INFLATION AND REHEATING

Juan García-Bellido
Departement de Physique Théorique, Université de Genève, 1211 Genève 04, Switzerland

Abstract
In these lectures I first review the Standard Cosmological Model, based on the hot Big Bang Theory and the Inflationary Paradigm and then describe in detail the event that marked the transition from one to the other, a process known as “reheating”, where a very rich phenomenology may allow us to open a new window into the early Universe. I will describe the recent developments in observational cosmology, mainly the acceleration of the universe, the precise measurements of the microwave background anisotropies, and the formation of structure like galaxies and clusters of galaxies from tiny primordial density fluctuations. Then I will address the main difficulties of the Big Bang Theory with initial conditions and the surprisingly elegant resolution in terms of the Inflationary Paradigm. I will go in detail over the ADM 3+1 splitting of cosmological spacetimes and the Hamilton-Jacobi formalism behind the slow-roll approximation of homogeneous scalar field cosmology during inflation. I will describe the process of particle production in quantum field theory in curved S.T. and explain the origin of the primordial spectrum of density perturbations and gravitational waves from scalar and tensor metric fluctuations. The reheating of the Universe soon after the end of inflation is an extremely violent process where the dominant vacuum energy during inflation gets transferred to relativistic particles, which rescatter and backreact until they thermalize at a high temperature. The process is very rich - non-thermal phase transitions, production of topological defects, black holes and dark matter, baryogenesis, primordial magnetic fields and gravitational waves - and could have left their signature in the present Universe.

1 INTRODUCTION
The last ten years have seen the coming of age of Modern Cosmology, a mature branch of science based on the hot Big Bang theory and the Inflationary Paradigm. In particular, we can now define rather precisely a Standard Model of Cosmology, where the basic parameters are determined within small uncertainties, of just a few percent, thanks to a host of experiments and observations. This precision era of cosmology has become possible thanks to important experimental developments in all fronts, from measurements of supernovae at high redshifts to the microwave background anisotropies, as well as to the distribution of matter in galaxies and clusters of galaxies. In the near future we will have new observational tools, like neutrino, high energy cosmic ray and gravitational wave astronomy to help explore the very Early Universe, where many new phenomena could have left their signature in the present Universe.

In these lecture notes I will first introduce the basic concepts and equations associated with the hot Big Bang cosmology, defining the main cosmological parameters and their corresponding relationships. Then I will address in detail the three fundamental observations that have shaped our present knowledge: the recent acceleration of the universe, the distribution of matter on large scales and the anisotropies in the microwave background. Together these observations allow the precise determination of a handful of cosmological parameters, in the context of the $\Lambda$ cold dark matter paradigm.

I will then present the shortcomings of the hot Big Bang theory, mainly initial conditions problems, which will lead to the introduction of inflationary cosmology as an extremely elegant solution of those
problems. For a proper discussion of homogeneous scalar field dynamics in an expanding universe I will introduce the Hamilton-Jacobi formalism in the context of the ADM 3+1 slicing, which will allow us to formulate the slow-roll approximation and introduce the gauge-invariant linear perturbation theory as the origin of perturbations in the present Universe.

I will describe briefly the basic concepts of quantum field theory in curved space-times and recall the “alarming phenomenon” of particle production in an expanding universe, envisioned in 1939 by Erwin Schrödinger. I will describe in depth the transition from quantum fluctuations to classical metric perturbations and apply it to scalar and tensor fluctuations produced during inflation. I will deduce the predicted primordial power spectra.

The next lecture will be devoted to the phenomenological signatures of inflation, with a description of a generic inflationary model building based on high energy particle physics, and the predicted spectral tilts and tensor-to-scalar ratios, as well as the origin of non-gaussianities. Then I will concentrate on the observational signatures of inflation in the CMB temperature and polarization anisotropies, specially the Sachs-Wolfe plateau, the acoustic oscillations and the temperature-polarization cross-correlations. I will then describe the transition from metric perturbations to large scale structures in our Universe, with emphasis on the galaxy power spectrum, the concept of Jeans length and Baryon Acoustic Oscillation scale, to end with a discussion of the extraordinary agreement between inflationary predictions and present cosmological observations.

Subsequent lectures will describe the violent process of reheating after inflation, with an explicit connection to high energy particle physics, in the masses, coupling and decay rates of particles that are created directly or indirectly from the inflaton decays at the end of inflation. I will differentiate between the perturbative decay of the inflaton, and a naive computation of the reheating temperature, and the non-perturbative phenomena associated with the Bose condensate nature of the oscillating inflaton at the end of inflation. I will present an account of QFT in background fields and in particular on the parametric amplification of the fields’ amplitudes, in the narrow and broad resonance regime. Then I will describe the violent phenomenon of particle production in hybrid or tachyonic preheating, associated with spinodal instabilities and symmetry breaking.

The final lecture will be devoted to the phenomenological signatures of reheating after inflation, with emphasis on the generation of entropy, the production of dark matter out of equilibrium, of non-thermal phase transitions and topological defects, of the possibility of electroweak baryogenesis, of the generation of primordial magnetic fields and a generic stochastic background of gravitational waves. I will make special emphasis on the specific signatures that this new and rich phenomenology might have in the present Universe which may allow us to open a new window into the physics of the early Universe.

2 FRIEDMANN-ROBERTSON-WALKER COSMOLOGY
Our present understanding of the universe is based upon the successful hot Big Bang theory, which explains its evolution from the first fraction of a second to our present age, around 13.6 billion years later. This theory rests upon four robust pillars, a theoretical framework based on general relativity, as put forward by Albert Einstein [1] and Alexander A. Friedmann [2] in the 1920s, and three basic observational facts: First, the expansion of the universe, discovered by Edwin P. Hubble [3] in the 1930s, as a recession of galaxies at a speed proportional to their distance from us. Second, the relative abundance of light elements, explained by George Gamow [4] in the 1940s, mainly that of helium, deuterium and lithium, which were cooked from the nuclear reactions that took place at around a second to a few minutes after the Big Bang, when the universe was a few times hotter than the core of the sun. Third, the cosmic microwave background (CMB), the afterglow of the Big Bang, discovered in 1965 by Arno A. Penzias and Robert W. Wilson [5] as a very isotropic blackbody radiation at a temperature of about 3 degrees Kelvin, emitted when the universe was cold enough to form neutral atoms, and photons decoupled from matter, approximately 380,000 years after the Big Bang. Today, these observations are confirmed to
within a few percent accuracy, and have helped establish the hot Big Bang as the preferred model of the universe.

Modern Cosmology begun as a quantitative science with the advent of Einstein’s general relativity and the realization that the geometry of space-time, and thus the general attraction of matter, is determined by the energy content of the universe [6]

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}.$$  \hspace{1cm} (1)

These non-linear equations are simply too difficult to solve without invoking some symmetries of the problem at hand: the universe itself.

2.1 Homogeneity and isotropy

Our models of the universe are based on the idea that the Universe is essentially identical at every space-time point; this is a generalization at the cosmological level of the Copernican Principle, otherwise known as the Cosmological Principle (CP). It is obviously wrong: the center of the sun is very different from interstellar space. We will however assume that the CP is applicable only on the largest scales, where local variations in matter are averaged out. Its validity at large scales is confirmed by present observations of diffuse gamma, X-rays, UV, visible, IR and radio backgrounds, but most importantly from the extraordinary isotropy of the cosmic microwave background (CMB). The irregularities (anisotropies) are only of the order of one part in $10^5$, as shown for the first time by the COsmic Background Explorer (COBE) of NASA en 1992.

From a mathematical point of view, the CP is related to two precise mathematical properties of space: isotropy and homogeneity. In differential geometry, isometries of the metric are related with conserved Killing vectors. A maximally symmetric space has the maximum allowed number of Killing vectors. For a space of $n$-dimensions the maximum number is $n(n+1)/2$. In 4 dimensions this corresponds to the 10 generators of the Poincare algebra (4 translations, 3 rotations and 3 boosts). However, the Universe is not static and therefore we cannot impose time translation invariance. Therefore, we are left with the only requirement that spatial sections should be maximally symmetric, according to the Cosmological Principle. This implies 6 Killing vectors, associated with 3 rotations and 3 translations (i.e. homogeneity and isotropy). In the context of GR, it is said that we can foliate the 4D manifold in $\mathbb{R} \times \Sigma$, where $\mathbb{R}$ represents the temporal direction and $\Sigma$ is a maximally symmetric 3-space.

We can thus chose our metric to be of the form

$$ds^2 = -dt^2 + a^2(t) \gamma_{ij}(x) dx^i dx^j,$$  \hspace{1cm} (2)

where $t$ is a time-like coordinate which labels the cosmological events in the 3-surface $\Sigma(x)$, and the metric $\gamma_{ij}$ is maximally symmetric in $\Sigma$. The function $a(t)$ is known as the scale factor and gives us information about the relative size of spatial sections of $\Sigma$ at time $t$. These coordinates, in which the metric has no cross-terms $dt dx^i$, are known as comoving coordinates, and an observer which remains at $x^i = constant$, is known as a comoving observer. Only comoving observers see the Universe around them as isotropic. In fact, no known astronomical object, like the Earth, the Sun or the Milky Way galaxy, are exactly comoving; they move with peculiar velocities due to the gravitational attraction of the matter concentrations in their vicinity. In the case of the Earth, this peculiar velocity is believed to be responsible for the observed dipolar anisotropy of the CMB, i.e. a conventional Doppler effect.

2.2 The Friedmann-Robertson-Walker metric

Let us come back to the characterization of the Universe geometry. We are interested in maximally symmetric 3-metrics $\gamma_{ij}$. Such metrics satisfy

$$^{(3)}R_{ijkl} = k \left( \gamma_{ik} \gamma_{jl} - \gamma_{il} \gamma_{jk} \right),$$  \hspace{1cm} (3)
where we have introduced, for later convenience, a constant \( k = (3)R/6 \). The 3D Ricci tensor is

\[
(3)R_{ij} = 2 k \gamma_{ij}.
\]  

(4)

If the space \( \Sigma \) is maximally symmetric it is necessarily spherically symmetric. For those spaces the most general metric can be written as

\[
d\sigma^2 = \gamma_{ij} dx^i dx^j = e^{2\beta(r)} dr^2 + r^2 \left( d\theta^2 + \sin^2 \theta d\phi^2 \right).
\]  

(5)

The components of the Ricci tensor associated with such a metric are

\[
(3)R_{11} = \frac{2}{r} \beta',
\]  

(6)

\[
(3)R_{22} = \frac{(3)R_{33}}{\sin^2 \theta} = 1 + e^{-2\beta} (r \beta' - 1),
\]  

(7)

\[
(3)R_{ij} = 0, \ i \neq j.
\]  

(8)

In order for them to be proportional to the metric, see (4), we must impose

\[
\beta(r) = -\frac{1}{2} \ln(1 - kr^2),
\]  

(9)

with which the metric becomes

\[
ds^2 = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - kr^2} + r^2 \left( d\theta^2 + \sin^2 \theta d\phi^2 \right) \right],
\]  

(10)

which is the famous Friedmann-Robertson-Walker metric (FRW). Note that this metric is invariant under the redefinition

\[
k \rightarrow \frac{k}{|k|}, \quad r \rightarrow r \sqrt{|k|}, \quad a \rightarrow \frac{a}{\sqrt{|k|}},
\]  

(11)

so that the only relevant parameter is \( K \equiv k/|k| = \text{sign} \ k \). There are three possible cases. An open Universe, with constant negative curvature, \( K = -1 \); a flat Universe of vanishing spatial curvature, \( K = 0 \); and a closed Universe, with constant positive curvature, \( K = +1 \), see Fig. 1.

Fig. 1: The 2D projections of the 3 possible curvatures in 3D space. From left to right: closed \((K = +1)\), flat \((K = 0)\) and open \((K = -1)\) space. Rather than “flat”, a spatial section with \( K = 0 \) should be called “Euclidean” since its distance element is \( d\sigma^2 = d\chi^2 + \sin^2 \chi d\Omega^2 \).

These spaces \( \Sigma \) have a metric

\[
d\sigma^2 = d\chi^2 + \sin^2 \chi d\Omega^2,
\]

where

\[
r = \sin \chi \equiv \begin{cases} 
\sinh \chi & K = -1 \\
\chi & K = 0 \\
\sin \chi & K = +1 
\end{cases}
\]  

(12)
With the FRW we can calculate the dynamics satisfied by the scale factor $a(t)$, solving the Einstein equations. For this we have to find first the non zero components of the affine connection,

$$
\Gamma^0_{ij} = \frac{\dot{a}}{a} g_{ij}, \quad \Gamma^i_{0j} = \frac{\dot{a}}{a} \delta^i_j, \quad \Gamma^2_{12} = \Gamma^3_{13} = \frac{1}{r},
$$

$$
\Gamma^1_{11} = \frac{Kr_{1}}{1 - Kr_{2}}, \quad \Gamma^1_{22} = -r(1 - Kr_{2}) = \frac{\Gamma^1_{33}}{\sin^2 \theta}, \quad \Gamma^2_{33} = -\sin \theta \cos \theta, \quad \Gamma^3_{23} = \cot \theta.
$$

With which the non zero components of the Ricci tensor can be written as

$$
R_{00} = -3\ddot{a}/a, \quad R_{ij} = (\ddot{a} + 2\dot{a}^2 + 2K) \gamma_{ij},
$$

and the 4D curvature scalar

$$
R = \frac{6}{a^2} \left( \ddot{a} + \dot{a}^2 + K \right).
$$

2.2.1 The matter and energy content of the universe

The most general matter fluid consistent with the assumption of homogeneity and isotropy is a perfect fluid, one in which an observer comoving with the fluid would see the universe around it as isotropic. The energy momentum tensor associated with such a fluid can be written as [6]

$$
T^{\mu\nu} = p g^{\mu\nu} + (p + \rho) u^\mu u^\nu,
$$

where $p(t)$ and $\rho(t)$ are the pressure and energy density of the fluid at a given time in the expansion, as measured by this comoving observer, and $u^\mu$ is the comoving four-velocity, satisfying $u^\mu u_\mu = -1$. For such a comoving observer, the matter content looks isotropic (in its rest frame),

$$
T^{\mu}_{\nu} = \text{diag}(\rho(t), p(t), p(t), p(t)).
$$

Before substituting into the Einstein equations, let us consider the covariant conservation of the energy-momentum tensor in the presence of an expanding universe, $T^{\mu}_{\nu;\mu} = 0$. The $\nu = 0$ component gives

$$
0 = \partial_\mu T^{\mu}_{0} + \Gamma^{\mu}_{\rho\delta} T^{\rho}_{0} - \Gamma^{\lambda}_{\mu\delta} T^{\mu}_{\lambda} = -\ddot{\rho} - 3\frac{\dot{a}}{a}(\rho + p).
$$

In order to find explicit solutions, one has to supplement the conservation equation with an equation of state relating the pressure and the density of the fluid, $p = p(\rho)$. The most relevant fluids in cosmology are barotropic, i.e. fluids whose pressure is linearly proportional to the density, $p = w \rho$, and therefore the speed of sound is constant in those fluids.

We will restrict ourselves in these lectures to three main types of barotropic fluids:

- **Radiation**, with equation of state $p_R = \rho_R/3$, associated with relativistic degrees of freedom (i.e. particles with temperatures much greater than their mass). In this case, the energy density of radiation decays as $\rho_R \sim a^{-4}$ with the expansion of the universe.

- **Matter**, with equation of state $p_M \approx 0$, associated with nonrelativistic degrees of freedom (i.e. particles with temperatures much smaller than their mass). In this case, the energy density of matter decays as $\rho_M \sim a^{-3}$ with the expansion of the universe.

- **Vacuum energy**, with equation of state $p_V = -\rho_V$, associated with quantum vacuum fluctuations. In this case, the vacuum energy density remains constant with the expansion of the universe.
This is all we need in order to solve the Einstein equations. Let us now write the equations of motion of observers comoving with such a fluid in an expanding universe. According to general relativity, these equations can be deduced from the Einstein equations (1), by substituting the FRW metric (10) and the perfect fluid tensor (16). The $\mu = i, \nu = j$ component of the Einstein equations, together with the $\mu = 0, \nu = 0$ component constitute the so-called Friedmann equations,

$$
\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho + \frac{\Lambda}{3} - \frac{K}{a^2},
$$

(19)

$$
\frac{\ddot{a}}{a} = - \frac{4\pi G}{3} (\rho + 3p) + \frac{\Lambda}{3}.
$$

(20)

These equations contain all the relevant dynamics, since the energy conservation equation (18) can be obtained from these.

### 2.3 The Cosmological Parameters

I will now define the most important cosmological parameters. Perhaps the best known is the **Hubble parameter** or rate of expansion today, $H_0 = \dot{a}/a(t_0)$. We can write the Hubble parameter in units of 100 km s$^{-1}$ Mpc$^{-1}$, which can be used to estimate the order of magnitude for the present size and age of the universe,

$$
H_0 \equiv 100 \, h \, \text{km s}^{-1} \text{Mpc}^{-1},
$$

(21)

$$
c \, H_0^{-1} = 3000 \, h^{-1} \, \text{Mpc},
$$

(22)

$$
H_0^{-1} = 9.773 \, h^{-1} \, \text{Gyr}.
$$

(23)

The parameter $h$ was measured to be in the range $0.4 < h < 1$ for decades, and only in the last few years has it been found to lie within 4% of $h = 0.70$. I will discuss those recent measurements in the next Section.

Using the present rate of expansion, one can define a **critical density** $\rho_c$, that which corresponds to a flat universe,

$$
\rho_c \equiv \frac{3H_0^2}{8\pi G} = 1.88 h^2 10^{-29} \, \text{g/cm}^3
$$

(24)

$$
= 2.77 h^{-1} 10^{11} \, M_{\odot}/(h^{-1} \, \text{Mpc})^3
$$

(25)

$$
= 11.26 h^2 \, \text{protons/m}^3,
$$

(26)

where $M_{\odot} = 1.989 \times 10^{33} \, \text{g}$ is a solar mass unit. The critical density $\rho_c$ corresponds to approximately 6 protons per cubic meter, certainly a very dilute fluid!

In terms of the critical density it is possible to define the density parameter

$$
\Omega_0 \equiv \frac{8\pi G}{3H_0^2} \rho(t_0) = \frac{\rho}{\rho_c}(t_0),
$$

(27)

whose sign can be used to determine the spatial (three-)curvature. The name of the critical density is due to the fact that, using the above definition, the Friedmann equation (19) can be written as

$$
\Omega - 1 = \frac{K}{a^2 H^2},
$$

(28)

valid for all times. Thus, the sign of $K$ determines whether $\Omega$ is greater, equal or smaller 1,

$$
\rho < \rho_c \iff \Omega < 1 \iff K = -1 \quad \Sigma \; \text{open},
$$

$$
\rho = \rho_c \iff \Omega = 1 \iff K = 0 \quad \Sigma \; \text{flat},
$$

$$
\rho > \rho_c \iff \Omega > 1 \iff K = +1 \quad \Sigma \; \text{closed}.
$$
The \( \Omega \) parameter, thus determines the geometry of our Universe. For example, from the observations of the CMB anisotropies we can deduce that our Universe has spatial sections very close to Euclidean, \( K = 0 \), and therefore the total density of our Universe is very close to critical, \( \Omega_0 = 1.005 \pm 0.005 \).

Moreover, we can define the individual ratios \( \Omega_i \equiv \rho_i/\rho_c \), for matter, radiation, cosmological constant and even curvature, as

\[
\begin{align*}
\Omega_M &= \frac{8\pi G \rho_M}{3H_0^2} \\
\Omega_R &= \frac{8\pi G \rho_R}{3H_0^2} \\
\Omega_\Lambda &= \frac{\Lambda}{3H_0^2} \\
\Omega_K &= -\frac{K}{a_0^2H_0^2}.
\end{align*}
\] (29)

For instance, we can evaluate today the radiation component \( \Omega_R \), corresponding to relativistic particles, from the density of microwave background photons, \( \rho_{CMB} = \pi^2 k^4 T_{CMB}^4/(15\hbar^3c^3) = 4.5 \times 10^{-34} \text{ g/cm}^3 \), which gives \( \Omega_{CMB} = 2.4 \times 10^{-5} \text{ h}^{-2} \). Three approximately massless neutrinos would contribute a similar amount. Therefore, we can safely neglect the contribution of relativistic particles to the total density of the universe today, which is dominated either by non-relativistic particles (baryons, dark matter or massive neutrinos) or by a cosmological constant, and write the rate of expansion in terms of its value today, as

\[
H^2(a) = H_0^2 \left( \Omega_R \frac{a_0^4}{a^4} + \Omega_M \frac{a_0^3}{a^3} + \Omega_\Lambda + \Omega_K \frac{a_0^2}{a^2} \right). \] (31)

An interesting consequence of these definitions is that one can now write the Friedmann equation today, \( a = a_0 \), as a cosmic sum rule,

\[
1 = \Omega_M + \Omega_\Lambda + \Omega_K, \] (32)

where we have neglected \( \Omega_R \) today. That is, in the context of a FRW universe, the total fraction of matter density, cosmological constant and spatial curvature today must add up to one. For instance, if we measure one of the three components, say the spatial curvature, we can deduce the sum of the other two.

Looking now at the second Friedmann equation (20), we can define another basic parameter, the deceleration parameter,

\[
q_0 = -\frac{a \ddot{a}}{\dot{a}^2}(t_0) = \frac{4\pi G}{3H_0^2} \left[ \rho(t_0) + 3p(t_0) \right], \] (33)

defined so that it is positive for ordinary matter and radiation, expressing the fact that the universe expansion should slow down due to the gravitational attraction of matter. We can write this parameter using the definitions of the density parameter for known and unknown fluids (with density \( \Omega_x \) and arbitrary equation of state \( w_x \)) as

\[
q_0 = \Omega_R + \frac{1}{2} \Omega_M - \Omega_\Lambda + \frac{1}{2} \sum_x (1 + 3w_x) \Omega_x. \] (34)

Uniform expansion corresponds to \( q_0 = 0 \) and requires a cancellation between the matter and vacuum energies. For matter domination, \( q_0 > 0 \), while for vacuum domination, \( q_0 < 0 \). As we will see in a moment, we are at present probing the time dependence of the deceleration parameter and can determine with some accuracy the moment at which the universe went from a decelerating phase, dominated by dark matter, into an acceleration phase at present, which seems to indicate the dominance of some kind of vacuum energy.

### 2.4 Cosmological solutions to the Einstein equations

Given a certain matter and energy composition of the Universe, it is possible to solve the Einstein equations for \( a(t) \). For example, for a flat Universe, \( K = 0 \), dominated by non-relativistic matter, \( \rho = \rho_M \), (i.e. a matter dominated (MD) Universe) the Friedmann equation gives:

\[
\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho_M = \frac{8\pi G}{3} \rho_M (\frac{a_0}{a})^3 \quad \rightarrow \quad a_M(t) \propto t^{2/3}. \] (35)
While for a Universe dominated by relativistic matter, $\rho = \rho_R$, (i.e. a radiation dominated (RD) Universe) we have
\[
\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho_R = \frac{8\pi G}{3} \rho_R^{(0)} \left( \frac{a_0}{a} \right)^4 \quad \rightarrow \quad a_R(t) \propto t^{1/2}.
\] (36)
Note that, in both cases, $\rho \propto t^{-2}$, and thus $H \propto t^{-1}$, valid for any power law expansion, $a(t) \propto t^p$, for all $p$.

On the other hand, for a Universe dominated by vacuum energy, $\rho = \rho_v$, (i.e. a vacuum dominated (VD) Universe)
\[
\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho_v = \frac{\Lambda}{3} \quad \rightarrow \quad a_v(t) \propto \exp(t \sqrt{\Lambda/3}),
\] (37)
and thus the scale factor grows exponentially while the rate of expansion is constant. Clearly in this case, the Universe accelerates: a cosmological constant is a gravitational repulsor. This may seem awkward, but in fact is a consequence of the second law of thermodynamics for adiabatic expansion, see Fig. 2.

### Normal Matter

\[ \rho_1 > \rho_2 \quad p > 0 \]
\[ d(\rho V) + pdV = TdS \]

### Vacuum Energy

\[ \rho_1 = \rho_2 \quad p < 0 \]

Fig. 2: Ordinary matter dilutes as it expands. According to the second law of thermodynamics, its pressure on the piston walls must be positive, since it exerts a force on them and makes work, thus losing energy in the expansion. On the other hand, the vacuum energy density is always the same, by definition, independently of the volume explored. Therefore, according to the second law of thermodynamics, its pressure must be negative and equal in magnitude to the energy density. This negative pressure makes the volume grow progressively more rapidly, which explains the exponential expansion of the Universe dominated by a cosmological constant.

2.5 The $(\Omega_M, \Omega_\Lambda)$ plane

Now that we know that the universe is accelerating, one can parametrize the matter/energy content of the universe with just two components: the matter, characterized by $\Omega_M$, and the vacuum energy $\Omega_\Lambda$. Different values of these two parameters completely specify the universe evolution. It is thus natural to plot the results of observations in the plane $(\Omega_M, \Omega_\Lambda)$, in order to check whether we arrive at a consistent picture of the present universe from several different angles (different sets of cosmological observations).

Moreover, different regions of this plane specify different behaviors of the universe. The boundaries between regions are well defined curves that can be computed for a given model. I will now describe the various regions and boundaries.
Fig. 3: Parameter space $(\Omega_M, \Omega_\Lambda)$. The green (dashed) line $\Omega_\Lambda = 1 - \Omega_M$ corresponds to a flat universe, $\Omega_K = 0$, separating open from closed universes. The blue (dotted) line $\Omega_\Lambda = \Omega_M / 2$ corresponds to uniform expansion, $q_0 = 0$, separating accelerating from decelerating universes. The violet (dot-dashed) line corresponds to critical universes, separating eternal expansion from recollapse in the future. Finally, the red (continuous) lines correspond to $t_0 H_0 = 0.5, 0.6, \ldots, \infty$, beyond which the universe has a bounce.
• **Uniform expansion** \((q_0 = 0)\). Corresponds to the line \(\Omega_A = \Omega_M/2\). Points above this line correspond to universes that are accelerating today, while those below correspond to decelerating universes, in particular the old cosmological model of Einstein-de Sitter (EdS), with \(\Omega_A = 0, \Omega_M = 1\). Since 1998, all the data from Supernovae of type Ia appear above this line, many standard deviations away from EdS universes.

• **Flat universe** \((\Omega_K = 0)\). Corresponds to the line \(\Omega_A = 1 - \Omega_M\). Points to the right of this line correspond to closed universes, while those to the left correspond to open ones. In the last few years we have mounting evidence that the universe is spatially flat (in fact Euclidean).

• **Bounce** \((t_0\dot{H}_0 = \infty)\). Corresponds to a complicated function of \(\Omega_A(\Omega_M)\), normally expressed as an integral equation, where

\[
t_0\dot{H}_0 = \int_0^1 da \left[ 1 + \Omega_M(1/a - 1) + \Omega_A(a^2 - 1) \right]^{-1/2}
\]

is the product of the age of the universe and the present rate of expansion. Points above this line correspond to universes that have contracted in the past and have later rebounced. At present, these universes are ruled out by observations of galaxies and quasars at high redshift (up to \(z = 10\)).

It is worth studying whether we can discard observationally the possibility that our Universe suffered in the past a contraction, before entering the present expansion phase. This case corresponds to a minimum of the scale factor \((\dot{a} = 0 \text{ and } \ddot{a} > 0 \text{ in } z = z_\ast)\),

\[
x^3\Omega_M + x^2\Omega_K + \Omega_A = 0 \\
2\Omega_A > x^3\Omega_M
\]

\[
\Rightarrow \quad x^3 - 3x + 2 = z_\ast^2(z_\ast + 3) \leq \frac{2}{\Omega_M} \tag{38}
\]

In general, these “bounce” cosmologies are ruled out by observations of galaxies and quasars at high redshift. If we take the robust bound on the matter content, \(\Omega_M > 0.05\), one finds \(z_\ast < 3\), while galaxies have been detected up to redshifts \(z \simeq 10\). Moreover, the observation of the microwave background, at \(z_{\text{dec}} \simeq 1100\), definitely rule out these bounce cosmologies.

• **Critical Universe** \((H = \ddot{H} = 0 \text{ and } \dddot{H} > 0)\). Corresponds to the boundary between eternal expansion in the future and recollapse. Since the scale factor is always positive, by definition, if we take \(f(x) = H^2(a)\), with \(x = a_0/a > 0\), and using \(x \frac{d}{dx} = -\frac{1}{H} \frac{d}{dt}\), we have that, for a critical universe, \(x_c\), the function \(f(x)\) must have a minimum at zero, \(f(x_c) = f'(x_c) = 0\) and \(f''(x_c) > 0\), where

\[
f(x) = x^3\Omega_M + x^2\Omega_K + \Omega_A = 0 \tag{39}
\]

\[
f'(x) = 3x^2\Omega_M + 2x\Omega_K = 0 \\
\left\{ \begin{array}{l} x_c = 0 \\ x_c = -2\Omega_K/3\Omega_M > 0 \end{array} \right. \tag{40}
\]

\[
f''(x) = 6x\Omega_M + 2\Omega_K = \left\{ \begin{array}{l} +2\Omega_K > 0 \\ -2\Omega_K > 0 \end{array} \right. \\
\left\{ \begin{array}{l} x_c = 0 \\ x_c = 2|\Omega_K|/3\Omega_M \end{array} \right. \tag{41}
\]

Using now the cosmic sum rule, \((\Omega_M - 1)/\Omega_M = 3(-\Omega_K/3\Omega_M) - 4(\Omega_A/4\Omega_M)\) and some trigonometry, \(
\sin 3u = 3\sin u - 4\sin^3 u \quad \text{with} \quad x_c = 2\sin u \quad \text{we find the solution}
\)

\[
\Omega_A = \left\{ \begin{array}{ll} 0 & \Omega_M \leq 1 \\ 4\Omega_M \sin^3 \left[ \frac{1}{3} \arcsin(1 - \Omega_M^{-1}) \right] & \Omega_M \geq 1 \end{array} \right. \tag{42}
\]

The first solution corresponds to the critical point \(x_c = 0 (a = \infty)\), and \(\Omega_K > 0\), while the second solution corresponds to \(x_c = 2|\Omega_K|/3\Omega_M\) and \(\Omega_K < 0\). Expanding around \(\Omega_M = 1\), we find \(\Omega_A \simeq \frac{4}{27}(\Omega_M - 1)^{3}/\Omega_M^2\), for \(\Omega_M \geq 1\). These critical solutions are asymptotic to the Einstein-de Sitter model \((\Omega_M = 1, \Omega_A = 0)\).
These boundaries, and the regions they delimit, can be seen in Fig. 3, together with the lines of equal \( t_0 H_0 \) values.

In summary, the basic cosmological parameters that are now been hunted by a host of cosmological observations are the following: the present rate of expansion \( H_0 \); the age of the universe \( t_0 \); the deceleration parameter \( q_0 \); the spatial curvature \( \Omega_K \); the matter content \( \Omega_M \); the vacuum energy \( \Omega_\Lambda \); the baryon density \( \Omega_B \); the neutrino density \( \Omega_\nu \), and many other that characterize the perturbations responsible for the large scale structure (LSS) and the CMB anisotropies.

### 2.6 Geodesic movement in FRW metric

In order to determine in which epoch do we live, and learn about the evolution of the Universe, it is necessary to measure distant processes via the properties of light and other particles (neutrinos, cosmic rays, gravitational waves) that reach us from far away, and thus deduce the different fractions of matter and energy that the Universe had in the past, at the time when the photon or the neutrino was emitted. That is, in order to interpret these measurements, we need to consider the geodesic movement of these particles in the (averaged) FRW metric.

Note that the FRW metric has 6 space-like Killing vectors (3 rotations y 3 translations, associated with the isometries of isotropy and homogeneity); However, there is no time-like Killing vector that may give us a notion of conserved energy (in fact, energy is not conserved, it gets diluted in the expansion).

Let us consider a particle non-comoving with the fluid, with 4-velocity

\[
u^\mu = \frac{dx^\mu}{ds} = (u^0, u^i) = (\gamma, \gamma v^i), \quad (43)
\]

and non vanishing peculiar velocity, \( v^i = \frac{dx^i}{dt} \rightarrow |v|^2 = g_{ij}v^iv^j \) and \( \gamma = (1 - |v|^2)^{-1/2} \). This particle will follow a geodesic trajectory

\[
\frac{du^\mu}{ds} + \Gamma^\mu_{\nu\lambda} u^\nu u^\lambda = 0. \quad (44)
\]

The \( \mu = 0 \) component is \( \frac{du^0}{ds} + \Gamma^0_{\nu\lambda} u^\nu u^\lambda = 0 \). In the FRW metric, the only non-zero element of the affin connection is \( \Gamma^0_{ij} = \frac{\dot{a}}{a} g^{ij} \), therefore

\[
\frac{du^0}{ds} + \frac{\dot{a}}{a} |u|^2 = 0. \quad (45)
\]

Now, the normalization \( u_\mu u^\mu = -1 \) leads us to \( u_0^2 = 1 + |u|^2 \), and thus \( u^0 du^0 = |u| d|u| \). Finally,

\[
\frac{1}{u^0} \frac{d|u|}{ds} + \frac{\dot{a}}{a} |u| = 0. \quad (46)
\]

Substituting \( u^0 = dt/ds \) and integrating,

\[
|u| \propto \frac{1}{a} \quad \Rightarrow \quad |p| \propto \frac{1}{a}, \quad (47)
\]

that is, the 3-moment magnitude of every particle evolves inversely proportional to the scale factor. For a photon, \( ds^2 = 0 \), we can use the Einstein-De Broglie relation, \( p = \frac{h}{\lambda} \), and write

\[
\frac{\lambda_1}{\lambda_0} = \frac{a(t_1)}{a(t_0)} \quad \Rightarrow \quad z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a_0}{a_1} - 1, \quad (48)
\]

which defines the cosmological redshift parameter, \( z \), and its relation with the scale factor,

\[
a(z) = \frac{a_0}{1 + z}. \quad (49)
\]
An immediate question that arises is this: If photons loose energy in the expansion, where does it go? In the covariant conservation of energy, we saw that the total energy (matter-radiation plus geometry) is conserved, so what happens is that kinetic energy is converted into potential energy; i.e., the energy lost by photons and matter is used to expand the universe. OK, but how did the Universe start expanding in the first place? That is a question we will have to leave for the next chapter of Inflation.

![Fig. 4: Spectra of galaxies at different redshifts](image)

Fig. 4: Spectra of galaxies at different redshifts, from \( z = 0.107 \) to \( z = 2.401 \). The characteristic spectral lines correspond to some elements like Hydrogen and Helium, which are more abundant, and some molecules, like CO, which allow us to determine the cosmological redshift of each galaxy.

In Fig. 4 we show a series of galaxy spectra at different redshifts, together with the restframe spectrum in the lab, with some characteristic lines that permit to compute the frequency shift due to the expansion of the Universe. Note that, although the wavelengths get stretched like in the ordinary Doppler effect, the cosmological redshift is not a Doppler shift: it is the expansion of space-time itself, not the relative velocity between observer and emitter, which is responsible for the redshift of wavelengths. Remember that in curved space-time we cannot perform a parallel transport of velocity vectors in order to compare them, because this depends on the chosen path. However, the redshift is something we can measure. We know, from numerous observations in the laboratory, the wavelengths of a huge number of spectral lines of atoms and molecules that generate the light we observe from distant galaxies, and therefore we can compare and determine how much these wavelengths have changed in the course of expansion since they were emitted, \( \lambda_1 \), until the moment we observe them here on Earth, \( \lambda_2 \). From here we deduce the ratio of the two scale factors, but we cannot determine the times at which they were
emitted: photon spectra do not contain enough information as to allow us to deduce how much coordinate
time has elapsed during its journey of hundreds of Mpc to Earth.

We are assuming, of course, that both the emitted and the restframe spectra are identical, so that we
can actually measure the effect of the intervening expansion, i.e. the growth of the scale factor from \( t_1 \) to
\( t_0 \), when we compare the two spectra. Note however, that if the emitting galaxy and our own participated
in the expansion, i.e. if our measuring rods (our rulers) also expanded with the universe, we would see
no effect! The reason we can measure the redshift of light from a distant galaxy is because our galaxy is
a gravitationally bounded object that has decoupled from the expansion of the universe. It is the distance
between galaxies that changes with time, not the sizes of galaxies, nor the local measuring rods.

2.7 Kinematics in FRW: Hubble’s Law

We can now evaluate the relationship between physical distance and redshift as a function of the rate of
expansion of the universe. In an approximate way, since a photon travels at the speed of light, its time of
flight should be simply the distance covered, measured in lightseconds, but what is the “distance” of a far
away galaxy in an expanding Universe? Let us suppose that we are comoving observers with coordinates
\((r_0, \theta_0, \phi_0)\). We can ask, From what coordinates \((r, \theta, \phi)\) comes a light signal emitted at \( t = t_1 \) and
detected at \( t = t_0 \)? The past light cone can be seen in Fig. 5.

![Fig. 5: The past lightcone of an event at \((t_0, r_0, \theta_0, \phi_0)\) goes through the point \((t_1, r_1, \theta_0, \phi_0)\) where a source emits a photon, which follows a null geodesic to us.]

Because of homogeneity we can always choose our position to be at the origin \( r = 0 \) of our
spatial section. Imagine an object (a star) emitting light at time \( t_1 \), at coordinate distance \( r_1 \) from the
origin. Because of isotropy we can ignore the angular coordinates \((\theta, \phi)\). Then the physical distance,
to first order, will be \( d = a_0 r_1 \). Since light travels along null geodesics, \( ds^2 = 0 \), we can write
\( 0 = -dt^2 + a^2(t) dr^2/(1 - Kr^2) \), and therefore,

\[
\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - Kr^2}} \equiv f(r_1) = \begin{cases} 
\arcsin r_1 & K = 1 \\
1 & K = 0 \\
\text{arcsinh } r_1 & K = -1 
\end{cases} \tag{50}
\]

If we now Taylor expand the scale factor to first order,

\[
\frac{1}{1 + z} = \frac{a(t)}{a_0} = 1 + H_0(t - t_0) + \mathcal{O}(t - t_0)^2,
\]

we find, to first approximation,

\[
r_1 \approx f(r_1) = \frac{1}{a_0} (t_0 - t_1) + \cdots = \frac{z}{a_0 H_0} + \cdots
\]
Putting all together we find the famous Hubble law

\[ H_0 d = a_0 H_0 r_1 = z \simeq v_c, \]

which is just a kinematical effect (we have not included yet any dynamics, i.e. the matter content of the universe). Note that at low redshift \( z \ll 1 \), one is tempted to associate the observed change in wavelength with a Doppler effect due to a hypothetical recession velocity of the distant galaxy. This is only an approximation. In fact, the redshift cannot be ascribed to the relative velocity of the distant galaxy because in general relativity (i.e. in curved spacetimes) one cannot compare velocities through parallel transport, since the value depends on the path! If the distance to the galaxy is small, i.e. \( z \ll 1 \), the physical spacetime is not very different from Minkowsky and such a comparison is approximately valid. As \( z \) becomes of order one, such a relation is manifestly false: galaxies cannot travel at speeds greater than the speed of light; it is the stretching of spacetime which is responsible for the observed redshift.

Hubble’s law has been confirmed by observations ever since the 1920s, with increasing precision, which have allowed cosmologists to determine the Hubble parameter \( H_0 \) with less and less systematic errors. Nowadays, the best determination of the Hubble parameter was made by the Hubble Space Telescope Key Project [8], \( H_0 = 72 \pm 8 \text{ km/s/Mpc} \). This determination is based on objects at distances up to 500 Mpc, corresponding to redshifts \( z \leq 0.1 \).

Nowadays, we are beginning to probe much greater distances, corresponding to \( z \simeq 1 \), thanks to type Ia supernovae. These are white dwarf stars at the end of their life cycle that accrete matter from a companion until they become unstable and violently explode in a natural thermonuclear explosion that out-shines their progenitor galaxy. The intensity of the distant flash varies in time, it takes about three weeks to reach its maximum brightness and then it declines over a period of months. Although the maximum luminosity varies from one supernova to another, depending on their original mass, their environment, etc., there is a pattern: brighter explosions last longer than fainter ones. By studying the characteristic light curves, see Fig. 6, of a reasonably large statistical sample, cosmologists from the Supernova Cosmology Project [7] and the High-redshift Supernova Project [9], are now quite confident that they can use this type of supernova as a standard candle. Since the light coming from some of these rare explosions has travelled a large fraction of the size of the universe, one expects to be able to infer from their distribution the spatial curvature and the rate of expansion of the universe.

The connection between observations of high redshift supernovae and cosmological parameters is done via the luminosity distance, defined as the distance \( d_L \) at which a source of absolute luminosity
(energy emitted per unit time) $\mathcal{L}$ gives a flux (measured energy per unit time and unit area of the detector) $\mathcal{F} = \mathcal{L}/4\pi d_L^2$. One can then evaluate, within a given cosmological model, the expression for $d_L$ as a function of redshift [10],

$$H_0 d_L(z) = \frac{(1+z)}{|\Omega_K|^{1/2}} \sinh \left[ \int_0^z \frac{d z'}{(1+z')^2(1+z'\Omega_M) - z'(2+z')\Omega_\Lambda} \right],$$

(53)

where $\sinh(x) = x$ if $K = 0$; $\sin(x)$ if $K = +1$ and $\sinh(x)$ if $K = -1$, and we have used the cosmic sum rule (32).

Astronomers measure the relative luminosity of a distant object in terms of what they call the effective magnitude, which has a peculiar relation with distance,

$$m(z) \equiv M + 5 \log_{10} \left[ \frac{d_L(z)}{\text{Mpc}} \right] + 25 = M + 5 \log_{10}[H_0 d_L(z)].$$

(54)

Since 1998, several groups have obtained serious evidence that high redshift supernovae appear fainter than expected for either an open ($\Omega_M < 1$) or a flat ($\Omega_M = 1$) universe, see Fig. 7. In fact, the universe

Fig. 7: Upper panel: The Hubble diagram in linear redshift scale. Supernovae with $\Delta z < 0.01$ of each other have been weighted-averaged binned. The solid curve represents the best-fit flat universe model, ($\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$). Two other cosmological models are shown for comparison, ($\Omega_M = 0.25$, $\Omega_\Lambda = 0$) and ($\Omega_M = 1$, $\Omega_\Lambda = 0$). Lower panel: Residuals of the averaged data relative to an empty universe. From Ref. [7].
appears to be accelerating instead of decelerating, as was expected from the general attraction of matter, see Eq. (34); something seems to be acting as a repulsive force on very large scales. The most natural explanation for this is the presence of a cosmological constant, a diffuse vacuum energy that permeates all space and, as explained above, gives the universe an acceleration that tends to separate gravitationally bound systems from each other. The best-fit results from the Supernova Cosmology Project \[11\] give a linear combination
\[
0.8 \Omega_M - 0.6 \Omega_{\Lambda} = -0.16 \pm 0.05 \text{ (1}\sigma\text{)},
\]
which is now many sigma away from an EdS model with \(\Lambda = 0\). In particular, for a flat universe this gives
\[
\Omega_{\Lambda} = 0.71 \pm 0.05 \quad \text{and} \quad \Omega_M = 0.29 \pm 0.05 \text{ (1}\sigma\text{)}.\]
Surprising as it may seem, arguments for a significant dark energy component of the universe were proposed long before these observations, in order to accommodate the ages of globular clusters, as well as a flat universe with a matter content below critical, which was needed in order to explain the observed distribution of galaxies, clusters and voids.

Taylor expanding the scale factor to third order,
\[
\frac{a(t)}{a_0} = 1 + H_0(t - t_0) - \frac{q_0}{2!} H_0^2(t - t_0)^2 + \frac{j_0}{3!} H_0^3(t - t_0)^3 + \mathcal{O}(t - t_0)^4 ,
\]
where
\[
q_0 = -\frac{\ddot{a}}{aH^2}(t_0) = \frac{1}{2} \sum_i (1 + 3w_i) \Omega_i = \frac{1}{2} \Omega_M - \Omega_{\Lambda} ,
\]
\[
j_0 = +\frac{\dddot{a}}{aH^3}(t_0) = \frac{1}{2} \sum_i (1 + 3w_i)(2 + 3w_i) \Omega_i = \Omega_M + \Omega_{\Lambda} ,
\]
are the deceleration and “jerk” parameters. Substituting into Eq. (69) we find
\[
H_0 d_L(z) = z + \frac{1}{2} (1 - q_0) z^2 - \frac{1}{6} (1 - q_0 - 3q_0^2 + j_0) z^3 + \mathcal{O}(z^4) .
\]
This expression goes beyond the leading linear term, corresponding to the Hubble law, into the second and third order terms, which are sensitive to the cosmological parameters \(\Omega_M\) and \(\Omega_{\Lambda}\). It is only recently that cosmological observations have gone far enough back into the early universe that we can begin to probe these terms, see Fig. 8.

This extra component of the critical density would have to resist gravitational collapse, otherwise it would have been detected already as part of the energy in the halos of galaxies. However, if most of the energy of the universe resists gravitational collapse, it is impossible for structure in the universe to grow. This dilemma can be resolved if the hypothetical dark energy was negligible in the past and only recently became the dominant component. According to general relativity, this requires that the dark energy have negative pressure, since the ratio of dark energy to matter density goes like \(a(t)^{-3p/\rho}\). This argument would rule out almost all of the usual suspects, such as cold dark matter, neutrinos, radiation, and kinetic energy, since they all have zero or positive pressure. Thus, we expect something like a cosmological constant, with a negative pressure, \(p \approx -\rho\), to account for the missing energy.

However, if the universe was dominated by dark matter in the past, in order to form structure, and only recently became dominated by dark energy, we must be able to see the effects of the transition from the deceleration into the acceleration phase in the luminosity of distant type Ia supernovae. This has been searched for since 1998, when the first convincing results on the present acceleration appeared. However, only recently \[12\] do we have clear evidence of this transition point in the evolution of the universe. This **coasting point** is defined as the time, or redshift, at which the deceleration parameter vanishes,
\[
q(z) = -1 + (1 + z) \frac{d}{dz} \ln H(z) = 0 ,
\]
where
\[ H(z) = H_0 \left[ \Omega_M (1 + z)^3 + \Omega_x e^{3 \int_0^z (1 + w_x(z')) \frac{dz'}{1+z'}} + \Omega_K (1 + z)^2 \right]^{1/2}, \]
and we have assumed that the dark energy is parametrized by a density \( \Omega_x \) today, with a redshift-dependent equation of state, \( w_x(z) \), not necessarily equal to \(-1\). Of course, in the case of a true cosmological constant, this reduces to the usual expression.

Let us suppose for a moment that the barotropic parameter \( w \) is constant, then the coasting redshift can be determined from

\[ q(z) = q_0 + z \frac{dq_0}{dz} \]

which, in the case of a true cosmological constant, reduces to

\[ z_c = \left( \frac{3|w| - 1}{\Omega_M} \right) \frac{1}{1 + |w|} - 1, \]

When substituting \( \Omega_\Lambda \simeq 0.74 \) and \( \Omega_M \simeq 0.26 \), one obtains \( z_c \simeq 0.8 \), in excellent agreement with recent observations \[12\]. The plane \((\Omega_M, \Omega_\Lambda)\) can be seen in Fig. 9, which shows a significant improvement with respect to previous data.

Now, if we have to live with this vacuum energy, we might as well try to comprehend its origin. For the moment it is a complete mystery, perhaps the biggest mystery we have in physics today \[13\]. We measure its value but we don’t understand why it has the value it has. In fact, if we naively predict it using the rules of quantum mechanics, we find a number that is many (many!) orders of magnitude off the mark. Let us describe this calculation in some detail. In non-gravitational physics, the zero-point energy of the system is irrelevant because forces arise from gradients of potential energies. However, we
know from general relativity that even a constant energy density gravitates. Let us write down the most general energy momentum tensor compatible with the symmetries of the metric and that is covariantly conserved. This is precisely of the form $T^{(\text{vac})}_{\mu\nu} = p_V g_{\mu\nu} = -\rho_V g_{\mu\nu}$, see Fig. 2. Substituting into the Einstein equations (1), we see that the cosmological constant and the vacuum energy are completely equivalent, $\Lambda = 8\pi G \rho_V$, so we can measure the vacuum energy with the observations of the acceleration of the universe, which tells us that $\Omega_\Lambda \simeq 0.7$.

On the other hand, we can estimate the contribution to the vacuum energy coming from the quantum mechanical zero-point energy of the quantum oscillators associated with the fluctuations of all quantum fields,

$$\rho_V^{th} = \sum_i \int_{0}^{\Lambda_{UV}} \frac{d^3k}{(2\pi)^3} \frac{1}{2} h\omega_i(k) = \frac{h\Lambda_{UV}^4}{16\pi^2} \sum_i (-1)^F N_i + O(m_i^2 \Lambda_{UV}^2),$$

where $\Lambda_{UV}$ is the ultraviolet cutoff signaling the scale of new physics. Taking the scale of quantum gravity, $\Lambda_{UV} = M_{Pl}$, as the cutoff, and barring any fortuitous cancellations, then the theoretical expectation
(64) appears to be 120 orders of magnitude larger than the observed vacuum energy associated with the acceleration of the universe,

\[ \rho_{th}^V \simeq 1.4 \times 10^{74} \text{ GeV}^4 = 3.2 \times 10^{91} \text{ g/cm}^3, \quad (65) \]

\[ \rho_{\text{obs}}^V \simeq 0.7 \rho_c = 0.66 \times 10^{-29} \text{ g/cm}^3 = 2.9 \times 10^{-11} \text{ eV}^4. \quad (66) \]

Even if we assumed that the ultraviolet cutoff associated with quantum gravity was as low as the electroweak scale (and thus around the corner, liable to be explored in the LHC), the theoretical expectation would still be 60 orders of magnitude too big. This is by far the worst mismatch between theory and observations in all of science. There must be something seriously wrong in our present understanding of gravity at the most fundamental level. Perhaps we don’t understand the vacuum and its energy does not gravitate after all, or perhaps we need to impose a new principle (or a symmetry) at the quantum gravity level to accommodate such a flagrant mismatch.

In the meantime, one can at least parametrize our ignorance by making variations on the idea of a constant vacuum energy. Let us assume that it actually evolves slowly with time. In that case, we do not expect the equation of state \( p = -\rho \) to remain true, but instead we expect the barotropic parameter \( w(z) \) to depend on redshift. Such phenomenological models have been proposed, and until recently produced results that were compatible with \( w = -1 \) today, but with enough uncertainty to speculate on alternatives to a truly constant vacuum energy. However, with the recent supernovae results [12], there seems to be little space for variations, and models of a time-dependent vacuum energy are less and less favoured. In the near future, the SNAP satellite [14] will measure several thousand supernovae at high redshift and therefore map the redshift dependence of both the dark energy density and its equation of state with great precision. This will allow a much better determination of the cosmological parameters \( \Omega_M \) and \( \Omega_\Lambda \).

2.8 Cosmological distances

Apart from the age of the Universe, it is possible to measure the main cosmological parameters by measuring various distances to far away objects and determining how the Universe has changed since light (or other messengers) was emitted by these objects. Perhaps the most used distance has been the luminosity distance, mentioned in the previous sections.

2.8.1 Luminosity distance

Suppose we have a source that emits light at a distance \( d_L \) from a detector of area \( dA \) on Earth. The luminosity distance \( L \) of the source is nothing more than the distance at which we would place a source that emits that amount of energy per unit time. A standard candle is a luminous object which can be calibrated with some precision and whose absolute luminosity is known or reliably estimated within some systematic error. For example, the "Cepheids" variable stars that Hubble used, or the type Ia supernovae used more recently, are considered by many astronomers as reasonably standard, i.e. whose calibration errors are within acceptable limits. On the other hand, the energy flux \( F \) received in the photon detector is a measure of the energy deposited in such a detector per unit time and unit area. Then we define the luminosity distance \( d_L \) as the radius of a sphere centered on the source, for whom the absolute luminosity of the source \( L \) would give the observed flux in the detector, \( F = L / (4\pi d_L^2) \). In a Universe described by a FRW metric, light travels along null geodesics, \( ds^2 = 0 \), and thus

\[ \frac{a_0 H_0 \, dr}{\sqrt{1 + a_0^2 H_0^2 r^2 \Omega_K}} = \frac{dz}{\sqrt{\Omega_M (1 + z)^3 + \Omega_\Lambda + \Omega_K (1 + z)^2}}, \quad (67) \]

which determines the coordinate distance \( r = r(z, H_0, \Omega_M, \Omega_\Lambda) \), as a function of redshift \( z \) and the other cosmological parameters. Let us consider now the effect of expansion on the flux coming from the distant source at a redshift \( z \) from us. First, the energy of the photon on its way to us has suffered a redshift and thus the observed energy should be \( E_0 = E / (1 + z) \). Then, the rate at which photons arrive
at the detector has suffered time dilation with respect to that at the source, \( dt_0 = (1 + z) dt \). Finally, the surface fraction of the 2-sphere centered at the source covered by the detector is \( dA/(4\pi a_0^2 r^2(z)) \). Thus the total flux at the detector is

\[
\mathcal{F} = \frac{\mathcal{L}}{4\pi a_0^2 (1 + z)^2 r^2(z)} \equiv \frac{\mathcal{L}}{4\pi d_L^2}.
\]  

(68)

The final expression of the luminosity distance \( d_L \) as a function of redshift is given by

\[
H_0 d_L(z) = \frac{(1 + z)}{\Omega_K^{1/2}} \sin \left[ \Omega_K^{1/2} \int_0^z \frac{dz'}{\sqrt{\Omega_M (1 + z')^3 + \Omega_\Lambda + \Omega_K (1 + z')^2}} \right].
\]  

(69)

Let us compute this distance in the two cases of previous sections. For an open universe without cosmological constant, it is possible to compute the integral explicitly,

\[
H_0 d_L(z) = \frac{1}{q_0^2} \left[ 1 - q_0 (1 - z) - (1 - q_0)(1 + 2q_0 z)^{1/2} \right]
\]  

(70)

\[\approx z + \frac{1}{2}(1 - q_0)z^2 + \frac{1}{2}(q_0 - 1)q_0 z^3 + \mathcal{O}(z^4), \]

(71)

where the deceleration parameter is given by \( q_0 = \Omega_M/2 \), and we have expanded to third order.

On the other hand, for a flat universe with cosmological constant, it is possible to write the integral in terms of hypergeometric functions,

\[
H_0 d_L(z) = \frac{2(1 + z)}{\sqrt{\Omega_M}} \left[ F\left(\frac{1}{6}, \frac{1}{2}, \frac{7}{6}; \frac{\Omega_M - 1}{\Omega_M}\right) - \frac{1}{\sqrt{1 + z}} F\left(\frac{1}{6}, \frac{1}{2}, \frac{7}{6}; \frac{\Omega_M - 1}{\Omega_M (1 + z)^3}\right) \right]
\]  

(72)

\[\approx z + \frac{1}{2}(1 - q_0)z^2 + \frac{1}{6}(3q_0^2 + q_0 - 2) z^3 + \mathcal{O}(z^4), \]

(73)

where the deceleration parameter is in this case given by \( q_0 = \Omega_M/2 - \Omega_\Lambda \). Both expressions go beyond first order, corresponding to Hubble law, see Eq. (52), and coincide at second order, being sensitive to the cosmological parameters \( \Omega_M \) and \( \Omega_\Lambda \). Only recently has it been possible, with cosmological observations of distant supernovae, to reach distances (and thus ages of the Universe) sufficiently far to be sensitive to higher order terms in the expansion.

### 2.8.2 Angular diameter distance

Another distance often used in cosmology is the angular diameter distance, when the energy output of a source is unknown (or too variable) like for distant galaxies and quasars, but for whom its physical dimensions are known or estimated, like for example in the case of the baryon acoustic oscillation scale. Thus, suppose we know the proper diameter \( D \) of an extended object. The angle subtended by that source, seen from Earth, will be

\[
\theta = \frac{D}{a(t_1) r_1(z)}
\]  

(74)

where \( t_1 \) is the cosmological time at which the object emitted its light, and \( r_1 \) is its coordinate distance (we occupy the position \( r = 0 \)). Then, we define the angular diameter distance as

\[
d_A(z) \equiv \frac{D}{\theta} = a(t_1) r_1(z) = \frac{a_0 r_1(z)}{1 + z} = \frac{d_L(z)}{(1 + z)^2}, \]

(75)

which can be written in terms of the cosmological parameters \( \Omega_M \) and \( \Omega_\Lambda \) as

\[
H_0 d_A(z) = \frac{(1 + z)^{-1}}{\Omega_K^{1/2}} \sin \left[ \Omega_K^{1/2} \int_0^z \frac{dz'}{\sqrt{\Omega_M (1 + z')^3 + \Omega_\Lambda + \Omega_K (1 + z')^2}} \right],
\]  

(76)
and expanded to second order in redshift,

$$H_0 d_A(z) = z - \frac{1}{2} (3 + q_0) z^2 + O(z^3). \quad (77)$$

Note that this distance has a different redshift dependence than the luminosity distance. Its complementarity can be used to determine the cosmological parameters $\Omega_M$ and $\Omega_\Lambda$.

### 2.8.3 Proper motion distance

When neither the absolute flux nor the absolute size of an object is known or reliably estimated, one can still use its proper motion to determine its distance from us. Suppose we know the absolute speed of an object, $u(t_1)$, at the source’s position, and we measure locally the angular motion across the sky, $\dot{\theta}$. Then the proper motion distance is defined as

$$d_M(z) \equiv \frac{u(t_1)}{\dot{\theta}} = a(t_0) r_1(z) = \frac{d_L(z)}{(1 + z)}, \quad (78)$$

which can be written in terms of the cosmological parameters $\Omega_M$ and $\Omega_\Lambda$ as

$$H_0 d_M(z) = \frac{1}{|\Omega_K|^{1/2}} \sin \left[ |\Omega_K|^{1/2} \int_0^z \frac{dz'}{\sqrt{\Omega_M (1 + z')^3 + \Omega_\Lambda + \Omega_K (1 + z')^2}} \right], \quad (79)$$

and expanded to second order in redshift,

$$H_0 d_M(z) = z - \frac{1}{2} (1 + q_0) z^2 + O(z^3). \quad (80)$$

Note that this distance has a different redshift dependence than the luminosity distance and the angular diameter distance, and thus can be used to determine the cosmological parameters $\Omega_M$ and $\Omega_\Lambda$.

### 2.8.4 Number counts of galaxies, quasars and clusters of galaxies

An extraordinarily simple (geometric) way of determining the cosmological parameters is counting objects (galaxies or other luminous sources like quasars) at large distances. Suppose we have a comoving volume element $dV_c(z)$, and let $dN_{\text{gal}}$ be the number of galaxies within that volume. Assuming we know the number density of galaxies per comoving volume element, as a function of time, $n_c(t)$, we have

$$dN_{\text{gal}} = n_c(t) dV_c = n_c(t) \frac{r^2 dr d\Omega}{\sqrt{1 - K r^2}}. \quad (81)$$

Using (67), we can write the differential counting of galaxies as

$$\frac{dN_{\text{gal}}}{dz d\Omega} = (a_H)^{-3} n_c(z) \left( a_H^2 H_0^2 r^2(z) a_H^2 \frac{dr}{dz} (1 - K r^2)^{-1/2} \right)$$

$$= (a_0 H_0)^{-3} n_c(z) z^2 \left( 1 - 2(1 + q_0) z + \ldots \right). \quad (82)$$

If we now suppose that $n_c(z) \simeq \text{const.}$. – i.e. that galaxies are neither created nor destroyed (which is a more or less reasonable approximation in the redshift range $0.2 < z < 6$), we can use $dN_{\text{gal}}/dz d\Omega$ to determine $q_0$. This way of measuring parameters can be used also with quasars, which allow one to reach even larger distances.
2.9 Cosmological Horizons and conformal time

Since the advent of special relativity we are aware of the fact that there can be events that are not accessible to us, simply because e.g. their distance to us is larger than what it takes light to travel in the time since it was emitted. Since nothing travels faster than the speed of light, this determines the past and future light cones. The maximum distance accessible to us is called a horizon, in analogy with the physical horizon due to the curvature of the Earth, which prevented people on shore to see the approaching ship until it was above the horizon.

In the case of cosmological horizons, the situation is slightly more complicated because, contrary to the case of Minkowski, cosmological space-time (S.T.) is dynamical, and distances between objects are not fixed but change over the course of cosmic time. For example, if the universe is decelerating, light can reach us from distances that are larger than in ordinary Minkowski S.T., see e.g. Fig. 10.

![Fig. 10: The past light cone of an event at time \( t = t_0 \) depends on the dynamics of space-time. In a FRW Universe dominated by radiation, the particle horizon has twice the size of the Minkowski horizon.](image)

2.9.1 Cosmological time as a function of redshift

Let us first compute the non-linear relation between redshift and time. While in Minkowski time and space are univocally determined, this is not true in cosmology. For a known dynamics of the expansion of the Universe (i.e. once known the cosmological parameters and the matter-energy content of the Universe), one can estimate the time at which a photon was emitted from a source at redshift \( z \),

\[
H_0 t(z) = \int_0^{(1+z)^{-1}} \frac{H_0 da}{a H(a)} = \int_0^{(1+z)^{-1}} \frac{da}{\sqrt{\Omega_M/a + \Omega_\Lambda a^2 + \Omega_K}}.
\]  

For an open Universe without cosmological constant (\( \Omega_\Lambda = 0 \)) we find

\[
H_0 t(z) = \frac{1}{(1+z)\sqrt{\Omega_K}} \sqrt{1 + \frac{\Omega_M(1+z)}{\Omega_K}} - \frac{\Omega_M}{\sqrt{\Omega_K^2}} \sinh^{-1} \left( \frac{\Omega_K}{\Omega_M(1+z)} \right)^{1/2},
\]  

which in the limit \( \Omega_K \to 0 \) becomes

\[
H_0 t(z) = \frac{2}{3\sqrt{\Omega_M}} (1+z)^{-3/2}.
\]  

On the other hand, for a Universe with Euclidean spatial sections (\( \Omega_K = 0 \)) we have

\[
H_0 t(z) = \frac{2}{3\sqrt{\Omega_\Lambda}} \sinh^{-1} \left( \frac{\Omega_\Lambda}{\Omega_M(1+z)^3} \right)^{1/2},
\]
which has the same limit (86) as $\Omega_A \to 0$.

For example, let us calculate the cosmic time in which light was emitted by a galaxy at redshift $z = 10$, for a flat Universe with $\Omega_M = 0.3$. Using (87) we find $H_0 t = 0.03335$, i.e. light was emitted when the Universe was 460 Myr old (after the Big Bang), approximately at 3% of the age of the Universe.

2.9.2 Particle Horizon

One of the most important distances in Cosmology is the particle horizon since it gives us the extent of the causal region we can access by the measurement of the most distant particles. The particle horizon is the proper distance travelled by a photon since the origin of the Universe until today,

$$d_{PH}(z) = a(z) \int_0^{t_H} \frac{dr}{\sqrt{1-Kr^2}} = a(t) \int_0^t \frac{dt'}{a(t')} = \left( \frac{a_0}{1+z} \right) \int_0^a \frac{da'}{a'^2H(a')}.$$  

For an open Universe without cosmological constant ($\Omega_A = 0$) we find

$$d_{PH}(z) = \frac{2}{H_0(1+z)} |\Omega_K|^{-1/2} \sin^{-1} \left( \frac{|\Omega_K|}{\Omega_M(1+z)} \right)^{1/2}.$$  

In the limit of flat EdS space, $\Omega_K \to 0$, we have

$$d_{PH} = \frac{2}{H_0 \sqrt{\Omega_M}} (1+z)^{-3/2} = 3t,$$  

which is directly proportional to the distance travelled by light since the origin of the Universe at $t = 0$. The factor 3 (for the matter era) simply indicates that the expansion is decelerated and thus a photon travels more distance than it would in an otherwise flat Minkowski S.T., $d_H = ct$, see Fig. 10.

On the other hand, for a Universe with Euclidean spatial sections ($\Omega_K = 0$) we have

$$d_{PH}(z) = \frac{2}{H_0 \sqrt{\Omega_M}} (1+z)^{-3/2} F \left( \frac{1}{6}, \frac{1}{2}, \frac{7}{6}; \frac{\Omega_M - 1}{\Omega_M(1+z)^3} \right).$$  

In the limit of flat EdS space, $\Omega_M \to 1$, we have

$$d_{PH} = \frac{2}{H_0 \sqrt{\Omega_M}} (1+z)^{-3/2} = 3t,$$  

which naturally coincides with the previous expression since in that limit the two models coincide with the EdS model.

2.9.3 Visible Horizon

In universes like ours, in which the expansion is cooling the hot plasma and particles decouple one after another, as their interactions become slower than the rate of expansion (see Section 3), there is a horizon which is different from the particle horizon, although of similar origin. The visible horizon is the proper distance travelled by a particle, relativistic or not, since the moment it decoupled from the rest of the plasma,

$$d_{VH}(t) = a(t) \int_{t_{dec,\gamma}}^t \frac{dt'}{a(t')}.$$  

and depends on the particle, naturally. For example, photons decoupled when the Universe was approximately $t_{dec,\gamma} \sim 380.000 \text{ yrs old}$, and since then have travelled unimpeded until they reached us. Due to its interactions with the plasma, the dispersion of photons in the charged plasma before decoupling prevents us from reaching beyond this moment, in a similar way as their dispersion in the water vapor of a cloud prevents us from seeing inside the cloud. The Visible Universe is that which comprises the whole
sphere of the visible horizon and its size differs from that of the particle horizon, since the latter takes into
count also the evolution of the Universe since its origin in the Big Bang until photon decoupling. On
the other hand, neutrinos have a visible horizon much larger than that of photons since they decoupled
when the Universe was approximately one second old; while gravitons (i.e. gravitational waves) have an
even larger visible horizon since they decoupled from the plasma at Planck scale, $t_{\text{dec, gw}} \sim 10^{-43}$ s.
For the moment, the only visible horizon we detected was that of photons in the cosmic microwave
background; perhaps in the future we may detect the cosmic neutrino background or even the primordial
gravitational wave background.

2.9.4 Event Horizon

In universes with accelerated expansion, like today or during inflation, two particles that are separated
more than a certain distance will never be in causal contact in the future: the expansion of space-time
will separate them irremissibly. That distance is known as event horizon and can be computed as

$$d_{\text{EH}}(t) = a(t) \int_t^\infty \frac{dt'}{a(t')}.$$  \hfill (94)

In the case of a Universe dominated by a cosmological constant, with constant rate of expansion, $H = \sqrt{\Lambda/3}$, the event horizon is also constant, $d_{\text{EH}} = H^{-1}$. Two particles in that Universe that are separated
more than a distance $H^{-1}$ will never see each other again, unless such a period of exponential expansion
ends and a decelerating stage follows, which can allow events to enter the horizon again, as we believe
happens after inflation, see Section 5.

2.9.5 Conformal time and Hubble radius

The horizon distance characterizes the size of a Universe at a given time. While in Minkowski this
distance is given by the size of the past light cone, see Fig. 10, in a Friedmann-Robertson-Walker space-
time, $a(t) \sim t^p$ with $p < 1$, this distance, $d_H(t) = t/(1-p)$, is significantly larger than the Minkowski
light cone, $d_M = t$. For example, for a flat RD universe, $p = 1/2$ and $d_H = 2t = H^{-1}(t)$, while
for a flat MD universe, $p = 2/3$ and $d_H = 3t = 2H^{-1}(t)$. The peculiar time-dependence of these
distances suggests we compare with the instantaneous rate of expansion, or what is usually called the Hubble radius,

$$d_H(t) \equiv H^{-1}(t)$$

In the case of a RD universe, the horizon distance coincides with the Hubble radius, and this is the reason
one usually interchanges one with the other, but it is important to know the conceptual difference.

On the other hand, the horizon distance (88) is a physical distance. Its corresponding comoving
distance is called the conformal time

$$\eta = \int_0^t \frac{dt'}{a(t')}.$$ \hfill (95)

since the FRW metric, written in terms of this new time coordinate, is conformally equivalent (for flat
universes) to the Minkowski metric,

$$ds^2 = a^2(\eta) \left[ -d\eta^2 + dx^2 \right].$$ \hfill (96)

with the scale factor as conformal factor.

In summary, it is possible to define several cosmological distances: particle and event horizons,
visible horizon and Hubble radius, etc. All these distances are of the order of the Hubble radius of the
Universe, $d_H \sim H^{-1}$, and we will use them indistinguishably, unless stated otherwise.
3 HOT BIG BANG THEORY

The evolution of the Universe is nowadays understood in the context of general relativity, as we have described in the previous chapter, together with a description of the matter and radiation content of the Universe since its origin in terms of perfect fluids with a given equation of state and in thermal equilibrium at a certain temperature. Most of the evolution of the Universe occurred at a rate sufficiently slow that the main components of the fluid expanded adiabatically, except during those exceptional periods (probably the most interesting ones) in which the Universe went through phase transitions, or out of equilibrium processes, accompanied by entropy production. This ensures that the various particle species can be described to a very good approximation as corresponding to an equilibrium distribution, in which the characteristic momentum suffers a redshift displacement as the universe expands, see Eq. (47).

3.1 Thermodynamics of an expanding plasma

In this section I will describe the main concepts associated with ensembles of particles in thermal equilibrium and the brief periods in which the universe fell out of equilibrium. To begin with, let me make contact between the covariant energy conservation law (18) and the second law of thermodynamics,

\[ T dS = dU + p dV , \]

where \( U = \rho V \) is the total energy of the fluid, and \( p = w \rho \) is its barotropic pressure. Taking a comoving volume for the universe, \( V = a^3 \), we find

\[ T \frac{dS}{dt} = \frac{d}{dt}(\rho a^3) + p \frac{d}{dt}(a^3) = 0 , \]

where we have used (18). Therefore, entropy is conserved during the expansion of the universe, \( dS = 0 \); i.e., the expansion is adiabatic even in those epochs in which the equation of state changes, like in the matter-radiation transition (not a proper phase transition). Using (18), we can write

\[ \frac{d}{dt} \ln(\rho a^3) = -3H w . \]

Thus, our universe expands like a gaseous fluid in thermal equilibrium at a temperature \( T \). This temperature decreases like that of any expanding fluid, in a way that is inversely proportional to the cubic root of the volume. This implies that in the past the universe was necessarily denser and hotter. As we go back in time we reach higher and higher temperatures, which implies that the mean energy of plasma particles is larger and thus certain fundamental reactions are now possible and even common, giving rise to processes that today we can only attain in particle physics accelerators. That is the reason why it is so important, for the study of early universe, to know the nature of the fundamental interactions at high energies, and the basic connection between cosmology and high energy particle physics. However, I should clarify a misleading statement that is often used: “high energy particle physics colliders reproduce the early universe” by inducing collisions among relativistic particles. Although the energies of some of the interactions at those collisions reach similar values as those attained in the early universe, the physical conditions are rather different. The interactions within the detectors of the great particle physics accelerators occur typically in the perturbative regime, locally, and very far from thermal equilibrium, lasting a minute fraction of a second; on the other hand, the same interactions occurred within a hot plasma in equilibrium in the early universe while it was expanding adiabatically and its duration could be significantly larger, with a distribution in energy that has nothing to do with those associated with particle accelerators. What is true, of course, is that the fundamental parameters corresponding to those interactions — masses and couplings — are assumed to be the same, and therefore present terrestrial experiments can help us imagine what it could have been like in the early universe, and make predictions about the evolution of the universe, in the context of an expanding plasma a high temperatures and high densities, and in thermal equilibrium.
3.1.1 Fluids in thermal equilibrium

In order to understand the thermodynamical behaviour of a plasma of different species of particles at high temperatures we will consider a gas of particles with \( g \) internal degrees of freedom weakly interacting. The degrees of freedom corresponding to the different particles can be seen in Table 1. For example, leptons and quarks have 4 degrees of freedom since they correspond to the two helicities for both particle and antiparticle. However, the nature of neutrinos is still unknown. If they happen to be Majorana fermions, then they would be their own antiparticle and the number of degrees of freedom would reduce to 2. For photons and gravitons (without mass) their 2 d.o.f. correspond to their states of polarization. The 8 gluons (also without mass) are the gauge bosons responsible for the strong interaction between quarks, and also have 2 d.o.f. each. The vector bosons \( W^\pm \) and \( Z^0 \) are massive and thus, apart from the transverse components of the polarization, they also have longitudinal components.

<table>
<thead>
<tr>
<th>Particle</th>
<th>Spin</th>
<th>Degrees of freedom (( g ))</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higgs</td>
<td>0</td>
<td>1</td>
<td>Massive scalar</td>
</tr>
<tr>
<td>photon</td>
<td>1</td>
<td>2</td>
<td>Massless vector</td>
</tr>
<tr>
<td>graviton</td>
<td>2</td>
<td>2</td>
<td>Massless tensor</td>
</tr>
<tr>
<td>gluon</td>
<td>1</td>
<td>2</td>
<td>Massless vector</td>
</tr>
<tr>
<td>( W ) &amp; ( Z )</td>
<td>1</td>
<td>3</td>
<td>Massive vector</td>
</tr>
<tr>
<td>leptons &amp; quarks</td>
<td>1/2</td>
<td>4</td>
<td>Dirac Fermion</td>
</tr>
<tr>
<td>neutrinos</td>
<td>1/2</td>
<td>4 (2)</td>
<td>Dirac (Majorana) Fermion</td>
</tr>
</tbody>
</table>

Table 1: The internal degrees of freedom of various fundamental particles.

For each of these particles we can compute the number density \( n \), the energy density \( \rho \) and the pressure \( p \), in thermal equilibrium at a given temperature \( T \),

\[
\begin{align*}
    n &= g \int \frac{d^3\mathbf{p}}{(2\pi)^3} f(\mathbf{p}) , \\
    \rho &= g \int \frac{d^3\mathbf{p}}{(2\pi)^3} E(\mathbf{p}) f(\mathbf{p}) , \\
    p &= g \int \frac{d^3\mathbf{p}}{(2\pi)^3} \frac{|\mathbf{p}|^2}{3E} f(\mathbf{p}) ,
\end{align*}
\]

(100) \hspace{1cm} (101) \hspace{1cm} (102)

where the energy is given by \( E^2 = |\mathbf{p}|^2 + m^2 \) and the momentum distribution in thermal (kinetic) equilibrium is

\[
f(\mathbf{p}) = \frac{1}{e^{(E-\mu)/T} \pm 1} \begin{cases} 
-1 & \text{Bose – Einstein} \\
+1 & \text{Fermi – Dirac}
\end{cases}
\]

(103)

The chemical potential \( \mu \) is conserved in these reactions if they are in chemical equilibrium. For example, for reactions of the type \( i + j \leftrightarrow k + l \), we have \( \mu_i + \mu_j = \mu_k + \mu_l \). For example, the chemical potential of the photon vanishes \( \mu_\gamma = 0 \), and thus particles and antiparticles have opposite chemical potentials.

From the equilibrium distributions one can obtain the number density \( n \), the energy \( \rho \) and the
pressure $p$, of a particle of mass $m$ with chemical potential $\mu$ at the temperature $T$,

$$ n = \frac{g}{2\pi^2} \int_m^\infty dE \frac{E(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1}, \quad (104) $$

$$ \rho = \frac{g}{2\pi^2} \int_m^\infty dE \frac{E^2(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1}, \quad (105) $$

$$ p = \frac{g}{6\pi^2} \int_m^\infty dE \frac{(E^2 - m^2)^{3/2}}{e^{(E-\mu)/T} \pm 1}. \quad (106) $$

For a non-degenerate ($\mu \ll T$) relativistic gas ($m \ll T$), we find

$$ n = \frac{g}{2\pi^2} \int_0^\infty dE \frac{E^2}{e^{E/T} \pm 1} = \begin{cases} \frac{\zeta(3)}{\pi^2} g T^3 & \text{Bosons} \\ \frac{3}{4} \frac{\zeta(3)}{\pi^2} g T^3 & \text{Fermions} \end{cases}, \quad (107) $$

$$ \rho = \frac{g}{2\pi^2} \int_0^\infty dE \frac{E^3}{e^{E/T} \pm 1} = \begin{cases} \frac{\pi^2}{30} g T^4 & \text{Bosons} \\ \frac{7}{8} \frac{\pi^2}{30} g T^4 & \text{Fermions} \end{cases}, \quad (108) $$

$$ p = \frac{1}{3} \rho, \quad (109) $$

where $\zeta(3) = 1.20206\ldots$ is the Riemann Zeta function. For relativistic fluids, the energy density per particle is

$$ \langle E \rangle \equiv \frac{\rho}{n} = \begin{cases} \frac{\pi^4}{30\zeta(3)} T \simeq 2.701T & \text{Bosons} \\ \frac{7\pi^4}{180\zeta(3)} T \simeq 3.151T & \text{Fermions} \end{cases} \quad (110) $$

For relativistic bosons or fermions with $\mu < 0$ and $|\mu| < T$, we have

$$ n = \frac{g}{\pi^2} T^3 e^{\mu/T}, \quad (111) $$

$$ \rho = \frac{3g}{\pi^2} T^4 e^{\mu/T}, \quad (112) $$

$$ p = \frac{1}{3} \rho. \quad (113) $$

For a bosonic particle, a positive chemical potential, $\mu > 0$, indicates the presence of a Bose-Einstein condensate, and should be treated separately from the rest of the modes.

On the other hand, for a non-relativistic gas ($m \gg T$), with arbitrary chemical potential $\mu$, we find

$$ n = g \left( \frac{mT}{2\pi} \right)^{3/2} e^{-(m-\mu)/T}, \quad (114) $$

$$ \rho = m n, \quad (115) $$

$$ p = nT \ll \rho. \quad (116) $$

The average energy density per particle is

$$ \langle E \rangle \equiv \frac{\rho}{n} = m + \frac{3}{2} T. \quad (117) $$
Note that, at any given temperature \( T \), the contribution to the energy density of the universe coming from non-relativistic particles in thermal equilibrium is exponentially suppressed with respect to that of relativistic particles, therefore we can write

\[
\rho_R = \frac{\pi^2}{30} g_* T^4, \quad p_R = \frac{1}{3} \rho_R, \quad (118)
\]

\[
g_* (T) = \sum_{\text{bosons}} g_i \left( \frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_i \left( \frac{T_i}{T} \right)^4, \quad (119)
\]

where the factor \( 7/8 \) takes into account the difference between the Fermi and Bose statistics; \( g_* \) is the total number of light d.o.f. \((m \ll T)\), and we have also considered the possibility that particle species \( i \) (bosons or fermions) have an equilibrium distribution at a temperature \( T_i \) different from that of photons, as happens for example when a given relativistic species decouples from the thermal bath, as we will discuss later. This number, \( g_* \), strongly depends on the temperature of the universe, since as it expands and cools, different particles go out of equilibrium or become non-relativistic \((m \gg T)\) and thus become exponentially suppressed from that moment on. A plot of the time evolution of \( g_*(T) \) can be seen in Fig. 11.

![Fig. 11: the light degrees of freedom \( g_* \) and \( g_* S \) as a function of the temperature of the universe. From Ref. [15].](image)

For example, for \( T \ll 1 \text{ MeV} \), i.e. after the time of primordial Big Bang Nucleosynthesis (BBN) and neutrino decoupling, the only relativistic species are the 3 light neutrinos and the photons; since the temperature of the neutrinos is \( T_\nu = (4/11)^{1/3} T_\gamma = 1.90 \text{ K} \), see below, we have \( g_* = 2 + 3 \times \frac{7}{4} \times \left( \frac{4}{11} \right)^{4/3} = 3.36 \), while \( g_* S = 2 + 3 \times \frac{7}{4} \times \left( \frac{4}{11} \right) = 3.91 \).

For \( 1 \text{ MeV} \ll T \ll 100 \text{ MeV} \), i.e. between BBN and the phase transition from a quark-gluon plasma to hadrons and mesons, we have, as relativistic species, apart from neutrinos and photons, also the electrons and positrons, so \( g_* = 2 + 3 \times \frac{7}{4} + 2 \times \frac{7}{4} = 10.75 \).

For \( T \gg 250 \text{ GeV} \), i.e. above the electroweak (EW) symmetry breaking scale, we have one photon (2 polarizations), 8 gluons (massless), the \( W^\pm \) and \( Z^0 \) (massive), 3 families of quarks and leptons, a Higgs (still undiscovered), with which one finds \( g_* = \frac{447}{4} = 106.75 \).
At temperatures well above the electroweak transition we ignore the number of d.o.f. of particles, since we have never explored those energies in particle physics accelerators. Perhaps in the near future, with the results of the Large Hadron Collider (LHC) at CERN, we may may predict the behaviour of the universe at those energy scales. For the moment we even ignore whether the universe was in thermal equilibrium at those temperatures. The highest energy scale at which we can safely say the universe was in thermal equilibrium is that of BBN, i.e. 1 MeV, due to the fact that we observe the present relative abundances of the light element produced at that time. For instance, we can’t even claim that the universe went through the quark-gluon phase transition, at $\sim 200$ MeV, since we have not observed yet any signature of such an event, not to mention the electroweak phase transition, at $\sim 1$ TeV.

Let us now use the relation between the rate of expansion and the temperature of relativistic particles to obtain the time scale of the universe as a function of its temperature,

$$H = 1.66 g_*^{1/2} \frac{T^2}{M_P} = \frac{1}{2t} \implies t = 0.301 g_*^{-1/2} \frac{M_P}{T^2} \sim \left( \frac{T}{\text{MeV}} \right)^{-2} s,$$

thus, e.g. at the EW scale (100 GeV) the universe was just $10^{-10}$ s old, while during the primordial BBN (1 – 0.1 MeV), it was 1 s to 3 min old.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>New particle threshold</th>
<th>$4g_* (T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{\nu} &lt; T &lt; m_e$</td>
<td>$\gamma^* \nu + \nu^* \gamma$</td>
<td>29</td>
</tr>
<tr>
<td>$m_e &lt; T &lt; m_{\mu}$</td>
<td>$e^\pm$</td>
<td>43</td>
</tr>
<tr>
<td>$m_{\mu} &lt; T &lt; m_{\pi}$</td>
<td>$\mu^\pm$</td>
<td>57</td>
</tr>
<tr>
<td>$m_{\pi} &lt; T &lt; T_c$</td>
<td>$\pi^\pm$</td>
<td>69</td>
</tr>
<tr>
<td>$T_c &lt; T &lt; m_s$</td>
<td>$u, \bar{u}, d, \bar{d} + g^* s$</td>
<td>205</td>
</tr>
<tr>
<td>$m_s &lt; T &lt; m_c$</td>
<td>$s, \bar{s}$</td>
<td>247</td>
</tr>
<tr>
<td>$m_c &lt; T &lt; m_{\tau}$</td>
<td>$c, \bar{c}$</td>
<td>289</td>
</tr>
<tr>
<td>$m_{\tau} &lt; T &lt; m_b$</td>
<td>$\tau^\pm$</td>
<td>303</td>
</tr>
<tr>
<td>$m_b &lt; T &lt; m_{W,Z}$</td>
<td>$b, \bar{b}$</td>
<td>345</td>
</tr>
<tr>
<td>$m_{W,Z} &lt; T &lt; m_H$</td>
<td>$W^\pm, Z^0$</td>
<td>381</td>
</tr>
<tr>
<td>$m_H &lt; T &lt; m_t$</td>
<td>$H^0$</td>
<td>385</td>
</tr>
<tr>
<td>$m_t &lt; T$</td>
<td>$t, \bar{t}$</td>
<td>427</td>
</tr>
</tbody>
</table>

Table 2: The effective degrees of freedom $g_* (T)$ of the cosmic fluid as a function of the temperature and depending on the new thresholds for particle production. The temperature $T_c \simeq 200$ MeV corresponds to the QCD transition between the quark-gluon plasma and the hadronic phase. The jump in effective degrees of freedom can be seen in Fig. 11.

### 3.1.2 The entropy of the universe

During most of the history of the universe, the rates of reaction, $\Gamma_{\text{int}}$, of particles in the thermal bath are much bigger than the rate of expansion of the universe, $H$, so that local thermal equilibrium was maintained. In this case, the entropy per comoving volume remained constant. In an expanding universe, the second law of thermodynamics, applied to the element of comoving volume, of unit coordinate volume and physical volume $V = a^3$, can be written as, see (97),

$$T dS = d(\rho V) + p dV = d[(\rho + p)V] - V dp = 0.$$

(121)
Using the Maxwell condition of integrability, \( \frac{\partial^2 S}{\partial T \partial V} = \frac{\partial^2 S}{\partial V \partial T} \), we find that \( dp = (\rho + p) dT/T \), so that

\[
dS = d \left[ (\rho + p) \frac{V}{T} + \text{const} \right],
\]

i.e. the entropy in a comoving volume is \( S = (\rho + p) \frac{V}{T} \), except for a constant. Using now the first law, the covariant conservation of energy, \( T_{\mu \nu}^{; \nu} = 0 \), we have

\[
d \left[ (\rho + p) a^3 \right] = a^3 dp \quad \implies \quad d \left[ (\rho + p) \frac{a^3}{T} \right] = 0,
\]

and thus, in thermal equilibrium, the total entropy in a comoving volume, \( S = a^3 (\rho + p)/T \), is conserved. During most of the evolution of the universe, this entropy was dominated by the contribution from relativistic particles,

\[
S = \frac{2\pi^2}{45} g_{*S} (aT)^3 = \text{const.},
\]

where \( g_{*S} \) is the number of “entropic” degrees of freedom, as we can see in Fig. 11. Above the electron-positron annihilation, all relativistic particles had the same temperature and thus \( g_{*S} = g_* \). It may be also useful to realize that the entropy density, \( s = S/a^3 \), is proportional to the number density of relativistic particles, and in particular to the number density of photons, \( s = 1.80 g_{*S} n_\gamma \); today, \( s = 7.04 n_\gamma \). However, since \( g_{*S} \) in general is a function of temperature, we can’t always interchange \( s \) and \( n_\gamma \).

The conservation of \( S \) implies that the entropy density satisfies \( s \propto a^{-3} \), and thus the physical size of the comoving volume is \( a^3 \propto s^{-1} \); therefore, the number of particles of a given species in a comoving volume, \( N = a^3 n \), is proportional to the number density of that species over the entropy density \( s \),

\[
N \sim \frac{n}{s} = \begin{cases} \frac{45\zeta(3)}{2\pi^4} g_{*S} & T \gg m, \mu \\ \frac{45 g}{4\pi^5 \sqrt{2} g_{*S}} \left( \frac{m}{T} \right)^{3/2} e^{-m/T} & T \ll m \end{cases}
\]

If this number does not change, i.e. if those particles are neither created nor destroyed, then \( n/s \) remains constant. As a useful example, we will consider the barionic number in a comoving volume,

\[
\frac{n_B}{s} \equiv \frac{n_b - n_{\bar{b}}}{s}.
\]

As long as the interactions that violate barion number occur sufficiently slowly, the barionic number per comoving volume, \( n_B/s \), will remain constant. Although

\[
\eta \equiv \frac{n_B}{n_\gamma} = 1.80 g_{*S} \frac{n_B}{s},
\]

the ratio between barion and photon numbers it does not remain constant during the whole evolution of the universe since \( g_{*S} \) varies; e.g. during the annihilation of electrons and positrons, the number of photons per comoving volume, \( N_\gamma = a^3 n_\gamma \), grows a factor \( 11/4 \), and \( \eta \) decreases by the same factor. After this epoch, however, \( g_* \) is constant so that \( \eta \approx 7 n_B/s \) and \( n_B/s \) can be used indistinctly.

Another consequence of Eq. (124) is that \( S = \text{const.} \) implies that the temperature of the universe evolves as

\[
T \propto g_{*S}^{-1/3} a^{-1}.
\]
As long as $g_S S$ remains constant, we recover the well known result that the universe cools as it expands according to $T \propto 1/a$. The factor $g_S^{-1/3}$ appears because when a species becomes non-relativistic (when $T \leq m$), and effectively disappears from the energy density of the universe, its entropy is transferred to the rest of the relativistic particles in the plasma, making $T$ decrease not as quickly as $1/a$, until $g_S S$ again becomes constant.

From the observational fact that the universe expands today one can deduce that in the past it must have been hotter and denser, and that in the future it will be colder and more dilute. Since the ratio of scale factors is determined by the redshift parameter $z$, we can obtain (to very good approximation) the temperature of the universe in the past with

$$T = T_0 (1 + z).$$

(130)

This expression has been spectacularly confirmed thanks to the absorption spectra of distant quasars [16]. These spectra suggest that the radiation background was acting as a thermal bath for the molecules in the interstellar medium with a temperature of 9 K at a redshift $z \sim 2$, and thus that in the past the photon background was hotter than today. Furthermore, observations of the anisotropies in the microwave background confirm that the universe at a redshift $z = 1089$ had a temperature of 0.3 eV, in agreement with Eq. (130).

### 3.2 The thermal evolution of the universe

In a strict mathematical sense, it is impossible for the universe to have been always in thermal equilibrium since the FRW model does not have a timelike Killing vector. In practice, however, we can say that the universe has been most of its history very close to thermal equilibrium. Of course, those periods in which there were deviations from thermal equilibrium have been crucial for its evolution thereafter (e.g. baryogenesis, QCD transition, primordial nucleosynthesis, recombination, etc.); without these the universe today would be very different and probably we would not be here to tell the story.

The key to understand the thermal history of the universe is the comparison between the rates of interaction between particles (microphysics) and the rate of expansion of the universe (macrophysics). Ignoring for the moment the dependence of $g_S$ on temperature, the rate of change of $T$ is given directly by the rate of expansion, $\dot{T}/T = -H$. As long as the local interactions — necessary in order that the particle distribution function adjusts adiabatically to the change of temperature — are sufficiently fast compared with the rate of expansion of the universe, the latter will evolve as a succession of states very close to thermal equilibrium, with a temperature proportional to $a^{-1}$. If we evaluate the interaction rates as

$$\Gamma_{int} \equiv \langle n \sigma | v | \rangle,$$

(131)

where $n(t)$ is the number density of target particles, $\sigma$ is the cross section on the interaction and $v$ is the relative velocity of the reaction, all averaged on a thermal distribution; then a rule of thumb for ensuring that thermal equilibrium is maintained is

$$\Gamma_{int} \gtrsim H.$$  

(132)

This criterium is understandable. Suppose, as often occurs, that the interaction rate in thermal equilibrium is $\Gamma_{int} \propto T^n$, with $n > 2$; then, the number of interactions of a particle after time $t$ is

$$N_{int} = \int_t^{\infty} \Gamma_{int}(t') dt' = \frac{1}{n - 2} \frac{\Gamma_{int}(t)}{H(t)},$$

(133)

to determine that the particle interacts less than once from the moment in which $\Gamma_{int} \approx H$. If $\Gamma_{int} \gtrsim H$, the species remains coupled to the thermal plasma. This doesn’t mean that, necessarily, the particle is out of local thermal equilibrium, since we have seen already that relativistic particles that have decoupled retain their equilibrium distribution, only at a different temperature from that of the rest of the plasma.
In order to obtain an approximate description of the decoupling of a particle species in an expanding universe, let us consider two types of interaction:

i) interactions mediated by massless gauge bosons, like for example the photon. In this case, the cross section for particles with significant momentum transfer can be written as \( \sigma \sim \alpha^2/T^2 \), with \( \alpha = g^2/4\pi \) the coupling constant of the interaction. Assuming local thermal equilibrium, \( n(t) \sim T^3 \) and thus the interaction rate becomes \( \Gamma \sim n \sigma |v| \sim \alpha^2 T \). Therefore,

\[
\frac{\Gamma}{H} \sim \alpha^2 \frac{M_P}{T},
\]

so that for temperatures of the universe \( T \lesssim \alpha^2 M_P \sim 10^{16} \text{ GeV} \), the reactions are fast enough and the plasma is in equilibrium, while for \( T \gtrsim 10^{16} \text{ GeV} \), reactions are too slow to maintain equilibrium and it is said that they are “frozen-out”. An important consequence of this result is that the universe could never have been in thermal equilibrium above the grand unification (GUT) scale.

ii) interactions mediated by massive gauge bosons, e.g. like the \( W^\pm \) and \( Z^0 \), or those responsible for the GUT interactions, \( X \) and \( Y \). We will generically call them \( X \) bosons. The cross section depends rather strongly on the temperature of the plasma,

\[
\sigma \sim \begin{cases} 
G_X^2 T^2 & T \ll M_X \\
\frac{\alpha^2}{T^2} & T \gg M_X
\end{cases}
\]

where \( G_X \sim \alpha/M_X^2 \) is the effective coupling constant of the interaction at energies well below the mass of the vector boson, analogous to the Fermi constant of the electroweak interaction, \( G_F = g^2/(4\sqrt{2}M_W^2) \) at tree level. Note that for \( T \gg M_X \) we recover the result for massless bosons, so we will concentrate here on the other case. For \( T \leq M_X \), the rate of thermal interactions is \( \Gamma \sim n \sigma |v| \sim G_X^2 T^5 \). Therefore,

\[
\frac{\Gamma}{H} \sim G_X^2 M_P T^3,
\]

such that at temperatures in the range

\[
M_X \lesssim T \gtrsim G_X^{-2/3} M_P^{1/3} \sim \left( \frac{M_X}{100 \text{ GeV}} \right)^{4/3} \text{MeV},
\]

reactions occur so fast that the plasma is in thermal equilibrium, while for \( T \lesssim (M_X/100 \text{ GeV})^{4/3} \text{ MeV} \), those reactions are too slow for maintaining equilibrium and they effective freeze-out, see Eq. (133).

3.2.1 The decoupling of relativistic particles

Those relativistic particles that have decoupled from the thermal bath do not participate in the transfer of entropy when the temperature of the universe falls below the mass threshold of a given species \( T \simeq m \); in fact, the temperature of the decoupled relativistic species falls as \( T \propto 1/a \), as we will now show. Suppose that a relativistic particle is initially in local thermal equilibrium, and that it decouples at a temperature \( T_D \) and time \( t_D \). The phase space distribution at the time of decoupling is given by the equilibrium distribution,

\[
f(\mathbf{p},t_D) = \frac{1}{e^{E/T_D} \pm 1}.
\]

After decoupling, the energy of each massless particle suffers redshift, \( E(t) = E_D (a_D/a(t)) \). The number density of particles also decreases, \( n(t) = n_D (a_D/a(t))^3 \). Thus, the phase space distribution at a time \( t > t_D \) is

\[
f(\mathbf{p},t) = \frac{dn}{d^3\mathbf{p}} = f(\mathbf{p} a_D/a_D, t_D) = \frac{1}{e^{E a_D T_D} \pm 1} = \frac{1}{e^{E/T} \pm 1},
\]
so that we conclude that the distribution function of a particle that has decoupled while being relativistic remains self-similar as the universe expands, with a temperature that decreases as

$$T = T_D \frac{a_D}{a} \propto a^{-1},$$

(140)

and not as $g_s S^{1/3} a^{-1}$, like the rest of the plasma in equilibrium (129).

### 3.2.2 The decoupling of non-relativistic particles

Those particles that decoupled from the thermal bath when they were non-relativistic ($m \gg T$) behave differently. Let us study the evolution of the distribution function of a non-relativistic particle that was in local thermal equilibrium at a time $t_D$, when the universe had a temperature $T_D$. The moment of each particle suffers redshift as the universe expands, $|\mathbf{p}| = |\mathbf{p}_D| (a_D/a)$, see Eq. (48). Therefore, their kinetic energy satisfies $E = E_D (a_D/a)^2$. On the other hand, the particle number density also varies, $n(t) = n_D (a_D/a(t))^3$, so that a decoupled non-relativistic particle will have an equilibrium distribution function characterized by a temperature

$$T = T_D \frac{a^2_D}{a^2} \propto a^{-2},$$

(141)

and a chemical potential

$$\mu(t) = m + (\mu_D - m) \frac{T}{T_D},$$

(142)

whose variation is precisely that which is needed for the number density of particle to decrease as $a^{-3}$.

In summary, a particle species that decouples from the thermal bath follows an equilibrium distribution function with a temperature that decreases like $T_R \propto a^{-1}$ for relativistic particles ($T_D \gg m$) or like $T_{NR} \propto a^{-2}$ for non-relativistic particles ($T_D \ll m$). On the other hand, for semi-relativistic particles ($T_D \sim m$), its phase space distribution does not maintain an equilibrium distribution function, and should be computed case by case.

### 3.2.3 Brief thermal history of the universe

I will briefly summarize here the thermal history of the Universe, from the Planck era to the present, see Fig. 12. As we go back in time, the universe becomes hotter and hotter and thus the amount of energy available for particle interactions increases. As a consequence, the nature of interactions goes from those described at low energy by long range gravitational and electromagnetic physics, to atomic physics, nuclear physics, all the way to high energy physics at the electroweak scale, gran unification (perhaps), and finally quantum gravity. The last two are still uncertain since we do not have any experimental evidence for those ultra high energy phenomena, and perhaps Nature has followed a different path.

The way we know about the high energy interactions of matter is via particle accelerators, which are unravelling the details of those fundamental interactions as we increase in energy. However, one should bear in mind that the physical conditions that take place in our high energy colliders are very different from those that occurred in the early universe. These machines could never reproduce the conditions of density and pressure in the rapidly expanding thermal plasma of the early universe. Nevertheless, those experiments are crucial in understanding the nature and rate of the local fundamental interactions available at those energies. What interests cosmologists is the statistical and thermal properties that such a plasma should have, and the role that causal horizons play in the final outcome of the early universe expansion. For instance, of crucial importance is the time at which certain particles decoupled from the plasma, i.e. when their interactions were not quick enough compared with the expansion of the universe, and they were left out of equilibrium with the plasma.

One can trace the evolution of the universe from its origin till today. There is still some speculation about the physics that took place in the universe above the energy scales probed by present colliders.
Nevertheless, the overall layout presented here is a plausible and hopefully testable proposal. According to the best accepted view, the universe must have originated at the Planck era ($10^{19}$ GeV, $10^{-43}$ s) from a quantum gravity fluctuation. Needless to say, we don’t have any experimental evidence for such a statement: Quantum gravity phenomena are still in the realm of physical speculation. However, it is plausible that a primordial era of cosmological inflation originated then. Its consequences will be discussed below. Soon after, the universe may have reached the Grand Unified Theories (GUT) era ($10^{16}$ GeV, $10^{-35}$ s). Quantum fluctuations of the inflaton field most probably left their imprint then as tiny perturbations in an otherwise very homogenous patch of the universe. At the end of inflation, the huge energy density of the inflaton field was converted into particles, which soon thermalized and became the origin of the hot Big Bang as we know it. Such a process is called reheating of the universe. Since then, the universe became radiation dominated. It is probable (although by no means certain) that the asymmetry between matter and antimatter originated at the same time as the rest of the energy of the universe, from the decay of the inflaton. This process is known under the name of baryogenesis since baryons (mostly quarks at that time) must have originated then, from the leftovers of their annihilation with antibaryons. It is a matter of speculation whether baryogenesis could have occurred at energies as low as the electroweak scale ($100$ GeV, $10^{-10}$ s). Note that although particle physics experiments...
have reached energies as high as 100 GeV, we still do not have observational evidence that the universe actually went through the EW phase transition. If confirmed, baryogenesis would constitute another “window” into the early universe. As the universe cooled down, it may have gone through the quark-gluon phase transition ($10^2$ MeV, $10^{-5}$ s), when baryons (mainly protons and neutrons) formed from their constituent quarks.

The furthest window we have on the early universe at the moment is that of primordial nucleosynthesis ($1 - 0.1$ MeV, $1$ s – $3$ min), when protons and neutrons were cold enough that bound systems could form, giving rise to the lightest elements, soon after neutrino decoupling: It is the realm of nuclear physics. The observed relative abundances of light elements are in agreement with the predictions of the hot Big Bang theory. Immediately afterwards, electron-positron annihilation occurs (0.5 MeV, 1 min) and all their energy goes into photons. Much later, at about (1 eV, $\sim 10^5$ yr), matter and radiation have equal energy densities. Soon after, electrons become bound to nuclei to form atoms (0.3 eV, $3 \times 10^5$ yr), in a process known as recombination: It is the realm of atomic physics. Immediately after, photons decouple from the plasma, travelling freely since then. Those are the photons we observe as the cosmic microwave background. Much later ($\sim 1 - 10$ Gyr), the small inhomogeneities generated during inflation have grown, via gravitational collapse, to become galaxies, clusters of galaxies, and superclusters, characterizing the epoch of structure formation. It is the realm of long range gravitational physics, perhaps dominated by a vacuum energy in the form of a cosmological constant. Finally (3K, 13 Gyr), the Sun, the Earth, and biological life originated from previous generations of stars, and from a primordial soup of organic compounds, respectively.

I will now review some of the more robust features of the Hot Big Bang theory of which we have precise observational evidence.

### 3.2.4 Primordial nucleosynthesis and light element abundance

In this subsection I will briefly review Big Bang nucleosynthesis and give the present observational constraints on the amount of baryons in the universe. In 1920 Eddington suggested that the sun might derive its energy from the fusion of hydrogen into helium. The detailed reactions by which stars burn hydrogen were first laid out by Hans Bethe in 1939. Soon afterwards, in 1946, George Gamow realized that similar processes might have occurred also in the hot and dense early universe and gave rise to the first light elements [4]. These processes could take place when the universe had a temperature of around $T_{\text{NS}} \sim 1 - 0.1$ MeV, which is about 100 times the temperature in the core of the Sun, while the density is $\rho_{\text{NS}} = \frac{\pi^2}{30} g_s T_{\text{NS}}^4 \sim 82 \text{ g cm}^{-3}$, about the same density as the core of the Sun. Note, however, that although both processes are driven by identical thermonuclear reactions, the physical conditions in star and Big Bang nucleosynthesis are very different. In the former, gravitational collapse heats up the core of the star and reactions last for billions of years (except in supernova explosions, which last a few minutes and creates all the heavier elements beyond iron), while in the latter the universe expansion cools the hot and dense plasma in just a few minutes. Nevertheless, Gamow reasoned that, although the early period of cosmic expansion was much shorter than the lifetime of a star, there was a large number of free neutrons at that time, so that the lighter elements could be built up quickly by successive neutron captures, starting with the reaction $n + p \rightarrow D + \gamma$. The abundances of the light elements would then be correlated with their neutron capture cross sections, in rough agreement with observations [6, 17].

Nowadays, Big Bang nucleosynthesis (BBN) codes compute a chain of around 30 coupled nuclear reactions [18], to produce all the light elements up to beryllium-7. ¹ Only the first four or five elements can be computed with accuracy better than 1% and compared with cosmological observations. These light elements are $H, ^4He, D, ^3He, ^7Li,$ and perhaps also $^6Li$. Their observed relative abundance to hydrogen is $[1 : 0.25 : 3 \cdot 10^{-5} : 2 \cdot 10^{-5} : 2 \cdot 10^{-10}]$ with various errors, mainly systematic. The BBN codes calculate these abundances using the laboratory measured nuclear reaction rates, the decay rate of

---

¹The rest of nuclei, up to iron (Fe), are produced in heavy stars, and beyond Fe in novae and supernovae explosions.
the neutron, the number of light neutrinos and the homogeneous FRW expansion of the universe, as a function of only one variable, the number density fraction of baryons to photons, \( \eta \equiv n_B/n_\gamma \). In fact, the present observations are only consistent, see Fig. 13 and Ref. [17, 18, 19], with a very narrow range of values of
\[
\eta_{10} \equiv 10^{10} \eta = 6.15 \pm 0.25 .
\]
(143)

Such a small value of \( \eta \) indicates that there is about one baryon per \( 10^9 \) photons in the universe today. Any acceptable theory of baryogenesis should account for such a small number. Furthermore, the present baryon fraction of the critical density can be calculated from \( \eta_{10} \) as
\[
\Omega_B h^2 = 3.6271 \times 10^{-3} \eta_{10} = 0.0224 \pm 0.0018 \quad (95\% \text{ c.l.})
\]
(144)

Clearly, this number is well below closure density, so baryons cannot account for all the matter in the universe, as I shall discuss below.

3.2.5 Neutrino decoupling

Just before the nucleosynthesis of the lightest elements in the early universe, weak interactions were too slow to keep neutrinos in thermal equilibrium with the plasma, so they decoupled. We can estimate the temperature at which decoupling occurred from the weak interaction cross section, \( \sigma_w \approx G_F^2 T^2 \) at finite temperature \( T \), where \( G_F = 1.2 \times 10^{-5} \text{ GeV}^{-2} \) is the Fermi constant. The neutrino interaction rate, via
W boson exchange in $n + \nu \leftrightarrow p + e^-$ and $p + \bar{\nu} \leftrightarrow n + e^+$, can be written as [15]

$$\Gamma_\nu = n_\nu \langle |\sigma_w| v |\rangle \simeq 2.1 G_F^2 T^5,$$

(145)

while the rate of expansion of the universe at that time ($g_* = 10.75$) was $H \simeq 5.4 \; T^2 / M_P$, where $M_P = 1.22 \times 10^{19} \; \text{GeV}$ is the Planck mass. Neutrinos decouple when their interaction rate is slower than the universe expansion, $\Gamma_\nu \leq H$, or, equivalently, at $T_{\nu-dec} \simeq 0.8 \; \text{MeV}$. Below this temperature, neutrinos are no longer in thermal equilibrium with the rest of the plasma, and their temperature continues to decay inversely proportional to the scale factor of the universe. Since neutrinos decoupled before $e^+e^-$ annihilation, the cosmic background of neutrinos has a temperature today lower than that of the microwave background of photons. Let us compute the difference. At temperatures above the mass of the electron, $T > m_e = 0.511 \; \text{MeV}$, and below 0.8 MeV, the only particle species contributing to the entropy of the universe are the photons ($g_* = 2$) and the electron-positron pairs ($g_* = 4 \times 7/2$); total number of degrees of freedom $g_* = 17/2$. At temperatures $T \simeq m_e$, electrons and positrons annihilate into photons, heating up the plasma (but not the neutrinos, which had decoupled already). At temperatures $T < m_e$, only photons contribute to the entropy of the universe, with $g_* = 2$ degrees of freedom. Therefore, from the conservation of entropy, we find that the ratio of $T_\gamma$ and $T_\nu$ today must be

$$\frac{T_\gamma}{T_\nu} = \left( \frac{11/4}{1} \right)^{1/3} = 1.401 \; \Rightarrow \; T_\nu = 1.945 \; \text{K},$$

(146)

where I have used $T_{CMB} = 2.725 \pm 0.002 \; \text{K}$. We still have not measured such a relic background of neutrinos, and probably will remain undetected for a long time, since they have an average energy of order $10^{-4} \; \text{eV}$, much below that required for detection by present experiments (of order GeV), precisely because of the relative weakness of the weak interactions. Nevertheless, it would be fascinating if, in the future, ingenious experiments were devised to detect such a background, since it would confirm one of the most robust features of Big Bang cosmology.

### 3.2.6 Matter-radiation equality

Relativistic species have energy densities proportional to the quartic power of temperature and therefore scale as $\rho_R \propto a^{-4}$, while non-relativistic particles have essentially zero pressure and scale as $\rho_M \propto a^{-3}$. Therefore, there will be a time in the evolution of the universe in which both energy densities are equal $\rho_R(t_{eq}) = \rho_M(t_{eq})$. Since then both decay differently, and thus

$$1 + z_{eq} = \frac{a_0}{a_{eq}} = \frac{\Omega_M}{\Omega_R} = 3.1 \times 10^4 \; \Omega_M h^2,$$

(147)

where I have used $\Omega_R h^2 = \Omega_{CMB} h^2 + \Omega_\nu h^2 = 3.24 \times 10^{-5}$ for three massless neutrinos at $T = T_\nu$. As I will show later, the matter content of the universe today is below critical, $\Omega_M \simeq 0.3$, while $h \simeq 0.73$, and therefore $(1 + z_{eq}) \simeq 3900$, or about $t_{eq} = 1300 (\Omega_M h^2)^{-2} \; \text{yr} \simeq 70,000 \; \text{years}$ after the origin of the universe. Around the time of matter-radiation equality, the rate of expansion (31) can be written as ($a_0 \equiv 1$)

$$H(a) = H_0 \left( \Omega_R a^{-4} + \Omega_M a^{-3} \right)^{1/2} = H_0 \; \Omega_M^{1/2} a^{-3/2} \left( 1 + \frac{a_{eq}}{a} \right)^{1/2}.$$  

(148)

The horizon size is the coordinate distance travelled by a photon since the beginning of the universe, $d_H \sim H^{-1}$, i.e. the size of causally connected regions in the universe. The comoving horizon size is then given by

$$d_H = \int \frac{c \; da}{a^2 H(a)} = 2c \; H_0^{-1} \; \Omega_M^{1/2} a_{eq}^{1/2} \left( \sqrt{1 + \frac{a_{eq}}{a}} - 1 \right).$$

(149)

Thus the horizon size at matter-radiation equality ($a = a_{eq}$) is

$$d_H(a_{eq}) = 2c \; H_0^{-1} \; \Omega_M^{1/2} a_{eq}^{1/2} (\sqrt{2} - 1) \simeq 18 \; (\Omega_M h)^{-1} h^{-1} \text{Mpc}.$$  

(150)

This scale plays a very important role in theories of structure formation.
3.2.7 Recombination and photon decoupling

As the temperature of the universe decreased, electrons could eventually become bound to protons to form neutral hydrogen. Nevertheless, there is always a non-zero probability that a rare energetic photon ionizes hydrogen and produces a free electron. The ionization fraction of electrons in equilibrium with the plasma at a given temperature is given by the Saha equation [15]

\[
\frac{1 - X_{e}^{eq}}{X_{e}^{eq} T^{2}} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \eta \left( \frac{T}{m_{e}} \right)^{3/2} e^{E_{ion}/T},
\]

where \( E_{ion} = 13.6 \text{ eV} \) is the ionization energy of hydrogen, and \( \eta \) is the baryon-to-photon ratio (143). If we now use Eq. (130), we can compute the ionization fraction \( X_{e}^{eq} \) as a function of redshift \( z \). Note that the huge number of photons with respect to electrons (in the ratio \( 4He : H : \gamma \simeq 1 : 4 : 10^{10} \)) implies that even at a very low temperature, the photon distribution will contain a sufficiently large number of high-energy photons to ionize a significant fraction of hydrogen. In fact, defining recombination as the time at which \( X_{e}^{eq} \approx 0.1 \), one finds that the recombination temperature is \( T_{rec} = 0.296 \text{ eV} \ll E_{ion} \), for \( \eta_{10} \approx 6 \). Comparing with the present temperature of the microwave background, we deduce the corresponding redshift at recombination, \( (1 + z_{rec}) \approx 1260 \).

Photons remain in thermal equilibrium with the plasma of baryons and electrons through elastic Thomson scattering, with cross section

\[
\sigma_{\gamma} = \frac{8\pi\alpha^{2}}{3m_{e}^{2}} = 6.65 \times 10^{-25} \text{ cm}^{2} = 0.665 \text{ barn},
\]

where \( \alpha = 1/137.036 \) is the dimensionless electromagnetic coupling constant. The mean free path of photons \( \lambda_{\gamma} \) in such a plasma can be estimated from the photon interaction rate, \( \lambda_{\gamma}^{-1} \sim \Gamma_{\gamma} = n_{e}\sigma_{\gamma} c \). For temperatures above a few eV, the mean free path is much smaller that the causal horizon at that time and photons suffer multiple scattering: the plasma is like a dense fog. Photons will decouple from the plasma when their interaction rate cannot keep up with the expansion of the universe and the mean free path becomes larger than the horizon size: the universe becomes transparent. We can estimate this moment by evaluating \( \Gamma_{\gamma} = H \) at photon decoupling. Using \( n_{e} = X_{e} \eta n_{\gamma} \), one can compute the decoupling temperature as \( T_{dec} = 0.26 \text{ eV} \), and the corresponding redshift as \( 1 + z_{dec} \approx 1100 \). Recently, WMAP measured this redshift to be \( 1 + z_{dec} \approx 1089 \pm 1 \) [20]. This redshift defines the so called last scattering surface, when photons last scattered off protons and electrons and travelled freely ever since. This decoupling occurred when the universe was approximately \( t_{\text{dec}} = 1.5 \times 10^{5} (\Omega_{M} h^{2})^{-1/2} \approx 380,000 \) years old.

3.2.8 The microwave background

One of the most remarkable observations ever made by mankind is the detection of the relic background of photons from the Big Bang. This background was predicted by George Gamow and collaborators in the 1940s, based on the consistency of primordial nucleosynthesis with the observed helium abundance. They estimated a value of about 10 K, although a somewhat more detailed analysis by Alpher and Herman in 1949 predicted \( T_{\gamma} \approx 5 \text{ K} \). Unfortunately, they had doubts whether the radiation would have survived until the present, and this remarkable prediction slipped into obscurity, until Dicke, Peebles, Roll and Wilkinson [22] studied the problem again in 1965. Before they could measure the photon background, they learned that Penzias and Wilson had observed a weak isotropic background signal at a radio wavelength of 7.35 cm, corresponding to a blackbody temperature of \( T_{\gamma} = 3.5 \pm 1 \text{ K} \). They published their two papers back to back, with that of Dicke et al. explaining the fundamental significance of their measurement [6].

Since then many different experiments have confirmed the existence of the microwave background. The most outstanding one has been the Cosmic Background Explorer (COBE) satellite, whose FIRAS
instrument measured the photon background with great accuracy over a wide range of frequencies ($\nu = 1 - 97$ cm$^{-1}$), see Ref. [21], with a spectral resolution $\Delta\nu = 0.0035$. Nowadays, the photon spectrum is confirmed to be a blackbody spectrum with a temperature given by [21]

$$ T_{\text{CMB}} = 2.725 \pm 0.002 \text{ K (systematic, 95\% c.l.)} \pm 7 \mu \text{K (1}\sigma \text{ statistical}) $$(153)

In fact, this is the best blackbody spectrum ever measured, see Fig. 14, with spectral distortions below the level of 10 parts per million (ppm).

Moreover, the differential microwave radiometer (DMR) instrument on COBE, with a resolution of about $7^\circ$ in the sky, has also confirmed that it is an extraordinarily isotropic background. The deviations from isotropy, i.e. differences in the temperature of the blackbody spectrum measured in different directions in the sky, are of the order of 20 $\mu$K on large scales, or one part in $10^5$, see Ref. [23]. There is, in fact, a dipole anisotropy of one part in $10^3$, $\delta T_1 = 3.372 \pm 0.007$ mK (95\% c.l.), in the direction of the Virgo cluster, $(l, b) = (264.14^\circ \pm 0.30, 48.26^\circ \pm 0.30)$ (95\% c.l.). Under the assumption that a Doppler effect is responsible for the entire CMB dipole, the velocity of the Sun with respect to the CMB rest frame is $v_\odot = 371 \pm 0.5$ km/s, see Ref. [21].\footnote{COBE even determined the annual variation due to the Earth’s motion around the Sun – the ultimate proof of Copernicus’ hypothesis.} When subtracted, we are left with a whole spectrum of anisotropies in the higher multipoles (quadrupole, octupole, etc.), $\delta T_2 = 18 \pm 2 \mu$K (95\% c.l.), see Ref. [23] and Fig. 15.

Soon after COBE, other groups quickly confirmed the detection of temperature anisotropies at around 30 $\mu$K and above, at higher multipole numbers or smaller angular scales. As I shall discuss below, these anisotropies play a crucial role in the understanding of the origin of structure in the universe.

### 3.2.9 Large-scale structure formation

Although the isotropic microwave background indicates that the universe in the past was extraordinarily homogeneous, we know that the universe today is not exactly homogeneous: we observe galaxies, clusters and superclusters on large scales. These structures are expected to arise from very small primordial inhomogeneities that grow in time via gravitational instability, and that may have originated from tiny ripples in the metric, as matter fell into their troughs. Those ripples must have left some trace as temperature anisotropies in the microwave background, and indeed such anisotropies were finally discovered...
Fig. 15: The Cosmic Microwave Background Spectrum seen by the DMR instrument on COBE. The top figure corresponds to the monopole, $T_0 = 2.725 \pm 0.002$ K. The middle figure shows the dipole, $\delta T_1 = 3.372 \pm 0.014$ mK, and the lower figure shows the quadrupole and higher multipoles, $\delta T_2 = 18 \pm 2$ $\mu$K. The central region corresponds to foreground by the galaxy. From Ref. [23].

by the COBE satellite in 1992. The reason why they took so long to be discovered was that they appear as perturbations in temperature of only one part in $10^5$.

3.2.10 Newtonian linear perturbation theory
Suppose we start with a perfect fluid with pressure $p$ and energy density $\rho$, in the presence of a weak gravitational field described by a Newtonian potential $\phi$. The metric in that case can be written, to first order, as

$$g_{00} = -1 - 2\phi, \quad g_{0i} = 0, \quad g_{ij} = (1 - 2\phi)\delta_{ij},$$

(154)

and the energy-momentum tensor as

$$T^{00} = \rho, \quad T^{0i} = (\rho + p) v^i, \quad T^{ij} = (\rho + p) v^i v^j + p \delta^{ij}.$$  

(155)
In this case, the covariant conservation of the energy momentum tensor becomes
\[
T_{\mu\nu} = \frac{1}{\sqrt{g}} \partial_{\nu}(\sqrt{g} T^{\mu\nu}) + \Gamma^\mu_{\nu\lambda} T^\nu_{\lambda} = \frac{\partial p}{\partial x^\nu} g^{\mu\nu} + \frac{1}{\sqrt{g}} \partial_\nu \left( \sqrt{g} u^\mu u^\nu (p + \rho) \right) + \Gamma^\mu_{\nu\lambda} (p + \rho) u^\nu u^\lambda = 0.
\]

Let us now assume that the dominant cosmic fluid during the period of structure formation (after photon decoupling) is a non-relativistic fluid with zero pressure \( p \ll \rho \), but whose pressure gradients can be important (e.g. to prevent gravitational collapse). Let us also assume that we write the comoving 4-velocity as \( u^\mu = (1, v) \), with peculiar velocity \( v \). In this case, the energy conservation equation, together with the Poisson equation determine the fundamental hydrodynamical equations of the fluid,
\[
\partial_0 \rho + \nabla (\rho v) = 0, \quad \text{(continuity eq.)}
\]
\[
\partial_0 v + (v \cdot \nabla) v + \nabla \phi = -\frac{\nabla p}{\rho}, \quad \text{(Euler eq.)}
\]
\[
\nabla^2 \phi = 4\pi G \rho. \quad \text{(Poisson eq.)}
\]

As long as we only consider scales that are much smaller than the size of the horizon at the time of structure formation \( (z \sim 20) \), these equations will give a reasonable description and we can ignore the relativistic effects of horizons, etc.

Since the universe is expanding, the physical coordinates can be written as \( r = a(t) x \), where \( x \) are comoving coordinates. The proper fluid velocity can then be separated into the Hubble flow plus peculiar velocities, \( u \equiv \dot{x} \),
\[
v = a \left( H x + u \right).
\]
Spatial gradients are now given by \( \nabla_r = a^{-1} \nabla_x \). On the other hand, energy density and pressure can be decomposed into a background average plus a perturbation,
\[
\rho(t, x) = \bar{\rho}(t) \left( 1 + \delta(t, x) \right), \quad p(t, x) = 0 + \delta p(t, x),
\]
where we have defined the density contrast as
\[
\delta(t, x) \equiv \frac{\rho(t, x) - \bar{\rho}(t)}{\bar{\rho}(t)},
\]
with \( \bar{\rho}(a) = \rho_0 a^{-3} \) the average density of the Universe. If we make a change of coordinates from \( r \) to \( x \), and we linearize the equations, we find
\[
\partial_0 \delta + \nabla_x u = 0, \quad \text{(continuity)}
\]
\[
\partial_0 u + 2H u + \frac{1}{a^2} \nabla_x \phi + \frac{\nabla_x p}{a^2 \bar{\rho}} = 0, \quad \text{(Euler)}
\]
\[
\frac{1}{a^2} \nabla_x^2 \phi - 4\pi G \bar{\rho} \delta = 0. \quad \text{(Poisson)}
\]
We can now combine the three equations to obtain a single second order PDE equation for the density contrast, which written in Fourier space
\[
\delta(t, x) \equiv \int d^3k \ \delta_k(t) e^{i k \cdot x},
\]
becomes
\[
\dot{\delta}_k + 2H \delta_k + (c_s^2 k_{ph}^2 - 4\pi G \bar{\rho}) \delta_k = 0,
\]
where \( c_s^2 = dp/d\rho \) is the speed of sound of the fluid (assuming adiabatic perturbations, \( \delta p/\delta \rho = dp/d\rho \)) and \( k_{ph} = k/a \) is the wavenumber, the inverse of the physical wavelength of the perturbation. This equation is that of a damped harmonic oscillator. The null frequency oscillator determines the Jeans wavenumber, and its associated wavelength,

\[
k_J = \sqrt{\frac{4\pi G}{c_s^2}}, \quad \lambda_J = 2\pi k_J = c_s \sqrt{\frac{\pi}{G \bar{\rho}}},
\]

which separate stable perturbations from gravitationally unstable ones. For modes with wavenumber \( k \ll k_J \), the perturbation amplitude \( \delta_k \) will grow exponentially on a dynamical timescale \( \tau_{grav} = \text{Im} \omega_k^{-1} = (4\pi G \bar{\rho})^{-1/2} \), which is the typical scale of gravitational collapse in this problem. If we define the pressure response time of the fluid as the size of the perturbation over the sound speed,

\[
\tau_{pres} \sim \lambda/c_s,
\]

then, when \( \tau_{pres} > \tau_{grav} \), gravitational collapse of the perturbation occurs before the pressure forces can respond restoring hydrostatic equilibrium. On the other hand, if \( \tau_{pres} < \tau_{grav} \), thermal pressure prevents gravitational collapse and the fluid enters a series of damped acoustic oscillations.

### 3.2.11 The matter power spectrum and future galaxy surveys

While the predicted anisotropies have finally been seen in the CMB, not all kinds of matter and/or evolution of the universe can give rise to the structure we observe today. If we define the density contrast as \[\delta(\mathbf{x}, a) \equiv \frac{\rho(\mathbf{x}, a) - \bar{\rho}(a)}{\bar{\rho}(a)} = \int d^3k \delta_k(a) e^{i\mathbf{k} \cdot \mathbf{x}}, \tag{167}\]

where \( \bar{\rho}(a) = \rho_0 a^{-3} \) is the average cosmic density, we need a theory that will grow a density contrast with amplitude \( \delta \sim 10^{-5} \) at the last scattering surface (\( z = 1100 \)) up to density contrasts of the order of \( \delta \sim 10^2 \) for galaxies at redshifts \( z \ll 1 \), i.e. today. This is a necessary requirement for any consistent theory of structure formation \[25\].

Furthermore, the anisotropies observed by the COBE satellite correspond to a small-amplitude scale-invariant primordial power spectrum of inhomogeneities

\[
P(k) = \langle |\delta_k|^2 \rangle \propto k^n, \quad \text{with} \quad n = 1, \tag{168}\]

where the brackets \( \langle \cdot \rangle \) represent integration over an ensemble of different universe realizations. These inhomogeneities are like waves in the space-time metric. When matter fell in the troughs of those waves, it created density perturbations that collapsed gravitationally to form galaxies and clusters of galaxies, with a spectrum that is also scale invariant. Such a type of spectrum was proposed in the early 1970s by Edward R. Harrison, and independently by the Russian cosmologist Yakov B. Zel’dovich, see Ref. \[26\], to explain the distribution of galaxies and clusters of galaxies on very large scales in our observable universe.

Today various telescopes – like the Hubble Space Telescope, the twin Keck telescopes in Hawaii and the European Southern Observatory telescopes in Chile – are exploring the most distant regions of the universe and discovering the first galaxies at large distances. The furthest galaxies observed so far are at redshifts of \( z \simeq 10 \) (at a distance of 13.7 billion light years from Earth), whose light was emitted when the universe had only about 3% of its present age. Only a few galaxies are known at those redshifts, but there are at present various catalogs like the CfA and APM galaxy catalogs, and more recently the IRAS Point Source redshift Catalog, see Fig. 16, and Las Campanas redshift surveys, that study the spatial distribution of hundreds of thousands of galaxies up to distances of a billion light years, or \( z < 0.1 \), or the 2 degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), which reach \( z < 0.5 \) and study millions of galaxies. These catalogs are telling us about the evolution of clusters and superclusters of galaxies in the universe, and already put constraints on the theory of structure formation. From these observations one can infer that most galaxies formed at redshifts of the
order of $2 - 6$; clusters of galaxies formed at redshifts of order 1, and superclusters are forming now. This fundamental difference is an indication of the type of matter that gave rise to structure.

We know from Big Bang nucleosynthesis that all the baryons in the universe cannot account for the observed amount of matter, so there must be some extra matter (dark since we don’t see it) to account for its gravitational pull. Whether it is relativistic (hot) or non-relativistic (cold) could be inferred from observations: relativistic particles tend to diffuse from one concentration of matter to another, thus transferring energy among them and preventing the growth of structure on small scales. This is excluded by observations, so we conclude that most of the matter responsible for structure formation must be cold. How much there is is a matter of debate at the moment. Some recent analyses suggest that there is not enough cold dark matter to reach the critical density required to make the universe flat. If we want to make sense of the present observations, we must conclude that some other form of energy permeates the universe. In order to resolve this issue, 2dFGRS and SDSS started taking data a few years ago. The first has already been completed, but the second one is still taking data up to redshifts $z \simeq 5$ for quasars, over a large region of the sky. These important observations will help astronomers determine the nature of the dark matter and test the validity of the models of structure formation.

Before COBE discovered the anisotropies of the microwave background there were serious doubts whether gravity alone could be responsible for the formation of the structure we observe in the universe today. It seemed that a new force was required to do the job. Fortunately, the anisotropies were found with the right amplitude for structure to be accounted for by gravitational collapse of primordial inhomogeneities under the attraction of a large component of non-relativistic dark matter. Nowadays, the standard theory of structure formation is a cold dark matter model with a non vanishing cosmological constant in a spatially flat universe. Gravitational collapse amplifies the density contrast initially through linear growth and later on via non-linear collapse. In the process, overdense regions decouple from the Hubble expansion to become bound systems, which start attracting each other to form larger bound structures. In fact, the largest structures, superclusters, have not yet gone non-linear.

The primordial spectrum (168) is reprocessed by gravitational instability after the universe be-
comes matter dominated and inhomogeneities can grow. Linear perturbation theory shows that the growing mode of small density contrasts go like \[ \delta(a) \propto a^{1+3\omega} = \begin{cases} a^2, & a < a_{eq} \\ a, & a > a_{eq} \end{cases} \tag{169} \]
in the Einstein-de Sitter limit (\(\omega = p/\rho = 1/3\) and 0, for radiation and matter, respectively). There are slight deviations for \(a \gg a_{eq}\), if \(\Omega_M \neq 1\) or \(\Omega_\Lambda \neq 0\), but we will not be concerned with them here. The important observation is that, since the density contrast at last scattering is of order \(\delta \sim 10^{-5}\), and the scale factor has grown since then only a factor \(z_{dec} \sim 10^3\), one would expect a density contrast today of order \(\delta_0 \sim 10^{-2}\). Instead, we observe structures like galaxies, where \(\delta \sim 10^2\). So how can this be possible? The microwave background shows anisotropies due to fluctuations in the baryonic matter component only (to which photons couple, electromagnetically). If there is an additional matter component that only couples through very weak interactions, fluctuations in that component could grow as soon as it decoupled from the plasma, well before photons decoupled from baryons. The reason why baryonic inhomogeneities cannot grow is because of photon pressure: as baryons collapse towards denser regions, radiation pressure eventually halts the contraction and sets up acoustic oscillations in the plasma that prevent the growth of perturbations, until photon decoupling. On the other hand, a weakly interacting cold dark matter component could start gravitational collapse much earlier, even before matter-radiation equality, and thus reach the density contrast amplitudes observed today. The resolution of this mismatch is one of the strongest arguments for the existence of a weakly interacting cold dark matter component of the universe.

![Fig. 17: The power spectrum for cold dark matter (CDM), tilted cold dark matter (TCDM), hot dark matter (HDM), and mixed hot plus cold dark matter (MDM), normalized to COBE, for large-scale structure formation. From Ref. [28].](image)

How much dark matter there is in the universe can be deduced from the actual power spectrum (the Fourier transform of the two-point correlation function of density perturbations) of the observed large scale structure. One can decompose the density contrast in Fourier components, see Eq. (160). This is very convenient since in linear perturbation theory individual Fourier components evolve independently. A comoving wavenumber \(k\) is said to “enter the horizon” when \(k = d_H^{-1}(a) = aH(a)\). If a certain perturbation, of wavelength \(\lambda = k^{-1} < d_H(a_{eq})\), enters the horizon before matter-radiation equality, the fast radiation-driven expansion prevents dark-matter perturbations from collapsing. Since light can only cross regions that are smaller than the horizon, the suppression of growth due to radiation is restricted to scales smaller than the horizon, while large-scale perturbations remain unaffected. This is the reason

\[3\text{The decaying modes go like } \delta(t) \sim t^{-1}, \text{ for all } \omega.\]
why the horizon size at equality, Eq. (150), sets an important scale for structure growth,

$$k_{eq} = d_H^{-1}(a_{eq}) \simeq 0.083 (\Omega_M h) h \text{ Mpc}^{-1}. \quad (170)$$

The suppression factor can be easily computed from (169) as $f_{sup} = (a_{enter}/a_{eq})^2 = (k_{eq}/k)^2$. In other words, the processed power spectrum $P(k)$ will have the form:

$$P(k) \propto \begin{cases} 
  k, & k \ll k_{eq} \\
  k^{-3}, & k \gg k_{eq}
\end{cases} \quad (171)$$

This is precisely the shape that large-scale galaxy catalogs are bound to test in the near future, see Fig. 17. Furthermore, since relativistic Hot Dark Matter (HDM) transfer energy between clumps of matter, they will wipe out small scale perturbations, and this should be seen as a distinctive signature in the matter power spectra of future galaxy catalogs. On the other hand, non-relativistic Cold Dark Matter (CDM) allow structure to form on all scales via gravitational collapse. The dark matter will then pull in the baryons, which will later shine and thus allow us to see the galaxies.

Naturally, when baryons start to collapse onto dark matter potential wells, they will convert a large fraction of their potential energy into kinetic energy of protons and electrons, ionizing the medium. As a consequence, we expect to see a large fraction of those baryons constituting a hot ionized gas surrounding large clusters of galaxies. This is indeed what is observed, and confirms the general picture of structure formation.

4 DETERMINATION OF COSMOLOGICAL PARAMETERS

In this Section, I will restrict myself to those recent measurements of the cosmological parameters by means of standard cosmological techniques, together with a few instances of new results from recently applied techniques. We will see that a large host of observations are determining the cosmological parameters with some reliability of the order of 10%. However, the majority of these measurements are dominated by large systematic errors. Most of the recent work in observational cosmology has been the search for virtually systematic-free observables, like those obtained from the microwave background anisotropies, and discussed in Section 4.4. I will devote, however, this Section to the more ‘classical’ measurements of the following cosmological parameters: The rate of expansion $H_0$; the matter content $\Omega_M$; the cosmological constant $\Omega_\Lambda$; the spatial curvature $\Omega_K$, and the age of the universe $t_0$.

4.1 The rate of expansion $H_0$

Over most of last century the value of $H_0$ has been a constant source of disagreement [29]. Around 1929, Hubble measured the rate of expansion to be $H_0 = 500 \text{ km s}^{-1}\text{Mpc}^{-1}$, which implied an age of the universe of order $t_0 \sim 2 \text{ Gyr}$, in clear conflict with geology. Hubble’s data was based on Cepheid standard candles that were incorrectly calibrated with those in the Large Magellanic Cloud. Later on, in 1954 Baade recalibrated the Cepheid distance and obtained a lower value, $H_0 = 250 \text{ km s}^{-1}\text{Mpc}^{-1}$, still in conflict with ratios of certain unstable isotopes. Finally, in 1958 Sandage realized that the brightest stars in galaxies were ionized HII regions, and the Hubble rate dropped down to $H_0 = 60 \text{ km s}^{-1}\text{Mpc}^{-1}$, still with large (factor of two) systematic errors. Fortunately, in the past 15 years there has been significant progress towards the determination of $H_0$, with systematic errors approaching the 10% level. These improvements come from two directions. First, technological, through the replacement of photographic plates (almost exclusively the source of data from the 1920s to 1980s) with charged couple devices (CCDs), i.e. solid state detectors with excellent flux sensitivity per pixel, which were previously used successfully in particle physics detectors. Second, by the refinement of existing methods for measuring extragalactic distances (e.g. parallax, Cepheids, supernovae, etc.). Finally, with the development of completely new methods to determine $H_0$, which fall into totally independent and very broad categories: a) Gravitational lensing; b) Sunyaev-Zel’dovich effect; c) Extragalactic distance scale, mainly
Cepheid variability and type Ia Supernovae; d) Microwave background anisotropies. I will review here the first three, and leave the last method for Section 4.4, since it involves knowledge about the primordial spectrum of inhomogeneities.

4.1.1 Gravitational lensing

Imagine a quasi-stellar object (QSO) at large redshift ($z \gg 1$) whose light is lensed by an intervening galaxy at redshift $z \sim 1$ and arrives to an observer at $z = 0$. There will be at least two different images of the same background variable point source. The arrival times of photons from two different gravitationally lensed images of the quasar depend on the different path lengths and the gravitational potential traversed. Therefore, a measurement of the time delay and the angular separation of the different images of a variable quasar can be used to determine $H_0$ with great accuracy. This method, proposed in 1964 by Refsdael [30], offers tremendous potential because it can be applied at great distances and it is based on very solid physical principles [31].

Unfortunately, there are very few systems with both a favourable geometry (i.e. a known mass distribution of the intervening galaxy) and a variable background source with a measurable time delay. That is the reason why it has taken so much time since the original proposal for the first results to come out. Fortunately, there are now very powerful telescopes that can be used for these purposes. The best candidate to-date is the QSO 0957 + 561, observed with the 10m Keck telescope, for which there is a model of the lensing mass distribution that is consistent with the measured velocity dispersion. Assuming a flat space with $\Omega_{M} = 0.25$, one can determine [32]

$$H_0 = 72 \pm 7 \ (1\sigma \ \text{statistical}) \pm 15\% \ (\text{systematic}) \ \text{km} \ \text{s}^{-1} \text{Mpc}^{-1}.$$

The main source of systematic error is the degeneracy between the mass distribution of the lens and the value of $H_0$. Knowledge of the velocity dispersion within the lens as a function of position helps constrain the mass distribution, but those measurements are very difficult and, in the case of lensing by a cluster of galaxies, the dark matter distribution in those systems is usually unknown, associated with a complicated cluster potential. Nevertheless, the method is just starting to give promising results and, in the near future, with the recent discovery of several systems with optimum properties, the prospects for measuring $H_0$ and lowering its uncertainty with this technique are excellent.

4.1.2 Sunyaev-Zel’dovich effect

As discussed in the previous Section, the gravitational collapse of baryons onto the potential wells generated by dark matter gave rise to the reionization of the plasma, generating an X-ray halo around rich clusters of galaxies, see Fig. 18. The inverse-Compton scattering of microwave background photons off the hot electrons in the X-ray gas results in a measurable distortion of the blackbody spectrum of the microwave background, known as the Sunyaev-Zel’dovich (SZ) effect. Since photons acquire extra energy from the X-ray electrons, we expect a shift towards higher frequencies of the spectrum, $(\Delta \nu/\nu) \simeq (k_B T_{\text{gas}}/m_e c^2) \sim 10^{-2}$. This corresponds to a decrement of the microwave background temperature at low frequencies (Rayleigh-Jeans region) and an increment at high frequencies, see Ref. [33].

Measuring the spatial distribution of the SZ effect (3 K spectrum), together with a high resolution X-ray map ($10^8$ K spectrum) of the cluster, one can determine the density and temperature distribution of the hot gas. Since the X-ray flux is distance-dependent ($F = L/4\pi d_L^2$), while the SZ decrement is not (because the energy of the CMB photons increases as we go back in redshift, $\nu = \nu_0(1 + z)$, and exactly compensates the redshift in energy of the photons that reach us), one can determine from there the distance to the cluster, and thus the Hubble rate $H_0$.

The advantages of this method are that it can be applied to large distances and it is based on clear physical principles. The main systematics come from possible clumpiness of the gas (which would
reduce $H_0$), projection effects (if the clusters are prolate, $H_0$ could be larger), the assumption of hydrostatic equilibrium of the X-ray gas, details of models for the gas and electron densities, and possible contaminations from point sources. Present measurements give the value $H_0 = 60 \pm 10 \, (1\sigma \text{ statistical}) \pm 20\% \, (\text{systematic}) \, \text{km s}^{-1}\text{Mpc}^{-1}$, \hfill (173)

compatible with other determinations. A great advantage of this completely new and independent method is that nowadays more and more clusters are observed in the X-ray, and soon we will have high-resolution 2D maps of the SZ decrement from several balloon flights, as well as from future microwave background satellites, together with precise X-ray maps and spectra from the Chandra X-ray observatory recently launched by NASA, as well as from the European X-ray satellite XMM launched a few months ago by ESA, which will deliver orders of magnitude better resolution than the existing Einstein X-ray satellite.

### 4.1.3 Cepheid variability

Cepheids are low-mass variable stars with a period-luminosity relation based on the helium ionization cycles inside the star, as it contracts and expands. This time variability can be measured, and the star’s absolute luminosity determined from the calibrated relationship. From the observed flux one can then deduce the luminosity distance, see Eq. (69), and thus the Hubble rate $H_0$. The Hubble Space Telescope (HST) was launched by NASA in 1990 (and repaired in 1993) with the specific project of calibrating the extragalactic distance scale and thus determining the Hubble rate with 10% accuracy. The most recent results from HST are the following $H_0 = 71 \pm 4 \, (\text{random}) \pm 7 \, (\text{systematic}) \, \text{km s}^{-1}\text{Mpc}^{-1}$, \hfill (174)

The main source of systematic error is the distance to the Large Magellanic Cloud, which provides the fiducial comparison for Cepheids in more distant galaxies. Other systematic uncertainties that affect the value of $H_0$ are the internal extinction correction method used, a possible metallicity dependence of the Cepheid period-luminosity relation and cluster population incompleteness bias, for a set of 21 galaxies within 25 Mpc, and 23 clusters within $z < 0.03$.

With better telescopes already taking data, like the Very Large Telescope (VLT) interferometer of the European Southern Observatory (ESO) in the Chilean Atacama desert, with 8 synchronized telescopes, and others coming up soon, like the Next Generation Space Telescope (NGST) proposed by NASA for 2008, and the Gran TeCan of the European Northern Observatory in the Canary Islands, for 2010, it is expected that much better resolution and therefore accuracy can be obtained for the determination of $H_0$. 

---

Fig. 18: The Coma cluster of galaxies, seen here in an optical image (left) and an X-ray image (right), taken by the recently launched Chandra X-ray Observatory. From Ref. [34].
4.2 Dark Matter

In the 1920s Hubble realized that the so called nebulae were actually distant galaxies very similar to our own. Soon afterwards, in 1933, Zwicky found dynamical evidence that there is possibly ten to a hundred times more mass in the Coma cluster than contributed by the luminous matter in galaxies [36]. However, it was not until the 1970s that the existence of dark matter began to be taken more seriously. At that time there was evidence that rotation curves of galaxies did not fall off with radius and that the dynamical mass was increasing with scale from that of individual galaxies up to clusters of galaxies. Since then, new possible extra sources to the matter content of the universe have been accumulating:

\[ \Omega_M = \Omega_{B, \text{lum}} \text{ (stars in galaxies)} + \Omega_{B, \text{dark}} \text{ (MACHOs?)} + \Omega_{CDM} \text{ (weakly interacting : axion, neutralino?)} + \Omega_{HDM} \text{ (massive neutrinos?)} \]

The empirical route to the determination of \( \Omega_M \) is nowadays one of the most diversified of all cosmological parameters. The matter content of the universe can be deduced from the mass-to-light ratio of various objects in the universe; from the rotation curves of galaxies; from microlensing and the direct search of Massive Compact Halo Objects (MACHOs); from the cluster velocity dispersion with the use of the Virial theorem; from the baryon fraction in the X-ray gas of clusters; from weak gravitational lensing; from the observed matter distribution of the universe via its power spectrum; from the cluster abundance and its evolution; from direct detection of massive neutrinos at SuperKamiokande; from direct detection of Weakly Interacting Massive Particles (WIMPs) at CDMS, DAMA or UKDMC, and finally from microwave background anisotropies. I will review here just a few of them.

4.2.1 Rotation curves of spiral galaxies

The flat rotation curves of spiral galaxies provide the most direct evidence for the existence of large amounts of dark matter. Spiral galaxies consist of a central bulge and a very thin disk, stabilized against gravitational collapse by angular momentum conservation, and surrounded by an approximately spherical halo of dark matter. One can measure the orbital velocities of objects orbiting around the disk as a function of radius from the Doppler shifts of their spectral lines.

The rotation curve of the Andromeda galaxy was first measured by Babcock in 1938, from the stars in the disk. Later it became possible to measure galactic rotation curves far out into the disk, and a trend was found [38, 39]. The orbital velocity rose linearly from the center outward until it reached a typical value of 200 km/s, and then remained flat out to the largest measured radii. This was completely unexpected since the observed surface luminosity of the disk falls off exponentially with radius [38], \( I(r) = I_0 \exp(-r/r_D) \). Therefore, one would expect that most of the galactic mass is concentrated within a few disk lengths \( r_D \), such that the rotation velocity is determined as in a Keplerian orbit, \( v_{\text{rot}} = (GM/r)^{1/2} \propto r^{-1/2} \). No such behaviour is observed. In fact, the most convincing observations come from radio emission (from the 21 cm line) of neutral hydrogen in the disk, which has been measured to much larger galactic radii than optical tracers. A typical case is that of the spiral galaxy NGC 6503, where \( r_D = 1.73 \) kpc, while the furthest measured hydrogen line is at \( r = 22.22 \) kpc, about 13 disk lengths away. Nowadays, thousands of galactic rotation curves are known, see Fig. 19, and all suggest the existence of about ten times more mass in the halos of spiral galaxies than in the stars of the disk. The connection with dark matter halos was emphasized in Ref. [40]. Recent numerical simulations of galaxy formation in a CDM cosmology [41] suggest that galaxies probably formed by the infall of material in an overdense region of the universe that had decoupled from the overall expansion.

The dark matter is supposed to undergo violent relaxation and create a virialized system, i.e. in hydrostatic equilibrium. This picture has led to a simple model of dark-matter halos as isothermal spheres, with density profile \( \rho(r) = \rho_c/(r_c^2 + r^2) \), where \( r_c \) is a core radius and \( \rho_c = v_{\infty}^2/4\pi G \), with
\( \nu_\infty \) equal to the plateau value of the flat rotation curve. This model is consistent with the universal rotation curves seen in Fig. 19. At large radii the dark matter distribution leads to a flat rotation curve. The question is for how long. In dense galaxy clusters one expects the galactic halos to overlap and form a continuum, and therefore the rotation curves should remain flat from one galaxy to another. However, in field galaxies, far from clusters, one can study the rotation velocities of substructures (like satellite dwarf galaxies) around a given galaxy, and determine whether they fall off at sufficiently large distances according to Kepler’s law, as one would expect, once the edges of the dark matter halo have been reached. These observations are rather difficult because of uncertainties in distinguishing between true satellites and interlopers. Recently, a group from the Sloan Digital Sky Survey Collaboration claim that they have seen the edges of the dark matter halos around field galaxies by confirming the fall-off at large distances of their rotation curves \([42]\). These results, if corroborated by further analysis, would constitute a tremendous support to the idea of dark matter as a fluid surrounding galaxies and clusters, while at the same time eliminates the need for modifications of Newtonian or even Einsteinian gravity at the scales of galaxies, to account for the flat rotation curves.

That’s fine, but how much dark matter is there at the galactic scale? Adding up all the matter in galactic halos up to a maximum radii, one finds

\[
\Omega_{\text{halo}} \simeq 10 \Omega_{\text{lum}} \geq 0.03 - 0.05 .
\]

Of course, it would be extraordinary if we could confirm, through direct detection, the existence of dark matter in our own galaxy. For that purpose, one should measure its rotation curve, which is much more
difficult because of obscuration by dust in the disk, as well as problems with the determination of reliable
galactocentric distances for the tracers. Nevertheless, the rotation curve of the Milky Way has been
measured and conforms to the usual picture, with a plateau value of the rotation velocity of 220 km/s.
For dark matter searches, the crucial quantity is the dark matter density in the solar neighbourhood, which
turns out to be (within a factor of two uncertainty depending on the halo model) \( \rho_{DM} = 0.3 \) GeV/cm\(^3\).
We will come back to direct searches of dark matter in a later subsection.

### 4.2.2 Baryon fraction in clusters

Since large clusters of galaxies form through gravitational collapse, they scoop up mass over a large
volume of space, and therefore the ratio of baryons over the total matter in the cluster should be rep-
resentative of the entire universe, at least within a 20% systematic error. Since the 1960s, when X-ray
telescopes became available, it is known that galaxy clusters are the most powerful X-ray sources in the
sky [43]. The emission extends over the whole cluster and reveals the existence of a hot plasma with
temperature \( T \sim 10^7 - 10^8 \) K, where X-rays are produced by electron bremsstrahlung. Assuming the
gas to be in hydrostatic equilibrium and applying the virial theorem one can estimate the total mass in
the cluster, giving general agreement (within a factor of 2) with the virial mass estimates. From these
estimates one can calculate the baryon fraction of clusters

\[
 f_B h^{3/2} = 0.08 \quad \Rightarrow \quad \frac{\Omega_B}{\Omega_M} \approx 0.14, \quad \text{for} \quad h = 0.70. \tag{180}
\]

Since \( \Omega_{\text{lum}} \approx 0.002 - 0.006 \), the previous expression suggests that clusters contain far more baryonic
matter in the form of hot gas than in the form of stars in galaxies. Assuming this fraction to be repre-
sentative of the entire universe, and using the Big Bang nucleosynthesis value of \( \Omega_B = 0.04 \pm 0.01 \), for
\( h = 0.7 \), we find

\[
 \Omega_M = 0.3 \pm 0.1 \text{ (statistical) } \pm 20\% \text{ (systematic)}. \tag{181}
\]

This value is consistent with previous determinations of \( \Omega_M \). If some baryons are ejected from the cluster
during gravitational collapse, or some are actually bound in nonluminous objects like planets, then the
actual value of \( \Omega_M \) is smaller than this estimate.

### 4.2.3 Weak gravitational lensing

Since the mid 1980s, deep surveys with powerful telescopes have observed huge arc-like features in
galaxy clusters. The spectroscopic analysis showed that the cluster and the giant arcs were at very different
redshifts. The usual interpretation is that the arc is the image of a distant background galaxy which
is in the same line of sight as the cluster so that it appears distorted and magnified by the gravitational
lens effect: the giant arcs are essentially partial Einstein rings. From a systematic study of the clus-
ter mass distribution one can reconstruct the shear field responsible for the gravitational distortion [44].
This analysis shows that there are large amounts of dark matter in the clusters, in rough agreement with
the virial mass estimates, although the lensing masses tend to be systematically larger. At present, the
estimates indicate \( \Omega_M = 0.2 - 0.3 \) on scales \( \lesssim 6 h^{-1} \) Mpc.

### 4.2.4 Large scale structure formation and the matter power spectrum

Although the isotropic microwave background indicates that the universe in the past was extraordinarily
homogeneous, we know that the universe today is far from homogeneous: we observe galaxies, clusters
and superclusters on large scales. These structures are expected to arise from very small primordial inho-
mogeneities that grow in time via gravitational instability, and that may have originated from tiny ripples
in the metric, as matter fell into their troughs. Those ripples must have left some trace as temperature
anisotropies in the microwave background, and indeed such anisotropies were finally discovered by the
COBE satellite in 1992. However, not all kinds of matter and/or evolution of the universe can give rise to the structure we observe today. If we define the density contrast as

\[
\delta(\vec{x},a) \equiv \frac{\rho(\vec{x},a) - \bar{\rho}(a)}{\bar{\rho}(a)} = \int d^3k \, \delta_k(a) \, e^{i\vec{k} \cdot \vec{x}},
\]

(182)

where \(\bar{\rho}(a) = \rho_0 a^{-3}\) is the average cosmic density, we need a theory that will grow a density contrast with amplitude \(\delta \sim 10^{-5}\) at the last scattering surface \((z = 1100)\) up to density contrasts of the order of \(\delta \sim 10^2\) for galaxies at redshifts \(z \ll 1\), i.e. today. This is a necessary requirement for any consistent theory of structure formation.

Furthermore, the anisotropies observed by the COBE satellite correspond to a small-amplitude scale-invariant primordial power spectrum of inhomogeneities

\[
P(k) = \langle |\delta_k|^2 \rangle \propto k^n, \quad \text{with} \quad n = 1,
\]

(183)

These inhomogeneities are like waves in the space-time metric. When matter fell in the troughs of those waves, it created density perturbations that collapsed gravitationally to form galaxies and clusters of galaxies, with a spectrum that is also scale invariant. Such a type of spectrum was proposed in the early 1970s by Edward R. Harrison, and independently by the Russian cosmologist Yakov B. Zel’ dovich [26], to explain the distribution of galaxies and clusters of galaxies on very large scales in our observable universe, see Fig. 20.

Since the primordial spectrum is very approximately represented by a scale-invariant Gaussian random field, the best way to present the results of structure formation is by working with the 2-point correlation function in Fourier space, the so-called power spectrum. If the reprocessed spectrum of inhomogeneities remains Gaussian, the power spectrum is all we need to describe the galaxy distribution. Non-Gaussian effects are expected to arise from the non-linear gravitational collapse of structure, and may be important at small scales. The power spectrum measures the degree of inhomogeneity in the mass distribution on different scales, see Fig. 21. It depends upon a few basic ingredients: a) the primordial spectrum of inhomogeneities, whether they are Gaussian or non-Gaussian, whether \textit{adiabatic} (perturbations in the energy density) or \textit{isocurvature} (perturbations in the entropy density), whether the primordial spectrum has \textit{tilt} (deviations from scale-invariance), etc.; b) the recent creation of inhomogeneities, whether \textit{cosmic strings} or some other topological defect from an early phase transition are responsible for the formation of structure today; and c) the cosmic evolution of the inhomogeneity,
whether the universe has been dominated by cold or hot dark matter or by a cosmological constant since the beginning of structure formation, and also depending on the rate of expansion of the universe.

The working tools used for the comparison between the observed power spectrum and the predicted one are very precise N-body numerical simulations and theoretical models that predict the shape but not the amplitude of the present power spectrum. Even though a large amount of work has gone into those analyses, we still have large uncertainties about the nature and amount of matter necessary for structure formation. A model that has become a working paradigm is a flat cold dark matter model with a cosmological constant and $\Omega_M \sim 0.3$. This model is now been confronted with the recent very precise measurements from 2dFGRS [45] and SDSS [46].

### 4.2.5 The new redshift catalogs, 2dF and Sloan Digital Sky Survey

Our view of the large-scale distribution of luminous objects in the universe has changed dramatically during the last 25 years: from the simple pre-1975 picture of a distribution of field and cluster galaxies, to the discovery of the first single superstructures and voids, to the most recent results showing an almost regular web-like network of interconnected clusters, filaments and walls, separating huge nearly empty volumes. The increased efficiency of redshift surveys, made possible by the development of spectrographs and – specially in the last decade – by an enormous increase in multiplexing gain (i.e. the ability to collect spectra of several galaxies at once, thanks to fibre-optic spectrographs), has allowed us not only to do cartography of the nearby universe, but also to statistically characterize some of its properties. At the same time, advances in theoretical modeling of the development of structure, with large high-resolution gravitational simulations coupled to a deeper yet limited understanding of how to form galaxies within the dark matter halos, have provided a more realistic connection of the models to the observable quantities. Despite the large uncertainties that still exist, this has transformed the study of...
4.2.6 Summary of the matter content

We can summarize the present situation with Fig. 22, for $\Omega_M$ as a function of $H_0$. There are four bands, the luminous matter $\Omega_{\text{lum}}$; the baryon content $\Omega_B$, from BBN; the galactic halo component $\Omega_{\text{halo}}$, and the dynamical mass from clusters, $\Omega_M$. From this figure it is clear that there are in fact three dark matter problems: The first one is where are 90% of the baryons? Between the fraction predicted by BBN and that seen in stars and diffuse gas there is a huge fraction which is in the form of dark baryons. They could be in small clumps of hydrogen that have not started thermonuclear reactions and perhaps constitute the dark matter of spiral galaxies’ halos. Note that although $\Omega_B$ and $\Omega_{\text{halo}}$ coincide at $H_0 \simeq 70 \text{ km/s/Mpc}$, this could be just a coincidence. The second problem is what constitutes 90% of matter, from BBN baryons to the mass inferred from cluster dynamics? This is the standard dark matter problem and could be solved in the future by direct detection of a weakly interacting massive particle in the laboratory. And finally, since we know from observations of the CMB that the universe is flat, the rest, up to $\Omega_0 = 1$, must be a diffuse vacuum energy, which affects the very large scales and late times, and seems to be responsible for the present acceleration of the universe, see Section 3. Nowadays, multiple observations seem to converge towards a common determination of $\Omega_M = 0.25 \pm 0.08$ (95% c.l.), see Fig. 23.

4.2.7 Massive neutrinos

One of the ‘usual suspects’ when addressing the problem of dark matter are neutrinos. They are the only candidates known to exist. If neutrinos have a mass, could they constitute the missing matter? We know from the Big Bang theory, see Section 2.6.5, that there is a cosmic neutrino background at a temperature of approximately 2K. This allows one to compute the present number density in the form of neutrinos, which turns out to be, for massless neutrinos, $n_\nu(T_\nu) = \frac{3}{11} n_\gamma(T_\gamma) = 112 \text{ cm}^{-3}$, per species of neutrino.
Fig. 23: Different determinations of $\Omega_M$ as a function of distance, from various sources: 1. peculiar velocities; 2. weak gravitational lensing; 3. shear autocorrelation function; 4. local group of galaxies; 5. baryon mass fraction; 6. cluster mass function; 7. virgo-centric flow; 8. mean relative velocities; 9. redshift space distortions; 10. mass power spectrum; 11. integrated Sachs-Wolfe effect; 12. angular diameter distance: SNe; 13. cluster baryon fraction. While a few years ago the dispersion among observed values was huge and strongly dependent on scale, at present the observed value of the matter density parameter falls well within a narrow range, $\Omega_M = 0.25 \pm 0.07$ (95% c.l.) and is essentially independent on scale, from 100 kpc to 5000 Mpc. Adapted from Ref. [48].

If neutrinos have mass, as recent experiments seem to suggest, see Fig. 24, the cosmic energy density in massive neutrinos would be $\rho_\nu = \sum n_\nu m_\nu = \frac{3}{10} n_\gamma \sum m_\nu$, and therefore its contribution today,

$$\Omega_\nu h^2 = \frac{\sum m_\nu}{93.2 \text{ eV}}.$$  

The discussion in the previous Sections suggest that $\Omega_M \leq 0.4$, and thus, for any of the three families of neutrinos, $m_\nu \leq 40$ eV. Note that this limit improves by six orders of magnitude the present bound on the tau-neutrino mass [19]. Supposing that the missing mass in non-baryonic cold dark matter arises from a single particle dark matter (PDM) component, its contribution to the critical density is bounded by $0.05 \leq \Omega_{PDM} h^2 \leq 0.4$, see Fig. 23.

I will now go through the various logical arguments that exclude neutrinos as the dominant component of the missing dark matter in the universe. Is it possible that neutrinos with a mass $4 \text{ eV} \leq m_\nu \leq 40$ eV be the non-baryonic PDM component? For instance, could massive neutrinos constitute the dark matter halos of galaxies? For neutrinos to be gravitationally bound to galaxies it is necessary that their velocity be less that the escape velocity $v_{\text{esc}}$, and thus their maximum momentum is $p_{\text{max}} = m_\nu v_{\text{esc}}$.

How many neutrinos can be packed in the halo of a galaxy? Due to the Pauli exclusion principle, the maximum number density is given by that of a completely degenerate Fermi gas with momentum $p_F = p_{\text{max}}$, i.e. $n_{\text{max}} = \frac{p_{\text{max}}^3}{3\pi^2}$. Therefore, the maximum local density in dark matter neutrinos is $\rho_{\text{max}} = n_{\text{max}} m_\nu = \frac{m_\nu^3 v_{\text{esc}}^3}{3\pi^2}$, which must be greater than the typical halo density $\rho_{\text{halo}} = 0.3 \text{ GeV cm}^{-3}$. For a typical spiral galaxy, this constraint, known as the Tremaine-Gunn limit, gives $m_\nu \geq 40$ eV, see Ref. [50]. However, this mass, even for a single species, say the tau-neutrino, gives a value for $\Omega_\nu h^2 = 0.5$, which is far too high for structure formation. Neutrinos of such a low mass would constitute a relativistic hot dark matter component, which would wash-out structure below the supercluster scale, against evidence from present observations, see Fig. 24. Furthermore, applying the same phase-space argument to the neutrinos as dark matter in the halo of dwarf galaxies gives
$m_\nu \geq 100$ eV, beyond closure density (184). We must conclude that the simple idea that light neutrinos could constitute the particle dark matter on all scales is ruled out. They could, however, still play a role as a sub-dominant hot dark matter component in a flat CDM model. In that case, a neutrino mass of order 1 eV is not cosmological excluded, see Fig. 24.

Another possibility is that neutrinos have a large mass, of order a few GeV. In that case, their number density at decoupling, see Section 2.5.1, is suppressed by a Boltzmann factor, $\sim \exp(-m_\nu/T_{\text{dec}})$. For masses $m_\nu > T_{\text{dec}} \simeq 0.8$ MeV, the present energy density has to be computed as a solution of the corresponding Boltzmann equation. Apart from a logarithmic correction, one finds $\Omega_\nu h^2 \simeq 0.1(10 \text{ GeV}/m_\nu)^2$ for Majorana neutrinos and slightly smaller for Dirac neutrinos. In either case, neutrinos could be the dark matter only if their mass was a few GeV. Laboratory limits for $\nu_\tau$ of around 18 MeV [19], and much more stringent ones for $\nu_\mu$ and $\nu_e$, exclude the known light neutrinos. However, there is always the possibility of a fourth unknown heavy and stable (perhaps sterile) neutrino. If it couples to the Z boson and has a mass below 45 GeV for Dirac neutrinos (39.5 GeV for Majorana neutrinos), then it is ruled out by measurements at LEP of the invisible width of the Z. There are two logical alternatives, either it is a sterile neutrino (it does not couple to the Z), or it does couple but has a larger mass. In the case of a Majorana neutrino (its own antiparticle), their abundance, for this mass range, is too small for being cosmologically relevant, $\Omega_\nu h^2 \leq 0.005$. If it were a Dirac neutrino there could be a lepton asymmetry, which may provide a higher abundance (similar to the case of baryogenesis). However, neutrinos scatter on nucleons via the weak axial-vector current (spin-dependent) interaction. For the small momentum transfers imparted by galactic WIMPs, such collisions are essentially coherent over an entire nucleus, leading to an enhancement of the effective cross section. The relatively large detection rate in this case allows one to exclude fourth-generation Dirac neutrinos for the galactic dark matter [51]. Anyway, it would be very implausible to have such a massive neutrino today, since it would have to be stable, with a life-time greater than the age of the universe, and there is no theoretical reason

Fig. 24: The neutrino parameter space, mixing angle against $\Delta m^2$, including the results from the different solar and atmospheric neutrino oscillation experiments. Note the threshold of cosmologically important masses, cosmologically detectable neutrinos (by CMB and LSS observations), and cosmologically excluded range of masses. Adapted from Refs. [49] and [94].
to expect a massive sterile neutrino that does not oscillate into the other neutrinos.

Of course, the definitive test to the possible contribution of neutrinos to the overall density of the universe would be to measure directly their mass in laboratory experiments. There are at present two types of experiments: neutrino oscillation experiments, which measure only differences in squared masses, and direct mass-searches experiments, like the tritium $\beta$-spectrum and the neutrinoless double-$\beta$ decay experiments, which measure directly the mass of the electron neutrino. The former experiments give a bound $m_{\nu_e} \lesssim 2.3$ eV (95% c.l.) [52], while the latter claim [53] they have a positive evidence for a Majorana neutrino of mass $m_{\nu} = 0.05 - 0.89$ eV (95% c.l.), although this result still awaits confirmation by other experiments. Neutrinos with such a mass could very well constitute the HDM component of the universe, $\Omega_{\text{HDM}} \lesssim 0.15$. The oscillation experiments give a range of possibilities for $\Delta m_{\nu}^2 = 0.3 - 3$ eV$^2$ from LSND (not yet confirmed by Miniboone), to the atmospheric neutrino oscillations from SuperKamiokande ($\Delta m_{\nu}^2 \simeq 2.2 \pm 0.5 \times 10^{-3}$ eV$^2$, $\tan^2 \theta = 1.0 \pm 0.3$) and the solar neutrino oscillations from KamLAND and the Sudbury Neutrino Observatory ($\Delta m_{\nu}^2 \simeq 8.2 \pm 0.3 \times 10^{-5}$ eV$^2$, $\tan^2 \theta = 0.39 \pm 0.05$), see Ref. [49]. Only the first two possibilities would be cosmologically relevant, see Fig. 24. Thanks to recent observations by WMAP, 2dFGRS and SDSS, we can put stringent limits on the absolute scale of neutrino masses, see below.

Fig. 25: The annual-modulation signal accumulated over 7 years is consistent with a neutralino of mass of $m_{\chi} = 59^{+17}_{-14}$ GeV and a proton cross section of $\xi \sigma_p = 7.0^{+0.4}_{-0.3} \times 10^{-6}$ pb, according to DAMA. From Ref. [54].
4.2.8 Weakly Interacting Massive Particles

Unless we drastically change the theory of gravity on large scales, baryons cannot make up the bulk of the dark matter. Massive neutrinos are the only alternative among the known particles, but they are essentially ruled out as a universal dark matter candidate, even if they may play a subdominant role as a hot dark matter component. There remains the mystery of what is the physical nature of the dominant cold dark matter component. Something like a heavy stable neutrino, a generic Weakly Interacting Massive Particle (WIMP), could be a reasonable candidate because its present abundance could fall within the expected range,

$$\Omega_{\text{PDM}} h^2 \sim \frac{G^{3/2} T_f^3 h^2}{H_0^2 \langle \sigma_{\text{ann}} v_{\text{rel}} \rangle} = \frac{3 \times 10^{-27} \text{ cm}^3 \text{s}^{-1}}{\langle \sigma_{\text{ann}} v_{\text{rel}} \rangle}.$$ (185)

Here $v_{\text{rel}}$ is the relative velocity of the two incoming dark matter particles and the brackets $\langle \cdot \rangle$ denote a thermal average at the freeze-out temperature, $T_f \approx m_{\text{PDM}} / 20$, when the dark matter particles go out of equilibrium with radiation. The value of $\langle \sigma_{\text{ann}} v_{\text{rel}} \rangle$ needed for $\Omega_{\text{PDM}} \approx 1$ is remarkably close to what one would expect for a WIMP with a mass $m_{\text{PDM}} = 100$ GeV, $\langle \sigma_{\text{ann}} v_{\text{rel}} \rangle \sim \alpha^2 / 8\pi m_{\text{PDM}} \sim 3 \times 10^{-27} \text{ cm}^3 \text{s}^{-1}$. We still do not know whether this is just a coincidence or an important hint on the nature of dark matter.

There are a few theoretical candidates for WIMPs, like the neutralino, coming from supersymmetric extensions of the standard model of particle physics, but at present there is no empirical evidence that such extensions are indeed realized in nature. In fact, the non-observation of supersymmetric particles at current accelerators places stringent limits on the neutralino mass and interaction cross section [55].

If WIMPs constitute the dominant component of the halo of our galaxy, it is expected that some may cross the Earth at a reasonable rate to be detected. The direct experimental search for them rely on elastic WIMP collisions with the nuclei of a suitable target. Dark matter WIMPs move at a typical galactic “virial” velocity of around $200 - 300$ km/s, depending on the model. If their mass is in the range $10 - 100$ GeV, the recoil energy of the nuclei in the elastic collision would be of order 10 keV. Therefore, one should be able to identify such energy depositions in a macroscopic sample of the target.

There are at present three different methods: First, one could search for scintillation light in NaI crystals or in liquid xenon; second, search for an ionization signal in a semiconductor, typically a very pure germanium crystal; and third, use a cryogenic detector at 10 mK and search for a measurable temperature increase of the sample. The main problem with such a type of experiment is the low expected signal rate, with a typical number below 1 event/kg/day. To reduce natural radioactive contamination one must use extremely pure substances, and to reduce the background caused by cosmic rays requires that these experiments be located deep underground.

The best limits on WIMP scattering cross sections come from some germanium experiments, like the Criogetic Dark Matter Search (CDMS) collaboration at Stanford and the Soudan mine [56], as well as from the NaI scintillation detectors of the UK dark matter collaboration (UKDMC) in the Boulby salt mine in England [57], and the DAMA experiment in the Gran Sasso laboratory in Italy [54]. Current experiments already touch the parameter space expected from supersymmetric particles, see Fig. 26, and therefore there is a chance that they actually discover the nature of the missing dark matter. The problem, of course, is to attribute a tentative signal unambiguously to galactic WIMPs rather than to some unidentified radioactive background.

One specific signature is the annual modulation which arises as the Earth moves around the Sun. Therefore, the net speed of the Earth relative to the galactic dark matter halo varies, causing a modulation of the expected counting rate. The DAMA/NaI experiment has actually reported such a modulation signal, from the combined analysis of their 7-year data, see Fig. 25 and Ref. [54], which provides a confidence level of 99.6% for a neutralino mass of $m_\chi = 52^{+10}_{-8}$ GeV and a proton cross section of

---

4For a review of Supersymmetry (SUSY), see Kazakov’s contribution to these Proceedings.

5The time scale of the Sun’s orbit around the center of the galaxy is too large to be relevant in the analysis.
$\xi\sigma_p = 7.2^{+0.4}_{-0.9} \times 10^{-6}$ pb, where $\xi = \rho_\chi / 0.3$ GeV cm$^{-3}$ is the local neutralino energy density in units of the galactic halo density. There has been no confirmation yet of this result from other dark matter search groups. In fact, the CDMS collaboration claims an exclusion of the DAMA region at the 3 sigma level, see Fig. 26. Hopefully in the near future we will have much better sensitivity at low masses from the Cryogenic Rare Event Search with Superconducting Thermometers (CRESST) experiment at Gran Sasso. The CRESST experiment [58] uses sapphire crystals as targets and a new method to simultaneously measure the phonons and the scintillating light from particle interactions inside the crystal, which allows excellent background discrimination. Very recently there has been also the proposal of a completely new method based on a Superheated Droplet Detector (SDD), which claims to have already a similar sensitivity as the more standard methods described above, see Ref. [59].

There exist other indirect methods to search for galactic WIMPs [60]. Such particles could self-annihilate at a certain rate in the galactic halo, producing a potentially detectable background of high energy photons or antiprotons. The absence of such a background in both gamma ray satellites and the Alpha Matter Spectrometer [61] imposes bounds on their density in the halo. Alternatively, WIMPs traversing the solar system may interact with the matter that makes up the Earth or the Sun so that a small fraction of them will lose energy and be trapped in their cores, building up over the age of the universe. Their annihilation in the core would thus produce high energy neutrinos from the center of the Earth or from the Sun which are detectable by neutrino telescopes. In fact, SuperKamiokande already covers a large part of SUSY parameter space. In other words, neutrino telescopes are already competitive with direct search experiments. In particular, the AMANDA experiment at the South Pole [62], which has approximately $10^3$ Cherenkov detectors several km deep in very clear ice, over a volume $\sim 1$ km$^3$, is competitive with the best direct searches proposed. The advantages of AMANDA are also directional, since the arrays of Cherenkov detectors will allow one to reconstruct the neutrino trajectory and thus its source, whether it comes from the Earth or the Sun. AMANDA recently reported the detection of TeV neutrinos [62].
4.3 The age of the universe \( t_0 \)

The universe must be older than the oldest objects it contains. Those are believed to be the stars in the oldest clusters in the Milky Way, globular clusters. The most reliable ages come from the application of theoretical models of stellar evolution to observations of old stars in globular clusters. For about 30 years, the ages of globular clusters have remained reasonably stable, at about 15 Gyr [63]. However, recently these ages have been revised downward [64].

During the 1980s and 1990s, the globular cluster age estimates have improved as both new observations have been made with CCDs, and since refinements to stellar evolution models, including opacities, consideration of mixing, and different chemical abundances have been incorporated [65]. From the theory side, uncertainties in globular cluster ages come from uncertainties in convection models, opacities, and nuclear reaction rates. From the observational side, uncertainties arise due to corrections for dust and chemical composition. However, the dominant source of systematic errors in the globular cluster age is the uncertainty in the cluster distances. Fortunately, the Hipparcos satellite recently provided geometric parallax measurements for many nearby old stars with low metallicity, typical of globular clusters, thus allowing for a new calibration of the ages of stars in globular clusters, leading to a downward revision to \( 10 - 13 \) Gyr [65]. Moreover, there were very few stars in the Hipparcos catalog with both small parallax errors and low metal abundance. Hence, an increase in the sample size could be critical in reducing the statistical uncertainties for the calibration of the globular cluster ages. There are already proposed two new parallax satellites, NASA’s Space Interferometry Mission (SIM) and ESA’s mission, called GAIA, that will give 2 or 3 orders of magnitude more accurate parallaxes than Hipparcos, down to fainter magnitude limits, for several orders of magnitude more stars. Until larger samples are available, however, distance errors are likely to be the largest source of systematic uncertainty to the globular cluster age [29].

The supernovae groups can also determine the age of the universe from their high redshift observations. The high confidence regions in the \((\Omega_M, \Omega_\Lambda)\) plane are almost parallel to the contours of constant age. For any value of the Hubble constant less than \( H_0 = 70 \) km/s/Mpc, the implied age of the universe is greater than 13 Gyr, allowing enough time for the oldest stars in globular clusters to evolve [65]. Integrating over \( \Omega_M \) and \( \Omega_\Lambda \), the best fit value of the age in Hubble-time units is \( H_0 t_0 = 0.93 \pm 0.06 \) or equivalently \( t_0 = 14.1 \pm 1.0 \) (0.65 h\(^{-1}\)) Gyr, see Ref. [7]. Furthermore, a combination of 8 independent recent measurements: CMB anisotropies, type Ia SNe, cluster mass-to-light ratios, cluster abundance evolution, cluster baryon fraction, deuterium-to-hydrogen ratios in quasar spectra, double-lobed radio sources and the Hubble constant, can be used to determine the present age of the universe [66]. The result is shown in Fig. 27, compared to other recent determinations. The best fit value for the age of the universe is, according to this analysis, \( t_0 = 13.4 \pm 1.6 \) Gyr, about a billion years younger than other recent estimates [66].

4.4 Cosmic Microwave Background Anisotropies

The cosmic microwave background has become in the last five years the Holy Grail of Cosmology, since precise observations of the temperature and polarization anisotropies allow in principle to determine the parameters of the Standard Model of Cosmology with very high accuracy. Recently, the WMAP satellite has provided with a very detailed map of the microwave anisotropies in the sky, see Fig. 28, and indeed has fulfilled our expectations, see Table 2.

The physics of the CMB anisotropies is relatively simple [67]. The universe just before recombination is a very tightly coupled fluid, due to the large electromagnetic Thomson cross section \( \sigma_T = 8\pi\sigma^2/3m_e^2 \simeq 0.7 \) barn. Photons scatter off charged particles (protons and electrons), and carry energy, so they feel the gravitational potential associated with the perturbations imprinted in the metric during inflation. An overdensity of baryons (protons and neutrons) does not collapse under the effect of gravity until it enters the causal Hubble radius. The perturbation continues to grow until radiation pressure opposes gravity and sets up acoustic oscillations in the plasma, very similar to sound waves. Since
Fig. 27: The recent estimates of the age of the universe and that of the oldest objects in our galaxy. The last three points correspond to the combined analysis of 8 different measurements, for $h = 0.64, 0.68$ and 7.2, which indicates a relatively weak dependence on $h$. The age of the Sun is accurately known and is included for reference. Error bars indicate 1σ limits. The averages of the ages of the Galactic Halo and Disk are shaded in gray. Note that there isn’t a single age estimate more than 2σ away from the average. The result $t_0 > t_{gal}$ is logically inevitable, but the standard EdS model does not satisfy this unless $h < 0.55$. From Ref. [66].

Moreover, since photons scatter off these baryons, the acoustic oscillations occur also in the photon field and induces a pattern of peaks in the temperature anisotropies in the sky, at different angular scales, see Fig. 29. There are three different effects that determine the temperature anisotropies we observe in the CMB. First, gravity: photons fall in and escape off gravitational potential wells, characterized by $\Phi$ in the comoving gauge, and as a consequence their frequency is gravitationally blue- or red-shifted, $\delta \nu / \nu = \Phi$. If the gravitational potential is not constant, the photons will escape from a larger or smaller potential well than they fell in, so their frequency is also blue- or red-shifted, a phenomenon known as the Rees-Sciama effect. Second, pressure: photons scatter off baryons which fall into gravitational potential wells and the two competing forces create acoustic waves of compression and rarefaction. Finally, velocity: baryons accelerate as they fall into potential wells. They have minimum velocity at maximum compression and rarefaction. That is, their velocity wave is exactly 90° off-phase with the acoustic waves. These waves induce a Doppler effect on the frequency of the photons. The temperature anisotropy induced by these three effects is therefore given by [67]

$$\frac{\delta T}{T}(r) = \Phi(r, t_{dec}) + 2 \int_{t_{dec}}^{t_0} \Phi(r, t) dt + \frac{1}{3} \frac{\delta \rho}{\rho} - \frac{\mathbf{r} \cdot \mathbf{v}}{c}. \quad (186)$$

Metric perturbations of different wavelengths enter the horizon at different times. The largest wavelengths, of size comparable to our present horizon, are entering now. There are perturbations with wavelengths comparable to the size of the horizon at the time of last scattering, of projected size about 1°
in the sky today, which entered precisely at decoupling. And there are perturbations with wavelengths much smaller than the size of the horizon at last scattering, that entered much earlier than decoupling, all the way to the time of radiation-matter equality, which have gone through several acoustic oscillations before last scattering. All these perturbations of different wavelengths leave their imprint in the CMB anisotropies.

The baryons at the time of decoupling do not feel the gravitational attraction of perturbations with wavelength greater than the size of the horizon at last scattering, because of causality. Perturbations with exactly that wavelength are undergoing their first contraction, or acoustic compression, at decoupling. Those perturbations induce a large peak in the temperature anisotropies power spectrum, see Fig. 29. Perturbations with wavelengths smaller than these will have gone, after they entered the Hubble scale, through a series of acoustic compressions and rarefactions, which can be seen as secondary peaks in the power spectrum. Since the surface of last scattering is not a sharp discontinuity, but a region of $\Delta z \sim 100$, there will be scales for which photons, travelling from one energy concentration to another, will erase the perturbation on that scale, similarly to what neutrinos or HDM do for structure on small scales. That is the reason why we don’t see all the acoustic oscillations with the same amplitude, but in fact they decay exponentially towards smaller angular scales, an effect known as Silk damping, due to photon diffusion [68, 67].

From the observations of the CMB anisotropies it is possible to determine most of the parameters of the Standard Cosmological Model with few percent accuracy, see Table 3. However, there are many degeneracies between parameters and it is difficult to disentangle one from another. For instance, as mentioned above, the first peak in the photon distribution corresponds to overdensities that have undergone half an oscillation, that is, a compression, and appear at a scale associated with the size of the horizon at last scattering, about $1^\circ$ projected in the sky today. Since photons scatter off baryons, they will also feel the acoustic wave and create a peak in the correlation function. The height of the peak is proportional to the amount of baryons: the larger the baryon content of the universe, the higher the peak. The position of the peak in the power spectrum depends on the geometrical size of the particle horizon at last scattering. Since photons travel along geodesics, the projected size of the causal horizon at decoupling depends on whether the universe is flat, open or closed. In a flat universe the geodesics are straight lines and, by looking at the angular scale of the first acoustic peak, we would be measuring the actual size of the hori-
zon at last scattering. In an open universe, the geodesics are inward-curved trajectories, and therefore the projected size on the sky appears smaller. In this case, the first acoustic peak should occur at higher multipoles or smaller angular scales. On the other hand, for a closed universe, the first peak occurs at smaller multipoles or larger angular scales. The dependence of the position of the first acoustic peak on the spatial curvature can be approximately given by $l_{\text{peak}} \approx 220 \Omega_0^{-1/2}$, where $\Omega_0 = \Omega_M + \Omega_\Lambda = 1 - \Omega_K$. Present observations by WMAP and other experiments give $\Omega_0 = 1.005 \pm 0.006$ at one standard deviation [20].

The other acoustic peaks occur at harmonics of this, corresponding to smaller angular scales. Since the amplitude and position of the primary and secondary peaks are directly determined by the sound speed (and, hence, the equation of state) and by the geometry and expansion of the universe, they can be used as a powerful test of the density of baryons and dark matter, and other cosmological parameters. With the joined data from WMAP, VSA, CBI and ACBAR, we have rather good evidence of the existence of the second and third acoustic peaks, which confirms one of the most important predictions of inflation — the non-causal origin of the primordial spectrum of perturbations —, and rules out cosmological defects as the dominant source of structure in the universe [69]. Moreover, since the observations of CMB anisotropies now cover almost three orders of magnitude in the size of perturbations, we can determine the much better accuracy the value of the spectral tilt, $n = 0.96 \pm 0.03$, which is compatible with the approximate scale invariant spectrum needed for structure formation, and is a prediction of the simplest models of inflation. Soon after the release of data from WMAP, there was some expectation at the claim of a scale-dependent tilt. Nowadays, with better resolution in the linear matter power spectrum from SDSS [70], we can not conclude that the spectral tilt has any observable dependence on scale.
The microwave background has become also a testing ground for theories of particle physics. In particular, it already gives stringent constraints on the mass of the neutrino, when analysed together with large scale structure observations. Assuming a flat $\Lambda$CDM model, the 2-sigma upper bounds on the sum of the masses of light neutrinos is $\sum m_\nu < 1.0$ eV for degenerate neutrinos (i.e. without a large hierarchy between them) if we don’t impose any priors, and it comes down to $\sum m_\nu < 0.34$ eV if one imposes the bounds coming from the HST measurements of the rate of expansion and the supernova data on the present acceleration of the universe [71]. The final bound on the neutrino density can be expressed as $\Omega_\nu h^2 = \sum m_\nu/93.2$ eV $\leq 0.01$. In the future, both with Planck and with the Atacama Cosmology Telescope (ACT) we will be able to put constraints on the neutrino masses down to the 0.1 eV level.

Moreover, the present data is good enough that we can start to put constraints on the models of inflation that give rise to structure. In particular, multifield models of inflation predict a mixture of adiabatic and isocurvature perturbations, and their signatures in the cosmic microwave background anisotropies and the matter power spectrum of large scale structure are specific and perfectly distinguishable. Nowa-

---

Table 3: **The parameters of the standard cosmological model.** The standard model of cosmology has about 20 different parameters, needed to describe the background space-time, the matter content and the spectrum of metric perturbations. We include here the present range of the most relevant parameters (with 1$\sigma$ errors), as recently determined by WMAP, and the error with which the Planck satellite will be able to determine them in the near future. The rate of expansion is written in units of $H = 100 h$ km/s/Mpc.

<table>
<thead>
<tr>
<th>physical quantity</th>
<th>symbol</th>
<th>WMAP5+SDSS</th>
<th>Planck</th>
</tr>
</thead>
<tbody>
<tr>
<td>total density</td>
<td>$\Omega_0$</td>
<td>$1.005 \pm 0.006$</td>
<td>0.7%</td>
</tr>
<tr>
<td>baryonic matter</td>
<td>$\Omega_B h^2$</td>
<td>$0.0227 \pm 0.0006$</td>
<td>0.6%</td>
</tr>
<tr>
<td>cosmological constant</td>
<td>$\Omega_A$</td>
<td>$0.723 \pm 0.029$</td>
<td>0.5%</td>
</tr>
<tr>
<td>cold dark matter</td>
<td>$\Omega_{CDM}$</td>
<td>$0.231 \pm 0.026$</td>
<td>0.6%</td>
</tr>
<tr>
<td>hot dark matter</td>
<td>$\Omega_{\nu} h^2$</td>
<td>$&lt; 0.0065$ (95% c.l.)</td>
<td>1.0%</td>
</tr>
<tr>
<td>sum of neutrino masses</td>
<td>$\sum m_\nu$</td>
<td>$&lt; 0.3$ eV (95% c.l.)</td>
<td>1.0%</td>
</tr>
<tr>
<td>CMB temperature</td>
<td>$T_0$ (K)</td>
<td>$2.725 \pm 0.002$</td>
<td>0.1%</td>
</tr>
<tr>
<td>baryon to photon ratio</td>
<td>$\eta = n_\nu/n_\gamma$</td>
<td>$(6.15 \pm 0.25) \times 10^{-10}$</td>
<td>0.5%</td>
</tr>
<tr>
<td>baryon to matter ratio</td>
<td>$\Omega_B/\Omega_M$</td>
<td>$0.17 \pm 0.01$</td>
<td>1.0%</td>
</tr>
<tr>
<td>spatial curvature</td>
<td>$</td>
<td>\Omega_K</td>
<td>$</td>
</tr>
<tr>
<td>rate of expansion</td>
<td>$h$</td>
<td>$0.704 \pm 0.024$</td>
<td>0.8%</td>
</tr>
<tr>
<td>age of the universe</td>
<td>$t_0$ (Gyr)</td>
<td>$13.72 \pm 0.14$</td>
<td>0.1%</td>
</tr>
<tr>
<td>age at decoupling</td>
<td>$t_{dec}$ (kyr)</td>
<td>$379 \pm 8$</td>
<td>0.5%</td>
</tr>
<tr>
<td>age at reionization</td>
<td>$t_r$ (Myr)</td>
<td>$180 \pm 100$</td>
<td>5.0%</td>
</tr>
<tr>
<td>spectral amplitude</td>
<td>$A$</td>
<td>$0.93 \pm 0.06$</td>
<td>0.1%</td>
</tr>
<tr>
<td>spectral tilt</td>
<td>$n_s$</td>
<td>$0.96 \pm 0.03$</td>
<td>0.2%</td>
</tr>
<tr>
<td>spectral tilt variation</td>
<td>$dn_s/d \ln k$</td>
<td>$-0.034 \pm 0.028$</td>
<td>0.5%</td>
</tr>
<tr>
<td>tensor-scalar ratio</td>
<td>$r$</td>
<td>$&lt; 0.36$ (95% c.l.)</td>
<td>5.0%</td>
</tr>
<tr>
<td>rms matter fluctuation</td>
<td>$\sigma_s$</td>
<td>$0.811 \pm 0.032$</td>
<td>1.0%</td>
</tr>
<tr>
<td>reionization optical depth</td>
<td>$\tau$</td>
<td>$0.083 \pm 0.016$</td>
<td>5.0%</td>
</tr>
<tr>
<td>redshift of equality</td>
<td>$z_{eq}$</td>
<td>$3262 \pm 137$</td>
<td>5.0%</td>
</tr>
<tr>
<td>redshift of decoupling</td>
<td>$z_{dec}$</td>
<td>$1088 \pm 1$</td>
<td>0.1%</td>
</tr>
<tr>
<td>width of decoupling</td>
<td>$\Delta z_{dec}$</td>
<td>$195 \pm 2$</td>
<td>1.0%</td>
</tr>
<tr>
<td>redshift of reionization</td>
<td>$z_r$</td>
<td>$10.8 \pm 1.4$</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

---

6This mixture is generic, unless all the fields thermalize simultaneously at reheating, just after inflation, in which case the entropy perturbations that would give rise to the isocurvature modes disappear.
days, thanks to precise CMB, LSS and SNIa data, one can put rather stringent limits on the relative fraction and correlation of the isocurvature modes to the dominant adiabatic perturbations [72].

We can summarize this Section by showing the region in parameter space where we stand nowadays, thanks to the recent cosmological observations. We have plotted that region in Fig. 30. One could also superimpose the contour lines corresponding to equal $t_0H_0$ lines, as a cross check. It is extraordinary that only in the last few months we have been able to reduce the concordance region to where it stands today, where all the different observations seem to converge. There are still many uncertainties, mainly systematic; however, those are quickly decreasing and becoming predominantly statistical. In the near future, with precise observations of the anisotropies in the microwave background temperature and polarization anisotropies, thanks to Planck satellite, we will be able to reduce those uncertainties to the level of one percent. This is the reason why cosmologists are so excited and why it is claimed that we live in the Golden Age of Cosmology.

5 THE INFLATIONARY PARADIGM

The hot Big Bang theory is nowadays a very robust edifice, with many independent observational checks: the expansion of the universe; the abundance of light elements; the cosmic microwave background; a predicted age of the universe compatible with the age of the oldest objects in it, and the formation of
structure via gravitational collapse of initially small inhomogeneities. Today, these observations are confirmed to within a few percent accuracy, and have helped establish the hot Big Bang as the preferred model of the universe. All the physics involved in the above observations is routinely tested in the laboratory (atomic and nuclear physics experiments) or in the solar system (general relativity).

However, this theory leaves a range of crucial questions unanswered, most of which are initial conditions’ problems. There is the reasonable assumption that these cosmological problems will be solved or explained by new physical principles at high energies, in the early universe. This assumption leads to the natural conclusion that accurate observations of the present state of the universe may shed light onto processes and physical laws at energies above those reachable by particle accelerators, present or future. We will see that this is a very optimistic approach indeed, and that there are many unresolved issues related to those problems. However, there might be in the near future reasons to be optimistic.

5.1 Shortcomings of Big Bang Cosmology

The Big Bang theory could not explain the origin of matter and structure in the universe; that is, the origin of the matter–antimatter asymmetry, without which the universe today would be filled by a uniform radiation continuously expanding and cooling, with no traces of matter, and thus without the possibility to form gravitationally bound systems like galaxies, stars and planets that could sustain life. Moreover, the standard Big Bang theory assumes, but cannot explain, the origin of the extraordinary smoothness and flatness of the universe on the very large scales seen by the microwave background probes and the largest galaxy catalogs. It cannot explain the origin of the primordial density perturbations that gave rise to cosmic structures like galaxies, clusters and superclusters, via gravitational collapse; the quantity and nature of the dark matter that we believe holds the universe together; nor the origin of the Big Bang itself.

A summary [10] of the problems that the Big Bang theory cannot explain is:

- The global structure of the universe.
  - Why is the universe so close to spatial flatness?
  - Why is matter so homogeneously distributed on large scales?
- The origin of structure in the universe.
  - How did the primordial spectrum of density perturbations originate?
- The origin of matter and radiation.
  - Where does all the energy in the universe come from?
  - What is the nature of the dark matter in the universe?
  - How did the matter-antimatter asymmetry arise?
- The initial singularity.
  - Did the universe have a beginning?
  - What is the global structure of the universe beyond our observable patch?

Let me discuss one by one the different issues:

5.1.1 The Flatness Problem

The Big Bang theory assumes but cannot explain the extraordinary spatial flatness of our local patch of the universe. In the general FRW metric, the parameter \( K \) that characterizes spatial curvature is a free parameter. There is nothing in the theory that determines this parameter a priori. However, it is directly related, via the Friedmann equation, to the dynamics, and thus the matter content, of the universe,

\[
K = \frac{8\pi G}{3} \rho a^2 - H^2 a^2 = \frac{8\pi G}{3} \rho a^2 \left( \frac{\Omega - 1}{\Omega} \right).
\]  

(187)

We can therefore define a new variable,

\[
x \equiv \frac{\Omega - 1}{\Omega} = \text{const.} \frac{1}{\rho a^2},
\]

(188)
whose time evolution is given by

\[
x' = \frac{dx}{dN} = (1 + 3\omega) x,
\]

where \( N = \ln(a/a_i) \) characterizes the number of e-folds of universe expansion \((dN = H dt)\) and where we have used Eq. (18) for the time evolution of the total energy, \( \rho a^3 \). It is only depends on the barotropic ratio \( \omega \). It is clear from Eq. (189) that the phase-space diagram \((x, x')\) presents an unstable critical (saddle) point at \( x = 0 \) for \( \omega > -1/3 \), i.e. for the radiation \((\omega = 1/3)\) and matter \((\omega = 0)\) eras. A small perturbation from \( x = 0 \) will drive the system towards \( x = \pm \infty \). Since we know the universe went through both the radiation era (because of primordial nucleosynthesis) and the matter era (because of structure formation), tiny deviations from \( \Omega = 1 \) would have grown since then, such that today

\[
x_0 = \frac{\Omega_0 - 1}{\Omega_0} = x_{\text{in}} \left(\frac{T_{\text{in}}}{T_{\text{eq}}} \right) (1 + z_{\text{eq}}).
\]

In order that today’s value be in the range \(|\Omega_K| < 0.005\), or \( x_0 \approx 10^{-3} \), it is required that at, say, primordial nucleosynthesis \((T_{\text{NS}} \approx 10^6 T_{\text{eq}})\) its value be

\[
\Omega(t_{\text{NS}}) = 1 \pm 10^{-18},
\]

which represents a tremendous finetuning. Perhaps the universe indeed started with such a peculiar initial condition, but it is epistemologically more satisfying if we give a fundamental dynamical reason for the universe to have started so close to spatial flatness. These arguments were first used by Robert Dicke in the 1960s, much before inflation. He argued that the most natural initial condition for the spatial curvature should have been the Planck scale curvature, \((3)R = 6K/l_P^2\), where the Planck length is \( l_P = (\hbar G/c^3)^{1/2} = 1.62 \times 10^{-33} \) cm, that is, 60 orders of magnitude smaller than the present size of the universe, \( a_0 = 1.38 \times 10^{28} \) cm. A universe with this immense curvature would have collapsed within a Planck time, \( t_P = (\hbar G/c^5)^{1/2} = 5.39 \times 10^{-44} \) s, again 60 orders of magnitude smaller than the present age of the universe, \( t_0 = 4.1 \times 10^{17} \) s. Therefore, the flatness problem is also related to the Age Problem, why is it that the universe is so old and flat when, under ordinary circumstances (based on the fundamental scale of gravity) it should have lasted only a Planck time and reached a size of order the Planck length? As we will see, inflation gives a dynamical reason to such a peculiar initial condition.

### 5.1.2 The Homogeneity Problem

An expanding universe has particle horizons, that is, spatial regions beyond which causal communication cannot occur. The horizon distance can be defined as the maximum distance that light could have travelled since the origin of the universe [15],

\[
d_H(t) = a(t) \int_0^t \frac{dt'}{a(t')} \sim H^{-1}(t),
\]

which is proportional to the Hubble scale.\(^7\) For instance, at the beginning of nucleosynthesis the horizon distance is a few light-seconds, but grows linearly with time and by the end of nucleosynthesis it is a few light-minutes, i.e. a factor 100 larger, while the scale factor has increased only a factor of 10. The fact that the causal horizon increases faster, \( d_H \sim t \), than the scale factor, \( a \sim t^{1/2} \), implies that at any given time the universe contains regions within itself that, according to the Big Bang theory, were never in causal contact before. For instance, the number of causally disconnected regions at a given redshift \( z \) present in our causal volume today, \( d_H(t_0) = a_0 \), is

\[
N_{\text{CD}}(z) \sim \left(\frac{a(t)}{d_H(t)}\right)^3 \sim (1 + z)^{3/2},
\]

which, for the time of decoupling, is of order \( N_{\text{CD}}(z_{\text{dec}}) \sim 10^5 \gg 1 \).

\(^7\)For the radiation era, the horizon distance is equal to the Hubble scale. For the matter era it is twice the Hubble scale.
This phenomenon is particularly acute in the case of the observed microwave background. Information cannot travel faster than the speed of light, so the causal region at the time of photon decoupling could not be larger than $d_H(t_{dec}) \sim 3 \times 10^5$ light years across, or about $1^\circ$ projected in the sky today. So why should regions that are separated by more than $1^\circ$ in the sky today have exactly the same temperature, to within $10^{-5}$ ppm, when the photons that come from those two distant regions could not have been in causal contact when they were emitted? This constitutes the so-called horizon problem, see Fig. 31, and was first discussed by Robert Dicke in the 1970s as a profound inconsistency of the Big Bang theory.

5.2 Cosmological Inflation

In the 1980s, a new paradigm, deeply rooted in fundamental physics, was put forward by Alan H. Guth [74], Andrei D. Linde [75] and others [76, 77, 78], to address these fundamental questions. According to the inflationary paradigm, the early universe went through a period of exponential expansion, driven by the approximately constant energy density of a scalar field called the inflaton. In modern physics, elementary particles are represented by quantum fields, which resemble the familiar electric, magnetic and gravitational fields. A field is simply a function of space and time whose quantum oscillations are interpreted as particles. In our case, the inflaton field has, associated with it, a large potential energy density, which drives the exponential expansion during inflation, see Fig. 32. We know from general relativity that the density of matter determines the expansion of the universe, but a constant energy density acts in a very peculiar way: as a repulsive force that makes any two points in space separate at exponentially large speeds. (This does not violate the laws of causality because there is no information carried along in the expansion, it is simply the stretching of space-time.)

This superluminal expansion is capable of explaining the large scale homogeneity of our observable universe and, in particular, why the microwave background looks so isotropic: regions separated today by more than $1^\circ$ in the sky were, in fact, in causal contact before inflation, but were stretched to
Fig. 32: The inflaton field can be represented as a ball rolling down a hill. During inflation, the energy density is approximately constant, driving the tremendous expansion of the universe. When the ball starts to oscillate around the bottom of the hill, inflation ends and the inflaton energy decays into particles. In certain cases, the coherent oscillations of the inflaton could generate a resonant production of particles which soon thermalize, reheating the universe. From Ref. [73].

cosmological distances by the expansion. Any inhomogeneities present before the tremendous expansion would be washed out. This explains why photons from supposedly causally disconneted regions have actually the same spectral distribution with the same temperature, see Fig. 31.

Moreover, in the usual Big Bang scenario a flat universe, one in which the gravitational attraction of matter is exactly balanced by the cosmic expansion, is unstable under perturbations: a small deviation from flatness is amplified and soon produces either an empty universe or a collapsed one. As we discussed above, for the universe to be nearly flat today, it must have been extremely flat at nucleosynthesis, deviations not exceeding more than one part in $10^{15}$. This extreme fine tuning of initial conditions was also solved by the inflationary paradigm, see Fig. 33. Thus inflation is an extremely elegant hypothesis that explains how a region much, much greater that our own observable universe could have become smooth and flat without recourse to ad hoc initial conditions. Furthermore, inflation dilutes away any “unwanted” relic species that could have remained from early universe phase transitions, like monopoles, cosmic strings, etc., which are predicted in grand unified theories and whose energy density could be so large that the universe would have become unstable, and collapsed, long ago. These relics are diluted by the superluminal expansion, which leaves at most one of these particles per causal horizon, making them harmless to the subsequent evolution of the universe.

The only thing we know about this peculiar scalar field, the inflaton, is that it has a mass and a self-interaction potential $V(\phi)$ but we ignore everything else, even the scale at which its dynamics determines the superluminal expansion. In particular, we still do not know the nature of the inflaton field itself, is it some new fundamental scalar field in the electroweak symmetry breaking sector, or is it just some effective description of a more fundamental high energy interaction? Hopefully, in the near future, experiments in particle physics might give us a clue to its nature. Inflation had its original inspiration in the Higgs field, the scalar field supposed to be responsible for the masses of elementary particles (quarks and leptons) and the breaking of the electroweak symmetry. Such a field has not been found yet, and its discovery at the future particle colliders would help understand one of the truly fundamental problems in physics, the origin of masses. If the experiments discover something completely new and unexpected, it would automatically affect the idea of inflation at a fundamental level.
Fig. 33: The exponential expansion during inflation made the radius of curvature of the universe so large that our observable patch of the universe today appears essentially flat, analogous (in three dimensions) to how the surface of a balloon appears flatter and flatter as we inflate it to enormous sizes. This is a crucial prediction of cosmological inflation that will be tested to extraordinary accuracy in the next few years. From Ref. [77, 73].

6 The Arnowitt-Deser-Misner formalism

The Arnowitt–Deser–Misner formalism gives a (3+1)-splitting of space-time, a foliation in which the four dimensional metric $g_{\mu\nu}$ is parametrized by the three-metric $h_{ij}$ and the lapse and shift functions, $N$ and $N^i$, which describe the evolution of time-like hypersurfaces, with proper interval $ds$, between $x^\alpha = (t, x^i)$ and $x^\alpha + dx^\alpha = (t + dt, x^i + dx^i)$, given by

$$ds^2 = -(Ndt)^2 + h_{ij}(dx^i + N^jdt)(dx^j + N^jdt).$$

The components of the metric thus become

$$g_{00} = -N^2 + h^{ij}N_iN_j, \quad g_{0i} = g_{i0} = N_i, \quad g_{ij} = h_{ij},$$

and inverse metric

$$g^{00} = -N^{-2}, \quad g^{0i} = g^{i0} = N^{-2}N^i, \quad g^{ij} = h^{ij} - N^{-2}N^iN^j,$$

where the 3-metric is used to raise and lower spatial indices, $N^i = h^{ij}N_j$, with $h^{ik}h_{kj} = \delta^i_j$. This splitting corresponds to a 3-hypersurface $\Sigma$ and a timelike unit vector normal to it, with components

$$n_\alpha = (-N, 0), \quad n^\alpha = (N^{-1}, -N^{-1}N^i),$$

satisfying $n_\alpha n^\alpha = -1$. We can then define an intrinsic curvature to the 3-surface, $(3)R_{ij}$, written in terms of the 3-metric $h_{ij}$, as well as an extrinsic curvature, related to the normal vector,

$$K_{ij} = -n_{[ij]} = -N\Gamma^0_{ij} = \frac{1}{2N} \left(2N_{(ij)} - \partial_0 h_{ij}\right),$$
where bars denote 3-space covariant derivatives with connections derived from $h_{ij}$, and subindices in parenthesis denote symmetrization, $2A_{(ij)} = A_{ij} + A_{ji}$, while brackets denote antisymmetrization, $2A_{[ij]} = A_{ij} - A_{ji}$. The traceless part of a tensor is denoted by an overbar. In particular, the trace and traceless parts of the extrinsic curvature are

$$
\bar{K}_{ij} = K_{ij} - \frac{1}{3} K h_{ij}, \quad K = K^i_i = \frac{1}{N} \left[ N_{iji} - \partial_0 \ln \sqrt{h} \right].
$$

(197)

The trace $K$ is a generalization of the Hubble parameter, as will be shown below.

Instead of the coordinate basis ($e_0 = \partial_0$, $e_i = \partial_i$), with 1-forms ($dt$, $dx^i$), we will use a basis with the normal 3-vector $n$ instead of the time vector,

vectors $\quad$ 1-forms

$$
e_n = \frac{1}{N} (\partial_0 - N^i \partial_i) \quad w^n = (n \cdot n) n = N dt$$

(198)

$$
e_i = \partial_i \quad w^i = dx^i + N^i dt.$$

(199)

In this case, for instance, the kinetic term of a scalar field is written as

$$
- \partial_{\mu} \phi \partial^{\mu} \phi = (\Pi^\phi)^2 - (\partial_i \phi)^2,
$$

(200)

where $\Pi^\phi$ is the scalar-field’s conjugate momentum

$$
\Pi^\phi = \frac{1}{N} (\ddot{\phi} - N^i \phi_{,i}).
$$

(201)

The gravitational Lagrangian can be written as

$$
L_G = \sqrt{-g} R = N \sqrt{h} \left( (3) R + K_{ij} K^{ij} - K^2 \right)
$$

(202)

from which we can obtain the conjugate momentum of the metric

$$
\Pi^{ij} = \frac{\partial L_G}{\partial \dot{h}_{ij}} = -\sqrt{h} (K^{ij} - K h^{ij}),
$$

(203)

with trace and traceless parts given by

$$
\Pi = 2 \sqrt{h} K, \quad \bar{\Pi}^{ij} = -\sqrt{h} \bar{K}^{ij}.
$$

(204)

After some algebra it can be shown that the gravitational Lagrangian (202) can be written as

$$
\mathcal{L}_G = N \sqrt{h} (3) R + \frac{N}{\sqrt{h}} \left( \Pi_{ij} \Pi^{ij} - \frac{1}{2} \Pi^2 \right) = \Pi^{ij} h_{ij} - N \mathcal{H} - N_i \mathcal{H}^i - 2 \nabla_i (\Pi^{ij} N_j),
$$

where the lapse and shift functions appear as Lagrange multipliers, and

$$
\mathcal{H}(h_{ij}, \Pi^{ij}) = -\sqrt{h} (3) R + \frac{1}{\sqrt{h}} \left( \Pi_{ij} \Pi^{ij} - \frac{1}{2} \Pi^2 \right),
$$

(205)

$$
\mathcal{H}^i(h_{ij}, \Pi^{ij}) = -2 \Pi^{ij}_{,i}.
$$

(206)

The gravitational Hamiltonian then becomes

$$
\mathcal{H}_G = \Pi^{ij} \dot{h}_{ij} - \mathcal{L}_G = -N \sqrt{h} (3) R + \frac{N}{\sqrt{h}} \left( \Pi_{ij} \Pi^{ij} - \frac{1}{2} \Pi^2 \right) + 2 \Pi^{ij} N_{(ij)},
$$

(207)
and the Hamiltonian and momentum constraints,
\[
\frac{\delta H_G}{\delta N} = \mathcal{H} = 0, \quad (208)
\]
\[
\frac{\delta H_G}{\delta N_i} = \mathcal{H}^i = 0. \quad (209)
\]

While the Hamiltonian evolution equations for the independent variables \(h_{ij}\) and \(\Pi^{ij}\) can be written as
\[
\dot{h}_{ij} = \frac{\delta H_G}{\delta \Pi^{ij}} = -2N K_{ij} + 2N_{\langle ij\rangle}, \quad (210)
\]
\[
\dot{\Pi}^{ij} = -\frac{\delta H_G}{\delta h_{ij}} = -N\sqrt{h}\left(3\Pi^{ij} - \frac{1}{2} R h^{ij}\right) + \frac{N}{2\sqrt{h}} h^{ij} \left(\Pi_{kl}\Pi^{kl} - \frac{1}{2} \Pi^2\right) - \frac{2N}{\sqrt{h}} \left(\Pi^{ij} \Pi^{j}_{\cdot k} \Lambda_{k}^{i} - \frac{1}{2} \Pi_{\cdot k} \Lambda_{k}^{j}\right) + \sqrt{h}\left(N_{\cdot i} h^{ij} - \frac{1}{2} \Pi_{\cdot i} \Lambda_{i}^{j}\right) + \left(N^{k} \Pi^{ij}\right)_{\cdot i} - 2\Pi^{i} N^{j}_{\cdot k}, \quad (211)
\]

With these equations we can evaluate the derivative of the trace \(\Pi\),
\[
\ddot{\Pi} = \dot{\Pi}^{ij} h_{ij} + \Pi^{ij} \dot{h}_{ij} = \frac{1}{2} N\sqrt{h}\left(3\Pi^{ij} - \frac{1}{2} R h^{ij}\right) + \frac{N}{2\sqrt{h}} h^{ij} \left(\Pi_{kl}\Pi^{kl} - \frac{1}{2} \Pi^2\right) - 2\sqrt{h} N^{i}_{\cdot i} + 2\sqrt{h} (K N^{i})_{\cdot i}, \quad (214)
\]

while from (204) we have
\[
\ddot{\Pi} = 2\sqrt{h} \dot{K} + \sqrt{h} K h^{ij} \dot{h}_{ij} = 2\sqrt{h} \dot{K} - 2N \sqrt{h} K^2 + 2\sqrt{h} K N^{i}_{\cdot i}, \quad (212)
\]

and therefore the derivative of the trace of the extrinsic curvature is
\[
\dot{K} - N^{i} K_{\cdot i} = -N^{i}_{\cdot i} + N\left(\frac{1}{4} R + \frac{3}{4} K_{ij} K^{ij} + \frac{1}{2} K^2\right). \quad (215)
\]

Now we can evaluate the derivative of the traceless part \(\Pi^{ij}\). Using the identity
\[
\frac{2N}{\sqrt{h}} \left(\Pi^{ik} \Pi^{j}_{\cdot k} - \frac{1}{2} \Pi \Pi^{ij}\right) = 2N\sqrt{h} \left(K^{ik} \dot{K}_{\cdot k}^{j} - \frac{1}{3} K K^{ij} - \frac{1}{2} K^2 h^{ij}\right),
\]

and \(2N K^{ik} \dot{K}_{\cdot k}^{j} = \dot{K}^{ik} \left(N^{j}_{\cdot k} + N^{k}_{\cdot j}\right) - \left(\frac{3}{2} N K K^{ij} + \dot{K}^{ik} h_{kl} \dot{h}^{lj}\right)\), we have, after some algebra,
\[
\dot{\Pi}^{ij} = \Pi^{ij} - \frac{1}{3} \Pi \dot{h}^{ij} = \Pi^{ij} - \frac{1}{3} \Pi \dot{h}^{ij} = -N\sqrt{h} \left(3\Pi^{ij} - \frac{1}{2} R h^{ij}\right) - 2\sqrt{h} K^{ik} \dot{K}_{\cdot k}^{j} - \frac{2}{3} N K \dot{K}^{ij}
\]
\[
+ \sqrt{h} \left(N_{\cdot i} h^{ij} - \frac{1}{2} \Pi_{\cdot i} \Lambda_{i}^{j}\right) - \sqrt{h} N^{k} \dot{K}^{ij}_{\cdot k} - \sqrt{h} K^{ij} N^{k}_{\cdot k} + 2\sqrt{h} K^{i} N^{j}_{\cdot k}. \quad (216)
\]

Also, from \(\ddot{\Pi}^{ij} = -\sqrt{h} \dot{K}^{ij}\), we obtain
\[
\ddot{\Pi}^{ij} = -\sqrt{h} \dot{K}^{ij} + N\sqrt{h} K \dot{K}^{ij} - \sqrt{h} K^{ij} N^{k}_{\cdot k} \quad (217)
\]

and therefore, comparing the two expressions, we deduce
\[
\dot{K}^{i} - N^{k} \dot{K}^{i}_{\cdot k} + N^{i}_{\cdot k} \dot{K}^{k} - N^{k}_{\cdot j} \dot{K}^{j} = -N^{i}_{\cdot j} + \frac{1}{3} N^{k}_{\cdot k} \delta_{j}^{i} + N\left(3\Pi^{j} + K \dot{K}^{j}\right). \quad (218)
\]
Let us consider now the matter content and write the gravitational action for a scalar field with potential $V(\phi)$ in the ADM formalism as

$$S = \int d^4x \sqrt{-g} \left[ \frac{1}{2\kappa^2} R - \frac{1}{2}(\partial\phi)^2 - V(\phi) \right]$$

$$= \int d^4x N \sqrt{h} \left[ \frac{1}{2\kappa^2} (\frac{3}{2}R + K_{ij}\bar{K}^{ij} - \frac{2}{3}K^2) + \frac{1}{2} [\Pi^0]^2 - \phi_i \phi^i + V(\phi) \right], \quad (219)$$

Variation of the action with respect to $N$ and $N^i$ yields the energy and momentum constraint equations respectively

$$-(\frac{3}{2}R + K_{ij}\bar{K}^{ij} - \frac{2}{3}K^2 + 2\kappa^2 T_{00} = 0, \quad (220)$$

$$\bar{K}^{ij}_{ij} - \frac{2}{3}K^i_k + \kappa^2 T^i_k = 0. \quad (221)$$

Variation with respect to $h_{ij}$ gives the dynamical gravitational field equations, which can be separated into the trace and traceless parts

$$\ddot{K} - N^i K_{i|j} = - N^i_{|j} + N \left( \frac{1}{4}(\frac{3}{2}R + \frac{3}{4}K_{ij}\bar{K}^{ij} + \frac{1}{2}K^2 + \frac{\kappa^2}{2} T \right), \quad (222)$$

$$\ddot{K}^i_j - N^k K^i_{jk|j} + N^i_{|k} \bar{K}^k_j - N^k_{|i} \bar{K}^k_i = - N^i_{|j} + \frac{1}{3} N^i_{|k} \delta^i_j + N \left( \frac{3}{2}\bar{R}^i_j + K \bar{K}^i_j - \kappa^2 \bar{T}^i_j \right). \quad (223)$$

The matter energy-momentum tensor is

$$T_{00} = \frac{1}{2} \Pi^0_i + \frac{1}{2} \phi_k \phi^k + V(\phi), \quad (224)$$

$$T^0_i = \Pi^0_i, \quad (225)$$

$$T^i_j = \phi^i_j + \delta^i_j \left[ \frac{1}{2} \Pi^0_i + \frac{1}{2} \phi_k \phi^k - V(\phi) \right], \quad (226)$$

$$\bar{T}^i_j = \phi^i_j - \frac{1}{3} \phi_k \phi^k \delta^i_j, \quad (227)$$

$$T = \frac{3}{2} \Pi^0_i - \frac{1}{2} \phi_k \phi^k - 3V(\phi) \quad (228)$$

Variation with respect to $\phi$ gives the scalar-field’s equation of motion

$$\frac{1}{N} \left( \bar{\Pi}^0 - N^i \Pi^0_i \right) - K \Pi^0 - \frac{1}{N} \left[ N^i \phi^i - \phi^i_i + \frac{\partial V}{\partial \phi} \right] = 0. \quad (229)$$

It is extremely difficult to solve these highly nonlinear coupled equations in a cosmological scenario without making some approximations. The usual approach is to assume homogeneity of the fields to give a background solution and then linearize the equations to study deviations from spatial uniformity.

The smallness of cosmic microwave background anisotropies gives some justification for this perturbative approach at least in our local part of the Universe. However, there is no reason to believe it will be valid on much larger scales. In fact, the stochastic approach to inflation suggests that the Universe is extremely inhomogeneous on very large scales. Fortunately, in this framework one can coarse-grain over a horizon distance and separate the short- from the long-distance behavior of the fields, where the former communicates with the latter through stochastic forces. The equations for the long-wavelength background fields are obtained by neglecting large-scale gradients, leading to a self-consistent set of equations, as we will discuss in the next section.
6.1 Spatial gradient expansion

It is reasonable to expand in spatial gradients whenever the forces arising from time variations of the fields are much larger than forces from spatial gradients. In linear perturbation theory one solves the perturbation equations for evolution outside of the horizon: a typical time scale is the Hubble time $H^{-1}$, which is assumed to exceed the gradient scale $a/k$, where $k$ is the comoving wave number of the perturbation. Since we are interested in structures on scales larger than the horizon, it is reasonable to expand in $k/(aH)$. In particular, for inflation this is an appropriate parameter of expansion since spatial gradients become exponentially negligible after a few $e$-folds of expansion beyond horizon crossing, $k = aH$.

It is therefore useful to split the field $\phi$ into coarse-grained long-wavelength background fields $\phi(t, x^j)$ and residual short-wavelength fluctuating fields $\delta\phi(t, x^j)$. There is a preferred timelike hypersurface within the stochastic inflation approach in which the splitting can be made consistently, but the definition of the background field will depend on the choice of hypersurface, i.e. the smoothing is not gauge invariant. For stochastic inflation the natural smoothing scale is the comoving Hubble length $(aH)^{-1}$ and the natural hypersurfaces are those on which $aH$ is constant. In that case a fundamental difference between $\phi$ and $\delta\phi$ is that the short-wavelength components are essentially uncorrelated at different times, while long-wavelength components are deterministically correlated through the equations of motion.

In order to solve the equations for the background fields, we will have to make suitable approximations. The idea is to expand in the spatial gradients of $\phi$ and to treat the terms that depend on the fluctuating fields as stochastic forces describing the connection between short- and long-wavelength components. In this Section we will neglect the stochastic forces due to quantum fluctuations of the scalar fields and will derive the approximate equation of motion for the background fields. We retain only those terms that are at most first order in spatial gradients, neglecting such terms as $\phi_i|_i$, $\phi_i\phi^i$, $(^3R)$, $(^3\bar{R})_i$, and $\bar{T}^i$.

We will also choose the simplifying gauge $N^i = 0$ [Note that for the evolution during inflation this is a consequence of the rapid expansion, more than a gauge choice]. The evolution equation (223) for the traceless part of the extrinsic curvature is then $\dot{\bar{K}}^i_j = NK\bar{K}^i_j$. Using $NK = -\partial_t \ln \sqrt{h}$ from (197), we find the solution $\bar{K}^i_j \propto h^{-1/2}$, where $h$ is the determinant of $h_{ij}$. During inflation $h^{-1/2} \equiv a^{-3}$, with $a$ the overall expansion factor, therefore $\bar{K}^i_j$ decays extremely rapidly and can be set to zero in the approximate equations. The most general form of the three-metric with vanishing $\bar{K}^i_j$ is

$$h_{ij} = a^2(t, x^k) \gamma_{ij}(x^k), \quad a(t, x^k) \equiv e^{\alpha(t, x^k)},$$

where the time-dependent conformal factor is interpreted as a space-dependent expansion factor. The time-independent three-metric $\gamma_{ij}$, of unit determinant, describes the three-geometry of the conformally transformed space. Since $a(t, x^k)$ is interpreted as a scale factor, we can substitute the trace $K$ of the extrinsic curvature for the local space-dependent Hubble parameter

$$H(t, x^i) \equiv \frac{1}{N(t, x^i)} \dot{a}(t, x^i) = -\frac{1}{3} K(t, x^i).$$

The energy and momentum constraint equations, (220) and (221), can now be written as

$$H^2 = \frac{\kappa^2}{3} \left[ \frac{1}{2} (\Pi^\phi)^2 + V(\phi) \right],$$  \hspace{2cm} (232)

$$H_{|i} = -\frac{\kappa^2}{2} \Pi^\phi \phi_{|i},$$  \hspace{2cm} (233)
together with the evolution equation (222)

$$\frac{1}{N} \dot{H} = \frac{3}{2} H^2 + \frac{\kappa^2}{6} T = \frac{\kappa^2}{2} (\Pi^\phi)^2,$$  

(234)

where $$T = \frac{3}{2} (\Pi^\phi)^2 - 3V(\phi).$$

In general, $$H$$ is a function of the scalar field and time, $$H(t, x^i) \equiv H(\phi(t, x^i), t)$$. From the momentum constraint (233) we find that the scalar-field's momentum must obey

$$\Pi^\phi = -\frac{2}{\kappa^2} \left( \frac{\partial H}{\partial \phi} \right)_t.$$  

(235)

Comparing Eq. (234) with the time derivative of $$H(\phi, t),$$

$$\frac{1}{N} \left( \frac{\partial H}{\partial t} \right)_x = \Pi^\phi \left( \frac{\partial H}{\partial \phi} \right)_t + \frac{1}{N} \left( \frac{\partial H}{\partial \phi} \right)_\phi = -\frac{\kappa^2}{2} (\Pi^\phi)^2 + \frac{1}{N} \left( \frac{\partial H}{\partial \phi} \right)_\phi,$$  

(236)

we find $$\frac{\partial H}{\partial \phi} = 0$$. In fact, we should not be surprised since this is actually a consequence of the general covariance of the theory.

On the other hand, the scalar field’s equation (229) can be written to first order in spatial gradients as

$$\frac{1}{N} \Pi^\phi + 3H \Pi^\phi + \frac{\partial V}{\partial \phi} = 0.$$  

(237)

We can also show that the conjugate momentum $$\Pi^\phi$$ does not depend explicitly on time, its only dependence comes through $$\phi$$. For this, differentiate Eq. (232) w.r.t. $$\phi$$ to obtain

$$\Pi^\phi \left( \frac{\partial \Pi^\phi}{\partial \phi} \right)_t + 3H \Pi^\phi + \frac{\partial V}{\partial \phi} = 0$$

and compare with (237), where

$$\frac{1}{N} \Pi^\phi = \Pi^\phi \left( \frac{\partial \Pi^\phi}{\partial \phi} \right)_t + \left( \frac{\partial \Pi^\phi}{\partial \phi} \right)_\phi,$$  

(238)

which implies $$\frac{\partial \Pi^\phi}{\partial t} = 0$$.

6.2 Hamilton-Jacobi formalism

We can now summarise what we have learned. The evolution of a general foliation of space-time in the presence of a scalar field fluid can be described solely in terms of the rate of expansion, which is a function of the scalar field only, $$H(\phi(t, x^i))$$, satisfying the Hamiltonian constraint equation:

$$3H^2(\phi) = \frac{2}{\kappa^2} \left( \frac{\partial H}{\partial \phi} \right)^2 + \kappa^2 V(\phi),$$  

(239)

together with the momentum constraint and the evolution of the scale factor,

$$\frac{1}{N} \dot{\phi} = -\frac{2}{\kappa^2} \left( \frac{\partial H}{\partial \phi} \right) = \Pi^\phi$$  

(240)

$$\frac{1}{N} \dot{\phi} = H(\phi),$$  

(241)
as well as the dynamical gravitational and scalar field evolution equations

\[
\frac{1}{N} \dot{H} = -\frac{2}{\kappa^2} \left( \frac{\partial H}{\partial \phi} \right)^2 = -\frac{\kappa^2}{2} (\Pi^\phi)^2, \tag{242}
\]

\[
\frac{1}{N} \dot{\Pi}^\phi = -3H \Pi^\phi - V'(\phi). \tag{243}
\]

Therefore, \( H(\phi) \) is all you need to specify (to second order in field gradients) the evolution of the scale factor and the scalar field during inflation.

These equations are still too complicated to solve for arbitrary potentials \( V(\phi) \). In the next section we will find solutions to them in the slow-roll approximation.

### 6.3 Slow-roll approximation and attractor

Given the complete set of constraints (232)-(233) and evolution equations (234)-(237), we can construct the following parameters,

\[
\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{2}{\kappa^2} \left( \frac{H'(\phi)}{H(\phi)} \right)^2 = -\frac{\partial \ln H}{\partial \ln a}, \tag{244}
\]

\[
\delta \equiv -\frac{\ddot{\phi}}{H \dot{\phi}} = \frac{2}{\kappa^2} \left( \frac{H''(\phi)}{H(\phi)} \right) = -\frac{\partial \ln H'}{\partial \ln a}, \tag{245}
\]

in terms of which we can define the number of e-folds \( N_e \) as

\[
N_e \equiv \ln \frac{a_{\text{end}}}{a(t)} = \int_t^{t_{\text{end}}} H dt = -\frac{\kappa^2}{2} \int_{\phi_i}^{\phi_{\text{end}}} \frac{H(\phi) d\phi}{H'(\phi)}. \tag{246}
\]

In order for inflation to be predictive, you need to ensure that inflation is independent of initial conditions. That is, one should ensure that there is an attractor solution to the dynamics, such that differences between solutions corresponding to different initial conditions rapidly vanish.

Let \( H_0(\phi) \) be an exact, particular, solution of the constraint equation (239), either inflationary or not. Add to it a homogeneous linear perturbation \( \delta H(\phi) \), and substitute into (239). The linear perturbation equation reads \( H'_0(\phi) \delta H'(\phi) = (3\kappa^2/2) H_0 \delta H \), whose general solution is

\[
\delta H(\phi) = \delta H(\phi_i) \exp \left(\frac{3\kappa^2}{2} \int_{\phi_i}^{\phi} \frac{H_0(\phi) d\phi}{H_0'(\phi)}\right) = \delta H(\phi_i) \exp(-3\Delta N), \tag{247}
\]

where \( \Delta N = N_i - N > 0 \), and we have used (246) with the particular solution \( H_0(\phi) \). This means that any deviation from the attractor dies away exponentially fast. This ensures that we can effectively reduce our two-dimensional space \( (\phi, \Pi^\phi) \) to just a single trajectory in phase space.

As a consequence, regardless of the initial condition, the attractor behaviour implies that late-time solutions are the same up to a constant time shift, which cannot be measured.

#### 6.3.1 An example: Power-law inflation

An exponential potential is a particular case where the attractor can be found explicitly and one can study the approach to it, for an arbitrary initial condition.

Consider the inflationary potential

\[
V(\phi) = V_0 e^{-\beta \kappa \phi}, \tag{248}
\]
with $\beta \ll 1$ for inflation to proceed. A particular solution to the Hamiltonian constraint equation (239) is

$$H_{\text{att}}(\phi) = H_0 e^{-\frac{1}{2} \beta \kappa \phi}, \quad (249)$$

$$H_0^2 = \frac{\kappa^2}{3} V_0 \left( 1 - \frac{\beta^2}{6} \right)^{-1}. \quad (250)$$

This model corresponds to an inflationary universe with a scale factor that grows like

$$a(t) \sim t^p, \quad p = \frac{2}{\beta^2} \gg 1. \quad (251)$$

The slow-roll parameters are both constant,

$$\epsilon = \frac{2}{\kappa^2} \left( \frac{H'(\phi)}{H(\phi)} \right)^2 = \frac{\beta^2}{2} = \frac{1}{p} \ll 1, \quad (252)$$

$$\delta = \frac{2}{\kappa^2} \left( \frac{H''(\phi)}{H(\phi)} \right) = \frac{\beta^2}{2} = \frac{1}{p} \ll 1, \quad (253)$$

$$\xi = \frac{4}{\kappa^4} \left( \frac{H' H'''(\phi)}{H^2(\phi)} \right) = \frac{\beta^4}{4} = \frac{1}{p^2} \ll 1. \quad (254)$$

All trajectories tend to the attractor (249), while we can also write down the solution corresponding to the slow-roll approximation, $\epsilon = \delta = 0$,

$$H^2_{\text{SR}}(\phi) = \frac{\kappa^2}{3} V_0 e^{-\beta \kappa \phi}, \quad (255)$$

which differs from the actual attractor by a tiny constant factor, $3p/(3p - 1) \simeq 1$, responsible for a constant time-shift which cannot be measured.

### 7 Homogeneous scalar field dynamics

In this subsection I will describe the theoretical basis for the phenomenon of inflation. Consider a scalar field $\phi$, a singlet under any given interaction, with an effective potential $V(\phi)$. The Lagrangian for such a field in a curved background is

$$S_{\text{inf}} = \int d^4 x \sqrt{-g} \mathcal{L}_{\text{inf}}, \quad \mathcal{L}_{\text{inf}} = -\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi), \quad (256)$$

whose evolution equation in a Friedmann-Robertson-Walker metric and for a homogeneous field $\phi(t)$ is given by

$$\ddot{\phi} + 3H \dot{\phi} + V'(\phi) = 0, \quad (257)$$

where $H$ is the rate of expansion, together with the Einstein equations,

$$H^2 = \frac{\kappa^2}{3} \left( \frac{1}{2} \dot{\phi}^2 + V(\phi) \right), \quad (258)$$

$$\dot{H} = -\frac{\kappa^2}{2} \ddot{\phi}, \quad (259)$$

where $\kappa^2 \equiv 8\pi G$. The dynamics of inflation can be described as a perfect fluid with a time dependent pressure and energy density given by

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad (260)$$

$$p = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (261)$$
The field evolution equation (257) can then be written as the energy conservation equation,
\[ \dot{\rho} + 3H(\rho + p) = 0. \] (262)
If the potential energy density of the scalar field dominates the kinetic energy, \( V(\phi) \gg \dot{\phi}^2 \), then we see that
\[ p \simeq -\rho \quad \Rightarrow \quad \rho \simeq \text{const.} \quad \Rightarrow \quad H(\phi) \simeq \text{const.}, \] (263)
which leads to the solution
\[ a(t) \sim \exp(Ht) \quad \Rightarrow \quad \frac{\dot{a}}{a} > 0 \quad \text{accelerated expansion}. \] (264)
Using the definition of the number of e-folds, \( N = \ln(a/a_i) \), we see that the scale factor grows exponentially, \( a(N) = a_i \exp(N) \). This solution of the Einstein equations solves immediately the flatness problem. Recall that the problem with the radiation and matter eras is that \( \Omega = 1 \) \((x = 0)\) is an unstable critical point in phase-space. However, during inflation, with \( p \simeq -\rho \Rightarrow \omega \simeq -1 \), we have that \( 1 + 3\omega \geq 0 \) and therefore \( x = 0 \) is a stable attractor of the equations of motion, see Eq. (189). As a consequence, what seemed an ad hoc initial condition, becomes a natural prediction of inflation. Suppose that during inflation the scale factor increased \( N \) e-folds, then
\[ x_0 = x_{\text{in}} e^{-2N} \left( \frac{T_{\text{rh}}}{T_{\text{eq}}} \right)^2 (1 + z_{\text{eq}}) \simeq e^{-2N} 10^{56} \leq 1 \quad \Rightarrow \quad N \geq 65, \] (265)
where we have assumed that inflation ended at the scale \( V_{\text{end}} \), and the transfer of the inflaton energy density to thermal radiation at reheating occurred almost instantaneously\(^8\) at the temperature \( T_{\text{rh}} \sim V_{\text{end}}^{1/4} \sim 10^{15} \) GeV. Note that we can now have initial conditions with a large uncertainty, \( x_{\text{in}} \simeq 1 \), and still have today \( x_0 \simeq 1 \), thanks to the inflationary attractor towards \( \Omega = 1 \). This can be understood very easily by realizing that the three curvature evolves during inflation as
\[ \left(3\right)R = \frac{6K}{a^2} = \left(3\right)R_{\text{in}} e^{-2N} \rightarrow 0, \quad \text{for } N \gg 1. \] (266)
Therefore, if cosmological inflation lasted over 65 e-folds, as most models predict, then today the universe (or at least our local patch) should be exactly flat, a prediction that has been tested with great (2\%) accuracy by WMAP from observations of temperature anisotropies in the microwave background.

Furthermore, inflation also solves the homogeneity problem in a spectacular way. First of all, due to the superluminal expansion, any inhomogeneity existing prior to inflation will be washed out,
\[ \delta_k \sim \left( \frac{k}{aH} \right)^2 \Phi_k \propto e^{-2N} \rightarrow 0, \quad \text{for } N \gg 1. \] (267)
Moreover, since the scale factor grows exponentially, while the horizon distance remains essentially constant, \( d_H(t) \sim H^{-1} = \text{const.} \), any scale within the horizon during inflation will be stretched by the superluminal expansion to enormous distances, in such a way that at photon decoupling all the causally disconnected regions that encompass our present horizon actually come from a single region during inflation, about 65 e-folds before the end. This is the reason why two points separated more than 1° in the sky have the same backbody temperature, as observed by the COBE satellite: they were actually in causal contact during inflation. There is at present no other proposal known that could solve the homogeneity problem without invoking an acausal mechanism like inflation.

\(^8\)There could be a small delay in thermalization, due to the intrinsic inefficiency of reheating, but this does not change significantly the required number of e-folds.
Finally, any relic particle species (relativistic or not) existing prior to inflation will be diluted by the expansion,
\[
\rho_M \propto a^{-3} \sim e^{-3N} \quad \rightarrow \quad 0, \quad \text{for } N \gg 1,
\]
\[
\rho_R \propto a^{-4} \sim e^{-4N} \quad \rightarrow \quad 0, \quad \text{for } N \gg 1.
\]
Note that the vacuum energy density \(\rho_v\) remains constant under the expansion, and therefore, very soon it is the only energy density remaining to drive the expansion of the universe.

### 7.1 The slow-roll approximation

In order to simplify the evolution equations during inflation, we will consider the slow-roll approximation (SRA). Suppose that, during inflation, the scalar field evolves very slowly down its effective potential, then we can define the slow-roll parameters,
\[
\epsilon \equiv -\frac{\ddot{H}}{H^2} = \frac{\kappa^2}{2} \frac{\dot{\phi}^2}{H^2} \ll 1,
\]
\[
\delta \equiv -\frac{\ddot{\phi}}{H \dot{\phi}} \ll 1,
\]
\[
\xi \equiv \frac{\dddot{\phi}}{H^2 \dot{\phi}} - \delta^2 \ll 1.
\]
It is easy to see that the condition
\[
\epsilon < 1 \iff \frac{\ddot{a}}{a} > 0
\]
characterizes inflation: it is all you need for superluminal expansion, i.e. for the horizon distance to grow more slowly than the scale factor, in order to solve the homogeneity problem, as well as for the spatial curvature to decay faster than usual, in order to solve the flatness problem.

The number of \(e\)-folds during inflation can be written with the help of Eq. (270) as
\[
N = \ln \frac{a_{\text{end}}}{a_i} = \int_{t_i}^{t_e} H dt = \int_{\phi_i}^{\phi_e} \frac{\kappa d\phi}{\sqrt{2 \epsilon(\phi)}},
\]
which is an exact expression in terms of \(\epsilon(\phi)\).

In the limit given by Eqs. (270), the evolution equations (257) and (258) become
\[
H^2 \left(1 - \frac{\epsilon}{3}\right) \simeq H^2 = \frac{\kappa^2}{3} V(\phi),
\]
\[
3H \dot{\phi} \left(1 - \frac{\delta}{3}\right) \simeq 3H \dot{\phi} = -V'(\phi).
\]
Note that this corresponds to a reduction of the dimensionality of phase-space from two to one dimensions, \(H(\phi, \dot{\phi}) \rightarrow H(\phi)\). In fact, it is possible to prove a theorem, for single-field inflation, which states that the slow-roll approximation is an attractor of the equations of motion, and thus we can always evaluate the inflationary trajectory in phase-space within the SRA, therefore reducing the number of initial conditions to just one, the initial value of the scalar field. If \(H(\phi)\) only depends on \(\phi\), then
\[ H'(\phi) = -\kappa^2 \dot{\phi}/2 \] and we can rewrite the slow-roll parameters \(270\) as

\[
\epsilon = \frac{2}{\kappa^2} \left( \frac{H'(\phi)}{H(\phi)} \right)^2 \simeq \frac{1}{2\kappa^2} \left( \frac{V'(\phi)}{V(\phi)} \right)^2 \equiv \epsilon_V \ll 1, \quad (277)
\]

\[
\delta = \frac{2}{\kappa^2} \frac{H''(\phi)}{H(\phi)} \simeq \frac{1}{\kappa^2} \frac{V''(\phi)}{V(\phi)} - \frac{1}{2\kappa^2} \left( \frac{V'(\phi)}{V(\phi)} \right)^2 \equiv \eta_V - \epsilon_V \ll 1, \quad (278)
\]

\[
\xi = \frac{4}{\kappa^4} \frac{H'(\phi) H'''(\phi)}{H^2(\phi)} \simeq \frac{1}{\kappa^4} \frac{V'(\phi) V'''(\phi)}{V^2(\phi)} - \frac{3}{2\kappa^4} \frac{V''(\phi)}{V(\phi)} \left( \frac{V'(\phi)}{V(\phi)} \right)^2 \\
+ \frac{3}{4\kappa^4} \left( \frac{V'(\phi)}{V(\phi)} \right)^4 \equiv \xi_V - 3\eta_V \epsilon_V + 3\epsilon_V^2 \ll 1. \quad (279)
\]

These expressions define the new slow-roll parameters \(\epsilon_V\), \(\eta_V\) and \(\xi_V\). The number of \(e\)-folds can also be rewritten in this approximation as

\[
N \simeq \int_{\phi_i}^{\phi_e} \frac{\kappa d\phi}{\sqrt{2\epsilon_V(\phi)}} = \kappa^2 \int_{\phi_i}^{\phi_e} \frac{V(\phi)}{V'(\phi)} \, d\phi, \quad (280)
\]

a very useful expression for evaluating \(N\) for a given effective scalar potential \(V(\phi)\).

8 The origin of density perturbations

If cosmological inflation made the universe so extremely flat and homogeneous, where did the galaxies and clusters of galaxies come from? One of the most astonishing predictions of inflation, one that was not even expected, is that quantum fluctuations of the inflaton field are stretched by the exponential expansion and generate large-scale perturbations in the metric. Inflaton fluctuations are small wave packets of energy that, according to general relativity, modify the space-time fabric, creating a whole spectrum of curvature perturbations. The use of the word spectrum here is closely related to the case of light waves propagating in a medium: a spectrum characterizes the amplitude of each given wavelength. In the case of inflation, the inflaton fluctuations induce waves in the space-time metric that can be decomposed into different wavelengths, all with approximately the same amplitude, that is, corresponding to a scale-invariant spectrum. These patterns of perturbations in the metric are like fingerprints that unequivocally characterize a period of inflation. When matter fell in the troughs of these waves, it created density perturbations that collapsed gravitationally to form galaxies, clusters and superclusters of galaxies, with a spectrum that is also scale invariant. Such a type of spectrum was proposed in the early 1970s (before inflation) by Harrison and Zel’dovich [26], to explain the distribution of galaxies and clusters of galaxies on very large scales in our observable universe. Perhaps the most interesting aspect of structure formation is the possibility that the detailed knowledge of what seeded galaxies and clusters of galaxies will allow us to test the idea of inflation.

8.1 Linear perturbations on a homogeneous background

We will now choose a different kind of expansion. Instead of assuming that gradients are small on large scales, we will expand in the amplitude of the metric and scalar field perturbations. Then the linearized equations will be valid both on scales larger and smaller than the horizon, as long as the amplitude of the perturbations is not too large. This approximation will fail both in the limit of very large scales, where the stochastic approach to inflation suggest that curvature perturbations become of order unity or even larger, and in the limit of very small scales, well inside the horizon, where non-linear gravitational collapse makes the density contrast of matter increase beyond perturbation theory to form collapsed structures like stars and galaxies.

The rapid expansion during inflation quickly makes the comoving trajectories orthogonal to the spatial hypersurfaces \(\Sigma\). This means that the shift function quickly becomes negligible, and we can
choose \( g_{0i} = 0 \). In that case, we can write the other metric components to first order in perturbations as

\[
g_{00} = -\left(1 + 2\Phi(t, x)\right),
\]
(281)

\[
g_{ij} = \left(1 - 2\Psi(t, x)\right) a^2(t) \delta_{ij},
\]
(282)

\[
\phi = \phi_0 + \delta\phi(t, x),
\]
(283)

where we are assuming a vanishing homogeneous 3-curvature, \( K = 0 \); i.e. the universe is spatially flat on large scales due to the tremendous expansion of the scale factor. These metric components correspond to a lapse function and a metric with intrinsic and extrinsic curvatures given to first order in perturbations by

\[
N = 1 + \Phi, \quad \sqrt{h} = (1 - 3\Psi) a^3,
\]
(284)

\[
K = -\frac{1}{N} \partial_t \ln \sqrt{h} = -3H(1 - \Phi) + 3\Psi,
\]
(285)

\[
\dot{K}^i_j = -\Phi \dot{\Psi} a^2 \delta^i_j = 0,
\]
(286)

\[
(3) R = \frac{4}{a^2} \Psi^i_i, \quad (3) \ddot{R}^i_j = \frac{1}{a^2} \left(\Psi^i_{|j} - \frac{1}{3} \Psi^{|k}_i \delta^i_j\right),
\]
(287)

\[
\Pi^\phi = (1 - \Phi)(\dot{\phi}_0 + \delta \phi) = \dot{\phi}_0 + \delta \phi - \Phi \dot{\phi}_0.
\]
(288)

The energy momentum tensor to first order, and taking into account all gradient terms is

\[
T_{00} = \frac{1}{2} \dot{\phi}_0^2 + \dot{\phi}_0 \dot{\phi} - \dot{\phi}_0^2 \Phi + V(\phi_0) + V'(\phi) \delta \phi,
\]
(289)

\[
T^0_i = \dot{\phi}_0 \delta \phi_i,
\]
(290)

\[
T^i_j = \delta \phi^j_i \delta \phi_{|j} - \frac{1}{3} \delta \phi^{|k}_i \delta \phi^k_j \delta^i_j = 0,
\]
(291)

\[
T = \frac{3}{2} \dot{\phi}_0^2 + 3\dot{\phi}_0 \delta \phi - 3\dot{\phi}_0^2 \Phi - 3V(\phi_0) - 3V'(\phi) \delta \phi
\]
(292)

With these ingredients, we can compute the constraint and evolution equations. Let us start with the energy constraint equation (220), to first order in perturbation theory, \( K^2/3 + (3) R/2 = \kappa^2 T_{00} \), with \( (3) R = (4/a^2) \Psi^i_i \),

\[
\frac{1}{a^2} \Psi^i_i - 3H \dot{\Psi} - \dot{H} \Phi - 3H^2 \Phi = \frac{\kappa_2}{2} \left( \dot{\phi}_0 \dot{\phi} + V'(\phi) \delta \phi \right)
\]
(293)

where we have used the zero-th order (homogeneous) equations

\[
H^2 = \frac{\kappa_2^2}{3} \left( \frac{1}{2} \dot{\phi}_0^2 + V(\phi_0) \right),
\]
(294)

\[
\dot{H} = -\frac{\kappa_2^2}{2} \dot{\phi}_0^2.
\]
(295)

In a similar way, the evolution equation for \( K \), (222), gives

\[
\ddot{\Psi} + 3H \dot{\Psi} + H \dot{\Phi} + 3H^2 \Phi = \frac{1}{3a^2} (\Psi - \Phi)^i_i + \frac{\kappa_2^2}{2} \left( \dot{\phi}_0 \delta \phi - V'(\phi) \delta \phi \right).
\]
(296)
The momentum constrain equation (221) reads, to first order,
\[(\dot{\Psi} + H \Phi)_i = \frac{\kappa^2}{2} \dot{\phi}_0 \delta \phi_i.\]  
(297)

While the evolution equation for the traceless part \( \bar{K}_{ij}, \) (223), is
\[(\Psi - \Phi)_{ij} - \frac{1}{3} (\Psi - \Phi)_k \delta_{ij} = \kappa^2 a^2 \bar{T}_{ij} = 0 \quad \Rightarrow \quad \Phi = \Psi,\]  
(298)
which constitutes an important constraint, implying that the two Newtonian potentials do not differ, to first order in perturbation theory, a result that agrees with the PPN formalism of general relativity.

Substituting into (293) and (296), we end up with the metric perturbations’ equations
\[\ddot{\Phi} + 4H \dot{\Phi} + H \dot{\Phi} + 3H^2 \Phi = \frac{\kappa^2}{2} (\dot{\phi}_0 \delta \phi - V'(\phi) \delta \phi),\]  
(299)

\[-\frac{1}{a^2} \ddot{\phi}_0^{|i|} + 3H \dot{\Phi} + H \dot{\Phi} + 3H^2 \Phi = -\frac{\kappa^2}{2} (\dot{\phi}_0 \dot{\phi} + V'(\phi) \delta \phi),\]  
(299)
\[\ddot{\Phi} + H \Phi = \frac{\kappa^2}{2} \dot{\phi}_0 \delta \phi,\]
together with the evolution equation for the scalar field perturbation,
\[\ddot{\delta \phi} + 3H \dot{\delta \phi} - \frac{1}{a^2} \ddot{\phi}_0^{|i|} + V'''(\phi_0) \delta \phi = 4 \dot{\phi}_0 \dot{\Phi} - 2V'(\phi_0) \Phi,\]  
(300)

where we have used the homogeneous field equation
\[\dddot{\phi}_0 + 3H \dot{\phi}_0 + V'(\phi_0) = 0.\]  
(301)

### 8.2 Gauge invariant linear perturbation theory

The unperturbed (background) FRW metric can be described by a scale factor \( a(t) \) and a homogeneous density field \( \rho(t) \),
\[ds^2 = a^2(\eta)[-d\eta^2 + \gamma_{ij} dx^i dx^j],\]  
(302)
where \( \eta \) is the conformal time \( \eta = \int \frac{dt}{a(t)} \) and the background equations of motion can be written as
\[\mathcal{H}^2 = a^2 H^2 = \frac{\kappa^2}{3} a^2 \rho - K,\]  
(303)
\[\mathcal{H}' - \mathcal{H}^2 = a^2 \dot{H} = K - \frac{\kappa^2}{2} a^2 (\rho + p),\]  
(304)
where \( \mathcal{H} = aH. \)

The most general line element, in linear perturbation theory, with both scalar, vector and tensor metric perturbations, is given by
\[ds^2 = a^2(\eta) \left\{- (1 + 2\psi) d\eta^2 + 2(B_{ij} - S_{ij}) dx^i dx^j \right\} + \left[ (1 - 2\psi) \gamma_{ij} + 2E_{ij} + 2 F_{(ij)} + h_{ij} \right] dx^i dx^j \} .\]  
(305)
The indices \( \{i, j\} \) label the three-dimensional spatial coordinates with metric \( \gamma_{ij}, \) and the \( |i \) denotes covariant derivative with respect to that metric. The vector perturbation is transverse \( \gamma^{ij} S_{ij} = \gamma^{ij} E_{ij} = 0, \) and the tensor perturbation \( h_{ij} \) corresponds to a symmetric transverse traceless gravitational wave, \( \nabla^i h_{ij} = \gamma^{ij} h_{ij} = 0. \) In total, these correspond to \( n(n + 1)/2 = 10 \) independent degrees of freedom, 4 scalars, 2 vectors (3 components − 1 transverse condition each) and 2 tensor (6 components − 3 transverse conditions − 1 trace condition).
8.3 General coordinate transformations and perturbation theory

Not all 10 degrees of freedom are physical. As we know, there is a gauge invariance of the theory under general coordinate transformations. The homogeneity of the FRW space-time gives a natural choice of coordinates in the absence of perturbations; however, the presence of first-order perturbations allow a general coordinate (gauge) transformation,

\[ x^\mu \rightarrow \tilde{x}^\mu = x^\mu + \xi^\mu \]

with arbitrary scalar functions \( \xi^0 \) and \( \xi \). The vector function \( \xi^i \) is a transverse field, \( \xi^i_i = 0 \). After this gauge transformation, \( \xi^0 \) determines the choice of constant—\( \eta \)—hypersurfaces, while \( \xi^i \) and \( \xi^i \) select the spatial coordinates within these hypersurfaces. The choice of coordinates is arbitrary to first order and definitions of first-order metric and matter perturbations are thus coordinate (gauge)-dependent. The result of the gauge transformation (306) acting on any tensor \( Q \) is that of the Lie derivative of the background value \( Q_0 \) of that physical quantity,

\[ \delta \tilde{Q} = \delta Q - L_\xi Q_0. \] (307)

Alternatively, we can obtain the transformed metric components by perturbing the line element,

\[
\begin{align*}
\eta' & = \eta + \xi^0(\eta, x^k), \\
\xi^i & = x^i + \gamma^{ij} \xi_j(\eta, x^k) + \xi^i(\eta, x^k), \\
\end{align*}
\] (306)

Substituting them into the line element and using \( a(\eta) = a(\eta) - a'(\eta)\xi^0 \), we get the line element in the new coordinate system, to first order in metric and coordinate transformations,

\[ ds^2 = a^2(\tilde{\eta}) \left\{ -1 + 2(\phi - \mathcal{H}\xi^0 - \xi^0') \right\} d\tilde{\eta}^2 + 2 \left\{ (B - \xi^0 - \xi^i)_{ij} - (S + \xi^i) \right\} d\tilde{\eta} d\tilde{x}^i + \left\{ 1 - 2(\psi + \mathcal{H}\xi^0) \right\} \gamma_{ij} + 2(E - \xi)_{ij} + 2 \left\{ F_{(ij)} - \xi_{(ij)} \right\} + h_{ij} \right\} d\tilde{x}^i d\tilde{x}^j. \]

Since \( ds^2 = d\tilde{s}^2 \) is invariant under general coordinate transformations, we can read off the transformation equations for the metric perturbations by writing down the new line element with the new metric perturbations as

\[ d\tilde{s}^2 = a^2(\tilde{\eta}) \left\{ -(1 + 2\tilde{\phi}) d\tilde{\eta}^2 + 2(\tilde{B}_i - \tilde{S}_i) d\tilde{\eta} d\tilde{x}^i + \left( 1 - 2\tilde{\psi} + \mathcal{H}\xi^0 \right) \left( \tilde{E}_{ij} + 2\tilde{F}_{(ij)} + \tilde{h}_{ij} \right) d\tilde{x}^i d\tilde{x}^j \right\}. \]

Thus, the gauge transformation of scalar perturbations becomes

\[ \tilde{\phi} = \phi - \mathcal{H}\xi^0 - \xi^0', \]

\[ \tilde{B} = B + \xi^0 - \xi^i, \]

\[ \tilde{\psi} = \psi + \mathcal{H}\xi^0, \]

\[ \tilde{\xi} = E - \xi, \]

and of vector perturbations is

\[ \tilde{S}_i = S_i + \xi^i, \]

\[ \tilde{F}_i = F_i - \xi_i, \]

and finally, tensor perturbations remain invariant

\[ \tilde{h}_{ij} = h_{ij}. \]

Alternatively, these metric transformations could have been obtained from the general expression \( \tilde{h}_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu \) for a coordinate change (306).
8.4 Time slicing and spatial hypersurfaces

Eventually we will have to relate metric perturbations to physical observable quantities. To linear perturbation theory, the unit time-like vector field orthogonal to constant $-\eta$ hypersurfaces $n^\mu$ is

$$n^\mu = (N^{-1}, -N^{-1} N^i) = \frac{1}{a} (1 - \phi, -B^i + S^i), \quad (316)$$

and therefore, the lapse and shift functions, to linear order, are given by

$$N = a(1 + \phi), \quad N^i = B^i - S^i. \quad (318)$$

With these, we can decompose the congruence of worldlines into

$$\Theta_{\mu\nu} = n_{\mu;\nu} = \frac{1}{3} \Theta P_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu} - a_\mu n_{\nu}, \quad (319)$$

where the projection operator, $P^\mu_\nu = \delta^\mu_\nu + n^\mu n_\nu$, projects any tensor into a plane orthogonal to $n_\mu$. Note that in general $n^\mu$ is not necessarily identical to $u^\mu = dx^\mu/d\tau$, the tangent vector to the worldline, and thus $n^\mu$ does not follow a geodesic. The tensor $\Theta_{\mu\nu}$ measures the extent to which neighbouring worldlines deviate from remaining parallel.

Imagine a sphere of test particles centered at some point, and describe the evolution of the particles with respect to the central worldline. The 4 components of the congruence have physical meaning. The trace gives the expansion rate of the congruence,

$$\Theta = n^\mu_{;\mu}, \quad (320)$$

and characterizes the growth in the volume of the sphere; it’s a scalar. The traceless symmetric part,

$$\sigma_{\mu\nu} = \frac{1}{2} P^\alpha_\mu P^\beta_\nu (n_{\alpha;\beta} + n_{\beta;\alpha}) - \frac{1}{3} \Theta P_{\mu\nu}, \quad (321)$$

is the shear tensor of the congruence, and represents the distortion in the shape of the set of test particles; it describes how the sphere is deformed into an ellipsoid. The antisymmetric part,

$$\omega_{\mu\nu} = \frac{1}{2} P^\alpha_\mu P^\beta_\nu (n_{\alpha;\beta} - n_{\beta;\alpha}), \quad (322)$$

is the vorticity tensor, and describes the rotation of the sphere of test particles around a center axis. In our case, $n_\mu = (-N, 0)$, the congruence is always orthogonal by definition to a family of hypersurfaces and therefore it cannot sustain vorticity.

Finally, if the normal vectors do not follow geodesics of the metric, then there is also an acceleration,

$$a_\mu = n^\nu n_{\mu;\nu}. \quad (323)$$

These geometric quantities allow one to obtain the evolution of the congruence. By computing the covariant derivative of $\Theta_{\mu\nu}$ along the worldlines, and taking the trace, ones finds the famous Raychaudhury equation,

$$\frac{d\Theta}{d\tau} = -\frac{1}{3} \Theta^2 - \sigma_{\mu\nu} \sigma^{\mu\nu} + \omega_{\mu\nu} \omega^{\mu\nu} - R_{\mu\nu} n^\mu n^\nu + a^{\mu}_{;\mu}. \quad (324)$$

Note that on spatial hypersurfaces, the shear, vorticity, expansion and acceleration coincide with their Newtonian counterparts. Moreover, with these definitions one can obtain the extrinsic curvature to the spatial hypersurface, $K_{\mu\nu} = -P^\alpha_\mu P^\beta_\nu n_{\alpha;\beta}$. 
Let us come back to linear perturbation theory over a FRW space-time. The scalar perturbations give

\[ \Theta = \frac{1}{a} \left[ 3\mathcal{H}(1 - \phi) - 3\psi' - \nabla^2(B - E') \right], \quad (325) \]

\[ a_i = \phi_i, \quad a_0 = 0, \quad (326) \]

\[ \sigma_{ij} = a \left( \sigma_{ij} - \frac{1}{3} \gamma_{ij} \nabla^2 \sigma \right), \quad \sigma = E' - B, \quad (327) \]

where \( \sigma \) is the scalar shear. The intrinsic spatial curvature on hypersurfaces of constant conformal time \( \eta \) is

\[ (3)R = \frac{6K}{a^2} + \frac{12K}{a^2} \psi + \frac{4}{a^2} \nabla^2 \psi. \quad (328) \]

For a perturbation with wavenumber \( k \), \( \nabla^2 \psi = -k^2 \psi \), we have

\[ \delta(3)R = \frac{4}{a^2} (3K - k^2) \psi. \quad (329) \]

The vector perturbations have \( a_{(v)}^\mu = 0 \), \( \sigma_{(v)}^{00} = \sigma_{(v)}^{0i} = 0 \), \( \Theta^{(v)} = 0 \). The only nonzero first order quantity is the shear,

\[ \tau_{ij} \equiv \sigma_{ij}^{(v)} = a \left( S_{(ij)} + F'_{(ij)} \right). \quad (330) \]

There is no tensor contribution to the expansion, acceleration and vorticity. There is only a nonzero contribution to shear,

\[ \sigma_{ij}^{(t)} = \frac{a}{2} h_{ij}' \quad (331) \]

As an aside, already at this purely geometrical level, we can obtain an evolution equation for the curvature perturbation \( \dot{\psi} \). Multiplying by \( (1 + \phi) \) the expression of the scalar expansion \( \Theta \) and writing it in terms of coordinate time \( t \), we find

\[ \ddot{\Theta} \equiv (1 + \phi) \theta = 3H - 3\dot{\psi} + \nabla^2 \tilde{\sigma}, \quad (332) \]

where \( \tilde{\sigma} = \dot{E} - B/a \). Writing the perturbation in the expansion as \( \delta \tilde{\Theta} \equiv \Theta - 3H \), we find

\[ \dot{\psi} = \frac{1}{3} \delta \tilde{\Theta} + \frac{1}{3} \nabla^2 \tilde{\sigma}, \quad (333) \]

which is independent of the field equations (follows simply from geometry). However, on large scales, \( \nabla^2 \tilde{\sigma} \to 0 \), the curvature perturbation satisfies \( \dot{\psi} = -\delta \tilde{\Theta}/3 \), a very interesting relation, which will allow us later to relate the curvature perturbation on a comoving hypersurface to the change in e-folds between different slicings.

### 8.5 Perturbations in Matter

We will consider here the stress-energy tensor for a matter fluid including anisotropic stresses. Like the metric, the coordinate representation of matter fields is gauge-dependent. The energy-momentum tensor of a fluid with density, isotropic pressure and 4-velocity \( u^\mu \) is

\[ T^\mu_\nu = (\rho + p)u^\mu u_\nu + p \delta^\mu_\nu + \Pi^\mu_\nu. \quad (334) \]

The anisotropic stress-tensor \( \Pi_{\mu\nu} \) decomposes into a trace-free part, \( \Pi \), a vector part, \( \Pi^i \), and a tensor part, \( \Pi_{ij} \),

\[ \Pi_{ij} = \Pi_{ij}^{\Pi} - \frac{1}{3} \nabla^2 \Pi \delta^i_j + \frac{1}{2} \left( \Pi_{ij}^{\|} + \Pi_{ij}^{\Pi^i} \right) + \Pi_{ij}^i. \quad (335) \]
The anisotropic stress-tensor has only spatial coordinates and is therefore gauge-invariant, since the background value is zero, \( \tilde{\Pi}_{ij} = \Pi_{ij} \).

The linearly perturbed velocity can be written as
\[
\begin{align*}
   u^\mu &= \frac{1}{a} \left( (1 - \phi), v^i + v^i \right), \\
   u_\mu &= a \left( - (1 + \phi), (v + B)_{ji} + v_i - S_i \right),
\end{align*}
\]
satisfying \( u_\mu u^\mu = -1 \). We have introduced a scalar velocity potential, \( v \), for irrotational flows, as well as a transverse vector field, \( v^i |_i = 0 \). We can then write the components of the energy-momentum tensor as
\[
\begin{align*}
   T_0^0 &= -(\rho_0 + \delta \rho), \\
   T_0^i &= (\rho_0 + p_0)(B_{ji} + v_{ji} + v_i - S_i), \\
   T_i^0 &= -(\rho_0 + p_0)(v^i + v^i), \\
   T_{ij} &= (p_0 + \delta p) \delta_{ij} + \Pi_{ij}.
\end{align*}
\]

Coordinate transformations affect the splitting between spatial and temporal components of matter fields and thus quantities like density, pressure and 3-velocity are gauge dependent. We have seen that the tensor perturbations are gauge independent, but let us see how scalar and vector matter fields change with a coordinate transformation \( \tilde{x}^\mu = x^\mu + \xi^\mu \). Any scalar function \( q \) which is homogeneous in the background FRW space-time can be written as \( q(\eta, x) = q_0(\eta) + \delta q(\eta, x) \). Thus the perturbed quantity transforms as \( \delta \tilde{q} = \delta q - \xi^0 q_0^\prime \). Physical scalars measured on spatial hypersurfaces like curvature, acceleration, shear or density perturbations, only depend on the choice of temporal gauge, \( \xi^0 \), but are independent of the coordinates within 3-dimensional hypersurfaces determined by \( \xi \). That is, the spatial gauge \( \xi \) can only affect the components of 3-vectors or 3-tensors on spatial hypersurfaces, but not 3-scalars. For example, the scalar shear transforms as
\[
\tilde{\sigma} = \sigma - \xi^0.
\]

Vector quantities that are derived from a potential, such as the velocity potential \( v \), only depend on the shift \( \xi \) within a hypersurface, and are independent of \( \xi^0 \). We thus find
\[
\tilde{v} = v + \xi^i.
\]

The vector function \( \xi^i \) only affects components of divergence-free (transverse) 3-vectors and 3-tensors in spatial hypersurfaces such as the velocity perturbation,
\[
\tilde{v}^i = v^i + \xi^i.
\]

### 8.6 Gauge invariant gravitational potentials and matter variables

The gauge dependence of metric perturbations had bothered people for a long time until Bardeen suggested the construction of gauge invariant combinations of perturbations. Out of 10 independent components of the metric perturbations we should be able to reduce them to 6 gauge independent degrees of freedom: 2 scalars, 2 (transverse) vectors and 2 (transverse traceless) tensors.

Let us construct first the scalars. The two scalar gauge functions, \( \xi^0 \) and \( \xi \), allow two of the four scalars to be eliminated, resulting 2 possible combinations
\[
\begin{align*}
   \Phi &\equiv \phi + \mathcal{H}(B - E') + (B - E')', \\
   \Psi &\equiv \psi - \mathcal{H}(B - E'),
\end{align*}
\]
which are related to those proposed by Bardeen (80), \( \Phi = \Phi_A Q^{(0)} \) and \( \Psi = -\Phi_H Q^{(0)} \).
These functions coincide with the metric perturbations in particular gauges, like the orthogonal zero-shear gauge, the conformal Newtonian gauge and the longitudinal gauge, but these are by no means preferred gauge choices, as we will see. In fact, any unambiguous choice of time slicing can be used to define gauge-invariant perturbations.

In general, for a gauge-dependent scalar quantity \( q \), whose perturbation transforms like \( \delta \bar{q} = \delta q - q' \xi_0 \), we can find a gauge-invariant combination as \( \delta q^{(g,1)} = \delta q + q' (B - E') \). For example, the scalar field perturbation with background value \( \varphi_0(\eta) \) has a gauge-independent expression

\[
\delta \varphi^{(g,1)} = \delta \varphi + \varphi_0'(\eta)(B - E') .
\]  

(342)

Let us now construct the gauge invariant vector combinations. There is one transverse vector gauge function \( \xi^i \), and there are two transverse vector metric perturbations, so we can construct a single gauge invariant transverse perturbation,

\[
\Sigma_i \equiv S_i + F'_i ,
\]  

(343)

which is directly related to the gauge invariant vector shear,

\[
\tau_{ij} = a \Sigma_{(ij)} .
\]  

(344)

Finally the tensors. For all tensor fields with vanishing or constant background contributions, such that \( \mathcal{L} Q_0 = 0 \), we can construct gauge invariant perturbation equations in terms of gauge-independent quantities. Since we can always write general relativity equations as \( Q = 0 \), it is always possible to construct these equations, which are then free from spurious gauge modes.

Before describing specific time-slicings, let us emphasize that the choice of Eq. (341) is not unique. Any combination of gauge invariant quantities will also be gauge invariant. For example, the following quantity will play a crucial role for testing the predictions of inflation,

\[
\zeta \equiv \psi - \mathcal{H}(v + B) ,
\]  

(345)

which coincides with the scalar curvature perturbation on comoving hypersurfaces, and can be written as a function of \( \Phi \) and \( \Psi \) as

\[
\zeta = \Psi + \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'(\Psi' + \mathcal{H}\Phi)} .
\]  

(346)

8.6.1 Gauge invariant matter variables

Let us now construct the gauge invariant matter variables. Again, the tensor matter fields are automatically gauge invariant, but the scalar and vector fields are not. If we write the density and pressure perturbations as

\[
\rho = \rho_0(1 + \delta) ,
\]  

(347)

\[
p = p_0(1 + \pi_L) ,
\]  

(348)

where \( \delta \) and \( \pi_L \) are the density and pressure contrasts, then we can readily construct a gauge-invariant quantity,

\[
\Gamma \equiv \pi_L - \frac{c_s^2}{w} \delta ,
\]  

(349)

which measures the intrinsic non-adiabaticity of the matter content, where

\[
w = \frac{p}{\rho} , \quad c_s^2 = \frac{p'}{\rho'} ,
\]  

(350)
are the barotropic index and the sound speed of the fluid. Actually, $\Gamma$ is related to the entropy production rate. If the pressure of the fluid is a function of local energy density only, $p = p(\rho)$, then
\[ \frac{\delta p}{\delta \rho} = \frac{p'}{\rho'} = c_s^2 \quad \Rightarrow \quad \Gamma = 0. \]
For instance, in the case of a single barotropic fluid, with $p = w\rho$, we have $\Gamma = 0$. Values different from zero may arise from the relative entropy (particle number) of a mixture of several fluid components.

Another scalar quantity, the density contrast, does not have a unique gauge-invariant expression because it requires the use of the metric variables, and that depends on the specific choice of slicing. For example, three different gauge invariant density contrasts, related to three different gauge choices, are
\begin{align*}
D_l &\equiv \delta - 3(1 + w)\mathcal{H}(B - E') \quad \text{longitudinal gauge} \tag{351} \\
D_f &\equiv \delta - 3(1 + w)\psi \quad \text{flat slicing gauge} \tag{352} \\
D_c &\equiv \delta - 3(1 + w)\mathcal{H}(v + B) \quad \text{comoving gauge} \tag{353}
\end{align*}
The distinction between them is only important on superhorizon scales, since on small (subhorizon) scales they reduce to the same variable, the Newtonian density contrast.

The velocity perturbations can be written also in gauge-invariant form
\begin{align*}
V &\equiv v + E', \tag{354} \\
V_i &\equiv v_i + F_i'. \tag{355}
\end{align*}
Some useful relations between gauge invariant quantities are
\begin{align*}
D_l &= D_f + 3(1 + w)\mathcal{H}(B - E') \tag{356} \\
D_c &= D_l - 3(1 + w)\mathcal{H}V \tag{357} \\
D_c &= D_f + 3(1 + w)\zeta \tag{358}
\end{align*}
which are further related by
\[ \zeta = \Psi - \mathcal{H}V \tag{359} \]

8.7 Different time slices and their gauge choices

8.7.1 The longitudinal gauge
It is also known as the conformal Newtonian gauge or the zero-shear gauge. It is defined by the condition $\tilde{\sigma} = 0$ in Eq. (338). If we start from arbitrary coordinates and perform a gauge transformation with
\[ \tilde{\xi}_l^0 = E' - B, \tag{360} \]
then we indeed find $\tilde{\sigma}_l = 0$. This is sufficient to find the two scalar metric perturbations, or any other scalar quantity on these hypersurfaces. However, in addition, we must supply the vector gauge transformation. The longitudinal gauge is completely determined by the spatial gauge choice
\[ \xi_l = E, \quad \xi_l^i = -S_i, \tag{361} \]
and thus $\tilde{E}_l = \tilde{B}_l = 0$, as well as $\tilde{S}_l^i = 0$. The remaining scalar metric perturbations and density perturbations become
\begin{align*}
\tilde{\phi}_l &= \phi + \mathcal{H}(B - E') + (B - E')', \tag{362} \\
\tilde{\psi}_l &= \psi - \mathcal{H}(B - E'), \tag{363} \\
\delta \tilde{\rho}_l &= \delta \rho + \rho_0'(B - E'), \tag{364}
\end{align*}
where the two scalar metric perturbations coincide with the Bardeen potentials (341). In this gauge the line element becomes
\[
ds^2 = a(\eta)^2 \left[ -(1 + 2\Phi) d\eta^2 + (1 - 2\Psi) \gamma_{ij} dx^i dx^j \right]
\] (365)
which is a cosmological (conformal) generalization of the Schwarzschild metric, where \( \Phi \) plays the role of the first-order Newtonian potential \( \phi \), and \( \Psi \) gives the Newtonian curvature perturbation. This is the reason why it is sometimes called the conformal Newtonian gauge.

8.7.2 The flat-slicing gauge
Also known as the uniform curvature gauge or the off-diagonal gauge. It is defined purely by local metric quantities. In this gauge one selects spatial hypersurfaces on which the induced 3-metric is left unperturbed: \( \tilde{\psi} = \tilde{E} = 0 \). This corresponds to a gauge transformation
\[
\xi_f^0 = -\frac{\psi}{\mathcal{H}}, \quad \xi_f = E.
\] (366)
The definitions of the remaining gauge invariant metric perturbations are
\[
\tilde{\phi}_f = \phi + \psi + \left(\frac{\psi}{\mathcal{H}}\right)',
\tilde{B}_f = B - E' - \frac{\psi}{\mathcal{H}},
\delta \tilde{\rho}_f = \delta \rho + \rho_0' \frac{\psi}{\mathcal{H}}.
\] (367) (368) (369)
The first two quantities coincide with the two potentials, \( \tilde{A} \) and \( \tilde{B} \), defined by Kodama and Sasaki (84). Note also that in this gauge, \( \tilde{\sigma}_f = -\tilde{B}_f \), while the difference between different choices of slicings gives gauge invariant metric perturbations, \( \xi_f^0 - \xi_l^0 = \tilde{B}_f = -\tilde{\psi}_l/\mathcal{H} \), and thus
\[
\tilde{\sigma}_f = \frac{\tilde{\psi}_l}{\mathcal{H}},
\]
i.e. the shear perturbation in the uniform-curvature gauge is equal to the curvature perturbation in the zero-shear (longitudinal) gauge over the rate of expansion.

In this gauge, the line element can be written as
\[
ds^2 = a(\eta)^2 \left[ -(1 + 2\tilde{\phi}_f) d\eta^2 + 2\tilde{B}_f dx^i dx^j + \gamma_{ij} dx^i dx^j \right]
\] (370)
In some cases, it is more convenient to use gauge invariant quantities like \( \tilde{\phi}_f \) and \( \tilde{B}_f \), rather than \( \Phi \) and \( \Psi \) when the background space-time behaves singularly (e.g. in a collapsing universe like in pre-Big-Bang cosmology), since they remain small.

Notice also that the scalar field perturbations in uniform-curvature (flat-slicing) hypersurfaces is the gauge invariant scalar field perturbation defined by Mukhanov (85),
\[
\delta \varphi_f = \delta \varphi + \varphi_0' \frac{\psi}{\mathcal{H}} = \frac{u}{a}.
\] (371)

8.7.3 The synchronous gauge
It is defined by the condition that the time vector is orthogonal to constant-time 3-space hypersurfaces: \( \tilde{\phi} = \tilde{B} = 0 \) and \( \tilde{B}^i = 0 \). The problem is that this gauge choice does not determine the time-slicing
unambiguously: there is a residual gauge freedom and it is not possible to define gauge-invariant quantities in general using this gauge condition. One could always, as Lifshitz did in 1963, eliminate the unphysical degrees of freedom by symmetry arguments.

The metric in this gauge becomes

\[ ds^2 = a(\eta)^2 \left[ -d\eta^2 + (\gamma_{ij} + h_{ij})dx^idx^j \right] , \]  

(372)

with the perturbation

\[ h_{ij}(\eta, x) \equiv h(\eta, x)_{ij} + \left( \nabla_i \nabla_j - \frac{1}{3} \gamma_{ij} \nabla^2 \right) 6\tau(\eta, x) , \]  

(373)

written in terms of two arbitrary functions \( h(\eta, x) \) and \( \tau(\eta, x) \). The gauge transformations that take an arbitrary set of coordinates to the synchronous gauge are

\[ \xi^0 = -\frac{1}{a} \int d\eta a\phi + \frac{\alpha(x)}{a} , \]  

(374)

\[ \xi = \int d\eta (\xi^0 - B) + \beta(x) , \]  

(375)

\[ \xi^i = -\int d\eta B^i + \beta^i(x) , \]  

(376)

with \( \beta^i |_i = 0 \). It has a residual gauge freedom: 4 integration constants \( (\alpha, \beta, \beta^i) \) that correspond to different choices of spatial hypersurfaces, that give rise to fictitious “gauge modes” in the perturbation equations, and must be removed by physical considerations.

8.7.4 The comoving orthogonal gauge

This gauge is defined by choosing spatial coordinates such that the 3-velocity of the cosmic fluid vanishes, \( \tilde{v} = 0 \). Orthogonality of the constant-\( -\eta \) hypersurfaces to the 4-velocity \( u^\mu \) then requires \( \tilde{v} + \tilde{B} = 0 \), which implies that these hypersurfaces have zero vorticity.

The gauge transformations needed to go to this gauge are

\[ \xi_0^c = -(v + B) , \]  

(377)

\[ \xi_c = -\int d\eta v + \xi(x) , \]  

(378)

where \( \xi \) is a residual gauge freedom corresponding to a constant shift of spatial coordinates. All 3-scalars like curvature, expansion, acceleration and shear, are independent of \( \xi \). The scalar perturbations in the comoving orthogonal gauge are then

\[ \tilde{\phi}_c = \phi + \mathcal{H}(v + B) + (v + B)' , \]  

(379)

\[ \tilde{\psi}_c = \psi - \mathcal{H}(v + B) , \]  

(380)

\[ \tilde{E}_c = E + \int d\eta v - \hat{\xi} . \]  

(381)

Apart from the residual gauge dependence of \( \tilde{E}_c \), these variables are gauge invariant under transformations of their component parts.

Note that the curvature perturbation in the comoving gauge, \( \tilde{\psi}_c \) has the same expression as the variable \( \zeta \) (345), used for the first time by Lukash (80) and later called \( \mathcal{R} \) by Lyth (85). The density perturbation on comoving orthogonal hypersurfaces is given in gauge invariant form by

\[ \delta\tilde{\rho}_c = \delta\rho + \rho_0'(v + B) , \]  

(382)
and is equivalent to the expression $\epsilon_m \rho_0 Q^{(0)}$ of Bardeen (80). The gauge-invariant scalar density perturbation of Ellis and Bruni (89) is $\Delta = \frac{1}{m_0} \delta \rho^i c^i$. The velocity perturbation in this gauge is $\tilde{v}_c$ is equivalent to the gauge-invariant velocity perturbation $V$.

One can always write these quantities in terms of metric perturbations, rather than the velocity potential, using the Einstein equations, to give

$$v + \tilde{B} = \frac{\Psi' + \mathcal{H}(B - E')}{\mathcal{H}' - \mathcal{H}^2 - K}.$$  \hspace{1cm} (383)

In particular, one can write the comoving curvature perturbation $\tilde{\psi}_c$ in terms of longitudinal gauge invariant quantities,

$$\tilde{\psi}_c = \Psi - \frac{\mathcal{H} (\Psi' + \mathcal{H} \Phi)}{\mathcal{H}' - \mathcal{H}^2 - K},$$  \hspace{1cm} (384)

which coincides (for $K = 0$) with the variable $\zeta$ in Eq. (346), defined for the first time by Bardeen, Steinhardt and Turner (83), and later by Mukhanov, Feldman and Brandenberger (92).

8.7.5 The comoving total matter gauge

This gauge extends the previous one to the case of a multi-component fluid. Here the total momentum vanishes,

$$(\rho + p)(\tilde{v} + \tilde{B}) \equiv \sum_\alpha (\rho_\alpha + p_\alpha)(\tilde{v}_\alpha + \tilde{B}_\alpha) = 0.$$  \hspace{1cm} (385)

Orthogonality of constant-$\eta$ hypersurfaces to the 4-velocity again requires $\tilde{B} = 0$. Note that the gauge-invariant scalar velocity perturbation, $V$ (the velocity in the longitudinal gauge), coincides in the total matter gauge with the shear $\tilde{\sigma}_c$ of constant-$\eta$ hypersurfaces.

8.7.6 The uniform density gauge

This gauge is defined by imposing that the matter density remains unperturbed, $\delta \tilde{\rho} = 0$, in these hypersurfaces. The coordinate transformation that takes us there is

$$\xi_0^d = \frac{\tilde{\rho}}{\rho_0^d}. $$  \hspace{1cm} (386)

On these uniform-density hypersurfaces, the gauge-invariant curvature perturbation is

$$\zeta \equiv \tilde{\psi}_d = \psi + \mathcal{H} \frac{\delta \tilde{\rho}}{\rho_0^d}. $$  \hspace{1cm} (387)

There is another degree of freedom left inside spatial hypersurfaces and we can choose either $\tilde{B}$, $\tilde{E}$ or $\tilde{v}$ to vanish. If we choose $\tilde{B} = \tilde{v} = 0$, then the gauge function is

$$\xi_d = - \int d\eta \tilde{v}. $$  \hspace{1cm} (388)

Note that the gauge-invariant curvature perturbation in the uniform density gauge is equal to the gauge-invariant curvature perturbation in the comoving gauge.

8.8 Perturbed Einstein field equations

We will now write down the Einstein field equations, $G_{\mu\nu} = \kappa^2 T_{\mu\nu}$ and separate between the background Einstein equations, $G^{(0)}_{\mu\nu} = \kappa^2 T^{(0)}_{\mu\nu}$, and the first order perturbed equations, $\delta G_{\mu\nu} = \kappa^2 \delta T_{\mu\nu}$. These equations can be written in specific gauges as well as in a gauge invariant form.
Einstein equations can be separated into the Hamiltonian and momentum contraint equations, (220) and (221), together with the evolution equation, which itself can be separated into the trace (222) and traceless (223) parts. The background equations give the Friedmann equations,

\[ \mathcal{H}^2 + K = \frac{\kappa^2}{3} a^2 \rho, \quad \text{Hamiltonian constraint} \quad (389) \]
\[ 0 = 0, \quad \text{Momentum constraint} \quad (390) \]
\[ \mathcal{H}' = -\frac{\kappa^2}{6} a^2 (\rho + 3p), \quad \text{Evolution equation (trace)} \quad (391) \]
\[ 0 = 0, \quad \text{Evolution equation (traceless)} \quad (392) \]

All the background equations are scalar equations since the background fields are also scalars. However, the perturbed equations can be separated into scalar, vector and tensor perturbation equations. We will write first the gauge-dependent equations, and then their gauge-invariant forms, in terms of the gauge-invariant potentials and matter fields.

The scalar perturbation equations can be separated into the energy and momentum constraint equations, plus the trace and traceless evolution equations,

\[ 3\mathcal{H}(\psi' + \mathcal{H}\phi) - (\nabla^2 + 3K)\psi - \mathcal{H}\nabla^2 \sigma = -\frac{\kappa^2}{2} a^2 \delta \rho, \quad \text{Hamiltonian constraint} \quad (393) \]
\[ \psi' + \mathcal{H}\phi + K\sigma = -\frac{\kappa^2}{2} a^2 (\rho + p)(v + B), \quad \text{Momentum constraint} \quad (394) \]
\[ \psi'' + 2\mathcal{H}\psi' - K\psi + \mathcal{H}\phi' + (2\mathcal{H}' + \mathcal{H}^2)\phi = \frac{\kappa^2}{2} a^2 \left( \delta p + \frac{2}{3} \nabla^2 \Pi \right), \quad \text{Trace equation} \quad (395) \]
\[ \sigma' + 2\mathcal{H}\sigma + \psi' - \phi = \kappa^2 a^2 \Pi, \quad \text{Traceless equation} \quad (396) \]

where \( \sigma = E' - B \) is the scalar shear. The gauge-invariant scalar equations for the 4 gauge invariant scalar quantities \( (\Phi, \Psi, V, D \equiv D_c) \) is

\[ (\nabla^2 + 3K)\Psi = \frac{\kappa^2}{2} a^2 \rho_0 D, \quad \text{Poisson equation} \quad (397) \]
\[ \Psi' + \mathcal{H}\Phi = -\frac{\kappa^2}{2} a^2 (\rho_0 + p_0)V, \quad \text{Constraint equation} \quad (398) \]
\[ \mathcal{H}U' + (2\mathcal{H}' + \mathcal{H}^2) U = \frac{\kappa^2}{2} a^2 \left( \rho_0 (c_s^2 D_f + w \Gamma) + \frac{2}{3} \nabla^2 \Pi \right), \quad \text{Evolution equation} \quad (399) \]
\[ \Psi - \Phi = \kappa^2 a^2 \Pi, \quad \text{Anisotropic stress} \quad (400) \]

where \( \delta p = \rho_0 (c_s^2 \delta + w \Gamma) \), and we have defined a new variable \( U \equiv \epsilon \zeta \), with \( \epsilon = (1 - \mathcal{H}'/\mathcal{H}^2) \), and \( D_f = D - 3(1+w)\zeta \). In order to solve these coupled equations we need to specify the matter content, i.e. \( w, c_s^2, \Gamma \) and \( \Pi \). For instance, for a perfect fluid without anisotropic stresses, \( \Gamma = \Pi = 0 \) automatically implies that the two Newtonian potentials are identical, \( \Psi = \Phi \).

The vector perturbation equations can be separated into a momentum constraint equation and an evolution equation

\[ (\nabla^2 + 2K)(S_i + F_i) = 2\kappa^2 a^2 (\rho_0 + p_0)(S_i - v_i), \quad \text{Constraint equation} \quad (401) \]
\[ \tau_{ij}' + \mathcal{H}\tau_{ij} = \frac{\kappa^2}{2} a^3 \Pi_{(ij)}, \quad \text{Evolution equation} \quad (402) \]
written in terms of the vector shear (330).

The *gauge-invariant vector* perturbation equations can be written in terms of the gauge-invariant vector fields, \( \Sigma_i \) and \( V_i \), (343) and (354),

\[
\left( \nabla^2 + 2K - 4\epsilon \mathcal{H}^2 \right) \Sigma_i = -2\kappa^2 a^2 (\rho_0 + p_0) V_i, \quad \text{Constraint equation} \quad (403)
\]
\[
\Sigma_i' + 2\mathcal{H}\Sigma_i = \kappa^2 a^2 \Pi_i, \quad \text{Evolution equation} \quad (404)
\]

For a perfect fluid with \( \Pi_i = 0 \), these equations, on large scales and for a flat universe, give

\[
\Sigma_i = V_i \propto \frac{1}{a^2} \rightarrow 0, \quad (405)
\]

which is equivalent to \( \delta u_i^{(g.i.)} \propto 1/a^3 \), and therefore, in the case of a matter fluid dominated e.g. by a scalar field, the vector perturbations are quickly driven to zero, even if initially there were anisotropic sources, as long as in the late time evolution there are no sources of vorticity (like due to magnetic fields or gravitational waves).

The *tensor* perturbation equations are already written in gauge-invariant form and arise from the \((ij)\) components of the Einstein equations,

\[
h''_{ij} + 2\mathcal{H}h_{ij} + (2K - \nabla^2)h_{ij} = 2\kappa^2 a^2 \Pi_{ij}. \quad (406)
\]

Since there is no tensor conservation equation for the matter variables, the tensor metric perturbation is solely determined by the Hubble parameter and the scale factor, and is only sourced by the matter anisotropic tensor perturbation. For scalar and vector matter fields, \( \Pi_{ij} = 0 \) for linear perturbations, and thus the evolution equation is source-free and homogeneous.

On superhorizon scales and zero curvature, the homogeneous equation in the radiation era \( (\mathcal{H} = 1/\eta) \) has a decaying \( (h_{ij} \propto 1/\eta) \) and a growing \( (h_{ij} \propto \text{const.}) \) solution. As the mode enters the horizon, the oscillatory behaviour takes over, and the gravitational wave propagates with frequency \( k^2 + 2K \), and amplitude damped as \( h \propto 1/a \).

In the absence of anisotropic stresses \( (\Pi_{ij} = 0) \) in a flat universe \( (K = 0) \), the general solution for \( h_{ij} = h(\eta, x) \epsilon_{ij} \), with \( \epsilon_{ij} \) the polarization tensor associated with the wave, in Fourier space is

\[
h(\eta, k) = \frac{1}{(k\eta)^{p-1}} \left[ A j_{p-1}(k\eta) + B n_{p-1}(k\eta) \right], \quad (407)
\]

with \( j_\nu \) and \( n_\nu \) the spherical Bessel and Von Neumann functions of order \( \nu \). The evolution of the scale factor is given by \( a \propto \eta^p \).

### 8.9 Conservation of energy-momentum tensor

The covariant conservation of the energy-momentum tensor, \( T^{\mu\nu}_{;\nu} = 0 \), gives the homogeneous conservation of energy in the isentropic expansion, \( T^{0\nu}_{;\nu} = 0 \),

\[
\rho_0' = -3\mathcal{H}(\rho_0 + p_0), \quad \text{or} \quad d(\rho_0 a^3) + p_0 \, d(a^2) = 0. \quad (408)
\]

There is no background momentum conservation equation, \( T^{\mu\nu}_{;\nu} = 0 \), since the momentum is zero by assumption of homogeneity.

The perturbed conservation equations follow from the Bianchi identity, \( \delta T^{\mu\nu}_{;\nu} = 0 \), and provide with evolution equations for the physical degrees of freedom that are often simpler and easier to deal with than the perturbed Einstein equations, since they are first order, rather than second order, as the latter usually are.
The first order scalar perturbation equation for the energy density perturbation is

\[ \delta \rho' + 3 \mathcal{H} (\delta \rho + \delta p) = (\rho_0 + p_0) [3 \psi' - \nabla^2 (v + E')] , \tag{409} \]

\[ [(\rho_0 + p_0)(v + B)]' + \delta p + \frac{2}{3} (\nabla^2 + 3 K) \Pi = -(\rho_0 + p_0) [\phi + 4 \mathcal{H}(v + B)] , \tag{410} \]

together with the evolution equation for the scalar momentum perturbation. These equations can be written in gauge-invariant form as two equations, one for the density contrast \( D_f \) (in the flat-slicing gauge) and the other for the velocity perturbation \( V \), using \( \delta p = \rho_0 (w \Gamma + c_s^2 \delta) \) and \( w' = -3 \mathcal{H} (1 + w) (c_s^2 - w) \),

\[ D'_f + 3 \mathcal{H} (c_s^2 - w) D_f = -3 \mathcal{H} \ w \Gamma - (1 + w) \nabla^2 V , \tag{411} \]

\[ V' + \mathcal{H} (1 - 3 c_s^2) V = -\Phi - 3 c_s^2 \Psi - \frac{1}{1 + w} \left[ w \Gamma + c_s^2 D_f + \frac{2 w}{3} (\nabla^2 + 3 K) \Pi \right] , \tag{412} \]

where we have redefined \( \Pi = p_0 \tilde{\Pi} = w \rho_0 \Pi \). Sometimes it is more convenient to write them in terms of the density contrast in the comoving gauge, \( D \equiv D_c \),

\[ D' - 3 w \mathcal{H} D = (\nabla^2 + 3 K) \left[ 2 \mathcal{H} \Pi - (1 + w) V \right] , \tag{413} \]

\[ V' + \mathcal{H} V = -\Phi - \frac{1}{1 + w} \left[ w \Gamma + c_s^2 D + \frac{2 w}{3} (\nabla^2 + 3 K) \Pi \right] . \tag{414} \]

There is no energy conservation equation in the case of vector perturbations since energy is a scalar in the spatial hypersurfaces, but we do have a momentum conservation equation for vector perturbations since momentum is a 3-vector,

\[ [(\rho_0 + p_0)(v_i - S_i)]' + 4 \mathcal{H} (\rho_0 + p_0)(v_i - S_i) = -(\nabla^2 + 2 K) \Pi_i . \tag{415} \]

This equation can be put in gauge-invariant form by defining a new variable, the vorticity vector \( \Omega_i \equiv \Sigma_i - V_i \),

\[ \Omega'_i + \mathcal{H} (1 - 3 c_s^2 \Omega_i = \frac{w}{2(1 + w)} \nabla^2 \Pi_i . \tag{416} \]

If the anisotropic stress vanishes, \( \Pi_i = 0 \), then there is a solution

\[ \Omega_i \propto a^{3 c_s^2 - 1} \quad \left\{ \begin{array}{l} \Omega_i \propto \text{const.} \quad \text{for radiation era} \\ \Omega_i \propto a^{-1} \quad \text{for matter era} \end{array} \right. \tag{417} \]

There is no energy or momentum conservation equation for tensor matter variables and thus they are decoupled from the other tensor quantities.

### 8.9.1 Three different gauge choices

To illustrate the gauge invariant approach, we will take the gauge-dependent equations and write them in terms of gauge-invariant quantities, which coincide with physical degrees of freedom in particular time-slicings.

The evolution equations in the longitudinal gauge arise when imposing \( \tilde{B} = \tilde{E} = 0 \), and thus \( \tilde{\sigma} = 0 \), as well as \( \tilde{S}^i = 0 \). The gauge-invariant equations of motion in this gauge are the following.

The conservation of the energy-momentum tensor gives the continuity equation and the constraint equation,

\[ \delta \tilde{\rho}' + 3 \mathcal{H} (\delta \tilde{\rho} + \delta \tilde{\rho}') = (\rho_0 + p_0) [3 \tilde{\psi}' - \nabla^2 \tilde{v}'] , \tag{418} \]

\[ [(\rho_0 + p_0) \tilde{v}_i]' + \delta \tilde{p}_i + \frac{2}{3} (\nabla^2 + 3 K) \tilde{\Pi}_i = -(\rho_0 + p_0) [\tilde{\phi}' + 4 \mathcal{H} \tilde{v}_i] . \tag{419} \]
The scalar metric perturbation evolution equations (trace and traceless parts) are

\[ \ddot{\psi}_t + 2\dot{H}\dot{\psi}_t - K\dot{\psi}_t + \mathcal{H}\ddot{\phi}_t + (2\dot{\mathcal{H}} + \mathcal{H}^2)\dot{\phi}_t = \frac{\kappa^2}{2}a^2\left(\delta\tilde{p}_t + \frac{2}{3}\nabla^2\tilde{\Pi}_t\right), \tag{420} \]

\[ \ddot{\psi}_t - \ddot{\phi}_t = \kappa^2a^2\tilde{\Pi}_t. \tag{421} \]

The energy and momentum constraint equations are

\[ 3\mathcal{H}(\ddot{\psi}_t + \ddot{\phi}_t) - (\nabla^2 + 3K)\ddot{\phi}_t = -\frac{\kappa^2}{2}a^2\delta\tilde{p}_t, \tag{422} \]

\[ \ddot{\psi}_t + \ddot{\phi}_t = -\frac{\kappa^2}{2}a^2(\rho_0 + p_0)\tilde{v}_t. \tag{423} \]

The case \(\tilde{\Pi}_t = 0\) for a perfect fluid, or a scalar field, gives \(\ddot{\psi}_t = \ddot{\phi}_t\), which simplifies the equations.

The evolution equations in the uniform density gauge arise when imposing \(\delta\tilde{p} = 0\), and also fixing spatial hypersurfaces with \(\dot{B} = 0\), so that \(\tilde{\sigma}_d = \tilde{E}_d\). The gauge-invariant equations of motion in this gauge are the following.

The conservation of the energy-momentum tensor gives the continuity equation and the constraint equation,

\[ \frac{3\mathcal{H}}{\rho_0 + p_0}\delta\tilde{p}_d = 3\tilde{v}_d - \nabla^2(\tilde{v}_d + \tilde{\sigma}_d), \tag{424} \]

\[ [(\rho_0 + p_0)\tilde{v}_d]' + \delta\tilde{p}_d + \frac{2}{3}(\nabla^2 + 3K)\tilde{\Pi}_d = -(\rho_0 + p_0)\tilde{v}_d' + 4\mathcal{H}\tilde{v}_d]. \tag{425} \]

The first equation is crucial for understanding the behaviour of perturbations on large scales, when the second term in the R.H.S. vanishes, and thus \(\ddot{\psi}_d = \text{constant}\) on large scales for vanishing non-adiabatic pressure perturbations.

The Einstein evolution equations (trace and traceless parts) give

\[ \ddot{\psi}_d'' + 2\dot{H}\dot{\psi}_d - K\dot{\psi}_d + \mathcal{H}\ddot{\phi}_d + (2\dot{\mathcal{H}} + \mathcal{H}^2)\dot{\phi}_d = \frac{\kappa^2}{2}a^2\left(\delta\tilde{p}_d + \frac{2}{3}\nabla^2\tilde{\Pi}_d\right), \tag{426} \]

\[ \ddot{\sigma}_d'' + 2\mathcal{H}\ddot{\sigma}_d + \ddot{\psi}_d - \ddot{\phi}_d = \kappa^2a^2\tilde{\Pi}_d. \tag{427} \]

The energy and momentum constraint equations are

\[ 3\mathcal{H}(\ddot{\psi}_d + \ddot{\phi}_d) - (\nabla^2 + 3K)\ddot{\phi}_d - \mathcal{H}\nabla^2\ddot{\phi}_d = 0, \tag{428} \]

\[ \ddot{\psi}_d + \ddot{\phi}_d + K\ddot{\phi}_d = -\frac{\kappa^2}{2}a^2(\rho_0 + p_0)\tilde{v}_d. \tag{429} \]

The case \(\tilde{\Pi}_d = 0\) for a perfect fluid or a scalar field does not imply in this case \(\ddot{\psi}_d = \ddot{\phi}_d\).

The evolution equations in the comoving gauge are defined by a vanishing 3-velocity \(\dot{v} = 0\), and also fixing spatial hypersurfaces with \(\dot{B} = 0\). The gauge-invariant equations of motion in this gauge become as follows.

The conservation of the energy-momentum tensor gives the continuity and the constraint equation,

\[ \delta\tilde{p}_c + 3\mathcal{H}(\delta\tilde{p}_c + \delta\tilde{p}_e) = (\rho_0 + p_0)[3\tilde{v}_c' - \nabla^2\tilde{E}_c'], \tag{430} \]

\[ \delta\tilde{p}_c + \frac{2}{3}(\nabla^2 + 3K)\tilde{\Pi}_c = -(\rho_0 + p_0)\ddot{\phi}_c. \tag{431} \]
The evolution equations (trace and traceless parts) give
\[ \ddot{\tilde{\psi}}^c + 2H\dot{\tilde{\psi}}^c - K\tilde{\psi}^c + \mathcal{H}\tilde{\phi}^c + (2\mathcal{H}' + \mathcal{H}^2)\tilde{\phi}^c = \frac{\kappa^2}{2} a^2 \left( \delta\tilde{p}^c + \frac{2}{3} \nabla^2 \tilde{\Pi}^c \right), \quad (432) \]
\[ \ddot{\tilde{\sigma}}_c + 2H\tilde{\sigma}_c - \tilde{\psi}_c - \tilde{\phi}_c = \kappa^2 a^2 \tilde{\Pi}_c. \quad (433) \]
The energy and momentum constraint equations are
\[ 3\mathcal{H}(\ddot{\tilde{\psi}}^c + \mathcal{H}'\tilde{\psi}^c) - (\nabla^2 + 3K)\tilde{\phi}_c - \mathcal{H}\nabla^2 \tilde{\sigma}_c = -\frac{\kappa^2}{2} a^2 \delta\tilde{p}_c, \quad (434) \]
\[ \ddot{\tilde{\sigma}}^d + \mathcal{H}\tilde{\phi}^d + K\tilde{\sigma}^d = 0. \quad (435) \]

8.9.2 The Bardeen equation
It is often convenient to have an evolution equation for the Bardeen potential written in terms of the total matter content. By combining the matter conservation equation (413) with the Poisson (397) and constraint (398) equations we obtain a second order equation for \( \Psi \),
\[ \Psi'' + 3\mathcal{H}(1 + c_s^2)\Psi' + [3(c_s^2 - w)\mathcal{H}^2 - (1 + 3c_s^2)K - c_s^2 \nabla^2]\Psi = \kappa^2 a^2 \left[ \mathcal{H}\Pi' + \left( 2\mathcal{H}' + \frac{3\mathcal{H}^2}{w}(w - c_s^2) \right)\Pi + \frac{1}{2} \nabla^2 \Pi + \frac{1}{2} \rho_0 \Gamma \right]. \quad (436) \]
This equation can be recast (using the variable \( U = \epsilon \zeta \)) into a first order evolution equation for the gauge invariant curvature perturbation \( \zeta = \Psi - \mathcal{H}V \). For hydrodynamical matter with \( \Pi = 0 \), and flat spatial sections \( (K = 0) \), we find
\[ \zeta' = \frac{1}{\epsilon \mathcal{H}} \left[ c_s^2 \nabla^2 \Psi + \frac{3}{2} \mathcal{H}^2 w \Gamma \right], \quad (437) \]
which is extremely useful when discussing the evolution of curvature and entropy perturbations in the case of multiple fluids.

8.10 Multiple fluids
Previous expressions assume the universe is filled with, or dominated by, a single fluid with energy density \( \rho \) and pressure \( p \). We know the universe is filled during most of its evolution by more than one fluid. In particular, during the epoch of photon decoupling, there are four different components, two relativistic (photons and neutrinos) and two non-relativistic (baryons and cold dark matter). Therefore, we should take into account these different components with different equations of state, and in some cases with exchange of energy between them, as occurs between baryons and photons in the tight-coupling regime before recombination.

If there is more than one matter component, then the total perturbation variables are actually weighted averages of the individual perturbation variables,
\[ \delta \equiv \sum_{\alpha} \frac{\rho_{\alpha}}{\rho} \delta_{\alpha}, \quad \Theta^i \equiv \sum_{\alpha} \left( \frac{\rho_{\alpha} + p_{\alpha}}{\rho + p} \right) \Theta^i_{\alpha}, \quad \Pi^{ij} \equiv \sum_{\alpha} \Pi_{\alpha}^{ij}. \quad (439) \]
The equation of state and adiabatic sound speed are defined for each component,
\[ w_{\alpha} \equiv \frac{p_{\alpha}}{\rho_{\alpha}}, \quad c_s^2_{\alpha} \equiv \frac{p_{\alpha}'}{\rho_{\alpha}}, \quad \rho_{\alpha}' + 3\mathcal{H}\rho_{\alpha}(1 + w_{\alpha}) = 0, \quad (440) \]
where we have assumed for the moment that there is no exchange of energy between different components and thus the total energy on each component is conserved in the expansion. For the overall mixture, we have
\[ w \equiv \frac{p}{\rho}, \quad c_s^2 \equiv \frac{p'}{\rho}, \quad \rho' + 3\mathcal{H}\rho(1 + w) = 0. \quad (441) \]
The transformation properties under general coordinate reparametrizations are the same as for total matter variables, e.g. \( \delta_\alpha = \delta_\alpha - 3H(1 + w_\alpha)\xi^0 \). For each matter component we can define gauge invariant quantities,

\[
\begin{align*}
\Gamma_\alpha & \equiv \pi_{L,\alpha} - \frac{c_{\alpha}^2}{w_\alpha} \delta_\alpha, \\
V_\alpha & \equiv v_\alpha + E', \\
D_{t,\alpha} & \equiv \delta_\alpha - 3(1 + w_\alpha)\mathcal{H}(B - E'), \quad \text{longitudinal gauge} \\
D_{f,\alpha} & \equiv \delta_\alpha - 3(1 + w_\alpha)\psi, \quad \text{flat slicing gauge} \\
D_{c,\alpha} & \equiv \delta_\alpha - 3(1 + w_\alpha)\mathcal{H}(v_\alpha + B), \quad \text{comoving gauge}
\end{align*}
\]

For multiple matter components, it is often useful to work with gauge-invariant density contrasts,

\[
\Delta_\alpha \equiv \delta_\alpha - 3(1 + w_\alpha)\mathcal{H}(B + v),
\]

which corresponds to the density contrast in the gauge where the total matter is at rest (the total comoving gauge). It is also related to the density contrast in the flat slicing gauge,

\[
\Delta_\alpha = D_{f,\alpha} + 3(1 + w_\alpha)\zeta,
\]

a very useful expression since it includes the gauge-invariant curvature perturbation \( \zeta \), which remains constant on superhorizon scales, for adiabatic perturbations.

### 8.10.1 Entropy perturbations

When more than one component is present, entropy perturbations can arise even for a mixture of perfect uncoupled fluids. Clearly a multicomponent fluid can exchange not only energy, but also particle number (entropy). The non-adiabaticity of the mixture is given by \( \Gamma \), see (349). Using definitions of matter variables, this can be expressed as

\[
p\Gamma = p\Gamma_{\text{int}} + \sum_\alpha \delta_\alpha \rho_\alpha (c_{\alpha}^2 - c_s^2) = p(\Gamma_{\text{int}} + \Gamma_{\text{rel}}),
\]

where the total intrinsic entropy is

\[
\Gamma_{\text{int}} = \sum_\alpha \frac{p_\alpha}{p} \Gamma_\alpha,
\]

while the relative entropy is given by

\[
p\Gamma_{\text{rel}} = \frac{1}{2} \sum_{\alpha,\beta} \frac{(1 + w_\alpha)(1 + w_\beta)\rho_\alpha \rho_\beta}{(1 + w)\rho} (c_{\alpha}^2 - c_{\beta}^2) \left( \frac{\delta_\alpha}{1 + w_\alpha} - \frac{\delta_\beta}{1 + w_\beta} \right).
\]

Note that we have assumed that the components have decoupled from each other, i.e. \( Q_\alpha^\nu = 0 \). We will describe later the case for interacting fluids.

The entropy perturbation between \( \alpha \) and \( \beta \) components is defined as

\[
S_{\alpha\beta} \equiv \frac{\delta_\alpha}{1 + w_\alpha} - \frac{\delta_\beta}{1 + w_\beta} = \delta \ln \frac{n_\alpha}{n_\beta},
\]

That is, the fluctuation in the local number density of different components. In fact, the entropy perturbation is itself a gauge-invariant quantity that can be expressed in terms of gauge invariant density contrasts as

\[
S_{\alpha\beta} = \frac{\Delta_\alpha}{1 + w_\alpha} - \frac{\Delta_\beta}{1 + w_\beta}.
\]
Since both the number density and the entropy density evolve with temperature as $T^3$, we have $\delta n_r/n_r = (3/4)\delta \rho_r/\rho_r$, while for matter $\delta n_m/n_m = \delta \rho_m/\rho_m$. Therefore,

$$S_{rm} = \frac{3}{4} \frac{\delta \rho_r}{\rho_r} - \frac{\delta \rho_m}{\rho_m}. \quad (454)$$

A non-zero entropy density thus corresponds to a spatial inhomogeneity in the relative number density, or equivalently, to spatial variations of the equation of state.

9 Quantum field theory in curved space-time

We will describe in this section the formalism used to describe the origin of metric fluctuations during inflation.

9.1 Scalar field perturbed equations

Consider the action (256) with line element

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

in the Longitudinal gauge, where $\Phi$ is the gauge-invariant gravitational potential (341). Then the gauge-invariant equations for the perturbations on comoving hypersurfaces (constant energy density hypersurfaces) are

$$\Phi'' + 3H\Phi' + (H' + 2H^2)\Phi = \frac{\kappa^2}{2} \left[\phi'\delta\phi' - a^2V'(\phi)\delta\phi\right], \quad (455)$$

$$-\nabla^2\Phi + 3H\Phi' + (H' + 2H^2)\Phi = -\frac{\kappa^2}{2} \left[\phi'\delta\phi' + a^2V'(\phi)\delta\phi\right], \quad (456)$$

$$\Phi' + H\Phi = \frac{\kappa^2}{2} \phi'\delta\phi, \quad (457)$$

$$\delta\phi'' + 2H\delta\phi' - \nabla^2\delta\phi + a^2V''(\phi)\delta\phi = 4\phi'\Phi' - 2a^2V'(\phi)\Phi. \quad (458)$$

This system of equations seem too difficult to solve at first sight. However, there is a gauge-invariant combination of Mukhanov variables

$$u \equiv a\delta\phi + z\Phi,$$

$$z \equiv a\frac{\phi'}{H},$$

for which the above equations simplify enormously,

$$u'' - \nabla^2u - \frac{z''}{z}u = 0,$$

$$\nabla^2\Phi = \frac{\kappa^2}{2} \frac{H}{a^2} (zu' - z' u), \quad (458)$$

$$\left(\frac{a^2\Phi}{H}\right)' = \frac{\kappa^2}{2} zu. \quad (459)$$

From these, we can find a solution $u(z)$, which can be integrated to give $\Phi(z)$, and together allow us to obtain $\delta\phi(z)$. 

9.2 Canonical quantization in perturbation theory

Until now we have treated the perturbations as classical, but we should in fact consider the perturbations $\Phi$ and $\delta \phi$ as quantum fields. Note that the perturbed action for the scalar mode $u$ can be written as

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

(460)

In order to quantize the field $u$ in the curved background defined by the metric (302), we can write the operator

$$\hat{u}(\eta, x) = \int \frac{d^3k}{(2\pi)^3/2} \left[ u_k(\eta) \hat{a}_k e^{ik \cdot x} + u_k^*(\eta) \hat{a}_k^\dagger e^{-ik \cdot x} \right],$$

(461)

where the creation and annihilation operators satisfy the commutation relation of bosonic fields, and the scalar field’s Fock space is defined through the vacuum condition,

$$[\hat{a}_k, \hat{a}_k^\dagger] = \delta^3(k - k'),$$

(462)

$$\hat{a}_k |0\rangle = 0.$$  

(463)

Note that we are not assuming that the inflaton is a fundamental scalar field, but that it can be written as a quantum field with its commutation relations (as much as a pion can be described as a quantum field at low energies).

If we impose the equal-time commutation relations on the fields themselves,

$$[\hat{u}(\eta, x), \hat{\Pi}^\dagger(\eta, x')] = i\hbar \delta^3(x - x'),$$

we find a normalization condition on the modes $u_k$

$$u_k u_k^* - u_k^* u_k = i,$$

(464)

that coincides with the Wronskian of the mode equation,

$$u_k'' + \left( k^2 - \frac{z''}{z} \right) u_k = 0.$$  

(465)

Note that the modes decouple in linear perturbation theory. The ratio $U(\eta) = z''/z$ acts like a time-dependent potential for this 1D Schrödinger like equation (with time ↔ space),

$$-u_k'' + U(\eta) u_k = k^2 u_k.$$

In order to find exact solutions to the mode equation, we will use the slow-roll parameters (270),

$$\epsilon = 1 - \frac{\mathcal{H}'}{\mathcal{H}^2} = \frac{\kappa^2 z^2}{2 a^2},$$

(466)

$$\delta = 1 - \frac{\phi''}{\mathcal{H}\phi'} = 1 + \epsilon - \frac{z'}{\mathcal{H} z},$$

(467)

$$\xi = - \left( 2 - \epsilon - 3\delta + \delta^2 - \frac{\phi''}{\mathcal{H}^2 \phi'} \right).$$

(468)

In terms of these parameters, the conformal time and the effective potential for the $u_k$ mode can be written as

$$\eta = -\frac{1}{\mathcal{H}} + \int \frac{\epsilon da}{a\mathcal{H}},$$

(469)

$$\frac{z''}{z} = \mathcal{H}^2 \left[ (1 + \epsilon - \delta)(2 - \delta) + \mathcal{H}^{-1}(\epsilon' - \delta') \right].$$

(470)
Note that the slow-roll parameters, \( (466) \) and \( (467) \), can be taken as constant,\(^9\) to order \( \mathcal{O}(\epsilon^2) \),

\[
\epsilon' = 2\mathcal{H} \left( \epsilon^2 - \epsilon\delta \right) = \mathcal{O}(\epsilon^2), \\
\delta' = \mathcal{H} \left( \epsilon\delta - \xi \right) = \mathcal{O}(\epsilon^2). \tag{471}
\]

In that case, for constant slow-roll parameters, we can write

\[
\eta = -\frac{1}{\mathcal{H}} \frac{1}{1 - \epsilon}, \tag{472}
\]

\[
\frac{z''}{z} = \frac{1}{\eta^2} \left( \nu^2 - \frac{1}{4} \right), \quad \text{where} \quad \nu = \frac{1 + \epsilon - \delta}{1 - \epsilon} + \frac{1}{2}. \tag{473}
\]

### 9.3 Exact solutions

We are now going to search for approximate solutions of the mode equation \( (465) \), where the effective potential \( (470) \) is of order \( z''/z \simeq 2\mathcal{H}^2 \) in the slow-roll approximation. In quasi-de Sitter there is a characteristic scale given by the (event) horizon size or Hubble scale during inflation, \( H^{-1} \). There will be modes \( u_k \) with physical wavelengths much smaller than this scale, \( k/a \gg H \), that are well within the de Sitter horizon and therefore do not feel the curvature of space-time. On the other hand, there will be modes with physical wavelengths much greater than the Hubble scale, \( k/a \ll H \). In these two asymptotic regimes, the solutions can be written as

\[
u k \gg aH,
\]

\[
u k \ll aH. \tag{475}
\]

In the limit \( k \gg aH \) the modes behave like ordinary quantum modes in Minkowsky space-time, appropriately normalized, while in the opposite limit, \( u/z \) becomes constant on superhorizon scales. For approximately constant slow-roll parameters one can find exact solutions to \( (465) \), with the effective potential given by \( (473) \), that interpolate between the two asymptotic solutions,

\[
\|u_k\| = \frac{\sqrt{\pi}}{2} e^{i\nu \frac{1}{2} \eta} (-\eta)^{1/2} H^{(1)}_{\nu} (-k\eta), \tag{476}
\]

where \( H^{(1)}_{\nu}(z) \) is the Hankel function of the first kind (e.g. \( H^{(1)}_{3/2}(x) = -e^{ix} \sqrt{2/\pi x} (1 + i/x) \)), and \( \nu \) is given by \( (473) \) in terms of the slow-roll parameters. In the limit \( k\eta \to 0 \), the solution becomes

\[
|u_k| = \frac{2^{\nu - 2}}{\sqrt{2\kappa}} \frac{\Gamma(\nu)}{\Gamma(\nu)} (-k\eta)^{\nu - 1} \sqrt{2\kappa} (\frac{k}{aH})^{\frac{1}{2} - \nu}, \tag{477}
\]

\[
C(\nu) = 2^{\nu - 3} \Gamma(\nu) \Gamma\left(\frac{3}{2}\right) (1 - \epsilon)^{\nu - \frac{1}{2}} \simeq 1 \quad \text{for} \quad \epsilon, \delta \ll 1. \tag{478}
\]

We can now compute \( \Phi \) and \( \delta\phi \) from the super-Hubble-scale mode solution \( (475) \), for \( k \ll aH \). Substituting into Eq. \( (459) \), we find

\[
\Phi = C_1 \left( 1 - \frac{\mathcal{H}}{a^2} \int a^2 d\eta \right) + C_2 \frac{\mathcal{H}}{a^2}, \tag{479}
\]

\[
\frac{\delta\phi}{\delta' t} = \frac{C_1}{a^2} \int a^2 d\eta - \frac{C_2}{a^2}. \tag{480}
\]

\(^9\)For instance, there are models of inflation, like power-law inflation, \( a(t) \sim t^p \), where \( \epsilon = \delta = 1/p < 1 \), that give constant slow-roll parameters.
The term proportional to $C_1$ corresponds to the growing solution, while that proportional to $C_2$ corresponds to the decaying solution, which can soon be ignored. These quantities are gauge invariant but evolve with time outside the horizon, during inflation, and before entering again the horizon during the radiation or matter eras. We would like to write an expression for a gauge invariant quantity that is also constant for superhorizon modes. Fortunately, in the case of adiabatic perturbations, there is such a quantity:

$$\zeta \equiv \Phi + \frac{1}{\epsilon \mathcal{H}} (\Phi' + \mathcal{H}\Phi) = \frac{u}{z},$$

(481)

which is constant, see Eq. (475), for $k \ll aH$. In fact, this quantity $\zeta$ is identical, for superhorizon modes, to the gauge invariant curvature metric perturbation $R_c$ on comoving (constant energy density) hypersurfaces,

$$\zeta = R_c + \frac{1}{\epsilon \mathcal{H}^2} \nabla^2 \Phi.$$  

(482)

Using Eq. (458) we can write the evolution equation for $\zeta = \frac{u}{z}$ as

$$\zeta'' + \left( k^2 - \frac{a''}{a} \right) \zeta = 0,$$

(489)

These expressions will be of special importance later.

### 9.4 Gravitational wave perturbations

Let us now compute the tensor or gravitational wave metric perturbations generated during inflation. The perturbed action for the tensor mode can be written as

$$\delta S = \frac{1}{2} \int d^3 x \, d\eta \, a^2 \frac{a^2}{2\kappa^2} \left[ (h_{ij}')^2 - (\nabla h_{ij})^2 \right],$$

(484)

with the tensor field $h_{ij}$ considered as a quantum field,

$$\hat{h}_{ij}(\eta, \mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \sum_{\lambda=1,2} \left[ h_k(\eta) e_{ij}(k, \lambda) \hat{a}_{k,\lambda} e^{ik \cdot x} + h.c. \right],$$

(485)

where $e_{ij}(k, \lambda)$ are the two polarization tensors, satisfying symmetric, transverse and traceless conditions

$$e_{ij} = e_{ji}, \quad k^i e_{ij} = 0, \quad e_{ii} = 0,$$

(486)

$$e_{ij}(-k, \lambda) = e_{ij}^*(k, \lambda), \quad \sum_{\lambda} e_{ij}^*(k, \lambda) e^{i\lambda} = 4,$$

(487)

while the creation and annihilation operators satisfy the usual commutation relation of bosonic fields, Eq. (462). We can now redefine our gauge invariant tensor amplitude as

$$v_k(\eta) = \frac{a}{\sqrt{2\kappa}} h_k(\eta),$$

(488)

which satisfies the following evolution equation, decoupled for each mode $v_k(\eta)$ in linear perturbation theory,

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

(489)
where we have used $R$ and the running of the tilt inflationary model, we can have significant departures from scale invariance. The scalar tilt which is either positive (see Eqs. (277), (278). Note from this equation that it is possible, in principle, to obtain from inflation a term of its amplitude at horizon-crossing, $A$.

We can solve equation (489) in the two asymptotic regimes,

\[ v_k = \frac{1}{2k} e^{-ik\eta} \quad k \gg aH, \]
\[ v_k = C_3(k) a \quad k \ll aH. \]

In the limit $k \gg aH$ the modes behave like ordinary quantum modes in Minkowsky space-time, appropriately normalized, while in the opposite limit, the metric perturbation $h_k$ becomes constant on superhorizon scales. For constant slow-roll parameters one can find exact solutions to (489), with effective potential given by (490), that interpolate between the two asymptotic solutions. These are identical to Eq. (476) except for the substitution $\nu \to \mu$. In the limit $k\eta \to 0$, the solution becomes

\[ |v_k| = \frac{C(\mu)}{\sqrt{2k}} \left( \frac{k}{aH} \right)^{\frac{1}{2} - \mu}. \]

Since the mode $h_k$ becomes constant on superhorizon scales, we can evaluate the tensor metric perturbation when it reentered during the radiation or matter era directly in terms of its value during inflation.

### 9.5 Power spectra of scalar and tensor metric perturbations

Not only do we expect to measure the amplitude of the metric perturbations generated during inflation and responsible for the anisotropies in the CMB and density fluctuations in LSS, but we should also be able to measure its power spectrum, or two-point correlation function in Fourier space. Let us consider first the scalar metric perturbations $R_k$, which enter the horizon at $a = k/H$. Its correlator is given by

\[ \langle 0|R_k^* R_{k'}|0 \rangle = \frac{|u_k|^2}{z^2} \delta^3(k - k') \equiv \frac{P_R(k)}{4\pi k^3} (2\pi)^3 \delta^3(k - k'), \]
\[ P_R(k) = \frac{k^3}{2\pi^2} \frac{|u_k|^2}{z^2} = \frac{\kappa^2}{2\epsilon} \left( \frac{H}{2\pi} \right)^2 \left( \frac{k}{aH} \right)^{3 - 2\nu} = A_S^3 \left( \frac{k}{aH} \right)^{n_s - 1}, \]

where we have used $R_k = \zeta_k = \frac{u_k}{z}$ and Eq. (477). This last equation determines the power spectrum in terms of its amplitude at horizon-crossing, $A_S$, and a tilt,

\[ n_s - 1 \equiv \frac{d\ln P_R(k)}{d\ln k} = 3 - 2\nu = 2 \left( \frac{\delta - 2\epsilon}{1 - \epsilon} \right) \approx 2\eta_V - 6\epsilon_V, \]

see Eqs. (277), (278). Note from this equation that it is possible, in principle, to obtain from inflation a scalar tilt which is either positive ($n > 1$) or negative ($n < 1$). Furthermore, depending on the particular inflationary model, we can have significant departures from scale invariance.

Note that at horizon entry $k\eta = -1$, and thus we can alternatively evaluate the tilt as

\[ n_s - 1 \equiv -\frac{d\ln P_R}{d\ln \eta} = -2\eta \dot{H} \left[ (1 - \epsilon) - (\epsilon - \delta) - 1 \right] = 2 \left( \frac{\delta - 2\epsilon}{1 - \epsilon} \right) \approx 2\eta_V - 6\epsilon_V, \]

and the running of the tilt

\[ \frac{dn_s}{d\ln k} = - \frac{dn_s}{d\ln \eta} = -\eta \dot{H} \left( 2\xi + 8\epsilon^2 - 10\epsilon\delta \right) \approx 2\xi_V + 24\epsilon_V^2 - 16\eta_V \epsilon_V, \]
where we have used Eqs. (471).

Let us consider now the tensor ( gravitational wave) metric perturbation, which enter the horizon at \( a = k/H \),

\[
\sum_{\lambda} \langle 0 | h^e_{\lambda,\lambda} h_{\lambda',\lambda'} | 0 \rangle = 4 \frac{2\kappa^2}{a^2} |v_k|^2 \delta^3(k - k') \equiv \frac{\mathcal{P}_g(k)}{4\pi k^3} (2\pi)^3 \delta^3(k - k'),
\]

where we have used Eqs. (471).

The fluctuations of a massless minimally-coupled scalar field are scale invariant.

Alternatively, we can evaluate the tensor tilt by

\[
n_T = - \frac{d\ln \mathcal{P}_T}{d\ln k} = -2\eta \mathcal{H} \left[ (1 - \epsilon) - 1 \right] = \frac{-2\epsilon}{1 - \epsilon} \simeq -2\epsilon_V,
\]

and its running by

\[
\frac{dn_T}{d\ln k} = - \frac{dn_T}{d\ln \eta} = -\eta \mathcal{H} \left( 4\epsilon^2 - 4\epsilon \delta \right) \simeq 8\epsilon_V^2 - 4\eta_V \epsilon_V,
\]

where we have used Eqs. (471).

**9.6 Massless minimally coupled scalar field fluctuations**

The fluctuations of a massless minimally-coupled scalar field \( \phi \) during inflation (quasi de Sitter) are quantum fields in a curved background. We will redefine \( y(x, t) = a(t) \delta \phi(x, t) \), whose action is

\[
S = \frac{1}{2} \int d^3x \, dt \left[ (y')^2 - (\nabla y)^2 + \alpha'' \frac{a''}{a} y^2 \right],
\]

where primes denote derivatives w.r.t. conformal time \( \eta = \int dt/a(t) = -1/(aH) \), with \( H \) the constant rate of expansion during inflation. Using the identity \( (y')^2 + \alpha'' \frac{a''}{a} y = (y' - \alpha' \frac{a'}{a} y)^2 + (\alpha' \frac{a'}{a} y')^2 \), we can define the conjugate momentum as \( p = \frac{\partial S}{\partial \dot{y}} = y' - \alpha' \frac{a'}{a} y \), and write the corresponding Hamiltonian as

\[
\mathcal{H} = \frac{1}{2} \int d^3x \left[ p^2 + (\nabla y)^2 + 2 \alpha' \frac{a'}{a} p y \right].
\]

We can now Fourier transform: \( \Phi(k, \eta) = \int \frac{d^3x}{(2\pi)^{3/2}} \Phi(x, \eta) e^{-i \cdot x \cdot k} \) all the fields and momenta. Since the scalar field is assumed real, we have: \( y(k, \eta) = \imath y^\dagger(-k, \eta) \) and \( p(k, \eta) = p^\dagger(-k, \eta) \), and the Hamiltonian becomes

\[
\mathcal{H} = \frac{1}{2} \int d^3k \left[ p(k, \eta) p^\dagger(k, \eta) + k^2 y(k, \eta) y^\dagger(k, \eta) \right.
\]

\[
+ \left. \alpha' \left( y(k, \eta) p^\dagger(k, \eta) + p(k, \eta) y^\dagger(k, \eta) \right) \right].
\]

(507)
As we will see later, it is the last term which is responsible for squeezing.

The Euler-Lagrange equations for this field can be written in terms of the field eigenmodes as a series of uncoupled oscillator equations:

\[
p' = -i \left[ p, \mathcal{H} \right] = -k^2 y - \frac{\alpha'}{\alpha} p, \\
y' = -i \left[ y, \mathcal{H} \right] = p + \frac{\alpha'}{\alpha} y.
\]

\[
y''(k, \eta) + \left( k^2 - \frac{\alpha''}{\alpha} \right) y(k, \eta) = 0.
\]

(508)

9.7 Heisenberg picture: The field operators

We can now treat each mode as a quantum oscillator, and introduce the corresponding creation and annihilation operators:

\[
a(k, \eta) = \sqrt{\frac{k}{2}} y(k, \eta) + i \frac{1}{\sqrt{2k}} p(k, \eta),
\]

\[
a^\dagger(-k, \eta) = \sqrt{\frac{k}{2}} y(k, \eta) - i \frac{1}{\sqrt{2k}} p(k, \eta),
\]

which can be inverted to give

\[
y(k, \eta) = \frac{1}{\sqrt{2k}} \left[ a(k, \eta) + a^\dagger(-k, \eta) \right],
\]

\[
p(k, \eta) = -i \sqrt{\frac{k}{2}} \left[ a(k, \eta) - a^\dagger(-k, \eta) \right].
\]

(511)

The usual equal-time commutation relations for fields ($\hbar = 1$ here and throughout),

\[
\left[ y(x, \eta), p(x', \eta) \right] = i \delta^3(x - x'),
\]

becomes a commutation relation for the creation and annihilation operators,

\[
\left[ y(k, \eta), p^\dagger(k', \eta) \right] = i \delta^3(k - k') \Rightarrow \left[ a(k, \eta), a^\dagger(k', \eta) \right] = \delta^3(k - k').
\]

(514)

In terms of these operators, the Hamiltonian becomes:

\[
\mathcal{H} = \frac{1}{2} \int d^3 \mathbf{k} \left[ k \left[ a(k, \eta) a^\dagger(k, \eta) + a^\dagger(k, \eta) a(k, \eta) \right] + \frac{\alpha'}{\alpha} \left[ a^\dagger(k, \eta) a^\dagger(k, \eta) + a(k, \eta) a(k, \eta) \right] \right].
\]

(515)

It is the last (non-diagonal) term which is responsible for squeezing.

The evolution equation can be written as

\[
\begin{pmatrix}
a'(k) \\
a^\dagger'(-k)
\end{pmatrix} = \begin{pmatrix}
-ik & \frac{\alpha'}{\alpha} \\\n\frac{\alpha'}{\alpha} & ik
\end{pmatrix} \begin{pmatrix}
a(k) \\
a^\dagger(-k)
\end{pmatrix},
\]

(516)

whose general solution is, in terms of the initial conditions $a(k, \eta_0)$,

\[
a(k, \eta) = u_k(\eta) a(k, \eta_0) + v_k(\eta) a^\dagger(-k, \eta_0),
\]

\[
a^\dagger(-k, \eta) = u^*_k(\eta) a^\dagger(-k, \eta_0) + v^*_k(\eta) a(k, \eta_0),
\]

(517) (518)
which correspond to a Bogoliubov transformation of the creation and annihilation operators, and characterizes the time evolution of the system of harmonic oscillators in the Heisenberg representation.

The commutation relation (514) is preserved under the unitary evolution if
\[
|u_k(\eta)|^2 - |v_k(\eta)|^2 = 1,
\]
which gives a normalization condition for these functions.

We can write the quantum fields \(y\) and \(p\) in terms of these as,
\[
y(k, \eta) = f_k(\eta) a(k, \eta_0) + f_k^*(\eta) a^\dagger(-k, \eta_0),
\]
\[
p(k, \eta) = -i \left[ g_k(\eta) a(k, \eta_0) - g_k^*(\eta) a^\dagger(-k, \eta_0) \right],
\]
where the functions
\[
f_k(\eta) = \frac{1}{\sqrt{2k}} [u_k(\eta) + v_k^*(\eta)],
\]
\[
g_k(\eta) = \sqrt{\frac{k}{2}} [u_k(\eta) - v_k^*(\eta)],
\]
are the field and momentum modes, respectively, satisfying the following equations and initial conditions,
\[
f''_k + \left( k^2 - \frac{a''}{a} \right) f_k = 0,
\]
\[
g_k = i \left( f'_k - \frac{a'}{a} f_k \right),
\]
as well as the Wronskian condition,
\[
i (f^*_k f_k - f^*_k f_k) = g_k f^*_k + g_k f_k = 1.
\]

### 9.8 Squeezing parameters

Since we have two complex functions, \(f_k\) and \(g_k\), plus a constraint (526), we can write these in terms of three real functions in the standard parametrization for squeezed states,
\[
u_k(\eta) = e^{-i \theta_k(\eta)} \cosh r_k(\eta),
\]
\[
v_k(\eta) = e^{i \theta_k(\eta) + 2i \phi_k(\eta)} \sinh r_k(\eta),
\]
where \(r_k\) is the squeezing parameter, \(\phi_k\) the squeezing angle, and \(\theta_k\) the phase.

We can also write its relation to the usual Bogoliubov formalism in terms of the functions \(\{\alpha_k, \beta_k\}\),
\[
u_k = \alpha_k e^{-ik\eta}, \quad v_k^* = \beta_k e^{ik\eta},
\]
which is useful for the adiabatic expansion, and allows one to write the average number of particles and other quantities,
\[
n_k = |\beta_k|^2 = |v_k|^2 = \frac{1}{2k} \left| g_k - k f_k \right|^2 = \sinh^2 r_k,
\]
\[
\sigma_k = 2 \text{Re} \left( \alpha_k \beta_k e^{2i\kappa \eta} \right) = 2 \text{Re} \left( u_k v_k^* \right) = \cos 2\phi_k \sinh 2r_k,
\]
\[
\tau_k = 2 \text{Im} \left( \alpha_k \beta_k e^{2i\kappa \eta} \right) = 2 \text{Im} \left( u_k v_k^* \right) = -\sin 2\phi_k \sinh 2r_k.
\]
We can invert these expressions to give \((r_k, \theta_k, \phi_k)\) as a function of \(u_k\) and \(v_k\),

\[
\begin{align*}
sinh r_k &= \sqrt{Re u_k^2 + Im u_k^2}, \\
\cosh r_k &= \sqrt{Re u_k^2 + Im u_k^2}, \\
\tan \theta_k &= -\frac{Im u_k}{Re u_k}, \\
\tan 2\phi_k &= \frac{Im u_k Re u_k + Im u_k Re u_k}{Re u_k Re u_k - Im u_k Im u_k}.
\end{align*}
\] (533)

We can now write Eqs. (520) and (521) in terms of the initial values,

\[
\begin{align*}
y(k, \eta) &= \sqrt{2k} f_{k1}(\eta) y(k, \eta_0) - \sqrt{\frac{2}{k}} f_{k2}(\eta) p(k, \eta_0), \\
p(k, \eta) &= \sqrt{\frac{2}{k}} g_{k1}(\eta) p(k, \eta_0) + \sqrt{2k} g_{k2}(\eta) y(k, \eta_0),
\end{align*}
\] (535) (536)

where subindices 1 and 2 correspond to real and imaginary parts, \(f_{k1} \equiv Re f_k\) and \(f_{k2} \equiv Im f_k\), and similarly for the momentum mode \(g_k\).

### 9.9 The Schrödinger picture: The vacuum wave function

Let us go now from the Heisenberg to the Schrödinger picture, and compute the initial state vacuum eigenfunction \(\Psi_0(\eta = \eta_0)\). The initial vacuum state \(|0, \eta_0\rangle\) is defined through the condition

\[
\forall k, \quad \hat{a}(k, \eta_0)|0, \eta_0\rangle = \left[\sqrt{\frac{k}{2}} \hat{y}_k(\eta_0) + i \frac{1}{\sqrt{2k}} \hat{p}_k(\eta_0)\right]|0, \eta_0\rangle = 0,
\]

\[
\left\{\begin{array}{l}
y_k^0 + \frac{1}{k} \frac{\partial}{\partial y_k^0}\end{array}\right\} \Psi_0 \left(y_k^0, y_k^{0*}, \eta_0\right) = 0 \implies \Psi_0 \left(y_k^0, y_k^{0*}, \eta_0\right) = N_0 e^{-k |y_k^0|^2}
\]

where we have used the position representation, \(\hat{y}_k(\eta_0) = y_k^0\), \(\hat{p}_k(\eta_0) = -i \frac{\partial}{\partial y_k^0}\), and \(N_0\) gives the corresponding normalization.

We will now study the time evolution of this initial wave function using the unitary evolution operator \(S = S(\eta, \eta_0)\), i.e. the state evolves in the Schrödinger picture as \(|0, \eta\rangle = S|0, \eta_0\rangle\). Now, inverting (520) and (521)

\[
\hat{a}(k, \eta_0) = g_k^0(\eta) \hat{y}(k, \eta) + i f_k^0(\eta) \hat{p}(k, \eta),
\] (537)

which, acting on the initial state becomes, \(\forall k, \forall \eta,\)

\[
S \left[\hat{y}(k, \eta) + i \frac{f_k^0(\eta)}{g_k^0(\eta)} \hat{p}(k, \eta)\right] S^{-1} |0, \eta_0\rangle = 0
\]

\[
\implies \left[\hat{y}_k(\eta_0) + i \frac{f_k^0(\eta)}{g_k^0(\eta)} \hat{p}_k(\eta_0)\right]|0, \eta\rangle = 0,
\]

\[
\implies \Psi_0 \left(y_k^0, y_k^{0*}, \eta\right) = \frac{1}{\sqrt{\pi |f_k(\eta)|}} e^{-\Omega_k(\eta) |y_k^0|^2},
\] (538)

where

\[
\Omega_k(\eta) = \frac{g_k^0(\eta)}{f_k^0(\eta)} = k \frac{u_k^2 - v_k}{u_k^2 + v_k} = \frac{1 - 2i F_k(\eta)}{2 |f_k(\eta)|^2},
\]

\[
F_k(\eta) = Im(f_k^0 g_k) = Im(u_k v_k) = \frac{1}{2} \sin 2\phi_k \sinh 2r_k.
\] (539) (540)

We see that the unitary evolution preserves the Gaussian form of the wave functional. The wave function (538) is called a 2-mode squeezed state.
The normalized probability distribution,

\[ P_0 (y(k, \eta_0), y(-k, \eta_0), \eta) = \frac{1}{\pi |f_k(\eta)|^2} \exp \left( -\frac{|y(k, \eta_0)|^2}{|f_k(\eta)|^2} \right), \]

is a Gaussian distribution, with dispersion given by \( |f_k|^2 \).

In fact, we can compute the vacuum expectation values,

\[
\begin{align*}
\langle \Delta y(k, \eta) \Delta y^\dagger(k', \eta) \rangle &= \Delta y^2(k) \delta^3(k - k') = |f_k|^2 \delta^3(k - k'), \\
\langle \Delta p(k, \eta) \Delta p^\dagger(k', \eta) \rangle &= \Delta p^2(k) \delta^3(k - k') = |g_k|^2 \delta^3(k - k'),
\end{align*}
\]

and therefore the Heisenberg uncertainty principle reads

\[
\Delta y^2(k) \Delta p^2(k) = |f_k|^2 |g_k|^2 = F_k^2(\eta) + \frac{1}{4} \geq \frac{1}{4}.
\]

It is clear that for \( \eta = \eta_0, \Omega_k(\eta_0) = k \) and \( F_k(\eta_0) = 0 \), and thus we have initially a minimum wave packet, \( \Delta y \Delta p = \frac{1}{2} \). However, through its unitary evolution, the function \( F_k \) grows exponentially, see (540), and we quickly find \( \Delta y \Delta p \gg \frac{1}{2} \), corresponding to the semiclassical regime, as we will soon demonstrate rigorously.

### 9.10 The squeezing formalism

Let us now use the squeezing formalism to describe the evolution of the wave function. The equations of motion for the squeezing parameters follow from those of the field and momentum modes,

\[
\begin{align*}
\phi_k' &= -k - \frac{a'}{a} \coth 2r_k \sin 2\phi_k, \\
\phi_k' &= k + \frac{a'}{a} \tanh 2r_k \sin 2\phi_k.
\end{align*}
\]

As we will see, the evolution is driven towards large \( r_k \propto N \gg 1 \), the number of e-folds during inflation. Thus, in that limit,

\[
(\theta_k + \phi_k)' = -\frac{a'}{a} \sin 2\phi_k \frac{\sinh 2r_k}{2r_k} \to 0,
\]

and therefore \( \theta_k + \phi_k \to \text{const.} \) We can always choose this constant to be zero, so that the real and imaginary components of the field and momentum modes become

\[
\begin{align*}
f_{k1} &= \frac{1}{\sqrt{2k}} e^{r_k} \cos \phi_k, & f_{k2} &= \frac{1}{\sqrt{2k}} e^{-r_k} \sin \phi_k, \\
g_{k1} &= \sqrt{\frac{k}{2}} e^{-r_k} \cos \phi_k, & g_{k2} &= \sqrt{\frac{k}{2}} e^{r_k} \sin \phi_k.
\end{align*}
\]

It is clear that, in the limit of large squeezing \( (r_k \to \infty) \), the field mode \( f_k \) becomes purely real, while the momentum mode \( g_k \) becomes pure imaginary.

This means that the field (535) and momentum (536) operators become, in that limit,

\[
\begin{align*}
\hat{y}(k, \eta) &\to \sqrt{2k} f_{k1}(\eta) \hat{y}(k, \eta_0) \\
\hat{p}(k, \eta) &\to \sqrt{2k} g_{k2}(\eta) \hat{y}(k, \eta_0)
\end{align*}
\]

\[
\Rightarrow \hat{p}(k, \eta) \simeq \frac{g_{k2}(\eta)}{f_{k1}(\eta)} \hat{y}(k, \eta).
\]
As a consequence of this squeezing, information about the initial momentum \( \hat{p}_0 \) distribution is lost, and the positions (or field amplitudes) at different times commute,

\[
\left[ \hat{y}(\mathbf{k}, \eta_1), \hat{y}(\mathbf{k}, \eta_2) \right] \approx \frac{1}{2} e^{-2r_k} \cos^2 \phi_k \approx 0.
\] (551)

This result defines what is known as a quantum non-demolition (QND) variable, which means that one can perform successive measurements of this variable with arbitrary precision without modifying the wave function. Note that \( y = a\delta \phi \) is the amplitude of fluctuations produced during inflation, so what we have found is: first, that the amplitude is distributed as a classical Gaussian random field with probability (541); and second that we can measure its amplitude at any time, and as much as we like, without modifying the distribution function.

In a sense, this problem is similar to that of a free non-relativistic quantum particle, described initially by a minimum wave packet, with initial expectation values \( \langle x \rangle_0 = x_0 \) and \( \langle \hat{p} \rangle_0 = p_0 \), which becomes broader by its unitary evolution, and at late times \((t \gg mx_0/p_0)\) this Gaussian state becomes an exact WKB state,

\[
\Psi(x) = \Omega_R^{-1/2} \exp(-\Omega x^2/2),
\]

with \( \text{Im} \Omega \gg \text{Re} \Omega \) (i.e. high squeezing limit). In that limit, \([\hat{x}, \hat{p}] \approx 0\), and we have lost information about the initial position \( x_0 \) (instead of the initial momentum like in the inflationary case), \( \hat{x}(t) \approx \hat{p}(t) t/m = p_0 t/m \) and \( \hat{p}(t) = p_0 \). Therefore, not only \([\hat{p}(t_1), \hat{p}(t_2)] = 0\), but also, at late times, \([\hat{x}(t_1), \hat{x}(t_2)] \approx 0\).

### 9.11 The Wigner function

The Wigner function is the best candidate for a probability density of a quantum mechanical system in phase-space. Of course, we know from QM that such a probability distribution function cannot exist, but the Wigner function is just a good approximation to that distribution. Furthermore, for a Gaussian state, this function is in fact positive definite.

Consider a quantum state described by a density matrix \( \hat{\rho} \). Then the Wigner function can be written as

\[
W(y_k^0, y_k^0^*, p_k^0, p_k^0^*) = \frac{1}{(2\pi)^2} \int \int dx_1 dx_2 e^{-i(p_1 x_1 + p_2 x_2)} \langle y - x/2, \eta | \hat{\rho} | y + x/2, \eta \rangle.
\]

If we substitute for the state our vacuum initial condition \( \hat{\rho} = |\Psi_0\rangle \langle \Psi_0| \), with \( \Psi_0 \) given by (538), we can perform the integration explicitly to obtain

\[
W_0(y_k^0, y_k^0^*, p_k^0, p_k^0^*) = \frac{1}{\pi^2} \exp \left( -\frac{|y|^2}{|f_k|^2} - 4|f_k|^2 p - \frac{F_k}{|f_k|^2} y^2 \right) \equiv \Phi(y_1, p_1) \Phi(y_2, p_2)
\]

\[
\Phi(y_1, p_1) = \frac{1}{\pi} \exp \left( -\left( \frac{y_1^2}{|f_k|^2} + 4|f_k|^2 \bar{p}_1 \right) \right),
\]

\[
\bar{p}_1 \equiv p_1 - \frac{F_k}{|f_k|^2} y_1.
\]

In general, \( W_0 \) describes an asymmetric Gaussian in phase space, whose \( 2\sigma \) contours satisfy

\[
\frac{y_1^2}{|f_k|^2} + 4|f_k|^2 \bar{p}_1^2 \leq 1.
\] (554)
For instance, at time $\eta = \eta_0$, we have $y_1^0 = \frac{1}{\sqrt{2k}} = |f_k(\eta_0)|$, $p_1^0 = \sqrt{\frac{k}{2}} = 1/2|f_k(\eta_0)|$, and $F_k(\eta_0) = 0$, so that $\bar{p}_1 = p_1^0$, and the $2\sigma$ contours become

$$\frac{y_1^2}{y_1^{02}} + \frac{p_1^2}{p_1^{02}} \leq 1,$$

which is a circle in phase space.

On the other hand, for time $\eta \gg \eta_0$, we have

$$|f_k| \to \frac{1}{\sqrt{2k}} e^{\gamma k} \sim y_k^0 e^N, \quad \text{growing mode}, \quad (555)$$

$$\frac{1}{2|f_k|} \to \sqrt{\frac{k}{2}} e^{-\gamma k} \sim p_k^0 e^{-N}, \quad \text{decaying mode}, \quad (556)$$

so that the ellipse (554) becomes highly “squeezed”.

Note that Liouville’s theorem implies that the volume of phase space is conserved under Hamiltonian (unitary) evolution, so that the area within the ellipse should be conserved. As the probability distribution compresses (squeezes) along the $p$-direction, it expands along the $y$-direction. At late times, the Wigner function is highly concentrated around the region

$$\hat{p}^2 = \left( p - \frac{F_k}{|f_k|^2} y \right)^2 < \frac{1}{4|f_k|^2} \sim e^{-2N} \ll 1. \quad (557)$$

We can thus take the above squeezing limit in the Wigner function (552) and write the exponential term as a Dirac delta function,

$$W_0(y, p) \xrightarrow{r_k \to \infty} \frac{1}{\pi^2} \exp \left\{ - \frac{|y|^2}{|f_k|^2} \right\} \delta \left( p - \frac{F_k}{|f_k|^2} y \right). \quad (558)$$

In this limit we have

$$\hat{p}_k(\eta) = \frac{F_k}{|f_k|^2} \hat{y}_k(\eta) \simeq \frac{g_{k2}(\eta)}{f_{k1}(\eta)} \hat{y}_k(\eta), \quad (559)$$

so we recover the previous result (550). This explains why we can treat the system as a classical Gaussian random field: the amplitude of the field $y$ is uncertain with probability distribution (541), but once a measurement of $y$ is performed, we can automatically assign to it a definite value of the momentum, according to (550).

Note that the condition $F_k^2 \gg 1$ is actually a condition between operators and their commutators/anticommutators. The Heisenberg uncertainty principle states that

$$\Delta_{\psi} A \Delta_{\psi} B \geq \frac{1}{2} \left| \langle \Psi | [A, B] | \Psi \rangle \right|,$$

for any two hermitian operators (observables) in the Hilbert space of the wave function $\Psi$. In our case, and in Fourier space, this corresponds to (544)

$$\Delta y^2(k) \Delta p^2(k) = F_k^2(\eta) + \frac{1}{4} \geq \frac{1}{4} \left| \langle \Psi | |y_k(\eta), \ p_k^0(\eta)\rangle |\Psi\rangle \right|^2, \quad (560)$$

with $|\Psi\rangle = |0, \eta\rangle$ the evolved wave function.

On the other hand, the phase $F_k$ can be written as

$$F_k = \frac{1}{2} \left( g_k f_k^* - f_k g_k^* \right) = \frac{1}{2} \left( \frac{g_k}{f_k} |f_k|^2 - |f_k|^2 \frac{g_k^*}{f_k} \right) = \frac{1}{2} \langle \Psi | p(k, \eta) y^\dagger(k, \eta) + y(k, \eta) p^\dagger(k, \eta) |\Psi\rangle,$$

(561)
and we have used that, in the semiclassical limit, we can write
\[ \langle \Psi | y_k(\eta) | \Psi \rangle = | f_k |^2, \text{ as well as } p(k, \eta) = -i \frac{\partial}{\partial \eta} y(k, \eta), \text{ see (550)}. \]

The above relation just indicates that, for any state \( \Psi \), the condition of classicality \( (F_k \gg 1) \) is satisfied whenever, for that state,
\[ \{ y_k(\eta), p_k^\dagger(\eta) \} \gg \{ [y_k(\eta), p_k^\dagger(\eta)] \} = \hbar, \]
which is an interesting condition.

9.12 \textbf{Massless scalar field fluctuations on superhorizon scales}

The gauge invariant tensor fluctuations (gravitational waves) act as a minimally-coupled massless scalar field during inflation, so we will study here the generation of its fluctuations during quasi de Sitter.

Let us consider here the exact solutions to the equation of motion of a minimally-coupled massless scalar field during inflation or quasi de Sitter, with scale factor \( a = -1/H\eta \),
\[ f_k = \frac{1}{\sqrt{2k}} e^{-ik\eta} \left( 1 - \frac{i}{k\eta} \right), \quad g_k = i \left( f_k' - \frac{a'}{a} f_k \right) = \sqrt{\frac{k}{2}} e^{-ik\eta}, \]
which satisfy the Wronskian condition, \( g_k f_k^* + g_k^* f_k = 1 \). The eigenmodes become
\[ u_k = e^{-ik\eta} \left( 1 - \frac{i}{2k\eta} \right) = e^{-ik\eta - i\delta_k} \cosh r_k, \]
\[ v_k = e^{ik\eta} \frac{i}{2k\eta} = e^{ik\eta + i\frac{\pi}{2}} \sinh r_k, \]
which comparing with (527) and (528) provides the squeezing parameter, the angle and the phase, as inflation proceeds towards \( k\eta \to 0^- \),
\[ \sinh r_k = \tan \delta_k = \frac{1}{2k\eta} \to -\infty, \]
\[ \theta_k = k\eta + \arctan \frac{1}{2k\eta} \to -\frac{\pi}{2}, \quad \phi_k = \frac{\pi}{4} - \frac{1}{2} \arctan \frac{1}{2k\eta} \to \frac{\pi}{2}, \]
while the imaginary part of the phase of the wave function becomes
\[ F_k(\eta) = \frac{1}{2} \sin 2\phi_k \sinh 2r_k = \frac{1}{2k\eta} \to -\infty. \]

The number of scalar field particles produced during inflation grow exponentially, \( n_k = |\beta_k|^2 = \sinh^2 r_k = (2k\eta)^{-2} \to \infty. \)

Thus, through unitary evolution, the fluctuations will very soon enter the semiclassical regime due to a highly squeezed wave function. The question which remains is when do fluctuations become classical?

9.13 \textbf{Hubble crossing}

As we will see, the field fluctuation modes will become semiclassical as their wavelength becomes larger than the only physical scale in the problem, the de Sitter horizon scale, \( \lambda_{\text{phys}} = 2\pi a/k \gg H^{-1} \).

Therefore, let us consider the general solution to Eq. (524) for the superhorizon modes \( (k \ll aH) \),
\[ f_k(\eta) = C_1(k) a + C_2(k) a \int^\eta \frac{d\eta'}{a^2(\eta')} = C_1(k) a - C_2(k) \frac{1}{a^2 H}. \]
We can always choose \( C_1(k) \) to be real, while \( C_2(k) \) will be complex in general. The first term corresponds to the growing mode, while the second term is the decaying mode.

Integrating out \( g_k \) from (525), one finds

\[
g_k(\eta) = i C_2(k) \frac{1}{a} - i C_1(k) k^2 \frac{1}{a} \int a^2 \, d\eta = i C_2(k) \frac{1}{a} - i C_1(k) \frac{k^2}{H},
\]

where we have added a \( k^2 \) term for completeness. To second order in \( k^2 \), the Wronskian becomes

\[
C_1(k) \text{Im} C_2(k) \left( 1 + \frac{k^2}{a^2 H^2} \right) \simeq C_1(k) \text{Im} C_2(k) = -\frac{1}{2}.
\]

Comparing with the exact solutions (562), we find, to first order,

\[
C_1(k) = \frac{H_k}{\sqrt{2} k^3}, \quad C_2(k) = -i \frac{k^{3/2}}{\sqrt{2} H_k},
\]

where \( H_k \) is the Hubble rate at horizon crossing, \( k\eta = -1 \), i.e. when the perturbation’s physical wavelength becomes of the same order as the de Sitter horizon size, \( k = aH = H \).

We are now prepared to answer the question of classicality of the modes. Let us compute the wave function phase shift

\[
|F_k| = |\text{Im}(f_k g_k)| = \left| C_1^2(k) \frac{k^2 a}{H} + |C_2(k)|^2 \frac{1}{a^3 H} \right| - C_1(k) \text{Re} C_2(k) \left( 1 + \frac{k^2}{a^2 H^2} \right).
\]

Since only the first term remains after \( k\eta \to 0 \), we see that \(|F_k| \gg 1\) whenever

\[
C_1^2(k) = \frac{H_k^2}{2k^3} \gg \frac{H}{k^2 a} \quad \Rightarrow \quad \lambda_{\text{phys}} = \frac{2\pi a}{k} \gg \lambda_{\text{HC}} = \frac{2\pi}{H_k}.
\]

Therefore, we confirm that modes that start as Minkowski vacuum well inside the de Sitter horizon are stretched by the expansion and become semiclassical soon after horizon crossing, and their amplitude can be described as a classical Gaussian random variable.

Furthermore, the fact that the momentum is immediately defined once the amplitude for a given wavelength is known, implies that there is a fixed temporal phase coherence for all perturbations with the same wavelength. As we know, this implies that inflationary perturbations will induce coherent acoustic oscillations in the plasma just before decoupling, which should be seen in the microwave background anisotropies as acoustic peaks in the angular power spectrum.

10 Anisotropies of the microwave background

The Universe just before recombination is a very tightly coupled fluid, due to the large electromagnetic Thomson cross section. Photons scatter off charged particles (protons and electrons), and carry energy, so they feel the gravitational potential associated with the perturbations imprinted in the metric during inflation. An overdensity of baryons (protons and neutrons) does not collapse under the effect of gravity until it enters the causal Hubble radius. The perturbation continues to grow until radiation pressure opposes gravity and sets up acoustic oscillations in the plasma. Since overdensities of the same size will enter the Hubble radius at the same time, they will oscillate in phase. Moreover, since photons scatter off these baryons, the acoustic oscillations occur also in the photon field and induces a pattern of peaks in the temperature anisotropies in the sky, at different angular scales.
Three different effects determine the temperature anisotropies we observe in the microwave background:

**Gravity**: photons fall in and escape off gravitational potential wells, characterized by $\Phi$ in the comoving gauge, and as a consequence their frequency is gravitationally blue- or red-shifted, $\delta \nu / \nu = \Phi$. If the gravitational potential is not constant, the photons will escape from a larger or smaller potential well than they fell in, so their frequency is also blue- or red-shifted, a phenomenon known as the Rees-Sciama effect.

**Pressure**: photons scatter off baryons which fall into gravitational potential wells, and radiation pressure creates a restoring force inducing acoustic waves of compression and rarefaction.

**Velocity**: baryons accelerate as they fall into potential wells. They have minimum velocity at maximum compression and rarefaction. That is, their velocity wave is exactly $90^\circ$ off-phase with the acoustic compression waves. These waves induce a Doppler effect on the frequency of the photons.

The temperature anisotropy induced by these three effects is therefore given by

$$
\frac{\delta T}{T}(\mathbf{r}) = \Phi(\mathbf{r}, t_{\text{dec}}) + 2 \int_{t_{\text{dec}}}^{t_{\text{LS}}} \dot{\Phi}(\mathbf{r}, t) dt + \frac{1}{3} \frac{\delta \rho}{\rho}(\mathbf{r}, t_{\text{dec}}) - \frac{\mathbf{r} \cdot \mathbf{v}}{c}. 
$$

(576)

Metric perturbations of different wavelengths enter the horizon at different times. The largest wavelengths, of size comparable to our present horizon, are entering now. There are perturbations with wavelengths comparable to the size of the horizon at the time of last scattering, of projected size about $1^\circ$ in the sky today, which entered precisely at decoupling. And there are perturbations with wavelengths much smaller than the size of the horizon at last scattering, that entered much earlier than decoupling, during the radiation era, which have gone through several acoustic oscillations before last scattering. All these perturbations of different wavelengths leave their imprint in the CMB anisotropies.

The baryons at the time of decoupling do not feel the gravitational attraction of perturbations with wavelength greater than the size of the horizon at last scattering, because of causality. Perturbations with exactly that wavelength are undergoing their first contraction, or acoustic compression, at decoupling. Those perturbations induce a large peak in the temperature anisotropies power spectrum. Perturbations with wavelengths smaller than these will have gone, after they entered the Hubble scale, through a series of acoustic compressions and rarefactions, which can be seen as secondary peaks in the power spectrum. Since the surface of last scattering is not a sharp discontinuity, but a region of $\Delta z \sim 100$, there will be scales for which photons, travelling from one energy concentration to another, will erase the perturbation on that scale, similarly to what neutrinos or HDM do for structure on small scales. That is the reason why we don’t see all the acoustic oscillations with the same amplitude, but in fact they decay exponentially towards smaller angular scales, an effect known as Silk damping, due to photon diffusion.

### 10.1 The Sachs-Wolfe effect

The anisotropies corresponding to large angular scales are only generated via gravitational red-shift and density perturbations through the Einstein equations, $\delta \rho / \rho = -2\Phi$ (for adiabatic perturbations); we can ignore the Doppler contribution, since the perturbation is non-causal. In that case, the temperature anisotropy in the sky today is given by

$$
\frac{\delta T}{T}(\theta, \phi) = \frac{1}{3} \Phi(\eta_{\text{LS}}, \eta_0, \theta, \phi) + 2 \int_{\eta_{\text{LS}}}^{\eta_0} dr \Phi'(\eta_0 - r) Q(r, \theta, \phi),
$$

(577)

where $\eta_0$ is the coordinate distance to the surface of last scattering, i.e. the present conformal time, while $\eta_{\text{LS}} \simeq 0$ determines its comoving hypersurface. The Sachs-Wolfe effect (577) contains two parts, the intrinsic and the Integrated Sachs-Wolfe (ISW) effect, due to the integration along the line of sight of time variations in the gravitational potential.
In linear perturbation theory, the scalar metric perturbations can be separated into \( \Phi(\eta, \mathbf{x}) \equiv \Phi(\eta) Q(\mathbf{x}) \), where \( Q(\mathbf{x}) \) are the scalar harmonics, eigenfunctions of the Laplacian in three dimensions,

\[
\nabla^2 Q_{klm}(r, \theta, \phi) = -k^2 Q_{klm}(r, \theta, \phi).
\]

These functions have the general form

\[
Q_{klm}(r, \theta, \phi) = \Pi_{kl}(r) Y_{lm}(\theta, \phi),
\]

where \( Y_{lm}(\theta, \phi) \) are the usual spherical harmonics, and the radial parts can be written (in a flat Universe) in terms of spherical Bessel functions, \( \Pi_{kl}(r) = \sqrt{2 \pi} k j_l(kr) \). On the other hand, the time evolution of the metric perturbation during the matter era is given by

\[
\Phi'' + 3H \Phi' + a^2 \Lambda \Phi - 2K \Phi = 0.
\]

In the case of a flat universe \( (K = 0) \) without cosmological constant, the Newtonian potential \( \Phi \) remains constant during the matter era and only the intrinsic SW effect contributes to \( \delta T/T \). In case of a non-vanishing \( \Lambda \), since its contribution is negligible in the past, most of the photon's trajectory towards us is unperturbed. We will consider here the approximation \( \Phi \simeq \text{constant} \) during the matter era.

The growing mode solution of the metric perturbation that left the Hubble scale during inflation contributes to the temperature anisotropies on large scales as

\[
\frac{\delta T}{T}(\theta, \phi) = \frac{1}{3} \Phi(\eta_{LS}) Q = \frac{1}{5} \mathcal{R} Q(\eta_0, \theta, \phi) \equiv \sum_{l=2}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta, \phi),
\]

where we have used the fact that, at horizon reentry during the matter era, the gauge-invariant Newtonian potential \( \Phi = \frac{2}{5} \mathcal{R} \) is related to the curvature perturbation \( \mathcal{R} \) at Hubble-crossing during inflation.

We can now compute the two-point correlation function or angular power spectrum, \( C(\theta) \), of the CMB anisotropies on large scales, defined as an expansion in multipole number,

\[
C(\theta) = \left\langle \frac{\delta T^*}{T}(\mathbf{n}) \frac{\delta T}{T}(\mathbf{n}') \right\rangle_{\mathbf{n} \cdot \mathbf{n}' = \cos \theta} = \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l + 1) C_l P_l(\cos \theta),
\]

where \( P_l(z) \) are the Legendre polynomials, and we have averaged over different universe realizations. Since the coefficients \( a_{lm} \) are isotropic (to first order), we can compute the \( C_l = \langle |a_{lm}|^2 \rangle \) as

\[
C_l^{(S)} = \frac{4\pi}{25} \int_0^\infty \frac{dk}{k} \mathcal{P}_R(k) j_l^2(k\eta_0).
\]

In the case of scalar metric perturbation produced during inflation, the scalar power spectrum at reentry is given by \( \mathcal{P}_R(k) = A_S^2(k\eta_0)^{n-1} \), in the power-law approximation. In that case, one can integrate (582) to give

\[
C_l^{(S)} = \frac{2\pi}{25} A_S^2 \frac{\Gamma\left(\frac{3}{2}\right) \Gamma[1 - \frac{n-1}{2}] \Gamma[l + \frac{n-1}{2}]}{\Gamma\left[\frac{3}{2} - \frac{n-1}{2}\right] \Gamma[l + 2 - \frac{n-1}{2}]},
\]

\[
\frac{l(l+1) C_l^{(S)}}{2\pi} = \frac{A_S^2}{25} = \text{constant}, \quad \text{for} \quad n = 1.
\]

This last expression corresponds to what is known as the Sachs-Wolfe plateau, and is the reason why the coefficients \( C_l \) are always plotted multiplied by \( l(l+1) \).
10.2 The tensor perturbations Sachs-Wolfe effect

Tensor metric perturbations also contribute with an approximately constant angular power spectrum, \(l(l+1)C_l\). The Sachs-Wolfe effect for a gauge-invariant tensor perturbation is given by

\[
\frac{\delta T}{T}(\theta, \phi) = \int_{\eta_{LS}}^{\eta_0} dr \ h'(\eta_0 - r) \ Q_{rr}(r, \theta, \phi),
\]

where \(Q_{rr}\) is the \(rr\)-component of the tensor harmonic along the line of sight. The tensor perturbation \(h_k(\eta)\) during the matter era satisfies

\[
h_k'' + 2\mathcal{H} h_k' + (k^2 + 2K) h_k = 0,
\]

which depends on the wavenumber \(k\), contrary to what happens with the scalar modes, see Eq. (579). For a flat \((K = 0)\) universe, the solution to this equation is \(h_k(\eta) = h G_k(\eta)\), where \(h\) is the constant tensor metric perturbation at horizon crossing and \(G_k(\eta) = 3 j_1(k\eta)/k\eta\), normalized so that \(G_k(0) = 1\) at the surface of last scattering. The radial part of the tensor harmonic \(Q_{rr}\) in a flat universe can be written as

\[
Q_{kl}^r(r) = \left[\frac{(l-1)(l+1)(l+2)}{\pi k^2}\right]^{1/2} \frac{j_l(kr)}{r^2}.
\]

The tensor angular power spectrum can finally be expressed as

\[
C_l^{(T)} = \frac{9\pi}{4} (l-1)(l+1)(l+2) \int_0^\infty \frac{dk}{k} \mathcal{P}_g(k) I_{kl}^2,
\]

where \(x \equiv k\eta\), and \(\mathcal{P}_g(k)\) is the primordial tensor spectrum. For a scale invariant spectrum, \(n_T = 0\), we can integrate (588) to give

\[
l(l+1) C_l^{(T)} = \frac{\pi}{36} \left(1 + \frac{48\pi^2}{385}\right) A_T^2 B_l,
\]

with \(B_l = (1.1184, 0.8789, \ldots, 1.00)\) for \(l = 2, 3, \ldots, 30\). Therefore, \(l(l+1) C_l^{(T)}\) also becomes constant for large \(l\). Beyond \(l \sim 30\), the Sachs-Wolfe expression is not a good approximation and the tensor angular power spectrum decays very quickly at large \(l\).

10.3 The consistency condition

In spite of the success of inflation in predicting a homogeneous and isotropic background on which to imprint a scale-invariant spectrum of inhomogeneities, it is difficult to test the idea of inflation. Before the 1980s anyone would have argued that ad hoc initial conditions could have been at the origin of the homogeneity and flatness of the universe on large scales, while most cosmologists would have agreed with Harrison and Zel’dovich that the most natural spectrum needed to explain the formation of structure was a scale-invariant spectrum. The surprise was that inflation incorporated an understanding of both the globally homogeneous and spatially flat background, and the approximately scale-invariant spectrum of perturbations in the same formalism. But that could have been a coincidence.

What is unique to inflation is the fact that inflation determines not just one but two primordial spectra, corresponding to the scalar (density) and tensor (gravitational waves) metric perturbations, from a single continuous function, the inflaton potential \(V(\phi)\). In the slow-roll approximation, one determines, from \(V(\phi)\), two continuous functions, \(\mathcal{P}_R(k)\) and \(\mathcal{P}_g(k)\), that in the power-law approximation reduces to two amplitudes, \(A_S\) and \(A_T\), and two tilts, \(n\) and \(n_T\). It is clear that there must be a relation between
the four parameters. Indeed, one can see from Eqs. (590) and (584) that the ratio of the tensor to scalar contribution to the angular power spectrum is proportional to the tensor tilt,

$$ R \equiv \frac{C_l^{(T)}}{C_l^{(S)}} = \frac{25}{9} \left( 1 + \frac{48\pi^2}{385} \right) 2\epsilon \simeq -2\pi n_T . $$

(591)

This is a unique prediction of inflation, which could not have been postulated a priori. If we finally observe a tensor spectrum of anisotropies in the CMB, or a stochastic gravitational wave background in laser interferometers like LIGO or VIRGO, with sufficient accuracy to determine their spectral tilt, one might have some chance to test the idea of inflation, via the consistency relation (591).

For the moment, observations of the microwave background anisotropies suggest that the Sachs-Wolfe plateau exists, but it is still premature to determine the tensor contribution. Perhaps in the near future, from the analysis of polarization as well as temperature anisotropies, with the CMB satellites MAP and Planck, we might have a chance of determining the validity of the consistency relation.

Assuming that the scalar contribution dominates over the tensor on large scales, i.e. $R \ll 1$, one can actually give a measure of the amplitude of the scalar metric perturbation from the observations of the Sachs-Wolfe plateau in the angular power spectrum,

$$ \left[ \frac{l(l+1)C_l^{(S)}}{2\pi} \right]^{1/2} = \frac{A_S}{5} = (1.03 \pm 0.07) \times 10^{-5} , $$

(592)

$$ n = 0.96 \pm 0.03 . $$

(593)

These measurements can be used to normalize the primordial spectrum and determine the parameters of the model of inflation. In the near future these parameters will be determined with much better accuracy.

10.4 The acoustic peaks

Before decoupling, the photons and the baryons are tightly coupled via Thomson scattering. The dynamics of the photon-baryon fluid is described by a forced and damped harmonic oscillator equation for the baryon density contrast,

$$ \delta''_k + \mathcal{H} \frac{R}{1+R} \delta'_k + k^2 c_s^2 \delta_k = F(\Phi_k) , $$

(594)

where $R = 3\rho_B/4\rho_\gamma$ is the baryon-to-photon ratio, $c_s^2 = c^2/3(1 + R)$ is the sound speed of the plasma, and $F(\Phi_k)$ is the external force due to the gravitational effect of dark matter and neutrinos. Baryons tend to collapse due to self-gravitation, while radiation pressure provides the restoring force, setting up acoustic oscillations in the plasma. Because of tight coupling, $\delta_k = 3\Theta_0(k, \eta)$, and the baryon oscillations give rise to oscillations in the temperature fluctuations $\Theta_0$. The higher the baryon fraction $R$, the higher the amplitude of the oscillations. The external gravitational force displaces the zero-point of oscillations, which makes higher the amplitude of compressions versus rarefactions.

At decoupling there is a freeze out of the oscillations. The microwave background is like a snapshot of the instant of last scattering, where each mode $k$ is at a different stage of oscillation,

$$ \Theta_0(k, \eta) \propto (1 + R)^{-1/4} \begin{cases} \zeta_k(0) \cos kr_s & \text{adiabatic}, \\ S_k(0) k \sin kr_s & \text{isocurvature} \end{cases} , $$

(595)

where $r_s = \int_0^{\eta_{\text{dec}}} c_s d\eta$ is the sound horizon at decoupling. These fluctuations induce acoustic peaks in the Angular Power Spectrum that correspond to maxima and minima of oscillations. For adiabatic and isocurvature perturbations, the harmonic peaks appear at wavenumber $k_n^{(A)} = n\pi/c_s\eta_{\text{dec}}$.
and $k_n^{(I)} = (n + 1/2)\pi/c_s\eta_{\text{dec}}$, respectively. In particular, the angle subtended by the sound horizon at decoupling, $\theta_s = r_s/d_A$, corresponds to a multipole number (e.g. for adiabatic perturbations)

$$l_n \approx \frac{n\pi}{\theta_s} = \frac{n\pi}{2c_s} \left(\frac{(1 + z_{\text{dec}})\Omega_M}{|\Omega_K|}\right)^{1/2} \sin n \int_0^{z_{\text{dec}}} \frac{|\Omega_K|^{1/2}dz}{[\Omega_A + \Omega_M(1 + z)^3 + \Omega_K(1 + z)^2]^{1/2}}$$

### 10.4.1 The new microwave anisotropy satellites, WMAP and Planck

The large amount of information encoded in the anisotropies of the microwave background is the reason why both NASA and the European Space Agency have decided to launch two independent satellites to measure the CMB temperature and polarization anisotropies to unprecedented accuracy. The Wilkinson Microwave Anisotropy Probe [91] was launched by NASA at the end of 2000, and has fulfilled most of our expectation, while Planck [92] is expected to be launched by ESA in 2007. There are at the moment other large proposals like CMB Pol [98], ACT [99], etc. which will see the light in the next few years, see Ref. [93].

As we have emphasized before, the fact that these anisotropies have such a small amplitude allow for an accurate calculation of the predicted anisotropies in linear perturbation theory. A particular cosmological model is characterized by a dozen or so parameters: the rate of expansion, the spatial curvature, the baryon content, the cold dark matter and neutrino contribution, the cosmological constant (vacuum energy), the reionization parameter (optical depth to the last scattering surface), and various primordial spectrum parameters like the amplitude and tilt of the adiabatic and isocurvature spectra, the amount of gravitational waves, non-Gaussian effects, etc. All these parameters can now be fed into very fast CMB codes called CMBFAST [96] and CAMB [97], that compute the predicted temperature and polarization anisotropies to better than 1% accuracy, and thus can be used to compare with observations.

These two satellites will improve both the sensitivity, down to $\mu$K, and the resolution, down to arc minutes, with respect to the previous COBE satellite, thanks to large numbers of microwave horns of various sizes, positioned at specific angles, and also thanks to recent advances in detector technology, with high electron mobility transistor amplifiers (HEMTs) for frequencies below 100 GHz and bolometers for higher frequencies. The primary advantage of HEMTs is their ease of use and speed, with a typical sensitivity of 0.5 mKs$^{1/2}$, while the advantage of bolometers is their tremendous sensitivity, better than 0.1 mKs$^{1/2}$, see Ref. [100]. This will allow cosmologists to extract information from around 3000 multipoles! Since most of the cosmological parameters have specific signatures in the height and position of the first few acoustic peaks, the higher the resolution, the more peaks one is expected to see, and thus the better the accuracy with which one will be able to measure those parameters, see Table 2.

Although the satellite probes were designed for the accurate measurement of the CMB temperature anisotropies, there are other experiments, like balloon-borne and ground interferometers [93]. Probably the most important objective of the future satellites (beyond WMAP) will be the measurement of the CMB polarization anisotropies, discovered by DASI in November 2002 [101], and confirmed a few months later by WMAP with greater accuracy [20], see Fig. 29. These anisotropies were predicted by models of structure formation and indeed found at the level of microKelvin sensitivities, where the new satellites were aiming at. The complementary information contained in the polarization anisotropies already provides much more stringent constraints on the cosmological parameters than from the temperature anisotropies alone. However, in the future, Planck and CMB pol will have much better sensitivities. In particular, the curl-curl component of the polarization power spectra is nowadays the only means we have to determine the tensor (gravitational wave) contribution to the metric perturbations responsible for temperature anisotropies, see Fig. 34. If such a component is found, one could constraint very precisely the model of inflation from its spectral properties, specially the tilt [94].
10.5 From metric perturbations to large scale structure

If inflation is responsible for the metric perturbations that gave rise to the temperature anisotropies observed in the microwave background, then the primordial spectrum of density inhomogeneities induced by the same metric perturbations should also be responsible for the present large scale structure [102]. This simple connection allows for more stringent tests on the inflationary paradigm for the generation of metric perturbations, since it relates the large scales (of order the present horizon) with the smallest scales (on galaxy scales). This provides a very large lever arm for the determination of primordial spectra parameters like the tilt, the nature of the perturbations, whether adiabatic or isocurvature, the geometry of the universe, as well as its matter and energy content, whether CDM, HDM or mixed CHDM.

10.5.1 The galaxy power spectrum

As metric perturbations enter the causal horizon during the radiation or matter era, they create density fluctuations via gravitational attraction of the potential wells. The density contrast $\delta$ can be deduced from the Einstein equations in linear perturbation theory, see Eq. (397),

$$\delta_k \equiv \frac{\delta\rho_k}{\rho} = \left(\frac{k}{aH}\right)^2 \frac{2}{3} \Phi_k = \left(\frac{k}{aH}\right)^2 \frac{2 + 2\omega}{5 + 3\omega} R_k,$$

(596)

where we have assumed $K = 0$, and used Eq. (483). From this expression one can compute the power spectrum, at horizon crossing, of matter density perturbations induced by inflation, see Eq. (495),

$$P(k) = \langle |\delta_k|^2 \rangle = A \left(\frac{k}{aH}\right)^n,$$

(597)

with $n$ given by the scalar tilt (497), $n = 1 + 2\eta - 6\epsilon$. This spectrum reduces to a Harrison-Zel’dovich spectrum (168) in the slow-roll approximation: $\eta, \epsilon \ll 1$.

Since perturbations evolve after entering the horizon, the power spectrum will not remain constant. For scales entering the horizon well after matter domination ($k^{-1} \gg k_{eq}^{-1} \approx 81$ Mpc), the metric perturbation has not changed significantly, so that $R_k(\text{final}) = R_k(\text{initial})$. Then Eq. (596) determines the final density contrast in terms of the initial one. On smaller scales, there is a linear transfer function $T(k)$, which may be defined as [79]

$$R_k(\text{final}) = T(k) R_k(\text{initial}).$$

(598)
To calculate the transfer function one has to specify the initial condition with the relative abundance of photons, neutrinos, baryons and cold dark matter long before horizon crossing. The most natural condition is that the abundances of all particle species are uniform on comoving hypersurfaces (with constant total energy density). This is called the adiabatic condition, because entropy is conserved independently for each particle species \( X \), i.e. \( \frac{\delta \rho_X}{\rho_X + p_X} = \frac{\delta \rho_Y}{\rho_Y + p_Y} \), given a perturbation in time from a comoving hypersurface, so

\[
\frac{\delta \rho_X}{\rho_X + p_X} = \frac{\delta \rho_Y}{\rho_Y + p_Y}, \tag{599}
\]

where we have used the energy conservation equation for each species, \( \dot{\rho}_X = -3H(\rho_X + p_X) \), valid to first order in perturbations. It follows that each species of radiation has a common density contrast \( \delta_r \), and each species of matter has also a common density contrast \( \delta_m \), with the relation \( \delta_m = \frac{3}{4} \delta_r \).

Given the adiabatic condition, the transfer function is determined by the physical processes occurring between horizon entry and matter domination. If the radiation behaves like a perfect fluid, its density perturbation oscillates during this era, with decreasing amplitude. The matter density contrast living in this background does not grow appreciably before matter domination because it has negligible self-gravity. The transfer function is therefore given roughly by, see Eq. (171),

\[
T(k) = \begin{cases} 
1, & k \ll k_{\text{eq}} \\
(k/k_{\text{eq}})^2, & k \gg k_{\text{eq}} 
\end{cases} \tag{600}
\]

The perfect fluid description of the radiation is far from being correct after horizon entry, because roughly half of the radiation consists of neutrinos whose perturbation rapidly disappears through free streaming. The photons are also not a perfect fluid because they diffuse significantly, for scales below the Silk scale, \( k_{\text{S}}^{-1} \sim 1 \text{ Mpc} \). One might then consider the opposite assumption, that the radiation has zero perturbation after horizon entry. Then the matter density perturbation evolves according to

\[
\ddot{\delta}_k + 2H\dot{\delta}_k + (c_s^2 k_{\text{ph}}^2 - 4\pi G \rho) \delta_k = 0, \tag{601}
\]

which corresponds to the equation of a damped harmonic oscillator. The zero-frequency oscillator defines the Jeans wavenumber, \( k_J = \sqrt{4\pi G \rho / c_s^2} \). For \( k \ll k_J \), \( \delta_k \) grows exponentially on the dynamical timescale, \( \tau_{\text{dyn}} = \text{Im} \omega^{-1} = (4\pi G \rho)^{-1/2} = \tau_{\text{grav}} \), which is the time scale for gravitational collapse. One can also define the Jeans length,

\[
\lambda_J = \frac{2\pi}{k_J} = c_s \sqrt{\frac{\pi}{G \rho}}, \tag{602}
\]

which separates gravitationally stable from unstable modes. If we define the pressure response timescale as the size of the perturbation over the sound speed, \( \tau_{\text{pres}} \sim \lambda/c_s \), then, if \( \tau_{\text{pres}} > \tau_{\text{grav}} \), gravitational collapse of a perturbation can occur before pressure forces can respond to restore hydrostatic equilibrium (this occurs for \( \lambda > \lambda_J \)). On the other hand, if \( \tau_{\text{pres}} < \tau_{\text{grav}} \), radiation pressure prevents gravitational collapse and there are damped acoustic oscillations (for \( \lambda < \lambda_J \)).

We will consider now the behaviour of modes within the horizon during the transition from the radiation (\( c_s^2 = 1/3 \)) to the matter era (\( c_s^2 = 0 \)). The growing mode solution increases only by a factor of 2 between horizon entry and the epoch when matter starts to dominate, i.e. \( y = 1 \). The transfer function is therefore again roughly given by Eq. (600). Since the radiation consists roughly half of neutrinos, which free stream, and half of photons, which either form a perfect fluid or just diffuse, neither the perfect fluid nor the free-streaming approximation looks very sensible. A more precise calculation is needed, including: neutrino free streaming around the epoch of horizon entry; the diffusion of photons around the same time, for scales below Silk scale; the diffusion of baryons along with the photons, and the establishment after matter domination of a common matter density contrast, as the baryons fall into the potential wells of cold dark matter. All these effects apply separately, to first order in the perturbations, to
each Fourier component, so that a linear transfer function is produced. There are several parametrizations
in the literature, but the one which is more widely used is that of Ref. [103],

\[ T(k) = \left[ 1 + \left( ak + (bk)^{3/2} + (ck)^2 \right)^\nu \right]^{-1/\nu}, \quad \nu = 1.13, \quad (603) \]

\[ a = 6.4 (\Omega_M h)^{-1} h^{-1} \text{ Mpc}, \quad (604) \]

\[ b = 3.0 (\Omega_M h)^{-1} h^{-1} \text{ Mpc}, \quad (605) \]

\[ c = 1.7 (\Omega_M h)^{-1} h^{-1} \text{ Mpc}. \quad (606) \]

We see that the behaviour estimated in Eq. (600) is roughly correct, although the break at \( k = k_{\text{eq}} \) is not at all sharp, see Fig. 35. The transfer function, which encodes the solution to linear equations, ceases to be valid when the density contrast becomes of order 1. After that, the highly nonlinear phenomenon of gravitational collapse takes place, see Fig. 35.

![Fig. 35: The CDM power spectrum \( P(k) \) as a function of wavenumber \( k \), in logarithmic scale, normalized to the local abundance of galaxy clusters, for an Einstein-de Sitter universe with \( h = 0.5 \). The solid (dashed) curve shows the linear (non-linear) power spectrum. While the linear power spectrum falls off like \( k^{-3} \), the non-linear power spectrum illustrates the increased power on small scales due to non-linear effects, at the expense of the large-scale structures. From Ref. [44].](image)

10.5.2 The new redshift catalogs, 2dF and Sloan Digital Sky Survey

Our view of the large-scale distribution of luminous objects in the universe has changed dramatically during the last 25 years: from the simple pre-1975 picture of a distribution of field and cluster galaxies, to the discovery of the first single superstructures and voids, to the most recent results showing an almost regular web-like network of interconnected clusters, filaments and walls, separating huge nearly empty volumes. The increased efficiency of redshift surveys, made possible by the development of spectrographs and – specially in the last decade – by an enormous increase in multiplexing gain (i.e. the ability to collect spectra of several galaxies at once, thanks to fibre-optic spectrographs), has allowed us not only to do cartography of the nearby universe, but also to statistically characterize some of its properties, see Ref. [104]. At the same time, advances in theoretical modeling of the development of structure, with large high-resolution gravitational simulations coupled to a deeper yet limited understanding of how to form galaxies within the dark matter halos, have provided a more realistic connection of the models to the observable quantities [105]. Despite the large uncertainties that still exist, this has transformed the study of cosmology and large-scale structure into a truly quantitative science, where theory and observations can progress side by side.
I will concentrate on two of the new catalogs, which are taking data at the moment and which have changed the field, the 2-degree-Field (2dF) Catalog and the Sloan Digital Sky Survey (SDSS). The advantages of multi-object fibre spectroscopy have been pushed to the extreme with the construction of the 2dF spectrograph for the prime focus of the Anglo-Australian Telescope [45]. This instrument is able to accommodate 400 automatically positioned fibres over a 2 degree in diameter field. This implies a density of fibres on the sky of approximately $130 \text{ deg}^{-2}$, and an optimal match to the galaxy counts for a magnitude $b_J \simeq 19.5$, similar to that of previous surveys like the ESP, with the difference that with such an area yield, the same number of redshifts as in the ESP survey can be collected in about 10 exposures, or slightly more than one night of telescope time with typical 1 hour exposures. This is the basis of the 2dF galaxy redshift survey. Its goal is to measure redshifts for more than 250,000 galaxies with $b_J < 19.5$. In addition, a faint redshift survey of 10,000 galaxies brighter than $R = 21$ will be done over selected fields within the two main strips of the South and North Galactic Caps. The survey has now finished, with a quarter of a million redshifts. The final result can be seen in Ref. [45].

The most ambitious and comprehensive galaxy survey currently in progress is without any doubt the Sloan Digital Sky Survey [46]. The aim of the project is, first of all, to observe photometrically the whole Northern Galactic Cap, $30^\circ$ away from the galactic plane (about $10^4 \text{ deg}^2$) in five bands, at limiting magnitudes from 20.8 to 23.3. The expectation is to detect around 50 million galaxies and around $10^8$ star-like sources. This has already led to the discovery of several high-redshift ($z > 4$) quasars, including the highest-redshift quasar known, at $z = 5.0$, see Ref. [46]. Using two fibre spectrographs carrying 320 fibres each, the spectroscopic part of the survey will then collect spectra from about $10^6$ galaxies with $r' < 18$ and $10^5$ AGNs with $r' < 19$. It will also select a sample of about $10^5$ red luminous galaxies with $r' < 19.5$, which will be observed spectroscopically, providing a nearly volume-limited sample of early-type galaxies with a median redshift of $z \simeq 0.5$, that will be extremely valuable to study the evolution of clustering. The data that is coming from these catalogs is so outstanding that already cosmologists are using them for the determination of the cosmological parameters of the standard model of cosmology. The main outcome of these catalogs is the linear power spectrum of matter fluctuations that give rise to galaxies, and clusters of galaxies. It covers from the large scales of order Gigaparsecs, the realm of the unvirialised superclusters, to the small scales of hundreds of kiloparsecs, where the Lyman-\alpha systems can help reconstruct the linear power spectrum, since they are less sensitive to the nonlinear growth of perturbations.

As often happens in particle physics, not always are observations from a single experiment sufficient to isolate and determine the precise value of the parameters of the standard model. We mentioned in the previous Section that some of the cosmological parameters created similar effects in the temperature anisotropies of the microwave background. We say that these parameters are degenerate with respect to the observations. However, often one finds combinations of various experiments/observations which break the degeneracy, for example by depending on a different combination of parameters. This is precisely the case with the cosmological parameters, as measured by a combination of large-scale structure observations, microwave background anisotropies, Supernovae Ia observations and Hubble Space Telescope measurements. It is expected that in the near future we will be able to determine the parameters of the standard cosmological model with great precision from a combination of several different experiments.

OBSERVATIONAL SIGNATURES OF INFLATION

11 Inflationary model building. The particle physics connection

For the moment, observations of the microwave background anisotropies suggest that the Sachs-Wolfe plateau exists, but it is still premature to determine the tensor contribution. Perhaps in the near future, from the analysis of polarization as well as temperature anisotropies, with the CMB satellites MAP and Planck, we might have a chance of determining the validity of the consistency relation. Assuming that the scalar contribution dominates over the tensor on large scales, i.e. $r \ll 1$, one can actually give a
measure of the amplitude of the scalar metric perturbation from the observations of the Sachs-Wolfe plateau in the angular power spectrum,

\[ \left( \frac{l(l+1)C_l^{(S)}}{2\pi} \right)^{1/2} = \frac{A_S}{\sqrt{5}} = (1.03 \pm 0.07) \times 10^{-5} , \]

\[ n = 0.98 \pm 0.03 . \]

These measurements can be used to normalize the primordial spectrum and determine the parameters of a particular model of inflation. In the near future these parameters will be determined with much better accuracy, to less than a percent.

In the next sections we will consider specific models of inflation. The formulae we will be using are

\[ \epsilon = \frac{1}{2\kappa^2} \left( \frac{V'}{V} \right)^2 , \quad \eta = \frac{1}{\kappa^2} \left( \frac{V''}{V} \right) , \quad \xi = \frac{1}{\kappa^2} \left( \frac{V'V'''}{V^2} \right) \]

\[ N = \int_{\phi_{end}}^{\phi} \frac{\kappa d\phi}{\sqrt{2\epsilon}} \]

together with the formula for the amplitude and tilt of scalar and tensor anisotropies

\[ A_S = \frac{\kappa}{\sqrt{2\epsilon}} \frac{H_2}{2\pi} , \quad n = 1 + 2\eta - 6\epsilon \]

\[ A_T = \frac{\sqrt{2}}{\pi} \kappa H , \quad n_T = -2\epsilon , \quad r = -2\pi n_T \]

### 11.1 Power-law inflation

\[ V(\phi) = V_0 e^{-\beta\kappa \phi} \quad \beta \ll 1 \quad \text{for inflation} \]

\[ 3H^2(\phi) = \frac{2}{\kappa^2} \left( \frac{\partial H}{\partial \phi} \right)^2 + \kappa^2 V(\phi) \]

\[ H(\phi) = H_0 e^{-\frac{1}{2}\beta\kappa \phi} \quad \Rightarrow \quad \frac{1}{H} \frac{\partial H}{\partial \phi} = -\frac{1}{2}\beta = \text{const} \]

\[ H_0^2 = \frac{\kappa^2}{3} V_0 \left( 1 - \frac{\beta^2}{6} \right)^{-1} \quad \text{where} \quad V_0 \equiv M^4 \]

\[ \epsilon = \frac{2}{\kappa^2} \left( \frac{H'}{H} \right)^2 = \frac{1}{2} \beta^2 < 1 \]

\[ \delta = \frac{2}{\kappa^2} \left( \frac{H''}{H} \right)^2 = \frac{1}{2} \beta^2 < 1 \]

\[ \epsilon = -\frac{\dot{H}}{H^2} = \frac{1}{2} \beta^2 \quad \Rightarrow \quad a \propto t^p \quad \epsilon = 1/p \quad p = \frac{2}{\beta^2} \]

\[ \epsilon = \delta = \frac{1}{p} = \text{const} \]
\[ N = \int_{\phi}^{\phi_{\text{end}}} \frac{\kappa d\phi}{\sqrt{2\epsilon}} = \frac{\kappa}{\beta} (\phi_{\text{end}} - \phi) = 65 \quad (618) \]

\[ A_S = \frac{\kappa}{\sqrt{2\epsilon}} \frac{H}{2\pi} = 5 \times 10^{-5} \quad \Rightarrow \quad M \simeq 10^{-3} M_P \quad (619) \]

\[ n - 1 = 2 \left( \frac{\delta - 2\epsilon}{1 - \epsilon} \right) = - \frac{2}{p - 1} \quad (620) \]

\[ |n - 1| < 0.05 \quad \Rightarrow \quad p > 41 \quad (621) \]

\[ n_T = - \frac{2\epsilon}{1 - \epsilon} = - \frac{2}{p - 1}, \quad r = -2\pi n_T < \frac{\pi}{10} = 0.314 \quad (622) \]

### 11.2 Chaotic inflation \((m^2 \phi^2)\)

\[ V(\phi) = \frac{1}{2} m^2 \phi^2 \quad \Rightarrow \quad H^2 \simeq \frac{\kappa^2}{6} m^2 \phi^2 \quad (623) \]

\[ \epsilon = \frac{1}{2\kappa^2} \left( \frac{V'}{V} \right)^2 = \frac{2}{\kappa^2 \phi^2} = 1 \quad \Rightarrow \quad \phi_{\text{end}} = \frac{M_P}{2\sqrt{\pi}} \simeq \frac{M_P}{3.5} \quad (624) \]

\[ \eta = \frac{1}{\kappa^2} \left( \frac{V''}{V} \right) = \frac{2}{\kappa^2 \phi^2} = \epsilon = \frac{1}{2N} \quad (625) \]

\[ N = \int_{\phi_{\text{end}}}^{\phi} \frac{\kappa d\phi}{\sqrt{2\epsilon}} = \left( \frac{\kappa \phi}{2} \right)^2 \bigg|_{\phi_{\text{end}}} \approx \frac{\kappa^2 \phi^2}{4} \quad \Rightarrow \quad \phi_{65} = 6.4 M_P \quad (626) \]

\[ A_S = \frac{\kappa m}{\sqrt{6} \frac{\kappa^2 \phi^2}{4\pi}} = N \sqrt{\frac{4}{3\pi} \frac{m}{M_P}} = 5 \times 10^{-5} \quad \Rightarrow \quad (627) \]

\[ m = 1.2 \times 10^{-6} M_P = 1.4 \times 10^{13} \text{GeV} \quad (628) \]

\[ n = 1 + 2\eta - 6\epsilon = 1 - \frac{2}{N} \approx 0.97 \quad (629) \]

\[ A_T = \frac{4}{\sqrt{\pi}} \frac{H}{M_P} < 10^{-5} \quad (630) \]

\[ n_T = -2\epsilon = - \frac{1}{N} \simeq -0.016 \quad (631) \]

\[ r = \frac{C_f^T}{C_f^P} = \frac{25}{9} \left( 1 + \frac{48\pi^2}{385} \right) 2\epsilon \simeq -2\pi n_T \approx \frac{2\pi}{N} \simeq 0.1 \quad (632) \]
11.3 Chaotic inflation \((\lambda \phi^4)\)

\[
V(\phi) = \frac{1}{4} \lambda \phi^4 \quad \Rightarrow \quad H^2 \simeq \frac{\kappa^2}{12} \lambda \phi^4
\]  

\[
\epsilon = \frac{1}{2\kappa^2} \left( \frac{V'}{V} \right)^2 = \frac{8}{\kappa^2 \phi^2} = 1 \quad \Rightarrow \quad \phi_{\text{end}} = \frac{M_P}{\sqrt{\pi}} \simeq \frac{M_P}{1.8}
\]  

\[
\eta = \frac{1}{\kappa^2} \left( \frac{V''}{V} \right) = \frac{12\epsilon}{2} - \frac{3}{2N}
\]  

\[
N = \int_{\phi_{\text{end}}}^{\phi} \frac{\kappa d\phi}{\sqrt{2\epsilon}} = \left( \frac{\kappa \phi}{8} \right)^2 \bigg|_{\phi_{\text{end}}} \approx \frac{\kappa^2 \phi^2}{8} \quad \Rightarrow \quad \phi_{65} = 4.5 M_P
\]

\[
A_S = \sqrt{\frac{\lambda}{3}} \frac{\kappa^3 \phi^3}{16\pi} = \sqrt{\frac{\lambda}{3}} \frac{(8N)^{3/2}}{16\pi} = 5 \times 10^{-5} \quad \Rightarrow
\]

\[
\lambda = 1.3 \times 10^{-13}
\]

\[
n = 1 + 2\eta - 6\epsilon = 1 - \frac{3}{N} \approx 0.95
\]

\[
A_T = \frac{4}{\sqrt{\pi}} \frac{H}{M_P} < 10^{-5}
\]

\[
n_T = -2\epsilon = -2 \frac{N}{N} \approx -0.03
\]

\[
r = \frac{C_T}{C_{\text{Sun}}} = \frac{25}{9} \left( 1 + \frac{48\pi^2}{385} \right) 2\epsilon \simeq -2n T = \frac{4\pi}{N} \simeq 0.2
\]

11.4 New inflation

\[
V(\phi) = \frac{\lambda}{4} (\phi^2 - v^2)^2 \quad \Rightarrow \quad H^2 \simeq \frac{\kappa^2}{12} \lambda (\phi^2 - v^2)^2
\]

complete here!

11.5 Natural inflation

\[
V(\phi) = M^2 f^2 \left( 1 - \cos \frac{\phi}{f} \right) = 2M^2 f^2 \sin^2 \frac{\phi}{2f}
\]

\[
= \frac{1}{2} M^2 \phi^2 - \frac{1}{4!} \frac{M^2}{f^2} \phi^4 + O(\phi^6)
\]

\[
\epsilon = \frac{1}{2\kappa^2 f^2} \left( \frac{\sin \frac{\phi}{f}}{1 - \cos \frac{\phi}{f}} \right)^2 = \cot^2 \frac{\phi}{2f} \ll 1
\]

\[
\eta = \frac{1}{\kappa^2 f^2} \cos \frac{\phi}{f} \ll 1
\]

\[
N = 2\kappa^2 f^2 \int_{x_{\text{end}}}^{x} dx \tan x = -2\kappa^2 f^2 \ln \cos \frac{\phi}{2f} \bigg|_{\phi_{\text{end}}}
\]
\[
\epsilon = 1 \quad \Rightarrow \quad \cos \frac{\phi_{\text{end}}}{2f} = \left( \frac{2\kappa^2 f^2}{1 + 2\kappa^2 f^2} \right)^{1/2} < 1 \quad (649)
\]

\[
\cos \frac{\phi}{2f} = \left( \frac{2\kappa^2 f^2}{1 + 2\kappa^2 f^2} \right)^{1/2} e^{-\frac{N}{2\kappa^2 f^2}} \quad (650)
\]

\[
\epsilon_{65} = \frac{1}{2\kappa^2 f^2} \left( e^{\frac{N}{2\kappa^2 f^2}} - 1 \right)^{-1} \ll \frac{1}{2\kappa^2 f^2} \quad (651)
\]

\[
\eta_{65} = \epsilon_{65} - \frac{1}{2\kappa^2 f^2} \quad \Rightarrow \quad n \approx 1 - \frac{1}{\kappa^2 f^2} \quad (652)
\]

\[
A_S = \sqrt{\frac{2}{3} \frac{\kappa M}{2\pi} (2\kappa^2 f^2) \sinh \frac{N}{2\kappa^2 f^2}} \quad (653)
\]

If \( f = M_P \) \( \Rightarrow \) \( M = 9 \times 10^{-7} M_P = 10^{13} \text{ GeV} \quad (654) \)

\[ n = 1 - \frac{1}{8\pi} = 0.96 \quad (655) \]

\[
r = -2\pi n_T = \frac{2\pi}{\kappa^2 f^2} \left( e^{\frac{N}{2\kappa^2 f^2}} - 1 \right)^{-1} \approx 0.02 \quad (656)
\]

### 11.6 Starobinski inflation

\[
R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \kappa^2 \langle T_{\mu\nu} \rangle_{\text{ren}} = \frac{1}{6M^2} (1) H_{\mu\nu} + \frac{1}{H_0^2} (3) H_{\mu\nu} , \quad (657)
\]

\[
^{(1)} H_{\mu\nu} = 2(\nabla_\mu \nabla_\nu - g_{\mu\nu} \nabla^2) R + 2R R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R^2 \quad (658)
\]

\[
^{(3)} H_{\mu\nu} = R^\lambda_{\mu\nu} R_{\lambda\rho} - \frac{2}{3} R R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R^\rho\sigma R_{\rho\sigma} + \frac{1}{4} g_{\mu\nu} R^2 \quad (659)
\]

Substituting FRW metric and using the Slow Roll Approximation,

\[
\dot{H} = -\frac{M^2}{6} \left( 1 - \frac{H^2}{H_0^2} \right) . \quad (660)
\]

At first stage: \( H_0^2 \gg M^2 \) \( \Rightarrow \) \( -\dot{H} < M^2/6 \ll H_0^2 \Rightarrow H \approx H_0 = \text{const.} \) However, \( H \) grows and becomes unstable. When \( H \sim M \) inflation ends. Alternatively, one can study the evolution in the effective action formalism, including higher derivatives,

\[
S_g = \int d^4x \sqrt{-g} \frac{1}{2\kappa^2} \left( R - \frac{R^2}{6M^2} \right) \equiv \int d^4x \sqrt{-g} \frac{1}{2\kappa^2} f(R) \quad (661)
\]

which gives rise to Eq. (658). One can then write this action as the usual Einstein-Hilbert action plus a scalar field, making use of the conformal transformation

\[
\tilde{g}_{\mu\nu} = F(R) g_{\mu\nu} \equiv e^{\alpha\phi} g_{\mu\nu} \quad \Rightarrow \quad \sqrt{-\tilde{g}} = e^{2\alpha\phi} \sqrt{-g} \quad (662)
\]

\[
\tilde{R}_{\mu\nu} = R_{\mu\nu} - \frac{\alpha K}{2} (g_{\mu\nu} \nabla^2 \phi + 2\nabla_\mu \nabla_\nu \phi) \quad (663)
\]

\[
\tilde{R} = e^{-\alpha\phi} \left[ R - 3\alpha \nabla^2 \phi + \frac{3}{2} \kappa^2 (\partial \phi)^2 \right] \quad (664)
\]
The scalar field $\phi$ will have canonical kinetic term for $\alpha^2 = 2/3$. From the equations of motion one finds the relationship $F(R) = f'(R)$, and therefore the effective scalar potential becomes

$$V(\phi) = \frac{1}{2\kappa^2} \frac{f(R) - R f'(R)}{(f'(R))^2} = \frac{R^2}{12\kappa^2 M^2} \left(1 - \frac{R}{3M^2}\right)^{-2}$$  \hspace{1cm} (665)

$$V(\phi) = \frac{3M^2}{4\kappa^2} \left(1 - e^{-\alpha\phi}\right)^2 = \frac{1}{2} M^2 \phi^2 \left(1 + \alpha\kappa \phi + \ldots\right)$$  \hspace{1cm} (666)

$$\epsilon = \frac{2\alpha^2}{(e^{\alpha\kappa\phi} - 1)^2} = 1 \quad \Rightarrow \quad \phi_{\text{end}} = \frac{\sqrt{3} M_P}{4\sqrt{\pi}} \ln \left(1 + \frac{2}{\sqrt{3}}\right) \simeq \frac{M_P}{5.33} \quad \Rightarrow \quad H_{\text{end}} = \frac{\sqrt{3} M}{2\sqrt{2 + \sqrt{3}}}$$

$$\eta = \frac{2\alpha^2 (2 - e^{\alpha\kappa\phi})}{(e^{\alpha\kappa\phi} - 1)^2} = 0 \quad \Rightarrow \quad \phi_* = \frac{\sqrt{3} M_P}{4\sqrt{\pi}} \ln 2 \simeq \frac{M_P}{5.90} < \phi_{\text{end}}$$

$$N = \frac{e^{\alpha\kappa\phi} - \alpha\kappa\phi}{2\alpha^2} \bigg|_{\phi_{\text{end}}} \simeq \frac{3}{4} e^{\alpha\kappa\phi} \quad \Rightarrow \quad \phi_{65} = 1.09 M_P$$

$$\epsilon_{65} \simeq \frac{1}{2\alpha^2 N^2} \quad \eta_{65} \simeq -\frac{1}{N}$$

$$\alpha\kappa \phi_{65} = 4.46 \gg 1 \quad \Rightarrow \quad V(\phi_{65}) \simeq \frac{M^2}{2\alpha^2 \kappa^2} \quad \Rightarrow \quad H_{65} \simeq \frac{M}{2}$$

$$A_S = \frac{\alpha N}{2\pi} \kappa H = 5 \times 10^{-5} \quad \Rightarrow \quad M \simeq 2.4 \times 10^{-6} M_P$$

$$n = 1 - \frac{2}{N} \simeq 0.97$$

$$A_T = \frac{\sqrt{2}}{\pi} \frac{H}{M_P} = \frac{2}{\sqrt{\pi}} \frac{M}{M_P} = 2.7 \times 10^{-6}$$

$$n_T = -2\epsilon \simeq -\frac{3}{2N^2} = -1.6 \times 10^{-4}$$

$$r = -2\pi n_T \simeq 10^{-3}$$

11.7 Hybrid inflation

$$V(\phi, \chi) = \frac{\lambda}{4} (\chi^2 - v^2)^2 + \frac{1}{2} g^2 \phi^2 \chi^2 + \frac{1}{2} m^2 \phi^2$$  \hspace{1cm} (675)

The effective Higgs mass in the false vacuum ($\chi = 0$):

$$m^2 = \frac{\partial^2 V}{\partial \chi^2} = g^2 \phi^2 - \lambda v^2 = 0 \quad \Rightarrow \quad \phi_c = \frac{M}{g} = \frac{\sqrt{\lambda} v}{g}$$  \hspace{1cm} (676)

For large values of the inflaton, the Higgs has a large mass and sits at its minimum, and therefore the effective potential during inflation is

$$V(\phi) = V_0 + \frac{1}{2} m^2 \phi^2 \equiv V_0(1 + \xi) \simeq V_0 = \text{const.}$$

$$H_0 \simeq \sqrt{\frac{2\pi}{3} \frac{M v}{M_P}}$$  \hspace{1cm} (677)
\[ \epsilon = \frac{m^2}{\kappa^2 V_0} \xi \ll \eta = \frac{m^2}{\kappa^2 V_0} \Rightarrow n = 1 + \frac{2m^2}{\kappa^2 V_0} > 1 \quad (679) \]

\[ N = \frac{\kappa^2 V_0}{m^2} \ln \frac{\phi}{\phi_c} \Rightarrow \phi = \phi_c \epsilon^{\eta N} \quad (680) \]

Inflation ends not because of the end of slow-roll (\( \epsilon = 1 \)) but because of symmetry breaking by the Higgs

\[ A_S = \frac{H^2}{2\pi \phi} = \frac{gH}{2\pi \eta M} e^{-\eta N} = 5 \times 10^{-5} \Rightarrow \quad (681) \]

\[ g = \sqrt{\frac{3\pi}{8}} (n - 1) 10^{-4} \frac{M_P}{\epsilon} e^{(n-1)\frac{\delta}{2}} \quad (682) \]

Negligible gravitational waves:

\[ r = -2\pi n_T = 4\pi \epsilon \ll 2\pi (n - 1) \quad (683) \]

Many possibilities of scales of inflation: e.g. GUT scale,

\[ v = 10^{-3} M_P, \lambda = 0.1, g = 0.01, \Rightarrow n - 1 = 0.035, \quad (684) \]

\[ M = 4 \times 10^{15} \text{ GeV}, m = 1.3 \times 10^{12} \text{ GeV}, r = 5 \times 10^{-4} \quad (685) \]

### 11.8 Radiative corrections on SUSY hybrid inflation

Coleman-Weinberg potential

\[ V_{1-\text{loop}} = \frac{1}{64\pi^2} \sum_i (-1)^F_i m_i^4 \ln \frac{m_i^2}{\Lambda^2} \quad (686) \]

Supergravity hybrid model (units \( \kappa = 1 \))

\[ W = \sqrt{\lambda} \Phi (\bar{\Sigma} \Sigma - v^2) \quad (687) \]

\[ V = \lambda \left| \bar{\sigma} \sigma - v^2 \right|^2 + \frac{\lambda}{2} \phi^2 \left( |\sigma|^2 + |\bar{\sigma}|^2 \right) + \text{D-term} \quad (688) \]

where \( \phi = \sqrt{2} \Phi \) is the canonically normalized field. The absolute minimum appears at \( \phi = 0 \), \( \sigma = \bar{\sigma} = v \). For \( \phi > \phi_c = \sqrt{2} v \), the fields \( \sigma, \bar{\sigma} \) have a positive mass squared and stay at the origin. Inflation takes place along that “flat” direction, which is lifted by radiative corrections. The masses of bosons are \( m_B^2 = \frac{1}{2} \lambda (\phi^2 \pm 2v^2) \), while that of the fermion is \( m_F^2 = \frac{1}{2} \lambda \phi^2 \). The loop corrected potential along the flat direction is

\[ V_{1-\text{loop}}(\phi) = \frac{\lambda^2}{128\pi^2} \left[ (\phi^2 - 2v^2)^2 \ln \left( \frac{\phi^2 - 2v^2}{\Lambda^2} \right) + (\phi^2 + 2v^2)^2 \ln \left( \frac{\phi^2 + 2v^2}{\Lambda^2} \right) - 2\phi^4 \ln \left( \frac{\phi^2}{\Lambda^2} \right) \right] \quad (689) \]

\[ \Rightarrow V(\phi) \simeq \lambda v^4 \left( 1 + \frac{\lambda}{8\pi^2} \ln \frac{\phi}{\phi_c} \right), \quad \phi \gg \phi_c \quad (690) \]

\[ \epsilon = \frac{\lambda^2}{128\pi^2 \phi^2} = \frac{\lambda}{32\pi^2 N}, \quad \eta = -\frac{\lambda}{8\pi^2 \phi^2} = -\frac{1}{2N} \quad (691) \]

\[ N = \int \frac{d\phi}{\sqrt{2\epsilon}} = \frac{4\pi^2 \phi^2}{\lambda} \quad (692) \]
\[ A_S = \sqrt{\frac{N}{3}} 16\pi \frac{v^2}{M_P^2} = 5 \times 10^{-5} \Rightarrow v = 5.6 \times 10^{15} \text{GeV} \] (693)

\[ n = 1 - \frac{1}{N} = 0.98, \quad r = -2\pi n_T = 4\pi \epsilon = \frac{\lambda}{8\pi N} \ll 1 \] (694)

12 REHEATING AFTER INFLATION

One of the fundamental quests of cosmology is to understand the origin of all the matter and radiation present in the universe today. We have seen how inflation produces a homogeneous and flat background space-time, and imprints on top of it a set of scalar and tensor quantum fluctuations that become classical Gaussian random fields outside the horizon, with an approximately scale invariant spectrum.

Inflation also dilutes any relic species left from a hypothetical earlier period of the universe, such that at the end of inflation there remains only a homogeneous zero mode of the inflaton field with tiny fluctuations on the homogeneous metric. That is, the universe is empty and very cold: the entropy of the universe is exponentially small and the temperature can be taken to be zero, \( S = T = 0 \).

Therefore we are left with the puzzle: How does the large entropy and energy of our present horizon, \( S \sim 10^{89} \) and \( M \sim 10^{23} M_\odot \), arise from such a cold and empty universe? The answer seems to lie in the process by which the large potential energy density present during inflation gets converted into radiation at the end of inflation, a process known as reheating of the universe.

This process was studied soon after the first models of inflation were proposed and considered the perturbative decay of the inflaton field into quanta of other fields to which it coupled, e.g. fermions, gauge fields, and other scalars. Such couplings exist during inflation but play no role (except for inducing radiative corrections, as we will discuss later), because even if those particles were produced during inflation the exponential expansion would dilute them almost instantaneously, and nothing would be left at the end of inflation.

Let us write down the most general Lagrangian with couplings of the inflaton to other fields and among themselves,

\[ \mathcal{L} = -\frac{1}{2}(\partial_\mu \phi)^2 - V(\phi) - \frac{1}{2}(\partial_\mu \chi)^2 - \frac{1}{2}m_\chi^2 \chi^2 - \frac{1}{2}\xi \chi^2 R - \bar{\psi}(i\gamma^\mu \partial_\mu - m_\psi)\psi - h\phi \bar{\psi} \psi - \frac{1}{2}g^2 \phi^2 \chi^2 - g^2 \sigma \phi \chi^2, \] (695)

where \( g, h, \xi, \) etc. are small couplings (to avoid large radiative corrections during inflation); \( \sigma \) is the possibly finite vev of the inflaton, and we have shifted the inflaton potential by \( \phi - \sigma \rightarrow \phi \), such that the minimum is at \( \phi = 0 \) and the potential can be expanded around the minimum as

\[ V(\phi) = \frac{1}{2} m^2 \phi^2 + O(\phi^4), \] (696)

where \( m \) is the mass of the inflaton at the minimum. In chaotic inflation of the type \( m^2 \phi^2 \) or \( \lambda \phi^4 \), this mass and self-coupling are bounded by observations of the CMB to be

\[ m \sim 10^{13} \text{GeV}, \quad \lambda \lesssim 10^{-13}. \] (697)

We will consider this mass to be much larger than that of the other fields to which it couples: \( m^2 \gg m_\chi^2, m_\psi^2 \gg g^2 \sigma \phi, h\phi \). Also, the end of inflation occurs in these models when \( H \sim m \), and subsequently, the rate of expansion decays as \( H \sim 1/t < m \).

Let us compute the evolution of the inflaton after inflation, whose amplitude satisfies the equation (we are neglecting here the couplings to other fields, but we will consider them later)

\[ \ddot{\phi} + 3H(t)\dot{\phi} + m^2 \phi = 0, \] (698)
whose solution is oscillatory,

$$\phi(t) = \Phi(t) \sin mt,$$

with the amplitude of oscillations decaying like $\Phi \sim a^{-3/2}$, as we will prove now. Consider the average energy density and pressure of the homogeneous inflaton field over one period of oscillations,

$$\langle \rho \rangle = \frac{1}{2} \langle \dot{\phi}^2 \rangle + \frac{1}{2} m^2 \langle \phi^2 \rangle = \frac{1}{2} m^2 \Phi^2(t) \left( \langle \cos^2 mt \rangle + \langle \sin^2 mt \rangle \right),$$

$$\langle p \rangle = \frac{1}{2} \langle \dot{\phi}^2 \rangle - \frac{1}{2} m^2 \langle \phi^2 \rangle = \frac{1}{2} m^2 \Phi^2(t) \left( \langle \cos^2 mt \rangle - \langle \sin^2 mt \rangle \right) \approx 0$$

where we have neglected the change in $\Phi(t)$ due to the condition $m \gg H$ during reheating. The fact that an oscillating homogeneous scalar field behaves like a pressureless fluid means that the universe during that period expands like a matter dominated universe,

$$\dot{\rho} + 3H (\rho + p) = 0 \quad \Rightarrow \quad \rho = \frac{1}{2} m^2 \Phi^2(t) \sim a^{-3},$$

and therefore $\Phi \sim a^{-3/2} \sim t^{-1}$. That is, a homogeneous scalar field oscillating with frequency equal to its mass can be considered as a coherent wave of $\phi$ particles with zero momenta and particle density

$$n_\phi = \rho_\phi / m = \frac{1}{2} m \Phi^2 \sim a^{-3},$$

oscillating coherently with the same phase.

Until now we have considered only the effects of expansion, and ignored the effects due to the production of particles from the inflaton. This can be accounted for by including, in the equation of motion, the denominator of the QFT propagator,

$$\ddot{\phi} + 3H(t) \dot{\phi} + \left( m^2 + \Pi(\omega) \right) \phi = 0,$$

where $\Pi(\omega)$ is the Minkowski space polarization operator for $\phi$ with four-momentum $k^\mu = (\omega, 0)$, with $\omega = m$. The real part of the polarization operator can be neglected (due to the small couplings), $\text{Re} \, \Pi(\omega) \ll m^2$. However, due to phase space, if the frequency of oscillations satisfies $\omega \gg \text{min}(2m_\chi, 2m_\psi)$, then the polarization operator acquires an imaginary part,

$$\text{Im} \, \Pi(m) = m \Gamma_\phi,$$

where $\Gamma_\phi$ is the total decay rate of the inflaton, and we have used the optical theorem (i.e. unitarity) to relate both quantities at the physical pole, $\omega = m$.

The total decay rate can be written as a sum over partial decays,

$$\Gamma_\phi = \sum_i \Gamma(\phi \rightarrow \chi_i \chi_i) + \sum_i \Gamma(\phi \rightarrow \bar{\psi}_i \psi_i),$$

$$\Gamma(\phi \rightarrow \chi_i \chi_i) = \frac{g_i^4 v^2}{8 \pi m}, \quad \Gamma(\phi \rightarrow \bar{\psi}_i \psi_i) = \frac{h_i^2 m}{8 \pi},$$

$$\Gamma_\phi \equiv \frac{h_{\text{eff}}^2 m}{8 \pi} \ll m, \quad h_{\text{eff}}^2 = \sum_i \left( k_i^2 + \frac{g_i^4 v^2}{m^2} \right)$$

The evolution of the inflaton during the period of oscillations after inflation can be described through the phenomenological equation

$$\ddot{\phi} + 3H(t) \dot{\phi} + \Gamma_\phi \dot{\phi} + m^2 \phi = 0,$$
which includes the decay rate $\Gamma_\phi$ as a friction term giving rise to the damping of the oscillations due to inflaton particle decay. It assumes the inflaton condensate (the homogeneous zero mode) is composed of very many inflaton particles, each of these decaying into other particles to which it couples. The solution to this equation is given by (699) with

$$\Phi(t) = \Phi_0 e^{-\frac{1}{2} \int 3H dt} e^{-\frac{1}{2} \Gamma_\phi t} = \Phi_0 \frac{1}{t} e^{-\frac{1}{2} \Gamma_\phi t},$$

(710)

where we have used $H = 2/3t$.

We can now compute the evolution of the energy and number density of the inflaton field under the effect of particle production,

$$\frac{d}{dt} \left( \rho_\phi a^3 \right) = -\Gamma_\phi \rho_\phi a^3, \quad (711)$$

$$\frac{d}{dt} \left( n_\phi a^3 \right) = -\Gamma_\phi n_\phi a^3, \quad (712)$$

which simply states the usual exponential decay law for particles with decay rate $\Gamma$. Initially, the total decay rate is much smaller than the rate of expansion, $\Gamma_\phi \ll 3H = 2/t \ll m$, and the total comoving energy and total number of inflaton particles is conserved, their energy and number densities decaying like a matter fluid, $\rho_\phi \simeq m n_\phi \sim a^{-3}$.

Eventually, the universe expands sufficiently (this may take many many inflaton oscillations) that the decay rate becomes larger than the rate of expansion, or alternatively, the inflaton life-time, $\eta_\phi = \frac{1}{\Gamma_\phi}$, becomes smaller than the age of the universe, $\eta_\phi < t_U = H^{-1}$, and the inflaton decays suddenly (in less than one Hubble time), releasing all its energy density $\rho_\phi$ into relativistic particles $\chi$ and $\psi$, in an exponential burst of energy. Subsequently, the produced particles interact among themselves and soon thermalize to a common temperature. This process is responsible for the present abundance of matter and radiation energy, and could be associated with the Big Bang of the “old” cosmology.

At first sight it may seem paradoxical that the universe may have to “wait” until it is old enough for the inflaton to decay, because we are accustomed to very rapid decays in our particle physics detectors, where life-times of order $10^{-17}$s are possible, while our universe is $10^{17}$s old! However, if inflation took place at energy densities of order the GUT scale, the Hubble time of a causal domain at the end of inflation would be of order $10^{-35}$s, which is many orders of magnitude smaller than even the fastest decays of the inflaton, $\sim 10^{-25}$s. So the probability that the inflaton decays in such a short Hubble time is negligible, and the universe has to wait until it is old enough that there is any probability of decay of a single inflaton particle. Eventually, of course, once the universe is older than the inflaton life-time, it (the inflaton) will decay exponentially fast due to its constant decay rate $\Gamma_\phi$.

Let us now compute the reheating temperature of the universe that arises from the thermalization of the products of decay of the inflaton. Note that the process of reheating, once possible, is essentially instantaneous and therefore the energy density at reheating can be estimated as that corresponding to a rate of expansion $H = \Gamma_\phi$. Since all that energy density will be quickly converted into a plasma of relativistic particles, we can estimate

$$\rho(t_{\text{rh}}) = \frac{3 \Gamma_\phi^2 M_P^2}{8 \pi} = \frac{\pi^2}{30} g(T_{\text{rh}}) T_{\text{rh}}^4,$$

(713)

$$\Rightarrow \quad T_{\text{rh}} \simeq 0.1 \sqrt{\frac{\Gamma_\phi M_P}{\rho}}, \quad (714)$$

where we have assumed $g(T_{\text{rh}}) \sim 10^2 - 10^3$. Let us estimate this temperature. If we substitute $\Gamma_\phi = h_{\text{eff}}^2 m/8\pi$ with $m \sim 10^{13}$ GeV, we find

$$T_{\text{rh}} \simeq 2 \times 10^{14} h_{\text{eff}} \text{ GeV} \lesssim 10^{11} \text{ GeV},$$

(715)
where we have imposed the constraint $h_{\text{eff}} \lesssim 10^{-3}$ from radiative corrections in chaotic type models. Let us estimate it: if we consider the quantum loop corrections to the inflaton potential due to its coupling to other fields like in the Lagrangian described above, we find

$$V(\phi) = \frac{1}{2} m^2 \phi^2 \left( 1 + \frac{3g^4}{16\pi^4\lambda} \right) + \frac{\lambda}{4} \phi^4 \left( 1 + \frac{3g^4}{16\pi^4\lambda} - \frac{h^4}{16\pi^4\lambda} \right) + \ldots$$

(716)

Therefore, the couplings of the inflaton to other fields cannot be very large otherwise they would modify the amplitude of CMB anisotropies. If we impose that the mass and self-coupling of the inflaton satisfy (697), then the other couplings are bound to

$$3g^4, h^4 < 16\pi^2\lambda \quad \Rightarrow \quad g, h \lesssim 10^{-3}. \quad (717)$$

For completeness, let us mention that in theories with only gravitational interactions, like e.g. in Starobinsky model, the decay of the inflaton is induced via gravity only and

$$\Gamma_{\text{grav}} \sim \frac{m^3}{M_P^2} \sim 10^{-18} M_P \quad \Rightarrow \quad T_{\text{rh}} \sim 10^9 \text{ GeV}.$$  

(718)

All this indicates that, although the energy density at the end of inflation may be large, the final reheating temperature $T_{\text{rh}}$ may not be higher than $10^{12}$ GeV, and thus the usually assumed thermal phase transition at Grand Unification, which was the basis for most of the early universe phenomenology, like production of topological defects, GUT baryogenesis, etc., could not have taken place.

We will see shortly that such phenomenology may be resuscitated in the context of preheating and non-thermal phase transitions, but for the moment let us focus our attention onto two well differentiated and concrete cases of ordinary reheating:

### 12.1 Reheating in chaotic inflation models

Consider a $m^2\phi^2$ model of inflation, for which the value of the inflaton at the end of inflation is $\phi_{\text{end}} = M_P/2\sqrt{\pi}$, and the corresponding energy density

$$\rho_{\text{end}} = \frac{3}{2} V(\phi_{\text{end}}) = \frac{3m^2M_P^2}{16\pi} = \left(6.5 \times 10^{15} \text{ GeV}\right)^4.$$ \quad (719)

On the other hand, CMB anisotropies require

$$A_S = N \sqrt{\frac{4}{3\pi}} \frac{m}{M_P} = 5 \times 10^{-5} \quad \Rightarrow \quad m \simeq 1.4 \times 10^{13} \text{ GeV},$$ \quad (720)

while radiative corrections impose the constraint $h_{\text{eff}} \lesssim 10^{-3}$.

We are thus left with three time scales:

$$\begin{align*}
t_{\text{osc}} &\sim m^{-1} \sim 10^{-36} \text{ s} \\
t_{\text{exp}} &\gtrsim H_{\text{end}}^{-1} \sim 10^{-35} \text{ s} \\
t_{\text{dec}} &\sim \Gamma_{\phi}^{-1} \sim 10^{-25} \text{ s}
\end{align*}$$

$$\Rightarrow \quad t_{\text{osc}} \ll t_{\text{exp}} \ll t_{\text{dec}}, \quad (721)$$

so there are several oscillations per Hubble time, and we also expect many oscillations of the inflaton field before it decays. This result is typical of most high-scale models of inflation.
12.2 Reheating in low-scale hybrid inflation models

In this case, reheating occurs in very different circumstances. Most models of inflation occur at scales of order the GUT scale, because their parameters are fixed by the amplitude of CMB anisotropies, $\delta T/T \sim m/M_P \sim 10^{-5}$. However, in models of hybrid inflation, which end due to the symmetry breaking of a field coupled to the inflaton, and not because of the end of slow-roll, it is possible to decouple the amplitude of CMB fluctuations from the scale of inflation. For instance, consider a hybrid model at the electroweak scale, where the symmetry breaking field is the SM Higgs field, with a vev $v = 246$ GeV, a relatively strong coupling to the inflaton, $g = 0.4$, and a Higgs self-coupling $\lambda = 0.12$, giving rise to the following masses in the true vacuum

$$m_{\text{inf}} = g v \sim 100 \text{ GeV} , \quad m_H = \sqrt{2\lambda} v \sim 120 \text{ GeV} ,$$

(722)

which are much larger than the rate of expansion at the end of inflation

$$H_{\text{end}} = \sqrt{\frac{\pi}{3} \frac{m_H v}{M_P}} \sim 2 \times 10^{-5} \text{ eV} \ll m_H ,$$

(723)

and therefore we can neglect it during the oscillations of the inflaton and Higgs fields around the minimum of their potential.

The energy density at the end of inflation is

$$\rho_{\text{end}} = \frac{1}{8} m_H^2 v^2 \sim (10^2 \text{ GeV})^4 ,$$

(724)

which is very low indeed.

The couplings of the Higgs to matter could be large, e.g. the top quark Yukawa $h_t \sim 1$, although for such a low mass Higgs there is no phase space for top perturbative production. On the other hand, the inflaton may couple to other particles, so it is expected that their decay widths be similar and both of order $\Gamma \sim 1$ GeV. Naively, using (714), one would thus expect that the reheating temperature be $T_{\text{rh}} \sim 10^9$ GeV, but that is impossible because it would correspond to an energy density during inflation much above the actual false vacuum energy, $\rho_{\text{end}} \sim (10^2 \text{ GeV})^4$.

Actually, since the rate of expansion is so low compared with the other scales, we can ignore the decay in energy due to the expansion of the universe, which was so important during chaotic inflation, and use energy conservation to estimate

$$\rho_{\text{end}} = \frac{\lambda v^4}{4} = \frac{\pi^2}{30} g_* T_{\text{rh}}^4 \Rightarrow T_{\text{rh}} \simeq \left( \frac{15\lambda}{2\pi^2 g_*} \right)^{1/4} v \sim 42 \text{ GeV} ,$$

(725)

where we have used $g_* = 106.75$ as the effective number of degrees of freedom of the SM particles. Note that this temperature is rather low, but in fact we have no observational evidence that the universe has actually gone through a thermal period with a temperature above this.

We are thus left with three time scales:

$$t_{\text{osc}} \sim m^{-1}_{\text{inf}} \sim 10^{-27} \text{ s} , \quad t_{\exp} \sim H^{-1} \sim 10^{-10} \text{ s} , \quad t_{\text{dec}} \sim \Gamma^{-1} \sim 10^{-23} \text{ s} \ \Rightarrow \quad t_{\text{osc}} \ll t_{\text{dec}} \ll t_{\exp} ,$$

(726)

so there are many oscillations per Hubble time, but contrary to the case of chaotic inflation models, here the decay time is much smaller than the expansion time, because the universe is already quite old, so once the inflaton and Higgs start oscillating they decay very soon via their usual perturbative decay.
The previous discussion falls under the name of \textit{perturbative reheating}, because it assumes that the coherently oscillating inflaton will decay as if it were composed on individual inflaton quanta, each one decaying as described by ordinary QFT, with the perturbative decay rate computed above. This was the standard lore during at least a decade since it was first proposed in 1982. However, it was soon realized that the inflaton at the end of inflation is actually a coherent wave, a zero mode, a condensate made out of many inflaton quanta, all oscillating with the same phase, and non-perturbative effects associated with this condensate were bound to be important for the problem of reheating. In fact, a few years ago, in a seminal paper, Linde, Kofman and Starobinsky proposed a new picture of reheating, which has become known as \textit{preheating}. I will describe these new developments in the following sections. They make use of the well studied problem of particle production in the presence of strong background fields, whose formalism we have already encountered for the analysis of the generation of metric fluctuations during inflation. In this case, instead of a quantum field evolving in a rapidly changing gravitational field (like during inflation), we have a field coupled to the inflaton, which has a rapidly changing frequency or mass due to the inflaton oscillations.

We will first describe the Bogolyubov formalism for a single scalar field with a time-dependent mass and then particularize to the case of the inflaton oscillations after inflation. Later on, we will also extended the formalism to fermions, which can also be produced at preheating.

Consider a massive scalar field $\phi$ with Lagrangian density
\begin{equation}
L = \frac{1}{2} \dot{\phi}^2 - \frac{1}{2} (\nabla \phi)^2 - \frac{1}{2} m^2 \phi^2,
\end{equation}
which gives a canonically conjugate momentum $\pi = \frac{\delta L}{\delta \dot{\phi}} = \dot{\phi}$, and the Hamiltonian
\begin{equation}
H = \pi \dot{\phi} - L = \frac{1}{2} \dot{\phi}^2 + \frac{1}{2} (\nabla \phi)^2 + \frac{1}{2} m^2 \phi^2.
\end{equation}
We can treat the fields as quantum fields and define the usual equal time commutation relation
\begin{equation}
[\phi(x,t), \pi(x',t)] = i \delta^{(3)}(x - x'),
\end{equation}
as well as expand in Fourier components,
\begin{equation}
\phi(x,t) = \int \frac{d^3k}{(2\pi)^3/2} \phi_k(t) e^{ik\cdot x}.
\end{equation}
The field mode $\phi_k(t)$ satisfies the harmonic oscillator equation
\begin{equation}
\ddot{\phi}_k + \omega_k^2 \phi_k = 0,
\end{equation}
\begin{equation}
\omega_k^2(t) = k^2 + m^2(t),
\end{equation}
where the time dependence of the oscillation frequency comes through that of the mass. We will assume that the field is real, so we should impose the constraint $\phi_k(t) = \phi^*_k(t)$. Following the quantization condition (729), we can write the field and momentum operators in terms of time-dependent creation and annihilation operators,
\begin{equation}
\phi_k(t) = \frac{1}{\sqrt{2\omega_k}} \left( a_k(t) + a_{-k}^+(t) \right),
\end{equation}
\begin{equation}
\pi_k(t) = -i \sqrt{\frac{\omega_k}{2}} \left( a_k(t) + a_{-k}^+(t) \right),
\end{equation}
satisfying the usual commutation relation, \( \forall t, \)
\[
[a_k(t), a_{k'}^\dagger(t)] = \delta^{(3)}(k - k'),
\]
and in terms of which the Hamiltonian becomes
\[
H = \frac{1}{2} \int d^3k \left[ \pi_k \pi_k^\dagger + \omega_k^2 \phi_k \phi_k^\dagger \right]
= \frac{1}{2} \int d^3k \omega_k \left( a_k^\dagger a_k + a_k a_k^\dagger \right) \equiv H_{\text{part}} + H_{\text{vac}}(t),
\]
where
\[
H_{\text{part}} = \int d^3k \omega_k a_k^\dagger a_k, \quad (735)
\]
\[
H_{\text{vac}}(t) = \frac{V}{(2\pi)^3} \int d^3k \frac{\omega_k}{2}. \quad (736)
\]
We can then define a number operator for these fields
\[
N = \int d^3k a_k^\dagger a_k, \quad (737)
\]
and a Fock space with vacuum state defined as
\[
a_k(t)|0_t\rangle = 0, \quad \langle 0_t|0_t\rangle = 1, \quad \langle n_k|0_t\rangle = 0, \quad \langle 0_t|n_k\rangle = 1, \quad (738)
\]
and particle states \( |n_k\rangle \propto (a_k^\dagger)^n|0_t\rangle \) satisfying
\[
H_{\text{part}}|n_k\rangle = n_k \omega_k |n_k\rangle \equiv E_k |n_k\rangle, \quad (739)
\]
\[
N|n_k\rangle = n_k |n_k\rangle. \quad (740)
\]
In the vacuum state \( |0_t\rangle \), the energy takes its lowest possible value, \( H_{\text{vac}}(t) = \langle 0_t|H|0_t\rangle \).

We can compute the equations of motion as usual with
\[
\frac{d}{dt} a_k = \frac{\partial a_k}{\partial t} + i [H, a_k]
\]
where we can invert the relations (733)
\[
a_k(t) = \sqrt{\frac{\omega_k}{2}} \phi_k(t) + \frac{i}{\sqrt{2\omega_k}} \pi_k(t), \quad (741)
\]
\[
a_{-k}(t) = \sqrt{\frac{\omega_k}{2}} \phi_k(t) - \frac{i}{\sqrt{2\omega_k}} \pi_k(t).
\]
In the Heisenberg picture, the original canonical operators \( \{\phi_k, \pi_k\} \) may have no explicit time-dependence, but \( \omega_k \) is indeed time-dependent, so
\[
\frac{d}{dt} a_k = -i \omega_k a_k + \frac{\omega_k}{2\omega_k} a_{-k}^\dagger.
\]
(742)
The solution to the equations of motion is
\[
\begin{pmatrix}
  a_k(t) \\
  a_{-k}(t)
\end{pmatrix} =
\begin{pmatrix}
  u_k(t) & v_k(t) \\
  v_k^*(t) & u_k^*(t)
\end{pmatrix}
\begin{pmatrix}
  a_k(0) \\
  a_{-k}^\dagger(0)
\end{pmatrix}
\]
(743)
The unitary evolution preserves the commutation relation (729) iff
\[ |u_k|^2 - |v_k|^2 = 1, \tag{744} \]
with initial condition:
\[ |u_k|^2 = 1, \quad |v_k|^2 = 0. \tag{745} \]

If the initial state is the vacuum, \(|0\rangle \equiv |0_{t=0}\rangle\), then
\[ a_k(0)|0\rangle = 0 \quad \Rightarrow \quad a_k(t)|0\rangle = v_k(t)a_{-k}^\dagger(0)|0\rangle \neq 0 \tag{746} \]
In particular, the number density of particles created from the vacuum is
\[ n(t) = \frac{1}{V} \langle 0 | N | 0 \rangle = \int \frac{d^3k}{(2\pi)^3} |v_k|^2(t). \tag{747} \]

In order to find the function \(n(t)\) explicitly, we have to solve for \(u_k\) and \(v_k\) as a solution of
\[
\begin{pmatrix}
\dot{u}_k(t) \\
\dot{v}_k^\ast(t)
\end{pmatrix} = 
\begin{pmatrix}
-i\omega_k & \frac{\dot{\omega}_k}{2\omega_k} \\
\frac{\omega_k}{2\omega_k} & i\omega_k
\end{pmatrix}
\begin{pmatrix}
u_k(t) \\
v_k^\ast(t)
\end{pmatrix}, \tag{748}
\]

It is customary to write the mode functions \(u_k\) and \(v_k\) in terms of the usual Bogolyubov coefficients, \(\{\alpha_k, \beta_k\}\),
\[ u_k = \alpha_k e^{-i \int^t \omega_k dt}, \quad v_k^\ast = \beta_k e^{i \int^t \omega_k dt}, \tag{749} \]
then the evolution equations (748) become
\[ \dot{\alpha}_k = \frac{\dot{\omega}_k}{2\omega_k} \beta_k e^{2i \int^t \omega_k dt}, \tag{750} \]
\[ \dot{\beta}_k = -\frac{\dot{\omega}_k}{2\omega_k} \alpha_k e^{-2i \int^t \omega_k dt}, \]
which can be integrated in the adiabatic approximation, to give
\[ n(t) = \int \frac{d^3k}{(2\pi)^3} n_k(t) = \int \frac{d^3k}{(2\pi)^3} |\beta_k|^2(t), \tag{751} \]
the number density of particles produced due to the time-dependent background field.

Alternatively, one can introduce the \(|\text{in}\rangle\) and \(|\text{out}\rangle\) states, and make the field decomposition over time-independent creation and annihilation operators \(\{a_k, a_{-k}^\dagger\}\),
\[ \phi(x, t) = \int \frac{d^3k}{(2\pi)^3/2} \left[ f_k(t) a_k e^{ik\cdot x} + h.c. \right] \]
\[ = \int \frac{d^3k}{(2\pi)^3/2} \left[ f_k(t) a_k + f_k^\ast(t) a_{-k}^\dagger \right] e^{ik\cdot x}, \tag{752} \]
\[ \pi(x, t) = \int \frac{d^3k}{(2\pi)^3/2} \left[ g_k(t) a_k + g_k^\ast(t) a_{-k}^\dagger \right] e^{ik\cdot x}, \tag{753} \]
where the mode functions \(f_k(t)\) and \(g_k(t)\) depend only on the modulus \(k = |k|\), thanks to the homogeneity and isotropy of the background fields. These functions satisfy the equations of motion
\[ \dot{f}_k + \omega_k^2 f_k = 0, \quad g_k = if_k. \tag{754} \]
Comparing with the former decomposition (733), we find the relation

\[ u_k = \frac{1}{\sqrt{2\omega_k}} (\omega_k f_k + g_k), \]
\[ v_k = \frac{1}{\sqrt{2\omega_k}} (\omega_k f_k - g_k), \]

and vice versa

\[ f_k = \frac{1}{\sqrt{2\omega_k}} (u_k + v_k^*), \]
\[ g_k = \sqrt{\frac{\omega_k}{2}} (u_k - v_k^*), \]

which gives for the occupation number

\[ n_k(t) = |\beta_k|^2 = \frac{1}{2\omega_k} |\dot{f}_k|^2 + \frac{\omega_k}{2} |f_k|^2 - \frac{1}{2}, \]

where we have used the Wronskian

\[ i(\dot{f}_k f_k^* - \dot{f}_k^* f_k) = 2 \text{Re} (f_k^* g_k) = 1 \Leftrightarrow |u_k|^2 - |v_k|^2 = 1. \]

13.1 Diagonalization of the Hamiltonian

With the above decomposition, we can write the Hamiltonian as

\[ H = \int d^3k \left[ E_k(t) \left( a_k^\dagger a_k + a_k a_k^\dagger \right) + F_k(t) a_k a_{-k} + F_k^*(t) a_k^\dagger a_{-k}^\dagger \right], \]

where

\[ E_k(t) = \frac{1}{2} (|\dot{f}_k|^2 + \omega_k^2 |f_k|^2) = \omega_k \left( n_k + \frac{1}{2} \right), \]
\[ F_k(t) = \frac{1}{2} (\dot{f}_k^2 + \omega_k^2 f_k^2), \]
\[ E_k^2(t) - |F_k(t)|^2 = \frac{\omega_k^2}{4}. \]

Let us now introduce a canonical Bogolyubov transformation

\[
\begin{pmatrix}
  a_k \\
  a_{-k}^\dagger
\end{pmatrix}
= 
\begin{pmatrix}
  u_k(t) & v_k(t) \\
  v_k^*(t) & u_k^*(t)
\end{pmatrix}
\begin{pmatrix}
  b_k \\
  b_{-k}^\dagger
\end{pmatrix}
\]

Then

\[ n_k = |\beta_k|^2 = \frac{2E_k - \omega_k}{2\omega_k}, \]
\[ \frac{u_k}{v_k} = \frac{2E_k + \omega_k}{2F_k^*}, \]

and the Hamiltonian becomes diagonal

\[ H = \int d^3k \frac{\omega_k}{2} \left( b_k^\dagger b_k + b_k b_k^\dagger \right), \]

which can be decomposed into \( H_{\text{part}} \) and \( H_{\text{vac}} \), as before, see (735).
### 13.2 The Schrödinger picture

We can define the unitary evolution operator $U^\dagger(t) = U^{-1}(t)$, where $i\hbar \partial_t U(t) = H U(t)$, such that time evolution determines

$$a_k(t) = U^\dagger(t) a_k(0) U(t).$$

The solution of the Schrödinger equation is the squeezed state

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle.$$  

The vacuum at time $t$ is given by $|0_t\rangle = U^\dagger(t)|0\rangle$. Let $|\psi(0)\rangle$ be the initial vacuum state $|0\rangle$. Then the operator

$$b_k(t) = U(t)a_k(0)U^\dagger(t) = u_k^* a_k(0) - v_k a_k^+(0)$$

annihilates the state $|\psi(t)\rangle$. Now let us use (741) to evaluate

$$a_k(0) = \sqrt{\frac{\omega_k}{2}} \phi_k(0) + i \frac{\pi_k(0)}{\sqrt{2\omega_k}},$$

and substitute into $b_k(t)|\psi(t)\rangle = 0$,

$$\frac{1}{\sqrt{2\omega_k}} [(u_k^* - v_k)\omega_k \phi_k(0) + i(u_k^* + v_k)\pi_k(0)] |\psi(t)\rangle = 0.$$ 

Therefore, the evolved state satisfies the Schrödinger equation

$$[\Omega_k(t)\phi_k(0) + i\pi_k(0)]|\psi(t)\rangle = 0,$$

where we have used $g_k f_k = \text{Re}(g_k f_k) - i\text{Im}(f_k^* g_k) = \frac{1}{2} - i\text{Re}(f_k^* \dot{f}_k)$. Using the operator definition $\pi_k = -i\frac{\partial}{\partial \phi_k}$, we find the solution

$$\psi(\phi_k, \dot{\phi}_k^*, t) \sim e^{-\Omega_k(t)|\phi_k|^2},$$

$$P(\phi_k, \dot{\phi}_k^*, t) = |\psi(\phi_k, \dot{\phi}_k^*, t)|^2 \sim e^{-i\frac{1}{\Omega_k(t)}|\phi_k|^2}.$$  

The phase $F_k(t) = \text{Re}(f_k^* \dot{f}_k) \gg 1$ quickly becomes very large during preheating, which ensures that the state becomes a squeezed state, with large occupation numbers, described by the Gaussian distribution (774).

### 13.3 Parametric resonance

We will consider here the case of a scalar field $\chi$ coupled to the inflaton $\phi$ with coupling $\frac{1}{2}g^2 \phi^2 \chi^2$, which induces an oscillating mass term

$$m^2_\chi(t) = m^2_\chi + g^2\dot{\phi}^2(t).$$

The inflaton is assumed to oscillate like (699) with a frequency given by its mass $m$, not necessarily much larger than the “bare” mass of the field $\chi$. In that case, the frequency can be written as

$$\omega_k^2(t) = k^2 + m^2_\chi + g^2\Phi^2(t)\sin^2 mt,$$

and the mode equation (754) can be written as a Mathieu equation, where $z = mt$, and primes denote differentiation w.r.t. $z$,

$$f_k'' + (A_k - 2q \cos 2z) f_k = 0,$$

$$A_k = \frac{k^2 + m^2_\chi}{m^2} + 2q, \quad q = \frac{g^2\Phi^2(t)}{4m^2}.$$  

The Mathieu equation is part of a large class of Hill equations (which includes also the Lamé equation and many others) that present unstable solutions for certain values of the momenta for a given set of parameters \( \{A_k, q\} \), with \( A \geq 2q \). These solutions fall into bands of instability that are narrow for small values of the resonant parameter \( q \leq 1 \), but can be very broad for larger values of \( q \).

The solutions to the Mathieu eq. have the form \( f_k(z) = e^{p_k z} p(z) \), with \( p_k \) the Floquet index, characterizing the exponential instability, and typically much smaller than one, although it could be as large as \( \mu_{\text{max}} = 0.28055 \); and where \( p(z) \) is a periodic function of \( z \). The occupation number can then be computed to be

\[
n_k(t) \sim e^{2\mu_k m t},
\]

which can grow significantly in a few oscillations, if the growth index \( \mu_k \) is not totally negligible.

The effect of parametric resonance is similar to the lasing effect (or light amplification by stimulated emission of radiation), where a large population of particles is produced from oscillations of a coherent source.

### 13.4 Narrow resonance

Let us consider first the case where \( m_\chi, g \Phi \ll m, \) or \( q \ll 1 \). Then the Mathieu equation instability chart shows that the resonance occurs only in some narrow bands around \( A_k \approx l^2 \), \( l = 1, 2, \ldots \), with widths in momentum space of order \( \Delta k \sim q^2 ; \) so, for \( q < 1 \), the most important band is the first one, \( A_k \sim 1 + q, \) centered around \( k = m/2 \).

The growth factor \( \mu_k \) for the first instability band is given by

\[
\mu_k = \sqrt{\frac{q^2}{2} - \left(\frac{2k}{m} - 1\right)^2}.
\]

The resonance occurs for \( k \) within the range \( \frac{m}{2}(1 \pm \frac{q}{2}) \). The index \( \mu_k \) vanishes at the edges of the resonance band and takes its maximum value \( \mu_k = \frac{q}{2} \) at \( k = \frac{m}{2} \). The corresponding modes grow at a maximal rate \( \chi_k \sim \exp(qz/2) \). This leads to a growth of the occupation numbers (779) as \( n_k \sim \exp(qm t) \).

We can interpret this as follows. In the limit \( q \ll 1 \), the effective mass of the \( \chi \) particles is much smaller than \( m \), and each decaying \( \phi \) particle creates two \( \chi \) particles with momentum \( k \sim m/2 \). The difference with respect to the perturbative decay \( \Gamma(\phi \rightarrow \chi \chi) \) is that, in the regime of parametric resonance, the rate of production of \( \chi \) particles is proportional to the number of particles produced earlier (which gives rise to an exponential growth in time). This is a non-perturbative effect, as we will discuss later, and we could not have obtained it by using the methods described in the previous section, at any finite order of perturbation theory with respect to the interaction term \( g^2 \Phi^2 \sin^2 mt \). It is by solving the mode equation (777) exactly that we have found this result.

Note that only a very narrow range of modes grow exponentially with time, so the spectrum of particles is dominated by these modes, while the rest are still in the vacuum, produced only through the ordinary perturbative decay process. Of course, the exponential production does not last for ever: the universe expansion is going to affect the resonant production of particles in two ways, leading to the end of the narrow resonance regime.

First, the time-dependent amplitude of oscillations \( \Phi(t) \), which determines \( q \), see (778), not only decays \( \sim t^{-1} \) due to the expansion of the universe, but also due to the perturbative decay of the inflaton field, \( \Phi(t) \sim \exp(\Gamma_\phi t/2) \). Therefore, the narrow resonance will end when the usual perturbative decay becomes important, i.e. when \( q m < \Gamma_\phi \).

Second, in the evolution equation (777), the momenta \( k \) are actually physical momenta, which redshift with the scale factor as \( k_{\text{phys}} = k/a \), and therefore, even if a given mode is initially within the narrow band, \( \Delta k \sim q m \), it will very quickly redshift away from it, within the time scale \( \Delta t \sim q H^{-1} \),
preventing its occupation numbers (779) from growing exponentially. Thus, the narrow resonance will end when \( q^2 m < H \).

Therefore, if the amplitude of inflaton oscillations decays like \( \Phi \sim 1/t \), there will always be a time (typically a dozen oscillations) for which one of the two conditions above will hold and the narrow resonance will end.

### 13.5 Broad resonance

If the initial amplitude of oscillations is very large, like in models of chaotic inflation, in which \( \Phi_0 \sim M_P/10 \) and \( m \sim 10^{-6} M_P \), then the initial \( q \)-parameter could be very large,

\[
q_0 = \frac{g^2 \Phi_0^2}{4m^2} \sim g^2 10^{10} \lesssim 10^4 ,
\]

where we have used the constraint due to radiative corrections (717). In this case, the \( \chi \) particle production due to stimulated emission by the oscillating inflaton field can be very efficient as it enters into the broad resonance regime.

Particles are produced only at the instances of maximum acceleration of the inflaton field, when \( \phi(t) \sim 0 \), and

\[
\left| \frac{\dot{\omega}_k}{\omega_k^2} \right| \gg 1 ,
\]

a relation known as the non-adiabaticity condition. When it holds, we cannot define a proper Fock space for the \( \chi \) particles, and the occupation numbers of those particles grow very quickly. We thus associate (782) with particle production.

We will now describe how to compute the growth of modes and the Floquet index in this regime, using the formalism developed above. We can expand the quantum field \( \chi \) in Fourier components \( f_k \) satisfying the mode equation (754) with time-dependent frequency (776) and initial conditions

\[
f_k(0) = \frac{1}{\sqrt{2\omega_k}} e^{-i\omega_k t} , \quad g_k(0) = i \dot{f}_k(0) = \omega_k f_k(0) ,
\]

whose evolution in terms of the Bogolyubov coefficients is

\[
f_k(t) = \frac{\alpha_k(t)}{\sqrt{2\omega_k}} e^{-i\omega_k t} + \frac{\beta_k(t)}{\sqrt{2\omega_k}} e^{+i\omega_k t} ,
\]

\[
\alpha_k(0) = 1 , \quad \beta_k(0) = 0 .
\]

And the occupation numbers are

\[
n_k(t) = |\beta_k(t)|^2 = \frac{1}{2\omega_k} |\dot{f}_k|^2 + \frac{\omega_k}{2} |f_k|^2 - \frac{1}{2} .
\]

The inflaton field has maximum acceleration at \( t = t_j = j\pi/m \), such that \( \sin mt_j = 0 \). Between \( t_j \) and \( t_{j+1} \), the amplitude \( \phi(t) \approx \phi_0 = \text{const} \), so that the frequency \( \omega_k(t) \) is approximately constant between successive zeros of the inflaton, and we can properly define a Fock space for \( \chi \). At \( t_j \), the amplitude changes rapidly, such that (782) is satisfied and we cannot define an adiabatic invariant like the occupation number (786). Therefore, let us study the behaviour of the modes \( \chi_k \) precisely at those instances \( t = t_j \). We can expand the time-dependent frequency (776) around those points (where the frequency has a minimum) as

\[
\omega_k^2(t) = \omega_k^2(t_j) + \frac{1}{2} \omega_k^2(t_j)(t - t_j)^2 + \cdots
\]
and make the change of variables

\[ \eta \equiv [2\omega_k^2(t)]^{1/4}(t - t_j), \]  
\[ \kappa^2 \equiv \frac{\omega_k^2(t_j)}{\sqrt{2\omega_k^2(t_j)}} = \frac{k^2 + m_k^2}{2gm\Phi} = A_k - 2q \frac{1}{4\sqrt{q}}. \]  

(788)  
(789)

The mode equation (754) around \( t = t_j \) then becomes

\[ \frac{d^2 f_k}{d\eta^2} + \left( \kappa^2 + \frac{\eta^2}{4} \right) f_k = 0, \]  
(790)

which can be interpreted as a Schrödinger equation for a wave function scattering in an inverted parabolic potential. The exact solutions are parabolic cylinder functions, \( W(\pm \kappa^2, \pm \eta) \), whose asymptotic expressions are well known. Thus we have substituted the problem of parametric resonance after chaotic inflation with that of partial waves scattering off successive inverted parabolic potentials.

Let the wave \( f_k(t) \) have the form of the adiabatic solution (784) before scattering at \( t_j \),

\[ f_k^j(t) = \frac{\alpha_k^j}{\sqrt{2\omega_k}} e^{-i\int_0^t \omega_k dt} + \frac{\beta_k^j}{\sqrt{2\omega_k}} e^{i\int_0^t \omega_k dt}, \]  
(791)

where the coefficients \( \{\alpha_k^j, \beta_k^j\} \) are constant, for \( t_{j-1} < t < t_j \).

After scattering the potential at \( t_j \), the wave \( f_k^j(t) \) takes the form

\[ f_k^{j+1}(t) = \frac{\alpha_k^{j+1}}{\sqrt{2\omega_k}} e^{-i\int_0^t \omega_k dt} + \frac{\beta_k^{j+1}}{\sqrt{2\omega_k}} e^{i\int_0^t \omega_k dt}, \]  
(792)

where the coefficients \( \{\alpha_k^{j+1}, \beta_k^{j+1}\} \) are again constant, for \( t_j < t < t_{j+1} \). These are essentially the asymptotic expressions for the incoming and the outgoing waves, scattered at \( t_j \). Therefore, the outgoing amplitudes \( \{\alpha_k^{j+1}, \beta_k^{j+1}\} \) can be expressed in terms of the incoming amplitudes \( \{\alpha_k^j, \beta_k^j\} \) with the help of the reflection \( R_k \) and transmission \( D_k \) coefficients of scattering at \( t_j \),

\[ \left( \begin{array}{c} \alpha_k^{j+1} e^{-i\theta_k^j} \\ \beta_k^{j+1} e^{+i\theta_k^j} \end{array} \right) = \left( \begin{array}{cc} \frac{1}{D_k} & \frac{R_k}{D_k} \\ \frac{R_k}{D_k} & \frac{1}{D_k} \end{array} \right) \left( \begin{array}{c} \alpha_k^j e^{-i\theta_k^j} \\ \beta_k^j e^{+i\theta_k^j} \end{array} \right), \]  
(793)

where \( \theta_k^j = \int_0^{t_j} \omega_k(t) dt \), and

\[ R_k = -i e^{-i\phi_k} [1 + e^{2\pi\kappa^2}]^{-1/2}, \]  
\[ D_k = e^{-i\phi_k} [1 + e^{-2\pi\kappa^2}]^{-1/2}, \]  
\[ |R_k|^2 + |D_k|^2 = 1. \]  

(794)

The \( k \)-dependent angle of scattering is

\[ \phi_k = \text{Arg} \Gamma \left[ \frac{1}{2} + i\kappa^2 \right] + \kappa^2 (1 - \ln \kappa^2). \]  
(795)

Simplifying (793), we find

\[ \left( \begin{array}{c} \alpha_k^{j+1} \\ \beta_k^{j+1} \end{array} \right) = \left( \begin{array}{cc} [1 + e^{-2\pi\kappa^2}]^{1/2} e^{i\phi_k} & i e^{-\pi\kappa^2 + 2i\theta_k} \\ -i e^{-\pi\kappa^2 - 2i\theta_k} & [1 + e^{-2\pi\kappa^2}]^{1/2} e^{-i\phi_k} \end{array} \right) \left( \begin{array}{c} \alpha_k^j \\ \beta_k^j \end{array} \right). \]  
(796)
and therefore, using \( n_k^j = |\beta_k^j|^2 \) and \( |\alpha_k^j|^2 |\beta_k^j|^2 = n_k^j (n_k^j + 1) \), we have
\[
n_k^{j+1} = e^{-2\pi \kappa^2} + (1 + 2e^{-2\pi \kappa^2}) n_k^j - 2e^{-\pi \kappa^2} [1 + e^{-2\pi \kappa^2}]^{1/2} [n_k^j (n_k^j + 1)]^{1/2} \sin \theta_{\text{tot}}^j ,
\]
where \( \theta_{\text{tot}}^j = 2\theta_k^j - \phi_k + \text{Arg} \beta_k^j - \text{Arg} \alpha_k^j \).

This expression is very enlightening. Let us describe its properties:

- **Step-like.** The number of created particles is a step-like function of time. The occupation number between successive scatterings is constant. In the first scattering (when \( n_k^0 = 0 \)), we have
  \[
n_k = e^{-2\pi \kappa^2} = e^{-\frac{\pi \kappa^2}{\sin \theta_k}} < 1 .
  \]

- **Non-perturbative.** The occupation number (798) cannot be expanded perturbatively, for small coupling, because the function \( e^{-1/g} \) is non-analytical at \( g = 0 \). This is the form that most non-perturbative effects take in quantum field theory.

- **Infrared effect.** For large momenta, the occupation number decays exponentially, so even if there are bands at low momenta, i.e. in the IR region, the high momentum modes will not be populated,
  \[
  \kappa^2 \gg \pi^{-1} \quad \Rightarrow \quad n_k^{j+1} \simeq n_k^j \simeq 0 .
  \]

- **Non-linear.** For small momenta one may have production of particles with mass greater than that of the inflaton:
  \[
  \kappa^2 = \frac{k^2 + m_X^2}{2gm\Phi_0} \lesssim \pi^{-1} \quad \Rightarrow \quad n_k \text{ large if } m^2 < m_X^2 \ll gm\Phi_0
  \]

- **Exponential boson production.** In the case of bosons (we will discuss the fermionic case later), the occupation number can grow exponentially due to Bose-Einstein statistics, \( n_k \sim \exp(2\mu_k z) \gg 1 \),
  \[
  n_k^{j+1} \simeq [(1 + 2e^{-2\pi \kappa^2}) - 2e^{-\pi \kappa^2} [1 + e^{-2\pi \kappa^2}]^{1/2} \sin \theta_{\text{tot}}^j] n_k^j \equiv e^{2\pi \mu_k} n_k^j
  \]
  which allows one to estimate the Floquet index \( \mu_k \).

- **Resonant production.** Valid only for periodic sources. If scattering occurs in phase, the incoming and outgoing waves add up constructively, and we can have resonant effects. This occurs when \( \theta_{\text{tot}}^j \) is a semi-integer multiple of \( \pi \). In that case, it is possible that, for some modes, \( n_k^{j+1} > n_k^j \). This gives rise to a particular band structure.

- **Stochastic preheating.** It may happen that the phase a mode has acquired in a given scattering exactly compensates for the universe expansion in that interval and the phases destructively interfere, decreasing the number of particles in that mode. This gives rise to a stochastic growth of particles, where approximately 3/4 of the time the particle number increases.

- **Band structure.** Different models of inflation give rise to different evolution laws for the amplitude of inflaton oscillations, and therefore to different mode equations (754). The corresponding Hill equations (linear second order differential equations with periodic coefficients) can have quite different band structures, e.g. those of Mathieu or Lamé equations.

Even if we compute the complete band structure of the Mathieu or Lamé equation and we determine the growth factors \( \mu_k \) with great accuracy, the universe expansion will shift any given mode from one band to the next, as the mode redshifts and the amplitude of inflaton oscillations decreases: A mode starts in a given band, its occupation numbers increase exponentially through several oscillations, and suddenly it falls out of the band, until the expansion makes it fall into the next band, and so on until it reaches the narrow resonance regime described above.
14 CONCLUSION

In the last ten years we have seen a true revolution in the quality and quantity of cosmological data that has allowed cosmologists to determine most of the cosmological parameters with a few percent accuracy and thus fix a Standard Model of Cosmology. The art of measuring the cosmos has developed so rapidly and efficiently that one may be tempted of renaming this science as Cosmonomy, leaving the word Cosmology for the theories of the Early Universe. In summary, we now know that the stuff we are made of — baryons — constitutes just about 4% of all the matter/energy in the Universe, while 25% is dark matter — perhaps a new particle species related to theories beyond the Standard Model of Particle Physics —, and the largest fraction, 70%, some form of diffuse tension also known as dark energy — perhaps a cosmological constant. The rest, about 1%, could be in the form of massive neutrinos.

Nowadays, a host of observations — from CMB anisotropies and large scale structure to the age and the acceleration of the universe — all converge towards these values, see Fig. 30. Fortunately, we will have, within this decade, new satellite experiments like Planck, CMBpol, SNAP as well as deep galaxy catalogs from Earth, to complement and precisely pin down the values of the Standard Model cosmological parameters below the percent level, see Table 1.

All these observations would not make much sense without the encompassing picture of the inflationary paradigm that determines the homogeneous and isotropic background on top of which it imprints an approximately scale invariant gaussian spectrum of adiabatic fluctuations. At present all observations are consistent with the predictions of inflation and hopefully in the near future we may have information, from the polarization anisotropies of the microwave background, about the scale of inflation, and thus about the physics responsible for the early universe dynamics.

ACKNOWLEDGEMENTS

I would like to thank Misha Shaposhnikov and Ruth Durrer for inviting me to give this course at the Troisieme Cycle de Physique de la Suisse Romand. This work was supported in part by a CICYT project AYA2009-13936-C06-06.

References


[34] The Chandra X-ray observatory Home Page: http://chandra.harvard.edu/


[92] Planck Surveyor Home Page: http://astro.estec.esa.nl/Planck/


[99] ACT experiment Home Page: http://www.hep.upenn.edu/act/


