sigma(\tau, \mathbf{x}) = \sigma_0(\tau) + \delta\sigma$, the linear perturbations satisfy on large scales

$$\delta\sigma'' + 2\,\mathcal{H}\delta\sigma' + a^2 \frac{\partial^2 V}{\partial\sigma^2}\,\delta\sigma = 0\,. \tag{526}$$

As a result on superHubble scales its fluctuations $\delta\sigma$ will be Gaussian distributed and with a nearly scale-invariant spectrum given by

$$\mathcal{P}^{\frac{1}{2}}_{\delta\sigma}(k) \approx \frac{H_*}{2\pi}\,, \tag{527}$$

where the subscript $*$ denotes the epoch of horizon exit $k = aH$. Once inflation is over the inflaton energy density will be converted to radiation (γ) and the curvaton field will remain approximately constant until $H^2 \sim m_\sigma^2$. At this epoch the curvaton field begins to oscillate around the minimum of its potential which can be safely approximated to be quadratic $V \approx \frac{1}{2} m_\sigma^2 \sigma^2$. During this stage the energy density of the curvaton field just scales as non-relativistic matter $\rho_\sigma \propto a^{-3}$. The energy density in the oscillating field is

$$\rho_\sigma(\tau, \mathbf{x}) \approx m_\sigma^2 \sigma^2(\tau, \mathbf{x})\,, \tag{528}$$

and it can be expanded into a homogeneous background $\rho_\sigma(\tau)$ and a second-order perturbation $\delta\rho_\sigma$ as

$$\rho_\sigma(\tau, \mathbf{x}) = \rho_\sigma(\tau) + \delta\rho_\sigma(\tau, \mathbf{x}) = m_\sigma^2 \sigma + 2 m_\sigma^2\, \sigma\, \delta\sigma + m_\sigma^2 \delta\sigma^2\,. \tag{529}$$

The ratio $\delta\sigma/\sigma$ remains constant and the resulting relative energy density perturbation is

$$\frac{\delta\rho_\sigma}{\rho_\sigma} = 2\left(\frac{\delta\sigma}{\sigma}\right)_* + \left(\frac{\delta\sigma}{\sigma}\right)^2_*\,, \tag{530}$$

where the $*$ stands for the value at horizon crossing. Such perturbations in the energy density of the curvaton field produce in fact a primordial density perturbation well after the end of inflation and a potentially large NG.

During the oscillations of the curvaton field, the total curvature perturbation can be written as a weighted sum of the single curvature perturbations

$$\zeta = (1 - f)\zeta_\gamma + f\zeta_\sigma\,, \tag{531}$$

where the quantity

$$f = \frac{3\rho_\sigma}{4\rho_\gamma + 3\rho_\sigma} \tag{532}$$

defines the relative contribution of the curvaton field to the total curvature perturbation. Working under the approximation of sudden decay of the curvaton field. Under this approximation the curvaton and the radiation components ρ_σ and ρ_γ satisfy separately the energy conservation equations

$$\begin{aligned}\rho_\gamma' &= -4\,\mathcal{H}\rho_\gamma\,,\\ \rho_\sigma' &= -3\,\mathcal{H}\rho_\sigma\,,\end{aligned} \tag{533}$$

and the curvature perturbations ζ_i remains constant on superHubble scales until the decay of the curvaton. In the curvaton scenario it is supposed that the curvature perturbation in the radiation produced at the end of inflation is negligible. From Eq. (531) the total curvature perturbation during the curvaton oscillations is given by

$$\zeta = f\zeta_\sigma \simeq \frac{f}{3}\frac{\delta\rho_\sigma}{\rho_\sigma} \simeq \frac{f}{3}\left[2\left(\frac{\delta\sigma}{\sigma}\right)_* + \left(\frac{\delta\sigma}{\sigma}\right)^2_*\right], \tag{534}$$

from which we deduce that

$$\zeta = \zeta_{\rm g} + \frac{3}{4f}(\zeta_{\rm g}^2 - \langle\zeta_{\rm g}^2\rangle)\,,\ \ \zeta_{\rm g} = (2f/3)(\delta\sigma/\sigma)_*\,, \tag{535}$$

and therefore

$$f_{\rm NL}^{\rm loc} = \frac{5}{4f}\,. \tag{536}$$

We discover that the NG can be very large if $f \ll 1$. Furthermore, the NG is of the local type. This is because it is generated not at horizon-crossing, but when the fluctuations are already outside the horizon.

It is nice to reproduce the same result with the δN formalism. In the absence of interactions, fluids with a barotropic equation of state, such as radiation ($P_\gamma = \rho_\gamma/3$) or the non-relativistic curvaton ($P_\sigma = 0$), have a conserved curvature perturbation (notice again a change of an irrelevant sign from Eq. (332))

$$\zeta_i = \delta N + \frac{1}{3}\int_{\bar\rho_i}^{\rho_i} \frac{{\rm d}\tilde\rho_i}{\tilde\rho_i + P_i(\tilde\rho_i)}\,. \tag{537}$$

We assume that the curvaton decays on a uniform-total density hypersurface corresponding to $H = \Gamma$, *i.e.*, when the local Hubble rate equals the decay rate for the curvaton (assumed constant). Thus on this hypersurface we have

$$\rho_\gamma(t_{\rm dec},\mathbf{x}) + \rho_\sigma(t_{\rm dec},\mathbf{x}) = \bar\rho(t_{\rm dec})\,, \tag{538}$$

where, for the sake of clarity, we use a bar to denote the homogeneous, unperturbed quantity. Note that we have $\zeta = \delta N$ on the decay surface, and we can interpret ζ as the perturbed expansion, or "δN". Assuming all the curvaton decay products are relativistic, we have that ζ is conserved after the curvaton decay since the total pressure is simply $P = \rho/3$.

By contrast the local curvaton and radiation densities on this decay surface may be inhomogeneous and we have from Eq. (537)

$$\zeta_\gamma = \zeta + \frac{1}{4}\ln\left(\frac{\rho_\gamma}{\bar{\rho}_\gamma}\right), \tag{539}$$

$$\zeta_\sigma = \zeta + \frac{1}{3}\ln\left(\frac{\rho_\sigma}{\bar{\rho}_\sigma}\right), \tag{540}$$

or, equivalently,

$$\rho_\gamma = \bar{\rho}_\gamma e^{4(\zeta_\gamma-\zeta)}, \tag{541}$$

$$\rho_\sigma = \bar{\rho}_\sigma e^{3(\zeta_\sigma-\zeta)}. \tag{542}$$

Requiring that the total density is uniform on the decay surface, we obtain the relation

$$(1-\Omega_{\sigma,\rm dec})e^{4(\zeta_\gamma-\zeta)} + \Omega_{\sigma,\rm dec}e^{3(\zeta_\sigma-\zeta)} = 1, \tag{543}$$

where $\Omega_{\sigma,\rm dec} = \bar{\rho}_\sigma/(\bar{\rho}_\gamma + \bar{\rho}_\sigma)$ is the dimensionless density parameter for the curvaton at the decay time.

For simplicity we will restrict the following analysis to the simplest curvaton scenario in which the curvature perturbation in the radiation fluid before the curvaton decays is negligible, *i.e.*, $\zeta_\gamma = 0$. After the curvaton decays the universe is dominated by radiation, with equation of state $P = \rho/3$, and hence the curvature perturbation, ζ, is non-linearly conserved on large scales. With $\zeta_\gamma = 0$ Eq. (543) reads

$$e^{4\zeta} - \left[\Omega_{\sigma,\rm dec}e^{3\zeta_\sigma}\right]e^{\zeta} + \left[\Omega_{\sigma,\rm dec} - 1\right] = 0. \tag{544}$$

At first-order Eq. (543) gives

$$4(1-\Omega_{\sigma,\rm dec})\zeta^{(1)} = 3\Omega_{\sigma,\rm dec}(\zeta_\sigma^{(1)} - \zeta^{(1)}), \tag{545}$$

and hence we can write

$$\zeta^{(1)} = f\zeta_\sigma^{(1)}, \tag{546}$$

where

$$f = \frac{3\Omega_{\sigma,\rm dec}}{4-\Omega_{\sigma,\rm dec}} = \left.\frac{3\bar{\rho}_\sigma}{3\bar{\rho}_\sigma + 4\bar{\rho}_\gamma}\right|_{t_{\rm dec}}. \tag{547}$$

At second order Eq. (543) gives

$$4(1-\Omega_{\sigma,\rm dec})\zeta^{(2)} - 16(1-\Omega_{\sigma,\rm dec})\zeta^{(1)2} = 3\Omega_{\sigma,\rm dec}(\zeta_\sigma^{(2)} - \zeta^{(2)}) + 9\Omega_{\sigma,\rm dec}(\zeta_\sigma^{(1)} - \zeta^{(1)})^2, \tag{548}$$

and hence

$$\zeta^{(2)} = \frac{3}{4f}\zeta^{(1)2}, \tag{549}$$

which gives again Eq. (536).

13.4.4 A test of multi-field models of inflation

The collapsed limit of the four-point correlator is particularly important because, together with the squeezed limit of the three-point correlator, it may lead to the so-called Suyama-Yamaguchi (SY) inequality. Consider a class of multi-field models which satisfy the following conditions: a) scalar fields are responsible for generating curvature perturbations and b) the fluctuations in scalar fields at the horizon crossing are scale invariant and Gaussian. The second condition amounts to assuming that the connected three- and four-point correlations of the σ^I fields vanish and that the NG is generated at superHubble scales. If so, the three- and four-point correlators of the comoving curvature perturbation (468) and (469) respectively reduce to

$$B_\zeta(k_1,k_2,k_3) = N_I N_{JK} N_L \left(P^{IK}_{k_1} P^{JL}_{k_2} + 2 \text{ permutations}\right) \tag{550}$$

and

$$\begin{aligned} T_\zeta(k_1,k_2,k_3,k_4) &= N_{IJ}N_{KL}N_M N_N \left(P^{JL}_{k_{12}} P^{IM}_{k_1} P^{KN}_{k_3} + 11 \text{ permutations}\right) \\ &+ N_{IJK}N_L N_M N_N \left(P^{IL}_{k_1} P^{JM}_{k_2} P^{KN}_{k_3} + 3 \text{ permutations}\right), \end{aligned} \tag{551}$$

Notice in particular that in the collapsed limit $k_{12} \simeq 0$ the last term of the four-point correlator (551) is subleading. By defining the nonlinear parameters $f_{\rm NL}$ and $\tau_{\rm NL}$ as

$$\begin{aligned} f_{\rm NL} &= \frac{5}{12}\frac{\langle \zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle'}{P^\zeta_{k_1}P^\zeta_{k_2}} \qquad (k_1 \ll k_2 \sim k_3), \\ \tau_{\rm NL} &= \frac{1}{4}\frac{\langle \zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\zeta_{\mathbf{k}_4}\rangle'}{P^\zeta_{k_1}P^\zeta_{k_3}P^\zeta_{k_{12}}} \qquad (k_{12} \simeq 0). \end{aligned} \tag{552}$$

From these expressions we deduce that

$$\frac{6}{5}f_{\rm NL} = \frac{N^I N_{IJ} N^J}{(N_I N^I)^2}, \tag{553}$$

and

$$\tau_{\rm NL} = \frac{N^I N_{JI} N^{JK} N_K}{(N_I N^I)^3}. \tag{554}$$

Defining now the vectors $V_I = N_{IJ}N^J$ and $N^I$and using the Cauchy-Schwarz inequality $(V\cdot N)^2 \leq V^2 N^2$, we may immediately deduce that

$$(V_I V^I)(N_I N^I) \geq (V_I N^I)^2 \Rightarrow (N^I N_{IJ} N_{JK} N^K)(N_I N^I) \geq (N^J N_{IJ} N^I)^2 \tag{555}$$

or

$$\boxed{\tau_{\rm NL} \geq \left(\frac{6}{5}f_{\rm NL}\right)^2}. \tag{556}$$

In fact, this inequality holds also if the light fields are not Gaussian at horizon-crossing. The SY inequality is more a consequence of fundamental physical principles rather than of pure mathematical arrangements. The observation of a strong violation of the inequality will then have profound implications for inflationary models as it will imply either that multi-field inflation cannot be responsible for generating the observed fluctuations independently of the details of the model or that some new non-trivial degrees of freedom play a role during inflation.

13.4.5 Non-Standard Vacuum

If inflation started in an excited state rather than in the Bunch-Davies vacuum, remnant non-Gaussianity may be observable (unless inflation lasted much more than the minimal number of e-folds, in which case the effect is exponentially diluted). The signal is peaked at folded triangles with a shape function

$$\mathcal{S}^{\rm folded}(k_1, k_2, k_3) \propto \frac{1}{K_{111}}(K_{12} - K_3) + 4\frac{K_2}{(\tilde{k}_1\tilde{k}_2\tilde{k}_3)^2}\,. \tag{557}$$

Part V

The impact of the non-Gaussianity on the CMB anisotropies

Statistics like the bispectrum and the trispectrum of the CMB can be used to assess the level of primordial NG (and possibly its shape) on various cosmological scales and to discriminate it from the one induced by secondary anisotropies and systematic effects. A positive detection of a primordial NG in the CMB at some level might therefore confirm and/or rule out a whole class of mechanisms by which the cosmological perturbations have been generated.

One should take into account that there are many sources of NG in CMB anisotropies, beyond the primordial one. The most relevant sources are the so-called secondary anisotropies, which arise after the last scattering epoch. These anisotropies can be divided into two categories: scattering secondaries, when the CMB photons scatter with electrons along the line of sight, and gravitational secondaries when effects are mediated by gravity. Among the scattering secondaries we may list the thermal Sunyaev-Zeldovich effect, where hot electrons in clusters transfer energy to the CMB photons, the kinetic Sunyaev-Zeldovich effect produced by the bulk motion of the electrons in clusters, the Ostriker-Vishniac effect, produced by bulk motions modulated by linear density perturbations, and effects due to reionization processes. The scattering secondaries are most significant on small angular scales as density inhomogeneities, bulk and thermal motions grow and become sizeable on small length-scales when structure formation proceeds.

Gravitational secondaries arise from the change in energy of photons when the gravitational potential is time-dependent, the ISW effect, and gravitational lensing. At late times, when the Universe becomes dominated by the dark energy, the gravitational potential on linear scales starts to decay, causing the ISW effect mainly on large angular scales. Other secondaries that result from a time dependent potential are the Rees-Sciama effect, produced during the matter-dominated epoch by the time evolution of the potential on non-linear scales.

The fact that the potential never grows appreciably means that most second order effects created by gravitational secondaries are generically small compared to those created by scattering ones. However, when a photon propagates from the last scattering to us, its path may be deflected because of the gravitational lensing. This effect does not create anisotropies, but only modifies existing ones. Since photons with large wavenumbers k are lensed over many regions ($\sim k/H$, where H is the Hubble rate) along the line of sight, the corresponding second-order effect may be sizeable. The three-point function arising from the correlation of the gravitational lensing and ISW effects generated by the matter distribution along the line of sight and the Sunyaev-Zeldovich effect are large and detectable by Planck. A crucial issue is the level of contamination to the extraction of the primordial NG the secondary effects can produce.

Another relevant source of NG comes from the physics operating at the recombination. A naive estimate would tell that these non-linearities are tiny being suppressed by an extra power of the gravitational potential. However, the dynamics at recombination is quite involved because all the non-linearities in the evolution of the baryon-photon fluid at recombination and the ones coming from general relativity should be accounted for. This complicated dynamics might lead to unexpected suppressions or enhancements of the NG at recombination. Recently the computation of the full system of Boltzmann equations, describing the evolution of the photon, baryon and Cold Dark Matter (CDM) fluids, at second order and neglecting polarization, has been performed. These equations allow to follow the time evolution of the CMB anisotropies at second order on all angular scales from the early epochs, when the cosmological perturbations were generated, to the present time, through the recombination era. These calculations set the stage for the computation of the full second-order radiation transfer function at all scales and for a a generic set of initial conditions specifying the level of primordial NG. Of course, for specific effects on small angular scales like Sunyaev-Zel'dovich, gravitational lensing, etc., fully non-linear calculations would provide a more accurate estimate of the resulting CMB anisotropy, however, as long as the leading contribution to second-order statistics like the bispectrum is concerned, second-order perturbation theory suffices.

While post-inflationary contributions to the NG in the CMB anisotropies are expected to be of order unity, as we shall describe later in a oversimplified example, if the primordial NG is much larger than unity one can safely use the linear transfer function. Indeed, in the evolution of the CMB anistropies, the primordial NG enters as an initial condition. Suppose for instance that at second-order one has an equation of the symbolic form

$$F[\ddot{\Phi}^{(2)}, \dot{\Phi}^{(1)}, \cdots] = \mathcal{S}[\Phi^{(1)2}, \cdots] \, . \tag{558}$$

The second-order gravitational potential $\Phi^{(2)}$ will be the sum of the homogeneous solution plus the inhomogeneous proportional to the source. The homogeneous solution resembles the first-order solution with some NG initial condition set on primordial epochs. If, for instance, $|f_{\rm NL}| \gg 1$, then the primordial NG dominates and one can effectively work at the linear level. This observation is crucial to assess the impact of large NG on the CMB anisotropies.

13.5 Why do we expect NG in the cosmological perturbations?

Before tackling the problem of interest – the computation of the cosmological perturbations at second-order after the inflationary era– we first provide a simple, but insightful computation, which illustrates why we expect that the cosmological perturbations develop some NG even if the latter is not present at some primordial epoch. This example will help the reader to understand why the cosmological perturbations are inevitably affected by nonlinearities, beyond those arising at some primordial epoch. The reason is clear: gravity is nonlinear and it feeds this property into the cosmological perturbations during the post-inflationary evolution of the universe. As gravity talks to all fluids, this transmission is inevitable. We will adopt the Poisson gauge which eliminates one scalar degree of freedom from the g_{0i} component of the metric and one scalar and two vector degrees of freedom from g_{ij}. We will use a metric of the form

$$\mathrm{d}s^2 = -e^{2\Phi}\mathrm{d}t^2 + 2a(t)\omega_i \mathrm{d}x^i dt + a^2(t)(e^{-2\Psi}\delta_{ij} + \chi_{ij})\mathrm{d}x^i \mathrm{d}x^j \, , \tag{559}$$

where ω_i and χ_{ij} are the vector and tensor peturbation modes respectively. Each metric perturbation can be expanded into a linear (first-order) and a second-order part, as for example, the gravitational potential $\Phi = \Phi^{(1)} + \Phi^{(2)}/2$. However in the metric (559) the choice of the exponentials greatly helps in computing the relevant expressions, and thus we will always keep them where it is convenient.

We now consider the long wavelength modes of the CMB anisotropies, *i.e.* we focus on scales larger than the horizon at last-scattering. We can therefore neglect vector and tensor perturbation modes in the metric. For the vector perturbations the reason is that we are they contain gradient terms being produced as non-linear combination of scalar-modes and thus they will be more important on small scales (remember linear vector modes are not generated in standard mechanisms for cosmological perturbations, as inflation). The tensor contribution can be neglected for two reasons. First, the tensor perturbations produced from inflation on large scales give a negligible contribution to the higher-order statistics of the Sachs-Wolfe effect being of the order of (powers of) the slow-roll parameters during inflation (this holds for linear tensor modes as well as for tensor modes generated by the non-linear evolution of scalar perturbations during inflation).

Since we are interested in the cosmological perturbations on large scales, that is in perturbations whose wavelength is larger than the Hubble radius at last scattering, a local observer would see them in the form of a classical – possibly time-dependent – (nearly zero-momentum) homogeneous and isotropic background. Therefore, it should be possible to perform a change of coordinates in such a way as to absorb the super-Hubble modes and work with a metric of an homogeneous and isotropic Universe (plus, of course, cosmological perturbations on scale smaller than the horizon). We split the gravitational potential Φ as

$$\Phi = \Phi_\ell + \Phi_s \,, \tag{560}$$

where Φ_ℓ stands for the part of the gravitational potential receiving contributions only from the super-Hubble modes; Φ_s receives contributions only from the sub-horizon modes

$$\begin{aligned} \Phi_\ell &= \int \frac{d^3k}{(2\pi)^3}\, \theta\,(aH-k)\; \Phi_{\mathbf{k}}\; e^{i\mathbf{k}\cdot\mathbf{x}} \,, \\ \Phi_s &= \int \frac{d^3k}{(2\pi)^3}\, \theta\,(k-aH)\; \Phi_{\mathbf{k}}\; e^{i\mathbf{k}\cdot\mathbf{x}} \,, \end{aligned} \tag{561}$$

where H is the Hubble rate computed with respect to the cosmic time, $H = \dot{a}/a$, and $\theta(x)$ is the step function. Analogous definitions hold for the other gravitational potential Ψ.

By construction Φ_ℓ and Ψ_ℓ are a collection of Fourier modes whose wavelengths are larger than the horizon length and we may safely neglect their spatial gradients. Therefore Φ_ℓ and Ψ_ℓ are only functions of time. This amounts to saying that we can absorb the large-scale perturbations in the metric (559) by the following redefinitions

$$d\bar{t} = e^{\Phi_\ell} dt \,, \tag{562}$$

$$\bar{a} = a\, e^{-\Psi_\ell} \,. \tag{563}$$

The new metric describes a homogeneous and isotropic Universe

$$ds^2 = -d\bar{t}^2 + \bar{a}^2 \delta_{ij}\, dx^i\, dx^j \,, \tag{564}$$

where for simplicity we have not included the sub-horizon modes. On super-horizon scales one can regard the Universe as a collection of regions of size of the Hubble radius evolving like unperturbed patches with metric (564).

Let us now go back to the quantity we are interested in, namely the anisotropies of the CMB as measured today by an observer $\mathcal{O}$. If she/he is interested in the CMB anisotropies at large scales, the effect of super-Hubble modes is encoded in the metric (564). During their travel from the last scattering surface – to be considered as the emitter point $\mathcal{E}$ – to the observer, the CMB photons suffer a redshift determined by the ratio of the emitted frequency $\overline{\omega}_{\mathcal{E}}$ to the observed one $\overline{\omega}_{\mathcal{O}}$

$$\overline{T}_{\mathcal{O}} = \overline{T}_{\mathcal{E}}\,\frac{\overline{\omega}_{\mathcal{O}}}{\overline{\omega}_{\mathcal{E}}}\,, \tag{565}$$

where $\overline{T}_{\mathcal{O}}$ and $\overline{T}_{\mathcal{E}}$ are the temperatures at the observer point and at the last scattering surface, respectively.

What is then the temperature anisotropy measured by the observer? The expression (565) shows that the measured large-scale anisotropies are made of two contributions: the intrinsic inhomogeneities in the temperature at the last scattering surface and the inhomogeneities in the scaling factor provided by the ratio of the frequencies of the photons at the departure and arrival points. Let us first consider the second contribution. As the frequency of the photon is the inverse of a time period, we get immediately the fully non-linear relation

$$\frac{\overline{\omega}_{\mathcal{E}}}{\overline{\omega}_{\mathcal{O}}} = \frac{\omega_{\mathcal{E}}}{\omega_{\mathcal{O}}} e^{-\Phi_{\ell\mathcal{E}}+\Phi_{\ell\mathcal{O}}}\,. \tag{566}$$

As for the temperature anisotropies coming from the intrinsic temperature fluctuation at the emission point, it maybe worth to recall how to obtain this quantity in the longitudinal gauge at first-order. By expanding the photon energy density $\rho_\gamma \propto T_\gamma^4$, the intrinsic temperature anisotropies at last scattering are given by $\delta^{(1)}T_{\mathcal{E}}/T_{\mathcal{E}} = (1/4)\delta^{(1)}\rho_\gamma/\rho_\gamma$. One relates the photon energy density fluctuation to the gravitational perturbation first by implementing the adiabaticity condition $\delta^{(1)}\rho_\gamma/\rho_\gamma = (4/3)\delta^{(1)}\rho_{\rm m}/\rho_{\rm m}$, where $\delta^{(1)}\rho_{\rm m}/\rho_{\rm m}$ is the relative fluctuation in the matter component, and then using the energy constraint of Einstein equations $\Phi^{(1)} = -(1/2)\delta^{(1)}\rho_{\rm m}/\rho_{\rm m}$. The result is $\delta^{(1)}T_{\mathcal{E}}/T_{\mathcal{E}} = -2\Phi^{(1)}_{\mathcal{E}}/3$. Summing this contribution to the anisotropies coming from the redshift factor (566) expanded at first order provides the standard (linear) Sachs-Wolfe effect $\delta^{(1)}T_{\mathcal{O}}/T_{\mathcal{O}} = \Phi^{(1)}_{\mathcal{E}}/3$. Following the same steps, we may easily obtain its full non-linear generalization.

Let us first relate the photon energy density $\overline{\rho}_\gamma$ to the energy density of the non-relativistic matter $\overline{\rho}_{\rm m}$ by using the adiabaticity conditon. Again here a bar indicates that we are considering quantities in the locally homogeneous Universe described by the metric (564). Using the energy continuity equation on large scales $\partial\overline{\rho}/\partial\overline{t} = -3\overline{H}(\overline{\rho}+\overline{P})$, where $\overline{H} = d\ln\overline{a}/d\overline{t}$ and $\overline{P}$ is the pressure of the fluid, we have shown that there exists a conserved quantity in time at any order in perturbation theory

$$-\zeta \equiv \ln\overline{a} + \frac{1}{3}\int^{\overline{\rho}} \frac{d\overline{\rho'}}{(\overline{\rho'}+\overline{P'})}\,. \tag{567}$$

As we know, the perturbation ζ is a gauge-invariant quantity representing the non-linear extension of the curvature perturbation on uniform energy density hypersurfaces on superHubble scales for adiabatic fluids. At the non-linear level the adiabaticity condition generalizes to

$$\frac{1}{3}\int\frac{{\rm d}\overline{\rho}_{\rm m}}{\overline{\rho}_{\rm m}} = \frac{1}{4}\int\frac{{\rm d}\overline{\rho}_\gamma}{\overline{\rho}_\gamma}\,, \tag{568}$$

or

$$\ln \overline{\rho}_{\rm m} = \ln \overline{\rho}_\gamma^{3/4}\,. \tag{569}$$

Next we need to relate the photon energy density to the gravitational potentials at the non-linear level. The energy constraint inferred from the (0-0) component of Einstein equations in the matter-dominated era with the "barred" metric (564) is

$$\overline{H}^2 = \frac{8\pi G_N}{3}\overline{\rho}_{\rm m}\,. \tag{570}$$

Using Eqs. (562) and (563) the Hubble parameter $\overline{H}$ reads

$$\overline{H} = \frac{1}{\overline{a}}\frac{{\rm d}\overline{a}}{{\rm d}\overline{t}} = e^{-\Phi_\ell}(H - \dot{\Psi}_\ell)\,, \tag{571}$$

where $H = {\rm d}\ln a/{\rm d}t$ is the Hubble parameter in the "unbarred" metric. Eq. (570) thus yields an expression for the energy density of the non-relativistic matter which is fully nonlinear, being expressed in terms of the gravitational potential Φ_ℓ

$$\overline{\rho}_{\rm m} = \rho_{\rm m} e^{-2\Phi_\ell}\,, \tag{572}$$

where we have dropped $\dot{\Psi}_\ell$ which is negligible on large scales.

The expression for the intrinsic temperature of the photons at the last scattering surface $\overline{T}_{\mathcal{E}} \propto \overline{\rho}_\gamma^{1/4}$ follows from Eqs. (569) and (572)

$$\overline{T}_{\mathcal{E}} = T_{\mathcal{E}}\, e^{-2\Phi_\ell/3}\,. \tag{573}$$

Plugging Eqs. (566) and (573) into the expression (565) we are finally able to provide the expression for the CMB temperature which is fully nonlinear and takes into account both the gravitational red-shift of the photons due to the metric perturbations at last scattering and the intrinsic temperature anisotropies

$$\overline{T}_{\mathcal{O}} = \left(\frac{\omega_{\mathcal{O}}}{\omega_{\mathcal{E}}}\right) T_{\mathcal{E}}\, e^{\Phi_\ell/3}\,. \tag{574}$$

From Eq. (574) we read the *non-perturbative* anisotropy corresponding to the Sachs-Wolfe effect

$$\frac{\delta_{np}\overline{T}_{\mathcal{O}}}{T_{\mathcal{O}}} = e^{\Phi_\ell/3} - 1\,. \tag{575}$$

Eq. (575) represents *at any order in perturbation theory* the extension of the linear Sachs-Wolfe effect. At first order one gets

$$\frac{\delta^{(1)}T_{\mathcal{O}}}{T_{\mathcal{O}}} = \frac{1}{3}\Phi^{(1)}\,, \tag{576}$$

and at second order

$$\frac{1}{2}\frac{\delta^{(2)}T_{\mathcal{O}}}{T_{\mathcal{O}}} = \frac{1}{6}\Phi^{(2)} + \frac{1}{18}\left(\Phi^{(1)}\right)^2\,. \tag{577}$$

This result shows that the CMB anisotropies is nonlinear on large scales and that a source of NG is inevitably sourced by gravity and that the corresponding nonlinearities are order unity in units of the linear gravitational potential.

13.6 Primordial non-Gaussianity and the CMB anisotropies

With the assumption of working with large primordial NG, one can estimate the impact of primordial NG on the on the CMB anisotropies as follows. The observed CMB temperature fluctuation field $\Delta T(\hat{\mathbf{n}})/T$ is expanded into the spherical harmonics:

$$a_{\ell m} \equiv \int d^2\hat{\mathbf{n}} \frac{\Delta T(\hat{\mathbf{n}})}{T} Y^*_{\ell m}(\hat{\mathbf{n}}) \,, \tag{578}$$

where hats denote unit vectors. The CMB angular bispectrum is given by

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \,, \tag{579}$$

and the angle-averaged bispectrum is defined by

$$B_{\ell_1 \ell_2 \ell_3} \equiv \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \,, \tag{580}$$

where the matrix is the Wigner-3j symbol. The bispectrum $B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ must satisfy the triangle conditions and selection rules: $m_1 + m_2 + m_3 = 0$, $\ell_1 + \ell_2 + \ell_3 = \text{even}$, and $|\ell_i - \ell_j| \le \ell_k \le \ell_i + \ell_j$ for all permutations of indices. Thus, $B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ consists of the Gaunt integral, $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$, defined by

$$\begin{aligned} \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} &\equiv \int d^2\hat{\mathbf{n}} Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}(\hat{\mathbf{n}}) Y_{\ell_3 m_3}(\hat{\mathbf{n}}) \\ &= \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} . \end{aligned} \tag{581}$$

$\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ is real, and satisfies all the conditions mentioned above.

Given the rotational invariance of the universe, $B_{\ell_1 \ell_2 \ell_3}$ is written as

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} = \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} b_{\ell_1 \ell_2 \ell_3} \,, \tag{582}$$

where $b_{\ell_1 \ell_2 \ell_3}$ is an arbitrary real symmetric function of ℓ_1, ℓ_2, and ℓ_3. This form of equation (582) is necessary and sufficient to construct generic $B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ under the rotational invariance. Thus, we shall frequently use $b_{\ell_1 \ell_2 \ell_3}$ instead of $B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ in this paper, and call this function the "reduced" bispectrum, as $b_{\ell_1 \ell_2 \ell_3}$ contains all physical information in $B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$. Since the reduced bispectrum does not contain the Wigner-3j symbol that merely ensures the triangle conditions and selection rules, it is easier to calculate and useful to quantify the physical properties of the bispectrum.

The observable quantity, the angle-averaged bispectrum $B_{\ell_1 \ell_2 \ell_3}$, is obtained by substituting equation (582) into (580),

$$B_{\ell_1 \ell_2 \ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1 \ell_2 \ell_3} \,, \tag{583}$$

where we have used the identity:

$$\sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} . \tag{584}$$

Alternatively, one can define the bispectrum in the flat-sky approximation,

$$\langle a(\boldsymbol{\ell}_1) a(\boldsymbol{\ell}_1) a(\boldsymbol{\ell}_3) \rangle = (2\pi)^2 \delta^{(2)} \left(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3 \right) B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) , \tag{585}$$

where $\boldsymbol{\ell}$ is the two dimensional wave-vector on the sky. This definition of $B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3)$ corresponds to equation (582), given the correspondence of $\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} \to \delta^{(2)} \left(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3 \right)$ in the flat-sky limit. Thus,

$$b_{\ell_1 \ell_2 \ell_3} \approx B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) \qquad \text{(flat-sky approximation)} \tag{586}$$

is satisfied.

If the primordial fluctuations are adiabatic scalar fluctuations, then

$$a_{\ell m} = 4\pi (-i)^\ell \int \frac{\mathrm{d}^3 \mathbf{k}}{(2\pi)^3} \Phi(\mathbf{k}) g_{T\ell}(k) Y_{\ell m}^*(\hat{\mathbf{k}}) , \tag{587}$$

where, as usual, $\Phi(\mathbf{k})$ is the primordial curvature perturbation in the Fourier space, and $g_{T\ell}(k)$ is the radiation transfer function. $a_{\ell m}$ thus takes over the non-Gaussianity, if any, from $\Phi(\mathbf{k})$.

In this subsection, we explore the simplest weak local model of NG non-linear case:

$$\Phi(\mathbf{x}) = \Phi_{\mathrm{g}}(\mathbf{x}) + f_{\mathrm{NL}} \left(\Phi_{\mathrm{g}}^2(\mathbf{x}) - \left\langle \Phi_{\mathrm{g}}^2(\mathbf{x}) \right\rangle \right) , \tag{588}$$

in real space, where $\Phi_{\mathrm{g}}(\mathbf{x})$ denotes as usual the linear Gaussian part of the perturbation.

In the Fourier space, $\Phi(\mathbf{k})$ is decomposed into two parts:

$$\Phi(\mathbf{k}) = \Phi_{\mathrm{g}}(\mathbf{k}) + \Phi_{\mathrm{NG}}(\mathbf{k}) , \tag{589}$$

and accordingly,

$$a_{\ell m} = a_{\ell m}^{\mathrm{g}} + a_{\ell m}^{\mathrm{NG}} , \tag{590}$$

where $\Phi_{\mathrm{NG}}(\mathbf{k})$ is the non-linear part defined by

$$\Phi_{\mathrm{NG}}(\mathbf{k}) \equiv f_{\mathrm{NL}} \left[\int \frac{\mathrm{d}^3 \mathbf{p}}{(2\pi)^3} \Phi_{\mathrm{g}}(\mathbf{k}+\mathbf{p}) \Phi_{\mathrm{g}}^*(\mathbf{p}) - (2\pi)^3 \delta^{(3)}(\mathbf{k}) \left\langle \Phi_{\mathrm{g}}^2(\mathbf{x}) \right\rangle \right] . \tag{591}$$

In this model, a non-vanishing component of the $\Phi(\mathbf{k})$-field bispectrum is

$$\langle \Phi_{\mathrm{g}}(\mathbf{k}_1) \Phi_{\mathrm{g}}(\mathbf{k}_2) \Phi_{\mathrm{NG}}(\mathbf{k}_3) \rangle = 2(2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) f_{\mathrm{NL}} P_\Phi(k_1) P_\Phi(k_2) , \tag{592}$$

where $P_\Phi(k)$ is as usual the linear power spectrum given by Substituting equation (587) into (579), using equation (592) for the $\Phi(\mathbf{k})$-field bispectrum, and then integrating over angles $\hat{\mathbf{k}}_1$, $\hat{\mathbf{k}}_3$, and $\hat{\mathbf{k}}_3$, we obtain the primary CMB angular bispectrum,

$$\begin{aligned} B_{\ell_1\ell_2\ell_3}^{m_1m_2m_3} &= \left\langle a_{\ell_1m_1}^{\rm g} a_{\ell_2m_2}^{\rm g} a_{\ell_3m_3}^{\rm NG} \right\rangle + \left\langle a_{\ell_1m_1}^{\rm g} a_{\ell_2m_2}^{\rm NG} a_{\ell_3m_3}^{\rm g} \right\rangle + \left\langle a_{\ell_1m_1}^{\rm NG} a_{\ell_2m_2}^{\rm g} a_{\ell_3m_3}^{\rm g} \right\rangle \\ &= 2\mathcal{G}_{\ell_1\ell_2\ell_3}^{m_1m_2m_3} \int_0^\infty r^2 {\rm d}r \left[b_{\ell_1}^{\rm g}(r) b_{\ell_2}^{\rm g}(r) b_{l_3}^{\rm NG}(r) + b_{\ell_1}^{\rm g}(r) b_{\ell_2}^{|rmNG}(r) b_{\ell_3}^{\rm g}(r) + b_{\ell_1}^{\rm NG}(r) b_{\ell_2}^{\rm g}(r) b_{\ell_3}^{\rm g}(r) \right], \end{aligned} \tag{593}$$

where

$$b_\ell^{\rm g}(r) \equiv \frac{2}{\pi} \int_0^\infty k^2 {\rm d}k P_\Phi(k) g_{T\ell}(k) j_\ell(kr), \tag{594}$$

$$b_\ell^{\rm NG}(r) \equiv \frac{2}{\pi} \int_0^\infty k^2 {\rm d}k f_{\rm NL} g_{T\ell}(k) j_\ell(kr). \tag{595}$$

Note that $b_\ell^{\rm g}(r)$ is a dimensionless quantity, while $b_\ell^{\rm NG}(r)$ has a dimension of L^{-3}. One confirms that the form of equation (582) holds. Thus, the reduced bispectrum, $b_{\ell_1\ell_2\ell_3} = B_{\ell_1\ell_2\ell_3}^{m_1m_2m_3} \left(\mathcal{G}_{\ell_1\ell_2\ell_3}^{m_1m_2m_3} \right)^{-1}$ (Eq.(582)), for the primordial non-Gaussianity is

$$b_{\ell_1\ell_2\ell_3} = 2\int_0^\infty r^2 {\rm d}r \left[b_{\ell_1}(r) b_{\ell_2}^{\rm g}(r) b_{\ell_3}^{\rm NG}(r) + \text{cyclic} \right]. \tag{596}$$

Therefore $b_{\ell_1\ell_2\ell_3}$ is fully specified by a single constant parameter $f_{\rm NL}$, as the cosmological parameters will be precisely determined by measuring the CMB angular power spectrum C_ℓ.

We now discuss the detectability of CMB experiments to the primary non-Gaussianity in the bispectrum. Suppose that we try to fit the observed bispectrum $B_{\ell_1\ell_2\ell_3}^{\rm obs}$ by theoretically calculated bispectra which include both primary and secondary sources. Then we minimize χ^2 defined by

$$\chi^2 \equiv \sum_{2\le\ell_1\le\ell_2\le\ell_3} \frac{\left(B_{\ell_1\ell_2\ell_3}^{\rm obs} - \sum_i A_i B_{\ell_1\ell_2\ell_3}^{(i)} \right)^2}{\sigma_{\ell_1\ell_2\ell_3}^2}, \tag{597}$$

where i denotes a component such as the primary, the SZ and lensing effects, extragalactic sources, and so on. Unobservable modes $\ell = 0$ and 1 are removed. In case that the non-Gaussianity is small, the cosmic variance of the bispectrum is given by the six-point function of $a_{\ell lm}$. The variance of $B_{\ell_1\ell_2\ell_3}$ is then calculated as

$$\sigma_{\ell_1\ell_2\ell_3}^2 \equiv \left\langle B_{\ell_1\ell_2\ell_3}^2 \right\rangle - \left\langle B_{\ell_1\ell_2\ell_3} \right\rangle^2 \approx \mathcal{C}_{\ell_1}\mathcal{C}_{\ell_2}\mathcal{C}_{\ell_3}\Delta_{\ell_1\ell_2\ell_3}, \tag{598}$$

where $\Delta_{\ell_1\ell_2\ell_3}$ takes values 1, 2, and 6 for cases of that all ℓ's are different, two of them are same, and all are same, respectively. $\mathcal{C}_\ell \equiv C_\ell + C_\ell^{\rm N}$ is the total CMB angular power spectrum, which includes the power spectrum of the detector noise $C_\ell^{\rm N}$. We do not include C_ℓ from secondary sources, as they are totally subdominant compared with the primary C_ℓ and $C_\ell^{\rm N}$ for relevant experiments.

Taking $\partial\chi^2/\partial A_i = 0$, we obtain the normal equation,

$$\sum_j \left[\sum_{2\le\ell_1\le\ell_2\le\ell_3} \frac{B_{\ell_1\ell_2\ell_3}^{(i)} B_{\ell_1\ell_2\ell_3}^{(j)}}{\sigma_{\ell_1\ell_2\ell_3}^2} \right] A_j = \sum_{2\le\ell_1\le\ell_2\le\ell_3} \frac{B_{\ell_1\ell_2\ell_3}^{\rm obs} B_{\ell_1\ell_2\ell_3}^{(i)}}{\sigma_{\ell_1\ell_2\ell_3}^2}. \tag{599}$$

Thus, we define the Fisher matrix F_{ij} as

$$
\begin{aligned}
F_{ij} \equiv \sum_{2\le\ell_1\le\ell_2\le\ell_3} \frac{B^{(i)}_{\ell_1\ell_2\ell_3} B^{(j)}_{\ell_1\ell_2\ell_3}}{\sigma^2_{\ell_1\ell_2\ell_3}} &= \frac{2}{\pi} \sum_{2\le\ell_1\le\ell_2\le\ell_3} \left(\ell_1+\frac{1}{2}\right)\left(\ell_2+\frac{1}{2}\right)\left(\ell_3+\frac{1}{2}\right) \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \\
&\times \frac{b^{(i)}_{\ell_1\ell_2\ell_3} b^{(j)}_{\ell_1\ell_2\ell_3}}{\sigma^2_{\ell_1\ell_2\ell_3}}, \qquad (600)
\end{aligned}
$$

where we have used equation (583) to replace $B_{\ell_1\ell_2\ell_3}$ by the reduced bispectrum $b_{\ell_1\ell_2\ell_3}$ (see Eq.(582) for definition). Since the covariance matrix of A_i is F^{-1}_{ij}, we define the signal-to-noise ratio $(S/N)_i$ for a component i, the correlation coefficient r_{ij} between different components i and j, and the degradation parameter d_i of $(S/N)_i$ due to r_{ij} as

$$
\left(\frac{S}{N}\right)_i \equiv \frac{1}{\sqrt{F^{-1}_{ii}}}, \qquad (601)
$$

$$
r_{ij} \equiv \frac{F^{-1}_{ij}}{\sqrt{F^{-1}_{ii} F^{-1}_{jj}}}, \qquad (602)
$$

$$
d_i \equiv F_{ii} F^{-1}_{ii}. \qquad (603)
$$

Note that r_{ij} does not depend on amplitudes of bispectra, but shapes. d_i is defined so as $d_i = 1$ for zero degradation, while $d_i > 1$ for degraded $(S/N)_i$.

An order of magnitude estimation of S/N as a function of a certain angular resolution l is possible as follows. Since the number of modes contributing to S/N increases as $\ell^{3/2}$ and $\ell^3 \begin{pmatrix} \ell & \ell & \ell \\ 0 & 0 & 0 \end{pmatrix}^2 \sim 0.36 \times \ell$, we estimate $(S/N)_i \sim (F_{ii})^{1/2}$ as

$$
\left(\frac{S}{N}\right)_i \sim \frac{1}{3\pi} \ell^{3/2} \times \ell^{3/2} \left| \begin{pmatrix} \ell & \ell & \ell \\ 0 & 0 & 0 \end{pmatrix} \right| \times \frac{\ell^3 b^{(i)}_{\ell\ell\ell}}{(\ell^2 C_\ell)^{3/2}} \sim \ell^5 b^{(i)}_{\ell\ell\ell} \times 4 \times 10^{12}, \qquad (604)
$$

where we have used $\ell^2 C_\ell \sim 6 \times 10^{-10}$. A full numerical computation leads to

$$
\left(\frac{S}{N}\right)_{\rm NG} \sim \ell \times 10^{-4} f_{\rm NL}. \qquad (605)
$$

For an experiment like Planck for which the maximum multipole is about 2000 we get that the minimum value of $f_{\rm NL}$ detectable is about $10^4/2000 \sim 5$. How can we estimate analytically the (S/N)? As we are interested in large multipoles, where the (S/N) is higher, it is convenient to make use of the flat-sky approximation and write

$$
a(\boldsymbol{\ell}) = \int {\rm d}^2\boldsymbol{\ell} \frac{\delta T}{T}(\hat{n}) e^{-i\boldsymbol{\ell}\cdot\hat{n}}, \qquad (606)
$$

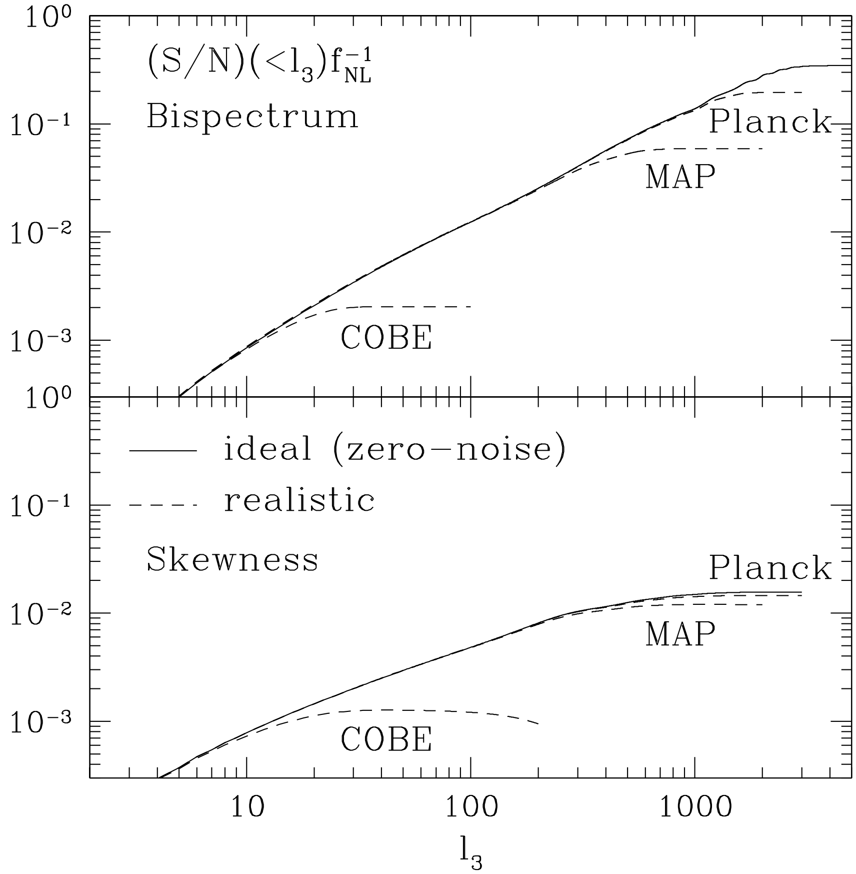

Figure 20: Signal-to-Noise ratio induced by the bispectrum and by the skewness (the bispectrum at three coincidence points) for the various experiments.

where we have decomposed $\hat{n}$ into a part orthogonal and parallel to the line of sight as $\hat{n} \simeq (\boldsymbol{\ell}, 1)$. Indeed, In the flat-sky formalism one chooses a fiducial direction $\hat{z}$ and expands at the lowest order in the angle θ between $\hat{z}$ and $\hat{n}$

$$\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \simeq (\boldsymbol{\ell}, 1)\,, \tag{607}$$

$\boldsymbol{\ell}$ being a two-dimensional vector normal to $\hat{z}$. it is convenient to separate $\mathbf{k}$ as the sum of a two-dimensional vector parallel to the flat sky and a component orthogonal to it, $\mathbf{k} = (\mathbf{k}^{\|}, k^z)$. The multipole is simply the two-dimensional Fourier transform with respect to $\boldsymbol{\ell}$

$$a(\boldsymbol{\ell}) = \int_0^{\tau_0} \mathrm{d}\tau \int \frac{\mathrm{d}^3k}{2\pi} \delta^{(2)}\left(\boldsymbol{\ell} - \mathbf{k}^{\|}(\tau_0 - \tau_{\rm ls})\right) e^{ik^z(\tau_0-\tau_{\rm ls})} S(\mathbf{k}, \tau) = \int \frac{\mathrm{d}k^z}{2\pi} e^{ik^z(\tau_0-\tau_{\rm ls})} S(\ell, k^z, \tau)\,, \tag{608}$$

where

$$S(\ell, k^z, \tau) = \int_0^{\tau_0} \frac{\mathrm{d}\tau}{(\tau_0 - \tau)^2} S(\sqrt{(k^z)^2 + \ell^2/(\tau_0 - \tau)^2}, \tau)\,, \tag{609}$$

is the radiation transfer function defined by the CMB source function $S(k, \tau)$. In this notation, τ_0 and τ_{ls} represent the present-day and the recombination conformal time, respectively and, as we have said, k^z and $\mathbf{k}^{\|}$ are the momentum components perpendicular and parallel respectively to the plane orthogonal to the line-of-sight. The radiation transfer function, as we know, is proportional to the gravitational potential $\Phi(\mathbf{k}')$, where $\mathbf{k}'$ means $\mathbf{k}$ evaluated such that $\mathbf{k}^{\|} = \boldsymbol{\ell}/(\tau_0 - \tau_{\mathrm{ls}})$.

The (S/N) ratio in the flat-sky formalism is

$$\left(\frac{S}{N}\right)^2 = \frac{f_{\mathrm{sky}}}{\pi} \frac{1}{(2\pi)^2} \int \mathrm{d}^2\boldsymbol{\ell}_1 \mathrm{d}^2\boldsymbol{\ell}_2 \mathrm{d}^2\boldsymbol{\ell}_3\, \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3)\, \frac{B^2(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3)}{6\, C(\ell_1)\, C(\ell_2)\, C(\ell_3)}\,, \tag{610}$$

where f_{sky} stands for the portion of the observed sky. The power spectrum in the flat-sky approximation is given by

$$\langle a(\boldsymbol{\ell}_1) a(\boldsymbol{\ell}_2)\rangle = (2\pi)^2 \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2) C(\ell_1)\,, \tag{611}$$

with

$$C(\ell) = \frac{(\tau_0 - \tau_{\mathrm{ls}})^2}{(2\pi)} \int \mathrm{d}k^z |S(\ell, k^z)|^2\,. \tag{612}$$

If we adopt a model with no radiative transfer, that is simply $S(\ell, k^z) = 1/3\Phi(\mathbf{k}')(\tau_0 - \tau_{\mathrm{ls}})^2$, we get

$$\frac{\ell^2 C(\ell)}{2\pi} = \frac{1}{9} \frac{A}{2\pi^2}\,, \tag{613}$$

where we have taken $P_\Phi(k) = A/k^3$. Likewise we can find the bispectrum

$$\begin{aligned} B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) &= \frac{2 f_{\mathrm{NL}} A^2}{3^3 \pi^2} \left(\frac{1}{\ell_1^2 \ell_2^2} + \text{cyclic} \right) \\ &= 6 f_{\mathrm{NL}} \left[C(\ell_1) C(\ell_2) + \text{cyclic} \right]\,. \end{aligned} \tag{614}$$

The (S/N) becomes

$$\left(\frac{S}{N}\right)^2 = \frac{f_{\mathrm{sky}} f_{\mathrm{NL}}^2 A}{6\pi^4} \int \mathrm{d}^2\boldsymbol{\ell}_1 \mathrm{d}^2\boldsymbol{\ell}_2 \mathrm{d}^2\boldsymbol{\ell}_3 \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \ell_1^2 \ell_2^2 \ell_3^2 \left(\frac{1}{\ell_1^2 \ell_2^2} + \text{cyclic} \right)^2\,, \tag{615}$$

and evaluating the above expression we find

$$\left(\frac{S}{N}\right)^2 = \frac{4}{\pi^2} f_{\mathrm{sky}} f_{\mathrm{NL}}^2 A \ell_{\max}^2 \ln \frac{\ell_{\max}}{\ell_{\min}}. \tag{616}$$

The logarithm is typical of scale invariant primordial power spectra. If the primordial perturbations were generated by a Poisson process so each point in space was statistically independent, the logarithm would be absent and the dependence on $\ell_{\max}$ would solely be $\ell_{\max}^2$. Equation (616) can be written in a more physical way by relating it to other observables,

$$\left(\frac{S}{N}\right)^2 = \frac{4}{\pi^2} f_{\mathrm{NL}}^2 A N_{\mathrm{pix}} \ln \frac{\ell_{\max}}{\ell_{\min}}. \tag{617}$$

where $N_{\rm pix} = f_{\rm sky}\ell_{\rm max}^2$ is the number of observed pixels. We thus reproduce the scaling $\left(\frac{S}{N}\right) \propto \ell$ that one can find with an exact numerical calculation.

There remains the question of why physical processes like Silk damping or cancellation due to oscillations during the finite width of the last scattering surface do not cause a strong change in the slope of (S/N) curve at high ℓ. The reason is that there are an equal number of transfer functions in the numerator and denominator of $(S/N)^2$, so there is a sense that the effects of radiative transfer cancel out. Of course the transfer functions are not simple multiplicative factors that can be cancelled, and one has to be careful. We will attempt to explore this in the model by including the effects of Silk damping by introducing an exponential cutoff to mimic the effects of Silk damping on the radiation transfer function, $S(\ell, k^z) = \Phi(\mathbf{k}')(\tau_0 - \tau_{\rm ls})^2 \exp(-k'^2/2k_{\rm D}^2)$.

Repeating the above steps we find the power spectrum can be formally evaluated in terms of Hypergeometric U-functions as

$$C(\ell) = \frac{\sqrt{\pi}A}{2\pi l^2} e^{-\ell^2/\ell_{\rm D}^2} U(1/2, 0, \ell^2/\ell_{\rm D}^2) \,, \tag{618}$$

where $\ell_{\rm D}$ is the Fourier multiple corresponding to the Silk damping scale. We can make an approximation in order to better understand the effects of Silk damping on the CMB power spectrum by cutting off the integral at $k \sim k_{\rm D}$, then

$$C(\ell) = \frac{A}{\pi\ell^2} \frac{e^{-\ell^2/\ell_{\rm D}^2}}{\sqrt{1+\ell^2/\ell_{\rm D}^2}} \,, \tag{619}$$

so when $\ell \ll \ell_{\rm D}$ we recover (apart from the factor 1/9) the no radiative case. Likewise we can evaluate the three-point functions again in order to facilitate the evaluation of this integral assume that the exponentials cutoff the region of integration at $k_1, k_2 \sim k_{\rm D}$.

$$B(\ell_1, \ell_2, \ell_3) = \frac{2 f_{\rm NL} A^2}{\pi^2} e^{-(\ell_1^2+\ell_2^2+\ell_3^2)/2\ell_{\rm D}^2} \left[\frac{1}{\ell_1^2\sqrt{1+\ell_1^2/\ell_{\rm D}^2}} \frac{1}{\ell_2^2\sqrt{1+\ell_2^2/\ell_{\rm D}^2}} + \text{cyclic} \right] . \tag{620}$$

Then using Eq. (619) and Eq. (620) and assuming $\ell \gg \ell_{\rm D}$, the (S/N) becomes

$$\left(\frac{S}{N}\right)^2 = \frac{f_{\rm sky} f_{\rm NL}^2 A \ell_{\rm D}}{6\pi^4} \int {\rm d}^2\boldsymbol{\ell}_1 {\rm d}^2\boldsymbol{\ell}_2 {\rm d}^2\boldsymbol{\ell}_3 \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \frac{(\ell_1^3 + \ell_2^3 + \ell_3^3)^2}{\ell_1^3 \ell_2^3 \ell_3^3} \,. \tag{621}$$

The leading term scales as

$$\left(\frac{S}{N}\right)^2 \propto f_{\rm sky} f_{\rm NL}^2 A \ell_{\rm max}^2 \,. \tag{622}$$

This shows that we can still expect to recover information about $f_{\rm NL}$ on scales where photon diffusion is exponentially damping the transfer functions. In practice, both detector noise, angular resolution and secondary anisotropies will limit the smallest scale that could be used. We see from Fig. (21) that the exact numerical result is well reproduced. We conclude that a primordial NG of the local type can be detected through the CMB by an experiment like Planck up to $f_{\rm NL} = \mathcal{O}(5)$.

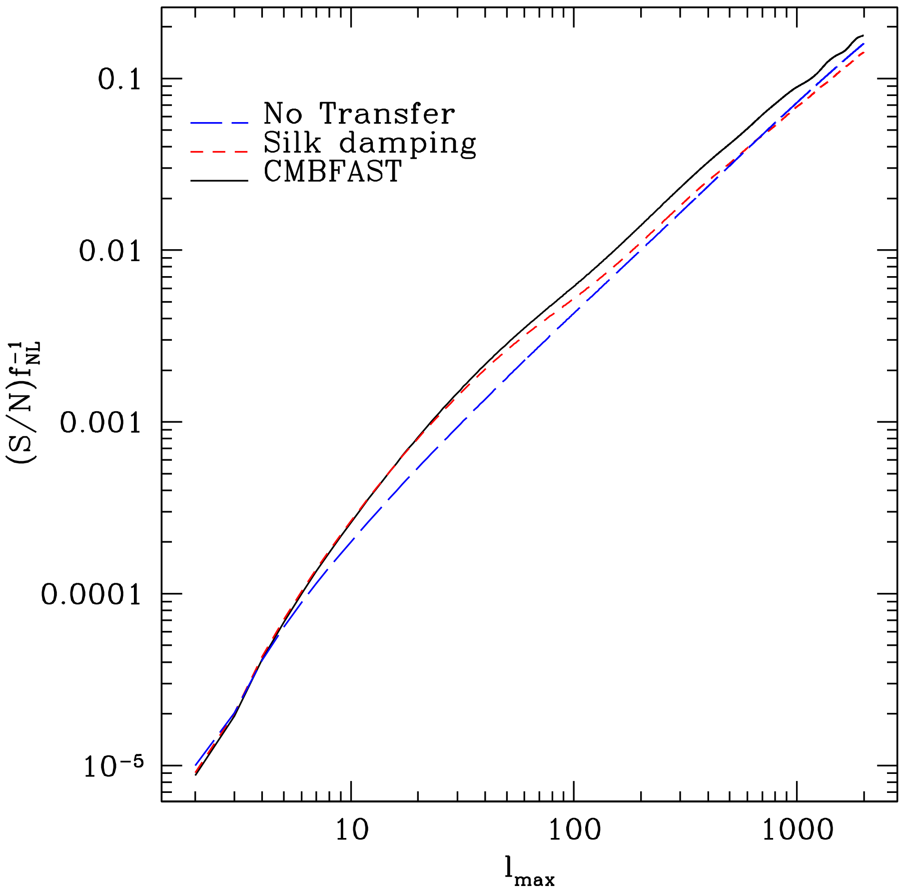

Figure 21: Signal-to-Noise ratio for the no radiative transfer, the Silk damping model and the exact numerical case.

13.7 Non-Gaussianity in the CMB anisotropies at recombination in the squeezed limit

In this subsection we come back to the question of how large is the contamination of NG coming from the inherently nonlinear evolution of the photon-baryon fluid. While the full computation requires the full set of second-order Boltzmann equations, here we show that, as long as we are interested in the squeezed limit of NG, we can indeed perform the computation entirely analytically. Indeed, a transparent computation of the bispectrum in the squeezed limit can be performed through a convenient coordinate rescaling. To understand such a rescaling, it is important to recall what is generally the origin of a squeezed non-Gaussian signal: typically the local-form bispectrum is generated when short-wavelength fluctuations are modulated by long-wavelength fluctuations. In particular we will focus on the temperature anisotropies at recombination when the long wavelength mode is outside the horizon, but observable at the present epoch. Thus, the effect of the long wavelength mode imprinted at recombination can be described simply by a coordinate transformation. In this way we can describe in a simple way the coupling of small scales to large scales that can generally produce the local form bispectrum. A similar cross-talk between large and small scales gives rise to the ISW-lensing cross-correlation bispectrum.

Our starting metric is

$$\mathrm{d}s^2 = a^2(\tau)\left[-e^{2\Phi}\mathrm{d}\tau^2 + e^{-2\Psi}\mathrm{d}\mathbf{x}^2\right] , \tag{623}$$

where $a(\tau)$ is the scale factor as a function of the conformal time τ, and we have neglected vector and tensor perturbations. Instead of solving the complicated network of second-order Boltzmann equations for the CMB temperature anisotropies, we use the following trick. As the wavenumber $k_1 \lesssim k_{\rm eq}$ corresponds to a perturbation which is almost larger than the horizon at recombination and the evolution in time of the corresponding gravitational potential is very moderate (one can easily check, for instance, that $\Phi^{(1)}_{\mathbf{k}_1}(\tau)$ changes its magnitude by at most 10% during the radiation epoch for $k_1 = k_{\rm eq}$), we can absorb the large-scale perturbation with wavelength $\sim k_1^{-1}$ in the metric by redefining the time and the space coordinates as follows. Let us indicate with Φ_ℓ and Ψ_ℓ the parts of the gravitational potentials that receive contributions only from the large-scale modes $k_1 \lesssim k_{\rm eq}$. If the scale factor is a power law $a(\tau) \propto \tau^\alpha$ ($\alpha = 1$ and $\alpha = 2$ for the period of radiation and matter domination, respectively), we can perform the redefinitions

$$a^2(\tau)e^{2\Phi_\ell}\mathrm{d}\tau^2 = \tau^{2\alpha}e^{2\Phi_\ell}\mathrm{d}\tau^2 = \overline{\tau}^{2\alpha}\mathrm{d}\overline{\tau}^2 = a^2(\overline{\tau})\mathrm{d}\overline{\tau}^2 \Rightarrow \overline{\tau} = e^{\frac{1}{1+\alpha}\Phi_\ell}\tau \, , \tag{624}$$

and

$$a^2(\tau)e^{-2\Psi_\ell}\mathrm{d}\mathbf{x}^2 = \tau^{2\alpha}e^{-2\Psi_\ell}\mathrm{d}\mathbf{x}^2 = \overline{\tau}^{2\alpha}e^{\frac{-2\alpha}{1+\alpha}\Phi_\ell}e^{-2\Psi_\ell}\mathrm{d}\mathbf{x}^2 = a^2(\overline{\tau})\mathrm{d}\overline{\mathbf{x}}^2 \Rightarrow \overline{\mathbf{x}} = e^{\frac{-\alpha}{1+\alpha}\Phi_\ell}e^{-\Psi_\ell}\mathbf{x} \, . \tag{625}$$

In particular, the combination

$$\overline{k}\overline{\tau} = e^{\Phi_\ell + \Psi_\ell}\, k\tau \, , \tag{626}$$

where $\overline{k}$ and k are the wavenumbers in the two coordinate systems. Obviously, if one wishes to account for the fact that at recombination the universe is not fully matter-dominated, one should perform a more involved coordinate transformation which will eventually depend also on the parameter R.

We make the simplifying assumption that $\tau_{\rm ls} \gg \tau_{\rm eq}$ in such a way that the coordinate transformations (624) and (625) can be performed in a matter-dominated period, that is we take $\alpha = 2$ and

$$\mathbf{x} \to e^{-5\Phi_\ell/3}\,\mathbf{x} \, , \quad \mathbf{k} \to e^{5\Phi_\ell/3}\,\mathbf{k} \, , \quad \tau_{\rm ls} \to e^{\Phi_\ell/3}\,\tau_{\rm ls} \, , \tag{627}$$

is the transformation for modes which were outside the horizon at recombination, but are subHubble at the time of observation. Notice that the rescaling (627) changes also the gravitational potential

$$\Phi^{(1)}_{\mathbf{k}} \to e^{5\Phi_\ell}\, \Phi^{(1)}_{e^{5\Phi_\ell/3}\mathbf{k}}\,. \tag{628}$$

Notice that for a long wavelength modulating mode that is outside the cosmological horizon today the power spectrum does not change. This is because in this case both $(\tau_0 - \tau_{\rm ls})$ and k must be rescaled, so that the integral (612) goes like $1/(\tau_0 - \tau_{\rm ls})^2 k^2$ and does not feel the coordinate rescaling, as one would expect for such a modulating mode. To compute the bispectrum, we go to the squeezed limit $\ell_1 \ll \ell_2, \ell_3$ (or $k_1 \ll k_2, k_3$). In this case $\Phi^{(1)}_{\mathbf{k}_1}$ acts as a background for the other two modes. One can therefore compute the three-point function in a two-step process: first compute the two-point function in the background of $\Phi^{(1)}_{\mathbf{k}_1}$ and then the result from the correlation induced by the background field. Using the Sachs-Wolfe limit for the multipole ℓ_1, this procedure leads to (notice that the coordinate rescaling, operating the the recombination point, is not relevant for the time as $\tau_0 \gg \tau_{\rm ls}$)

$$\langle a(\boldsymbol{\ell}_2)a(\boldsymbol{\ell}_3)\rangle_{\Phi^{(1)}_{\mathbf{k}_1}} = \langle a(\boldsymbol{\ell}_2)a(\boldsymbol{\ell}_3)\rangle_0 + 5\, a(\boldsymbol{\ell}_2+\boldsymbol{\ell}_3)C(\ell_2)\frac{{\rm d}\ln\left[\ell_2^2 C(\ell_2)\right]}{{\rm d}\ln \ell_2}\,. \tag{629}$$

In fact, in general, think of a function $F(\mathbf{x}_2, \mathbf{x}_3)$ that depends on the short distance $(\mathbf{x}_2 - \mathbf{x}_3)$, but also modulated by a long wavelength mode background function $F_{\rm B}(|\mathbf{x}_2 + \mathbf{x}_3|/2)$. One can expand

$$F(\mathbf{x}_2, \mathbf{x}_3) = F_0(|\mathbf{x}_2 - \mathbf{x}_3|) + F_{\rm B}(|\mathbf{x}_2 + \mathbf{x}_3|/2)\, \frac{\rm d}{{\rm d}F_{\rm B}} F_0(|\mathbf{x}_2 - \mathbf{x}_3|)\bigg|_0 + \cdots\,. \tag{630}$$

If the long wavelength background modulates the amplitude of the two point function is equivalent to a rescaling of the spatial coordinates, one can trade the derivative with respect to $F_{\rm B}$ for a derivative with respect to the log-distance between the points

$$F(\mathbf{x}_2, \mathbf{x}_3) \simeq F_0(|\mathbf{x}_2 - \mathbf{x}_3|) + \int \frac{{\rm d}^3 k}{(2\pi)^3} F_{\rm B}(k) e^{i\mathbf{k}\cdot(\mathbf{x}_1+\mathbf{x}_2)/2}\, \frac{\rm d}{{\rm d}\ln|\mathbf{x}_2 - \mathbf{x}_3|} F_0(|\mathbf{x}_2 - \mathbf{x}_3|)\bigg|_0\,. \tag{631}$$

If we now integrate over $\mathbf{x}_1$ and $\mathbf{x}_2$, or better over $(\mathbf{x}_1 + \mathbf{x}_2)/2$ and $(\mathbf{x}_1 - \mathbf{x}_2)/2 = \mathbf{x}_{\rm S}$, the second piece becomes proportional to (being $\mathbf{k}_{\rm S} = (\mathbf{k}_2 - \mathbf{k}_3)/2$)

$$\begin{aligned}
\int {\rm d}^3\mathbf{x}_{\rm S}\, \frac{\rm d}{{\rm d}\ln x_{\rm S}} F_0(x_{\rm S})\, e^{-i\mathbf{k}_{\rm S}\cdot\mathbf{x}_{\rm S}} &\sim \int {\rm d}\ln x_{\rm S}\, x_{\rm S}^3\, \frac{\rm d}{{\rm d}\ln x_{\rm S}} F_0(x_{\rm S})\, e^{-i\mathbf{k}_{\rm S}\cdot\mathbf{x}_{\rm S}} \\
&\sim -\int {\rm d}\ln x_{\rm S}\, F_0(x_{\rm S})\, \frac{\rm d}{{\rm d}\ln x_{\rm S}} \left(x_{\rm S}^3\, e^{-i\mathbf{k}_{\rm S}\cdot\mathbf{x}_{\rm S}}\right) \\
&\sim -\int {\rm d}\ln x_{\rm S}\, F_0(x_{\rm S})\, x_{\rm S}^3\, \frac{\rm d}{{\rm d}\ln k_{\rm S}}\, e^{-i\mathbf{k}_{\rm S}\cdot\mathbf{x}_{\rm S}} \\
&- 3\int {\rm d}\ln x_{\rm S}\, F_0(x_{\rm S})\, x_{\rm S}^3\, e^{-i\mathbf{k}_{\rm S}\cdot\mathbf{x}_{\rm S}} \\
&= -\frac{1}{k_{\rm S}^3}\, \frac{\rm d}{{\rm d}\ln k_{\rm S}} \left(k_{\rm S}^3 F_0(k_{\rm S})\right)\,.
\end{aligned} \tag{632}$$

Now, repeating these steps for the two-dimensional problem, being the the rescaling given by $\mathbf{x} \to e^{-5\Phi_\ell/3}\,\mathbf{x}$, and remembering that $a(\boldsymbol{\ell}_2+\boldsymbol{\ell}_3)=a(-\boldsymbol{\ell}_1)$ has a coefficient 1/3 for the SW effect, we get the expression (629). The corresponding bispectrum therefore reads

$$B(\ell_1,\ell_2,\ell_3)=\Big\langle a(\boldsymbol{\ell}_1)\langle a(\boldsymbol{\ell}_2)a(\boldsymbol{\ell}_3)\rangle\Big\rangle=(2\pi)^2\,\delta^{(2)}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3)\,5\;C(\ell_1)C(\ell_2)\,\frac{\mathrm{d}\ln\left[\ell_2^2C(\ell_2)\right]}{\mathrm{d}\ln\ell_2}\,. \tag{633}$$

In multipole space the bispectrum induced by a local primordial NG in the squeezed limit is given by

$$B_{\rm loc}(\ell_1,\ell_2,\ell_3)=6\,f_{\rm NL}^{\rm loc}\left[C(\ell_1)C(\ell_2)+{\rm cycl.}\right]\,. \tag{634}$$

Since at large multipoles the exponential of the transfer function allows to cut off the integral for $k\simeq k_*\sim 750$

$$C(\ell)\simeq 9\frac{A}{\pi}\frac{\ell_*}{\ell^3}\,e^{-(\ell/\ell_*)^{1.2}}\,, \tag{635}$$

which holds for $\ell\gg\ell_*$, we see that, roughly speaking, the effective $f_{\rm NL}^{\rm loc}$ coming from the second-order effects at recombination in the squeezed limit is

$$f_{\rm NL}^{\rm rec}\simeq\frac{5}{6}\frac{\mathrm{d}\ln\left[\ell_2^2C(\ell_2)\right]}{\mathrm{d}\ln\ell_2}=\mathcal{O}(1)\,, \tag{636}$$

which confirms our expectation that second-order effects lead to a contamination of order unity in $f_{\rm NL}^{\rm loc}$.

Part VI

Matter perturbations

We now discuss how matter perturbations evolve since inflation to the present. We have shown that inflation generates the seeds for the growth of perturbations with a power spectrum which is nearly scale invariant on superHubble scales. After perturbations become sub-Hubble again, their evolution can be described by simple Newtonian analysis and once can show that, at the linear level, the CDM perturbations grow with a growth function function $D_+(a)$ which reduces to the scale factor a in a MD epoch

$$\delta_{\rm m}(\mathbf{k},\tau>\tau_{\rm eq})=-\left(2+\frac{k^2\tau^2}{6}\right)D_1(\mathbf{k}), \tag{637}$$

This equation, at least at the linear level, is the essence of the phenomenon called gravitational instability responsible for structure formation. From this equation we see immediately that perturbation theory however is limited to ranges of wavenumber and times such that $k\tau\lesssim 1$. Perturbation theory breaks down either when we go to large times and/or when we probe the system at short wavelength. This is hardly a surprise. Perturbations grow because gravity is is in action and gravity, being a derivative theory, leads to ultraviolet divergences at large momenta. What happens when we go to higher order in perturbation theory? To study (or better just to introduce it) the problem we adopt again the metric in the Poisson gauge and derive a generic equation for the gravitational potential which is an alternative to the standard continuity and Euler equations we have found in the previous section.

The write the metric in the following form

$$\mathrm{d}s^2 = -a^2(\tau)\left[-(1+2\Phi)\,\mathrm{d}\tau^2 + 2\Omega_{\rm I}\,\mathrm{d}\tau\mathrm{d}x^i + ((1-2\Psi)\delta_{ij} + \chi_{ij})\,\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{638}$$

Here Φ and Ψ are the gravitational potential as usual, $\Omega_{\rm I}$ are the vector perturbations and χ_{ij} the tensor ones. As we are interested in subHubble scales, we will retain first-order terms in Ψ and Φ and only the would-be second-order terms with two space derivatives. In particular, the difference $(\Psi-\Phi)$ is second-order in Ψ, so we need to keep $\boldsymbol{\nabla}(\Psi-\Phi)$.

We do not assume that the vector perturbations vanish $\Omega_{\rm I}=0$, rather use the fact that the divergence of $\Omega_{\rm I}$ vanishes, $\boldsymbol{\nabla}^i\Omega_{\rm I}=0$. Similarly with the tensor modes χ_{ij} which are transverse and traceless. This leads us to use the traceless-longitudinal projection operator

$$\mathcal{P}^{ij} = \delta^{ij} - 3\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2} \tag{639}$$

to determine Φ in terms of Ψ from the Einstein equations. This procedure should be compared to the lowest order procedure in which one simply uses the $i\neq j$ Einstein equation to determine that $\Phi=\Psi$. We also use only the longitudinal part of the $(i0)$ equations so the contribution of the vector perturbations drops.

We take the energy-momentum tensor to be that of a cold pressureless fluid of density ρ and three-velocity vector $\mathbf{v}$,

$$T_{\mu\nu} = \rho\, v_\mu v_\nu\,. \tag{640}$$

Finally, and for simplicity, we choose units for which $8\pi G_{\rm N}=1$. With these assumptions, the relevant Einstein equations are

$$(00):\qquad 3\,\mathcal{H}^2 + 2\boldsymbol{\nabla}^2\Psi = \rho\, v_0^2\,, \tag{641}$$

$$(i0):\qquad \boldsymbol{\nabla}_i\Psi' + \mathcal{H}\boldsymbol{\nabla}_i\Psi = \frac{1}{2}\rho\, v_0\, v_i\,, \tag{642}$$

$$\begin{aligned}(ij):\qquad &\left(2\Psi'' + 6\,\mathcal{H}\Psi' - 2\left(\mathcal{H}^2 - 2\frac{a''}{a}\right)\Psi + (\boldsymbol{\nabla}\Psi)^2 - \boldsymbol{\nabla}^2(\Psi-\Phi)\right)\delta_{ij}\\ &+\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j(\Psi-\Phi) - 2\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi = \rho\, v_i v_j\,.\end{aligned} \tag{643}$$

The density contrast is as usual expressed in terms of the gravitational potential through the Poisson equation

$$\delta = \frac{2\boldsymbol{\nabla}^2\Psi}{3\,\mathcal{H}^2}. \tag{644}$$

The traceless-longitudinal part of the (ij) equation is given by

$$\mathcal{P}^{ij}\left(\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j(\Psi-\Phi)-2\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\right)=\mathcal{P}^{ij}\left(\rho\, v_i v_j\right)\,. \tag{645}$$

We may evaluate explicitly some of the terms (and divide the whole equation by -2),

$$\boldsymbol{\nabla}^2(\Psi-\Phi)+(\boldsymbol{\nabla}\Psi)^2-3\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\right)=-\frac{1}{2}\rho\mathbf{v}^2+\frac{3}{2}\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\rho v_i v_j\right)\,. \tag{646}$$

We will use Eqs. (646) and (642) to express $\boldsymbol{\nabla}^2(\Psi-\Phi)$ in terms of Ψ only which will allow us to derive an equation for Ψ that does not involve Φ. Despite appearances, Eq. (646) is completely local. In fact, all terms have at least two space derivatives acting on the fields.

Let us now look at 1/6 of the trace of Eq. (643)

$$\Psi''+3\,\mathcal{H}\Psi'-\left(\mathcal{H}^2-2\frac{a''}{a}\right)\Psi+\frac{1}{6}(\boldsymbol{\nabla}\Psi)^2-\frac{1}{3}\boldsymbol{\nabla}^2\left(\Psi-\Phi\right)=\frac{1}{6}\rho\mathbf{v}^2\,. \tag{647}$$

The next step is to substitute Eq. (646) into Eq. (647) and obtain

$$\Psi''+3\,\mathcal{H}\Psi'-\left(\mathcal{H}^2-2\frac{a''}{a}\right)\Psi+\frac{1}{2}(\boldsymbol{\nabla}\Psi)^2-\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\right)=\frac{1}{2}\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\rho\, v_i v_j\right)\,. \tag{648}$$

Finally, we can use the $(0i)$ equation (642) to integrate out the three-velocity $U_{\mathrm{m,i}}$ from Eq. (648)

$$\begin{aligned}\Psi''+3\,\mathcal{H}\Psi'-\left(\mathcal{H}^2-2\frac{a''}{a}\right)\Psi+\frac{1}{2}(\boldsymbol{\nabla}\Psi)^2-\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\right)&\\ =2\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\frac{\left(\boldsymbol{\nabla}_i\Psi'+\mathcal{H}\boldsymbol{\nabla}_i\Psi\right)\left(\boldsymbol{\nabla}_j\Psi'+\mathcal{H}\boldsymbol{\nabla}_j\Psi\right)}{\rho v_0^2}\right)&\,,\end{aligned} \tag{649}$$

and the (00) equation (641) to integrate out the energy density ρ

$$\Psi''+3\,\mathcal{H}\Psi'-\left(\mathcal{H}^2-2\frac{a''}{a}\right)\Psi+\frac{1}{2}(\boldsymbol{\nabla}\Psi)^2-\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\right)=\frac{2}{3\,\mathcal{H}^2}\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\left(\frac{\left(\boldsymbol{\nabla}_i\Psi'+\mathcal{H}\boldsymbol{\nabla}_i\Psi\right)\left(\boldsymbol{\nabla}_j\Psi'+\mathcal{H}\boldsymbol{\nabla}_j\Psi\right)}{1+2\boldsymbol{\nabla}^2\Psi/3\,\mathcal{H}^2}\right)\,. \tag{650}$$

We now verify that the perturbative solution of Eq. (650) reproduces the known perturbative solutions for Ψ, the density contrast δ and the velocity $\mathbf{u}$. First, we linearize Eq. (650) and obtain the standard equation

$$\Psi''+3\,\mathcal{H}\Psi'-2\left(\mathcal{H}^2-2\frac{a''}{a}\right)\Psi=0\,, \tag{651}$$

whose solution in matter domination does not evolve with time

$$\Psi(\boldsymbol{x},\tau)=\Psi_{\mathrm{L}}(\boldsymbol{x})\,. \tag{652}$$

For future use we recall that the resulting solution for $\mathbf{u}$ is

$$\mathbf{u}_{\mathrm{L}}=\frac{2\boldsymbol{\nabla}\Psi_{\mathrm{L}}}{3\,\mathcal{H}}=\frac{\tau}{3}\boldsymbol{\nabla}\Psi_{\mathrm{L}}\,, \tag{653}$$

where in the second equality we have used the explicit form of $\mathcal{H}$ for matter domination, $\mathcal{H} = 2/\tau$. The corresponding solution for δ can be obtained from its relation to Ψ, Eq. (644), so

$$\delta_{\rm L} = \frac{1}{6}\tau^2 \boldsymbol{\nabla}^2 \Psi_{\rm L} = \frac{1}{6} a(\tau) \boldsymbol{\nabla}^2 \Psi_{\rm L} \,. \tag{654}$$

It agrees of course with what obtained in the previous sections once we have dropped the initial condition (which is subleading at larger times). Next we solve for Ψ to second-order in perturbation theory. The idea is to substitute in the second-order terms the linear solution. Therefore, we take the non-perturbative Eq. (650), expand it to second-order in Ψ and take only terms that do not vanish on the constant leading order solution (652),

$$\Psi'' + 3\,\mathcal{H}\Psi' - \Psi\left(\mathcal{H}^2 - 2\frac{a''}{a}\right) + \frac{1}{2}(\boldsymbol{\nabla}\Psi)^2 - \frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\Big(\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\Big) = \frac{2}{3}\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\Big(\boldsymbol{\nabla}_i\Psi\boldsymbol{\nabla}_j\Psi\Big). \tag{655}$$

For matter domination it reduces to

$$\Psi'' + 3\,\mathcal{H}\Psi' = \frac{5}{3}\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\Big(\boldsymbol{\nabla}_i\Psi_{\rm L}\boldsymbol{\nabla}_j\Psi_{\rm L}\Big) - \frac{1}{2}(\boldsymbol{\nabla}\Psi_{\rm L})^2\,, \tag{656}$$

whose solution is

$$\Psi_2 = \frac{1}{14}\tau^2\left[\frac{5}{3}\frac{\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j}{\boldsymbol{\nabla}^2}\Big(\boldsymbol{\nabla}_i\Psi_{\rm L}\boldsymbol{\nabla}_j\Psi_{\rm L}\Big) - \frac{1}{2}(\boldsymbol{\nabla}\Psi_{\rm L})^2\right]. \tag{657}$$

To evaluate Ψ_2 in an explicit way it is easier to calculate $\boldsymbol{\nabla}^2\Psi_2$,

$$\boldsymbol{\nabla}^2\Psi_2 = \frac{1}{14}\tau^2\left(\frac{5}{3}\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j\left(\boldsymbol{\nabla}_i\Psi_{\rm L}\boldsymbol{\nabla}_j\Psi_{\rm L}\right) - \frac{1}{2}\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\Psi_{\rm L})^2\right)\right). \tag{658}$$

Thus, we need

$$\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j\left(\boldsymbol{\nabla}_i\Psi_{\rm L}\boldsymbol{\nabla}_j\Psi_{\rm L}\right) = (\boldsymbol{\nabla}^2\Psi_{\rm L})^2 + 2\boldsymbol{\nabla}^i\Psi_{\rm L}\boldsymbol{\nabla}^2\boldsymbol{\nabla}_i\Psi_{\rm L} + \boldsymbol{\nabla}^i\boldsymbol{\nabla}^j\Psi_{\rm L}\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j\Psi_{\rm L} \tag{659}$$

and

$$\boldsymbol{\nabla}^i\boldsymbol{\nabla}_i\left(\boldsymbol{\nabla}^j\Psi_{\rm L}\boldsymbol{\nabla}_j\Psi_{\rm L}\right) = 2\boldsymbol{\nabla}^i\Psi_{\rm L}\boldsymbol{\nabla}^2\boldsymbol{\nabla}_i\Psi_{\rm L} + 2\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j\Psi_{\rm L}\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j\Psi_{\rm L}\,. \tag{660}$$

Substituting into Eq. (658) we find

$$\boldsymbol{\nabla}^2\Psi_2 = \frac{1}{14}\tau^2\left(\frac{5}{3}(\boldsymbol{\nabla}^2\Psi_{\rm L})^2 + \frac{7}{3}\boldsymbol{\nabla}^i\Psi_{\rm L}\boldsymbol{\nabla}^2\boldsymbol{\nabla}_i\Psi_{\rm L} + \frac{2}{3}\boldsymbol{\nabla}^i\boldsymbol{\nabla}^j\Psi_{\rm L}\boldsymbol{\nabla}_i\boldsymbol{\nabla}_j\Psi_{\rm L}\right) \tag{661}$$

In momentum space it gives

$$\begin{aligned}\delta_2(\mathbf{k}_3,\tau) &= -\frac{1}{6}k_3^2\tau^2\Psi_2(\mathbf{k}_3,\tau)\\ &= \int\frac{{\rm d}^3k_1}{(2\pi)^3}\int\frac{{\rm d}^3k_2}{(2\pi)^3}\left[\frac{5}{7} + \frac{1}{2}(\mathbf{k}_1\cdot\mathbf{k}_2)\frac{k_1^2+k_2^2}{k_1^2k_2^2} + \frac{2}{7}\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{k_1^2k_2^2}\right]\\ &\times\ \delta^{(3)}\left(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3\right)\,\delta_{\rm L}(\mathbf{k}_1,\tau)\delta_{\rm L}(\mathbf{k}_2,\tau)\,,\end{aligned} \tag{662}$$

which reproduces the standard second-order kernel for the density contrast. Using Eqs. (642) and (661), it is also straightforward to recover the kernel for the (divergence of the) velocity $\theta = \boldsymbol{\nabla} \cdot \mathbf{v}$. One expands Eq. (642) to second-order, using the time dependence of Ψ_2

$$4\,\mathcal{H}\nabla\Psi_2 = \overline{\rho}\,(\mathbf{v}_2 + \delta_{\rm L}\mathbf{v}_{\rm L})\,. \tag{663}$$

Using the lowest order relations, the equation for the divergence of the velocity at second-order becomes

$$\frac{\boldsymbol{\nabla}\cdot\mathbf{v}}{\mathcal{H}} = 2\,\delta_2 - (\delta_{\rm L})^2 - \boldsymbol{\nabla}\delta_{\rm L}\cdot\frac{\boldsymbol{\nabla}\delta_{\rm L}}{\boldsymbol{\nabla}^2}. \tag{664}$$

In momentum space, using the explicit solution (661) the final result is,

$$\begin{aligned} -\frac{\theta_2(\mathbf{k}_3,\tau)}{\mathcal{H}} &= \int\frac{{\rm d}^3k_1}{(2\pi)^3}\int\frac{{\rm d}^3k_2}{(2\pi)^3}\left[\frac{3}{7}+\frac{1}{2}(\mathbf{k}_1\cdot\mathbf{k}_2)\frac{k_1^2+k_2^2}{k_1^2k_2^2}+\frac{4}{7}\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{k_1^2k_2^2}\right] \\ &\times\;\; \delta^{(3)}\left(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3\right)\delta_{\rm L}(\mathbf{k}_1,\tau)\,\delta_{\rm L}(\mathbf{k}_2,\tau)\,, \end{aligned} \tag{665}$$

which agrees with the standard Newtonian kernel. The expressions (664) and (665) render explicit what we have said at the beginning of this section. Once we go to higher orders in perturbation theory the series does not converge. Indeed, already at the second order we see that formally $\delta_2 \sim (\delta_{\rm L})^2 \sim (\mathbf{k}\tau)^4$. In general one has

$$\delta_{n-{\rm th\,order}} \sim (\mathbf{k}\tau)^{2n}, \tag{666}$$

where the $\mathbf{k}$ is a short-hand notation for the kernel in momentum space. This leads clearly to a breakdown of perturbation theory. Again, this is not a surprise, after all the structures we see today in the universe have $\delta \gg 1$, they are highly non-linear.

At what scales one expects perturbation theory to break down? Let us choose to work at the present time. The power spectrum has dimensions of volume and so a quantity that lends itself more easily to direct interpretation is the dimensionless combination

$$\Delta^2_{\rm m}(k) \equiv \frac{k^3P_{\rm m}(k)}{2\pi^2}. \tag{667}$$

In the standard CDM scenario $\Delta^2_{\rm m}(k)$ increases with wavenumber (at least until some exceedingly small scale determined by the physics of the production of the CDM in the early universe), but we observe the density field smoothed with some resolution. Therefore, a quantity of physical interest is the density field smoothed on a particular scale R,

$$\delta_{\rm m}(\mathbf{x},R) \equiv \int {\rm d}^3x'\; W(|\mathbf{x}'-\mathbf{x}|,R)\delta_{\rm m}(\mathbf{x}') \tag{668}$$

The function $W(\mathbf{x},R)$ is the window function that weights the density field in a manner that is relevant for the particular application. According to the convention used in Eq. (668), the window function (sometimes called filter function) has units of inverse volume by dimensional arguments. The Fourier transform of the smoothed field is

$$\delta_{\rm m}(\mathbf{k}, R) \equiv W(\mathbf{k}, R)\delta_{\rm m}(\mathbf{k}), \tag{669}$$

where $W(\mathbf{k}, R)$ is the Fourier transform of the window function. Once can chose for instance the tophat in Fourier space as

$$W(\mathbf{k}, R) = \begin{cases} 1 & (k \leq R^{-1}), \\ 0 & (k > R^{-1}), \end{cases} \tag{670}$$

and is

$$W(\mathbf{x}, R) = \frac{1}{2\pi^2 R^3} \frac{(\sin(xR^{-1}) - xR^{-1}\cos(xR^{-1}))}{(xR^{-1})^3} \tag{671}$$

in real space. A disadvantage of this window is that it does not have a well-defined volume. Therefore the associated mass is simply defined as $M = 4\pi\bar{\rho}_{\rm m}R^3/3$. The density fluctuation field is assumed to be a Gaussian random variable so the smoothed density fluctuation field $\delta_{\rm m}(\mathbf{x}, R)$ is then a Gaussian random variable as well because it represents a sum of Gaussian random variables. The variance of $\delta_{\rm m}(\mathbf{x}, R)$ is

$$\boxed{\sigma^2_{\rm m}(R) = \left\langle \delta^2_{\rm m}(\mathbf{x}, R) \right\rangle = \int {\rm d}\ln k \; \Delta^2_{\rm m}(k)|W(k, R)|^2.} \tag{672}$$

From Fig. 22 we learn two basics things about the CDM paradigm. First, the variance of the density contrasted smoothed over a radius R becomes of the order of unity when $R = \mathcal{O}(10)\, h^{-1}$ Mpc. This means that perturbation theory breaks down when $k \sim 1/R = \mathcal{O}(10^{-1})\, h/$Mpc. Historically, one set the amplitude of the perturbations by setting σ_8, that is the value of the variance at $8\, h^{-1}$ Mpc. Secondly, the fact that $\Delta_{\rm m}(k)$ has more power at large wavenumber means that the first scales to go no-linear are the small ones (as repeatedly said), that is in the CDM paradigm the first objects to form are the ones on small scales. Largers structures may form because of merging of smaller structures. This is the so-called hierarchical paradigm: big DM halos form from the merging of small DM halos. In the next section we discuss a classical example on how to deal with the nonlinearities of the DM perturbations.

14 Spherical collapse

One of the simplest and best studied models of nonlinear gravitational instability is the spherical model. In this model one ignores the tidal effects of neighbouring density perturbations upon the evolution of an isolated, homogeneous, spherical density perturbation. To justify this we can appeal to Birkhoff's theorem in General Relativity, or Gauss's law in Newtonian Gravity. Under these simplifying assumptions an exact analytical treatment is possible.

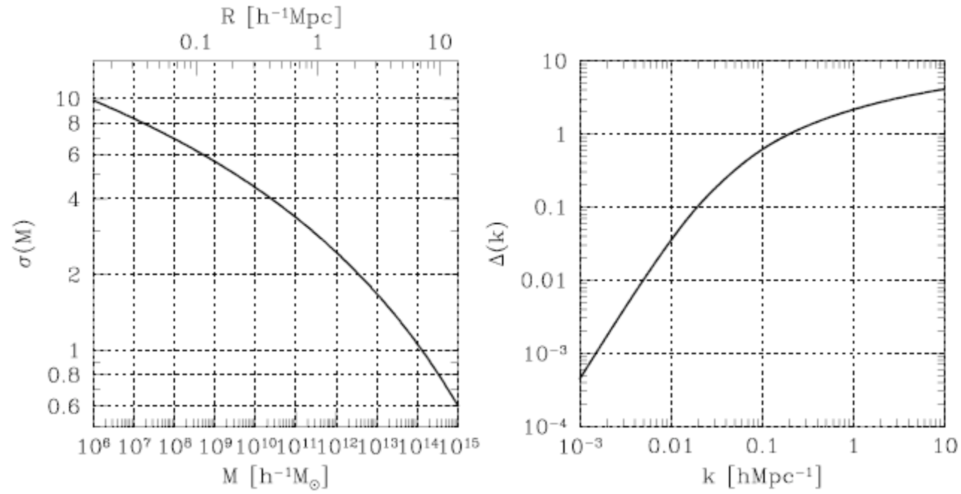

Figure 22: Power spectra in the standard ΛCDM cosmology with $\Omega_{\rm m} = 1 - \Omega_\Lambda = 0.3$, $h = 0.7$, $\sigma_8 = 0.93$, and $\Omega_{\rm b} h^2 = 0.022$. The left panel shows the mass variance smoothed with a real space tophat window as a function of the smoothing mass or smoothing radius. The right panel shows the rms density fluctuation per logarithmic interval of wavenumber as a function of wavenumber.

In order to understand the dynamics of non–linear spherical collapse, consider a spherical density perturbation expanding in the background of a homogeneous and isotropic background universe. The density of the fluctuation is characterised by $\Omega'_{\rm m}$ whereas that of the background universe by $\Omega_{\rm m}$ ($\Omega'_{\rm m} > \Omega_{\rm m}$ will correspond to an overdensity and $\Omega'_{\rm m} < \Omega_{\rm m}$ to an underdensity). The subsequent fate of the spherical density perturbation will depend crucially upon the value of $\Omega'_{\rm m}$. For $\Omega'_{\rm m} > 1$ the perturbation will behave just like a part of a closed FRW universe and will therefore expand to a maximum radius, turn around at a time $t_{\rm ta}$, and thereafter collapse to a point at $t_{\rm coll} \simeq 2t_{\rm ta}$. A spherical density perturbation with $\Omega'_{\rm m} < 1$ on the other hand, will mimic an open universe and never recollapse (if $\Omega'_{\rm m} < \Omega_{\rm m}$ then such an underdensity will correspond to a void). In an idealised cosmological scenario spherical overdensities might be thought of as progenitors of clusters of galaxies, whereas underdensities would correspond to voids.

In order to treat the collapse of a spherical overdensity quantitatively let us consider a spherical shell of radius R with an initial overdensity $\delta_{\rm I}$ and a constant mass $M = 4\pi R^3 \bar\rho_{\rm m}(1+\delta_{\rm I})/3$, where $\bar\rho_{\rm m}$ is the density of the backround universe. Conservation of energy guarantees

$$\frac{1}{2}\dot R^2 - \frac{G_{\rm N} M}{R} = E = \text{ const.} \tag{673}$$

At early times the expansion of the shell is virtually indistinguishable from that of the rest of the universe so that $\dot{R}_{\rm m,i} = H_{\rm I} R_{\rm m,i}$, $R_{\rm m,i}$ being the radius of the shell and $H_{\rm I}$ is the Hubble parameter at an initial time $t = t_{\rm I}$. The kinetic energy of the shell is therefore $K_{\rm m,i} = \frac{1}{2} H_{\rm I}^2 R_{\rm m,i}^2$ and its potential energy is $U_{\rm m,i} = -GM/R_{\rm m,i} = -K_{\rm m,i}\Omega_{\rm I}(1+\delta_{\rm I})$ where $\Omega_{\rm I}$ is the density parameter at $t_{\rm I}$: $3H_{\rm I}^2\Omega_{\rm I}/2 = 4\pi G\bar{\rho}_{\rm m,i}$. As a result we obtain

$$E = K_{\rm m,i} + U_{\rm m,i} = K_{\rm m,i}\Omega_{\rm I}[\Omega_{\rm I}^{-1} - (1+\delta_{\rm I})]. \tag{674}$$

The requirement for collapse $E < 0$ leads to the condition $1 + \delta_{\rm I} > \Omega_{\rm I}^{-1}$. Substituting $\Omega_{\rm I} \equiv \Omega_{\rm m}(z) = \Omega_{\rm m}(1+z)/(1+\Omega_{\rm m} z)$, $\delta_{\rm I} \equiv \delta_{\rm m}(z)$ we get

$$\delta_{\rm m}(z) > \frac{1-\Omega_{\rm m}}{\Omega_{\rm m}(1+z)} \tag{675}$$

as a precondition for collapse to occur. Equation (675) indicates that in flat or closed cosmological models an infinitesimal initial density perturbation is sufficient to give rise to collapsed objects. In open models on the other hand $\delta_{\rm m}(z)$ must exceed a critical positive value in order for collapse to occur.

It is relatively straightforward to relate the maximum expansion radius reached by an overdensity at turnaround $R_{\rm ta}$ to its "seed" values $R_{\rm m,i}$, and $\delta_{\rm I}$ (equivalently $R_{\rm m}(z)$ and $\delta_{\rm m}(z)$). Since the mass of a perturbation is conserved, and $\dot{R}\big|_{\rm ta} = 0$ we get

$$E = U_{\rm ta} = -\frac{G_{\rm N}M}{R_{\rm ta}} = -\frac{R_{\rm m,i}}{R_{\rm ta}} K_{\rm m,i}\Omega_{\rm I}(1+\delta_{\rm I}). \tag{676}$$

Equating (674) and (676) we get

$$\frac{R_{\rm ta}}{R_{\rm m,i}} = \frac{1+\delta_{\rm I}}{\delta_{\rm I} - (\Omega_{\rm I}^{-1} - 1)} \equiv \frac{1+\delta_{\rm m}(z)}{\delta_{\rm m}(z) - \frac{1-\Omega_{\rm m}}{\Omega_{\rm m}(1+z)}}. \tag{677}$$

The time evolution of a spherical mass shell is identical to that of a spatially open or closed FRW universe. The resulting equations of motion may be obtained by integrating Eq. (673) giving

$$\begin{aligned} R &= A(1-\cos\theta), \\ t &= B(\theta - \sin\theta), \end{aligned} \tag{678}$$

for the case $E < 0$, and

$$\begin{aligned} R &= A(\cosh\theta - 1), \\ t &= B(\sinh\theta - \theta), \end{aligned} \tag{679}$$

for the case $E > 0$. In Eqs. (678) and (679) we have $A^3 = G_{\rm N}MB^2$. The behaviour of the background universe is described by similar equations.

Setting $\theta = \pi$ in equation (678) we can express the constants A and B in terms of the turnaround radius $R_{\rm ta}$ and the turnaround time $t_{\rm ta}$: $A = R_{\rm ta}/2, B = t_{\rm ta}/\pi$. Next using Eq. (677) and the relationships $B^2 = A^3/G_\mathrm{N}M$, $M = 4\pi R^3\rho_\mathrm{m}/3$ and $8\pi G_\mathrm{N}\rho = 3H^2\Omega_\mathrm{m}$, we can re-express A and B in terms of $R_\mathrm{m,i}$ and δ_I

$$\begin{aligned} A &= \left(\frac{R_\mathrm{m,i}}{2}\right)\frac{1+\delta_\mathrm{I}}{\delta_\mathrm{I}-(\Omega_\mathrm{I}^{-1}-1)} \\ B &= \frac{1+\delta_\mathrm{I}}{2H\Omega_\mathrm{I}^{1/2}[\delta_\mathrm{I}-(\Omega_\mathrm{I}^{-1}-1)]^{3/2}}. \end{aligned} \tag{680}$$

In a spatially flat universe, Eq. (680) becomes

$$A \simeq \frac{R_\mathrm{m,i}}{2\delta_\mathrm{I}}, \quad B \simeq \frac{3}{4}t_\mathrm{I}\delta_\mathrm{I}^{-3/2}, \tag{681}$$

where we assume $\delta_\mathrm{I} \ll 1$. It is now relatively straightforward to compute the overdensity in each mass shell. Since mass is conserved we get, using $M = 4\pi R^3\rho_\mathrm{m}/3$ and Eq. (678),

$$\rho_\mathrm{m}(t) = \frac{3M}{4\pi A^3(1-\cos\theta)^3}. \tag{682}$$

In a spatially flat matter dominated universe the background density scales as

$$\bar{\rho}_\mathrm{m}(t) = \frac{1}{6\pi G_\mathrm{N}t^2} = \frac{1}{6\pi G_\mathrm{N}B^2(\theta-\sin\theta)^2}. \tag{683}$$

So, combining Eqs. (682) and (683), we get

$$\delta_\mathrm{m}(\theta) \equiv \frac{\rho_\mathrm{m}(t)}{\bar{\rho}_\mathrm{m}(t)} - 1 = \frac{9}{2}\frac{(\theta-\sin\theta)^2}{(1-\cos\theta)^3} - 1, \tag{684}$$

for positive density fluctuations, and

$$\delta_\mathrm{m}(\theta) = \frac{9}{2}\frac{(\theta-\sinh\theta)^2}{(\cosh\theta-1)^3} - 1 \tag{685}$$

for negative density fluctuations. From equations (684) and (681) we recover the linear limit for small θ, t:

$$\lim_{\theta\to 0}\delta_\mathrm{m}(\theta) \simeq \frac{3\theta^2}{20} \simeq \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3} = \frac{3}{5}\delta_\mathrm{I}\left(\frac{t}{t_\mathrm{I}}\right)^{2/3}, \tag{686}$$

indicating that only 3/5th of the initial amplitude is in the growing mode. In view of eq. (686) the criticality condition (675) translates into $\delta_\mathrm{I} > 3(\Omega_\mathrm{I}^{-1}-1)/5$ or, equivalently,

$$\delta_\mathrm{m}(z) > \frac{3}{5}\frac{1-\Omega_\mathrm{m,0}}{\Omega_\mathrm{m}(1+z)}. \tag{687}$$

From eq. (684) we find $\delta_\mathrm{m}(\theta=\pi) \simeq 4.6$ at the radius of maximum expansion ("turnaround"), and $\delta_\mathrm{m}(2\pi) \to \infty$ at recollapse. The corresponding extrapolated linear density contrast can be found from equations (686), (678) & (681):

$$\delta_\mathrm{L}(\theta) \simeq \frac{3}{5}\left(\frac{3}{4}\right)^{\frac{2}{3}}(\theta-\sin\theta)^{\frac{2}{3}}. \tag{688}$$

We thus obtain $\delta_{\rm L}(\pi) \simeq 1.063$ for the linear density contrast at turnaround, and $\delta_{\rm L}(2\pi) \equiv \delta_{\rm c} \simeq 1.686$ at recollapse. Knowing the linear density contrast corresponding to a given perturbation, the redshift at which that perturbation "turned around" and "collapsed" can be found from

$$\begin{aligned} 1 + z_{\rm ta} &\simeq \frac{\delta_{\rm L}}{1.063} \\ 1 + z_{\rm coll} &\simeq \frac{\delta_{\rm L}}{1.686}. \end{aligned} \tag{689}$$

In reality $\delta_{\rm coll} \to \infty$ will never be achieved since exact spherical collapse is at best a rather crude approximation, which will break down as the overdensity begins to contract, dynamical relaxation and shocks both ensuring that the system reaches virial equilibrium at a finite density. The maximum density at recollapse can be estimated using the virial theorem and the fact that at $R = R_{\rm ta}$ all the energy in the system is potential:

$$U(R = R_{\rm vir}) = 2E = 2U(R = R_{\rm ta}), \tag{690}$$

since $U = -G_{\rm N}M/R$ we get $R_{\rm vir} = R_{\rm ta}/2$ and $\rho_{\rm vir} = 8\rho_{\rm m,ta}$. The mean density of an object at turnaround is $\rho_{\rm m,ta}/\bar{\rho}_{\rm m} = \delta_{\rm m,ta} + 1 \simeq 5.6$ so that $\rho_{\rm m,ta} \simeq 5.6\bar{\rho}_{\rm m,ta}$. We therefore get $\rho_{\rm m,vir} \simeq 8 \cdot 5.6 \bar{\rho}_{\rm m,ta}$. Since $\bar{\rho}_{\rm m} = (6\pi G_{\rm N} t^2)^{-1}$ and setting $t_{\rm vir} \simeq t_{\rm coll} \simeq 2t_{\rm ta}$ we finally get $\rho_{\rm m,vir} \simeq 8 \times 5.6 \times 4\bar{\rho}_{\rm vir}$ or since $\bar{\rho}_{\rm m,vir} = (1+z)^3 \rho_{\rm m,0}$

$$\rho_{\rm m,vir} \simeq 179.2(1 + z_{\rm vir})^3 \rho_{\rm m,0}, \tag{691}$$

where $z_{\rm vir}$ is the collapse redshift, and ρ_0 the present matter density. Equation (691) permits us to relate the virialised density of a collapsed object to the epoch of its formation $z_{\rm vir} \simeq 0.18(\rho/\rho_0)^{1/3} - 1$. Since the present overdensity in clusters is between 10^2 and 10^4, the above arguments might indicate that clusters formed relatively recently at redshifts $z \leq 3$, provided they formed from spherical density enhancements.

Generalisations of these arguments show that the addition of a cosmological constant to the Einstein equations does not significantly affect the dynamics of a spherical overdensity. The final (virial) radius of a spherical overdensity in this case turns out to be

$$\frac{R_{\rm vir}}{R_{\rm ta}} \simeq \frac{1 - \eta/2}{2 - \eta/2} < \frac{1}{2}, \tag{692}$$

where $\eta = \Lambda / 8\pi G_{\rm N} \bar{\rho}(t = t_{\rm ta})$ is the ratio of the cosmological constant to the background density at turnaround. Equation (692) indicates that the presence of a positive cosmological constant leads to a somewhat smaller final radius (and consequently higher density) of a collapsed object. This effect is clearly larger for objects that collapse later, when the turnaround density is lower.

According to the spherical collapse model, CDM overdensities collapse to highly nonlinear objects, the spherical DM halos, whose potential wells will make the baryons fall into them and

$$\boxed{\text{HALOS ARE PEAKS OF THE UNDERLYING MATTER DISTRIBUTION.}} \tag{693}$$

It is therefore clear that one basic object in cosmology is the number of DM halos as function of their mass or radius. This is the quantity we would like to discuss in the next section.

15 The dark matter halo mass function and the excursion set method

The computation of the mass function of dark matter halos is a central problem in modern cosmology. The halo mass function is both a sensitive probe of cosmological parameters and a crucial ingredient when one studies the dark matter distribution, as well as the formation, evolution and distribution of galaxies, so its accurate prediction is obviously important.

The formation and evolution of dark matter halos is a highly complex dynamical process, and a detailed understanding of it can only come through large-scale N-body simulations. Some analytical understanding is however also desirable, both for obtaining a better physical intuition, and for the flexibility under changes of models or parameters.

The halo mass function $\mathrm{d}n/\mathrm{d}M$ can be written as (we drop from now on the subscript $_{\mathrm{m}}$ for DM)

$$\boxed{\frac{\mathrm{d}n(M)}{\mathrm{d}M} = f(\sigma)\frac{\bar{\rho}}{M^2}\frac{\mathrm{d}\ln\sigma^{-1}(M)}{\mathrm{d}\ln M},} \tag{694}$$

where $n(M)$ is the number density of dark matter halos of mass M, σ^2 is the variance of the linear density field smoothed on a scale R corresponding to a mass M, and $\bar{\rho}$ is the average density of the universe.

Now, to compute $\mathrm{d}n/\mathrm{d}M$ or $f(\sigma)$ we use the famous Press-Schecther (PS) formalism. Press and Schechter observed that the fraction of mass in collapsed objects more massive than some mass M is related to the fraction of volume samples in which the smoothed initial density fluctuations are above some density threshold. This yields a formula for the mass function (distribution of masses) of objects at any given time. In other words the philosophy is the following. We know that the collapse of DM halo is a very complicated phenomenon. Nevertheless we are not interested in describing the dynamics itself, but only in computing the probability that at a given point $\mathbf{x}$ a halo will form. But when does it? Press and Schechter assumed that the collapse is spherical and argued that a halo of mass M and radius R is formed when the corresponding smoothed linear density contrast at recollapse is larger than the critical value $\delta_{\mathrm{c}} \simeq 1.68$ computed in the previous section. Of course the real density contrast will be much larger, of the order of 200, but we are not interested in it. Notice that in the PS formalism therefore there is no dynamics in time, it only provides the probability that at a given point a halo of mass M will form. In order to compute such a probability we turn to a beautiful statistical tool, the excursion set method.

15.1 The computation of the halo mass function as a stochastic problem

The computation of the halo mass function can be formulated in terms of a stochastic process

One considers the density contrast $\delta(\mathbf{x})$ and smooths it on some scale R

$$\delta(\mathbf{x},R)=\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta_{\mathbf{k}}W(k,R)e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{695}$$

We focus on the evolution of $\delta(\mathbf{x},R)$ with R at a fixed value of $\mathbf{x}$, that we can choose without loss of generality as $\mathbf{x}=0$, and we write $\delta(\mathbf{x}=0,R)$ simply as $\delta(R)$. Taking the derivative of Eq. (695) with respect to R we get

$$\frac{\partial\delta(R)}{\partial R}=\zeta(R), \tag{696}$$

where

$$\zeta(R)\equiv\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta_{\mathbf{k}}\frac{\partial W(k,R)}{\partial R}. \tag{697}$$

Since the modes $\delta_{\mathbf{k}}$ are stochastic variables, $\zeta(R)$ is a stochastic variable too, and Eq. (696) has the form of a Langevin equation, with R playing the role of time, and $\zeta(R)$ playing the role of noise. When $\delta(R)$ is a Gaussian variable, only its two-point connected correlator is non-vanishing. In this case, we see from Eq. (zetadelta) that also ζ is Gaussian. The two-point function of δ defines the power spectrum $P(k)$,

$$\langle\delta(\mathbf{k})\delta_(\mathbf{k}')\rangle=(2\pi)^3\delta^{(3)}(\mathbf{k}+\mathbf{k}')P(k)\,. \tag{698}$$

From this it follows that

$$\langle\zeta(R_1)\zeta(R_2)\rangle=\int_{-\infty}^{\infty}\mathrm{d}(\ln k)\;\Delta^2(k)\frac{\partial W(k,R_1)}{\partial R_1}\frac{\partial W(k,R_2)}{\partial R_2}, \tag{699}$$

where, as usual, $\Delta^2(k)=k^3P(k)/2\pi^2$. For a generic filter function the right-hand side is a function of R_1 and R_2, different from a Dirac delta $\delta_{\mathrm{D}}(R_1-R_2)$. In the literature on stochastic processes this case is known as *colored Gaussian noise*. Things simplify considerably for a sharp k-space filter $W(k,R)=\theta(k-k_{\mathrm{f}})$. Using $k_{\mathrm{f}}=1/R$ instead of R, and defining $Q(k_{\mathrm{f}})=-(1/k_{\mathrm{f}})\zeta(k_{\mathrm{f}})$, we find

$$\frac{\partial\delta(k_{\mathrm{f}})}{\partial\ln k_{\mathrm{f}}}=Q(k_{\mathrm{f}}), \tag{700}$$

and

$$\langle Q(k_{\mathrm{f}1})Q(k_{\mathrm{f}2})\rangle=\Delta^2(k_{\mathrm{f}1})\delta^{(3)}(\ln k_{\mathrm{f}1}-\ln k_{\mathrm{f}2}). \tag{701}$$

Therefore, we have a Dirac delta noise. We can write these equations in an even simpler form using as “pseudotime” variable the variance $S=\sigma^2$

$$S(R)=\int_{-\infty}^{\infty}d(\ln k)\,\Delta^2(k)|W(k,R)|^2. \tag{702}$$

For a sharp k-space filter, S becomes

$$S(k_{\rm f}) = \int_{-\infty}^{\ln k_{\rm f}} d(\ln k)\, \Delta^2(k), \tag{703}$$

so

$$\frac{\partial S}{\partial \ln k_{\rm f}} = \Delta^2(k_{\rm f}). \tag{704}$$

Thus, redefining finally $\eta(k_{\rm f}) = Q(k_{\rm f})/\Delta^2(k_{\rm f})$, we get

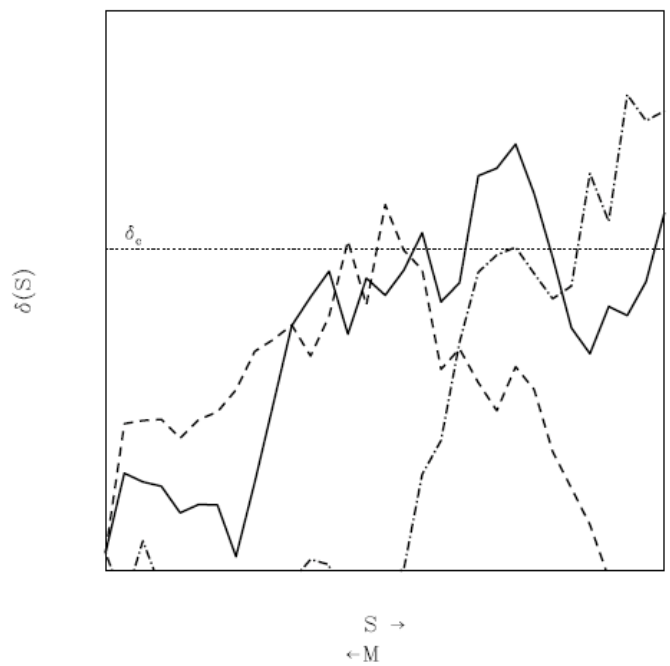

Figure 23: The stochastic motion of the smoothed density contrast.

$$\boxed{\frac{\partial \delta(S)}{\partial S} = \eta(S),} \tag{705}$$

with

$$\boxed{\langle \eta(S_1)\eta(S_2) \rangle = \delta(S_1 - S_2),} \tag{706}$$

which is a the Langevin equation with Dirac-delta noise, with S playing the role of time. In hierarchical power spectra, at $R = \infty$ we have $S = 0$, and S increases monotonically as R decreases. Therefore we can start from $R = \infty$, corresponding to "time" $S = 0$, where d $= 0$, and follow the evolution of $\delta(S)$ as we decrease R, *i.e.* as we increase S. The fact that this evolution is governed by the Langevin equation means that $\delta(S)$ performs a random walk, with respect to the "time" variable

S. We may refer to the evolution of δ as a function of S as a "trajectory". In the spherical collapse model, a virialized object forms as soon as the trajectory exceeds the threshold $\delta = \delta_c$.

According to the PS formalism an halo is formed when the stochastic quantity $\delta(R)$ in its random walk crosses the barrier at δ_c for the first time. It is therefore a problem known in statistical mechanics as the first-time passage problem. We are not interested in subsequent crossing of the barrier as the halo has formed after the first crossing.

We therefore consider an ensemble of trajectories, all starting from the initial value $\delta = 0$ at initial "time" $S = 0$, and have compute the function that gives the probability distribution of reaching a value δ_c at "time" S for the first time.

Notice that or a Gaussian theory, the only non-vanishing connected correlator is then the two-point correlator $\langle\delta(S_1)\delta(S_2)\rangle_c$, where the subscript c stands for connected.

We consider an ensemble of trajectories all starting at $S_0 = 0$ from an initial position $\delta(0) = \delta_0$, and we follow them for a time S. We discretize the interval $[0, S]$ in steps $\Delta S = \epsilon$, so $S_k = k\epsilon$ with $k = 1, \ldots n$, and $S_n \equiv S$. A trajectory is defined by the collection of values $\{\delta_1, \ldots, \delta_n\}$, such that $\delta(S_k) = \delta_k$. There is no absorbing barrier, i.e. $\delta(S)$ is allowed to range freely from $-\infty$ to $+\infty$. The probability density in the space of trajectories is

$$W(\delta_0; \delta_1, \ldots, \delta_n; S_n) \equiv \langle \delta_\mathrm{D}(\delta(S_1) - \delta_1) \ldots \delta_\mathrm{D}(\delta(S_n) - \delta_n)\rangle, \tag{707}$$

In terms of W we define

$$\Pi(\delta_0; \delta_n; S_n) \equiv \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \int_{-\infty}^{\delta_c} \mathrm{d}\delta_{n-1}\, W(\delta_0; \delta_1, \ldots, \delta_{n-1}, \delta_n; S_n)\,, \tag{708}$$

where $S_n = n\epsilon$. So, $\Pi(\delta_0; \delta; S)$ is the probability density of arriving at the "position" δ in a "time" S, starting from δ_0 at time $S_0 = 0$, through trajectories that never exceeded δ_c. Observe that the final point δ ranges over $-\infty < \delta < \infty$.

The usefulness of Π is that it allows us to compute the first-crossing rate from first principles, without the need of postulating the existence of an absorbing barrier. Simply, the quantity

$$F(S) = 1 - \int_{-\infty}^{\delta_c} \mathrm{d}\delta\, \Pi(\delta_0; \delta; S) \tag{709}$$

gives the fraction of trajectories that crossed the barrier at time S. The rate of change of this quantity is therefore equal to minus the rate at which trajectories cross for the first time the barrier, so the first-crossing rate is

$$\mathcal{F}(S) = -\frac{\mathrm{d}F}{\mathrm{d}S} = -\int_{-\infty}^{\delta_c} \mathrm{d}\delta\, \frac{\partial}{\partial S}\Pi(\delta_0; \delta; S). \tag{710}$$

The halo mass function follows if one has a relation $M = M(R)$ that gives the mass associated to the smoothing of d over a region of radius R. Once $M(R)$ is given, we can consider F as a function of

M rather than of $S(R)$. Then $|\mathrm{d}F/dM|\mathrm{d}M$ is the fraction of volume occupied by virialized objects with mass between M and $M+\mathrm{d}M$. Since each one occupies a volume $V = M/\bar{\rho}$, where $\bar{\rho}$ is the average density of the universe, the number of virialized object $n(M)$ with mass between M and $M+dM$ is given by

$$\frac{\mathrm{d}n}{\mathrm{d}M}\mathrm{d}M = \frac{\bar{\rho}}{M}\left|\frac{\mathrm{d}F}{\mathrm{d}M}\right|\mathrm{d}M, \tag{711}$$

so

$$\frac{\mathrm{d}n}{\mathrm{d}M} = \frac{\bar{\rho}}{M}\frac{\mathrm{d}F}{\mathrm{d}S}\left|\frac{\mathrm{d}S}{\mathrm{d}M}\right| = \frac{\bar{\rho}}{M^2}\mathcal{F}(S)2\sigma^2\frac{\mathrm{d}\ln\sigma^{-1}}{\mathrm{d}\ln M}, \tag{712}$$

where we used $S=\sigma^2$. Therefore, in terms of the first-crossing rate $\mathcal{F}(S) = \mathrm{d}F/\mathrm{d}S$, the function $f(\sigma)$ defined from Eq. (769) is given by

$$f(\sigma) = 2\sigma^2\mathcal{F}(\sigma^2). \tag{713}$$

To deduce $\Pi(\delta_0;\delta_n;S)$ we use a path integral formulation. We use the integral representation of the Dirac delta

$$\delta_{\mathrm{D}}(x) = \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{2\pi}\,e^{-i\lambda x}, \tag{714}$$

and we write Eq. (741) as

$$W(\delta_0;\delta_1,\ldots,\delta_n;S_n) = \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda_1}{2\pi}\ldots\frac{\mathrm{d}\lambda_n}{2\pi}\,e^{i\sum_{i=1}^n\lambda_i\delta_i}\langle e^{-i\sum_{i=1}^n\lambda_i\delta(S_i)}\rangle. \tag{715}$$

Observe that the dependence on δ_0 here is hidden in the correlators of δ, *e.g.* $\langle\delta^2(S=0)\rangle = \delta_0^2$. It is convenient to set for simplicity $\delta_0 = 0$ from now on. For Gaussian fluctuations,

$$\langle e^{-i\sum_{i=1}^n\lambda_i\delta(S_i)}\rangle = e^{-\frac{1}{2}\sum_{i,j=1}^n\lambda_i\lambda_j\langle\delta(S_i)\delta(S_j)\rangle_c}, \tag{716}$$

as can be checked immediately by performing the Taylor expansion of the exponential on the left-hand side, and using the fact that, for Gaussian fluctuations, the generic correlator factorizes into sum of products of two-points correlators. This gives

$$W(\delta_0;\delta_1,\ldots,\delta_n;S_n) = \int\mathcal{D}\lambda\;e^{i\sum_{i=1}^n\lambda_i\delta_i-\frac{1}{2}\sum_{i,j=1}^n\lambda_i\lambda_j\langle\delta_i\delta_j\rangle_c}, \tag{717}$$

where

$$\int\mathcal{D}\lambda \equiv \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda_1}{2\pi}\ldots\frac{\mathrm{d}\lambda_n}{2\pi}\,, \tag{718}$$

and $\delta_i\equiv\delta(S_i)$. Then

$$\Pi(\delta_n;S_n) = \int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_{n-1}\int\mathcal{D}\lambda\exp\left\{i\sum_{i=1}^n\lambda_i\delta_i - \frac{1}{2}\sum_{i,j=1}^n\lambda_i\lambda_j\langle\delta_i\delta_j\rangle_c\right\}. \tag{719}$$

Let us now compute the two-point correlator. Using as initial condition Eq. (705) integrates to

$$\delta(S) = \int_0^S\mathrm{d}S'\,\eta(S')\,, \tag{720}$$

so the two-point correlator is given by

$$\langle \delta(S_i)\delta(S_j)\rangle_c = \int_0^{S_i} \mathrm{d}S \int_0^{S_j} \mathrm{d}S' \langle \eta(S)\eta(S')\rangle = \min(S_i, S_j). \tag{721}$$

Let us now take the derivative of Eq. (753) with respect to S_n

$$\begin{aligned} \frac{\partial}{\partial S_n}\Pi(\delta_n; S_n) &= \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_{n-1} \int \mathcal{D}\lambda \left(-\frac{\lambda_n^2}{2}\right) \exp\left\{ i\sum_{i=1}^n \lambda_i \delta_i - \frac{1}{2}\sum_{i,j=1}^n \lambda_i\lambda_j \langle \delta_i \delta_j\rangle_c \right\} \\ &= \frac{1}{2}\frac{\partial^2}{\partial \delta_n^2} \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_{n-1} \int \mathcal{D}\lambda \exp\left\{ i\sum_{i=1}^n \lambda_i \delta_i - \frac{1}{2}\sum_{i,j=1}^n \lambda_i\lambda_j \langle \delta_i \delta_j\rangle_c \right\}. \end{aligned} \tag{722}$$

We discover that the probability $Pi(\delta_0; \delta_n; S_n)$ satisfies a Fokker-Planck equation (setting $S_n = S$ and $\delta_n = \delta$)

$$\boxed{\frac{\partial}{\partial S}\Pi(\delta; S) = \frac{1}{2}\frac{\partial^2}{\partial \delta^2}\Pi(\delta; S),} \tag{723}$$

which has to be solved with the following boundary conditions

$$\Pi(\delta; 0) = \delta_{\rm D}(\delta - \delta_0), \ \text{ and } \ \Pi(\delta_{\rm c}; S) = 0, \tag{724}$$

where we have restored a non vanishing initial condition. The first condition says that the trajectory has to start from $\delta(0) = \delta_0$, while the second condition simply states that as the random walk reaches the barrier at $\delta_{\rm c}$ for the first time the motion should stop, or in other words, the trajectory should be removed.

The solution of the Fokker-Planck equation with such boundary conditions is given by

$$\Pi(\delta; S) = \frac{1}{\sqrt{2\pi S}}\left(e^{-(\delta-\delta_0)^2/2S} - e^{(2\delta_{\rm c}-\delta+\delta_0)^2/2S}\right). \tag{725}$$

Correspondingly

$$\mathcal{F}(S) = -\frac{1}{2}\left.\frac{\partial \Pi}{\partial \delta}\right|_{\delta=\delta_{\rm c}} = \frac{1}{\sqrt{2\pi}S^{3/2}} e^{-(\delta_{\rm c}-\delta_0)^2/2S}, \tag{726}$$

and (now setting δ_0 back to zero)

$$\boxed{\left(\frac{\mathrm{d}n}{\mathrm{d}M}\right)_{\rm PS} = \left(\frac{2}{\pi}\right)^{1/2} \frac{\delta_{\rm c}}{\sigma}\, e^{-\delta_{\rm c}^2/(2\sigma^2)}\, \frac{\bar{\rho}}{M^2}\frac{\mathrm{d}\ln\sigma^{-1}}{\mathrm{d}\ln M}.} \tag{727}$$

This result can be extended to arbitrary redshift z reabsorbing the evolution of the variance into $\delta_{\rm c}$, so that $\delta_{\rm c}$ in the above result is replaced by $\delta_{\rm c}(z) = \delta_{\rm c}(0)/D_+(z)$, where $D_+(z)$ is the linear growth factor.

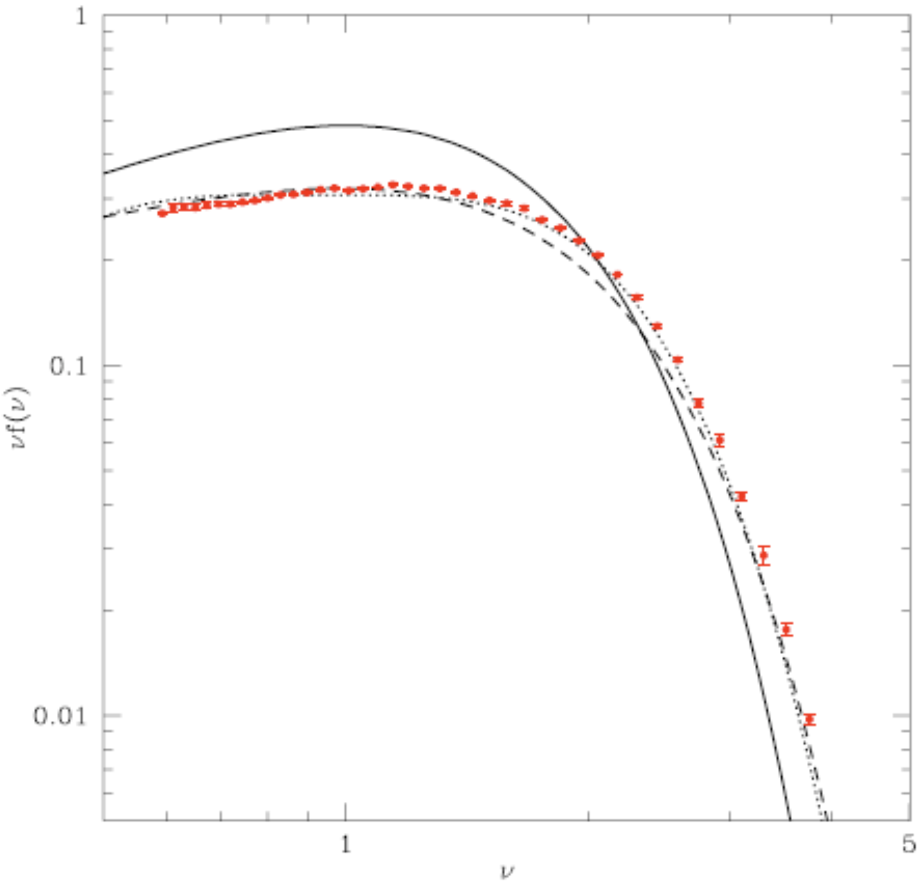

Figure 24: The Press-Schechter prediction (solid line) versus N-body data.

One can see from Fig. 24, where we plot the PS halo mass function as a function of the parameter $\nu = \delta_c/S^{1/2}$, that PS theory predicts a little bit too many low-mass halos, roughly by a factor of two, and too few high-mass halos, by a factor $\mathcal{O}(10)$ or so. Nevertheless, the simple logic behind the PS theory work surprisingly well if we think of how complicated the collapse is. In particular, the exponential tail of the halo mass function is obtained (even though not with quite the correct coefficient in the exponent) and due to the Gaussian nature of the perturbations. Nowdays analytic techniques generally go beyond the PS approach and model the collapse as, *e.g.* ellipsoidal. However, the PS theory is able to reproduce, at least qualitatively, several properties of dark matter halos such as their conditional and unconditional mass function, halo accretion histories, merger rates and halo bias.

16 The bias

In order to make full use of the cosmological information encoded in large-scale structure, it is essential to understand the relation between the number density of galaxies and the mass density field. It was first appreciated during the 1980s that these two fields need not be strictly proportional. So it is useful to introduce the linear bias parameter

$$\left(\frac{\delta\rho}{\rho}\right)_{\rm g} = \left(\frac{\delta\rho}{\rho}\right)_{\rm m}. \tag{728}$$

This seems a reasonable assumption when $\delta\rho/\rho \ll 1$, although it leaves open the question of how the effective value of b would be expected to change on nonlinear scales. Galaxy clustering on large scales therefore allows us to determine mass fluctuations only if we know the value of the bias parameter b. We now consider the central mechanism of biased clustering, in which a rare high density fluctuation, corresponding to a massive object, collapses sooner if it lies in a region of large-scale overdensity. This helping hand from the long-wavelength modes means that overdense regions contain an enhanced abundance of massive objects with respect to the mean, so that these systems display enhanced clustering. The basic mechanism can be immediately understood via the Fig. 25 which explains the peak-background split model. If we decompose a density field into a fluctuating component on galaxy scales, together with a long-wavelength as well, then those regions of density that lie above a threshold in density of ν times the variance σ will be strongly clustered. If proto-objects are presumed to form at the sites of these high peaks, then this is a population with Lagrangian bias, *i.e.* a non-uniform spatial distribution even prior to dynamical evolution of the density field. The key question is the physical origin of the threshold: for massive objects such as clusters, the requirement of collapse by the present imposes a threshold of $\nu \gtrsim 2$. For galaxies, there will be no bias without additional mechanisms to cause star formation to favour those objects that collapse first. The excursion set formalism provides a neat framework to understand how the clustering of dark matter halos differs from the overall clustering of matter. Consider the solution to the excursion set problem in we derived in the previous section. This gives the probability distribution of δ given that on a smoothing scale S_0, the smoothed density fluctuation is δ_0. Notice that the important quantity is the relative height of the density threshold $(\delta_{\rm c} - \delta_0)$ so that in regions with $\delta_0 > 0$ on large scales, trajectories are more likely to penetrate the barrier at $\delta_{\rm c}$ and conversely for $\delta_0 < 0$.

The fraction of mass in collapsed halos of mass greater than M in a region that has a smoothed density fluctuation δ_0 on scale S_0 (corresponding to mass M_0 and volume V_0) is given

$$F(M|\delta_0, S_0) = {\rm Erfc}\left(\frac{\delta_{\rm c} - \delta_0}{\sqrt{2}(S - S_0)}\right). \tag{729}$$

Notice that as the density of the region increases, F increases because smaller upward excursions

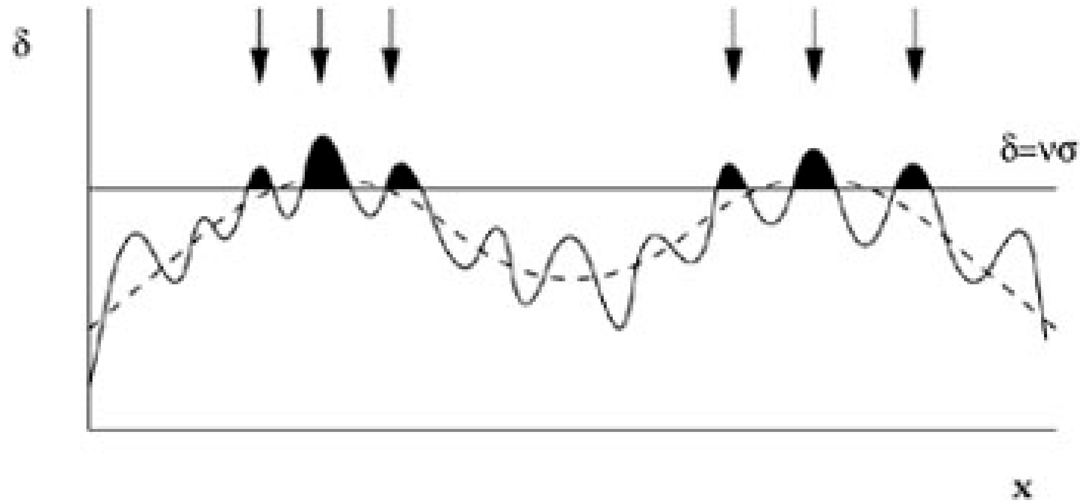

Figure 25: The peak-background split model.

are needed to cross the threshold. When $\delta_0 \to \delta_c$, $F \to 1$ because the entire region will then be interpreted as a collapsed halo of mass M_0. The fraction of mass in halos with mass in the range M to $M + \mathrm{d}M$ is

$$f(M|\delta_0, S_0)\left|\frac{\mathrm{d}S}{\mathrm{d}M}\right|\mathrm{d}M \equiv \frac{\mathrm{d}F(M|\delta_0, S_0)}{\mathrm{d}M}\mathrm{d}M$$
$$= \frac{1}{\sqrt{2\pi}}\frac{\delta_c - \delta_0}{(S - S_0)^{3/2}}\left|\frac{\mathrm{d}S}{\mathrm{d}M}\right| \exp\left[-\frac{(\delta_c - \delta_0)^2}{2(S - S_0)}\right]\mathrm{d}M, \quad (730)$$

so that regions with smoothed density δ_0 on scale S_0 contain, on average,

$$\mathcal{N}(M|\delta_0, S_0)\mathrm{d}M = \frac{M_0}{M}\, f(M|\delta_0, S_0)\left|\frac{\mathrm{d}S}{\mathrm{d}M}\right|\mathrm{d}M \quad (731)$$

halos in this mass range. The quantity of interest is the relative over-abundance of halos in dense regions compared to the mean abundance of halos,

$$\delta^{\mathrm{L}}_{\mathrm{halo}} = \frac{\mathcal{N}(M|\delta_0, S_0)}{(\mathrm{d}n(M)/\mathrm{d}M)V_0} - 1, \quad (732)$$

where $\mathrm{d}n(M)/\mathrm{d}M$ is the mean number density of halos in a mass range of width $\mathrm{d}M$ about M. The superscript L indicates that this is the overdensity in the initial Lagrangian space determined by the mass distribution at some very early time, ignoring the dynamical evolution of the overdense patch.

The relative overdensity of halos in large overdense and underdense patches is easy to compute. In sufficiently large regions, $S_0 \ll S$, $\delta_0 \ll \delta_c$. Expanding Eq. (732) to first order in the variables S_0/S and δ_0/δ_c gives a simple relation between halo abundance and dark matter density

$$\delta^{\mathrm{L}}_{\mathrm{halo}} = \frac{\nu^2 - 1}{\delta_c}\delta_0, \quad (733)$$

where $\nu = \delta_c/S^{1/2} = \delta_c/\sigma(M)$ as before. The overdensity in the initial Lagrangian space is proportional to the dark matter overdensity and is a function of halo mass through ν. The final ingredient needed to relate the abundance of halos to the matter density is a model for the dynamics that can map the initial Lagrangian volume to the final Eulerian space. Let V and δ represent the Eulerian space variables corresponding to the Lagrangian space variables V_0 and δ_0. The final halo abundance is

$$\delta_h = \frac{\mathcal{N}(M|\delta_0, S_0)}{(\mathrm{d}n(M)/\mathrm{d}M)V} - 1. \tag{734}$$

Mass conservation implies $V(1+\delta) = V_0(1+\delta_0)$, but In the limit of a small overdensity, $\delta_0 \ll 1$, $\delta \simeq \delta_0$, and $V \simeq V_0$, and therefore

$$\delta_h = \left(1 + \frac{\nu^2 - 1}{\delta_c}\right)\delta \tag{735}$$

$$\equiv b_h \delta. \tag{736}$$

This expression states that the overdensity of halos is linearly proportional to the overdensity of the mass. The constant of proportionality $b_h(M, z)$ depends on the masses of the halos, and the redshifts they virialized, but is independent of the size of the cells. Furthermore, its says that low-mass haloes are antibiased and high mass haloes are positively biased. We can now understand the observation that there are much more strongly clustered than galaxies in general: regions of large-scale overdensity contain systematically more high-mass haloes than expected if the haloes traced the mass. Indeed, by defining the correlation function

$$\begin{aligned} \xi(r) &= \Big\langle \delta_m(\mathbf{x}+\mathbf{r})\delta_m(\mathbf{x}) \Big\rangle = \int \frac{\mathrm{d}k^3}{(2\pi)^3 V} \, |\delta^m_{\mathbf{k}}|^2 \; e^{-i\mathbf{k}\cdot\mathbf{r}} = \int_0^\infty \mathrm{d}\ln k \, \frac{k^3 \, |\delta^m_{\mathbf{k}}|^2}{2\pi^2} \frac{\sin kr}{kr}, \\ &= \int_0^\infty \mathrm{d}\ln k \, \Delta_m(k) \frac{\sin kr}{kr} \end{aligned} \tag{737}$$

for which, if $\Delta^2_m(k) \sim k^{n+3}$ we have $\xi(r) \sim r^{-n-3}$, it turns out observationally that

$$\xi_{cc}(r) \simeq \left(\frac{r}{25\, h^{-1}\,\mathrm{Mpc}}\right)^{-1.8} \simeq 20\xi_{gg}(r) \simeq 20 \left(\frac{r}{5\, h^{-1}\,\mathrm{Mpc}}\right)^{-1.8}, \tag{738}$$

that is clusters are more correlated than galaxies. However, one should be careful that applying the idea to galaxies is not straightforward: we have shown that enhanced clustering is only expected for massive fluctuations with $\sigma \lesssim 1$, but galaxies at $z = 0$ fail this criterion. The high-peak idea applies will at high redshift, where massive galaxies are still assembling, but today there has been time for galaxy-scale haloes to collapse in all environments.

Part VII

The impact of the non-Gaussianity on the halo mass function

Non-Gaussianities are particularly relevant in the high-mass end of the power spectrum of perturbations, *i.e.* on the scale of galaxy clusters, since the effect of non-Gaussian fluctuations becomes especially visible on the tail of the probability distribution. As a result, both the abundance and the clustering properties of very massive halos are sensitive probes of primordial non-Gaussianities, and could be detected or significantly constrained by the various planned large-scale galaxy surveys, both ground based (such as DES, PanSTARRS and LSST) and on satellite (such as EUCLID and ADEPT). Furthermore, the primordial non-Gaussianity alters the clustering of dark matter halos inducing a scale-dependent bias on large scales while even for small primordial non-Gaussianity the evolution of perturbations on super-Hubble scales yields extra contributions on smaller scales. This will be the subject of the next subsection.

At present, there exist already various N-body simulations where non-Gaussianity has been included in the initial conditions and which are useful to test the accuracy of the different theoretical predictions for the dark matter halo mass function with non-Gaussianity.

Various attempts at computing analytically the effect of primordial non-Gaussianities on the mass function exist in the literature and here we follow what we have developed in Section 47 based on the excursion set method.

In this section we extend to non-Gaussian fluctuations the path integral approach that we developed in Section 47 whose notation we follow. In particular, we consider the density field δ smoothed over a radius R with a tophat filter in momentum space. We denote by S the variance of the smoothed density field and, as usual in excursion set theory, we consider δ as a variable evolving stochastically with respect to the "pseudotime" S. The statistical properties of a random variable $\delta(S)$ are specified by its connected correlators

$$\langle \delta(S_1) \ldots \delta(S_p) \rangle_c \,, \tag{739}$$

where the subscript c stands for "connected". We will also use the notation

$$\langle \delta^p(S) \rangle_c \equiv \mu_p(S) \,, \tag{740}$$

when all arguments $S_1, S_2, \ldots$ are equal. The quantities $\mu_p(S)$ are also called the cumulants. As in Section 47, we consider an ensemble of trajectories all starting at $S_0 = 0$ from an initial position $\delta(0) = \delta_0$ (we will typically choose $\delta_0 = 0$ but the computation can be performed in full generality) and we follow them for a "time" S. We discretize the interval $[0, S]$ in steps $S_k = k\epsilon$ with $k = 1, \dots n$, and $S_n \equiv S$. A trajectory is then defined by the collection of values $\{\delta_1, \dots, \delta_n\}$, such that $\delta(S_k) = \delta_k$.

The probability density in the space of trajectories is

$$W(\delta_0; \delta_1, \dots, \delta_n; S_n) \equiv \langle \delta_D(\delta(S_1) - \delta_1) \dots \delta_D(\delta(S_n) - \delta_n) \rangle \,, \tag{741}$$

where δ_D denotes the Dirac delta. Our basic object will be

$$\Pi_\epsilon(\delta_0; \delta_n; S_n) \equiv \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \dots \int_{-\infty}^{\delta_c} \mathrm{d}\delta_{n-1} \, W(\delta_0; \delta_1, \dots, \delta_{n-1}, \delta_n; S_n). \tag{742}$$

The usefulness of Π_ϵ is that it allows us to compute the first-crossing rate from first principles, without the need of postulating the existence of an absorbing barrier. In fact, the quantity

$$\int_{-\infty}^{\delta_c} \mathrm{d}\delta_n \, \Pi_\epsilon(\delta_0; \delta_n; S_n) \tag{743}$$

gives the probability that at "time" S_n a trajectory always stayed in the region $\delta < \delta_c$, for all times' smaller than S_n. The rate of change of this quantity is therefore equal to minus the rate at which trajectories cross for the first time the barrier, so the first-crossing rate is

$$\mathcal{F}(S_n) = -\frac{\partial}{\partial S_n} \int_{-\infty}^{\delta_c} \mathrm{d}\delta_n \, \Pi_\epsilon(\delta_0; \delta_n; S_n) \,. \tag{744}$$

The halo mass function is then obtained from the first-crossing rate

$$f(\sigma) = 2\sigma^2 \mathcal{F}(\sigma^2) \,, \tag{745}$$

where $S = \sigma^2$.

The first problem that we address is how to express $\Pi_\epsilon(\delta_0; \delta_n; S)$, in terms of the correlators of the theory. Using the integral representation of the Dirac delta

$$\delta_D(x) = \int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda}{2\pi} \, e^{-i\lambda x} \,, \tag{746}$$

we may write

$$W(\delta_0; \delta_1, \dots, \delta_n; S_n) = \int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda_1}{2\pi} \dots \frac{\mathrm{d}\lambda_n}{2\pi} \, e^{i \sum_{i=1}^n \lambda_i \delta_i} \langle e^{-i \sum_{i=1}^n \lambda_i \delta(S_i)} \rangle \,. \tag{747}$$

We must therefore compute

$$e^Z \equiv \langle e^{-i \sum_{i=1}^n \lambda_i \delta(S_i)} \rangle \,. \tag{748}$$

This is a well-known object both in quantum field theory and in statistical mechanics, since it is the generating functional of the connected Green's functions. To a field theorist this is even more clear if we define the "current" J from $-i\lambda = \epsilon J$, and we use a continuous notation, so that

$$e^Z = \langle e^{i \int \mathrm{d}S \, J(S) \delta(S)} \rangle \,. \tag{749}$$

Therefore

$$
\begin{aligned}
Z &= \sum_{p=2}^{\infty} \frac{(-i)^p}{p!} \sum_{i_1=1}^{n} \ldots \sum_{i_p=1}^{n} \lambda_{i_1} \ldots \lambda_{i_p} \langle \delta_{i_1} \ldots \delta_{i_p} \rangle_c \\
&= -\frac{1}{2} \lambda_i \lambda_j \langle \delta_i \delta_j \rangle_c + \frac{(-i)^3}{3!} \lambda_i \lambda_j \lambda_k \langle \delta_i \delta_j \delta_k \rangle_c \\
&\quad + \frac{(-i)^4}{4!} \lambda_i \lambda_j \lambda_k \lambda_l \langle \delta_i \delta_j \delta_k \delta_l \rangle_c + \ldots , \qquad (750)
\end{aligned}
$$

where $\delta_i = \delta(S_i)$ and the sum over $i, j, \ldots$ is understood. This gives

$$
W(\delta_0; \delta_1, \ldots, \delta_n; S_n) = \mathcal{D}\lambda \exp\left\{ i \sum_{i=1}^{n} \lambda_i \delta_i + \sum_{p=2}^{\infty} \frac{(-i)^p}{p!} \sum_{i_1=1}^{n} \ldots \sum_{i_p=1}^{n} \lambda_{i_1} \ldots \lambda_{i_p} \langle \delta_{i_1} \ldots \delta_{i_p} \rangle_c \right\} , \qquad (751)
$$

where

$$
\mathcal{D}\lambda \equiv \int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda_1}{2\pi} \ldots \frac{\mathrm{d}\lambda_n}{2\pi} . \qquad (752)
$$

Therefore we get

$$
\Pi_\epsilon(\delta_0; \delta_n; S_n) = \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_{n-1} \, \mathcal{D}\lambda \exp\left\{ i \sum_{i_1=1}^{n} \lambda_i \delta_i + \sum_{p=2}^{\infty} \frac{(-i)^p}{p!} \sum_{i=1}^{n} \ldots \sum_{i_p=1}^{n} \lambda_{i_1} \ldots \lambda_{i_p} \langle \delta_{i_1} \ldots \delta_{i_p} \rangle_c \right\} . \qquad (753)
$$

If we retain only the three-point correlator, and we use the tophat filter in momentum space, we have

$$
\Pi_\epsilon(\delta_0; \delta_n; S_n) = \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots d\delta_{n-1} \, \mathcal{D}\lambda \exp\left\{ i\lambda_i \delta_i - \frac{1}{2} \min(S_i, S_j) \lambda_i \lambda_j + \frac{(-i)^3}{6} \langle \delta_i \delta_j \delta_k \rangle \lambda_i \lambda_j \lambda_k \right\} . \qquad (754)
$$

Expanding to first order in NG, we must compute

$$
\Pi_\epsilon^{(3)}(\delta_0; \delta_n; S_n) \equiv -\frac{1}{6} \sum_{i,j,k=1}^{n} \langle \delta_i \delta_j \delta_k \rangle \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots d\delta_{n-1} \, \partial_i \partial_j \partial_k W^{\mathrm{gm}} , \qquad (755)
$$

where the superscript (3) in $\Pi_\epsilon^{(3)}$ refers to the fact that this is the contribution linear in the three-point correlator. To proceed, we remember that the non-Gaussianities are particularly interesting at large masses. Large masses correspond to small values of the variance $S = \sigma^2(M)$. Each of the integrals over $\mathrm{d}S_i, \mathrm{d}S_j, \mathrm{d}S_k$ must therefore be performed over an interval $[0, S_n]$ that shrinks to zero as $S_n \to 0$. In this limit it is not necessary to take into account the exact functional form of $\langle \delta(S_i) \delta(S_j) \delta(S_k) \rangle$. Rather, to lowest order we can replace it simply by $\langle \delta^3(S_n) \rangle$. More generally, we can expand the three-point correlator in a triple Taylor series around the point $S_i = S_j = S_k = S_n$. We introduce the notation

$$G_3^{(p,q,r)}(S_n) \equiv \left[\frac{\mathrm{d}^p}{\mathrm{d}S_i^p}\frac{\mathrm{d}^q}{\mathrm{d}S_j^q}\frac{\mathrm{d}^r}{\mathrm{d}S_k^r}\langle\delta(S_i)\delta(S_j)\delta(S_k)\rangle\right]_{S_i=S_j=S_k=S_n}. \tag{756}$$

Then

$$\langle\delta(S_i)\delta(S_j)\delta(S_k)\rangle = \sum_{p,q,r=0}^{\infty}\frac{(-1)^{p+q+r}}{p!q!r!}(S_n-S_i)^p(S_n-S_j)^q(S_n-S_k)^r G_3^{(p,q,r)}(S_n)\,. \tag{757}$$

We expect that the leading contribution to the halo mass function will be given by the term in with $p = q = r = 0$. The leading term in $\Pi^{(3)}$ is therefore

$$\Pi_\epsilon^{(3)}(\delta_0;\delta_n;S_n) = -\frac{\langle\delta_n^3\rangle}{6}\sum_{i,j,k=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots d\delta_{n-1}\partial_i\partial_j\partial_k W^{\mathrm{gm}}\,. \tag{758}$$

This expression can be computed very easily by making use of the following trick. Namely, we consider the derivative of $\Pi_\epsilon^{\mathrm{gm}}$ with respect to δ_c. The first derivative with respect to δ_c can be written as

$$\frac{\partial}{\partial\delta_c}\Pi_\epsilon^{\mathrm{gm}}(\delta_0;\delta_n;S_n) = \sum_{i=1}^{n-1}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_{n-1}\,\partial_i W^{\mathrm{gm}}\,, \tag{759}$$

since, when $\partial/\partial\delta_c$ acts on the upper integration limit of the integral over $\mathrm{d}\delta_i$, it produces $W(\delta_1,\ldots,\delta_i = \delta_c,\ldots,\delta_n;S_n)$, which is the same as the integral of $\partial_i W$ with respect to $\mathrm{d}\delta_i$ from $\delta_i = -\infty$ to $\delta_i = \delta_c$. Similarly

$$\frac{\partial^2}{\partial\delta_c^2}\Pi_\epsilon^{\mathrm{gm}}(\delta_0;\delta_n;S_n) = \sum_{i,j=1}^{n-1}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_{n-1}\,\partial_i\partial_j W^{\mathrm{gm}}\,, \tag{760}$$

In the same way we find that

$$\frac{\partial^3}{\partial\delta_c^3}\Pi_\epsilon^{\mathrm{gm}}(\delta_0;\delta_n;S_n) = \sum_{i,j,k=1}^{n-1}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_{n-1}\,\partial_i\partial_j\partial_k W^{\mathrm{gm}}\,. \tag{761}$$

The right-hand side of this identity is not yet equal to the quantity that appears in Eq. (758), since there the sums run up to n while the above identities only run up to $n-1$. However, what we need is not really $\Pi_\epsilon^{(3)}(\delta_0;\delta_n;S_n)$, but rather its integral over $\mathrm{d}\delta_n$. Then we consider

$$\int_{-\infty}^{\delta_c}\mathrm{d}\delta_n\,\Pi_\epsilon^{(3,\mathrm{L})}(\delta_0;\delta_n;S_n) = -\frac{1}{6}\,\langle\delta_n^3\rangle\sum_{i,j,k=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots d\delta_{n-1}\mathrm{d}\delta_n\partial_i\partial_j\partial_k W^{\mathrm{gm}}\,, \tag{762}$$

and we can now use the identity

$$\begin{aligned}\sum_{i,j,k=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_{n-1}\mathrm{d}\delta_n\,\partial_i\partial_j\partial_k W^{\mathrm{gm}} &= \frac{\partial^3}{\partial\delta_c^3}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_{n-1}d\delta_n\,W^{\mathrm{gm}}\\ &= \frac{\partial^3}{\partial\delta_c^3}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_n\;\Pi_\epsilon^{\mathrm{gm}}(\delta_0;\delta_n;S_n)\,,\end{aligned} \tag{763}$$

so

$$\int_{-\infty}^{\delta_c} \mathrm{d}\delta_n\, \Pi_\epsilon^{(3}(\delta_0;\delta_n;S_n) = -\frac{\langle\delta_n^3\rangle}{6}\frac{\partial^3}{\partial\delta_c^3}\int_{-\infty}^{\delta_c} \mathrm{d}\delta_n\, \Pi_\epsilon^{\rm gm}(\delta_0;\delta_n;S_n)\,. \tag{764}$$

Using

$$\Pi_{\epsilon=0}^{\rm gm}(\delta_0=0;\delta_n;S_n) = \frac{1}{\sqrt{2\pi S_n}}\,[e^{-\delta_n^2/(2S_n)} - e^{-(2\delta_c-\delta_n)^2/(2S_n)}]\,, \tag{765}$$

we immediately find the result in the continuum limit,

$$\int_{-\infty}^{\delta_c} d\delta_n\, \Pi_{\epsilon=0}^{(3)}(0;\delta_n;S_n) = \frac{\langle\delta_n^3\rangle}{3\sqrt{2\pi}\,S_n^{3/2}}\left(1-\frac{\delta_c^2}{S_n}\right)\,e^{-\delta_c^2/(2S_n)}\,. \tag{766}$$

We may express the result in terms of the normalized skewness

$$\mathcal{S}_3(\sigma) \equiv \frac{1}{S^2}\langle\delta^3(S)\rangle \simeq \frac{2.4\times 10^{-4}}{S^{0.45}} f_{\rm NL}\,. \tag{767}$$

Putting the contribution of $\Pi^{(3)}$ together with the gaussian contribution, we find

$$f(\sigma) = \left(\frac{2}{\pi}\right)^{1/2}\,\frac{\delta_c}{\sigma}\,e^{-\delta_c^2/(2\sigma^2)}\left\{1+\frac{\sigma^2}{6\delta_c}\left[\mathcal{S}_3(\sigma)\left(\frac{\delta_c^4}{\sigma^4}-\frac{2\delta_c^2}{\sigma^2}-1\right)+\frac{\mathrm{d}\mathcal{S}_3}{\mathrm{d}\ln\sigma}\,\left(\frac{\delta_c^2}{\sigma^2}-1\right)\right]\right\}\,. \tag{768}$$

The halo mass function in the presence of NG can therefore be written as can be written as

$$\frac{\mathrm{d}n(M)}{\mathrm{d}M} = f(\sigma)\frac{\bar\rho}{M^2}\frac{\mathrm{d}\ln\sigma^{-1}(M)}{\mathrm{d}\ln M} = \left.\frac{\mathrm{d}n(M)}{\mathrm{d}M}\right|_{\rm Gaussian}\left\{1+\frac{\sigma^2}{6\delta_c}\left[\mathcal{S}_3(\sigma)\left(\frac{\delta_c^4}{\sigma^4}-\frac{2\delta_c^2}{\sigma^2}-1\right)+\frac{\mathrm{d}\mathcal{S}_3}{\mathrm{d}\ln\sigma}\,\left(\frac{\delta_c^2}{\sigma^2}-1\right)\right]\right\}\,. \tag{769}$$

From Fig. (26) one sees that the halo mass function is considerably affected by NG for large halo masses (rare events) and at high redshifts.

Part VIII

The impact of the non-Gaussianity on the halo clustering

Let us conclude this set of lectures by studying the impact of NG on the halo clustering. As we have seen in Section 48, in the biased clustering idea, a rare high density fluctuation, corresponding to a massive object, collapses sooner if it lies in a region of large-scale overdensity. This helping hand from the long-wavelength modes means that overdense regions contain an enhanced abundance of massive objects with respect to the mean, so that these systems display enhanced clustering. This is the essence of the peak-background split model. If we decompose a density field into a fluctuating component on galaxy scales, together with a long-wavelength as well, then those regions

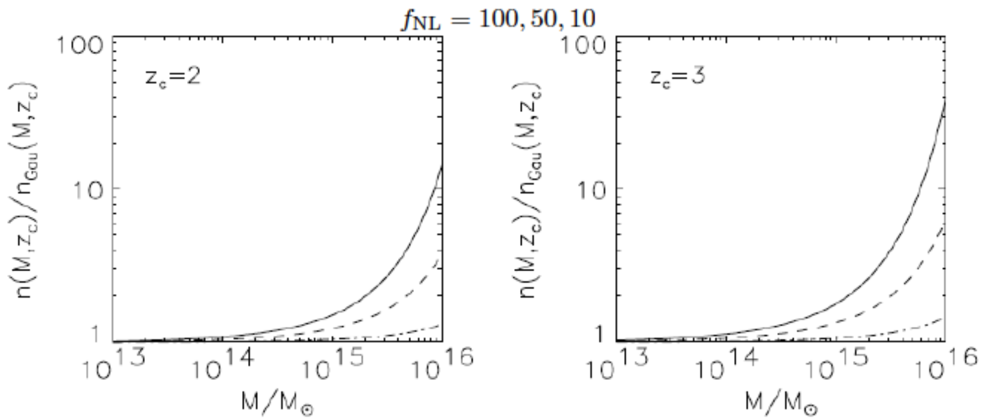

Figure 26: The ratio between the halo mass function with and without NG for three different values of $f_{\rm NL}$.

of density that lie above a threshold in density of ν times the variance σ will be strongly clustered. If proto-objects are presumed to form at the sites of these high peaks, then this is a population with Lagrangian bias, *i.e.* a non-uniform spatial distribution even prior to dynamical evolution of the density field. By extending the classical calculation for calculating the clustering of rare peaks in a Gaussian field to the local type non-Gaussianity, one can show that clustering of rare peaks exhibits a very distinct scale-dependent bias on the largest scales. The analytical result has been tested using N-body simulations, which confirm this basic picture. Following the peak-background split one can split the density field into a long-wavelength piece δ_ℓ and a short-wavelength piece δ_s as in

$$\rho(\mathbf{x}) = \bar{\rho}\,(1 + \delta_\ell + \delta_s)\ . \tag{770}$$

The local Lagrangian number density of haloes $n(\mathbf{x})$ at position $\mathbf{x}$ can then be written as a function of the local value of the long-wavelength perturbation $\delta_\ell(\mathbf{x})$ and the statistics of the short-wavelength fluctuations $P_s(k_s)$. The sufficiently averaged local density of halos follows the large scale matter perturbations, that is its average is a function of $(1 + b_{\rm L}\delta_\ell)$

$$n(\mathbf{x}) = \bar{n}\,(1 + b_{\rm L}\delta_\ell) \tag{771}$$

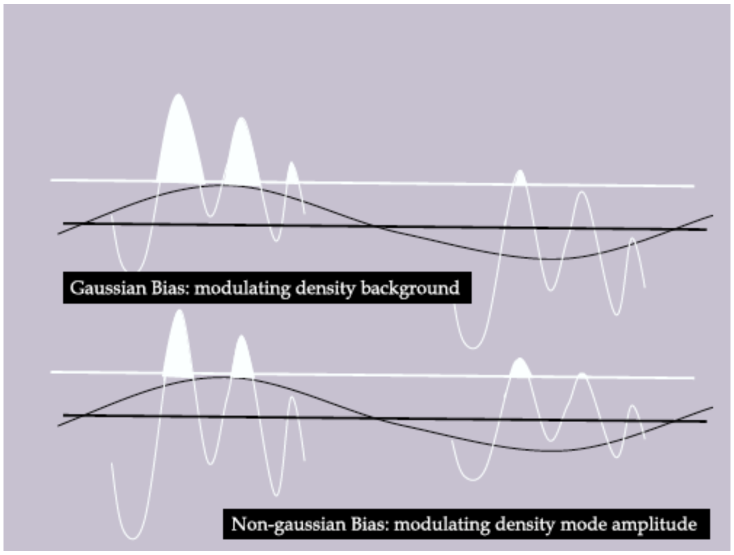

Figure 27: Non-Gaussianity of the local type modulates the peaks.

and so the Lagrangian bias is then

$$b^{\rm L}_{\rm halo} = \bar{n}^{-1}\frac{\partial n}{\partial \delta_\ell}\,. \tag{772}$$

For Eulerian space bias one needs to add the Eulerian space clustering, so the total or Eulerian bias is $b = (b^{\rm L}_{\rm halo} + 1)$. Essentially, in the presence of a long wavelength mode perturbation δ_ℓ it is easier to form a halo, the barrier value is no longer $\delta_{\rm c}$, but $(\delta_{\rm c} - \delta_\ell)$. The corresponding number of halos is therefore shifted

$$n \rightarrow n - \frac{{\rm d}n}{{\rm d}\delta_{\rm c}}\delta_\ell\,. \tag{773}$$

This leads to a revised density

$$\rho \to \overline{n}\left(1+\frac{\delta n}{\overline{n}}\right)(1+\delta_\ell) \tag{774}$$

where the first piece comes from the change in the halo number density and the second directly from the large scale mode. To first order this leads to the Eulerian bias

$$\delta_s = \left(1+\frac{\delta n}{\overline{n}}\right)\delta_\ell \;\text{ and }\; b = 1 - \frac{\mathrm{d}\ln n}{\mathrm{d}\delta_\mathrm{c}}\,. \tag{775}$$

This argument leads to a generically scale-independent bias at sufficiently large scales. The specific function $b(M)$ is obtained by constructing a specific function $n[\delta_\ell(\mathbf{x}), P_s(k_s); M]$, generally fit to simulations, and then differentiating it.

The non-Gaussian case is complicated by the fact that large and small-scale density fluctuations are no longer independent. Instead, one may separate long- and short-wavelength Gaussian potential fluctuations,

$$\Phi = \Phi_\ell + \Phi_s, \tag{776}$$

which are independent. For the local non-Gaussian potential fluctuations,

$$\Phi = \Phi_s + \Phi_\ell + f_\mathrm{NL}\Phi_\ell^2 + (1+2f_\mathrm{NL}\Phi_\ell)\Phi_s + f_\mathrm{NL}\Phi_s^2\,. \tag{777}$$

We can then convert this to a density field using the expression $\delta(k) = \alpha(k)\Phi(k)$, with

$$\alpha(k) = \frac{2k^2T(k)D(z)}{3\Omega_\mathrm{m}H_0^2}. \tag{778}$$

Here $T(k)$ is the transfer function, $D(z)$ the linear growth factor normalised to be $(1+z)^{-1}$ in the matter domination, Ω_m the matter density today and H_0 the Hubble parameter today. The operator $\alpha(k)$ makes it non-local on scales of $\sim$ 100 Mpc, so this can also be thought of as a convolution operator in real space. For long-wavelength modes of the density field, one may write

$$\delta_\ell(k) = \alpha(k)\Phi_\ell(k)\,. \tag{779}$$

The remaining terms in Equation (777) are either much smaller like $f_\mathrm{NL}\Phi_\ell^2$, have only short-wavelength pieces like $(1+2f_\mathrm{NL}\Phi_\ell)\Phi_s$, or simply add a small white noise contribution on large scales, like $f_\mathrm{NL}\Phi_s^2$.

Within a region of given large-scale over-density δ_ℓ and potential Φ_ℓ, the short-wavelength modes of the density field are

$$\delta_s = \alpha\left[(1+2f_\mathrm{NL}\Phi_\ell)\Phi_s + f_\mathrm{NL}\Phi_s^2\right]\,. \tag{780}$$

This is a special case of

$$\delta_s = \alpha\left[X_1\Phi_s + X_2\Phi_s^2\right]\,, \tag{781}$$

where $X_1 = 1 + 2f_{\rm NL}\Phi_\ell$ and $X_2 = f_{\rm NL}$. In the non-Gaussian case, the local number density of haloes of mass M is a function of not just δ_ℓ, but also X_1 and X_2: $n[\delta_\ell, X_1, X_2; P_s(k_s); M]$. The halo bias is then

$$b^{\rm L}_{\rm halo}(M,k) = \bar{n}^{-1}\left[\frac{\partial n}{\partial\delta_\ell(\mathbf{x})} + 2f_{\rm NL}\frac{{\rm d}\Phi_\ell(k)}{{\rm d}\delta_\ell(k)}\frac{\partial n}{\partial X_1}\right]\,, \tag{782}$$

where the derivative is taken at the mean value $X_1 = 1$. There is no X_2 term since X_2 is not spatially variable. The first term here is the usual Gaussian bias, which has no dependence on k.

Equation (781) shows that the effect on non-Gaussianity is a local rescaling of amplitude of (small scale) matter fluctuations. To keep the cosmologist's intuition we write this in terms of σ_8:

$$\sigma_8^{\rm local}(\mathbf{x}) = \sigma_8 X_1(\mathbf{x})\,, \tag{783}$$

so $\delta\sigma_8^{\rm local} = \sigma_8\delta X_1$. This allows us to rewrite Equation (782) as

$$b^{\rm L}_{\rm halo}(M,k) = b_{\rm L}^{\rm Gaussian}(M) + 2f_{\rm NL}\frac{{\rm d}\Phi_\ell(k)}{{\rm d}\delta_\ell(k)}\frac{\partial\ln n}{\partial\ln\sigma_8^{\rm local}}\,. \tag{784}$$

Substituting in ${\rm d}\Phi_\ell(k)/{\rm d}\delta_\ell(k) = \alpha^{-1}(k)$ and dropping the local label, we find

$$\Delta b(M,k) = \frac{3\Omega_{\rm m}H_0^2}{k^2T(k)D(z)}f_{\rm NL}\frac{\partial\ln n}{\partial\ln\sigma_8}\,. \tag{785}$$

This formula is extremely useful because it applies to the bias of any type of object and is expressible entirely in terms of quantities in Gaussian cosmologies, which have received enormous attention from N-body simulators. Within the peak-background split model, the task of performing non-Gaussian calculations is thus reduced to an ensemble of Gaussian simulations with varying amplitude of matter fluctuations.

We now apply Eq. (785) to halo abundance models with a universal mass function. Universal mass functions are those that depend only significance $\nu(M)$, *i.e.*

$$n(M) = n(M,\nu) = 2M^{-2}\nu^2 f(\nu)\frac{{\rm d}\ln\nu}{{\rm d}\ln M}\,, \tag{786}$$

where, as usual, we have defined $\nu = \delta_c/\sigma(M)$ and $f(\nu)$ is the fraction of mass that collapses into haloes of significance between ν and $(\nu + d\nu)$. Universality of the halo mass function has been tested in numerous simulations, with results generally confirming the assumption even if the specific functional forms for $f(\nu)$ may differ from one another.

The significance of a halo of mass M depends on the background density field δ_ℓ, so one can compute $\partial n/\partial\delta_\ell(\mathbf{x})$ to compute the bias

$$b^{\rm L}_{\rm halo} = 1 - \frac{1}{\delta_c}\nu\frac{\rm d}{{\rm d}\nu}\ln[\nu^2 f(\nu)]\,. \tag{787}$$

The derivative $\partial\ln n/\partial\ln\sigma_8$ appearing in Equation (785) can be obtained under the same universality assumption. In fact, the calculation is simpler. The definition of the significance implies $\nu^2 \propto \sigma_8^{-2}$, so that $\mathrm{d}\ln\nu/\mathrm{d}\ln M$ does not depend on σ_8 at fixed M. Therefore $n \propto \nu^2 f(\nu)$ and

$$\frac{\partial \ln n}{\partial \ln \sigma_8} = \frac{\partial \ln \nu}{\partial \ln \sigma_8}\frac{\partial \ln[\nu^2 f(\nu)]}{\partial \ln \nu} = -\nu \frac{\mathrm{d}}{d\nu}\ln[\nu^2 f(\nu)]\,. \tag{788}$$

Thus by comparison to Equation (787), we find

$$\Delta b(M,k) = 3 f_{\rm NL}(b^{\rm L}_{\rm halo} - 1)\delta_c \frac{\Omega_{\rm m}}{k^2 T(k) D(z)} H_0^2\,. \tag{789}$$

The strong $1/k^2$ dependence on large scales of the halo bias is a prediction of the local models of NG and can help to measure values of $f_{\rm NL}$ of order of unity.

This concludes this series of lectures on non-Gaussianity. As we mentioned at the beginning, they are not intended at all to be complete. Th reader is invited to consult more literature on the subject if interested.

Part IX

Exercises

Exercise 1: Determine the inflationary prediction for a model of inflation with potential $V(\phi) = \lambda\phi^4$.

Exercise 2: In one-single field models of inflation relate the prediction for the tensor-to-scalar ratio r to the field excursion $\Delta\phi$ in units of $M_{\rm Pl}$.

Exercise 3: Show that gravity waves are not sourced by the scalar field during inflation.

Exercise 4: By corse-graining the inflaton field on a scale (aH) compute the equation of motion for the long wavelength part of the field and compute its variance. This approach is called stochastic inflation.

Exercise 5: Photons are no produced during inflation. Find why.

Exercise 6: Describe the non-Gaussianity generated in the modulated decay scenario, that is when the decay rate of the inflaton field depends on a light field and its fluctuations.

Exercise 7: Discuss the origin in configuration space of the various shapes of NG.

Exercise 8: Extend the SY inequality to higher-order correlators.

Exercise 9: Compute through the in-in formalism the four-point correlator of a light field σ with potential $\lambda\sigma^4/4$.

Exercise 10: Compute the CMB (S/N) ratio for the equilateral bispectrum.

Exercise 11: Discuss the (S/N) ratio for the local NG coming from the bias.