Cosmology II

Hannu Kurki-Suonio

Fall 2020

Preface

In Cosmology I we discussed the universe in terms of a homogenous and isotropic approximation to it. In Cosmology II we add the inhomogeneity and anisotropy (Chapters 8 and 9). The mathematical background required includes Fourier analysis (taught in Fysiikan matemaattiset menetelmät I) and spherical harmonic analysis (taught in Fysiikan matemaattiset menetelmät II). We will take some results from Quantum Field Theory and Cosmological Perturbation Theory, but students are not expected to have them as background – they are more advanced courses. We begin with Inflation, but postpone the discussion of generation of perturbations during it to after we have discussed inhomogeneity in general and its later evolution – the chapter on Structure Formation. Thus in the Inflation chapter we still assume the homogeneous FRW model. We end with the Cosmic Microwave Background Anisotropy, which forms an important part of observational data in cosmology.

7 Inflation

7.1 Motivation

In Cosmology I we discussed how the universe began with a Hot Big Bang. This leaves open the question of initial conditions – how did the Hot Big Bang begin, and why did it begin with such a state of high density and temperature, rapid expansion, and a high level of isotropy and homogeneity. Inflation is a *scenario* to address this question, at least to some extent. Inflation is a period in the very early universe, before the events discussed in Cosmology I, when the expansion of the universe was *accelerating*.

Inflation is not really a specific theory; rather it is a more general idea of a certain kind of behavior (i.e., a "scenario") for the universe. It is not known for sure whether inflation occurred, but it makes a number of predictions that agree with observations. Inflation has been more successful than competing ideas for the very early universe and it has become part of the standard model for cosmology. The most important property of inflation is that it provides a mechanism for generating the initial density fluctuations, the *primordial perturbations*, from which the structure of the universe, stars and galaxies, grew. However, this property was discovered later, and the original motivation for inflation was to explain the initial flatness and homogeneity of the universe and the lack of certain relics that could have been produced at the very high temperatures of the very early universe[1][1]. This chapter discusses inflation in the homogeneous and isotropic approximation. Perturbations are discussed in Chapter 8.

Much of this chapter follows Chapter 3 of the book by Liddle&Lyth.[2]

7.1.1 Flatness problem

The Friedmann equation can be written as

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

If the universe has the critical density, $\Omega = 1$, it stays that way (since $K = 0$). But if $\Omega \neq 1$, it evolves in time. The difference $\Omega_k = 1 - \Omega$ grows with time during both the radiation-dominated and matter-dominated epochs. If Ω_k is small, its time evolution takes the form

$$\text{mat.dom} \qquad a \propto t^{2/3},\ H \propto t^{-1} \Rightarrow \frac{1}{aH} \propto t^{1/3} \qquad \Rightarrow \Omega_k \propto t^{2/3} \tag{2}$$

$$\text{rad.dom} \qquad a \propto t^{1/2},\ H \propto t^{-1} \Rightarrow \frac{1}{aH} \propto t^{1/2} \qquad \Rightarrow \Omega_k \propto t\,. \tag{3}$$

Since today, and at the end of the matter-dominated epoch, $\Omega_0 = \mathcal{O}(1)$ – and it is not essential here that Ω_0 is very close to 1, it would be enough that, say, $0.1 < \Omega_0 < 10$ – we can calculate backwards in time to, e.g., Big Bang Nucleosynthesis (BBN) and we find that the density parameter must have been extremely close to 1 then:

$$|\Omega(t_{\rm BBN}) - 1| \;=\; |\Omega_k(t_{\rm BBN})| \;\lesssim\; 10^{-16} \tag{4}$$

Thus we get as an initial condition to Big Bang, that Ω must have been initially extremely close to 1. The flatness problem is to explain why it was so. Otherwise, if we start the FRW universe in a radiation-dominated state with some initial value of the density parameter Ω_i not extremely close to 1, one of two things happens:

- $\Omega_i > 1 \quad \Rightarrow$ the universe recollapses almost immediately

[1] Guth[1] was not the first to propose a period of accelerating expansion in the early universe, but it was his proposal that became widely known and made inflation popular.

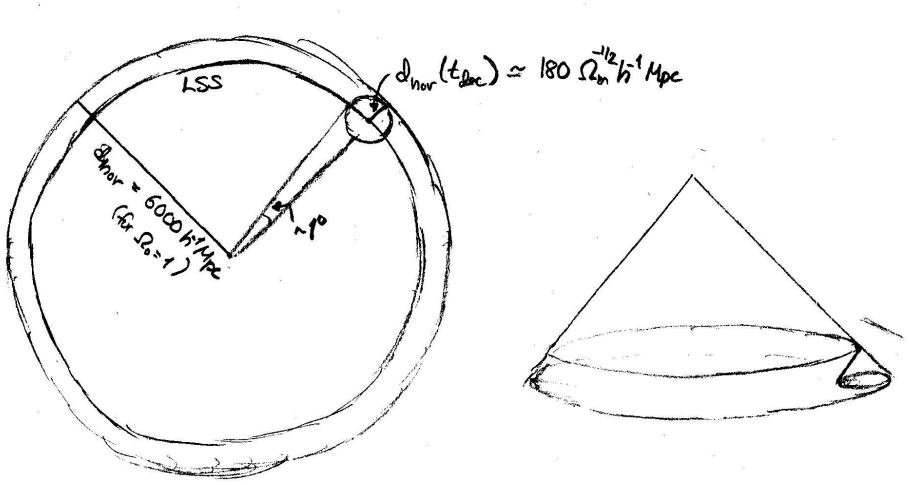


Figure 1: The horizon problem: regions on the CMB sky separated by more than about 1° had not had time to interact, yet their temperature is the same with an accuracy of $\lesssim 10^{-4}$.

- $\Omega_i < 1 \quad \Rightarrow$ the universe expands very fast and cools to $T < 3\,\mathrm{K}$ in a very short time.

Thus the flatness problem can also be called the oldness problem: why did it take so long, 14 billion years, for the universe to cool to $T = 2.7\,\mathrm{K}$.

Exercise: Oldness problem. Assume $\Omega_k(T = 100\,\mathrm{keV}) = 0.1$. (BBN takes place near $T = 100\,\mathrm{keV}$). Include just curvature and radiation (with $g_* = 3.384$) in the Friedmann equation. How long does it take for the universe to cool to $T = 2.7\,\mathrm{K}$? Why would inclusion of matter (with, say $\eta = 6\times10^{-10}$, and $\rho_m = 6\rho_b$) not change the answer?

7.1.2 Horizon problem

The horizon problem can also be called the homogeneity problem. The cosmic microwave background (CMB), which shows the universe at $z = 1090$ (age 370 000 years), is remarkably isotropic, the relative temperature variations being only $\mathcal{O}(10^{-4})$. This implies that density variations at that time must have been also very small, so the early universe was very homogeneous. Calculated according to the standard Hot Big Bang model, the horizon distance at that time was much smaller than the part of the early universe we see in the CMB, corresponding to only about 1° on the sky. Thus there could not have been any process to homogenize conditions over scales larger than this. This implies that this level of homogeneity must have been an initial condition.

Even the small CMB anisotropies show correlations at larger scales that 1°, a fact discovered after inflation was proposed.

7.1.3 Unwanted relics

If the Hot Big Bang begins at very high T it may produce objects surviving to the present, that are ruled out by observations.

- **Gravitino.** The supersymmetric partner of the graviton. $m \sim 100\,\mathrm{GeV}$. They interact very weakly (gravitational strength) $\Rightarrow$ they decay late, after BBN, and ruin the success of BBN.

- **Magnetic monopoles.** If the symmetry of a Grand Unified theory (GUT) is broken in a spontaneous symmetry breaking phase transition, magnetic monopoles are produced. These are point-like *topological defects* that are stable and very massive, $m \sim T_{\rm GUT} \sim 10^{14}\,$GeV. Their expected number density is such that their contribution to the energy density today $\gg$ the critical density.

- **Other topological defects** (cosmic strings, domain walls). These may also be produced in a GUT phase transition, and may also be a problem, but this is model-dependent. On the other hand, *cosmic strings* had been suggested as a possible explanation for the initial density perturbations—but this scenario fell later in trouble with the observational data (especially the anisotropy of the CMB).

These relics are produced very early, at extremely high temperatures, typically $T \gtrsim 10^{14}\,$GeV. From BBN, we only know that we should have standard Hot Big Bang for $T \lesssim 1\,$MeV.

7.1.4 What is needed

The word "problem" in the preceding is not to be taken to imply that the Hot Big Bang theory for the early universe would be in trouble. The theory by itself just does not contain answers to some questions one may pose about its initial conditions, for which we thus need additional ideas. We are perfectly happy if we can produce as an "initial condition" for Big Bang a universe with temperature $1\,{\rm MeV} < T < 10^{14}\,{\rm GeV}$, which is almost homogeneous and has $\Omega = 1$ with extremely high precision.

7.2 Inflation introduced

7.2.1 Accelerated expansion

Inflation is not a replacement for the Hot Big Bang, but an addition to it, occurring at very early times (e.g., $t \sim 10^{-35}$ s), without disturbing any of its successes. Thus we have first inflation, then Hot Big Bang; so that inflation produces the initial conditions for the Hot Big Bang.

The origin of the flatness problem is that $|\Omega - 1| = |K|/(aH)^2$ grows with time. Now

$$\frac{d}{dt}|\Omega - 1| \;=\; |K|\frac{d}{dt}\left(\frac{1}{a^2H^2}\right) \;=\; |K|\frac{d}{dt}\left(\frac{1}{\dot{a}^2}\right) \;=\; \frac{-2|K|}{\dot{a}^3}\ddot{a}\,. \tag{5}$$

For an expanding universe, $aH = a\left(\frac{\dot{a}}{a}\right) = \dot{a} > 0$. Thus $\dot{a}^3 > 0$, and

$$\frac{d}{dt}|\Omega - 1| > 0 \qquad \Leftrightarrow \qquad \ddot{a} < 0\,. \tag{6}$$

Thus the reason for the flatness problem is that the expansion of the universe is decelerating, i.e., slowing down. If we had an early period in the history of the universe, where the expansion was accelerating, it could make an initially arbitrary value of $|\Omega - 1| = |K|/(aH)^2$ very small.

Definition: Inflation = any epoch when the expansion is accelerating.

$$\text{Inflation} \;\Leftrightarrow\; \ddot{a} > 0 \tag{7}$$

Consider then the horizon problem. The horizon at photon decoupling, $d^p_{\rm hor}(t_{\rm dec})$ is somewhere between the radiation-dominated and matter-dominated values, H^{-1} and $2H^{-1}$. For comparing sizes of regions at different times, we should use their comoving sizes, $d^c \equiv d^p/a$. We have

$$d^c_{\rm hor}(t_{\rm dec}) \sim \frac{1}{a_{\rm dec}H_{\rm dec}}\,, \tag{8}$$

whereas the size of the observable universe today is of the order of the present Hubble length

$$d^c_{\rm hor}(t_0) \sim H_0^{-1}\,. \tag{9}$$

The horizon problem arises because the first is much smaller than the second,

$$\frac{d^c_{\rm hor}(t_{\rm dec})}{d^c_{\rm hor}(t_0)} \sim \frac{a_0 H_0}{a_{\rm dec} H_{\rm dec}} \ll 1\,. \tag{10}$$

Thus the problem is that aH, whose inverse gives roughly the comoving size of the horizon, decreases with time,

$$\frac{d}{dt}(aH) = \frac{d}{dt}(\dot a) = \ddot a < 0\,. \tag{11}$$

Having a period with $\ddot a > 0$ could solve the problem.

In the preceding we referred to the (comoving) horizon distance at some time t, defined as the comoving distance light has traveled from the beginning of the universe until time t. If there are no surprises at early times, we can calculate or estimate it; like in the preceding where we assumed radiation-dominated or matter-dominated behavior (standard Big Bang). If we now start adding other periods, like accelerating expansion at early times, the calculation of $d_{\rm hor}$ will depend on them. In principle, $d^c_{\rm hor}(t_0) > d^c_{\rm hor}(t_{\rm dec})$ always, since $t_0 > t_{\rm dec}$, so $(0, t_{\rm dec}) \subset (0, t_0)$. But note that in the horizon problem, the relevant present horizon is how far we can *see*: the observable universe is given just by the integrated comoving distance the photon has traveled in the interval $(t_{\rm dec}, t_0)$, which is not affected by what happens before $t_{\rm dec}$. Thus the relevant present horizon is still $\sim H_0^{-1}$.

What is the relation between $d^c_{\rm hor}(t)$ and $1/(aH)$ for arbitrary expansion laws? Introduce the comoving, or conformal, Hubble parameter,

$$\mathcal{H} \equiv aH = \frac{1}{a}\frac{da}{d\eta} \equiv \dot a\,, \tag{12}$$

where η is the conformal time, defined by $d\eta = dt/a$. The Hubble length is

$$l_H \equiv H^{-1}\,, \qquad \text{where } H \equiv \frac{\dot a}{a}\,, \tag{13}$$

and the *comoving Hubble length* is

$$l^c_H \equiv \frac{l_H}{a} = \frac{1}{aH} = \frac{1}{\dot a} = \mathcal{H}^{-1}\,. \tag{14}$$

Roughly speaking, $\mathcal{H}^{-1}$ gives the comoving distance light travels in a "cosmological timescale", i.e., the Hubble time. This statement cannot be exact, since both the comoving Hubble length and the Hubble time change with time. However, if $\mathcal{H}^{-1}$ is increasing with time, the comoving distances traveled at earlier "epochs" are shorter, and thus $\mathcal{H}^{-1}(t)$ is a good estimate for the total comoving distance light has traveled since the beginning of time (the horizon). On the other hand, if $\mathcal{H}^{-1}$ is shrinking, then at earlier epochs light was traveling longer comoving distances, and we expect the horizon at time t to be larger than $\mathcal{H}^{-1}$. In any case

$$d^c_{\rm hor}(t) \gtrsim \mathcal{H}(t)^{-1}\,. \tag{15}$$

Since the Hubble length is more easily "accessible" (less information needed to figure it out) than the horizon distance it has become customary in cosmology to use the word "horizon" also for the Hubble distance. We shall also adopt this practice. The Hubble length gives the distance over which we have causal interaction in cosmological timescales. The comoving Hubble length gives this distance in comoving units.

If aH is decreasing (Eq. 11) then $\mathcal{H}^{-1}$ increases, and vice versa.

$\therefore$ Inflation = any epoch when the comoving Hubble length is shrinking.

$$\text{Inflation} \;\Leftrightarrow\; \frac{d}{dt}\mathcal{H}^{-1} < 0 \tag{16}$$

Thus the comoving distance over which we have causal connection is *decreasing* during inflation: causal contact to other parts of the Universe is being lost.

Inflation can be discussed either 1) in terms of physical distances or 2) in terms of comoving distances.

1) In terms of physical distance, the distance between any two points in the Universe is increasing, with an accelerating rate. The distance over which causal connection can be maintained is increasing (much) more slowly.

2) In terms of comoving distance, i.e., viewed in comoving coordinates, the distance between two points stays fixed; regions of the universe corresponding to present structures maintain fixed size. From this viewpoint (the one normally adopted), the region causally connected to a given location in the Universe is shrinking.

To connect with dynamics, look at the second Friedmann equation,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \tag{17}$$

$$\therefore \text{Inflation} \;\Leftrightarrow\; \rho + 3p < 0 \;\Leftrightarrow\; w < -\tfrac{1}{3} \tag{18}$$

Thus inflation requires negative pressure, $p < -\frac{1}{3}\rho$ (we assume $\rho \geq 0$).

There is a huge class of models to realize the inflation scenario. These models rely on so-far-unknown physics of very high energies. Some models are just "toy models", with a hoped-for resemblance to the actual physics of the early universe. Others are connected to proposed extensions (like supersymmetry) to the standard model of particle physics.

The important point is that inflation makes many *generic*[2] predictions, i.e., predictions that are independent of the particular model of inflation. Present observational data agrees with these predictions. Thus it is widely believed—or considered probable—by cosmologists that inflation indeed took place in the very early universe. There are also numerical predictions of cosmological observables that differ from one model of inflation to another, allowing future observations to rule out classes of such models. (Many inflation models are already ruled out.)

7.2.2 Solving the problems

Inflation can solve[3] all the problems discussed in Sec. 7.1. The idea is that during inflation the universe expands by a large factor (at least by a linear factor of something like $\sim e^{70} \sim 10^{30}$ to solve the problems). This cools the universe to $T \sim 0$ (if the concept of temperature is applicable). When inflation ends, the universe is heated to a high temperature, and the usual Hot Big Bang history follows. This heating at the end of inflation is called *reheating*, since originally the thinking was that inflation started at an earlier hot epoch, but it is actually not clear whether that was the case.

[2] I was once in a conference where a speaker began his talk on inflation by promising not to use the words "generic" or "scenario". He failed in one but not the other.

[3] This is not to be taken too rigorously. The problems are related to the question of initial conditions of the universe at some very early time, whose physics we do not understand. Thus theorists are free to have different views on what kind of initial conditions are "natural". Inflation makes the flatness and horizon problems "exponentially smaller" in some sense, but inflation still places requirements—on the level of homogeneity—for the initial conditions *before* inflation, so that inflation can begin.

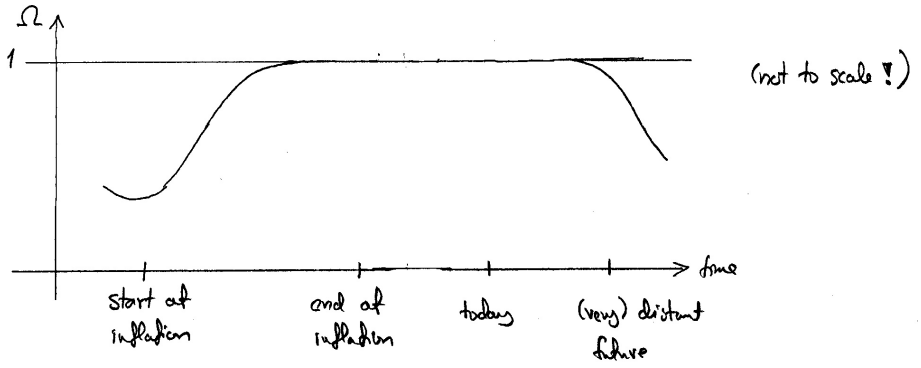


Figure 2: Solving the flatness problem. This figure is for a universe with no dark energy, where the expansion keeps decelerating after inflation ended in the early universe. Present observational evidence indicates that actually the expansion began accelerating again (supposedly due to the mysterious dark energy) a few billion years ago. Thus the universe is, technically speaking, inflating again, and Ω is again being driven towards 1. However, this current epoch of inflation is not enough to solve the flatness problem, or the other problems, since the universe has only expanded by about a factor of 2 during it.

Solving the flatness problem: The flatness problem is solved, since during inflation

$$|1-\Omega| = \frac{|K|}{\mathcal{H}^2} \quad \text{is shrinking.} \tag{19}$$

Thus inflation drives $\Omega \to 1$. Starting with an arbitrary Ω, inflation drives $|1-\Omega|$ so small that, although it has grown all the time from the end of inflation to the recent onset of dark energy domination, it is still very small today. See Fig. 2. In fact, inflation predicts that $\Omega_0 = 1$ to high accuracy, since it would be an unnatural coincidence for inflation to last just the right amount so that Ω would begin to deviate from 1 just at the current epoch.[4]

Solving the horizon problem: The horizon problem is solved, since during inflation the causally connected region is shrinking. It was very large before inflation; much larger than the present horizon. Thus the present observable universe has evolved from a small patch of a much larger causally connected region; and it is natural that the conditions were (or became) homogeneous in that patch then. See Fig. 3.

Getting rid of relics: If unwanted relics are produced before inflation, they are diluted to practically zero density by the huge expansion during inflation. We just have to take care they are not produced after inflation, i.e., the reheating temperature has to be low enough. This is an important constraint on models of inflation.

Did we really solve the problem of initial conditions? Actually solving the flatness and horizon problems is more complicated. We discussed them in terms of a FRW universe, which by assumption is already homogeneous. In fact, for inflation to get started, a sufficiently large region which is not too inhomogeneous and not too curved, is needed. We shall not discuss this in more detail, since the solution of these problems is not the most important aspect of inflation.

If inflation happened, we expect that the early universe after inflation was very homogeneous except for fluctuations generated during inflation and that $\Omega_0 = 1$. Thus inflation leads to predictions that can be tested with observations. More important than flatness and homogeneity

[4]Thus, if it were discovered by observations, that actually $\Omega_0 \neq 1$, this would be a blow to the credibility of inflation. However, there is a version of inflation, called *open inflation*, for which it is natural that $\Omega_0 < 1$. The existence of such models of inflation have led critics of inflation to complain that inflation is "unfalsifiable"—no matter what the observation, there is a model of inflation that agrees with it. Nevertheless, most models of inflation give the same "generic" predictions, including $\Omega_0 = 1$.

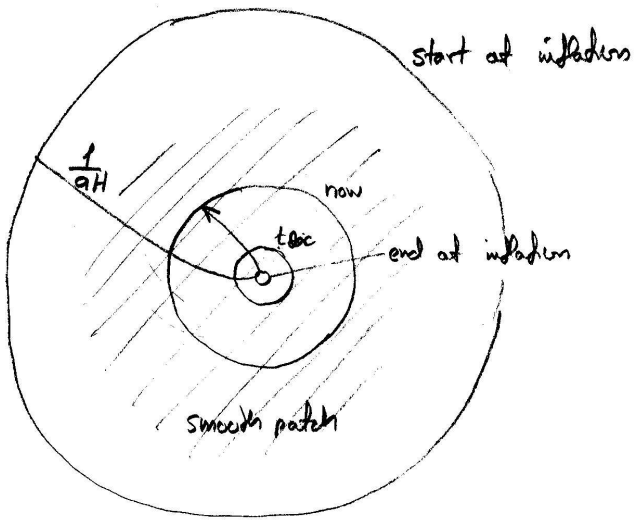


Figure 3: Evolution of the comoving Hubble radius (length, distance) during and after inflation (schematic).

are the predictions inflation makes about primordial perturbations, the "seeds" for structure formation, discussed in the next chapter.

Thus we assume that sufficient inflation has already taken place to make the universe (within a horizon volume) flat and homogeneous, and follow the inflation in detail after that, working in the flat FRW universe.

7.3 Quantum field theory for children

The theories (known and hypothetical) needed to describe the (very) early universe are *quantum field theories* (QFT). The fundamental entities of these theories are *fields*, i.e., functions of space and time. For each particle species, there is a corresponding field, having at least as many (real) components φ_i as the particle has internal degrees of freedom. For example, for the photon, the corresponding field is the vector field $A^\mu = (A^0, A^1, A^2, A^3) = (\phi, \vec{A})$, which you are probably familiar with from electrodynamics.[5] The photon has two internal degrees of freedom. The larger number of components in A^μ is related to the *gauge freedom* of electrodynamics. Since A^μ is a (Lorentz) vector field, it has the same number of components as there are spacetime dimensions, but other types of fields do not have this correspondence.

In classical field theory the evolution of the field is governed by the *field equation*. From the field equation one can identify a field potential, an expression in terms of the field, which helps to understand the field dynamics. Quantizing a field theory gives a quantum field theory. *Particles are quanta of the oscillations of the field around the minimum of its potential.* The state where the field values are constant at the potential minimum is called the *vacuum*. Up to now, we have described the events in the early universe in terms of the *particle picture*. However, the particle picture is not fundamental, and can be used only when the fields are doing small oscillations. For many possible events and objects in the early universe (inflation, topological defects, spontaneous symmetry breaking phase transitions) the field behavior is different, and we need to describe them in terms of field theory. In some of these topics classical field theory is already sufficient for a reasonable and useful description.

In this section we discuss "low-temperature" field theory in Minkowski space, i.e., we forget high-temperature effects and the curvature of spacetime.

The starting point in field theory is the *Lagrangian density*, a function of space and time, which is a scalar quantity constructed from the fields and their derivatives:

$$\mathcal{L}(\varphi_i, \partial_\mu \varphi_i)\,. \tag{20}$$

The Lagrangian density can be expressed as a sum of two parts, the *kinetic term*, which depends on field derivatives (gradients), and the *field potential* $V(\varphi_1, \ldots, \varphi_N)$ (for a theory with N fields), which does not. This expression for the Lagrangian density as a function of the fields and their derivatives defines the field theory, and one can derive the field equations (differential equations governing the field evolution) and the energy-momentum tensor (energy density and pressure of the fields) from the Lagrangian density. Usually the kinetic term has a simple form, called the canonical kinetic term, and we assume that here. The remaining freedom in defining the field theory is in defining the potential.

The simplest case is a theory with one scalar field φ, for which

$$\mathcal{L} \;=\; -\tfrac{1}{2}\partial_\mu\varphi\,\partial^\mu\varphi - V(\varphi)\,. \tag{21}$$

(We use the Einstein summation convention, where a repeated index implies summation over it, here $\mu = 0, 1, 2, 3$. Also $\partial_\mu \equiv \partial/\partial x^\mu$, where $x^0 = t$. Here we are in Minkowski space and use Cartesian coordinates, so that $\partial^0 = -\partial_0$ and $\partial^j = \partial_j$ for $j = 1, 2, 3$.) We write

$$V'(\varphi) \equiv \frac{dV}{d\varphi} \qquad \text{and} \qquad V''(\varphi) \equiv \frac{d^2V}{d\varphi^2}\,. \tag{22}$$

The *field equation*, which determines the classical evolution of the field, is obtained from the Lagrangian by minimizing (or extremizing) the action

$$\int \mathcal{L}\, d^4x\,, \tag{23}$$

[5] ϕ and $\vec{A}$ are the electromagnetic scalar and vector potentials. This is a different use of the word "potential" than what we use here. In our terminology, $A^\mu = (\phi, \vec{A})$ are fields.

which leads to the *Euler–Lagrange equation*

$$\frac{\partial \mathcal{L}}{\partial \varphi_i(x)} - \partial_\mu \frac{\partial \mathcal{L}}{\partial[\partial_\mu \varphi_i(x)]} = 0\,. \tag{24}$$

For the above scalar field we get the field equation

$$\partial_\mu \partial^\mu \varphi - V'(\varphi) = 0\,. \tag{25}$$

where $\partial_\mu \partial^\mu \varphi = -\ddot{\varphi} + \nabla^2 \varphi$, so that the field equation is

$$\boxed{\ddot{\varphi} - \nabla^2 \varphi \;=\; -V'(\varphi)\,.} \tag{26}$$

Here we use the overdot to denote partial derivative with respect to time: $\dot{} = \partial_0 = \partial/\partial t$.

The Lagrangian also gives us the energy tensor

$$T^{\mu\nu} = -\frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi)} \partial^\nu \varphi + g^{\mu\nu} \mathcal{L}\,. \tag{27}$$

For the scalar field

$$T^{\mu\nu} = \partial^\mu \varphi \partial^\nu \varphi - g^{\mu\nu} \left[\frac{1}{2} \partial_\rho \varphi \partial^\rho \varphi + V(\varphi)\right]. \tag{28}$$

In particular, the energy density $\rho = T^{00}$ and pressure $p = \frac{1}{3}(T^{11} + T^{22} + T^{33})$ of a scalar field are

$$\rho \;=\; \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}(\nabla\varphi)^2 + V(\varphi) \tag{29}$$

$$p \;=\; \frac{1}{2}\dot{\varphi}^2 - \frac{1}{6}(\nabla\varphi)^2 - V(\varphi)\,. \tag{30}$$

(We are in Minkowski space, so that $g^{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$). We see that the pressure due to a scalar field may be negative. The minimum value of $V(\varphi)$ is the *vacuum energy.* In principle it could be negative, acting like a negative cosmological constant. Any other contribution to ρ is positive. Since there is no evidence for a negative vacuum energy or cosmological constant, let us assume that $V(\varphi) \geq 0$.

If $V(\phi) \equiv const$ (typically assumed to be 0) the field equation becomes the wave equation

$$\ddot{\varphi} \;=\; \nabla^2 \varphi\,, \tag{31}$$

whose solutions are waves propagating at the speed of light.

For the corresponding quantum theory, the potential gives information about the masses and interactions of the particles that are the quanta of the field oscillations. The particles corresponding to scalar fields are spin-0 bosons. Spin-$\frac{1}{2}$ particles correspond to spinor fields and spin-1 particles to vector fields. The case $V(\varphi) \equiv 0$ corresponds to massless noninteracting particles. If the potential has the form

$$V(\varphi) = \frac{1}{2} m^2 \varphi^2\,, \tag{32}$$

the particle corresponding to the field φ will have mass m and it will have no interactions. In general, the mass of the particle is given by $m^2 = V''(\varphi)$.

Interactions between particles of two different species are due to terms in the Lagrangian which involve both fields. For example, in the Lagrangian of quantum electrodynamics (QED) the term

$$-ie\psi^\dagger \gamma^0 \gamma^\mu A_\mu \psi \tag{33}$$

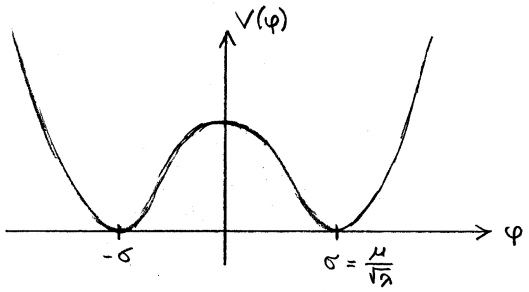
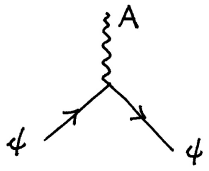


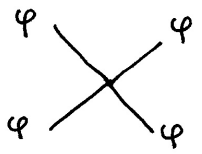
Figure 4: Potential giving rise to spontaneous symmetry breaking.

is responsible for the interaction between photons (A^μ) and electrons (ψ). (The γ^μ are Dirac matrices). A graphical representation of this interaction is the Feynman diagram



A higher power (third or fourth) of a field, e.g.,

$$V(\varphi) = \frac{1}{4}\lambda\varphi^4\,, \tag{34}$$



represents self-interaction, i.e., φ particles interacting with each other directly (as opposed to, e.g., electrons, who interact with each other indirectly, via photons). In QCD, gluons have this property.

Some theories exhibit *spontaneous symmetry breaking* (SSB). For example, the potential

$$V(\varphi) = V_0 - \frac{1}{2}\mu^2\varphi^2 + \frac{1}{4}\lambda\varphi^4 \tag{35}$$

has two minima, at $\varphi = \pm\sigma$, where $\sigma = \mu/\sqrt{\lambda}$. At low temperatures, the field is doing small oscillations around one of these two minima (see Fig. 4). Thus the vacuum value of the field is nonzero. If the Lagrangian has interaction terms, $c\varphi\psi^2$, with other fields ψ, these can now be separated into a mass term, $c\sigma\psi^2$, for ψ and an interaction term, by redefining the field φ as

$$\varphi = \sigma + \tilde{\varphi} \quad \Rightarrow \quad c\varphi\psi^2 = c\sigma\psi^2 + c\tilde{\varphi}\psi^2\,. \tag{36}$$

Thus spontaneous symmetry breaking gives the ψ particles a mass $\sqrt{2c\sigma}$. This kind of a field φ is called a *Higgs field*. In electroweak theory the fermion masses[6] are due to a Higgs field.

[6]The origin of neutrino masses is not clear.

7.4 Inflaton field

As we saw in Sec. 7.2, inflation requires negative pressure. In Chapter 4 we considered systems of particles where interaction energies can be neglected (ideal gas approximation). For such systems the pressure is always nonnegative. However, negative pressure is possible in systems with attractive interactions. In the field picture, negative pressure comes from the potential term. In many models of inflation, the inflation is caused by a scalar field. This scalar field (and the corresponding spin-0 particle) is called the *inflaton*.

Historical note. The idea of scalar fields playing an important role in the very early universe was very natural at the time inflation was proposed by Guth[1]. We already mentioned how a scalar field, the Higgs field, is responsible for the electroweak phase transition at $T \sim 100\,\mathrm{GeV}$. It is thought that at a much higher temperature, $T \sim 10^{14}\,\mathrm{GeV}$, another spontaneous symmetry breaking phase transition occurred, the GUT (Grand Unified Theory) phase transition, so that above this temperature the strong and electroweak forces were unified. This GUT phase transition gives rise to the monopole problem. Guth realized that the Higgs field associated with the GUT transition might lead the universe to "inflate" (the term was coined by Guth), solving this monopole problem. It was soon found out, however, that inflation based on the GUT Higgs field is not a viable inflation model, since in this model too strong inhomogeneities were created. So the inflaton field must be some other scalar field. The supersymmetric extensions of the standard model contain many inflaton field candidates.

During inflation the inflaton field is *almost* homogeneous.[7] The energy density and pressure of the inflaton are thus those of a homogeneous scalar field,

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\varphi}^2 + V(\varphi) \\ p &= \frac{1}{2}\dot{\varphi}^2 - V(\varphi), \end{aligned} \tag{37}$$

where $V(\varphi) \geq 0$. For the equation-of-state parameter $w \equiv p/\rho$ we have

$$w = \frac{\dot{\varphi}^2 - 2V(\varphi)}{\dot{\varphi}^2 + 2V(\dot{\varphi})} = \frac{1 - (2V/\dot{\varphi}^2)}{1 + (2V/\dot{\varphi}^2)}, \tag{38}$$

so that

$$-1 \leq w \leq 1\,. \tag{39}$$

If the kinetic term $\dot{\varphi}^2$ dominates, $w \approx 1$; if the potential term $V(\varphi)$ dominates, $w \approx -1$.

For the present discussion, the potential $V(\varphi)$ is some arbitrary non-negative function. Different inflaton models correspond to different $V(\varphi)$. From Eq. (37), we get the useful combinations

$$\begin{aligned} \rho + p &= \dot{\varphi}^2 \\ \rho + 3p &= 2\left[\dot{\varphi}^2 - V(\varphi)\right]\,. \end{aligned} \tag{40}$$

We already had the field equation for a scalar field in Minkowski space,

$$\ddot{\varphi} - \nabla^2\varphi = -V'(\varphi)\,. \tag{41}$$

For the homogenous case it is just

$$\ddot{\varphi} = -V'(\varphi)\,. \tag{42}$$

We get a working mental picture of the evolution of a homogeneous field by comparing it to the classical mechanics equation for a particle in a gravitational potential $V(\mathbf{r})$ whose accelaration

[7] Inflation makes the inflaton field homogenous. Again, a sufficient level of initial homogeneity of the field is required to get inflation started. We start our discussion when a sufficient level of inflation has already taken place to make the gradients negligible.

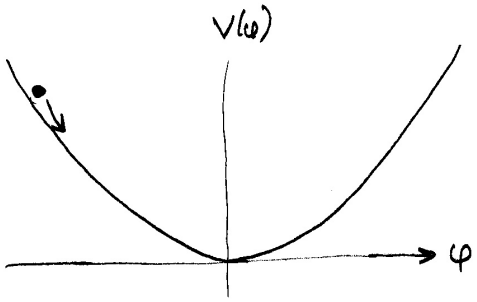


Figure 5: The inflaton and its potential.

is given by $\ddot{\mathbf{r}} = -\nabla V(\mathbf{r})$. Thus we can think of the field "rolling down" its potential like a stone rolling down a hillside, see Fig. 5; this motion is governed by Eq. (42).

We need to modify (42) for the expanding universe. We do not need to go to the GR formulation of field theory, since the modification for the present case can be simply obtained by sticking the ρ and p from Eq. (37) into the energy continuity equation

$$\dot{\rho} = -3H(\rho + p)\,. \tag{43}$$

This gives

$$\dot{\varphi}\ddot{\varphi} + V'(\varphi)\dot{\varphi} = -3H\dot{\varphi}^2 \;\Rightarrow\; \boxed{\ddot{\varphi} + 3H\dot{\varphi} = -V'(\varphi)}\,, \tag{44}$$

the field equation for a homogeneous φ in an expanding (FRW) universe. We see that the effect of expansion is to add the term $3H\dot{\varphi}$, which acts like a *friction term*, slowing down the evolution of φ.

The condition for inflation, $\rho + 3p = 2\dot{\varphi}^2 - 2V(\dot{\varphi}) < 0$, is satisfied, if

$$\dot{\varphi}^2 < V(\varphi)\,. \tag{45}$$

The idea of inflation is that φ is initially far from the minimum of $V(\varphi)$. The potential then pulls φ towards the minimum. See Fig. 5. If the potential has a suitable (sufficiently flat) shape, the friction term soon makes $\dot{\varphi}$ small enough to satisfy Eq. (45), even if it was not satisfied initially.

We shall also need the Friedmann equation for the flat universe,

$$H^2 = \frac{8\pi G}{3}\rho = \frac{1}{3M_{\rm Pl}^2}\rho\,. \tag{46}$$

where we have introduced the *reduced Planck mass*

$$M_{\rm Pl} \equiv \frac{1}{\sqrt{8\pi}} m_{\rm Pl} \equiv \frac{1}{\sqrt{8\pi G}} = 2.4353 \times 10^{18}\,{\rm GeV}\,. \tag{47}$$

Inserting Eq. (37), this becomes

$$\boxed{H^2 = \frac{1}{3M_{\rm Pl}^2}\left[\frac{1}{2}\dot{\varphi}^2 + V(\varphi)\right]}\,. \tag{48}$$

We have ignored other components to energy density and pressure besides the inflaton. During inflation, the inflaton φ moves slowly, so that the inflaton energy density, which is dominated by $V(\varphi)$ also changes slowly. If there are matter and radiation components to the energy density, they decrease fast, $\rho \propto a^{-3}$ or $\propto a^{-4}$ and soon become negligible. Again, this puts some initial conditions for inflation to get started, for the inflaton to become dominant. But once inflation gets started, we can soon forget the other components to the universe besides the inflaton.

7.5 Slow-roll inflation

The friction (expansion) term tends to slow down the evolution of φ, so that we may easily reach a situation where:

$$\dot{\varphi}^2 \ll V(\varphi) \tag{49}$$

$$|\ddot{\varphi}| \ll |3H\dot{\varphi}| \tag{50}$$

These are the *slow-roll conditions*.

If the slow-roll conditions are satisfied, we may approximate (the *slow-roll approximation*) Eqs. (48) and (44) by the *slow-roll equations*:

$$H^2 = \frac{V(\varphi)}{3M_{\rm Pl}^2} \tag{51}$$

$$3H\dot{\varphi} = -V'(\varphi) \tag{52}$$

The shape of the potential $V(\varphi)$ determines the *slow-roll parameters*:

$$\begin{aligned} \varepsilon(\varphi) &\equiv \frac{1}{2}M_{\rm Pl}^2\left(\frac{V'}{V}\right)^2 \\ \eta(\varphi) &\equiv M_{\rm Pl}^2\frac{V''}{V} \end{aligned} \tag{53}$$

Exercise: Show that

$$\varepsilon \ll 1 \quad \text{and} \quad |\eta| \ll 1 \qquad \Leftarrow \qquad \text{Eqs. (49) and (50)} \tag{54}$$

Note that the implication goes only in this direction. The conditions $\varepsilon \ll 1$ and $|\eta| \ll 1$ are necessary, but not sufficient for the slow-roll approximation to be valid (i.e., the slow-roll conditions to be satisfied).

The conditions $\varepsilon \ll 1$ and $|\eta| \ll 1$ are just *conditions on the shape of the potential*, and identify from the potential a *slow-roll section*, where the slow-roll approximation *may* be valid. Since the initial field equation, Eq. (44) was second order, it accepts arbitrary φ and $\dot{\varphi}$ as initial conditions. Thus Eqs. (49) and (50) may not hold initially, even if φ is in the slow-roll section. However, it turns out that the *slow-roll solution*, the solution of the slow-roll equations (51) and (52), is an *attractor* of the full equations, (48) and (44). This means that the solution of the full equations rapidly approaches it, starting from arbitrary initial conditions. Well, not fully arbitrary, the initial conditions need to lie in the *basin of attraction*, from which they are then attracted into the attractor. To be in the basin of attraction, means that φ must be in the slow-roll section, and that if $\dot{\varphi}$ is very large, φ needs to be deeper in the slow-roll section.

Once we have reached the attractor, where Eqs. (51) and (52) hold, $\dot{\varphi}$ is determined by φ (since we replaced the second-order differential equation with a first-order one). In fact everything is determined by φ (assuming a known form of $V(\varphi)$). The value of φ is the *single parameter describing the state of the universe*, and φ evolves down the potential $V(\varphi)$ as specified by the slow-roll equations.

This language of "attractor" and "basin of attraction" can be taken further. **If** the universe (or a region of it) finds itself initially (or enters) the basin of attraction of slow-roll inflation, meaning that: there is a sufficiently large region, where the curvature is sufficiently small, the inflaton makes a sufficient contribution to the total energy density, the inflaton is sufficiently homogeneous, and lies sufficiently deep in the slow-roll section, **then** this region begins inflating,

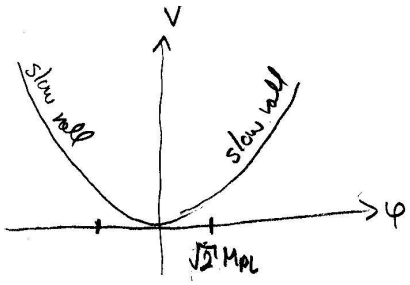


Figure 6: The potential $V(\varphi) = \frac{1}{2}m^2\varphi^2$ and its two slow-roll sections.

it becomes rapidly very homogeneous and flat, all other contributions to the energy density besides the inflaton become negligible, and the inflaton begins to follow the slow-roll solution.

Thus inflation *erases all memory of the initial conditions*, and we can predict the later history of the universe just from the shape of $V(\varphi)$ and the assumption that φ started out far enough in the slow-roll part of it.

Example: The simplest model of inflation (see Fig. 6) is the one where

$$V(\varphi) = \frac{1}{2}m^2\varphi^2 \quad \Rightarrow \quad V'(\varphi) = m^2\varphi\,, \quad V''(\varphi) = m^2\,. \tag{55}$$

The slow-roll parameters are

$$\left.\begin{aligned} \varepsilon(\varphi) &= \frac{1}{2}M_{\rm Pl}^2\left(\frac{2}{\varphi}\right)^2 \\ \eta(\varphi) &= M_{\rm Pl}^2\frac{2}{\varphi^2} \end{aligned}\right\} \quad \Rightarrow \quad \varepsilon = \eta = 2\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \tag{56}$$

and the slow-roll section of the potential is given by the condition

$$\varepsilon, \eta \ll 1 \quad \Rightarrow \quad \varphi^2 \gg 2M_{\rm Pl}^2\,. \tag{57}$$

7.5.1 Relation between inflation and slow roll

From the definition of the Hubble parameter,

$$H = \frac{\dot{a}}{a} \quad \Rightarrow \quad \dot{H} = \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} \quad \Rightarrow \quad \boxed{\frac{\ddot{a}}{a} = \dot{H} + H^2} \tag{58}$$

Thus the condition for inflation is $\dot{H} + H^2 > 0$. This would be satisfied, if $\dot{H} > 0$, but this is not possible here, since it would require $p < -\rho$, i.e., $w \equiv p/\rho < -1$, which is not allowed by Eq. (37).[8] Thus

$$\boxed{\dot{H} \leq 0} \tag{59}$$

and

$$\text{Inflation} \;\Leftrightarrow\; -\frac{\dot{H}}{H^2} < 1 \tag{60}$$

[8] From the Friedmann eqs.,

$$\left.\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{K}{a^2} \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p) \end{aligned}\right\} \quad \Rightarrow \quad \dot{H} \;=\; \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} \;=\; -4\pi G(\rho + p) + \frac{K}{a^2}$$

In the above, we are assuming space is already flat, i.e., $K = 0$. Then $\dot{H} > 0 \;\Rightarrow\; \rho + p < 0$.

If the slow-roll approximation is valid,

$$H^2 = \frac{V}{3M_{\rm Pl}^2} \quad \Rightarrow \quad 2H\dot{H} = \frac{V'\dot{\varphi}}{3M_{\rm Pl}^2} \quad \Rightarrow \quad H^2\dot{H} = \frac{V'H\dot{\varphi}}{6M_{\rm Pl}^2} \overset{3H\dot{\varphi}=-V'}{=} -\frac{V'^2}{18M_{\rm Pl}^2}$$
$$\Rightarrow \quad -\frac{\dot{H}}{H^2} = \frac{V'^2}{18M_{\rm Pl}^2}\frac{9M_{\rm Pl}^4}{V^2} = \frac{1}{2}M_{\rm Pl}^2\left(\frac{V'}{V}\right)^2 = \varepsilon \ll 1$$

Therefore, *if the slow-roll approximation is valid, inflation is guaranteed*. This is a sufficient, not necessary condition. The above result for slow-roll inflation, $-\dot{H}/H^2 \ll 1$ can also be written as

$$\left|\frac{\dot{H}}{H}\right| \ll \frac{\dot{a}}{a}\,. \tag{61}$$

During slow-roll inflation, the Hubble parameter H changes much more slowly than the scale factor a. For a constant H, the universe expands exponentially, since

$$\frac{\dot{a}}{a} = \frac{d\ln a}{dt} = H = const \quad \Rightarrow \quad \ln\frac{a}{a_1} = H(t-t_1) \quad \Rightarrow \quad a \propto e^{Ht}\,. \tag{62}$$

Thus, in slow-roll inflation, the universe expands "almost exponentially".

Note that accelerated expansion, which is defined to mean that $\ddot{a} > 0$, does not mean that the *expansion rate*, as given by H, would increase. Even during inflation, $\dot{H} < 0$, so the expansion rate decreases. (There may be some ambiguity in what is meant by an increasing/decreasing expansion rate. The Hubble parameter is a better quantity to be called the expansion rate than $\dot{a}$, since the value of the latter depends on the normalization of a_0. With the normalization $a_0 = 1$, $H = \dot{a}$ "today".)

Sometimes it is carelessly said that inflation was a period of very rapid expansion. Rapid compared to what? Certainly the expansion rate was larger than today, or indeed larger than during any period after inflation (since $\dot{H} < 0$ always). But note that in the original Hot Big Bang picture $H \to \infty$ (and also $\dot{a} \to \infty$) as $t \to 0$. When we replace the earliest part of Hot Big Bang with inflation, we replace it with *slower* expansion, H almost constant (and $\dot{a}$ becoming smaller towards earlier times—this is what acceleration means), instead of $H \to \infty$.

It is possible to have inflation without the slow-roll parameters being small (fast-roll inflation), but we will see that slow-roll inflation produces the observed primordial perturbation spectrum naturally (unlike fast-roll inflation).

7.5.2 Models of inflation

A model of inflation[9] consists of

1. a potential $V(\varphi)$
2. a way of ending inflation

There are two ways of ending inflation:

1. Slow-roll approximation is no more valid, as φ approaches the minimum of the potential with $V(\varphi_{\rm min}) = 0$ or very small. For a reasonable approximation we can assume inflation ends, when $\varepsilon(\varphi) = 1$ or $|\eta(\varphi)| = 1$. Denote this value of the inflaton field by $\varphi_{\rm end}$.
2. Extra physics intervenes to end inflation (e.g., *hybrid* inflation). In this case inflation may end while the slow-roll approximation is valid.

[9] There are also models of inflation which are not based on a scalar field.

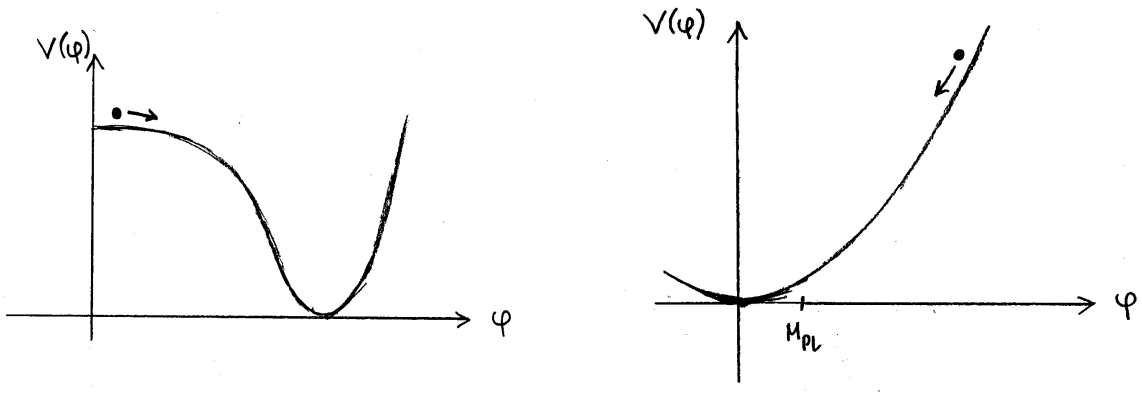


Figure 7: Potential for small-field (a) and large-field (b) inflation. For a typical small-field model, the entire range of φ shown is $\ll M_{\rm Pl}$.

Inflation models can be divided into two classes:

1. small-field inflation, $\Delta\varphi < M_{\rm Pl}$ in the slow-roll section
2. large-field inflation, $\Delta\varphi > M_{\rm Pl}$ in the slow-roll section

Here $\Delta\varphi$ is the range in which φ varies during (the observationally relevant part of) inflation. See Fig. 7 for typical shapes of potentials for large-field and small-field models.

Example of small-field inflation:

$$V(\varphi) = V_0 \left[1 - \frac{\lambda}{4}\left(\frac{\varphi}{M_{\rm Pl}}\right)^4 + \dots\right], \tag{63}$$

where the omitted terms, responsible for keeping $V \geq 0$ for larger φ are assumed negligible in the region of interest. We assume further that the second term is small in the slow-roll section, so that we can approximate $V(\varphi) \approx V_0$ except for its derivatives. The slow-roll parameters are then

$$\varepsilon = \frac{1}{2}\lambda^2 \left(\frac{\varphi}{M_{\rm Pl}}\right)^6 \qquad \text{and} \qquad \eta = -3\lambda \left(\frac{\varphi}{M_{\rm Pl}}\right)^2, \tag{64}$$

so that

$$\frac{\varepsilon}{|\eta|} = \frac{1}{6}\lambda \left(\frac{\varphi}{M_{\rm Pl}}\right)^4 \ll 1\,. \tag{65}$$

Thus $\eta < 0$ and $\varepsilon \ll |\eta|$, which is typical for small-field inflation, and inflation ends when

$$|\eta| = 1 \quad \Rightarrow \quad \varphi_{\rm end} = \frac{M_{\rm Pl}}{\sqrt{3\lambda}}\,. \tag{66}$$

The assumption that the second term in the potential is still small at $\varphi_{\rm end}$, requires that $\lambda \gtrsim 1$, and thus $|\eta| \ll 1$ requires $\varphi \ll M_{\rm Pl}/\sqrt{3}$, so this is indeed a small-field model.

Example of large-field inflation: A simple monomial potential of the form

$$V(\varphi) = A\varphi^n \qquad (n > 1)\,. \tag{67}$$

The slow-roll parameters are

$$\varepsilon = \frac{n^2}{2}\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \qquad \text{and} \qquad \eta = n(n-1)\left(\frac{M_{\rm Pl}}{\varphi}\right)^2, \tag{68}$$

so that $\eta > 0$ and ε and η are of similar size, typical for large-field inflation. This is a large-field model, since $\varepsilon \ll 1$ requires $\varphi^2 \gg \frac{1}{2}n^2 M_{\rm Pl}^2$.

For the special case of $V(\varphi) = \frac{1}{2}m^2\varphi^2$, $\varepsilon = \eta$, and inflation ends at $\varphi_{\rm end} = \sqrt{2}M_{\rm Pl}$. To get inflation to end, e.g., at energy scale $V(\varphi_{\rm end}) \equiv m^2 M_{\rm Pl}^2 = \left(10^{14}\,{\rm GeV}\right)^4$, we need $m = \left(10^{14}\,{\rm GeV}\right)^2/M_{\rm Pl} \approx 4 \times 10^9\,{\rm GeV}$.

7.5.3 Exact solutions

Usually the slow-roll approximation is sufficient. It fails near the end of inflation, but this just affects slightly our estimate of the total amount of inflation. It is much easier to solve the slow-roll equations, (51) and (52), than the full equations, (44) and (48). However, it is useful to have some exact solutions to the full equations, for comparison. For some special cases, exact analytical solutions exist.

One such case is *power-law inflation*, where the potential is

$$V(\varphi) = V_0 \exp\left(-\sqrt{\frac{2}{p}}\frac{\varphi}{M_{\rm Pl}}\right), \qquad p > 1\,, \tag{69}$$

where V_0 and p are constants.

An exact solution for the full equations, (44) and (48), is (**exercise**)

$$a(t) \propto t^p \tag{70}$$

$$\varphi(t) = \sqrt{2p} M_{\rm Pl} \ln\left(\sqrt{\frac{V_0}{p(3p-1)}}\frac{t}{M_{\rm Pl}}\right). \tag{71}$$

The general solution approaches rapidly this particular solution (i.e., it is an attractor). You can see that the expansion, $a(t)$, is power-law, giving the model its name.

The slow-roll parameters for this model are

$$\varepsilon = \frac{1}{2}\eta = \frac{1}{p}\,, \tag{72}$$

independent of φ. In this model inflation never ends, unless other physics intervenes.

7.6 Reheating

During inflation, practically all the energy in the universe is in the inflaton potential $V(\varphi)$, since the slow-roll condition says $\frac{1}{2}\dot\varphi^2 \ll V(\varphi)$. When inflation ends, this energy is transferred in the reheating process to a thermal bath of particles produced in the reheating. Thus reheating creates, from $V(\varphi)$, all the stuff there is in the later universe!

Note that *re*heating may be a misnomer, since we don't know whether the universe was in a thermodynamical equilibrium ever before.

In single-field models of inflation, reheating does not affect the primordial density perturbations,[10] except that it affects the relation of φ_k and k/H_0 given in (85), i.e., how much the distance scale of the perturbations is stretched between inflation and today (these will be discussed later).

Reheating is important for the question of whether unwanted—or wanted—relics are produced after inflation. The reheating temperature must be high enough so that we get standard Big Bang Nucleosynthesis (BBN) after reheating, but sufficiently low so that we do not produce unwanted relics. The latter constraint depends on the extended theory, but it should at least be below the GUT scale. Thus we can take that

$$1\,{\rm MeV} < T_{\rm reh} < 10^{14}\,{\rm GeV}\,. \tag{73}$$

[10] In more complicated models of inflation, involving several fields, reheating may also change the nature of primordial density perturbations.

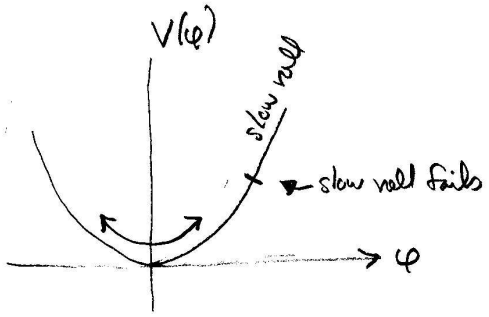


Figure 8: After inflation, the inflaton field is left oscillating at the bottom.

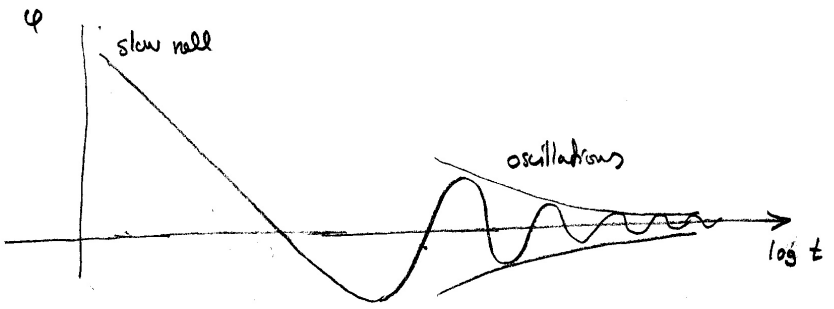


Figure 9: The time evolution of φ as inflation ends.

7.6.1 Scalar field oscillations

After inflation, the inflaton field φ begins to oscillate at the bottom of the potential $V(\varphi)$, see Fig. 8. The inflaton field is still homogeneous, $\varphi(t,\vec{x}) = \varphi(t)$, so it oscillates in the same phase everywhere (we say the oscillation is *coherent*). The expansion time scale H^{-1} soon becomes much longer than the oscillation period.

Assume the potential can be approximated as $\propto \varphi^2$ near the minimum of $V(\varphi)$, so that we have a harmonic oscillator. Write $V(\varphi) = \frac{1}{2}m^2\varphi^2$:

$$\left.\begin{aligned} \ddot{\varphi} + 3H\dot{\varphi} &= -V'(\varphi) \\ \rho &= \frac{1}{2}\dot{\varphi}^2 + V(\varphi) \end{aligned}\right\} \quad \text{become} \quad \begin{cases} \ddot{\varphi} + 3H\dot{\varphi} = -m^2\varphi \\ \rho = \frac{1}{2}\left(\dot{\varphi}^2 + m^2\varphi^2\right) \end{cases}$$

What is $\rho(t)$?

$$\dot{\rho} + 3H\rho \;=\; \dot{\varphi}\overbrace{\left(\ddot{\varphi} + m^2\varphi\right)}^{-3H\dot{\varphi}} + 3H\cdot\frac{1}{2}\left(\dot{\varphi}^2 + m^2\varphi^2\right) \;=\; \frac{3}{2}H\overbrace{\left(m^2\varphi^2 - \dot{\varphi}^2\right)}^{\text{oscillates}}$$

The oscillating factor on the right hand side averages to zero over one oscillation period (in the limit where the period is $\ll H^{-1}$).

Averaging over the oscillations, we get that the long-time behavior of the energy density is

$$\dot{\rho} + 3H\rho = 0 \quad \Rightarrow \quad \rho \propto a^{-3}\,, \tag{74}$$

just like in a matter-dominated universe (we use this result in Sec. 7.7.2). The fall in the energy density shows as a decrease of the oscillation amplitude, see Fig. 9.

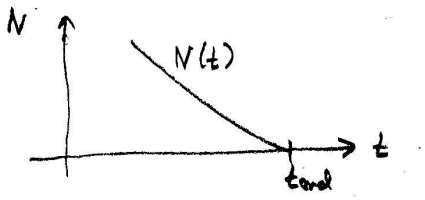


Figure 10: Remaining number of e-foldings $N(t)$ as a function of time.

7.6.2 Inflaton decays

Now that the inflaton field is doing small oscillations around the potential minimum, the particle picture becomes appropriate, and we can consider the energy density ρ_φ to be due to inflaton particles. These inflatons decay into other particles, once the Hubble time ($\sim$ the time after inflation ended) reaches the inflaton decay time.

If the decay is slow (which is the case if the inflaton can only decay into fermions) the inflaton energy density follows the equation

$$\dot{\rho}_\varphi + 3H\rho_\varphi \;=\; -\Gamma_\varphi \rho_\varphi \,, \tag{75}$$

where $\Gamma_\varphi = 1/\tau_\varphi$, the *decay width*, is the inverse of the inflaton decay time τ_φ, and the term $-\Gamma_\varphi\rho_\varphi$ represents energy transfer to other particles.

If the inflaton can decay into bosons, the decay may be very rapid, involving a mechanism called *parametric resonance*. This kind of rapid decay is called *preheating*, since the bosons thus created are far from thermal equilibrium (occupation numbers of states are huge—not possible for fermions).

7.6.3 Thermalization

The particles produced from the inflatons will interact, create other particles through particle reactions, and the resulting particle soup will eventually reach thermal equilibrium with some temperature $T_{\rm reh}$. This *reheating temperature* is determined by the energy density $\rho_{\rm reh}$ at the end of the reheating epoch:

$$\rho_{\rm reh} = \frac{\pi^2}{30} g_*(T_{\rm reh}) T_{\rm reh}^4 \,. \tag{76}$$

Necessarily $\rho_{\rm reh} < \rho_{\rm end}$ (end = end of inflation). If reheating takes a long time, we may have $\rho_{\rm reh} \ll \rho_{\rm end}$. After reheating, we enter the standard Hot Big Bang history of the universe.

7.7 Scales of inflation

7.7.1 Amount of inflation

During inflation, the scale factor $a(t)$ grows by a huge factor. We define the *number of e-foldings* from time t to end of inflation ($t_{\rm end}$) by

$$N(t) \equiv \ln \frac{a(t_{\rm end})}{a(t)} \tag{77}$$

See Fig. 10.

As we saw in Sec. 7.5.1, $a(t)$ changes much faster than $H(t)$ (when the slow-roll approximation is valid), so that the comoving Hubble length $\mathcal{H}^{-1} = 1/aH$ shrinks by almost as many e-foldings. ($a(t)$ grows fast, $H(t)$ decreases slowly.)

We can calculate $N(t) \equiv N(\varphi(t)) \equiv N(\varphi)$ from the shape of the potential $V(\varphi)$ and the value of φ at time t:

$$N(\varphi) \equiv \ln \frac{a(t_{\rm end})}{a(t)} = \int_t^{t_{\rm end}} H(t)dt = \int_\varphi^{\varphi_{\rm end}} \frac{H}{\dot\varphi} d\varphi \overset{\text{slow roll}}{\approx} \boxed{\frac{1}{M_{\rm Pl}^2} \int_{\varphi_{\rm end}}^{\varphi} \frac{V}{V'} d\varphi} . \tag{78}$$

where we used

$$d\ln a = \frac{da}{a} = Hdt = H\frac{d\varphi}{\dot\varphi} . \tag{79}$$

Example: For the simple inflation model $V(\varphi) = \frac{1}{2}m^2\varphi^2$,

$$N(\varphi) = \frac{1}{M_{\rm Pl}^2} \int_{\varphi_{\rm end}}^{\varphi} \frac{V}{V'} d\varphi = \frac{1}{M_{\rm Pl}^2} \int_{\varphi_{\rm end}}^{\varphi} \frac{\varphi}{2} = \frac{1}{4M_{\rm Pl}^2} \left(\varphi^2 - {\varphi_{\rm end}}^2\right) = \frac{1}{4}\left(\frac{\varphi}{M_{\rm Pl}}\right)^2 - \frac{1}{2} . \tag{80}$$

The largest initial value of φ we may contemplate is that which gives the Planck density, $V(\varphi) = M_{\rm Pl}^4 \quad \Rightarrow \quad \varphi = \sqrt{2}M_{\rm Pl}^2/m$. Starting from this value we get $\frac{1}{2}\left[(M_{\rm Pl}/m)^2 - 1\right]$ e-foldings of inflation. With $m = 4 \times 10^9\,\text{GeV}$ (see the earlier example with this model), this gives 1.85×10^{17} e-foldings, i.e., expansion by a factor $e^{1.85\times10^{17}} \sim 10^{8\times10^{16}} = 10^{80\,000\,000\,000\,000\,000}$. That's quite a lot!

7.7.2 Evolution of scales

When discussing (next chapter) evolution of density perturbations and formation of structure in the universe, we will be interested in the history of each comoving distance scale, or each *comoving wave number* k (from a Fourier expansion in comoving coordinates).

$$k = \frac{2\pi}{\lambda} , \qquad k^{-1} = \frac{\lambda}{2\pi}$$

An important question is, whether a distance scale is larger or smaller than the Hubble length at a given time.

We define a scale to be

- superhorizon, when $k < \mathcal{H} \quad \left(k^{-1} > \mathcal{H}^{-1}\right)$
- at horizon (exiting or entering horizon), when $k = \mathcal{H}$
- subhorizon, when $k > \mathcal{H} \quad \left(k^{-1} < \mathcal{H}^{-1}\right)$

Note that *large* scales (large k^{-1}) correspond to *low* k, and *vice versa*, although we often talk about "scale k". This can easily cause confusion, so watch for this, and be careful with wording! To avoid confusion, use the words *high*/*low* instead of large/small for k. Notice also that we are here using the word "horizon" to refer to the Hubble length.[11] Recall that:

$$\text{Inflation} \Rightarrow \mathcal{H}^{-1} \text{ shrinking}$$
$$\text{All other times} \Rightarrow \mathcal{H}^{-1} \text{ growing}$$

See Fig. 11.

[11] As discussed in Cosmology I, there are (at least) three different usages for the word "horizon":

1. particle horizon
2. event horizon (not used in Cosmology I/II)
3. Hubble length

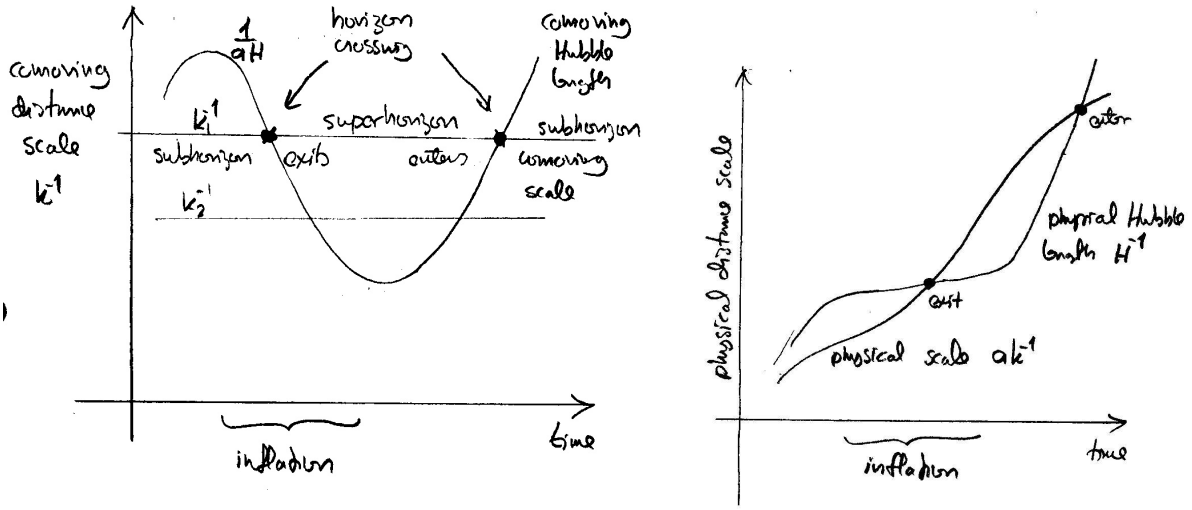


Figure 11: The evolution of the Hubble length, and two scales, k_1^{-1} and k_2^{-1}, seen in comoving coordinates (left) and in terms of physical distance (right).

We shall later find that the amplitude of primordial density perturbations at a given comoving scale is determined when this scale exits the horizon during inflation. The largest observable scales, $k \approx \mathcal{H}_0 = H_0$, are "at horizon" today. (Since the universe has recently began accelerating again, these scales have just barely entered, and are actually now exiting again.)

To identify the distance scales *during inflation* with the corresponding distance scales in the *present universe*, we need a complete history from inflation to the present. We divide it into the following periods:

1. **From** the time the scale k of interest exits the horizon during inflation **to** the end of inflation (t_k to $t_{\rm end}$).

2. **From** the end of inflation **to** reheating. We assume (as discussed in Sec. 7.6.1) that the universe behaves as if matter-dominated, $\rho \propto a^{-3}$, during this period ($t_{\rm end}$ to $t_{\rm reh}$).

3. **From** reheating **to** the present time ($t_{\rm reh}$ to t_0).

Consider now some scale k, which exits at $t = t_k$, when $a = a_k$ and $H = H_k$

$$\Rightarrow \quad k = \mathcal{H}_k = a_k H_k \,.$$

To find out how large this scale is today, we relate it to the present "horizon", i.e., the Hubble scale:

$$\frac{k}{H_0} \;=\; \frac{a_k H_k}{a_0 H_0} \;=\; \frac{a_k}{a_{\rm end}}\frac{a_{\rm end}}{a_{\rm reh}}\frac{a_{\rm reh}}{a_0}\frac{H_k}{H_0} \;=\; e^{-N(k)}\left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{\frac{1}{3}}\left(\frac{\rho_{r0}}{\rho_{\rm reh}}\right)^{\frac{1}{4}}\left(\frac{\rho_k}{\rho_{\rm cr0}}\right)^{\frac{1}{2}}, \qquad (81)$$

where $\rho_k \approx V(\varphi_k) \equiv V_k$ (since $\frac{1}{2}\dot{\varphi}^2 \ll V(\varphi)$ during slow roll) is the energy density when scale k exited and $N(k) \equiv$ number of e-foldings of inflation after that. Eq. (78) allows us to relate φ_k to $N(k)$. The factor $a_{\rm reh}/a_0 = a_{\rm reh}$ is related to the change in energy density from $t_{\rm reh} \to t_0$. The behavior of the total energy density as a function of a changed from the radiation-dominated to the matter-dominated to the dark-energy-dominated era, but we can keep things simpler by considering just the radiation component ρ_r, which was equal to the total energy density $\rho_{\rm reh}$ at end of reheating and behaves after that as $\rho_r \propto a^{-4}$. This is slightly inaccurate, since $\rho_r \propto a^{-4}$

does not take into account the change in g_*. However, the $\propto a^{-4}$ approximation is good enough[12] for us—we are making other comparable approximations also. From end of inflation to reheating

$$\rho \propto a^{-3} \quad \Rightarrow \quad \frac{a_{\rm end}}{a_{\rm reh}} = \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{\frac{1}{3}} ,$$

and the ratio H_k/H_0 we got from

$$H_k = \sqrt{\frac{8\pi G}{3}\rho_k}\,,\ H_0 = \sqrt{\frac{8\pi G}{3}\rho_{\rm cr0}} \quad \Rightarrow \quad \frac{H_k}{H_0} = \left(\frac{\rho_k}{\rho_{\rm cr0}}\right)^{\frac{1}{2}} .$$

Thus we get that

$$\frac{k}{H_0} = e^{-N(k)} \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{1/12} \left(\frac{V_k}{\rho_{\rm end}}\right)^{1/4} \frac{V_k^{1/4}\rho_{r0}^{1/4}}{\rho_{\rm cr0}^{1/2}} .$$

We can now relate $N(k)$ to k/H_0 as

$$N(k) \;=\; -\ln\frac{k}{H_0} - \frac{1}{3}\ln\frac{\rho_{\rm end}^{1/4}}{\rho_{\rm reh}^{1/4}} + \ln\frac{V_k^{1/4}}{\rho_{\rm end}^{1/4}} + \ln\frac{V_k^{1/4}}{10^{16}\,{\rm GeV}} + \ln\frac{10^{16}\,{\rm GeV}\cdot\rho_{r0}^{1/4}}{\rho_{\rm cr0}^{1/2}}\,, \tag{84}$$

where $10^{16}\,{\rm GeV}$ serves as a reference scale for V_k. This is roughly an upper limit to V_k due to lack of observation of primordial gravitational waves (discussed in the next chapter). Sticking in the known values of $\rho_{r0}^{1/4} = 2.4\times10^{-13}\,{\rm GeV}$ (assuming massless neutrinos; however, neutrino masses will not change the result for $N(\varphi_k)$) and $\rho_{\rm cr0}^{1/4} = 3.000\times10^{-12}\,{\rm GeV}\cdot h^{1/2}$, the last term becomes $60.85 - \ln h \approx 61$.

The final result is

$$N(\varphi_k) \;=\; -\ln\frac{k}{H_0} + 61 + \ln\frac{V_k^{1/4}}{\rho_{\rm end}^{1/4}} - \frac{1}{3}\ln\frac{\rho_{\rm end}^{1/4}}{\rho_{\rm reh}^{1/4}} - \ln\frac{10^{16}\,{\rm GeV}}{V_k^{1/4}}\,, \tag{85}$$

where the terms have been arranged so that they are all positive (when the sign in front of them is not included). Since the potential V_k changes slowly during slow roll, the k-dependence is dominated by the first term and the third term is small. The fourth term depends on how fast the reheating was. If it was instantaneous, this term is zero. The last term can be large if the inflation scale is much lower than $10^{16}\,{\rm GeV}$.

[12] Accurately this would go as:

$$g_{*s}a^3T^3 = \text{const.} \quad \Rightarrow \quad \frac{a_{\rm reh}}{a_0} = \left[\frac{g_{*s}(T_0)}{g_{*s}(T_{\rm reh})}\right]^{\frac{1}{3}} \frac{T_0}{T_{\rm reh}} \tag{82}$$

Eq. (81) approximates this with

$$\left(\frac{\rho_{r0}}{\rho_{\rm reh}}\right)^{\frac{1}{4}} = \left[\frac{g_*(T_0)}{g_*(T_{\rm reh})}\right]^{\frac{1}{4}} \frac{T_0}{T_{\rm reh}} \tag{83}$$

Taking $g_{*s}(T_{\rm reh}) = g_*(T_{\rm reh}) \sim 100$, the ratio of these two becomes

$$\frac{(82)}{(83)} = \frac{g_{*s}(T_0)^{\frac{1}{3}}}{g_*(T_0)^{\frac{1}{4}} g_*(T_{\rm reh})^{\frac{1}{12}}} \approx \frac{3.909^{\frac{1}{3}}}{3.363^{\frac{1}{4}}100^{\frac{1}{12}}} = 0.79 \sim 1$$

Note that $a \propto \rho_r^{-1/4}$ is a better approximation than $a \propto T^{-1}$, since these two differ by

$$\left[\frac{g_*(T_{\rm reh})}{g_*(T_0)}\right]^{\frac{1}{4}} \sim \left(\frac{100}{3.363}\right)^{\frac{1}{4}} \sim 2.33\,.$$

For any given present scale, given as a fraction of the present Hubble distance,[13] Eq. (85) identifies the value φ_k the inflaton had, when this scale exited the horizon during inflation. The last three terms give the dependence on the energy scales connected with inflation and reheating. In typical inflation models, they are relatively small. Usually, the precise value of N is not that important; we are more interested in the derivative dN/dk, or rather $d\varphi_k/dk$. We can see that typically (for high-energy-scale inflation) about 60 e-foldings of inflation occur *after* the largest observable scales exit the horizon. The number of e-foldings *before* that can be very large (e.g., billions or much more), depending on the inflation model and how the inflation is assumed to begin.

7.8 Initial conditions for inflation

Inflation provides the initial conditions for the Hot Big Bang. What about initial conditions for inflation? As we discussed earlier, inflation erases all memory of these initial conditions, removing this question from the reach of observational verification. However, a complete picture of the history of the universe should also include some idea about the conditions before inflation. To weigh how plausible inflation is as an explanation we may contemplate how easy it is for the universe to begin inflating.

Although inflation differs radically from the other periods of the history of the universe we have discussed, two qualitative features still hold true also during inflation: 1) the universe is expanding and 2) the energy density is decreasing[14] (although slowly during inflation).

Thus the energy density should be higher before inflation than during it or after it. Often it is assumed that inflation begins right at the Planck scale, $\rho \sim M_{\rm Pl}^4$, which is the limit to how high energy densities we can extend our discussion, which is based on classical GR. Consider one such scenario:

When $\rho > M_{\rm Pl}^4$, quantum gravitational effects should be important. We can imagine that the universe at that time, the *Planck era*, is some kind of "spacetime foam", where the spacetime itself is subject to large quantum fluctuations. When the energy density of some region, larger than H^{-1}, falls below $M_{\rm Pl}^4$, spacetime in that region begins to behave in a classical manner. See Fig. 12. The initial conditions, i.e, conditions at the time when "our universe" (referring to one such region) emerges from the spacetime foam, are usually assumed *chaotic* (term due to Linde [3], does not refer to chaos theory), i.e., φ takes different, random, values at different regions. Since $\rho \geq \rho_\varphi$, and

$$\rho_\varphi = \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}\nabla\varphi^2 + V(\varphi)\,, \tag{86}$$

we must have

$$\dot{\varphi}^2 \lesssim M_{\rm Pl}^4\,, \quad \nabla\varphi^2 \lesssim M_{\rm Pl}^4\,, \quad V(\varphi) \lesssim M_{\rm Pl}^4 \tag{87}$$

in a region for it to emerge from the spacetime foam. If the conditions are suitable such a region may then begin to inflate. Thus inflation may begin at many different parts of the spacetime foam. Our observable universe would be just a small part of one such region which has inflated to a huge size.

It is also possible that during inflation, for some part of the potential, quantum fluctuations of the inflaton (not of the spacetime) dominate over the classical evolution, pushing φ higher in some regions. These regions will then expand faster, and dominate the volume. This gives rise to *eternal inflation*, where, at any given time, most of the volume of the universe is inflating. (This possibility depends on the shape of the potential.) But our observable universe would be a

[13] For example, $k/H_0 = 10$ means that we are talking about a scale corresponding to a wavelength λ, where $\lambda/2\pi$ is one tenth of the Hubble distance.

[14] Not necessarily true in all cases, e.g., eternal inflation.

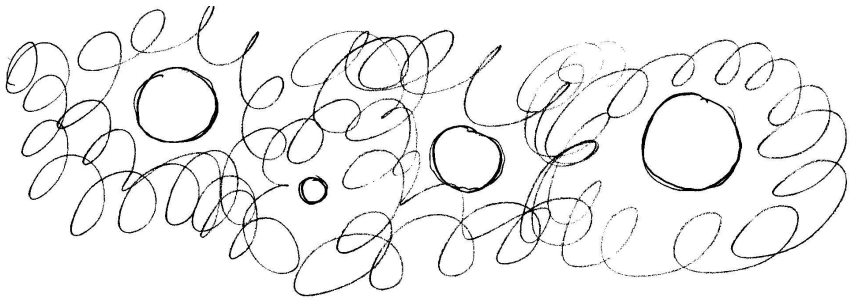


Figure 12: Spacetime foam and some regions emerging from it.

part of a region, where φ came down to a region of the potential where the quantum fluctuations of φ were small and the classical behavior began to dominate and eventually inflation ended.

Thus we see that the very, very, very large scale structure of the universe may be very complicated. But we will never discover this, since our entire observable universe is just a small homogeneous part of a patch which inflated, and then the inflation ended in that patch. All the observable features of the Universe can be explained in terms of what happened in this patch during and after inflation.

These ideas of spacetime foam and eternal inflation are rather speculative and there are also other suggestions for the initial stages of the universe.

7.9 Inflation and Big Bang

There is some confusion about what exactly is meant by the term "Big Bang". Initially the term was introduced to refer to the beginning of the Universe (specifically, Hoyle compared "the hypothesis that all matter of the universe was created in one big bang at a particular time in the remote past"[4] to his own steady-state model of the universe which required a continuous creation of matter). Matter- and/or radiation-dominated FRW models have a beginning where $a = 0$, and we can set this to correspond to the origin ($t = 0$) of the time coordinate, so it became customary to refer to this $t = 0$ as the Big Bang. In classical GR this is a singularity (the spacetime curvature becomes infinite) with infinite energy density. It is clear that classical GR must break down near this singularity, so instead of singularity, there should be something else, which is unknown. Moreover, we have well-tested physics theories only below the energy scale of the electroweak transition. This fuzziness about the very beginning was not an obstacle to calling it a Big Bang and thinking that it referred to a specific time, for as long as it was thought that the expansion was slowing down, because that made the time scale of the earlier events shorter, so that the "fuzzy part" was thought to be an extremely short time compared to the times for which there was rigorous scientific discussion.

Inflation changed this. In inflation the expansion is accelerating and the earlier times are no longer negligibly short compared to the time scales under discussion (e.g., of the late part of inflation during which the observable scales exit the horizon). In fact, we have no upper limit on how long inflation may have lasted, and no way to know what happened before inflation. Also, the conditions during inflation are very different from the hot dense particle soup that was thought to be created in the Big Bang. When Guth proposed inflation, he had in mind the earlier idea of Big Bang, where at first the temperature and density were extremely high so that cooling lead to a GUT transition giving rise to inflation. Thus he still had a "Big Bang" before inflation. This order of events has persisted in many popular presentations of the early universe. However, the modern view is that this is not correct. We have no knowledge of what happened before inflation, no evidence of an initial singularity or a Planck epoch, and thus we cannot place with confidence anything before inflation. If we go back to Hoyle's original phrase about

creation of all matter, then in the inflation scenario the "Big Bang" would be the reheating. On the other hand, we cannot be certain that inflation really happened, it is just a favorite scenario, not something proved beyond doubt. The clear evidence about "Big Bang" is the abundances of light element isotopes, which tell us about big bang nucleosynthesis, and the cosmic microwave background, which tell us about recombination and photon decoupling. When we say that we know that there was a Big Bang in the early universe, we really mean that we know that these events took place, in a way that we described in Cosmology I. Thus my personal favorite for the modern meaning of "Big Bang" is that it refers, not to a specific instant of time, but to this epoch in the early universe, ending at photon decoupling, when the universe was filled with an almost homogeneous hot soup of interacting particles. Not all cosmologists use the term this way. But in any case, within modern cosmology, "Big Bang" can sensibly refer only to something after inflation, not before it.

References

[1] A.H. Guth, *Inflationary universe: A possible solution to the horizon and flatness problems*, Phys. Rev. D **23**, 347 (1981)

[2] A.R. Liddle and D.H. Lyth: Cosmological Inflation and Large-Scale Structure (Cambridge University Press 2000)

[3] A.D. Linde, *Chaotic Inflation*, Phys. Lett. **129B**, 177 (1983)

[4] H. Kragh, *Big Bang: the etymology of a name*, Astronomy & Geophysics, **54**, Issue 2, p 2.28–30 (2013), `https://doi.org/10.1093/astrogeo/att035`

8 Structure Formation

Up to this point we have discussed the universe in terms of a homogeneous and isotropic model (which we shall now refer to as the "unperturbed" or the "background" universe). Clearly the universe is today rather inhomogeneous. By *structure formation* we mean the generation and evolution of this inhomogeneity. We are here interested in distance scales from galaxy size to the size of the whole observable universe. The structure is manifested in the existence of luminous galaxies and in their uneven distribution, their *clustering*. This is the obvious inhomogeneity, but we understand it reflects a density inhomogeneity also in other, nonluminous, components of the universe, especially the *cold dark matter*. The structure has formed by gravitational amplification of a small primordial inhomogeneity. There are thus two parts to the theory of structure formation:

1) The generation of this primordial inhomogeneity, "the seeds of galaxies". This is the more speculative part of structure formation theory. We cannot claim that we know how this primordial inhomogeneity came about, but we have a good candidate scenario, *inflation*, whose predictions agree with the present observational data, and can be tested more thoroughly by future observations. In inflation, the structure originates from *quantum fluctuations* of the inflaton field φ near the time the scale in question exits the horizon.

2) The growth of this small inhomogeneity into the present observable structure of the universe. This part is less speculative, since we have a well established theory of gravity, *general relativity*. However, there is uncertainty in this part too, since we do not know the precise nature of the dominant components to the energy density of the universe, the *dark matter* and the *dark energy*. The gravitational growth depends on the equations of state and the streaming lengths (particle mean free path between interactions) of these density components. Besides gravity, the growth is affected by pressure forces.

We shall do the second part first. But before that we discuss statistical measures of inhomogeneity: correlation functions and power spectra.

8.1 Inhomogeneity

We write all our inhomogeneous quantities as a sum of a homogeneous background value, and a perturbation, the deviation from the background value. For example, for energy density and pressure we write

$$\begin{aligned} \rho(t,\mathbf{x}) &= \bar{\rho}(t) + \delta\rho(t,\mathbf{x}) \\ p(t,\mathbf{x}) &= \bar{p}(t) + \delta p(t,\mathbf{x})\,, \end{aligned} \tag{1}$$

where $\bar{\rho}$ and $\bar{p}$ are the background density and pressure, $\mathbf{x}$ is the *comoving* 3D space coordinate, and $\delta\rho$ and δp are the density and pressure perturbations. We further define the relative density perturbation

$$\delta(t,\mathbf{x}) \equiv \frac{\delta\rho(t,\mathbf{x})}{\bar{\rho}(t)}\,. \tag{2}$$

Since $\rho \geq 0$, necessarily $\delta \geq -1$. These quantities can be defined separately for different components to the energy density, e.g., matter, radiation, and dark energy. Perturbations in dark energy are expected to be small, and if it is just vacuum energy, it has no perturbations. When we discuss the later history of the universe, the main interest is in the matter density perturbation,

$$\delta_m(t,\mathbf{x}) \equiv \frac{\delta\rho_m(t,\mathbf{x})}{\bar{\rho}_m(t)}\,, \tag{3}$$

and *then we will often write just* δ *for* δ_m.

We do the split into the background and perturbation so that the background is equal to the mean (volume average) of the full quantity. An important question is, whether the $\bar{\rho}(t)$ and $\bar{p}(t)$ defined this way correspond to a (homogeneous and isotropic) solution of General Relativity, i.e., an FRW universe. We expect the exact answer to be negative, since GR is a nonlinear theory, so that perturbations affect the evolution of the mean. This effect is called *backreaction*.

However, if the perturbations are small, we can make an approximation, where we drop from our equations all those terms which contain a product of two or more perturbations, as these are "higher-order" small. The resulting approximate theory is called *first-order perturbation theory* or *linear* perturbation theory. As the second name implies, the theory is now linear in the perturbations, meaning that the effect of overdensities cancel the effect of underdensities on, e.g., the average expansion rate. In this case the mean values evolve just like they would in the absence of perturbations.

While the perturbations at large scales have remained small, during the later history of the universe the perturbations have grown large at smaller scales. How big is the effect of backreaction, is an open research question in cosmology, since the calculations are difficult, but a common view is that the effect is small compared to the present accuracy of observations. For this course, we adopt this view, and assume that the background universe simultaneously represents a FRW universe ("the universe we would have if we did not have the perturbations") and the mean values of the quantities in the true universe at each time t.

Moreover, in Cosmology II we shall (mostly) assume that the background universe is flat ($K = 0$).

8.1.1 Statistical homogeneity and isotropy

We assume that the origin of the perturbations is some random process in the early universe. Thus over- ($\delta > 0$) and underdensities ($\delta < 0$) occur at randomly determined locations and we cannot expect to theoretically predict the values of $\delta(t, \mathbf{x})$ for particular locations $\mathbf{x}$. Instead, we can expect theory to predict statistical properties of the inhomogeneity field $\delta(t, \mathbf{x})$. The statistical properties are typically defined as averages of some quantities. We will deal with two kind of averages: *volume average* and *ensemble average*; the ensemble average is a theoretical concept, whereas the volume average is more observationally oriented.

We denote the volume average of some quantity $f(\mathbf{x})$ with the overbar, $\bar{f}$, and it is defined as

$$\bar{f} \equiv \frac{1}{V} \int_V d^3x f(\mathbf{x}) \,. \tag{4}$$

The integration volume V in question will depend on the situation.

For the ensemble average we assume that our universe is just one of an *ensemble* of an infinite number of possible universes (*realizations* of the random process) that could have resulted from the random process producing the initial perturbations. To know the random process means to know the probability distribution $\mathrm{Prob}(\gamma)$ of the quantities γ produced by it. (At this stage we use the abstract notation of γ to denote the infinite number of these quantities. They could be the generated initial density perturbations at all locations, $\delta(\mathbf{x})$, or the corresponding Fourier coefficients $\delta_{\mathbf{k}}$. We will be more explicit later.) The ensemble average of a quantity f depending on these quantities γ as $f(\gamma)$ is denoted by $\langle f \rangle$ and defined as the (possibly infinite-dimensional) integral

$$\langle f \rangle \equiv \int d\gamma \mathrm{Prob}(\gamma) f(\gamma) \,. \tag{5}$$

Here f could be, e.g., the value of $\rho(\mathbf{x})$ at some location $\mathbf{x}$. The ensemble average is also called the *expectation value*. Thus the ensemble represents a probability distribution of universes. A

cosmological theory predicts such a probability distribution, but it does not predict in which realization from this distribution we live in. Thus the theoretical properties of the universe we will discuss (e.g., statistical homogeneity and isotropy, and ergodicity, see below) will be properties of this ensemble.

We now make the assumption that, although the universe is inhomogeneous, it is *statistically homogeneous and isotropic*. This is the second version of the *Cosmological Principle*. Statistical homogeneity means that the expectation value $\langle f(\mathbf{x})\rangle$ must be the same at all $\mathbf{x}$, and thus we can write it as $\langle f\rangle$. Statistical isotropy means that for quantities which involve a direction, the statistical properties are independent of the direction. For example, for vector quantities $\mathbf{v}$, all directions must be equally probable. This implies that $\langle \mathbf{v}\rangle = 0$. The assumption of statistical homogeneity and isotropy is justified by inflation: inflation makes the background universe homogeneous and isotropic so that the external conditions for quantum fluctuations are everywhere the same.

If the theoretical properties of the universe are those of an ensemble, and we can only observe one universe from that ensemble, how can we compare theory and observation? It seems reasonable that the statistics we get by comparing different parts of the universe should be similar to the statistics of a given part of the universe over different realizations, i.e., that they provide a *fair sample* of the probability distribution. This is called *ergodicity*. Fields $f(\mathbf{x})$ that satisfy

$$\bar{f} = \langle f\rangle \tag{6}$$

for an infinite volume V (for $\bar{f}$) and an arbitrary location $\mathbf{x}$ (for $\langle f\rangle$) are called *ergodic*. We assume that cosmological perturbations are ergodic. The equality does not hold for a finite volume V; the difference is called *sample variance* or *cosmic variance*. The larger the volume, the smaller is the difference. Since cosmological theory predicts $\langle f\rangle$, whereas observations probe $\bar{f}$ for a limited volume, cosmic variance limits how accurately we can compare theory with observations.[1]

8.1.2 Density autocorrelation function

From ergodicity,

$$\langle\rho\rangle = \bar{\rho} \quad \Rightarrow \quad \langle\delta\rho\rangle = 0 \quad \text{and} \quad \langle\delta\rangle = 0\,. \tag{7}$$

Thus we cannot use $\langle\delta\rangle$ as a measure of the inhomogeneity. Instead we can use the square of δ, which is necessarily nonnegative everywhere, so it cannot average out like δ did. Its expectation value

$$\langle\delta^2\rangle = \frac{\langle\delta\rho^2\rangle}{\bar{\rho}^2} \tag{8}$$

is the *variance* of the density perturbation, and the square root of the variance,

$$\delta_{\rm rms} \equiv \sqrt{\langle\delta^2\rangle} \tag{9}$$

the *root-mean-square* (rms) density perturbation, is a typical expected absolute value of δ at an arbitrary location.[2] It tells us about how strong the inhomogeneity is, but nothing about the shapes or sizes of the inhomogeneities. To get more information, we introduce the correlation function ξ.

We define the *density 2-point autocorrelation function* (often called just *correlation function*) as

$$\xi(\mathbf{x}_1,\mathbf{x}_2) \equiv \langle\delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\rangle\,. \tag{10}$$

[1] Another notation I will use for volume average is $\widehat{f}$, for smaller volumes, e.g., the volume observed in a galaxy survey. I try to reserve $\bar{f}$ for situations where we can assume $\bar{f} = \langle f\rangle$, whereas cosmic variance is the difference between $\widehat{f}$ and $\langle f\rangle$.

[2] In other words, $\delta_{\rm rms}$ is the standard deviation of $\rho/\bar{\rho}$.

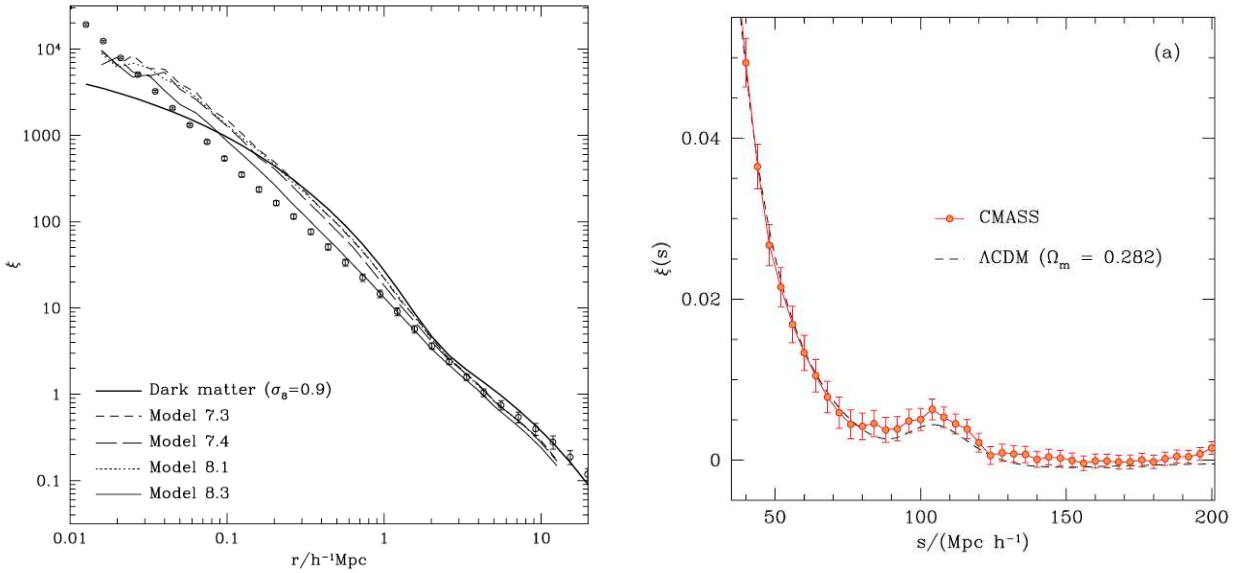


Figure 1: The 2-point correlation function $\xi(r)$ from galaxy surveys. Left: Small scales shown in a log-log plot. The circles with error bars show the observational determination from the APM galaxy survey [1]. The different lines are theoretical predictions by [2] (this is Fig. 9 from [2]). Right: Large scales shown in a linear plot. Red circles with error bars show the observational determination from the CMASS Data Release 9 (DR9) sample of the Baryonic Oscillation Spectroscopic Survey (BOSS). The dashed line is a theoretical prediction from the ΛCDM model. The bump near $100\,h^{-1}$Mpc is the baryon acoustic oscillation (BAO) peak that will be discussed later. This is Fig. 2a from [3].

It is positive if the density perturbation is expected to have the same sign at both $\mathbf{x}_1$ and $\mathbf{x}_2$, and negative for an overdensity at one and underdensity at the other. Thus it probes how density perturbations at different locations are correlated with each other. Due to statistical homogeneity, $\xi(\mathbf{x}_1, \mathbf{x}_2)$ can only depend on the *separation* $\mathbf{r} \equiv \mathbf{x}_2 - \mathbf{x}_1$, so we redefine ξ as

$$\xi(\mathbf{r}) \equiv \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle \,. \tag{11}$$

From statistical isotropy, $\xi(\mathbf{r})$ is independent of direction, i.e., spherically symmetric (isotropic),

$$\xi(\mathbf{r}) = \xi(r) \,. \tag{12}$$

We will have use for both the 3D, $\xi(\mathbf{r})$, and 1D, $\xi(r)$, versions. The correlation function is large and positive for r smaller than the size of a typical over- or underdense region, and becomes small for larger separations.

The correlation function at zero separation gives the variance of the density perturbation,

$$\langle \delta^2 \rangle \equiv \langle \delta(\mathbf{x})\delta(\mathbf{x}) \rangle \equiv \xi(0) \,. \tag{13}$$

We can also define a correlation function $\widehat{\xi}(\mathbf{r})$ for a single realization as a volume average,

$$\widehat{\xi}(\mathbf{r}) \equiv \frac{1}{V} \int d^3x\, \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \,. \tag{14}$$

Integrating over $\mathbf{r}$ and assuming periodic boundary conditions[3] we get the *integral constraint*

$$\int d^3r\, \widehat{\xi}(\mathbf{r}) \;=\; \frac{1}{V} \int d^3r d^3x\, \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \;=\; \frac{1}{V} \int d^3x\, \delta(\mathbf{x}) \int d^3r\, \delta(\mathbf{x}+\mathbf{r}) = 0 \,, \tag{16}$$

[3]The other option is not to use periodic boundary conditions, but to understand the integral in (14) to go over only those $\mathbf{x}$, for which both $\mathbf{x}$ and $\mathbf{x}+\mathbf{r} \in V$. This is what we have to do when V refers to an actual survey.

since the latter integral is $\bar{\delta} = 0$. Since $\xi(\mathbf{r}) = \langle\widehat{\xi}(\mathbf{r})\rangle$ the integral constraint applies to it likewise. Therefore $\xi(r)$ must become negative at some point, so that at such a distance from an overdense region we are more likely to find an underdense region. Going to ever larger separations, ξ as a function of r may oscillate around zero, the oscillation becoming ever smaller in amplitude. Most of the interest in $\xi(r)$ is for the small r within the initial positive region.

8.1.3 Fourier space

The evolution of perturbations is best discussed in Fourier space. Fourier analysis is a method for separating out different distance scales, so that the dependence of the physics on distance scale becomes clear and easy to handle.

For mathematical convenience, we assume the observable part of the universe lies within a fiducial cubic box, volume $V = L^3$, with periodic boundary conditions. This box may be assumed to be much larger than the region of interest, so that these boundary conditions should have no effect. Since the infinite universe is now assumed periodic, the volume average over the infinite universe will be equal to the volume average over the fiducial box. Thus also the ergodicity assumption requires the fiducial volume to be large, so that it can provide a fair sample of the ensemble.

We can now expand any function of space $f(\mathbf{x})$ as a Fourier series

$$f(\mathbf{x}) = \sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{17}$$

where the wave vectors $\mathbf{k} = (k_1, k_2, k_3)$ take values

$$k_i = n_i \frac{2\pi}{L}\,, \qquad n_i = 0, \pm 1, \pm 2, \ldots \tag{18}$$

The Fourier coefficients $f_{\mathbf{k}}$ are obtained as

$$f_{\mathbf{k}} = \frac{1}{V}\int_V f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x\,. \tag{19}$$

If $f(\mathbf{x})$ is a perturbation so that its mean value vanishes, then the term $\mathbf{k} = 0$ does not occur.

The Fourier coefficients are complex numbers even though we are dealing with real quantities $f(\mathbf{x})$. From the reality of $f(\mathbf{x})$ follows that

$$f_{-\mathbf{k}} = f^*_{\mathbf{k}}\,. \tag{20}$$

This means that for each pair of terms $f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + f_{-\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}$ the imaginary parts cancel. The real part of $f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}$ is

$$\Re\left(f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}\right) = \Re f_{\mathbf{k}} \cos\mathbf{k}\cdot\mathbf{x} - \Im f_{\mathbf{k}} \sin\mathbf{k}\cdot\mathbf{x}\,. \tag{21}$$

To visualize a Fourier component $f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}$ one may thus visualize just this real part, which is a sinusoidal plane wave in the $\mathbf{k}$ direction.

The double integral in (16) then goes over all pairs $(\mathbf{x}_1, \mathbf{x}_2)$ in V and can be written as

$$\frac{1}{V}\int_V d^3x_1\, \delta(\mathbf{x}_1) \int_V d^3x_2\, \delta(\mathbf{x}_2) = 0 \cdot 0\,. \tag{15}$$

For an actual survey, V is not really large enough to make cosmic variance negligible. This has two effects: First, the integral does not extend to large enough separations r to capture the full $\xi(r)$, so that typically negative $\xi(r)$ values at large separations r are missed. This would tend to make the integral positive. However, for an actual survey, also the mean density has to be estimated from the survey, so that δ will actually refer to the deviation from the mean density of the survey, which again forces $\int_V d^3x\, \delta(x) = 0$. Thus this forced integral constraint makes the survey typically underestimate the true correlation function.

The Fourier expansion works only if the background universe is flat, although it can be used as an approximation in open and closed universes,[4] if the region of interest is much smaller than the curvature radius.

The separation of neighboring k_i values is $\Delta k_i = 2\pi/L$, so we can write

$$f(\mathbf{x}) = \sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \left(\frac{L}{2\pi}\right)^3 \Delta k_1 \Delta k_2 \Delta k_3 \approx \frac{1}{(2\pi)^3} \int f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d^3k , \tag{22}$$

where

$$f(\mathbf{k}) \equiv L^3 f_{\mathbf{k}} . \tag{23}$$

replacing the Fourier series with the Fourier integral. The size of the Fourier coefficients depends on the fiducial volume V – increasing V tends to make the $f_{\mathbf{k}}$ smaller to compensate for the denser sampling of $\mathbf{k}$ in Fourier space.

In the limit $V \to \infty$, the approximation in (22) becomes exact, and we have the *Fourier transform* pair

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^3} \int f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d^3k \\ f(\mathbf{k}) &= \int f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x . \end{aligned} \tag{24}$$

Note that this assumes that the integrals converge, which requires that $f(\mathbf{x}) \to 0$ for $|\mathbf{x}| \to \infty$. Thus we use only the Fourier series for, e.g., $\delta(\mathbf{x})$, but for, e.g., the correlation function $\xi(\mathbf{x})$ the Fourier transform is appropriate.

Even with a finite V we can use the Fourier integral as an approximation. Often it is conceptually simpler to work first with the Fourier series (so that one can, e.g., use the Kronecker delta $\delta_{\mathbf{k}\mathbf{k}'}$ instead of the Dirac delta function $\delta_D(\mathbf{k} - \mathbf{k}')$), replacing it with the integral in the end, when it needs to be calculated. The recipe for going from the series to the integral is

$$\begin{aligned} \left(\frac{2\pi}{L}\right)^3 \sum_{\mathbf{k}} &\to \int d^3k \\ L^3 f_{\mathbf{k}} &\to f(\mathbf{k}) \\ \left(\frac{L}{2\pi}\right)^3 \delta_{\mathbf{k}\mathbf{k}'} &\to \delta_D^3(\mathbf{k} - \mathbf{k}') . \end{aligned} \tag{25}$$

so that, e.g.,

$$\sum_{\mathbf{k}} f_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{x}} \to \frac{1}{(2\pi)^3} \int f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d^3k . \tag{26}$$

8.1.4 Power spectrum

We now expand the density perturbation as a Fourier series

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} , \tag{27}$$

with

$$\delta_{\mathbf{k}} = \frac{1}{V} \int_V \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x \tag{28}$$

[4] An exact treatment in open and closed universes requires expansion in terms of suitable other functions instead of the plane waves $e^{i\mathbf{k}\cdot\mathbf{x}}$.

and $\delta_{-\mathbf{k}} = \delta^*_{\mathbf{k}}$. Note that

$$\langle \delta(\mathbf{x}) \rangle = 0 \quad \Rightarrow \quad \langle \delta_{\mathbf{k}} \rangle = 0\,. \tag{29}$$

In analogy with the correlation function $\xi(\mathbf{x}, \mathbf{x}')$, we may ask what is the corresponding correlation in Fourier space, $\langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle$. Note that due to the mathematics of complex numbers, correlations of Fourier coefficients are defined with the complex conjugate *. This way the correlation of $\delta_{\mathbf{k}}$ with itself, $\langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}} \rangle = \langle |\delta_{\mathbf{k}}|^2 \rangle$ is a real (and nonnegative) quantity, the expectation value of the absolute value (modulus) of $\delta_{\mathbf{k}}$ squared, i.e., the variance of $\delta_{\mathbf{k}}$. Calculating

$$\begin{aligned}
\langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle &= \frac{1}{V^2} \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} \int d^3x' e^{-i\mathbf{k}'\cdot\mathbf{x}'} \langle \delta(\mathbf{x}) \delta(\mathbf{x}') \rangle \\
&= \frac{1}{V^2} \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} \int d^3r e^{-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})} \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle \\
&= \frac{1}{V^2} \int d^3r e^{-i\mathbf{k}'\cdot\mathbf{r}} \xi(\mathbf{r}) \int d^3x e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\
&= \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}) \equiv \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} P(\mathbf{k}),
\end{aligned} \tag{30}$$

where we used $\langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle = \xi(\mathbf{r})$, i.e., independent of $\mathbf{x}$, which results *from statistical homogeneity*, and the orthogonality of plane waves

$$\boxed{\int d^3x e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} = V \delta_{\mathbf{k}\mathbf{k}'} \;\to\; (2\pi)^3 \delta_D^3(\mathbf{k}-\mathbf{k}')\,.} \tag{31}$$

Note that here $\delta_{\mathbf{k}\mathbf{k}'}$ is the Kronecker delta, 1 for $\mathbf{k} = \mathbf{k}'$, 0 otherwise – nothing to do with the density perturbation! In the limit $V \to \infty$ we get the Dirac delta function $\delta_D^3(\mathbf{k}-\mathbf{k}')$.

Written in terms of $\delta(\mathbf{k}) = V\delta_{\mathbf{k}}$, the result (30) reads as

$$\langle \delta(\mathbf{k})^* \delta(\mathbf{k}') \rangle = V \delta_{\mathbf{k}\mathbf{k}'} P(\mathbf{k}) \;\to\; (2\pi)^3 \delta_D^3(\mathbf{k}-\mathbf{k}') P(\mathbf{k})\,, \tag{32}$$

Thus, *from statistical homogeneity follows that the Fourier coefficients* $\delta_{\mathbf{k}}$ *are uncorrelated.* The quantity

$$P(\mathbf{k}) \equiv V \langle |\delta_{\mathbf{k}}|^2 \rangle = \int d^3r\, e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r})\,, \tag{33}$$

which gives the *variance* of $\delta_{\mathbf{k}}$, is called the *power spectrum* of $\delta(\mathbf{x})$. Since the correlation function $\to 0$ for large separations, we can replace the integration volume V in (33) with an infinite volume.[5] We see that the power spectrum is the 3D Fourier transform of $\xi(\mathbf{r})$, and therefore also

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{r}} P(\mathbf{k})\,. \tag{34}$$

Unlike the correlation function, the power spectrum $P(\mathbf{k})$ is positive everywhere. Perturbations at large distance scales are more commonly discussed in terms of $P(\mathbf{k})$ than $\xi(\mathbf{r})$.

From statistical isotropy

$$\xi(\mathbf{r}) = \xi(r) \quad \Rightarrow \quad P(\mathbf{k}) = P(k) \tag{35}$$

(the 3D Fourier transform of a spherically symmetric function is also spherically symmetric), so that the variance of $\delta_{\mathbf{k}}$ depends only on the magnitude k of the wave vector $\mathbf{k}$, i.e., on the corresponding distance scale. Using spherical coordinates and doing the angular integrals

[5] We want to avoid discussing $\xi(r)$ for $r \geq L$, since the artificially assumed periodicity would cause artifacts in the behavior of $\xi(\mathbf{r})$ at such separations. Thus L is assumed so large that ξ is completely negligible at such huge separations.

we obtain (**exercise**) the relation between the 1D correlation function $\xi(r)$ and the 1D power spectrum $P(k)$,

$$\begin{aligned} P(k) &= \int_0^\infty \xi(r)\frac{\sin kr}{kr}4\pi r^2 dr \\ \xi(r) &= \frac{1}{(2\pi)^3}\int_0^\infty P(k)\frac{\sin kr}{kr}4\pi k^2 dk\,, \end{aligned} \tag{36}$$

For the density variance we get

$$\langle\delta^2\rangle \equiv \xi(0) = \frac{1}{(2\pi)^3}\int_0^\infty P(k)4\pi k^2 dk = \frac{1}{2\pi^2}\int_{-\infty}^\infty k^3 P(k) d\ln k \equiv \int_{-\infty}^\infty \mathcal{P}(k) d\ln k\,. \tag{37}$$

where we have defined

$$\boxed{\mathcal{P}(k) \equiv \frac{k^3}{2\pi^2}P(k)\,.} \tag{38}$$

Another common notation for $\mathcal{P}(k)$ is $\Delta^2(k)$. The word "power spectrum" is used to refer to both $P(k)$ and $\mathcal{P}(k)$. Of these two, $\mathcal{P}(k)$ has the more obvious physical meaning: it gives the contribution of a logarithmic interval of scales, i.e., from k to ek, to the density variance. $\mathcal{P}(k)$ is dimensionless, whereas $P(k)$ has the dimension of Mpc^3 (when discussing observed values, it is usually given in units of $h^{-3}\mathrm{Mpc}^3$ as distance determinations are proportional to the Hubble constant).

Exercise: Define $\widehat{\xi}(\mathbf{r})$ as the volume average of $\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})$, i.e., integrate $\mathbf{x}$ over the box V with periodic boundary conditions, and show that

$$\widehat{\xi}(\mathbf{r}) = \frac{V}{(2\pi)^3}\int d^3k\,|\delta_\mathbf{k}|^2 e^{i\mathbf{k}\cdot\mathbf{r}}\,, \tag{39}$$

for a single realization. Note that here we do not need any statistical assumptions (like statistical homogeneity or ergodicity). Contrast this result with (34).

8.1.5 Scales of interest and window functions

In (37) we integrated over all scales, from the infinitely large ($k = 0$ and $\ln k = -\infty$) to the infinitely small ($k = \infty$ and $\ln k = \infty$) to get the density variance.[6] Perhaps this is not really what we want. The average matter density today is $3\times10^{-27}\,\mathrm{kg/m^3}$. The density of the Earth is $5.5\times10^3\,\mathrm{kg/m^3}$ and that of an atomic nucleus $2\times10^{17}\,\mathrm{kg/m^3}$, corresponding to $\delta\approx2\times10^{30}$ and $\delta\approx10^{44}$. Probing the density of the universe at such small scales finds a huge variance in it, but this is no longer the topic of cosmology – we are not interested here in planetary science or nuclear physics.

Even the study of the structure of individual galaxies is not considered to belong to cosmology, so the smallest (comoving) scale of cosmological interest, at least when we discuss the present universe,[7] is that of a typical separation between neighboring galaxies, of the order of 1 Mpc.

To exclude scales smaller than R ($r < R$ or $k > R^{-1}$) we *filter* the density field with a *window function*. This can be done in **k**-space or **x**-space.

---

[6] Note that large scales correspond to small k and vice versa. To avoid confusion, it is better to use the words *low* and *high* for k, so that *large scales correspond to low k, and small scales correspond to high k*.

[7] In early universe cosmology we may study events, or possible events, related to also smaller comoving scales.

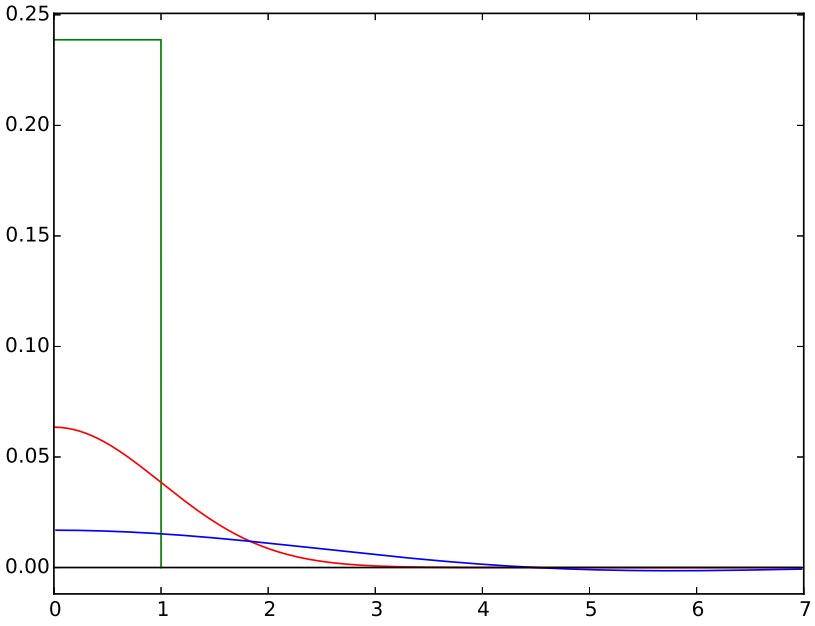


Figure 2: The 3D window functions $W(r)$, top-hat (green), Gaussian (red), and k (blue), for $R = 1$.

The filtering in $\mathbf{x}$-space is done by convolution. We introduce a (usually spherically symmetric) window function $W(\mathbf{r})$ such that $W(\mathbf{r})$ is relatively large for $|\mathbf{r}| \leq R$ and $W \sim 0$ for $|\mathbf{r}| \gg R$. We use normalization

$$\int d^3r\, W(\mathbf{r}) = 1 \tag{40}$$

and define the filtered density field

$$\delta(\mathbf{x}, R) \equiv (\delta * W)(\mathbf{x}) \equiv \int d^3\mathbf{x}'\, \delta(\mathbf{x}') W(\mathbf{x}' - \mathbf{x})\,. \tag{41}$$

The simplest window function is the top-hat window function

$$W_T(\mathbf{r}) \equiv \left(\frac{4\pi}{3}R^3\right)^{-1} \qquad \text{for} \quad |\mathbf{r}| \leq R \tag{42}$$

and $W_T(\mathbf{r}) = 0$ elsewhere, i.e., $\delta(\mathbf{x})$ is filtered by replacing it with its mean value within the distance R. Mathematically more convenient is the Gaussian window function

$$W_G(\mathbf{r}) \equiv \frac{1}{(2\pi)^{3/2}R^3} e^{-\frac{1}{2}|\mathbf{r}|^2/R^2}\,. \tag{43}$$

By the convolution theorem, the filtering in Fourier space becomes just multiplication:

$$\delta(\mathbf{k}, R) = \delta(\mathbf{k}) W(\mathbf{k})\,, \tag{44}$$

where $W(\mathbf{k})$ is the Fourier transform of the window function. For W_T and W_G we have (**exercise**)

$$\begin{aligned} W_T(\mathbf{k}) &= \frac{3(\sin kR - kR\cos kR)}{(kR)^3} \\ W_G(\mathbf{k}) &= e^{-\frac{1}{2}(kR)^2}\,. \end{aligned} \tag{45}$$

We can also define the **k**-space top-hat window function

$$W_k(\mathbf{k}) \equiv 1 \qquad \text{for} \quad k \le 1/R \tag{46}$$

and $W_k(\mathbf{k}) = 0$ elsewhere. In **x**-space this becomes (**exercise**)

$$W_k(\mathbf{r}) = \frac{1}{2\pi^2 R^3} \frac{\sin y - y\cos y}{y^3}, \qquad \text{where} \quad y \equiv |\mathbf{r}|/R\,. \tag{47}$$

The variance of the filtered density field (**Exercise:** derive the second equalities of both expressions)

$$\begin{aligned} \sigma^2(R) &\equiv \langle \delta(\mathbf{x},R)^2 \rangle = \frac{1}{(2\pi)^3} \int d^3k\, P(k) |W(\mathbf{k})|^2 \\ \widehat{\sigma}^2(R) &\equiv \frac{1}{V} \int d^3x\, \delta(\mathbf{x},R)^2 = \frac{V}{(2\pi)^3} \int d^3k\, |\delta_{\mathbf{k}}|^2 |W(\mathbf{k})|^2\,. \end{aligned} \tag{48}$$

is a measure of the inhomogeneity at scale R. For the **k**-space top-hat window this becomes simply

$$\sigma^2(R) = \frac{1}{(2\pi)^3} \int_0^{R^{-1}} 4\pi k^2 P(k) dk = \int_{-\infty}^{-\ln R} \mathcal{P}(k) d\ln k\,. \tag{49}$$

One may also ask, whether scales larger than the observed universe (the lower limit $k = 0$ or $\ln k = -\infty$ in the k integrals) are relevant, since we cannot observe the inhomogeneity at such scales. Due to such very-large-scale inhomogeneities, the average density in the observed universe may deviate from the average density of the entire universe. Inhomogeneities at scales somewhat larger than the observed universe could appear as an anisotropy in the observed universe. The importance of such large scales depends on how strong the inhomogeneities at these scales are, i.e., how the power spectrum behaves as $k \to 0$. The present understanding, supported by observations, is that the contribution of such large scales is small.[8]

8.1.6 Power-law spectra

We have observational information and theoretical predictions for $\xi(r)$ and $P(k)$ for a wide range of scales. (We will discuss the theory in detail later.) For certain intervals, they can be approximated by a power-law form,

$$\xi(r) \propto r^{-\gamma} \qquad \text{or} \qquad P(k) \propto k^n\,. \tag{50}$$

When plotted on a log-log scale, such functions appear as straight lines with slope $-\gamma$ and n. The proportionality constant can be given in terms of a reference scale. For $\xi(r)$ we usually choose the scale r_0 where $\xi(r_0) = 1$, so that

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}. \tag{51}$$

For $P(k)$ we may write

$$P(k) = A^2 \left(\frac{k}{k_p}\right)^n \qquad \text{or} \qquad \mathcal{P}(k) = A^2 \left(\frac{k}{k_p}\right)^{n+3}, \tag{52}$$

[8] The inflation scenario predicts that the universe outside the current horizon is similar to the observable universe up to distances very much larger than the current horizon distance. However at "very very far away", beyond the pre-inflation horizon, the universe may be quite different. For this section, we exclude such "very very far" regions from the concept of "universe". (They are "other universes" in a "multiverse".)

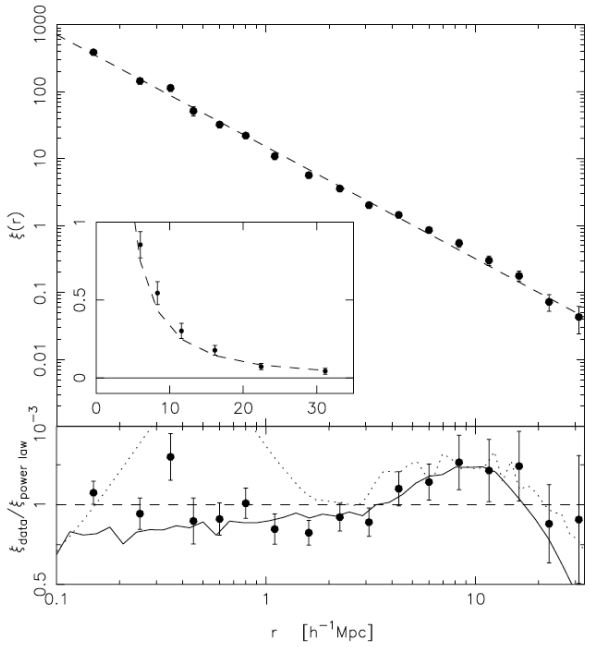


Figure 3: Top panel: The correlation function from the 2dFGRS galaxy survey in log-log scale. The dashed line is the best-fit power law ($r_0 = 5.05\,h^{-1}$Mpc, $\gamma = 1.67$). The inset shows the same in linear scale. Bottom panel: 2dFGRS data (solid circles with error bars) divided by the power-law fit. The solid line is the result from the APM survey and the dotted line from an N-body simulation. This is Fig. 11 from [4].

where k_p is called a *pivot scale* (whose choice depends on the application) and $A \equiv \sqrt{P(k_p)}$ or $\sqrt{\mathcal{P}(k_p)}$ is the amplitude of the power spectrum at the pivot scale.

We define the spectral index $n(k)$ as

$$n(k) \;\equiv\; \frac{d\ln P}{d\ln k}\,. \tag{53}$$

It gives the slope of $P(k)$ on a log-log plot. For a power-law $P(k)$, $n(k) = const = n$. We can study power-law $\xi(r)$ and $P(k)$ as a playground to get a feeling what different values of the spectral index mean, and, e.g., how γ and n are related.[9]

The Fourier transform of a power law is a power law. For the correlation function of (51) we get (**exercise**)

$$\begin{aligned} P(k) &= \frac{4\pi}{k^3}\Gamma(2-\gamma)\sin\frac{(2-\gamma)\pi}{2}\,(kr_0)^\gamma \\ \mathcal{P}(k) &= \frac{2}{\pi}\Gamma(2-\gamma)\sin\frac{(2-\gamma)\pi}{2}\,(kr_0)^\gamma \end{aligned} \tag{54}$$

for $1 < \gamma < 2$ or $2 < \gamma < 3$, and

$$\begin{aligned} P(k) &= \frac{2\pi^2}{k^3}(kr_0)^2 \\ \mathcal{P}(k) &= (kr_0)^2 \end{aligned} \tag{55}$$

for $\gamma = 2$. Thus

$$n \;=\; \gamma - 3 \qquad \text{for } 1 < \gamma < 3, \text{ i.e., } -2 < n < 0\,. \tag{56}$$

[9] In reality the spectral index is very different at small scales than at large scales. Observationally, for small scales, $\gamma \sim 1.8$, and for large scales, $n \sim 1$. We discuss this later.

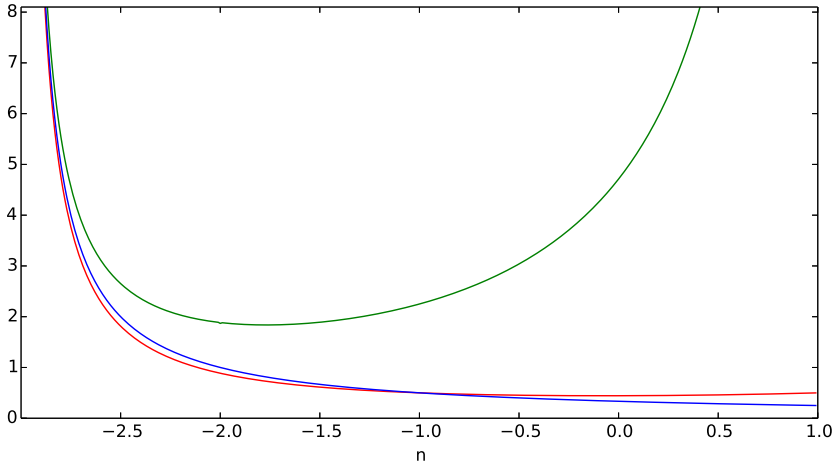


Figure 4: The ratio of $\sigma^2(R)$ to $\mathcal{P}(R^{-1})$ in the case of a power-law spectrum $\mathcal{P}(k) \propto k^{n+3}$ for the three different window functions: top-hat (green), Gaussian (red), and k (blue). They all diverge in the limit $n \to -3$ due to the contributions of ever larger scales ($\ln k \to -\infty$).

The variance

$$\langle \delta^2 \rangle = \xi(0) = \int_0^\infty \mathcal{P}(k) \frac{dk}{k} \propto \int_0^\infty k^{n+2} dk = \frac{1}{n+3} \left[k^{n+3} \right]_0^\infty \qquad \text{for} \quad n \neq -3 \tag{57}$$

diverges at small scales (high k) for $n \geq -3$ and at large scales (low k) for $n \leq -3$. In practice we encounter only the small-scale divergence.

We cure the small-scale divergence with filtering as discussed in Sec. 8.1.5, replacing (57) with (see Eq. 48)

$$\sigma^2(R) \equiv \langle \delta(\mathbf{x}, R)^2 \rangle = \int_0^\infty \mathcal{P}(k) |W(k)|^2 \frac{dk}{k} \,. \tag{58}$$

For the three window functions given in Sec. 8.1.5, power-law spectra give

$$\begin{aligned}
\sigma_T^2(R) &= \frac{9}{2^n}(n+1) \sin\frac{n\pi}{2} \, \frac{\Gamma(n-1)}{n-3} \mathcal{P}(R^{-1}) \qquad \text{for } -3 < n < 1 \\
\sigma_G^2(R) &= \frac{1}{2}\Gamma\left(\frac{n+3}{2}\right) \mathcal{P}(R^{-1}) \qquad \text{for } n > -3 \\
\sigma_k^2(R) &= \frac{1}{n+3} \mathcal{P}(R^{-1}) \qquad \text{for } n > -3.
\end{aligned} \tag{59}$$

For $n \geq 1$, the top-hat window is not able to cure the small-scale divergence, since its Fourier transform does not die out fast enough at high k (this is related to the sharp boundary of the window). For integers $-3 < n < 1$ the formula for $\sigma_T^2(R)$ in (59) is not defined, since either $n+1$ or $\sin n\pi/2$ gives 0 and $\Gamma(n-1)$ gives infinity. For these cases

$$\sigma_T^2(R) = \frac{3\pi}{5} \mathcal{P}(R^{-1}) \,, \quad \frac{9}{4} \mathcal{P}(R^{-1}) \,, \quad \frac{3\pi}{2} \mathcal{P}(R^{-1}) \qquad \text{for } n = -2, \; -1, \; 0 \,. \tag{60}$$

For $n = 1$, $\sigma_G^2(R) = \frac{1}{2}\mathcal{P}(R^{-1})$.

Exercise: Derive these results for $\sigma_G^2(R)$ and $\sigma_k^2(R)$. ($\sigma_T^2(R)$ is more difficult.)

8.1.7 Galaxy 2-point correlation function

The most obvious way to try to measure the cosmological density perturbations is to observe the spatial distribution of galaxies. We treat individual galaxies as mathematical points, so

that each galaxy has a comoving coordinate value $\mathbf{x}$. We define the *galaxy 2-point correlation function* $\xi_g(\mathbf{r})$ as the *excess probability* of finding a galaxy at separation $\mathbf{r}$ from another galaxy:

$$dP \equiv \bar{n}\,[1 + \xi_g(\mathbf{r})]\,dV \tag{61}$$

where $\bar{n}$ is the mean galaxy number density, dV is a volume element that is a separation $\mathbf{r}$ away from a chosen reference galaxy, and dP is the probability that there is a galaxy within dV. (Here dV is assumed so small that there is at most one galaxy in it.)

If the galaxy number density $n(\mathbf{x})$ faithfully traces the underlying matter density, so that

$$\delta_g \equiv \frac{\delta n}{\bar{n}} = \delta \equiv \frac{\delta\rho_m}{\bar{\rho}_m}\,, \tag{62}$$

then ξ_g becomes equal to the matter density autocorrelation function ξ: The probability of finding a galaxy in volume dV_1 at a random location $\mathbf{x}$ is

$$dP_1 \;=\; \langle n(\mathbf{x})\rangle dV_1 \;=\; \langle \bar{n} + \delta n(\mathbf{x})\rangle dV_1 \;=\; \bar{n} dV_1\,. \tag{63}$$

The probability of finding a galaxy pair at $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$ is

$$\begin{aligned} dP_{12} &= \langle n(\mathbf{x})n(\mathbf{x}+\mathbf{r})\rangle dV_1 dV_2 \;=\; \bar{n}^2\langle[1+\delta(\mathbf{x})][1+\delta(\mathbf{x}+\mathbf{r})]\rangle dV_1 dV_2 \\ &= \bar{n}^2\left[1 + \langle\delta(\mathbf{x})\rangle + \langle\delta(\mathbf{x}+\mathbf{r})\rangle + \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\right] dV_1 dV_2 \\ &= \bar{n}^2\left[1 + \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\right] dV_1 dV_2\,, \end{aligned} \tag{64}$$

since $\langle\delta(\mathbf{x})\rangle = \langle\delta(\mathbf{x}+\mathbf{r})\rangle = 0$. Dividing dP_{12} with dP_1 we get the probability dP_2 of finding the second galaxy once we have found the first one

$$dP_2 \;=\; \bar{n}\left[1 + \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\right] dV_2 \;=\; \bar{n}\left[1 + \xi(\mathbf{r})\right] dV_2\,. \tag{65}$$

Thus $\xi_g = \xi$.

It is probable that the galaxy number density does not trace the matter density faithfully, since galaxy formation is likely to be more efficient in high-density regions. This is called *bias*. Specifically the bias, or *galaxy bias* b_g, is defined as the ratio

$$b_g \equiv \frac{\delta_g}{\delta_m} \quad\Rightarrow\quad \xi_g = b_g^2\xi\,, \tag{66}$$

where the expectation is that $b_g > 1$. In principle the bias could depend on the scale k, the time t (or redshift z), and/or the strength of the density perturbation δ_m. The simplest treatment of bias is to assume b_g is a constant over the observationally relevant ranges of these quantities.

The bias will depend on the type of tracer (all galaxies, specific types of galaxies, galaxy clusters) and is typically larger for more massive objects.

For the galaxy number density field, observationally $\sigma_T(R) \approx 1$ at $R = 8\,h^{-1}$Mpc. This has motivated the definition (will come later) of the quantity σ_8 as a cosmological parameter related to the amplitude of large-scale structure.

8.2 Newtonian perturbation theory

We shall now study the evolution of perturbations during the history of the universe. Initially the perturbations were small and we restrict the quantitative treatment to that part of the evolution when they remained small (for large scales, this extends to the present time and the future). This allows us to use *first-order perturbation theory*, where we drop from our equations all those terms which contain a product of two or more perturbations (as these products are even smaller). The remaining equations will then contain only terms which are either *zeroth order*, i.e., contain only background quantities, or *first order*, i.e., contain exactly one perturbation. If we kept only the zeroth order parts, we would be back to the equations of the homogenous and isotropic universe. Subtracting these from our equations we arrive at the *perturbation equations* where every term is first-order in the perturbation quantities, i.e., it is a *linear equation* for them. This makes the equations easy to handle, we can, e.g., Fourier transform them.

As we discovered in our discussion of inflation, the different cosmological distance scales first exit the horizon during inflation, then enter the horizon during various epochs of the later history. Matter perturbations at subhorizon scales, i.e., after horizon entry, can be treated with *Newtonian perturbation theory*, but scales which are close to horizon size or superhorizon require *relativistic perturbation theory*, which is based on general relativity.

The Newtonian equations for (perfect gas)[10] fluid dynamics with gravity are

$$\frac{\partial\rho}{\partial t'} + \nabla_{\mathbf{r}}\cdot(\rho\mathbf{u}) = 0 \tag{67}$$

$$\frac{\partial\mathbf{u}}{\partial t'} + (\mathbf{u}\cdot\nabla_{\mathbf{r}})\mathbf{u} + \frac{1}{\rho}\nabla_{\mathbf{r}}p + \nabla_{\mathbf{r}}\tilde{\Phi} = 0 \tag{68}$$

$$\nabla^2_{\mathbf{r}}\tilde{\Phi} = 4\pi G\rho \tag{69}$$

Here ρ is the mass density, p is the pressure, and $\mathbf{u}$ is the flow velocity of the fluid. We write $\tilde{\Phi}$ for the Newtonian gravitational potential, since we want to reserve Φ for its perturbation. The subscript $\mathbf{r}$ in $\nabla_{\mathbf{r}}$ emphasizes that the space derivatives are taken with respect to the Newtonian space coordinate $\mathbf{r}$ (instead of a comoving coordinate). Although the Newtonian time coordinate t' is equal to the cosmic time coordinate t, we need to make a distinction between t' and t in partial derivatives as will become clear soon.

The first equation is the law of mass conservation. The second equation is called the *Euler equation*, and it is just "$F = ma$" for a fluid element, whose mass is ρdV. Here the acceleration of a fluid element is not given by $\partial\mathbf{u}/\partial t'$ which just tells how the velocity field changes at a given position, but by $d\mathbf{u}/dt'$, where

$$\frac{d}{dt'} \equiv \frac{\partial}{\partial t'} + (\mathbf{u}\cdot\nabla_{\mathbf{r}}) \tag{70}$$

is the *convective time derivative*, which follows the fluid element as it moves. The two other terms give the forces due to pressure gradient and gravitational field.

We can apply Newtonian physics if:

1) Distance scales considered are $\ll$ the scale of curvature of spacetime (given by the Hubble length in cosmology[11])

2) The fluid flow is nonrelativistic, $u \ll c \equiv 1$.

3) We are considering nonrelativistic matter, $|p| \ll \rho$

[10] perfect gas = no internal friction $\Rightarrow$ pressure is isotropic

[11] As discussed in Chapter 3, the spacetime curvature has two distance scales, the Hubble length H^{-1} and the curvature radius $R_{\rm curv} \equiv a|K|^{-1/2}$. From observations we know that the curvature radius is larger than the Hubble length (at all times of interest), possibly infinite.

The last condition corresponds to particle velocities being nonrelativistic, if the matter is made out of particles. Although the pressure is small compared to mass density, the pressure gradient can be important if the pressure varies at small scales.

Note: Energy density and mass density. In Newtonian gravity, the source of gravity is mass density ρ_m, not energy density ρ. For nonrelativistic matter, the kinetic energies of particles are negligible compared to their masses, and thus so is the energy density compared to mass density, if we don't count the rest energy in it. The Newtonian equations for mass density and energy density are

$$\frac{\partial \rho_m}{\partial t'} + \nabla_{\mathbf{r}} \cdot (\rho_m \mathbf{u}) = 0 \tag{71}$$

$$\frac{\partial \rho_u}{\partial t'} + \nabla_{\mathbf{r}} \cdot (\rho_u \mathbf{u}) + p \nabla_{\mathbf{r}} \cdot \mathbf{u} = 0\,, \tag{72}$$

where $\nabla_{\mathbf{r}} \cdot \mathbf{u}$ gives the rate of change of the volume of the fluid element and $p\nabla_{\mathbf{r}} \cdot \mathbf{u}$ is the work done by pressure. In Newtonian physics, rest energy (mass) is not included in the energy density. Eq. (72) applies whether we include it or not. Define total energy density as

$$\rho \equiv \rho_m + \rho_u\,,$$

where ρ_u is the Newtonian energy density and ρ_m is the mass density. Adding Eqs. (71) and (72) gives

$$\frac{\partial \rho}{\partial t'} + \nabla_{\mathbf{r}} \cdot (\rho \mathbf{u}) + p \nabla_{\mathbf{r}} \cdot \mathbf{u} = 0\,. \tag{73}$$

For nonrelativistic matter $\rho_u \ll \rho_m$ and $p \ll \rho_m$. We can thus drop the last term in (73) and ignore the distinction between mass density and total energy density.

A homogeneously expanding fluid,

$$\rho = \rho(t_0) a^{-3} \tag{74}$$

$$\mathbf{u} = \frac{\dot{a}}{a}\mathbf{r} \tag{75}$$

$$\tilde{\Phi} = \frac{2\pi G}{3}\rho r^2 \tag{76}$$

is a solution to these equations (**exercise**), with a condition to the function $a(t)$ giving the expansion law. It is the Newtonian version of the matter-dominated Friedmann model. Writing $H(t) \equiv \dot{a}/a$ we find that the homogeneous solution satisfies

$$\dot{\rho} + 3H\rho = 0\,, \tag{77}$$

and the condition for $a(t)$ (from the exercise) can be written as

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = -\frac{4\pi G}{3}\rho\,. \tag{78}$$

You should recognize these equations as the energy-continuity equation and the second Friedmann equation for a matter-dominated FRW universe.[12] The result for $\tilde{\Phi}$, Eq. (76), has no relativistic counterpart, the whole concept of gravitational potential does not exist in relativity (except in special cases; like here in perturbation theory, where we introduce potentials related to perturbations).

[12] The freedom of choosing the initial value of the expansion rate leaves the connection between H and ρ open up to a constant. This constant has the same effect on the time evolution of $a(t)$ as the curvature constant K in the first Friedmann equation, but of course in the Newtonian treatment it is not interpreted as curvature, and it does not otherwise have the same physical effects. We shall (unless otherwise noted) choose this constant so that the background solution matches the *flat* FRW universe. Then we have

$$H^2 = \frac{8\pi G}{3}\rho \qquad \text{or} \qquad 4\pi G\rho = \frac{3}{2}H^2\,. \tag{79}$$

8.2.1 Comoving coordinates

Introduce now a new (comoving) coordinate system $(t, \mathbf{x})$ which is related to the Newtonian coordinate system $(t', \mathbf{r})$ by

$$t' = t \qquad \mathbf{r} = a(t)\mathbf{x}\,. \tag{80}$$

Thus the time coordinate is the same in both coordinate systems, but we need to distinguish between the partial derivatives $\partial/\partial t$ and $\partial/\partial t'$, since in the first $\mathbf{x}$ is kept constant and in the second $\mathbf{r}$ is kept constant. Relate now the partial derivatives:

$$\begin{aligned}
\frac{\partial}{\partial t} &= \frac{\partial t'}{\partial t}\frac{\partial}{\partial t'} + \sum_i \frac{\partial r_i}{\partial t}\frac{\partial}{\partial r_i} = \frac{\partial}{\partial t'} + \sum \dot{a} x_i \frac{\partial}{\partial r_i} = \frac{\partial}{\partial t'} + H\mathbf{r}\cdot\nabla_{\mathbf{r}} \\
\frac{\partial}{\partial x_i} &= \frac{\partial t'}{\partial x_i}\frac{\partial}{\partial t'} + \sum_j \frac{\partial r_j}{\partial x_i}\frac{\partial}{\partial r_j} = \sum_j \delta_{ij} a \frac{\partial}{\partial r_j} = a\frac{\partial}{\partial r_i} \quad \Rightarrow \quad \nabla_{\mathbf{x}} = a\nabla_{\mathbf{r}}\,.
\end{aligned} \tag{81}$$

Thus

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial t} - H\mathbf{x}\cdot\nabla_{\mathbf{x}} \qquad \text{and} \qquad \nabla_{\mathbf{r}} = \frac{1}{a}\nabla_{\mathbf{x}}\,. \tag{82}$$

(Later we will work exclusively in the comoving coordinates and write just ∇ for $\nabla_{\mathbf{x}}$. The "original" coordinates $\mathbf{r}$ are just an artifact of the Newtonian approach and do not appear in relativistic perturbation theory.)

8.2.2 The perturbation

Now, consider a small perturbation, so that

$$\rho(t', \mathbf{r}) = \bar{\rho}(t') + \delta\rho(t', \mathbf{r}) \tag{83}$$
$$p(t', \mathbf{r}) = \bar{p}(t') + \delta p(t', \mathbf{r}) \tag{84}$$
$$\mathbf{u}(t', \mathbf{r}) = H(t')\mathbf{r} + \mathbf{v}(t', \mathbf{r}) \tag{85}$$
$$\tilde{\Phi}(t', \mathbf{r}) = \frac{2\pi G}{3}\bar{\rho}(t')r^2 + \Phi(t', \mathbf{r})\,, \tag{86}$$

where $\bar{\rho}$, $\bar{p}$, and H denote homogeneous background quantities (solutions of the background, or zeroth-order, equations) and $\delta\rho$, δp, $\mathbf{v}$, Φ are small inhomogeneous perturbations.

Inserting these into the Eqs. (67,68,69) and subtracting the homogeneous equations (76,77,78) we get (**exercise**) the *perturbation equations*

$$\frac{\partial\delta\rho}{\partial t'} + 3H\delta\rho + H\mathbf{r}\cdot\nabla_{\mathbf{r}}\delta\rho + \bar{\rho}\nabla_{\mathbf{r}}\cdot\mathbf{v} = 0 \tag{87}$$
$$\frac{\partial\mathbf{v}}{\partial t'} + H\mathbf{v} + H\mathbf{r}\cdot\nabla_{\mathbf{r}}\mathbf{v} + \frac{1}{\bar{\rho}}\nabla_{\mathbf{r}}\delta p + \nabla_{\mathbf{r}}\Phi = 0 \tag{88}$$
$$\nabla^2_{\mathbf{r}}\Phi = 4\pi G\delta\rho\,. \tag{89}$$

In terms of the comoving coordinates these become (**exercise**):

$$\frac{\partial\delta\rho}{\partial t} + 3H\delta\rho + \frac{\bar{\rho}}{a}\nabla_{\mathbf{x}}\cdot\mathbf{v} = 0 \tag{90}$$
$$\frac{\partial\mathbf{v}}{\partial t} + H\mathbf{v} + \frac{1}{a\bar{\rho}}\nabla_{\mathbf{x}}\delta p + \frac{1}{a}\nabla_{\mathbf{x}}\Phi = 0 \tag{91}$$
$$\nabla^2_{\mathbf{x}}\Phi = 4\pi G a^2\delta\rho\,. \tag{92}$$

In terms of the relative density perturbation $\delta \equiv \delta\rho/\bar{\rho}$ we have $\delta\rho = \bar{\rho}\cdot\delta$ and

$$\frac{\partial\delta\rho}{\partial t} = \dot{\bar{\rho}}\cdot\delta + \bar{\rho}\frac{\partial\delta}{\partial t} \qquad \text{where} \qquad \dot{\bar{\rho}}\cdot\delta = -3H\bar{\rho}\,\delta\,, \tag{93}$$

and we can write

$$\frac{\partial \mathbf{v}}{\partial t} + H\mathbf{v} = \frac{1}{a}\frac{\partial}{\partial t}(a\mathbf{v}) \tag{94}$$

so that the set of perturbation equations becomes

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla_{\mathbf{x}} \cdot \mathbf{v} = 0 \tag{95}$$

$$\frac{\partial}{\partial t}(a\mathbf{v}) + \frac{1}{\bar{\rho}}\nabla_{\mathbf{x}}\delta p + \nabla_{\mathbf{x}}\Phi = 0 \tag{96}$$

$$\nabla^2_{\mathbf{x}}\Phi = 4\pi G a^2 \bar{\rho}\delta \tag{97}$$

Finally, we Fourier expand the perturbations,

$$\delta(t, \mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}} \qquad \text{etc.} \tag{98}$$

In Fourier space the perturbation equations become

$$\dot{\delta}_{\mathbf{k}} + \frac{i\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}}}{a} = 0 \tag{99}$$

$$\frac{d}{dt}(a\mathbf{v}_{\mathbf{k}}) + i\mathbf{k}\frac{\delta p_{\mathbf{k}}}{\bar{\rho}} + i\mathbf{k}\Phi_{\mathbf{k}} = 0 \tag{100}$$

$$\Phi_{\mathbf{k}} = -4\pi G \left(\frac{a}{k}\right)^2 \bar{\rho}\delta_{\mathbf{k}} \,. \tag{101}$$

Solving the evolution of the perturbations is a two-step process:

1) Solve the background equations to obtain the functions $a(t)$, $H(t)$, and $\bar{\rho}(t)$. After this, these are *known functions* in the perturbation equations.

2) Solve the perturbation equations.

8.2.3 Vector and scalar perturbations

We now divide the velocity perturbation field $\mathbf{v}(t, \mathbf{r})$ into its rotational (solenoidal, divergence-free) and irrotational (curl-free) parts,

$$\mathbf{v} = \mathbf{v}_\perp + \mathbf{v}_\parallel \,, \tag{102}$$

where $\nabla \cdot \mathbf{v}_\perp = 0$ and $\nabla \times \mathbf{v}_\parallel = 0$. For Fourier components this simply means that $\mathbf{k} \cdot \mathbf{v}_{\perp\mathbf{k}} = 0$ and $\mathbf{k} \times \mathbf{v}_{\parallel\mathbf{k}} = 0$. That is, we divide $\mathbf{v}_\mathbf{k}$ into the components perpendicular and parallel to the wave vector $\mathbf{k}$. The parallel part we can write in terms of a scalar function v, whose Fourier components $v_\mathbf{k}$ are given by

$$\mathbf{v}_{\parallel\mathbf{k}} \equiv v_{\mathbf{k}}\hat{\mathbf{k}} \,, \tag{103}$$

where $\hat{\mathbf{k}}$ denotes the unit vector in the $\mathbf{k}$ direction.

We can now take the perpendicular and parallel parts of Eq. (100),

$$\frac{d}{dt}(a\mathbf{v}_{\perp\mathbf{k}}) = 0 \tag{104}$$

$$\frac{d}{dt}(av_{\mathbf{k}}) + ik\frac{\delta p_{\mathbf{k}}}{\bar{\rho}} + ik\Phi_{\mathbf{k}} = 0 \,. \tag{105}$$

We see that the rotational part of the velocity perturbation has a simple time evolution,

$$\mathbf{v}_\perp \propto a^{-1} \,, \tag{106}$$

i.e., it *decays* from whatever initial value it had, inversely proportional to the scale factor.

The other perturbation equations involve only the irrotational part of the velocity perturbation. Thus we can divide the total perturbation into two parts, commonly called the *vector* and *scalar* perturbations, which evolve independent of each other:

1) The vector perturbation: $\mathbf{v}_\perp$.

2) The scalar perturbation: δ, δp, v, Φ, which are all coupled to each other.

The vector perturbations are thus not related to the density perturbations, or the structure of the universe. Also, any primordial vector perturbation should become rather small as the universe expands, at least while first-order perturbation theory applies.[13] They are thus not very important, and we shall have no more to say about them. The rest of our discussion focuses on the scalar perturbations.

8.2.4 The equations for scalar perturbations

We summarize here the Fourier space equations for scalar perturbations:

$$\dot{\delta}_\mathbf{k} + \frac{ikv_\mathbf{k}}{a} = 0 \quad \Rightarrow \quad v_\mathbf{k} = i\frac{a}{k}\dot{\delta}_\mathbf{k} \tag{107}$$

$$\frac{d}{dt}(av_\mathbf{k}) + ik\frac{\delta p_\mathbf{k}}{\bar{\rho}} + ik\Phi_\mathbf{k} = 0 \tag{108}$$

$$\Phi_\mathbf{k} = -4\pi G\left(\frac{a}{k}\right)^2 \bar{\rho}\delta_\mathbf{k}\,. \tag{109}$$

Inserting $v_\mathbf{k}$ from (107) and $\Phi_\mathbf{k}$ from (109) into (108) we get

$$\boxed{\ddot{\delta}_\mathbf{k} + 2H\dot{\delta}_\mathbf{k} = -\frac{k^2}{a^2}\frac{\delta p_\mathbf{k}}{\bar{\rho}} + 4\pi G\bar{\rho}\delta_\mathbf{k}\,.} \tag{110}$$

8.2.5 Adiabatic and entropy perturbations

Suppose the equation of state is *barotropic*,

$$p = p(\rho) \tag{111}$$

i.e., pressure is uniquely determined by the energy density. Then the perturbations δp and $\delta\rho$ are necessarily related by the derivative $dp/d\rho$ of this function $p(\rho)$,

$$p = \bar{p} + \delta p = \bar{p}(\bar{\rho}) + \frac{dp}{d\rho}(\bar{\rho})\delta\rho \quad \Rightarrow \quad \delta p = \frac{dp}{d\rho}\delta\rho\,.$$

The time derivatives of the background quantities $\bar{p}$ and $\bar{\rho}$ are related by this same derivative,

$$\dot{\bar{p}} = \frac{d\bar{p}}{dt} = \frac{dp}{d\rho}(\bar{\rho})\frac{d\bar{\rho}}{dt} = \frac{dp}{d\rho}\dot{\bar{\rho}}\,.$$

Assuming this derivative $dp/d\rho$ is nonnegative, we call its square root the *speed of sound*

$$c_s \equiv \sqrt{\frac{dp}{d\rho}}\,. \tag{112}$$

[13] Thus we end up with an irrotational velocity field. The rotational motion (e.g., rotation of galaxies) which is common in the present universe at small scales has arisen from higher-order effects from the primordial scalar perturbations, not from the primordial vector perturbations.

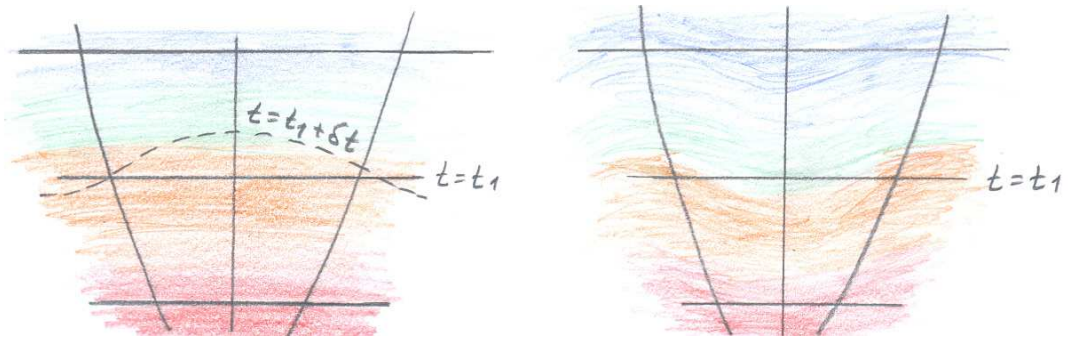


Figure 5: For adiabatic perturbations, the conditions in the perturbed universe (right) at $(t_1, \mathbf{x})$ equal conditions in the (homogeneous) background universe (left) at some time $t_1 + \delta t(\mathbf{x})$.

(We shall indeed find that sound waves propagate at this speed.) We thus have the relation

$$\frac{\delta p}{\delta \rho} = \frac{\dot{\bar{p}}}{\dot{\bar{\rho}}} = c_s^2 \,.$$

In general, when p may depend on other variables besides ρ, the speed of sound in a fluid is given by

$$c_s^2 = \left(\frac{\partial p}{\partial \rho}\right)_S \tag{113}$$

where the subscript S indicates that the derivative is taken so that the entropy of the fluid element is kept constant. Since the background universe expands adiabatically (meaning that there is no entropy production), we have that

$$\frac{\dot{\bar{p}}}{\dot{\bar{\rho}}} = \left(\frac{\partial p}{\partial \rho}\right)_S = c_s^2 \,. \tag{114}$$

Perturbations with the property

$$\frac{\delta p}{\delta \rho} = \frac{\dot{\bar{p}}}{\dot{\bar{\rho}}} \tag{115}$$

are called *adiabatic perturbations* in cosmology.

If $p = p(\rho)$, perturbations are necessarily adiabatic. In the general case the perturbations may or may not be adiabatic. In the latter case, the perturbation can be divided into an adiabatic component and an *entropy perturbation*. An entropy perturbation is a perturbation in the entropy-per-particle ratio.

For adiabatic perturbations we thus have

$$\delta p = c_s^2 \delta \rho = \frac{\dot{\bar{p}}}{\dot{\bar{\rho}}} \delta \rho \,. \tag{116}$$

Adiabatic perturbations have the property that the local state of matter (determined here by the quantities p and ρ) at some spacetime point $(t, \mathbf{x})$ of the perturbed universe is the same as in the background universe at some slightly different time $t + \delta t$, this time difference being different for different locations $\mathbf{x}$. See Fig. 5.

Thus we can view adiabatic perturbations as some parts of the universe being "ahead" and others "behind" in the evolution.

Adiabatic perturbations are the simplest kind of perturbations. Single-field inflation produces adiabatic perturbations, since perturbations in all quantities are proportional to a perturbation $\delta\varphi$ in a single scalar quantity, the inflaton field.

Adiabatic perturbations stay adiabatic while they are outside horizon, but may develop entropy perturbations when they enter the horizon. This happens for many-component fluids (discussed a little later).

Present observational data is consistent with the primordial (i.e., before horizon entry) perturbations being adiabatic.

8.2.6 Adiabatic perturbations in matter

Consider now adiabatic perturbations of a non-relativistic single-component fluid. The equation for the density perturbation is now

$$\boxed{\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \left[\frac{c_s^2 k^2}{a^2} - 4\pi G\bar{\rho}\right]\delta_{\mathbf{k}} = 0\,.} \tag{117}$$

I shall call this the *Jeans equation*[14] (although Jeans considered a static, not an expanding fluid).

This is a second-order differential equation from which we can solve the time evolution of the Fourier amplitudes $\delta_{\mathbf{k}}(t)$ of the perturbation. Before solving this equation we need to first find the background solution which gives the functions $a(t)$, $H(t) = \dot{a}/a$, and $\bar{\rho}(t)$.

The nature of the solution to Eq. (117) depends on the sign of the factor in the brackets. The first term in the brackets is due to pressure gradients. Pressure tries to resist compression, so if this term dominates, we get an oscillating solution, standing density (sound) waves. The second term in the brackets is due to gravity. If this term dominates, the perturbations grow. The wavenumber for which the terms are equal,

$$k_J \;=\; \frac{a\sqrt{4\pi G\bar{\rho}}}{c_s} \;=\; \sqrt{\frac{3}{2}}\frac{1}{c_s}\mathcal{H}\,, \tag{118}$$

is called the *Jeans wave number*, and the corresponding wavelength

$$\lambda_J \;=\; \frac{2\pi}{k_J} \;=\; 2\pi c_s\sqrt{\frac{2}{3}}\mathcal{H}^{-1} \tag{119}$$

the *Jeans length*. Here $\mathcal{H} \equiv aH$, the comoving Hubble parameter. In the latter equalities we assumed that the background solution is the flat FRW universe, so that

$$4\pi G\bar{\rho} = \tfrac{3}{2}H^2\,. \tag{120}$$

For nonrelativistic matter $c_s \ll 1$, so that the Jeans length is much smaller than the Hubble length, $k_J \gg \mathcal{H}$. Thus we can apply Newtonian theory for scales both larger and smaller than the Jeans length.

For **scales much smaller that the Jeans length**, $k \gg k_J$, we can approximate the Jeans equation by

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \frac{c_s^2 k^2}{a^2}\delta_{\mathbf{k}} = 0\,. \tag{121}$$

The solutions are oscillating, i.e., we get sound waves. The exact solutions of (121) are Bessel functions, but for small scales we can make a further approximation by first ignoring the middle term (which is smaller than the other two) and the time-dependence of a and c_s to get that $\delta_{\mathbf{k}}(t) \sim e^{\pm i\omega t}$, where $\omega = c_s k/a$. These oscillations are damped by the $2H\dot{\delta}_{\mathbf{k}}$ term, so the amplitude of the oscillations decreases with time. There is no growth of structure for sub-Jeans scales.

[14] In the literature, there is usually no name given to this equation, but the terms *Jeans length* etc. are standard.

Exercise: Sound waves. For short-wavelength modes $k \gg k_J$, density perturbations in the matter-dominated universe satisfy (121). Switch to conformal time, $d\eta = dt/a$, and solve $\delta_{\mathbf{k}}(\eta)$ for the $\Omega_m = 1$, $\Omega_\Lambda = 0$ cosmology, assuming $c_s = const.$ How does the amplitude and frequency of the oscillations change with time and scale factor? (Hint: The solutions are spherical Bessel functions.)

For **scales much longer than the Jeans length** (but still subhorizon), $\mathcal{H} \ll k \ll k_J$, we can approximate the Jeans equation by

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}\delta_{\mathbf{k}} = 0\,. \tag{122}$$

We dropped the pressure gradient term, which means that this equation applies also to nonadiabatic perturbations for scales where pressure gradients can be ignored. Note that Eq. (122) is the same for all $\mathbf{k}$, i.e., there is no k-dependence in the coefficients. This means that the equation applies also in coordinate space, i.e. for $\delta(\mathbf{x})$, as long as we ignore contributions from scales that do not satisfy $\mathcal{H} \ll k \ll k_J$.

For a matter-dominated universe, the background solution is $a \propto t^{2/3}$, so that

$$H = \frac{\dot{a}}{a} = \frac{2}{3t} \tag{123}$$

and

$$\frac{8\pi G}{3}\bar{\rho} = H^2 = \frac{4}{9t^2} \qquad \Rightarrow \qquad \bar{\rho} = \frac{1}{6\pi G t^2}\,, \tag{124}$$

so the Jeans equation becomes

$$\ddot{\delta}_{\mathbf{k}} + \frac{4}{3t}\dot{\delta}_{\mathbf{k}} - \frac{2}{3t^2}\delta_{\mathbf{k}} = 0\,. \tag{125}$$

The general solution is

$$\delta_{\mathbf{k}}(t) = b_1 t^{2/3} + b_2 t^{-1}\,. \tag{126}$$

The first term is the *growing mode* and the second term the *decaying mode*. After some time the decaying mode has died out, and the perturbation grows

$$\boxed{\delta \propto t^{2/3} \propto a\,.} \tag{127}$$

Thus *density perturbations in matter grow proportional to the scale factor.*

From Eq. (101) we have that

$$\Phi \propto a^2\bar{\rho}\delta \propto a^2 a^{-3} a = \text{const}.$$

The gravitational potential perturbation is constant in time during the matter-dominated era.

8.2.7 Radiation

Since radiation is a relativistic form of energy, we cannot apply the preceding Newtonian discussion to perturbations in radiation. However, the qualitative results are similar.

The equation of state for radiation is $p = \rho/3$, and the speed of sound in a radiation fluid is given by

$$c_s^2 = \frac{dp}{d\rho} = \frac{1}{3}\,.$$

Thus the Jeans length for radiation is comparable to the Hubble length, and the subhorizon scales are also sub-Jeans scales for radiation. Thus for subhorizon radiation perturbations we only get oscillatory solutions. During the radiation-dominated epoch they are not damped by expansion, but the oscillation amplitude stays roughly constant.

Since

$$\Phi \propto a^2 \bar{\rho}\delta \propto a^{-2}\delta\,,$$

the amplitude of the gravitational potential oscillation decays.

Relativistic perturbations in non-expanding space. While the full treatment of relativistic perturbations is beyond the level of this course, we can obtain the limit where we ignore the effect of expansion by combining special relativity and the Newtonian limit of general relativity.

Special relativistic fluid dynamics follows from the energy-momentum continuity equation

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu} \equiv \partial_\nu T^{\mu\nu} \equiv T^{\mu\nu}{}_{,\nu} \;=\; 0\,. \tag{128}$$

For a perfect fluid

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu \pm p g^{\mu\nu}\,, \tag{129}$$

where the metric is now that of Minkowski space, $g^{\mu\nu} = \mathrm{diag}(-1,1,1,1)$. The 4-velocity u^μ is related to the 3-velocity $\mathbf{v} = v^i$ by

$$u^\mu = (\gamma, \gamma\mathbf{v})\,, \tag{130}$$

where $\gamma = 1/\sqrt{1-v^2}$.

By contracting the energy tensor $T^{\mu\nu}$ with the 4-velocity u_μ we obtain $u_\nu T^{\mu\nu}{}_{,\nu} = 0$, which gives

$$(\rho u^\mu)_{,\mu} + p u^\mu{}_{,\mu} \;=\; 0\,, \tag{131}$$

the energy continuity equation. Subtracting u^ν times this from (128) we get the special relativistic Euler equation

$$(\rho + p)u^\mu u^\nu{}_{,\mu} + (g^{\mu\nu} + u^\mu u^\nu)p_{,\mu} \;=\; 0\,, \tag{132}$$

where

$$u^\mu u^\nu{}_{,\mu} \equiv a^\nu \tag{133}$$

is the 4-acceleration.

For small velocities, $v \ll 1$, we can approximate $\gamma \approx 1$, so that

$$u^\mu \approx (1, \mathbf{v}) \tag{134}$$

and (131),(132) become

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) &= -p\nabla\cdot\mathbf{v} \\ (\rho+p)\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\vec{v} &= -\nabla p - \mathbf{v}(\mathbf{v}\cdot\nabla p) \;\approx\; -\nabla p\,. \end{aligned} \tag{135}$$

In the Newtonian limit of general relativity, but without the assumption $p \ll \rho$, the passive gravitational mass density is given by $\rho + p$, so that the gravitational force on a volume element of fluid is given by $-(\rho+p)\nabla\Phi$ and the active gravitational mass density by $\rho + 3p$, so that the gravitational potential is given by

$$\nabla^2\Phi = 4\pi G(\rho + 3p)\,. \tag{136}$$

Thus the Euler equation with gravity becomes

$$(\rho+p)\left(\frac{\partial}{\partial t} \pm \mathbf{v}\cdot\nabla\right)\mathbf{v} \;\equiv\; -\nabla p - (\rho+p)\nabla\Phi\,. \tag{137}$$

For several fluid components, not interacting with each other except gravitationally, the fluid equations become thus

$$\begin{aligned} \frac{\partial \rho_i}{\partial t} + \nabla\cdot(\rho_i\mathbf{v}_i) &= -p_i\nabla\cdot\mathbf{v}_i \\ (\rho_i+p_i)\left(\frac{\partial}{\partial t} + \mathbf{v}_i\cdot\nabla\right)\mathbf{v}_i &= -\nabla p_i - (\rho_i+p_i)\nabla\Phi \\ \nabla^2\Phi &= 4\pi G\sum_i(\rho_i + 3p_i)\,. \end{aligned} \tag{138}$$

For perturbations $\rho_i = \bar{\rho}_i + \delta\rho_i = \bar{\rho}_i(1+\delta_i)$, $p_i = \bar{p}_i + \delta p_i$, where the background density and pressure are now constant both in space and time, we get to first order in perturbations

$$\begin{aligned} \frac{\partial \delta_i}{\partial t} &= -(1+w_i)\nabla\cdot\mathbf{v}_i \\ (\bar{\rho}_i+\bar{p}_i)\frac{\partial \mathbf{v}_i}{\partial t} &= -\nabla\delta p_i - (\bar{\rho}_i+\bar{p}_i)\nabla\Phi \\ \nabla^2\Phi &= 4\pi G\sum_i(\bar{\rho}_i\delta_i + 3\delta p_i)\,, \end{aligned} \tag{139}$$

where $w_i \equiv \bar{p}_i/\bar{\rho}_i$. For Fourier components this becomes

$$\begin{aligned} \dot{\delta}_{i\mathbf{k}} &= -ik(1+w_i)\mathbf{k}\cdot\mathbf{v}_{i\mathbf{k}} \\ (\bar{\rho}_i+\bar{p}_i)\dot{\mathbf{v}}_{i\mathbf{k}} &= -i\mathbf{k}\delta p_{i\mathbf{k}} - i\mathbf{k}(\bar{\rho}_i+\bar{p}_i)\Phi_{\mathbf{k}} \\ \Phi_{\mathbf{k}} &= \frac{-4\pi G}{k^2}\sum_i(\bar{\rho}_i\delta_{i\mathbf{k}} + 3\delta p_{i\mathbf{k}})\,. \end{aligned} \tag{140}$$

For vector perturbations the second equation gives

$$\dot{\mathbf{v}}_{i\perp\mathbf{k}} = 0 \quad \Rightarrow \quad \mathbf{v}_{i\perp\mathbf{k}} = const\,, \tag{141}$$

and for scalar perturbations the first and second equations become

$$\begin{aligned} \dot{\delta}_{i\mathbf{k}} &= -i(1+w_i)kv_{i\mathbf{k}} \\ \dot{v}_{i\mathbf{k}} &= -ik\frac{\delta p_{i\mathbf{k}}}{\bar{\rho}_i+\bar{p}_i} - ik\Phi_{\mathbf{k}}\,. \end{aligned} \tag{142}$$

If $w_i = const$, so that $\dot{w}_i = 0$, we get the Jeans equation

$$\ddot{\delta}_{i\mathbf{k}} + k^2\frac{\delta p_{i\mathbf{k}}}{\bar{\rho}_i} + k^2(1+w_i)\Phi_{\mathbf{k}} = 0\,. \tag{143}$$

8.2.8 Many fluid components

Assume now that the "cosmic fluid" contains several components i (different types of matter or energy) which *do not interact with each other*, except gravitationally. This means that each component feels only its own pressure[15], and that the components can have different flow velocities. Then the Newtonian equations for each component i are

$$\frac{\partial \rho_i}{\partial t'} + \nabla_{\mathbf{r}}\cdot(\rho_i\mathbf{u}_i) = 0 \tag{144}$$

$$\frac{\partial \mathbf{u}_i}{\partial t'} + (\mathbf{u}_i\cdot\nabla_{\mathbf{r}})\mathbf{u}_i + \frac{1}{\rho_i}\nabla_{\mathbf{r}}p_i + \nabla_{\mathbf{r}}\tilde{\Phi} = 0 \tag{145}$$

$$\nabla^2_{\mathbf{r}}\tilde{\Phi} = 4\pi G\rho\,, \tag{146}$$

where $\rho = \sum\rho_i$. Note that there is only one gravitational potential $\tilde{\Phi}$, due to the total density, and this way the different components do interact gravitationally.

We again have the homogeneous solution, where now each component has to satisfy

$$\dot{\rho}_i + 3H\rho_i = 0\,, \tag{147}$$

[15] In standard cosmology, we actually have just one component, the baryon-photon fluid, which feels its own pressure, and the other components do not feel even *their own* pressure (neutrinos after decoupling) or do not even *have* pressure (cold dark matter). But we shall first do this general treatment, and do the application to standard cosmology later.

and the expansion law

$$\dot{H} + H^2 = -\frac{4\pi G}{3}\rho\,, \tag{148}$$

is determined by the total density.

We can now introduce the density, pressure, and velocity perturbations for each component separately,

$$\begin{aligned}
\rho_i(t',\mathbf{r}) &= \bar{\rho}_i(t) + \delta\rho_i(t',\mathbf{r}) && (149)\\
p_i(t',\mathbf{r}) &= \bar{p}_i(t) + \delta p_i(t',\mathbf{r}) && (150)\\
\mathbf{u}_i(t',\mathbf{r}) &= H(t)\mathbf{r} + \mathbf{v}_i(t',\mathbf{r})\,, && (151)
\end{aligned}$$

but there is only one gravitational potential perturbation,

$$\tilde{\Phi}(t',\mathbf{r}) = \frac{2\pi G}{3}\bar{\rho}r^2 + \Phi(t',\mathbf{r})\,. \tag{152}$$

Following the earlier procedure, we obtain the perturbation equations for the fluid components,

$$\begin{aligned}
\frac{\partial}{\partial t'}\delta\rho_i + 3H\delta\rho_i + H\mathbf{r}\cdot\nabla_{\mathbf{r}}\delta\rho_i + \bar{\rho}_i\nabla\cdot\mathbf{v}_i &= 0 && (153)\\
\frac{\partial}{\partial t'}\mathbf{v}_i + H\mathbf{v}_i + H\mathbf{r}\cdot\nabla_{\mathbf{r}}\mathbf{v}_i + \frac{1}{\bar{\rho}_i}\nabla_{\mathbf{r}}\delta p_i + \nabla_{\mathbf{r}}\Phi &= 0 && (154)\\
\nabla^2_{\mathbf{r}}\Phi &= 4\pi G\delta\rho && (155)
\end{aligned}$$

in Newtonian coordinate space, and

$$\begin{aligned}
\dot{\delta}_{i\mathbf{k}} + \frac{i\mathbf{k}\cdot\mathbf{v}_{i\mathbf{k}}}{a} &= 0 && (156)\\
\frac{d}{dt}\left(a\mathbf{v}_{i\mathbf{k}}\right) + i\mathbf{k}\frac{\delta p_{i\mathbf{k}}}{\bar{\rho}_i} + i\mathbf{k}\Phi_{\mathbf{k}} &= 0 && (157)\\
\Phi_{\mathbf{k}} &= -4\pi G\frac{a^2}{k^2}\sum\bar{\rho}_i\delta_{i\mathbf{k}} && (158)
\end{aligned}$$

in comoving Fourier space. Here $\delta\rho = \sum\delta\rho_i$ and

$$\delta_i \equiv \frac{\delta\rho_i}{\bar{\rho}_i}\,. \tag{159}$$

Separating out the scalar perturbations we finally get

$$\ddot{\delta}_{i\mathbf{k}} + 2H\dot{\delta}_{i\mathbf{k}} = -\frac{k^2}{a^2}\frac{\delta p_{i\mathbf{k}}}{\bar{\rho}_i} + 4\pi G\delta\rho_{\mathbf{k}}\,, \tag{160}$$

where

$$\delta\rho_{\mathbf{k}} = \sum_j \bar{\rho}_j\delta_{j\mathbf{k}}\,. \tag{161}$$

8.2.9 Adiabatic and entropy perturbations again

The simplest inflation models predict that the primordial perturbations are adiabatic. This means that locally the perturbed universe at some $(t,\mathbf{x})$ looks like the background universe at some time $t + \delta t(\mathbf{x})$. See Sec. 8.2.5.

$$\left.\begin{aligned}\delta\rho_i(\mathbf{x}) &= \dot{\bar{\rho}}_i\delta t(\mathbf{x})\\ \delta p_i(\mathbf{x}) &= \dot{\bar{p}}_i\delta t(\mathbf{x})\end{aligned}\right\} \quad\Rightarrow\quad \begin{cases}\dfrac{\delta p_i}{\delta\rho_i} = \dfrac{\dot{\bar{p}}_i}{\dot{\bar{\rho}}_i}\\[2ex] \dfrac{\delta\rho_i}{\delta\rho_j} = \dfrac{\dot{\bar{\rho}}_i}{\dot{\bar{\rho}}_j}\end{cases} \quad\Rightarrow\quad \frac{\delta_i}{\delta_j} = \frac{\dot{\bar{\rho}}_i}{\bar{\rho}_i}\frac{\bar{\rho}_j}{\dot{\bar{\rho}}_j} \tag{162}$$

If there is no energy transfer between the fluid components at the background level, the energy continuity equation is satisfied by them separately,

$$\dot{\bar{\rho}}_i \;=\; -3H(\bar{\rho}_i + \bar{p}_i) \;\equiv\; -3H(1+w_i)\bar{\rho}_i\,, \tag{163}$$

where $w_i \equiv \bar{p}_i/\bar{\rho}_i$ is the *equation-of-state parameter* of fluid component i. Thus for adiabatic perturbations,

$$\frac{\delta_i}{1+w_i} = \frac{\delta_j}{1+w_j} \tag{164}$$

(which is thus related to $\bar{\rho}_i \propto a^{-3(1+w_i)}$). For matter components $w_i \approx 0$, and for radiation components $w_i = \frac{1}{3}$. Thus, for adiabatic perturbations, all matter components have the same perturbation

$$\delta_i = \delta_m$$

and all radiation perturbations have likewise

$$\delta_i = \delta_r = \frac{4}{3}\delta_m\,.$$

We can define a *relative entropy perturbation*[16] between two components

$$S_{ij} \;\equiv\; -3H\left(\frac{\delta\rho_i}{\dot{\bar{\rho}}_i} - \frac{\delta\rho_j}{\dot{\bar{\rho}}_j}\right) \;=\; \frac{\delta_i}{1+w_i} - \frac{\delta_j}{1+w_j} \tag{165}$$

to describe a deviation from the adiabatic case. The relative entropy perturbation is a perturbation in the ratio of the number densities of the two species. For a nonrelativistic species

$$\rho_i = m_i n_i \quad\Rightarrow\quad \delta\rho_i = m_i\delta n_i \qquad \text{and} \qquad \delta_i \equiv \frac{\delta\rho_i}{\bar{\rho}_i} = \frac{\delta n_i}{\bar{n}_i}\,, \tag{166}$$

whereas for an ultrarelativistic species ($\mu \ll T$ and $m \ll T$)

$$\begin{aligned} \rho_i \;&\propto\; T_i^4 \quad\Rightarrow\quad \delta\rho_i = \bar{\rho}_i \cdot 4\frac{\delta T_i}{T_i} \\ n_i \;&\propto\; T_i^3 \quad\Rightarrow\quad \delta n_i = \bar{n}_i \cdot 3\frac{\delta T_i}{T_i} \\ &\Rightarrow\quad \delta_i \equiv \frac{\delta\rho_i}{\bar{\rho}_i} = \frac{4}{3}\frac{\delta n_i}{\bar{n}_i}\,. \end{aligned} \tag{167}$$

For both cases

$$\delta_i = (1+w_i)\frac{\delta n_i}{\bar{n}_i}\,. \tag{168}$$

Thus

$$S_{ij} \;=\; \frac{\delta n_i}{\bar{n}_i} - \frac{\delta n_j}{\bar{n}_j} \;=\; \frac{\delta(n_i/n_j)}{\bar{n}_i/\bar{n}_j}\,. \tag{169}$$

Even if perturbations are initially adiabatic, relative entropy perturbation may develop inside the horizon. We shall encounter such a case in Sec. 8.3.4.

[16] There is a connection to entropy/particle of the different components, but we need not concern ourselves with it now. It is not central to this concept, and it is perhaps somewhat unfortunate that it has become customary, for historical reasons, to use the word "entropy" for these perturbations.

8.2.10 The effect of a homogeneous component

The energy density of the real universe consists of several components. In many cases it is reasonable to ignore the perturbations in some components (since they are relatively small in the scales of interest). We call such components *smooth* and we can add them together into a single smooth component $\rho_s = \bar{\rho}_s$.

Consider the case where we have perturbations in a nonrelativistic ("matter") component ρ_m, and the other components are smooth. Then

$$\rho = \rho_m + \rho_s \tag{170}$$

but

$$\delta\rho = \delta\rho_m \equiv \bar{\rho}_m \delta \,. \tag{171}$$

We write just δ for $\delta\rho_m/\bar{\rho}_m$, since there is no other density perturbation, but note that now $\delta \neq \delta\rho/\bar{\rho}$ (beware of this trap!).

Assuming adiabatic perturbations, we have then from Eq. (160) that

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \left[\frac{c_s^2 k^2}{a^2} - 4\pi G\bar{\rho}_m\right]\delta_{\mathbf{k}} = 0 \,. \tag{172}$$

The difference from Eq. (117) is that now the background energy density in the "gravity" term still contains only the matter component $\bar{\rho}_m$, but the expansion law, $a(t)$ and $H(t)$ comes from the full background energy density $\bar{\rho} = \bar{\rho}_m + \bar{\rho}_s$.

Newtonian perturbation theory can be applied even with the presence of relativistic energy components, like radiation and dark energy, as long as they can be considered as smooth components and their perturbations can be ignored. Then they contribute only to the background solution. In this case we have to calculate the background solution using general relativity, i.e., the background solution is a FRW universe, but the perturbation equations are the Newtonian perturbation equations. We can also consider a non-flat (open or closed) FRW universe, as long as we only apply perturbation theory to scales much shorter than the curvature radius (and the Hubble length). Thus the background quantities are to be solved from the Friedmann and energy continuity equations

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho \tag{173}$$

$$\dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3p) \tag{174}$$

$$\dot{\rho} = -3H(\rho + p) \,. \tag{175}$$

Example: Matter perturbations in flat vacuum-dominated universe. Consider the case where $\rho_s = \rho_{\rm vac} \gg \rho_m$ and matter is approximated as pressureless (we do not then have to make a separate adiabaticity assumption, since the pressure term does not appear). Then the Jeans equation becomes

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}_m\delta_{\mathbf{k}} = 0 \,. \tag{176}$$

To estimate the relative order of magnitude of the three terms it is better to divide the equation by H^2,

$$H^{-2}\ddot{\delta}_{\mathbf{k}} + 2H^{-1}\dot{\delta}_{\mathbf{k}} - \frac{4\pi G\bar{\rho}_m}{H^2}\delta_{\mathbf{k}} = 0 \,, \tag{177}$$

so that the Hubble time H^{-1} provides the time scale for the time derivatives. Now

$$H^2 = \frac{8\pi G}{3}\rho_{\rm cr} \approx \frac{8\pi G}{3}\rho_{\rm vac} = const \tag{178}$$

and in the last term $\delta_{\mathbf{k}}$ is multiplied with $\frac{3}{2}\bar{\rho}_m/\rho_{\rm vac} \ll 1$, so that we can drop the last term and approximate the Jeans equation by

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} \;=\; 0\,. \tag{179}$$

We see immediately that $\delta_{\mathbf{k}} = const$ is a solution. For the other solution, solve first $\dot{\delta}_{\mathbf{k}}$:

$$\frac{d\dot{\delta}_{\mathbf{k}}}{dt} = -2H\dot{\delta}_{\mathbf{k}} \quad \Rightarrow \quad \frac{d\dot{\delta}_{\mathbf{k}}}{\dot{\delta}_{\mathbf{k}}} = -2H dt\,, \tag{180}$$

whose solution is $\ln\dot{\delta}_{\mathbf{k}} = -2Ht + const$ or $\dot{\delta}_{\mathbf{k}} = Ce^{-2Ht}$. Integrating this gives

$$\delta_{\mathbf{k}} \;=\; Ae^{-2Ht} + B\,, \tag{181}$$

with a constant term and an exponentially decaying term. Thus in a vacuum-dominated universe matter perturbations stay constant (after the decaying term has died out); or to be more precise and referring to the original equation (177), the relative change in $\delta_{\mathbf{k}}$ in a Hubble time is of order $\bar{\rho}_m/\rho_{\rm vac} \ll 1$

We shall do this calculation more accurately later, including the transition from matter domination to vacuum domination. The main lesson now is that the increased expansion rate due to the presence of a smooth component slows down the growth of perturbations.

Exercise: Find the solution for the Jeans equation for pressureless matter perturbations when a) the energy density is dominated by a smooth radiation component b) when there is no other energy component, but the universe has the open geometry ($K < 0$) and is curvature dominated, considering only scales $\ll$ curvature radius.

8.2.11 Meszaros equation

Consider now a flat universe with just cold dark matter (CDM) and radiation, and ignore perturbations in radiation so that radiation can be taken as a smooth component. This approximation may be motivated by noting that subhorizon radiation perturbations do not grow. The CDM is pressureless, and thus the CDM sound speed is zero, and so is the CDM Jeans length. Thus, for CDM, all scales are larger than the Jeans scale, and we don't get an oscillatory behavior. Instead, perturbations grow at all scales.

We get the equation for the CDM perturbation from Eq. (172) by setting $c_s = 0$ (or rather, $\delta p = 0$; we need not invoke the assumption of adiabaticity, since CDM is pressureless),

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}_m\delta_{\mathbf{k}} = 0\,. \tag{182}$$

Note that the equation is the same for all $\mathbf{k}$ and therefore it applies also in the coordinate space, i.e., for $\delta(\mathbf{x})$. To simplify notation, we drop the subscript $_{\mathbf{k}}$.

The Friedmann equation is

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\bar{\rho}\,,$$

where $\bar{\rho} = \bar{\rho}_m + \bar{\rho}_r$ and $\bar{\rho}_m \propto a^{-3}$ and $\bar{\rho}_r \propto a^{-4}$.

A useful trick is to study this as a function of a instead of t or η. We define a new time coordinate,

$$y \equiv \frac{a}{a_{\rm eq}} = \frac{\bar{\rho}_m}{\bar{\rho}_r}\,. \tag{183}$$

($y = 1$ at $t = t_{\rm eq}$.) Now

$$4\pi G\bar{\rho}_m \;=\; 4\pi G\frac{y}{y+1}\bar{\rho} \;=\; \frac{3}{2}\frac{y}{y+1}H^2 \tag{184}$$

and Eq. (182) becomes

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\frac{y}{y+1}H^2 = 0\,. \tag{185}$$

Performing the change of variables from t to y (**Exercise**; you may need the 2nd Friedmann equation), we arrive at the equation (where $' \equiv d/dy$)

$$\boxed{\delta'' + \frac{2+3y}{2y(1+y)}\delta' - \frac{3}{2y(1+y)}\delta \;=\; 0\,,} \tag{186}$$

known as the *Meszaros equation*.

It has two solutions, one growing, the other one decaying. The growing solution is

$$\boxed{\delta \;=\; \delta_{\rm prim}\left(1+\frac{3y}{2}\right) \;=\; \delta_{\rm prim}\left(1+\frac{3}{2}\frac{a}{a_{\rm eq}}\right)\,.} \tag{187}$$

We see that the perturbation remains frozen to its primordial value, $\delta \approx \delta_{\rm prim}$, during the radiation-dominated period. By $t = t_{\rm eq}$, it has grown to $\delta = \frac{5}{2}\delta_{\rm prim}$.

During the matter-dominated period, $y \gg 1$, the CDM perturbation grows proportional to the scale factor,

$$\delta \propto y \propto a \propto t^{2/3}\,. \tag{188}$$

8.3 Perturbations at subhorizon scales in the real universe

8.3.1 Horizon entry

Newtonian perturbation theory is valid only at subhorizon scales, $k \gg \mathcal{H}$, or $k^{-1} \ll \mathcal{H}^{-1}$. During "normal", decelerating expansion, i.e., after inflation but before the recent onset of dark energy domination, scales are entering the horizon. Short scales enter first, large scales enter later. We have not yet studied what happens to perturbations outside the horizon (for that we need (general) relativistic perturbation theory, to be discussed In Sec. 8.4). So, for the present discussion, whatever values the perturbation amplitudes $\delta_{\mathbf{k}}$ have soon after horizon entry, are to be taken as an initial condition, the *primordial perturbation*[17]. Observations actually suggest that different scales enter the horizon with approximately equal perturbation amplitude, whose magnitude is characterized by the number[18] few $\times\, 10^{-5}$.

The history of the different scales after horizon entry, and thus their present perturbation amplitude, depends on at what epoch they enter. The scales which enter during transitions between epochs are thus special scales which should characterize the present structure of the universe. Such important scales are the scale (**exercise**)

$$k_{\rm eq}^{-1} \;=\; (\mathcal{H}_{\rm eq})^{-1} \;\sim\; 13.7\,\Omega_m^{-1} h^{-2}\,{\rm Mpc} \;\equiv\; 13.7\,\omega_m^{-1}\,{\rm Mpc}\,, \tag{189}$$

which enters at the time $t_{\rm eq}$ ($1 + z_{\rm eq} = 23\,902\,\omega_m$) of matter-radiation equality, and the scale

$$\begin{aligned} k_{\rm dec}^{-1} \;&=\; (\mathcal{H}_{\rm dec})^{-1} \;\sim\; 91\,\Omega_m^{-1/2}\left[1 + \frac{\Omega_r}{\Omega_m}(1+z_{\rm dec})\right]^{-1/2} h^{-1}{\rm Mpc} \\ &\equiv\; 91\,\omega_m^{-1/2}\left[1 + \frac{\omega_r}{\omega_m}(1+z_{\rm dec})\right]^{-1/2} {\rm Mpc}\,, \end{aligned} \tag{190}$$

which enters at the time $t_{\rm dec}$ ($z_{\rm dec} = 1090$) of photon decoupling. Here $\omega_r = 4.184 \times 10^{-5}$ includes relativistic neutrinos, since the result above only requires them to be relativistic at $t_{\rm dec}$. For $\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$, $h = 0.7$, these scales are

$$\begin{aligned} k_{\rm eq}^{-1} \;&=\; 65\,h^{-1}{\rm Mpc} \;=\; 93\,{\rm Mpc} \\ k_{\rm dec}^{-1} \;&=\; 145\,h^{-1}{\rm Mpc} \;=\; 207\,{\rm Mpc}\,. \end{aligned} \tag{191}$$

The smallest "cosmological" scale is that corresponding to a typical distance between galaxies, about 1 Mpc.[19] This scale entered during the radiation-dominated epoch (well after Big Bang nucleosynthesis).

The scale corresponding to the present "horizon" (i.e. Hubble length) is

$$k_0^{-1} \;=\; (\mathcal{H}_0)^{-1} \;=\; 2998\,h^{-1}{\rm Mpc} \;\sim\; 4300\,{\rm Mpc}\,. \tag{192}$$

Because of the acceleration due to dark energy, this scale is actually *exiting* now, and there are scales, somewhat larger than this, that have briefly entered, and then exited again in the recent past. The horizon entry is not to be taken as an instantaneous process, so these scales were

[17] We shall later redefine *primordial perturbation* to refer to the perturbation at the epoch when all cosmologically interesting scales were well outside the horizon, which is the standard meaning of this concept in cosmology.

[18] Although in coordinate space the relative density perturbation $\delta(\mathbf{x})$ is a dimensionless number, the Fourier quantity $\delta_{\mathbf{k}}$ is not. The size of $\delta_{\mathbf{k}}$ is characterized by the dimensionless value $\mathcal{P}(k)^{1/2}$.

[19] In the present universe, structure at smaller scales has been messed up by galaxy formation, so that it bears little relation to the primordial perturbations at these scales. However, observations of the high-redshift universe, especially so-called Lyman-α observations (absorption spectra of high-z quasars, which reveal distant gas clouds along the line of sight), can reveal these structures when they are closer to their primordial state. With such observations, the "cosmological" range of scales can be extended down to ~ 0.1 Mpc.

never really subhorizon enough for the Newtonian theory to apply to them. Thus we shall just consider scales $k^{-1} < k_0^{-1}$. The largest observable scales, of the order of k_0^{-1}, are essentially at their "primordial" amplitude now.

We shall now discuss the evolution of the perturbations at these scales ($k^{-1} < k_0^{-1}$) after horizon entry, using the Newtonian perturbation theory presented in the previous section.

8.3.2 Composition of the real universe

The present understanding is that there are five components to the energy density of the universe,

1. cold dark matter (c)
2. baryonic matter (b)
3. photons (γ)
4. neutrinos (ν)
5. dark energy (d)

(during the time of interest for this section, i.e., from some time after BBN until the present). Thus

$$\rho = \underbrace{\rho_c + \rho_b}_{\rho_m} + \underbrace{\rho_\gamma + \rho_\nu}_{\rho_r} + \rho_d \,. \tag{193}$$

(Note that ρ_c here is the CDM density, not the critical density, for which we write $\rho_{\rm cr}$.)

Baryons and photons interact with each other until $t = t_{\rm dec}$, so for $t < t_{\rm dec}$ they have to be discussed as a single component,

$$\rho_{b\gamma} = \rho_b + \rho_\gamma \,. \tag{194}$$

The other components do not interact with each other, except gravitationally, during the time of interest. The fluid description of Sec. 8.2 can only be applied to components whose particle mean free paths are shorter than the scales of interest. After decoupling, photons "free stream" and cannot be discussed as a fluid. On the other hand, the photon component becomes then rather homogeneous quite soon, so we can approximate it as a "smooth" component[20]. The same applies to neutrinos for the whole time since the BBN epoch, until the neutrinos become nonrelativistic. This will happen to at least two of the three neutrino species, and then they should be treated as matter (hot dark matter), not radiation. According to observations, the neutrino masses are small enough, not to have a major impact on structure formation. (However, for accurate work this must be taken into account and this effect on structure formation provides the tightest cosmological limits to neutrino masses.) Thus we shall here approximate neutrinos as a smooth radiation component. Dark energy is believed to be relatively smooth. If it is a cosmological constant (vacuum energy) then it is perfectly homogeneous.

The discussion in Sec. 8.2 applies to the case, where ρ can be divided into two components,

$$\rho = \rho_m + \rho_s \,, \tag{195}$$

where the perturbation is only in the matter component ρ_m and $\rho_s = \bar{\rho}_s$ is homogeneous. For perturbations in radiation components and dark energy the Newtonian treatment is not

[20] As long as we are interested in density perturbations only. When we are interested in the CMB anisotropy, the momentum distribution of these photons becomes the focus of our attention.

enough. Unfortunately, we do not have quite this two-component case here. Based on the above discussion, a reasonable approximation is given by a separation into three components:

$$t < t_{\rm dec}: \quad \rho = \rho_c + \rho_{b\gamma} + \rho_s \qquad (\rho_s = \rho_\nu + \rho_d) \tag{196}$$

$$t > t_{\rm dec}: \quad \rho = \rho_c + \rho_b + \rho_s \qquad (\rho_s = \rho_\gamma + \rho_\nu + \rho_d)\,. \tag{197}$$

After decoupling, both ρ_c and ρ_b are matter-like ($p \ll \rho$) and we'll discuss in Sec. 8.3.4 how this case is handled. Before decoupling, the situation is more difficult, since $\rho_{b\gamma}$ is not matter-like, as the pressure provided by photons is large. Here we shall be satisfied with a crude approximation for this period.

The most difficult period is that close to decoupling, where the photon mean free path λ_γ is growing rapidly. The fluid description, which we are here using for the perturbations, applies only to scales $\gg \lambda_\gamma$, whereas the photons are smooth only for scales $\ll \lambda_\gamma$. Thus this period can be treated properly only with large numerical "Boltzmann" codes, such as CMBFAST or CAMB.

8.3.3 CDM density perturbations

Cold dark matter is the dominant structure-forming component in the universe (dark energy dominates the energy density at late times, but does not form structure, or, if it does, these structures are very weak, not far from homogeneous). Observations indicate that $\rho_b \sim 0.2\rho_c$. Thus we get a first approximation to the behavior of the CDM perturbations by ignoring the baryon component and equating

$$\rho_m \approx \rho_c\,.$$

The CDM is pressureless, and thus the CDM sound speed is zero, and so is the CDM Jeans length. Thus, for CDM, all scales are larger than the Jeans scale, and we don't get an oscillatory behavior. Instead, perturbations grow at all scales. On the other hand, as we shall discuss in Sec. 8.3.4, perturbations in $\rho_{b\gamma}$ oscillate before decoupling. Therefore the perturbations in $\rho_{b\gamma}$ will be smaller than those in ρ_c, and we can make a (crude) approximation where we treat $\rho_{b\gamma}$ as a homogeneous component before decoupling. This is important, since although $\rho_b \ll \rho_c$, this is not true for $\rho_{b\gamma}$ at earlier, radiation-dominated, times. At decoupling $\rho_b < \rho_\gamma < \rho_c$. Before matter-radiation equality, there is an epoch when $\rho_c < \rho_\gamma$, but $\delta\rho_c > \delta\rho_{b\gamma}$. For simplicity, we now approximate

$$\rho = \rho_m + \rho_r + \rho_d \tag{198}$$

where $\rho_m = \rho_c$ and $\rho_r = \rho_\gamma + \rho_\nu$ is a smooth component (ρ_ν truly smooth, ρ_γ truly smooth after decoupling, and (crudely) approximated as smooth before decoupling). We have ignored baryons, since they are a subdominant part of $\rho_{b\gamma}$ before decoupling, and a subdominant matter component after decoupling. Likewise, ρ_d is also smooth, and becomes important only close to present times.

We can now study the growth of CDM perturbations even during the radiation-dominated period, as the radiation-component is taken as smooth and affects only the expansion rate. We can study it all the way from horizon entry to the present time, or until the perturbations become nonlinear ($\delta_c = \delta\rho_c/\bar\rho_c \sim 1$).

For the radiation- and matter-dominated epochs, including the transition in between, we then have the case of Sec. 8.2.11, and the matter perturbation grows as (187),

$$\delta \;=\; \delta_{\rm prim}\left(1 + \frac{3y}{2}\right) \;=\; \delta_{\rm prim}\left(1 + \frac{3}{2}\frac{a}{a_{\rm eq}}\right)\,.$$

The perturbation remains frozen to its primordial value, $\delta \approx \delta_{\rm prim}$, during the radiation-dominated period. By $t = t_{\rm eq}$, it has grown to $\delta = \frac{5}{2}\delta_{\rm prim}$. During the matter-dominated

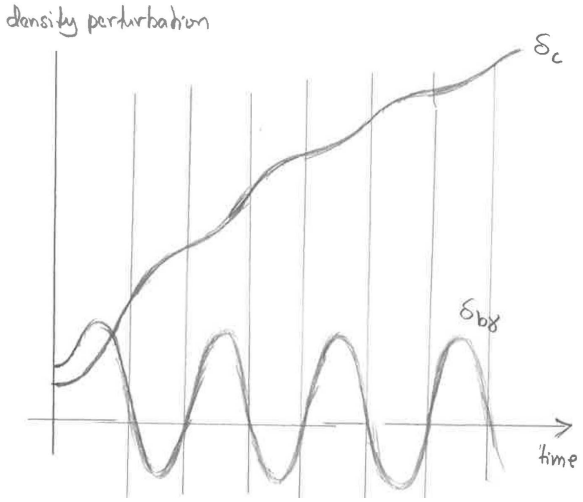


Figure 6: Growth of CDM perturbation during radiation-dominated epoch for the case of adiabatic primordial perturbations (qualitative). The time axis represents conformal time.

period, the CDM perturbation grows proportional to the scale factor,

$$\delta \propto y \propto a \propto t^{2/3} .$$

However, in the case of adiabatic primordial perturbations, the above approximation misses an important effect: an additional logarithmic growth factor $\sim \ln(k/k_{\rm eq})$ the CDM perturbations get from the gravitational effect (ignored in the above) of the oscillating radiation perturbation during the radiation-dominated epoch. To get this boost the CDM perturbation must initially be in the same direction (positive or negative) as the radiation perturbation, which is the case for adiabatic primordial perturbations.

For adiabatic primordial perturbations, the baryon, CDM, and radiation perturbations are related at horizon entry as $\delta_c = \delta_b = \frac{3}{4}\delta_\gamma$. Consider scales that enter during the radiation-dominated epoch ($t < t_{\rm eq} < t_{\rm dec}$). The gravitational effect is dominated initially by the radiation perturbation, which begins to oscillate after horizon entry; the baryon perturbation will oscillate with it until $t_{\rm dec}$. CDM, on the other hand, does not feel the radiation pressure responsible for the oscillation, it sees only the gravitational effect of the baryon-photon fluid. In the first phase of the oscillation period δ_c is of the same sign as $\delta_{b\gamma}$, so $\delta_{b\gamma}$ adds to the gravitational pull to increase δ_c. Since at first $\delta\rho_{b\gamma} > \delta\rho_c$, this additional pull is larger than that of CDM itself, leading to a much faster growth of δ_c (which otherwise would grow very little during the radiation domination). The flow of CDM is accelerated toward CDM overdensities. In the next phase of the oscillation, the sign of $\delta_{b\gamma}$ reverses, and now the pull of $\delta\rho_{b\gamma}$ on CDM is in the opposite direction, and will slow down the flow of CDM toward overdensities. But this is not enough to reverse the CDM flow before the sign of $\delta_{b\gamma}$ changes again and begins to accelerate CDM again toward CDM overdensities.

Thus the effect of the radiation oscillations is to increase δ_c stepwise, one step for each oscillation period. See Fig. 6. As the $\bar\rho_\gamma/\bar\rho_c$ ratio decreases the relative increases per step decrease; but this effect keeps adding steps until $t_{\rm dec}$. The smaller the scale (the higher the k) the more steps there are between horizon entry (t_k) and $t_{\rm dec}$, and the larger the first steps. An analytic calculation (too complicated for this course, but it is done in [9] and I do it in Cosmological Perturbation Theory) of this effect, in the small-scale limit $k \gg k_{\rm eq}$ and still ignoring baryons, so that the oscillating radiation perturbation is just photons, gives that it

leads to a boost by a factor $\sim 7.5\ln(0.17k/k_{\rm eq})$, so that (187) is modified to

$$\delta_c \;\approx\; \delta_{\rm prim}\left(1+\frac{3}{2}\frac{a}{a_{\rm eq}}\right)7.5\ln\left(0.17\frac{k}{k_{\rm eq}}\right) \qquad \text{for } k\gg 6k_{\rm eq} \text{ and } t>t_{\rm dec} \tag{199}$$

(for $k \lesssim 6k_{\rm eq}$ the logarithm is negative; this approximate result does not apply for such large scales). There is more discussion of this result in Sec. 8.4.4, where we compare this approximate analytical result to a more accurate numerical result from CAMB.

8.3.4 Baryon density perturbations

Although CDM is the dominant matter component in the universe, we cannot directly see it. The main method to observe the density perturbation today is to study the distribution of galaxies. But the part of galaxies that we can see is baryonic. Thus to compare the theory of structure formation to observations, we need to study how the perturbation in the baryonic component evolves.

Baryon Jeans length and speed of sound. We define the baryon Jeans length as $\lambda_J = 2\pi k_J^{-1}$, where

$$k_J^{-1} = \frac{c_s}{a\sqrt{4\pi G\bar\rho_b}}\,, \tag{200}$$

and c_s is the speed of sound for baryons (i.e., in the baryon-photon fluid before decoupling, and in the baryon fluid after decoupling). This definition compares the pressure felt by baryons to baryon gravity, so it addresses the question whether the baryon density perturbation can grow under its own gravity. This is not the question we face in reality, since at early times the gravity was dominated by the radiation perturbation and later by the CDM perturbation. The baryon Jeans length can still be used for order-of-magnitude estimates of at what scales the baryon perturbation can grow, and for the argument that we cannot match observations without CDM.

In general,

$$c_s^2 = \left(\frac{\partial p}{\partial\rho}\right)_\sigma\,, \tag{201}$$

where σ refers to constant entropy per baryon. Since in our case the entropy is completely dominated by photons,

$$s_{b\gamma} \sim s_\gamma = \frac{4\pi^2}{45}T^3 = \frac{2\pi^4}{45\zeta(3)}n_\gamma\,, \tag{202}$$

we have

$$\sigma \equiv \frac{s_{b\gamma}}{n_b} \sim \frac{s_\gamma}{n_b} = \frac{2\pi^4}{45\zeta(3)}\frac{n_\gamma}{n_b} \approx 3.6016\frac{1}{\eta}\,, \tag{203}$$

where η is the baryon-to-photon ratio.

We find the speed of sound by varying $\rho_{b\gamma}$ and $p_{b\gamma}$ *adiabatically*, (i.e., keeping σ, the entropy/baryon constant), which in this case means keeping η constant. Now

$$\begin{aligned}
&\rho_b = mn_b = m\eta n_\gamma = m\eta\frac{2\zeta(3)}{\pi^2}T^3 \quad\Rightarrow\quad \delta\rho_b = \bar\rho_b\cdot 3\frac{\delta T}{T}\\
&\rho_\gamma = \frac{\pi^2}{15}T^4 \quad\Rightarrow\quad \delta\rho_\gamma = \bar\rho_\gamma\cdot 4\frac{\delta T}{T}\\
&p_\gamma = \frac{\pi^2}{45}T^4 \quad\Rightarrow\quad \delta p_\gamma = \bar p_\gamma\cdot 4\frac{\delta T}{T} = \bar\rho_\gamma\cdot\frac{4}{3}\frac{\delta T}{T}\,.
\end{aligned}$$

Since $p_b \ll p_\gamma \quad\Rightarrow\quad \delta p_b \ll \delta p_\gamma$, we get

$$c_s^2 = \frac{\delta p}{\delta\rho} = \frac{\delta p_\gamma}{\delta\rho_\gamma+\delta\rho_b} = \frac{\frac{4}{3}\bar\rho_\gamma}{4\bar\rho_\gamma+3\bar\rho_b} = \frac{1}{3}\frac{1}{1+\frac{3}{4}\frac{\bar\rho_b}{\bar\rho_\gamma}}\,. \tag{204}$$

This was a *calculation* of the speed of sound, which one gets by varying the pressure and density adiabatically. It is independent of whether the actual perturbations we study are adiabatic or not.

This result, Eq. (204), applies before decoupling. As we go back in time, $\bar{\rho}_b/\bar{\rho}_\gamma \to 0$ and $c_s^2 \to 1/3$. As we approach decoupling, $\bar{\rho}_b$ becomes comparable to (but still smaller than) $\bar{\rho}_\gamma$ and the speed of sound falls, but not by a large factor.

Newtonian perturbation theory applies only to subhorizon scales. The ratio of the (comoving) baryon Jeans length

$$\lambda_J = \frac{2\pi c_s}{a\sqrt{4\pi G\bar{\rho}_b}}$$

to the comoving Hubble length

$$\mathcal{H}^{-1} = \frac{1}{a\sqrt{\frac{8\pi G}{3}\bar{\rho}}}$$

is

$$\frac{\lambda_J}{\mathcal{H}^{-1}} = \mathcal{H}\lambda_J = 2\pi\sqrt{\frac{2\bar{\rho}}{3\bar{\rho}_b}}c_s\,.$$

Thus we see that before decoupling the baryon Jeans length is comparable to the Hubble length, and thus all scales for which our present discussion applies are sub-Jeans. Therefore, if baryon perturbations are adiabatic[21], they oscillate before decoupling[22].

After decoupling, the baryon component sees just its own pressure. This component is now a gas of hydrogen and helium. This gas is monatomic for the epoch we are now interested in. Hydrogen forms molecules only later. For a non-relativistic monatomic gas,

$$c_s^2 = \frac{5T_b}{3m}\,, \tag{205}$$

where we can take $m \approx 1\,\mathrm{GeV}$, since hydrogen dominates. Down to $z \sim 100$, residual free electrons maintain enough interaction between the baryon and photon components to keep $T_b \approx T_\gamma$. After that the baryon temperature falls faster,

$$T_b \propto (1+z)^2 \qquad \text{whereas} \qquad T_\gamma \propto 1+z \tag{206}$$

(as shown in an exercise in Chapter 4). For example, at $1+z = 1000$, soon after decoupling, $T_b = 2725\,\mathrm{K} = 0.2348\,\mathrm{eV}$ and the speed of sound is $c_s = 5930\,\mathrm{m/s}$. The baryon density is $\bar{\rho}_b = \Omega_b(1+z)^3\rho_{\rm cr} = \omega_b(1+z)^3\,1.88\times10^{-26}\,\mathrm{kg/m^3}$, and we get for the Jeans length

$$\lambda_J = (1+z)\frac{\sqrt{\pi}c_s}{\sqrt{G\bar{\rho}_b}} \tag{207}$$

that soon after decoupling

$$\lambda_J(1+z=1000) \;=\; \omega_b^{-1/2}\,0.96\times10^3\,\mathrm{pc} \;=\; \eta_{10}\,0.016\,\mathrm{Mpc} \;\sim\; 0.095\,\mathrm{Mpc}\,, \tag{208}$$

where $\eta_{10} \equiv 10^{10}\eta = 274\,\omega_b$ or $\omega_b = 0.00365\,\eta_{10}$, and the last number is for $\eta_{10} \sim 6$.

We define the baryon Jeans mass

$$M_J \equiv \bar{\rho}_{b0}\frac{\pi}{6}\lambda_J^3 \tag{209}$$

as the mass of baryonic matter within a sphere whose diameter is λ_J. Note that since λ_J is defined as a comoving distance, we must use here the present (mean) baryon density $\bar{\rho}_{b0}$. At $1+z = 1000$, the baryon Jeans mass is $\omega_b^{-1/2}\,1.3\times10^5\,\mathrm{M}_\odot = \eta_{10}^{-1/2}\,2.1\times10^6\,\mathrm{M}_\odot \sim 9\times10^5\,\mathrm{M}_\odot$ for $\eta_{10} \sim 6$. This corresponds to

[21] If there is an initial baryon entropy perturbation, i.e., a perturbation in baryon density without an accompanying radiation perturbation, it will initially begin to grow in the same manner as a CDM perturbation, since the pressure perturbation provided by the photons is missing. (Such a baryon entropy perturbation corresponds to a perturbation in the baryon-photon ratio η.) But as the movement of baryons drags the photons with them, a radiation perturbation is generated, and the baryon perturbation begins to oscillate around its initial value (instead of oscillating around zero).

[22] We have not calculated this exactly, since all our calculations have been idealized, i.e., we have used perturbation theory which applies only to matter-dominated perturbations, and here we have ignored the CDM component. But this qualitative feature will hold also in the exact calculation, and this will be enough for us now.

the mass of a globular cluster and is much less than the mass of a galaxy. Thus, for our purposes, the baryonic component is pressureless after decoupling, i.e., baryon pressure can be ignored in the evolution of perturbations at cosmological scales (greater than $\sim 1\,\mathrm{Mpc}$). (The pressure cannot be ignored for smaller scale physics like the formation of individual galaxies.)

The baryon Jeans length after decoupling is $\ll$ Mpc. It would be relevant if we were interested in the process of the formation of individual galaxies, but here we are interested in the larger scales reflected in perturbations in the galaxy number density. Thus for our purposes, the baryonic component is pressureless after decoupling.

After decoupling, the evolution of the baryon density perturbation is governed by the gravitational effect of the dominant matter component, the CDM.

We now have the situation of Sec. 8.2.10, except that we have two matter components,

$$\rho = \rho_c + \rho_b + \rho_s \,, \tag{210}$$

where we approximate $\rho_s = \rho_\gamma + \rho_\nu + \rho_d$ as homogeneous. With the help of Sec. 8.2.8, the discussion is easy to generalize for the present case.

We can ignore the pressure of both ρ_b and ρ_c. Therefore their perturbation equations are

$$\ddot{\delta}_c + 2H\dot{\delta}_c = 4\pi G\bar{\rho}_m\delta \tag{211}$$

$$\ddot{\delta}_b + 2H\dot{\delta}_b = 4\pi G\bar{\rho}_m\delta \tag{212}$$

where $\bar{\rho}_m = \bar{\rho}_c + \bar{\rho}_b$ is the total background matter density and

$$\delta = \frac{\delta\rho_c + \delta\rho_b}{\bar{\rho}_c + \bar{\rho}_b} \tag{213}$$

is the total matter density perturbation.

We can now define the *baryon-CDM entropy perturbation*,

$$S_{cb} \equiv \delta_c - \delta_b \,, \tag{214}$$

which expresses how the perturbations in the two components deviate from each other. Subtracting Eq. (212) from (211) we get an equation for this entropy perturbation,

$$\ddot{S}_{cb} + 2H\dot{S}_{cb} = 0 \,. \tag{215}$$

We assume that the primordial perturbations were adiabatic, so that we had $\delta_b = \delta_c$, i.e, $S_{cb} = 0$ at horizon entry. For large scales, which enter the horizon after decoupling, an S_{cb} never develops, so the evolution of the baryon perturbations is the same as CDM perturbations.

But for scales which enter before decoupling, an S_{cb} develops because the baryon perturbation is then coupled to the photon perturbation, whereas the CDM perturbation is not. After decoupling, $\delta_b \ll \delta_c$, since δ_c has been growing, while δ_b has been oscillating. The initial condition for Eqs. (211,212,215) is then $S_{cb} \sim \delta_c$ ("initial" time here being the time of decoupling $t_{\rm dec}$). During the matter-dominated epoch, when $a \propto t^{2/3}$, so that $H = 2/3t$, the solution for S_{cb} is

$$S_{cb} = A + Bt^{-1/3} \,, \tag{216}$$

whereas for δ_c it is, neglecting the effect of baryons on it, from Eq. (126),

$$\delta_c = Ct^{2/3} + Dt^{-1} \sim Ct^{2/3} \,. \tag{217}$$

We call the first term the "growing" and the second term the "decaying" mode (although for S_{cb} the "growing" mode is actually just constant). For δ_c the growing and decaying modes have been growing and decaying since horizon entry, so we can now drop the decaying part of δ_c.

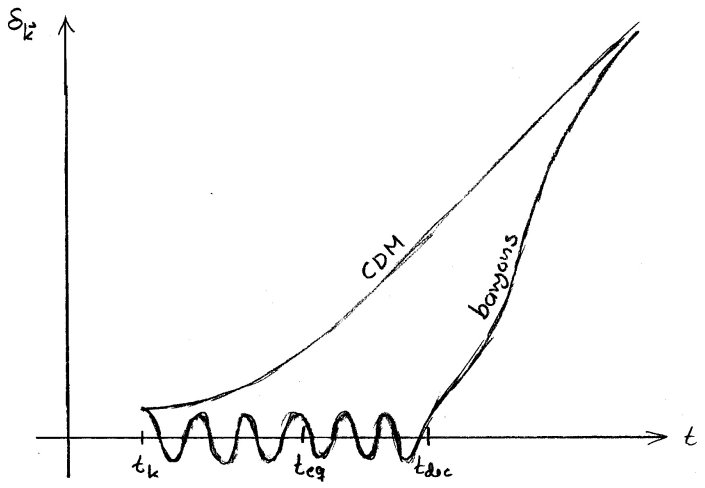


Figure 7: Evolution of the CDM and baryon density perturbations after horizon entry (at $t = t_k$). The figure is just schematic; the upper part is to be understood as having a $\sim$ logarithmic scale; the difference $\delta_c - \delta_b$ stays roughly constant, but the fractional difference becomes negligible as both δ_c and δ_b grow by a large factor.

To work out the precise initial conditions, we would need to work out the behavior of S_{cb} during decoupling. However, we really only need to assume that initially there is no strong cancellation between the growing and decaying modes in (216), so that $S_{cb} = \delta_c - \delta_b$ either shrinks or stays roughly constant near the initial value of δ_c. While δ_c grows by a large factor, δ_b must follow it to keep the difference close to the initial small value of δ_c, so that $\delta_b/\delta_c \to 1$.

Thus the baryon density contrast δ_b grows to match the CDM density contrast δ_c (see Fig. 7), and we have eventually $\delta_b = \delta_c = \delta$ to high accuracy.

The baryon density perturbation begins to grow only after $t_{\rm dec}$. Before decoupling the radiation pressure prevents it. Without CDM it would grow only as $\delta_b \propto a \propto t^{2/3}$ after decoupling (during the matter-dominated period; the growth stops when the universe becomes dark energy dominated). Thus it would have grown at most by the factor $a_0/a_{\rm dec} = 1 + z_{\rm dec} \sim 1100$ after decoupling. In the anisotropy of the CMB we observe the baryon density perturbations at $t = t_{\rm dec}$. They are too small (about 10^{-4}) for a growth factor of 1100 to give the present observed large scale structure[23].

With CDM this problem was solved. The CDM perturbations begin to grow earlier, at $t \sim t_{\rm eq}$, and by $t = t_{\rm dec}$ they are much larger than the baryon perturbations. After decoupling the baryons have lost the support from photon pressure and fall into the CDM gravitational potential wells, catching up with the CDM perturbations.

This allows the baryon perturbations to be small at $t = t_{\rm dec}$ and to grow after that by much more than the factor 10^3, matching observations. This is one of the reasons we are convinced that CDM exists.[24]

The whole subhorizon evolution history of all the different cosmological scales of perturbations is summarized by Fig. 8.

[23] This assumes adiabatic primordial perturbations, since we are seeing δ_γ, not δ_b. For a time, primordial baryon entropy perturbations $S_{b\gamma} = \delta_b - \frac{3}{4}\delta_\gamma$ were considered a possible explanation, but more accurate observations have ruled this model out.

[24] Historically, the above situation became clear in the 1980's when the upper limits to CMB anisotropy (which was finally discovered by COBE in 1992) became tighter and tighter. By today we have accurate detailed measurements of the structure of the CMB anisotropy which are compared to detailed calculations including the CDM so the argument is raised to a different level—instead of comparing just two numbers we are now comparing entire power spectra (to be discussed later).

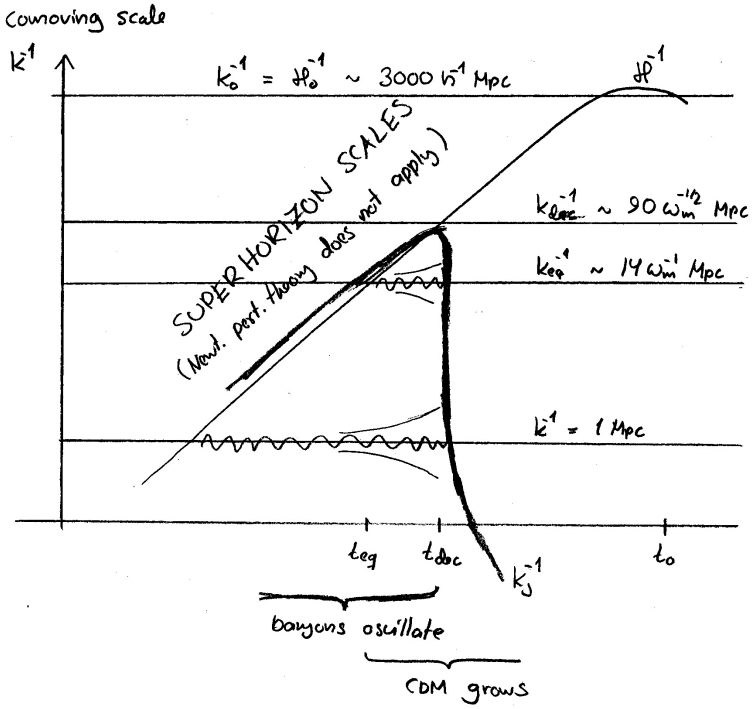


Figure 8: A figure summarizing the evolution of perturbations at different subhorizon scales. The baryon Jeans length k_J^{-1} drops precipitously at decoupling so that all cosmological scales became super-Jeans after decoupling, whereas all subhorizon scales were sub-Jeans before decoupling. The wavy lines symbolize the oscillation of baryon perturbations before decoupling, and the opening pair of lines around them symbolize the $\propto a$ growth of CDM perturbations after $t_{\rm eq}$. There is also an additional weaker (logarithmic) growth of CDM perturbations between horizon entry and $t_{\rm eq}$.

8.3.5 Late-time growth in the ΛCDM model

At late times, dark energy begins to accelerate the expansion, which will slow down the growth of the density perturbation. In the ΛCDM model dark energy is just a constant vacuum energy, so it has no perturbations and thus affects just the background. The perturbations are in CDM and baryons, and we can ignore the pressure term in the Jeans equation, since at such small scales where baryon pressure gradients would be important, first-order perturbation theory is not valid anyway at late times. Thus we are facing a similar calculation as we did in Sec. 8.2.11, the solution of Eq. (182),

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}_m\delta_{\mathbf{k}} = 0\,, \tag{218}$$

where $4\pi G\bar{\rho}_m = \frac{3}{2}\Omega_m H_0^2 a^{-3}$, with $\delta_b = \delta_c = \delta$, but instead of radiation we have now vacuum energy contributing to the background solution, which is the Concordance Model discussed in Cosmology I (Chapter 3):

$$a(t) \;=\; \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} \sinh^{2/3}\left(\tfrac{3}{2}\sqrt{\Omega_\Lambda}H_0 t\right)\,. \tag{219}$$

The Hubble parameter is given by

$$H = H_0\sqrt{\Omega_m a^{-3} + \Omega_\Lambda}\,. \tag{220}$$

Again, it is better to use the scale factor as time coordinate. The difference in the power of a in the behavior of the two density components is now 3 instead of 1, which makes the calculation more difficult. We follow here Dodelson[9]. After the change of variable from t to a,

(218) becomes (**exercise**)

$$\delta'' + \left(\frac{H'}{H} + \frac{3}{a}\right)\delta' - \frac{3\Omega_m}{2a^5}\left(\frac{H_0}{H}\right)^2 \delta = 0\,, \tag{221}$$

where $' \equiv d/da$. The decaying solution is

$$\delta \;\propto\; H \;\propto\; \sqrt{\Omega_m a^{-3} + \Omega_\Lambda} \tag{222}$$

and the growing solution is

$$\delta \;\propto\; H\int^a \frac{dx}{H^3x^3} \;\propto\; \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}\int^a \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}} \tag{223}$$

The effect of changing the lower limit of integration can be incorporated in the decaying solution; so we can set the lower limit to 0. (Equation (221) is valid in general for matter perturbations with an additional smooth background component. The first forms of the solutions (222) and (223) are valid when the smooth component is vacuum energy or negative curvature.)

In the limit $a \ll 1$, or rather, $\Omega_\Lambda \ll \Omega_m a^{-3}$, the decaying solution becomes

$$\delta \propto a^{-3/2} \propto t^{-1} \tag{224}$$

and the growing solution becomes

$$\delta \propto a \propto t^{2/3} \tag{225}$$

the familiar results for the matter-dominated universe from Sec. 8.2.6. We can ignore the decaying mode, since it has become completely negligible when the vacuum energy begins to have an effect.

To fix the proportionality coefficient in the growing mode, we write it as

$$\delta = A\left(\Omega_m a^{-3} + \Omega_\Lambda\right)^{1/2}\int_0^a \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}} \tag{226}$$

and note that in the limit $\Omega_\Lambda \ll \Omega_m a^{-3}$ it becomes

$$\delta \;\approx\; A\Omega_m^{1/2}a^{-3/2}\int_0^a x^{3/2}dx \;=\; \frac{2}{5}\Omega_m^{1/2}Aa\,. \tag{227}$$

At $a = a_0 = 1$ this would give

$$\frac{2}{5}\Omega_m^{1/2}A \equiv \tilde{\delta} \quad\Rightarrow\quad A = \frac{5}{2}\Omega_m^{-1/2}\tilde{\delta}\,, \tag{228}$$

where we have defined $\tilde{\delta}$ as the value δ would have "now"[25] if there were no vacuum energy, i.e., the universe had stayed matter dominated.

[25]Note that we defined "now" as $a = a_0 = 1$, or in more physical terms as $T = T_0 = 2.7255\,\mathrm{K}$; not as $t = t_0$. The comparison situation ($\tilde{}$) we have in mind is that the early universe (where vacuum energy has no effect) is the same as in the ΛCDM model, but there is no vacuum energy to accelerate the expansion at late times, so that by "now" the expansion rate, i.e., H_0, is smaller than we observe in reality. The present matter density $\rho_{m0} = (3/8\pi G)\Omega_m H_0^2$ is the same as in the ΛCDM model, but $\tilde{\Omega}_m = 1$, so $\tilde{H}_0 = \Omega_m^{1/2}H_0$. The age of the universe is $\tilde{t}_0 = \frac{2}{3}\tilde{H}_0^{-1} = \frac{2}{3}\Omega_m^{-1/2}H_0^{-1}$, which for $h = 0.7$ and $\Omega_m = 0.3$ gives $\tilde{t}_0 = 17.0\times10^9$ years, instead of the $t_0 = 13.5\times10^9$ years of the ΛCDM model.

Thus we write (226) as

$$\boxed{\delta \;=\; \tilde{\delta}\,\frac{5}{2}\left(a^{-3}+\frac{\Omega_\Lambda}{\Omega_m}\right)^{1/2}\int_0^a \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}}\,.} \tag{229}$$

Unfortunately, the integral in (229) does not give an elementary function. (I think it is a so-called hypergeometric function, which does not give much useful information compared to just integrating (229) numerically.) We can see that at late (future) times, when $a \gg 1$, there is very little growth, since the factor outside the integral approaches a constant and for any $a_1 \gg 1$ and $a_2 \gg 1$, the contribution to the integral,

$$\int_{a_1}^{a_2} \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}} \;\approx\; \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{3/2}\int_{a_1}^{a_2} x^{-3}\,dx \;=\; \frac{1}{2}\left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{3/2}\left(a_1^{-2}-a_2^{-2}\right) \tag{230}$$

is very small.

It turns out that the integral can be done if we extend it to the infinite future (**exercise**) : As $a \to \infty$,

$$\delta \;\to\; \delta(\infty) \;\equiv\; \tfrac{5}{2}\tilde{\delta}\left(a^{-3}+\frac{\Omega_\Lambda}{\Omega_m}\right)^{1/2}\int_0^\infty \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}} \;=\; \tfrac{5}{6}\tilde{\delta}\left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} B(\tfrac{5}{6},\tfrac{2}{3})\,, \tag{231}$$

where

$$B(p,q) \;\equiv\; \int_0^1 t^{p-1}(1-t)^{q-1}dt \;=\; \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \tag{232}$$

is the beta function and

$$B\left(\tfrac{5}{6},\tfrac{2}{3}\right) \approx 1.725\,. \tag{233}$$

Thus the perturbations "freeze", i.e., approach a final value

$$\delta(\infty) = 1.437\left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3}\tilde{\delta}\,. \tag{234}$$

which for $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ gives

$$\delta(\infty) = 1.084\,\tilde{\delta}\,, \tag{235}$$

i.e., the perturbations will never become much stronger than what they in the matter-dominated model would be already "now". To get the present density perturbation $\delta(a=1)$ one has to do (229) numerically. This is done in Fig. 9, from which one can read that $\delta(a=1) \approx 0.78\,\tilde{\delta}$.

For perturbations that entered horizon well before matter-radiation equality $t_{\rm eq}$, we have from (199) that

$$\tilde{\delta} \approx \delta_{\rm prim}\left(1+\frac{3}{2a_{\rm eq}}\right)7.5\ln\left(0.17\frac{k}{k_{\rm eq}}\right)\,, \tag{236}$$

assuming that this is still $\ll 1$ so that first-order perturbation theory remains valid.

For $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$, we have $k_{\rm eq}^{-1} = 65\,h^{-1}$Mpc and $a_{\rm eq} = 1/3514$. Equation (236) gives then for the scale $k^{-1} = 8\,h^{-1}$Mpc,

$$\tilde{\delta} \approx 13\,000\,\delta_{\rm prim} \qquad \text{and} \qquad \delta(a=1) \approx 10\,000\,\delta_{\rm prim}. \tag{237}$$

The contribution of the boost from radiation oscillation is a factor $7.5\ln(0.17k/k_{\rm eq}) \approx 2.4$ at this scale (and for smaller scales it is more). Actually, this scale is still too large (too low

k) for (236) to apply; in reality the factor is somewhat larger.[26] We chose the reference scale $k^{-1} = 8\,h^{-1}$Mpc, since observationally, the variance σ_T^2 of the top-hat-filtered density field of the galaxy distribution today is ≈ 1 at this scale. Because of the galaxy bias b_g, the corresponding variance for the matter distribution is less by a factor b_g^{-2}, but still not far from 1, meaning that the linear perturbation theory approximation is beginning to break. Because of nonlinear effects, the perturbation today is somewhat larger than our prediction from linear theory. We noted earlier that the perturbations entered the horizon with amplitude $\mathcal{P}(k)^{1/2} \approx$ few $\times\, 10^{-5}$. Thus today the amplitude at $k = 1/8\,h^{-1}$Mpc should be somewhat more than few $\times\, 10^{-1}$. From Fig. 4 we see that σ_T^2 is typically a bit more that 2 times $\mathcal{P}$ at the same scale (depending on the shape of $\mathcal{P}(k)$). So indeed we get a prediction that it should be somewhat less than 1. We will do this comparison more quantitatively later.

8.3.6 Growth function

Inside the horizon, after photon decoupling the linear growth of matter perturbations is independent of scale (once the decaying mode has died out and ignoring the subcosmological scales where pressure gradients have a role). Thus it can be described by a function that depends on time (or scale factor, or redshift) only, called the *growth function*,

$$D(a) \equiv \frac{\delta(a)}{\delta_{\rm ref}} \tag{238}$$

where $\delta(a)$ is the density perturbation ($\delta_{\mathbf{k}}$ or $\delta(\mathbf{x})$; $D(a)$ is the same function for any $\mathbf{k}$ or $\mathbf{x}$) when scale factor is a and $\delta_{\rm ref}$ is it at some reference time. The choice of reference time fixes the normalization of D. During matter domination, $D(a) \propto a$ and a common normalization is to normalize so that $D(a) = a$ during matter domination. This corresponds to setting $\delta_{\rm ref} = \tilde\delta$. So that in the ΛCDM model

$$D(a) \;=\; \frac{5}{2}\left(a^{-3} + \frac{\Omega_\Lambda}{\Omega_m}\right)^{1/2} \int_0^a \frac{x^{3/2}dx}{\left(1 + \frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}}\,, \tag{239}$$

(from the onset of matter domination).

We define the *growth rate*

$$f \equiv \frac{d\ln D}{d\ln a} = \frac{d\ln\delta}{d\ln a} = \frac{a}{\delta}\frac{d\delta}{da}\,, \tag{240}$$

which is independent of this normalization.

For the ΛCDM model of Sec. 8.3.5, we get from (229) (**exercise**)

$$f(a) \;=\; \frac{1}{1+\frac{\Omega_\Lambda}{\Omega_m}a^3}\left(\tfrac{5}{2}a\frac{\tilde\delta}{\delta} - \tfrac{3}{2}\right) \;=\; \frac{1}{1+\frac{\Omega_\Lambda}{\Omega_m}a^3}\left[\frac{a}{\left(a^{-3}+\frac{\Omega_\Lambda}{\Omega_m}\right)^{1/2}\int_0^a \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}}} - \frac{3}{2}\right]. \tag{241}$$

It turns out that a good approximation to (241) is

$$f(a) \approx \Omega_m(a)^\gamma\,, \qquad \text{where } \gamma = 0.55\,, \tag{242}$$

where γ is called the *growth index*. (We have assumed General Relativity, and the measurement of the growth index from galaxy surveys is a way of testing gravity theory.) We plot D, f, and the approximation (242) for ΛCDM in Fig. 9.

[26] One should instead use the BBKS transfer function (which still ignores effect of baryons) here, which gives a larger factor. On the other hand, baryonic effects decrease the result somewhat; but the net effect is a larger factor than 2.4. These corrections will be discussed in Sec. 8.4.4.

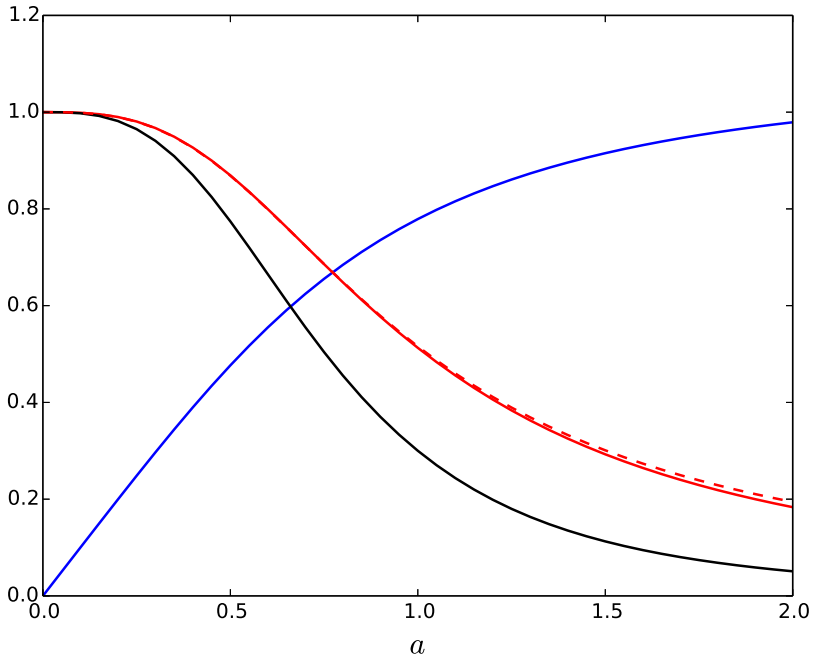


Figure 9: The growth function $D(a)$ (blue, with normalization $\delta_{\rm ref} = \tilde{\delta}$), matter density parameter $\Omega_m(a)$ (black), growth rate $f(a)$ (red), and the approximation (242) (red, dashed) for ΛCDM with $\Omega_m = 0.3$.

8.4 Relativistic perturbation theory

For scales comparable to, or larger than the Hubble scale, Newtonian perturbation theory does not apply, because we can no more ignore the curvature of spacetime. Therefore we need to use (general) relativistic perturbation theory. Instead of the Newtonian equations of gravity and fluid mechanics, the fundamental equation is now the *Einstein equation* of general relativity (GR). We assume a background solution, which is homogeneous and isotropic, i.e., a solution of the Friedmann equations, and study small perturbations around it. This particular choice of the background solution means that we are doing a particular version of relativistic perturbation theory, called *cosmological perturbation theory*.

The evolution of the perturbations while they are well outside the horizon is simple, but the mathematical machinery needed for its description is complicated. This is due to the coordinate freedom of general relativity. For the background solution we had a special coordinate system (time slicing) of choice, the one where the $t = const$ slices are homogeneous. The perturbed universe is no more homogeneous, it is just "close to homogeneous", and therefore we no more have a unique choice for the coordinate system. We should now choose a coordinate system where the universe is close to homogeneous on the time slices, but there are many different possibilities for such slicing. This freedom of choosing the coordinate system in the perturbed universe is called *gauge freedom*, and a particular choice is called a *gauge*.[27] The most important part of the choice of gauge is the choice of the time coordinate, because it determines the slicing of the spacetime into $t = const$ slices, "universe at time t". Sometimes the term 'gauge' is used to refer only to this slicing.

[27] If you are familiar with *gauge field theories*, like electrodynamics, the concept of 'gauge' may look different here. The mathematical similarity appears when the perturbation equations are developed. In relativistic perturbation theory gauge has this geometric origin (this is where the use of the word "gauge" comes from), unlike in electrodynamics.

Because the perturbations are defined in terms of the chosen coordinate system, they look different in different gauges. We can, for example, choose the gauge so that the perturbation in one scalar quantity, e.g., proper energy density, disappears, by choosing the $\rho = const$ 3-surfaces as the time slices (this is called "the uniform energy density gauge").

The true nature of gravitation is spacetime curvature, so perturbations should be described in terms of curvature.

We leave the actual development of cosmological perturbation theory to a more advanced course (Cosmological Perturbation Theory, lectured in spring 2020), and just summarize here some basic concepts and results.

In the Newtonian theory gravity was represented by a single function, the gravitational potential Φ. In GR, gravity is manifested in the geometry of spacetime, described in terms of the metric. Thus in addition to the density, pressure, and velocity perturbations, we have a perturbation in the metric. The perturbed metric tensor is

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \, . \tag{243}$$

For the background metric, $\bar{g}_{\mu\nu}$, we choose that of the flat Friedmann-Robertson-Walker universe,

$$ds^2 = \bar{g}_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a(t)^2 (dx^2 + dy^2 + dz^2) \tag{244}$$

The restriction to the flat case is an important simplification, because it allows us to Fourier expand our perturbations in terms of plane waves.[28] Fortunately the real universe appears to be flat, or at least close to it. And earlier it was even flatter. Inflation predicts a very flat universe.

For the metric perturbation, we have now 10 functions $\delta g_{\mu\nu}(t, \mathbf{x})$. So there appears to be ten degrees of freedom. Four of them are not physical degrees of freedom, since they just correspond to our freedom in choosing the four coordinates. So there are 6 real degrees of freedom.

Two of these metric degrees of freedom couple to density and pressure perturbations and the irrotational velocity perturbation. These are the *scalar perturbations*. Two couple to the rotational velocity perturbation to make up the *vector perturbations*. The remaining two are not coupled to the cosmic fluid at all[29], and are called *tensor perturbations*. They are gravitational waves, which do not exist in Newtonian theory.

The vector perturbations decay in time, and are not produced by inflation, so they are the least interesting.

Although the tensor perturbations also are not related to growth of structure, they are produced in inflation and affect the cosmic microwave background (CMB) anisotropy and polarization. Different inflation models produce tensor perturbations with different amplitudes and spectral indices (to be explained later), so they are an important diagnostic of inflation. No tensor perturbations have been detected in the CMB so far, but they could be detected in the future with more sensitive instruments if their amplitude is large enough.[30]

The three kinds of perturbations evolve independently of each other in linear perturbation theory, so they can be studied separately. We shall first concentrate on the scalar perturbations, returning to the tensor perturbations later.

[28] In the Newtonian case this restriction was not necessary, and we could apply it to any Friedmann model, as there is no curvature of spacetime in the Newtonian view, and only the expansion law $a(t)$ of the Friedmann model is used. The Newtonian theory of course is only valid for small scales were the curvature can indeed be ignored.

[29] This is true in first-order perturbation theory in the perfect fluid approximation, but not in general.

[30] Typically, large-field inflation models produce tensor perturbations with much larger amplitude than small-field inflation models. In the latter case they are likely to be too small to be detectable.

8.4.1 Gauges for scalar perturbations

Consider now scalar perturbations. The gauges discussed in the following assume scalar perturbations.

Perturbations appear different in different gauges. When needed, we use superscripts to indicate in which gauge the quantity is defined: C for the comoving gauge and N for the Newtonian gauge. Some other gauges are the synchronous gauge (S), spatially flat gauge (Q), and the uniform energy density gauge (U).

There are two common ways to specify a gauge, i.e., the choice of coordinate system in the perturbed universe:

- A statement about the relation of the coordinate system to the fluid perturbation. This will lead to some condition on the metric perturbation.
- A statement about the metric perturbation. This will then lead to some condition on the coordinate system.

The two gauges (C and N) we shall refer to in the following, give an example of each.

The *comoving gauge* is defined so that the space coordinate lines $\mathbf{x} = const$ follow fluid flow lines, and the time slice, the $t = const$ hypersurface, is orthogonal to them. Thus the velocity perturbation is zero in this gauge,

$$v^C = 0\,. \tag{245}$$

The *conformal-Newtonian* gauge, also called the longitudinal gauge, or the zero-shear gauge, and sometimes, for short, just the Newtonian gauge, is defined by requiring the metric to be of the form

$$ds^2 = -(1+2\Phi)dt^2 + a^2(1-2\Psi)(dx^2+dy^2+dz^2)\,. \tag{246}$$

This means that we require

$$\delta g_{0i} = 0, \quad \delta g_{11} = \delta g_{22} = \delta g_{33}, \quad \text{and} \quad g_{ij} = 0 \quad \text{for} \quad i \neq j\,. \tag{247}$$

(This is possible for scalar perturbations). The two metric perturbations, $\Phi(t,\mathbf{x})$ and $\Psi(t,\mathbf{x})$ are called *Bardeen potentials.*[31] Φ is also called the Newtonian potential, since in the Newtonian limit ($k \gg \mathcal{H}$ and $p \ll \rho$), it becomes equal to the Newtonian gravitational potential perturbation. Thus we can use the same symbol for it. Ψ is also called the Newtonian curvature perturbation, because it determines the curvature of the 3-dimensional $t = const$ subspaces, which are flat in the unperturbed universe (since it is the flat FRW universe).

It turns out that the difference $\Phi - \Psi$ is caused only by anisotropic stress (or anisotropic pressure). We shall here consider only the case of a perfect fluid. For a perfect fluid the pressure (or stress) is necessary isotropic. Thus we have only a single metric perturbation[32]

$$\Psi = \Phi \tag{248}$$

The density perturbations in these two gauges become equal in the limit $k \gg \mathcal{H}$ (inside horizon), and we can then identify them with the "usual" density perturbation δ of Newtonian theory.

[31] Warning: The sign conventions for Ψ differ, and many authors call them Ψ and Φ instead.

[32] In reality, neutrinos develop anisotropic pressure after neutrino decoupling. Therefore the two Bardeen potentials actually differ from each other by about 10 % between the times of neutrino decoupling and matter-radiation equality. After the universe becomes matter-dominated, the neutrinos become unimportant, and Ψ and Φ rapidly approach each other. The same happens to photons after photon decoupling, but the universe is then already matter-dominated, so they do not cause a significant $\Psi - \Phi$ difference.

8.4.2 Evolution at superhorizon scales

When the perturbations are outside the horizon (meaning that the wavelength of the Fourier mode we are considering is much longer than the Hubble length), very little happens to them, and we can find quantities which remain constant for superhorizon scales. Such a quantity is the (comoving gauge) curvature perturbation $\mathcal{R}(\mathbf{x})$, which describes how curved is the $t = const$ slice in the comoving gauge.[33] *For adiabatic perturbations, the curvature perturbation $\mathcal{R}$ stays constant in time outside the horizon.*

Using gauge transformation equations $\mathcal{R}$ can be related to the metric in the Newtonian gauge. The result is

$$\mathcal{R} = -\frac{5+3w}{3+3w}\Phi - \frac{2}{3+3w}H^{-1}\dot{\Phi}\,, \tag{251}$$

where $w \equiv \bar{p}/\bar{\rho}$.

Because $\mathcal{R}_\mathbf{k}$ stays constant while $k \ll \mathcal{H}$, it is a very useful quantity for "carrying" the perturbations from their generation at horizon exit during inflation to horizon entry at later times. We now define *the primordial perturbation* to refer to the perturbation at the epoch when it is well outside the horizon. For adiabatic perturbations, the primordial perturbation is completely characterized by the set of these constant values $\mathcal{R}_\mathbf{k}$. We shall later discuss how the primordial perturbation is generated by inflation, and how these superhorizon values $\mathcal{R}_\mathbf{k}$ are determined by it.

However, we would like to describe the perturbation in more "familiar" terms, the gravitational potential perturbation Φ and the density perturbation δ. While $\mathcal{R}_k$ remains constant this turns out to be easy. Eq. (251) can be written as a differential equation for $\Phi_\mathbf{k}$,

$$\frac{2}{3}H^{-1}\dot{\Phi}_\mathbf{k} + \frac{5+3w}{3}\Phi_\mathbf{k} = -(1+w)\mathcal{R}_\mathbf{k}\,. \tag{252}$$

During any period, when also $w = const$, the solution of this equation is

$$\Phi_\mathbf{k} = -\frac{3+3w}{5+3w}\mathcal{R}_\mathbf{k} + \text{a decaying part}\,. \tag{253}$$

Thus, after w has stayed constant for some time, the Bardeen potential has settled to the constant value

$$\boxed{\Phi_\mathbf{k} = -\frac{3+3w}{5+3w}\mathcal{R}_\mathbf{k}} \qquad (w = const)\,. \tag{254}$$

In particular, we have the relations

$$\Phi_\mathbf{k} = -\frac{2}{3}\mathcal{R}_\mathbf{k} \qquad (\text{rad.dom}, w = \tfrac{1}{3}) \tag{255}$$

$$\Phi_\mathbf{k} = -\frac{3}{5}\mathcal{R}_\mathbf{k} \qquad (\text{mat.dom}, w = 0)\,. \tag{256}$$

[33] Technically, $\mathcal{R}$ is defined in terms of the trace of the space part of the comoving gauge metric perturbation ($-\Psi$ is the corresponding quantity in the Newtonian gauge), and it is related to the scalar curvature ${}^{(3)}R^C$ of the *comoving gauge* time slice (the ${}^{(3)}$ reminds us that we are considering a 3-dimensional subspace, and the C refers to the comoving gauge) so that

$${}^{(3)}R^C = -4a^{-2}\nabla^2\mathcal{R}\,. \tag{249}$$

For Fourier components we have then that

$$\mathcal{R}_\mathbf{k} \equiv \frac{1}{4}\left(\frac{a}{k}\right)^2 {}^{(3)}R^C_\mathbf{k}\,. \tag{250}$$

Another similar quantity is the (uniform-density-gauge) curvature perturbation ζ that is defined the same way, but for the uniform-density-gauge time slice. For superhorizon scales they are equal, $\mathcal{R} = \zeta$ (in the limit $k \ll \mathcal{H}$).

8.4.3 From outside to inside horizon

After the perturbation has entered horizon, we can use the Newtonian perturbation theory result, Eq. (109), which gives the density perturbation as

$$\delta_{\mathbf{k}} = -\left(\frac{k}{a}\right)^2 \frac{\Phi_{\mathbf{k}}}{4\pi G\bar{\rho}} = -\frac{2}{3}\left(\frac{k}{aH}\right)^2 \Phi_{\mathbf{k}} = -\frac{2}{3}\left(\frac{k}{\mathcal{H}}\right)^2 \Phi_{\mathbf{k}}\,, \tag{257}$$

where we used the background relation

$$H^2 = \frac{8\pi G}{3}\bar{\rho} \quad \Rightarrow \quad \boxed{4\pi G\bar{\rho} = \frac{3}{2}H^2}\,. \tag{258}$$

The problem is to get $\Phi_{\mathbf{k}}$ from the superhorizon epoch where it is constant (as long as $w = const$) through the horizon entry to the subhorizon epoch where it evolves according to Newtonian theory. We do this for the two cases, large ($k \ll k_{\rm eq}$) and small ($k \gg k_{\rm eq}$) scales, below.

Large scales. For scales k which enter while the universe is matter dominated, this is easy, since in this case Φ_k stays constant the whole time (until dark energy becomes important). Thus we can relate these constant values of $\Phi_{\mathbf{k}}$, and the corresponding subhorizon density perturbations $\delta_{\mathbf{k}}$ during the matter-dominated epoch to the primordial perturbations $\mathcal{R}_{\mathbf{k}}$ by

$$\begin{aligned} \Phi_{\mathbf{k}} &= -\frac{3}{5}\mathcal{R}_{\mathbf{k}} \qquad \text{(mat.dom)} \\ \delta_{\mathbf{k}} &= -\frac{2}{3}\left(\frac{k}{\mathcal{H}}\right)^2 \Phi_{\mathbf{k}} = \frac{2}{5}\left(\frac{k}{\mathcal{H}}\right)^2 \mathcal{R}_{\mathbf{k}} \propto \frac{1}{(aH)^2} \propto t^{2/3} \propto a \end{aligned} \tag{259}$$

Note that by $\mathcal{R}_{\mathbf{k}}$ we refer always to the constant primordial value, when we use it in equations, like (259), that give other quantities at later times.

Small scales. For perturbations which enter during the radiation-dominated epoch, the potential $\Phi_{\mathbf{k}}$ does not stay constant. We learned earlier, that in this case the radiation density perturbation oscillates with roughly constant amplitude, which means that the amplitude for the potential Φ must decay $\propto a^2\bar{\rho} \propto a^{-2}$. This oscillation applies to the baryon-photon fluid, whereas the CDM density perturbation grows slowly. After the universe becomes matter dominated, it is these CDM perturbations that matter.

We shall now make a crude estimate how the amplitudes of these smaller-scale perturbations during the matter-dominated epoch are related to the primordial perturbations (in particular, we ignore the slow growth of the CDM perturbation during the radiation-dominated epoch). These perturbations enter during the radiation-dominated epoch. Assume that the relation $\Phi_{\mathbf{k}} = -\frac{2}{3}\mathcal{R}_{\mathbf{k}}$ holds all the way to horizon entry ($k = \mathcal{H}$). Assume then that the Newtonian relation (257) holds already. Then

$$\delta_{\mathbf{k}} \approx -\frac{2}{3}\left(\frac{k}{\mathcal{H}}\right)^2 \Phi_{\mathbf{k}} = -\frac{2}{3}\Phi_k \approx \frac{4}{9}\mathcal{R}_{\mathbf{k}} \tag{260}$$

at horizon entry. The universe is now radiation-dominated, and therefore $\delta_{r\mathbf{k}} = \delta_{\mathbf{k}}$. We are assuming primordial adiabatic perturbations and therefore the adiabatic relations $\delta_c = \frac{3}{4}\delta_r$, $\delta_\gamma = \delta_r$ hold at superhorizon scales. Assume that these relations hold until horizon entry. After that $\delta_{\gamma\mathbf{k}}$ begins to oscillate, whereas $\delta_{c\mathbf{k}}$ grows slowly. Thus we have that at horizon entry

$$\delta_{c\mathbf{k}} \approx \frac{3}{4}\delta_{\mathbf{k}} \approx \frac{1}{3}\mathcal{R}_{\mathbf{k}}\,. \tag{261}$$

Ignoring the slow growth of δ_c we get that $\delta_{c\mathbf{k}}$ stays at this value until the universe becomes matter-dominated at $t = t_{\rm eq}$, after which we can approximate $\delta_{\mathbf{k}} \approx \delta_{c\mathbf{k}}$ and $\delta_{\mathbf{k}}$ begins to grow according to the matter-dominated law, $\propto a \propto 1/\mathcal{H}^2$.

Thus

$$\delta_{\mathbf{k}}(t_{\rm eq}) \approx \frac{1}{3}\mathcal{R}_{\mathbf{k}} \tag{262}$$

and

$$\delta_{\mathbf{k}}(t) \approx \frac{1}{3}\mathcal{R}_{\mathbf{k}} \left(\frac{\mathcal{H}_{\rm eq}}{\mathcal{H}}\right)^2 = \frac{1}{3}\mathcal{R}_{\mathbf{k}} \left(\frac{k_{\rm eq}}{\mathcal{H}}\right)^2 \qquad \text{for } t > t_{\rm eq}\,, \tag{263}$$

as long as the universe stays matter dominated.

8.4.4 Transfer function

For large scales ($k \ll k_{\rm eq}$) which enter the horizon during the matter-dominated epoch, we got

$$\delta_{\mathbf{k}}(t) = \frac{2}{5}\left(\frac{k}{\mathcal{H}}\right)^2 \mathcal{R}_{\mathbf{k}} \qquad (k \ll k_{\rm eq})\,, \tag{264}$$

for as long as the universe stays matter dominated.

This is a simple result, and we use this as a reference for the more complicated result at smaller scales. That is, we define a *transfer function* $T(k,t)$ so that

$$\delta_{\mathbf{k}}(t) = \frac{2}{5}\left(\frac{k}{\mathcal{H}}\right)^2 T(k,t)\mathcal{R}_{\mathbf{k}} \tag{265}$$

where $\mathcal{R}_{\mathbf{k}}$ refers to the primordial perturbation. Thus by definition $T(k,t) = 1$ for $k \ll k_{\rm eq}$.[34]

Using the rough estimate from the previous subsection we get, comparing (263) to (259), that

$$T(k,t) \approx \frac{5}{6}\left(\frac{k_{\rm eq}}{k}\right)^2 \tag{266}$$

during the matter-dominated epoch, where we can drop the factor $\frac{5}{6}$, since this is anyway just a rough estimate.

Once we are well into the matter-dominated era, perturbations at all scales grow $\propto a \propto 1/(aH)^2$ and the transfer function becomes independent of time,[35]

$$\begin{aligned} T(k) &= 1 && k \ll k_{\rm eq} \\ T(k) &\sim \left(\frac{k_{\rm eq}}{k}\right)^2 && k \gg k_{\rm eq} \end{aligned} \tag{267}$$

A more accurate calculation, including the gravitational effect of radiation perturbation oscillations on the CDM perturbation (see Sec. 8.3.3), assuming adiabatic primordial perturbations and still ignoring baryons, adds a logarithmic growth factor, the ratio between (199) and the $\delta = \delta_{\rm prim}(a/a_{\rm eq})$ of (263), and gives[36]

$$T(k) \approx \frac{3}{2} \times 7.5 \left(\frac{k_{\rm eq}}{k}\right)^2 \ln\left(\frac{0.17k}{k_{\rm eq}}\right) \qquad k \gg k_{\rm eq}\,, \tag{269}$$

[34] With the given definition for $T(k,t)$, this holds for $t \ll t_0$, i.e, before we entered the present dark-energy-dominated epoch.

[35] We shall later define other transfer functions, but this is *the* transfer function $T(k)$ of structure formation theory. It relates the perturbations inside the horizon during the matter-dominated epoch to the primordial perturbations, and it is independent of time.

[36] This calculation is presented in Dodelson[9] (Sections 7.3 and 7.4). Dodelson (7.69) gives the result as

$$T(k) = \frac{12k_{\rm eq}^2}{k^2} \ln\left(\frac{k}{8k_{\rm eq}}\right)\,. \tag{268}$$

For some reason I get the somewhat different numerical factors in (269) when I do the same calculation.

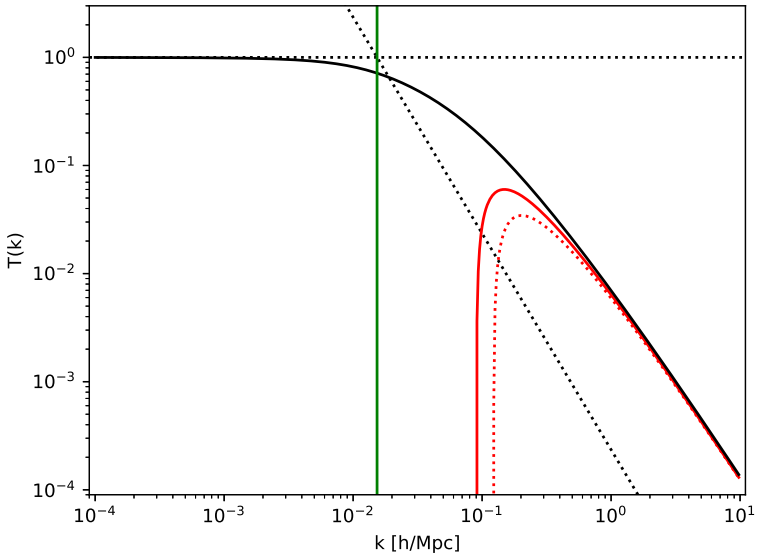


Figure 10: Transfer function $T(k)$ for CDM, adiabatic primordial fluctuations. The black curve is the BBKS transfer function (270), the red curve is the small-scale approximate analytical result (269) (the dotted red curve is the Dodelson version (268)), the two black dotted lines correspond to (267), and the green vertical line gives $k = k_{\rm eq}$. The k scale is for our reference model, $\Omega_m = 0.3$, $h = 0.7$, for which $k_{\rm eq} = 0.0153\,h/{\rm Mpc} = 1/(65\,h^{-1}{\rm Mpc})$.

where we approximated $1 + a/a_{\rm eq} \approx a/a_{\rm eq}$ (appropriate for application at late times, $a \gg a_{\rm eq}$). Note that that logarithm is negative for $k \lesssim 6k_{\rm eq}$; the equation is not supposed to apply yet for this low k.

To include the intermediate scales, which enter close matter-radiation equality, requires numerical computation. For $\omega_b \ll \omega_c$ (i.e., still essentially ignoring baryons), Bardeen, Bond, Kaiser, and Szalay [10] gave a fitting formula, the BBKS transfer function

$$T(k) \;=\; \frac{\ln(1+2.34q)}{2.34q}\,\frac{1}{[1+3.89q+(16.1q)^2+(5.64q)^3+(6.71q)^4]^{1/4}}\,, \qquad (270)$$

where $q = 0.073(k/k_{\rm eq})$, to such numerical results. See Fig. 10 for these results. The slope of the BBKS transfer function is

$$\frac{d\ln T}{d\ln q} \;=\; \frac{2.34q}{(1+2.34q)\ln(1+2.34q)} - \frac{1}{4}\frac{3.89q+2(16.1q)^2+3(5.64q)^3+4(6.71q)^4}{1+3.89q+(16.1q)^2+(5.64q)^3+(6.71q)^4} - 1\,. \qquad (271)$$

For later reference, we note that for our reference model, $\Omega_m = 0.3$, $h = 0.7$, this gives $d\ln T/d\ln q = -1.184$ at $k = 1/(8\,h^{-1}{\rm Mpc})$.

According to present understanding, the universe becomes dark energy dominated as we approach the present time. The equation-of-state parameter w begins to decrease (becomes negative) and therefore Φ begins to change again. The growth of the density perturbations is slowed down as we saw in Secs. 8.3.5 and 8.3.6. Since this affect all scales the same way, we can model this with the growth function $D(a)$, and keep the transfer function $T(k)$ unaffected.

We are still missing the effect of baryons. There are publicly available computer programs (such as CMBFAST, CAMB[37], and CLASS[38]; you give your favorite values for the cosmological

[37] `https://camb.info/`
[38] `http://class-code.net/`

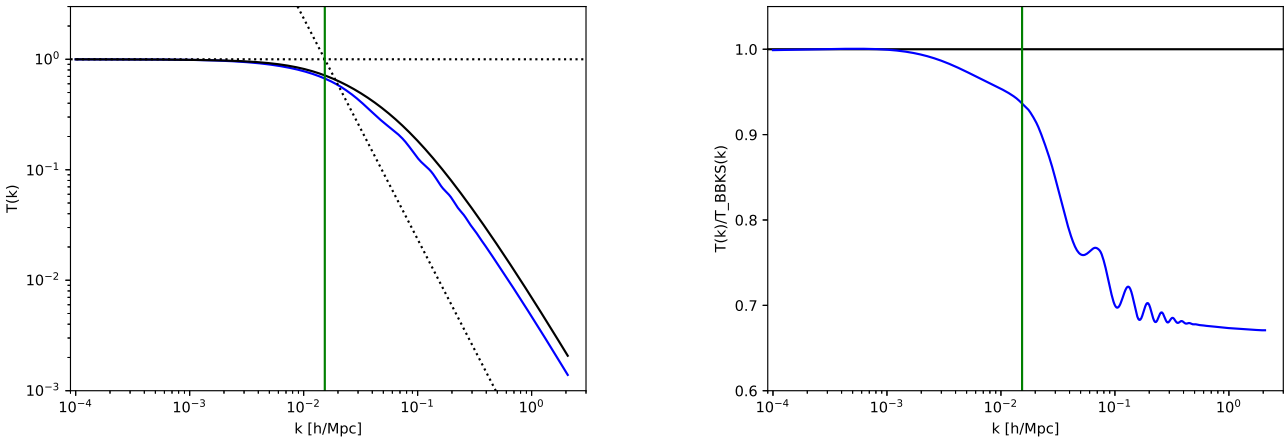


Figure 11: Left: Transfer function $T(k)$ calculated with CAMB (blue curve) for adiabatic primordial fluctuations in the flat ΛCDM model with $\omega_b = 0.023$, $\omega_c = 0.124$, $h = 0.7$ (so that $\Omega_m = 0.3$), and massless neutrinos (a neutrino mass 0.06 eV for one neutrino species changes the transfer function by less than the width of the curve). The black curve is the BBKS transfer function (270), the two black dotted lines correspond to (267), and the green vertical line gives $k = k_{\rm eq}$, as in Fig. 10. The main difference from BBKS is due to baryons. Right: The ratio (blue) of the $T(k)$ from CAMB to the BBKS transfer function.

parameters as input) that include it and other small physical effects we have ignored. They represent the current state of the art. The exact result can be given in form of the transfer function $T(k)$ we defined above. We show in Fig. 11 a transfer function calculated with CAMB. The effect of *baryon acoustic oscillations* (i.e., the oscillations of $\delta_{b\gamma}$ before decoupling, which leave a trace in δ_b) shows up as a small-amplitude wavy pattern in the $k > k_{\rm eq}$ part of the transfer function, since different modes k were at a different phase of the oscillation when that ended around $t_{\rm dec}$.

Everything has still been calculated using linear perturbation theory. Linear perturbation theory breaks down when the perturbations become large, $\delta(\mathbf{x}) \sim 1$. We say that the perturbation becomes nonlinear. This has happened for the smaller scales, $k^{-1} < 10$ Mpc by now. Nonlinear effects speed up the growth of density perturbations. They cannot be captured in a transfer function and a growth function, since now the ratio between the present-day and primordial power spectra depends on the primordial power spectrum. CAMB can also calculate nonlinear effects but within a more restricted set of cosmological models, because this is based on results from N-body simulations.

When the perturbation becomes sufficiently nonlinear, i.e., an overdense region becomes significantly (a few times) denser as the average density of the universe (see Sec. 8.5.2), it collapses and forms a gravitationally bound structure, e.g. a galaxy or a cluster of galaxies. Further collapse is prevented by the angular momentum of the structure. Galaxies in a cluster and stars (and CDM particles) in a galaxy orbit around the center of mass of the bound structure.

8.4.5 Tensor perturbations

In addition to scalar and vector perturbations, in general relativistic perturbation theory we have tensor perturbations. They have the nice property that we do not have to worry about different gauges, since they are gauge invariant in the sense that, if we first do a gauge transformation and then separate out the scalar, vector, and tensor parts, the tensor part has remained unchanged.

These are perturbations of the metric that for one Fourier mode take the form

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2\left[(1+h)dx^2 + (1-h)dy^2 + dz^2\right] \\ &= a(\eta)^2\left[-d\eta^2 + (1+h)dx^2 + (1-h)dy^2 + dz^2\right] \end{aligned} \tag{272}$$

where

$$h = h_{\mathbf{k}}(t)e^{ikz} \tag{273}$$

is the perturbation and η is conformal time. In (272) we have chosen the z axis in the direction of the wave vector, so that $\mathbf{k} = k\hat{k}$ and $\mathbf{k}\cdot\mathbf{x} = kz$. Since the metric is a real quantity, in (272) and (278) h should be interpreted as the real part of h; like one should always do when one makes physical interpretations for a single Fourier mode. Remember that when one sums over Fourier components the imaginary parts of $h_{\mathbf{k}}(t)e^{ikz} + h_{-\mathbf{k}}(t)e^{-ikz}$ cancel since $h_{-\mathbf{k}} = h^*_{\mathbf{k}}$, and thus the imaginary parts have no physical significance, they are just a mathematical convenience.

The effect of the tensor perturbation is to stretch space in one direction (here x if h is positive) and compress it in the other direction (here y) orthogonal to the wave vector of the Fourier mode. In (272) we also chose the orientation of the x and y axes so that they correspond to these stretch/compress directions. But of course the perturbation could be oriented differently. We get the other possibilities by rotating the pattern around the wave vector $\mathbf{k}$ by some angle φ, which is mathematically equivalent to rotating the coordinate system by angle $-\varphi$.

In matrix form the metric is

$$[g_{\mu\nu}] = a^2\begin{bmatrix} -1 & & & \\ & 1+h & & \\ & & 1-h & \\ & & & 1 \end{bmatrix} \tag{274}$$

After rotation by φ around the z axis it becomes

$$[g_{\mu\nu}] = a^2\begin{bmatrix} 1 & & & \\ & \cos\varphi & -\sin\varphi & \\ & \sin\varphi & \cos\varphi & \\ & & & 1 \end{bmatrix}\begin{bmatrix} -1 & & & \\ & 1+h & & \\ & & 1-h & \\ & & & 1 \end{bmatrix}\begin{bmatrix} 1 & & & \\ & \cos\varphi & \sin\varphi & \\ & -\sin\varphi & \cos\varphi & \\ & & & 1 \end{bmatrix} \tag{275}$$

Rotation by 45°, i.e., $\cos\varphi = \sin\varphi = 1/\sqrt{2}$, gives

$$[g_{\mu\nu}] = a^2\begin{bmatrix} -1 & & & \\ & 1 & h & \\ & h & 1 & \\ & & & 1 \end{bmatrix} \tag{276}$$

We call (274) the + mode and (276) the × mode. An arbitrary orientation of the stretch/compress pattern can be obtained as a linear combination of these two modes, so that the general form of the tensor perturbation is

$$[g_{\mu\nu}] = a^2\begin{bmatrix} -1 & & & \\ & 1+h_+ & h_\times & \\ & h_\times & 1-h_+ & \\ & & & 1 \end{bmatrix} \tag{277}$$

or

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2 \left[(1+h_+)dx^2 + 2h_\times dxdy + (1-h_+)dy^2 + dz^2\right] \\ &= a(\eta)^2 \left[-d\eta^2 + (1+h_+)dx^2 + 2h_\times dxdy + (1-h_+)dy^2 + dz^2\right] \end{aligned} \tag{278}$$

for a Fourier mode in the z direction. Thus we have two Fourier amplitudes $h_{+\mathbf{k}}(t)$ and $h_{\times\mathbf{k}}(t)$ for each wave vector $\mathbf{k}$. In the following we mostly write just $h(t)$ to represent an arbitrary such mode.

The evolution equation for $h(t)$,

$$\ddot{h} + 3H\dot{h} + \left(\frac{k}{a}\right)^2 h = 0 \qquad \Leftrightarrow \qquad H^{-2}\ddot{h} + 3H^{-1}\dot{h} + (k/\mathcal{H})^2\, h = 0\,, \tag{279}$$

can be obtained from the Einstein equation. This derivation is beyond the level of this course, but the equation has a simple and plausible form: it is the wave equation with a damping term $3H\dot{h}$; the wave velocity is the speed of light $= 1$.

For superhorizon scales we can ignore the last term, and we get $h = const$ as a solution and another solution where $\dot{h} \equiv dh/dt \propto a^{-3}$ so it also approaches a constant. Thus tensor perturbations remain essentially constant outside the horizon.

For evolution inside the horizon we get oscillatory solutions and then it is better to work with conformal time. The $h(\eta)$ evolution equation is

$$h'' + 2\mathcal{H}h' + k^2 h = 0 \qquad \Leftrightarrow \qquad \mathcal{H}^{-2}h'' + 2\mathcal{H}^{-1}h' + (k/\mathcal{H})^2\, h = 0\,, \tag{280}$$

where $' \equiv d/d\eta$. If we first ignore the middle term, we get solutions of the form $h \propto e^{\pm ik\eta}$, where $-$ represents a wave moving in the $\mathbf{k}$ direction and $+$ in the $-\mathbf{k}$ direction. These are *gravitational waves.* They propagate at the speed of light and they are transverse waves. During one half-period of the wave oscillation, space is stretched in one direction orthogonal to the direction of propagation, and compressed in the other orthogonal direction. During the next half-period the opposite happens. The amplitude of the stretching is given by h, meaning that the maximum stretching is by factor $1 + |h|$ and the maximum compression is by factor $1 - |h|$.

The middle term in (280) represents the damping of gravitational waves due to the expansion of the universe. Write

$$h(\eta) = A(\eta)e^{-ik\eta} \tag{281}$$

and insert this into (280) to get

$$A'' + 2\mathcal{H}A' - 2ik(A' + \mathcal{H}A) = 0\,. \tag{282}$$

For $k \gg \mathcal{H}$, the part $2ik(A' + \mathcal{H}A)$ dominates the left-hand side, and we get

$$A' + \mathcal{H}A \;=\; A' + \frac{a'}{a}A \;=\; \frac{1}{a}(aA)' \;=\; 0 \quad \Rightarrow \quad aA = const \quad \Rightarrow \quad A \propto a^{-1}\,. \tag{283}$$

Thus gravitational waves are damped inside the horizon as a^{-1} independent of the expansion law.

For simple expansion laws one can also solve Eq. (280) exactly, covering also horizon entry/exit. These solutions are Bessel functions.

8.5 Nonlinear growth

When δ grows the evolution becomes nonlinear, requiring a more complicated discussion. One can get further with higher-order perturbation theory or something called the Zeldovich approximation, but eventually one has to resort to numerical simulations. We shall not discuss these in this course. The spherically symmetric special case can be done analytically by basing it on solutions for FRW universes with different densities. We do it below for an overdensity in a flat matter-dominated background universe.

8.5.1 Closed Friedmann model

In Cosmology I we derived the expansion law for the closed ($\Omega > 1$) matter-dominated FRW universe. It cannot be given in closed form as $a(t)$, but can be given in terms of an auxiliary variable, the *development angle* ψ, as

$$\begin{aligned} a(\psi) &= a_i \frac{\Omega_i}{2(\Omega_i - 1)}(1 - \cos\psi) = a(\psi)\frac{\Omega(\psi)}{2[\Omega(\psi) - 1]}(1 - \cos\psi) \\ t(\psi) &= H_i^{-1}\frac{\Omega_i}{2(\Omega_i - 1)^{3/2}}(\psi - \sin\psi) = H(\psi)^{-1}\frac{\Omega(\psi)}{2[\Omega(\psi) - 1]^{3/2}}(\psi - \sin\psi)\,, \end{aligned} \tag{284}$$

where a_i, Ω_i, and H_i are the scale factor, density parameter, and Hubble parameter at some reference time t_i (usually chosen as the present time t_0, but below we will instead choose t_i to be some early time, when Ω is still very close to 1). In the second forms we took advantage of the fact that we can choose t_i to be any time during the development and replaced it with the "current" time. See Fig. 12 for the shape of $a(t)$. This curve is called a *cycloid*. (It is the path made by a point at the rim of a wheel.) From (284) we solve

$$\Omega(\psi) = \frac{2}{1 + \cos\psi}\,. \tag{285}$$

Calculating $da/dt = da/d\psi \times d\psi/dt$ we find (**exercise**)

$$H(\psi) = 2H_i\frac{(\Omega_i - 1)^{3/2}}{\Omega_i}\frac{\sin\psi}{(1 - \cos\psi)^2}\,. \tag{286}$$

The matter density is given by

$$\rho(\psi) = \rho_i\left(\frac{a_i}{a(\psi)}\right)^3 = 8\rho_i\frac{(\Omega_i - 1)^3}{\Omega_i^3(1 - \cos\psi)^3}\,. \tag{287}$$

The scale factor reaches a maximum $a_{\rm ta}$ (and the density a minimum) at the "turnaround" time $t_{\rm ta}$, when $\psi = \pi$, so that

$$a_{\rm ta} = a_i\frac{\Omega_i}{\Omega_i - 1}\,, \qquad t_{\rm ta} = \frac{\pi}{2}H_i^{-1}\frac{\Omega_i}{(\Omega_i - 1)^{3/2}}\,, \qquad \text{and} \qquad \rho(t_{\rm ta}) = \rho_i\frac{(\Omega_i - 1)^3}{\Omega_i^3}\,. \tag{288}$$

At this point $H = 0$ and then the universe begins to shrink. Since

$$\rho_i = \frac{3\Omega_i H_i^2}{8\pi G} \qquad \text{we have} \qquad \rho(t_{\rm ta}) = \frac{3\pi}{32Gt_{\rm ta}^2}\,. \tag{289}$$

The universe collapses at $t_{\rm coll} = 2t_{\rm ta}$, when $\psi = 2\pi$ and $a = 0$ again.

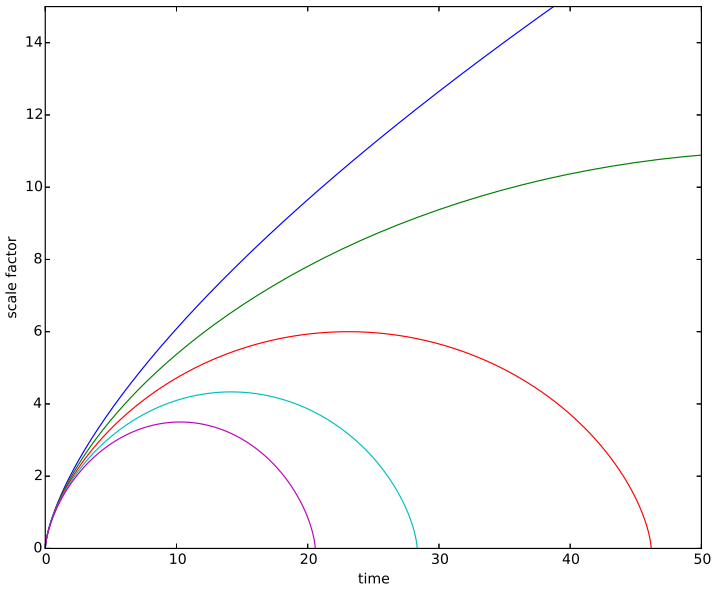


Figure 12: The expansion law for the flat matter-dominated universe (blue) and for closed matter-dominated universes with different initial values $\Omega_i > 1$ for the density parameter. Both axes are linear, the units are arbitrary.

8.5.2 Spherical collapse

The expansion law (284) will hold also for a spherically symmetric overdense region within a flat ($\Omega = 1$) matter-dominated FRW universe. Denote the quantities for this flat background universe by $\bar{a}, \bar{H}, \bar{\rho}$. (Time t is the same for both solutions and $\bar{\Omega} = 1$, so we don't need notations for them.) The background universe has

$$\bar{H}^2 = \frac{8\pi G}{3}\bar{\rho} = \left(\frac{2}{3t}\right)^2 \quad \Rightarrow \quad \bar{\rho} = \frac{1}{6\pi G t^2} \tag{290}$$

Thus we see that at $t_{\rm ta}$, the density of the overdense region is

$$\rho(t_{\rm ta}) = \frac{9\pi^2}{16}\bar{\rho}(t_{\rm ta}) \approx 5.5517\bar{\rho}(t_{\rm ta})\,, \tag{291}$$

i.e., at the turnaround time the density contrast has the value

$$\delta_{\rm ta} = \frac{9\pi^2}{16} - 1 \approx 4.5517\,. \tag{292}$$

Until then the overdense region has been expanding, although slower than the surrounding background universe. At turnaround the overdense region begins to shrink (in terms of proper distance).

The preceding applies both for an overdense region with homogeneous density and for one with a spherically symmetric density profile. In the latter case, we have to apply it separately for each spherical shell, and the density ρ refers, not to the density of the shell, but to the mean density within the shell, as it is the total mass within the shell that is responsible for the gravity affecting the expansion or contraction of the shell. To avoid shell crossing the density profile

has to decrease outward, so that outer shells do not collapse before inner shells.[39]

In linear perturbation theory, which applies when $\delta \ll 1$, density perturbations in the flat matter-dominated universe grow as

$$\delta^{\rm lin} \propto a \propto t^{2/3} \,. \tag{293}$$

When the density contrast δ becomes large it begins to grow faster. Compare now the linear growth law to the above result for δ at turnaround.

The initial density contrast δ_i is given by $\rho_i = (1+\delta_i)\bar\rho_i$. On the other hand

$$\bar H_i^2 = \frac{8\pi G}{3}\bar\rho_i \qquad \text{and} \qquad H_i^2 = \frac{8\pi G}{3}\Omega_i\rho_i \tag{294}$$

so that

$$1+\delta_i = \Omega_i \frac{H_i^2}{\bar H_i^2} \qquad \text{or at any time} \qquad 1+\delta = \Omega\frac{H^2}{\bar H^2}\,. \tag{295}$$

Thus the density contrast is not simply given by $\Omega - \bar\Omega = \Omega - 1$, since also the Hubble parameters are different for the two solutions. We can sort out the separate contributions from $\Omega_i - 1$ and $(H_i/\bar H_i)^2$ at an early time when $\Omega - 1 \ll 1$ and $\psi \ll 1$, by expanding Ω, H and $\bar H$ from (285), (286) and (290&284) in terms of ψ (**exercise**) to get

$$\Omega_i \approx 1+\tfrac{1}{4}\psi_i^2 \qquad \text{and} \qquad \frac{H_i^2}{\bar H_i^2} \approx 1-\tfrac{1}{10}\psi^2 \quad \Rightarrow \quad 1+\delta_i \approx 1+\tfrac{3}{20}\psi^2 \quad \Rightarrow \quad \delta_i \approx \tfrac{3}{5}(\Omega_i - 1)\,. \tag{296}$$

We can now give the linear prediction for the density contrast at turnaround time[40]:

$$\delta_{\rm ta}^{\rm lin} \;=\; \frac{\bar a_{\rm ta}}{\bar a_i}\delta_i \;=\; \left(\frac{t_{\rm ta}}{t_i}\right)^{2/3}\delta_i \;\approx\; \left(\frac{3\pi}{4}\right)^{2/3}\frac{\delta_i}{\Omega_i - 1} \;\approx\; \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \;\approx\; 1.0624, \tag{297}$$

where we approximated

$$t_{\rm ta} \;\approx\; \frac{\pi}{2}\bar H_i^{-1}\frac{1}{(\Omega_i - 1)^{3/2}} \qquad \text{and} \qquad t_i = \tfrac{2}{3}\bar H_1^{-1}\,. \tag{298}$$

Thus we conclude that density perturbations begin to collapse when the linear prediction is $\delta \sim 1$, at which time the true density perturbation is already over 4 times stronger.

The collapse is completed at $t_{\rm coll} = 2t_{\rm ta}$, when the linear prediction gives

$$\delta_{\rm coll}^{\rm lin} \;=\; 2^{2/3}\delta_{\rm ta}^{\rm lin} \;\approx\; 1.6865\,. \tag{299}$$

The above special case can be extended to the situation where the background universe is a closed or open Friedmann model (i.e., a matter-dominated FRW universe), and to the ΛCDM model, with more complicated math.

8.5.3 Without spherical symmetry

I suppose these idealized cases would lead to a supermassive black hole at the center of symmetry (for perturbations at cosmological scales, for a smaller scale perturbation we might end up with a star). In reality overdensities are never exactly spherically symmetric. The deviation from spherical symmetry increases as the collapse progresses. For an ellipsoidal overdensity the flattest direction collapses first leading first to a "Zeldovich pancake", and the second flattest

[39] More precisely, the density of an outer shell must not be more than the mean density inside it. We should also include in our model an underdense region around our overdense region so that their combined mean density equals that of the background universe, so as not to affect the evolution of the surroundings.

[40] Note that Kolb&Turner[11], p. 328, misses the factor 3/5 .

next leading then to an elongated structure. In the situation where the density refers to a number density of galaxies instead of a smooth continuous density, the galaxies will pass the center point at various distances (instead of colliding at the center as in the perfectly spherically symmetric case), after which they will move away from the center and will be decelerated, eventually falling back in and ending up orbiting the center, forming a cluster of galaxies.

For the real universe the different distance scales are in a different stage of the collapse. The largest distance scales are still "falling in", leading to flattened structures at the largest scales and elongated structures, "filaments", at somewhat smaller scales. These structures surround rounder underdense regions, "voids". Smaller scales have already collapsed into galaxy clusters.

8.6 Perturbations during inflation

So far we have developed perturbation theory describing the substance filling the universe in fluid terms, i.e., giving the perturbations in terms of $\delta\rho$ and δp. During inflation the universe is dominated by a scalar field, the inflaton φ, so it is better to give the perturbation directly as a perturbation in the inflaton field,

$$\varphi(t,\mathbf{x}) = \bar{\varphi}(t) + \delta\varphi(t,\mathbf{x})\,. \tag{300}$$

8.6.1 Evolution of inflaton perturbations

In Minkowski space the field equation for a scalar field is

$$\ddot{\varphi} - \nabla^2\varphi + V'(\varphi) = 0\,. \tag{301}$$

In the flat Friedmann-Robertson-Walker universe (the background universe) the field equation is

$$\ddot{\varphi} + 3H\dot{\varphi} - a^{-2}\nabla^2\varphi + V'(\varphi) = 0\,. \tag{302}$$

(Here $\nabla = \nabla_{\mathbf{x}}$, i.e., with respect to the comoving coordinates $\mathbf{x}$, and therefore the factor $1/a$ appears in front of it.)

We ignore for the moment the perturbation in the spacetime metric and just insert (300) into Eq. (302),

$$(\bar{\varphi}+\delta\varphi)^{\cdot\cdot} + 3H(\bar{\varphi}+\delta\varphi)^{\cdot} - a^{-2}\nabla^2(\bar{\varphi}+\delta\varphi) + V'(\bar{\varphi}+\delta\varphi) = 0\,. \tag{303}$$

Here $V'(\bar{\varphi}+\delta\varphi) = V'(\bar{\varphi}) + V''(\bar{\varphi})\delta\varphi$ and $\bar{\varphi}(t)$ is the homogeneous background solution from our earlier discussion of inflation. Thus $\nabla^2\bar{\varphi} = 0$, and $\bar{\varphi}$ satisfies the background equation

$$\ddot{\bar{\varphi}} + 3H\dot{\bar{\varphi}} + V'(\bar{\varphi}) = 0\,. \tag{304}$$

Subtracting the background equation from the full equation (303) we get the perturbation equation

$$\delta\ddot{\varphi} + 3H\delta\dot{\varphi} - a^{-2}\nabla^2\delta\varphi + V''(\bar{\varphi})\delta\varphi = 0 \tag{305}$$

In Fourier space we have

$$\delta\ddot{\varphi}_{\mathbf{k}} + 3H\delta\dot{\varphi}_{\mathbf{k}} + \left[\left(\frac{k}{a}\right)^2 + m^2(\bar{\varphi})\right]\delta\varphi_{\mathbf{k}} = 0\,, \tag{306}$$

or

$$H^{-2}\delta\ddot{\varphi}_{\mathbf{k}} + 3H^{-1}\delta\dot{\varphi}_{\mathbf{k}} + \left[\left(\frac{k}{aH}\right)^2 + \frac{m^2}{H^2}\right]\delta\varphi_{\mathbf{k}} = 0\,, \tag{307}$$

where

$$m^2(\bar{\varphi}) \equiv V''(\bar{\varphi})\,. \tag{308}$$

During inflation, H and m^2 change slowly. Thus we make now an approximation where we treat them as constants. If the slow-roll approximation is valid, $m^2 \ll H^2$, since

$$\frac{m^2}{H^2} = 3M_{\rm Pl}^2\frac{V''}{V} = 3\eta \ll 1\,. \tag{309}$$

Thus we can ignore the m^2/H^2 in Eq. (307)[41]. The general solution becomes then

$$\delta\varphi_{\mathbf{k}}(t) = A_{\mathbf{k}}w_k(t) + B_{\mathbf{k}}w_k^*(t)\,, \tag{312}$$

[41] The general solution to (306), when H and m^2 are constants, is

$$\delta\varphi_{\mathbf{k}}(t) = a^{-3/2}\left[A_{\mathbf{k}}J_{-\nu}\left(\frac{k}{aH}\right) + B_{\mathbf{k}}J_{\nu}\left(\frac{k}{aH}\right)\right]\,, \tag{310}$$

where

$$w_k(t) = \left(i + \frac{k}{aH}\right) \exp\left(\frac{ik}{aH}\right) . \tag{313}$$

(**Exercise:** Show that this is a solution of (306) when $H = const$ and $m^2 = 0$.) The time dependence of (312) is in

$$a = a(t) \propto e^{Ht} . \tag{314}$$

Well before horizon exit, $k \gg aH$, the argument of the exponent is large. As $a(t)$ increases the solution oscillates rapidly and its amplitude is damped. *After horizon exit*, $k \ll aH$, the solution stops oscillating and *approaches the constant value* $i(A_{\mathbf{k}} - B_{\mathbf{k}})$.

We have cheated by ignoring the metric perturbation. We should use GR and write the curved-spacetime field equation using the perturbed metric. Perturbations in a scalar field couple only to scalar perturbations, so we need to consider scalar perturbations only. For example, in the conformal-Newtonian gauge the correct perturbation equation is

$$\delta\ddot{\varphi}^N_{\mathbf{k}} + 3H\delta\dot{\varphi}^N_{\mathbf{k}} + \left[\left(\frac{k}{a}\right)^2 + V''(\bar{\varphi})\right]\delta\varphi^N_{\mathbf{k}} = -2\Phi_{\mathbf{k}} V(\bar{\varphi}) + \left(\dot{\Phi}_{\mathbf{k}} + 3\dot{\Psi}_{\mathbf{k}}\right)\dot{\bar{\varphi}} . \tag{315}$$

That is, there are additional terms which are first order in the metric and zeroth order (background) in the scalar field φ.

Fortunately, it is possible to choose the gauge so that the terms with the metric perturbations are negligible *during inflation*[42], and the previous calculation applies in such a gauge. The comoving gauge is not such a gauge, so a gauge transformation is required to obtain the comoving gauge curvature perturbation $\mathcal{R}$. Gauge transformations are beyond the scope of these lectures, but the result is

$$\mathcal{R} = -H\frac{\delta\varphi}{\dot{\bar{\varphi}}} . \tag{316}$$

Thus it is clear what we want from inflation. We want to find the inflaton perturbations $\delta\varphi_{\mathbf{k}}$ some time after horizon exit. We can use the constant value the solution (312) approaches after horizon exit. Then Eq. (316) gives us $\mathcal{R}_k$, which remains constant while the scale k is outside the horizon, and is indeed the primordial $\mathcal{R}_{\mathbf{k}}$ discussed in the previous section. And then we can use the results of Sec. 8.4 to get $\delta_{\mathbf{k}}$.

We are still missing the initial conditions for the solution (312). These are determined by quantum fluctuations, which we shall discuss in Sec. 8.6.3. Quantum fluctuations produce the initial conditions in a random manner, so that we can predict only their statistical properties. It turns out that the quantum fluctuations are a *Gaussian process*, a term which specifies certain statistical properties, which we shall discuss next before returning to the application to inflaton fluctuations.

8.6.2 Statistical properties of Gaussian perturbations

The statistical (Gaussian) nature of the inflaton perturbations $\delta\varphi(\mathbf{x})$ are inherited later by other perturbations, which depend linearly on them. Let us therefore discuss a generic Gaussian

where J_ν is the Bessel function of order ν and

$$\nu = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}} . \tag{311}$$

With $m^2 = 0$, $\nu = \frac{3}{2}$. Bessel functions of half-integer order are spherical Bessel functions which can be expressed in terms of trigonometric functions, or $e^{\pm ikx}$.

[42] One such gauge is the *spatially flat gauge* Q. For scalar perturbations it is possible to choose the time coordinate so that the time slices have Euclidean geometry. This leads to the spatially flat gauge. (There are still perturbations in the spacetime curvature; they show up when one considers the time direction).

perturbation

$$g(\mathbf{x}) = \sum_{\mathbf{k}} g_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{317}$$

where the set of Fourier coefficients $\{g_{\mathbf{k}}\}$ is a result of a *statistically homogeneous and isotropic Gaussian random process*. We assume $g(\mathbf{x})$ is real, so that $g_{-\mathbf{k}} = g^*_{\mathbf{k}}$. We write $g_{\mathbf{k}}$ in terms of its real and imaginary part,

$$g_{\mathbf{k}} = \alpha_{\mathbf{k}} + i\beta_{\mathbf{k}}\,. \tag{318}$$

For Fourier analysis of statistically homogeneous and isotropic random perturbations, see sections (8.1.1, 8.1.3, 8.1.4), where the probability distribution was treated as unknown. The new ingredient (in addition to the assumption that the perturbations are small, allowing the use of first-order perturbation theory, which we introduced in Sec. 8.2), is that the probability distribution is known to be Gaussian. This means that

$$\begin{aligned}\mathrm{Prob}(g_{\mathbf{k}}) &= \frac{1}{2\pi s^2_{\mathbf{k}}} \exp\left(-\frac{1}{2}\frac{|g_{\mathbf{k}}|^2}{s^2_{\mathbf{k}}}\right) \\ &= \frac{1}{\sqrt{2\pi} s_{\mathbf{k}}} \exp\left(-\frac{1}{2}\frac{\alpha^2_{\mathbf{k}}}{s^2_{\mathbf{k}}}\right) \times \frac{1}{\sqrt{2\pi} s_{\mathbf{k}}} \exp\left(-\frac{1}{2}\frac{\beta^2_{\mathbf{k}}}{s^2_{\mathbf{k}}}\right)\,,\end{aligned} \tag{319}$$

i.e., the real and imaginary parts are independent Gaussian random variables[43] with equal variance $s^2_{\mathbf{k}}$.

The *expectation value* of a quantity which depends on $g_{\mathbf{k}}$ as $f(g_{\mathbf{k}})$ is given by

$$\langle f(g_{\mathbf{k}})\rangle \equiv \int f(g_{\mathbf{k}})\mathrm{Prob}(g_{\mathbf{k}}) d\alpha_{\mathbf{k}} d\beta_{\mathbf{k}}\,, \tag{320}$$

where the integral is over the complex plane, i.e.,

$$\int_{-\infty}^{\infty} d\alpha_{\mathbf{k}} \int_{-\infty}^{\infty} d\beta_{\mathbf{k}}\,.$$

We immediately get (**exercise**) the *mean*

$$\langle g_{\mathbf{k}}\rangle = 0 \tag{321}$$

and *variance*

$$\langle |g_{\mathbf{k}}|^2\rangle = 2s^2_{\mathbf{k}} = \langle \alpha^2_{\mathbf{k}} + \beta^2_{\mathbf{k}}\rangle \tag{322}$$

of $g_{\mathbf{k}}$.

The distribution has one free parameter, the real positive number $s_{\mathbf{k}}$ which gives the width (determines the variance) of the distribution. From statistical isotropy and homogeneity follows that $s_{\mathbf{k}} = s(k)$ and

$$\langle g^*_{\mathbf{k}} g_{\mathbf{k}'}\rangle = 0 \qquad \text{for} \quad \mathbf{k} \neq \mathbf{k}'\,. \tag{323}$$

We can combine Eqs. (322) and (323) into a single equation,

$$\langle g^*_{\mathbf{k}} g_{\mathbf{k}'}\rangle \;=\; 2\delta_{\mathbf{k}\mathbf{k}'} s^2_{\mathbf{k}} \;=\; \delta_{\mathbf{k}\mathbf{k}'}\langle |g_{\mathbf{k}}|^2\rangle \;\equiv\; \frac{\delta_{\mathbf{k}\mathbf{k}'}}{V} P(k) \;\equiv\; \frac{2\pi^2 \delta_{\mathbf{k}\mathbf{k}'}}{Vk^3}\mathcal{P}(k), \tag{324}$$

where

$$\mathcal{P}(k) \equiv \left(\frac{L}{2\pi}\right)^3 4\pi k^3 \langle |g_{\mathbf{k}}|^2\rangle = \frac{V}{2\pi^2} k^3 \langle |g_{\mathbf{k}}|^2\rangle\,, \tag{325}$$

[43] $g_{\mathbf{k}}$ is a complex Gaussian random variable and $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$ are real Gaussian random variables.

which gives the dependence of the variance of $g_{\mathbf{k}}$ on the wave number k, is the *power spectrum* of g.

Going back to coordinate space, we find

$$\langle g(\mathbf{x})\rangle = \left\langle \sum_{\mathbf{k}} g_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \right\rangle = \sum_{\mathbf{k}} \langle g_{\mathbf{k}}\rangle e^{i\mathbf{k}\cdot\mathbf{x}} = 0 \tag{326}$$

The square of the perturbation can be written as

$$g(\mathbf{x})^2 = \sum_{\mathbf{k}} g^*_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} \sum_{\mathbf{k}'} g_{\mathbf{k}'} e^{i\mathbf{k}'\cdot\mathbf{x}} \tag{327}$$

since $g(\mathbf{x})$ is real. The typical amplitude of the perturbation is described by the variance, the expectation value of this square,

$$\begin{aligned} \langle g(\mathbf{x})^2\rangle &= \sum_{\mathbf{k}\mathbf{k}'} \langle g^*_{\mathbf{k}} g_{\mathbf{k}'}\rangle e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}} = \sum_{\mathbf{k}} \langle |g_{\mathbf{k}}|^2\rangle = 2\sum_{\mathbf{k}} s^2_{\mathbf{k}}\,. = \left(\frac{2\pi}{L}\right)^3 \sum_{\mathbf{k}} \frac{1}{4\pi k^3}\mathcal{P}(k) \\ &\to \frac{1}{4\pi}\int \frac{d^3k}{k^3}\mathcal{P}(k) = \int_0^\infty \frac{dk}{k}\mathcal{P}(k) = \int_{-\infty}^{\infty} \mathcal{P}(k) d\ln k\,. \end{aligned} \tag{328}$$

Note that there is no $\mathbf{x}$-dependence in the result, since this is an expectation value. $g(\mathbf{x})^2$ of course varies from place to place, but its expectation value from the random process is the same everywhere—the perturbed universe is statistically homogeneous. Thus the power spectrum of g gives the contribution of a logarithmic scale interval to the variance of $g(\mathbf{x})$.

An alternative definition for the power spectrum is

$$P(k) \equiv V\langle |g_{\mathbf{k}}|^2\rangle \tag{329}$$

While this definition is simpler, the result for the variance of $g(\mathbf{x})$ in terms of it and thus the interpretation is more complicated. Because of the common use of this latter definition, we shall make reference to both power spectra, and distinguish them by the different typeface. They are related by

$$P(k) = \frac{2\pi^2}{k^3}\mathcal{P}(k)\,. \tag{330}$$

For Gaussian perturbations, the power spectrum gives a complete statistical description. All statistical quantities can be calculated from it. In particular, (**exercise**)

$$\langle |g_{\mathbf{k}}|^4\rangle = 2\langle |g_{\mathbf{k}}|^2\rangle^2\,. \tag{331}$$

Exercise: Show that, if α is a real Gaussian random variable, with $\langle\alpha\rangle = 0$, then

$$\langle |\alpha|^4\rangle = 3\langle |\alpha|^2\rangle^2,$$

and that, if g is a complex Gaussian random variable (real and imaginary parts independent of each other), with $\langle g\rangle = 0$, then

$$\langle |g|^4\rangle = 2\langle |g|^2\rangle^2.$$

For a single realization, define $\hat{P}(\mathbf{k}) \equiv V|g_{\mathbf{k}}|^2$. From (331)

$$\langle \hat{P}(\mathbf{k})^2\rangle = 2P(k)^2\,. \tag{332}$$

The typical deviation of $\hat{P}(\mathbf{k})$ from its expectation value is given by the square root of the variance

$$\langle |\hat{P}(\mathbf{k}) - P(k)|^2\rangle \equiv \langle \hat{P}(\mathbf{k})^2\rangle - P(k)^2 = P(k)^2\,, \tag{333}$$

which (the last equality) only holds for Gaussian perturbations. For a single mode $\mathbf{k}$, this variance is large, but we can define

$$\hat{P}(k) \equiv \frac{V}{N_k} \sum_{\mathbf{k}} |g_{\mathbf{k}}|^2 , \tag{334}$$

where the sum is over all $\mathbf{k}$ for which $k - \frac{1}{2}\Delta k < |\mathbf{k}| \leq k + \frac{1}{2}\Delta k$, where Δk is the width of a k-bin (a shell in $\mathbf{k}$-space) over which we average, and N_k is the number of Fourier modes $\mathbf{k}$ in the bin. We then get (**exercise**)

$$\langle |\hat{P}(k) - P(k)|^2 \rangle \;=\; \frac{2}{N_k} P(k)^2 \, . \tag{335}$$

This is an example of the cosmic variance discussed in Sec. 8.1.1: the power spectrum $\hat{P}(k)$ measured from a finite volume V deviates from its expectation value $P(k)$, and thus we can measure $P(k)$ only with finite accuracy. The estimate $\hat{P}(k)$ is an average over N_k modes and its variance is reduced by the factor $N_k/2$ (since $g_{-\mathbf{k}} = g_{\mathbf{k}}^*$ we have $N_k/2$ independent modes, i.e., $N_k/2$ independent complex random variables or N_k independent real random variables.) The density of $\mathbf{k}$ modes in $\mathbf{k}$-space goes as $1/V$, so the larger the survey volume V, the more modes we have in a $\mathbf{k}$-bin (the larger os N_k). Also, for higher k, there are more modes in a k-bin: a given survey volume samples small scales better than large scales.

Exercise: Derive Eq. (335). Note that the $g_{\mathbf{k}}$ are independent variables so that $\langle |g_{\mathbf{k}}|^2 |g_{\mathbf{k}'}|^2 \rangle = \langle |g_{\mathbf{k}}|^2 \rangle \langle |g_{\mathbf{k}'}|^2 \rangle$ for $\mathbf{k}' \neq \pm\mathbf{k}$, but since $g_{-\mathbf{k}} = g_{\mathbf{k}}^*$, $\langle |g_{\mathbf{k}}|^2 |g_{\mathbf{k}'}|^2 \rangle = \langle |g_{\mathbf{k}}|^4 \rangle$.

It can be shown (under weak assumptions about the power spectrum), that statistically homogeneous and isotropic Gaussian perturbations are ergodic, so that we do not need to make a separate assumption of ergodicity.[44]

8.6.3 Generation of primordial perturbations from inflation

Subhorizon scales during inflation are microscopic[45] and therefore quantum effects are important. Thus we should study the inflaton field using quantum field theory.

This goes beyond the level of this course, so we have relegated the discussion into an appendix. The basic idea is that for scales that are inside horizon there are quantum fluctuations, called *vacuum fluctuations*, in the inflaton field. For a homogeneous inflaton field, the Fourier amplitudes $\delta\varphi_{\mathbf{k}}$ of its perturbations would be identically zero, but analogous to a quantum harmonic oscillator, it is not possible for them to stay there, but instead they fluctuate around this value.

We saw in Sec. 8.6.1 that the classical solutions to the evolution of $\delta\varphi_{\mathbf{k}}$ reach a constant value after horizon exit (in the approximation $H = const$ during horizon exit). The quantum treatment gives that at this stage we can neglect further quantum fluctuations and treat $\delta\varphi_{\mathbf{k}}$ classically—the fluctuations "freeze".

The final result is that well after horizon exit, $k \ll \mathcal{H}$, the Fourier amplitudes $\delta\varphi_{\mathbf{k}}$ have acquired a power spectrum

$$\mathcal{P}_\varphi(k) \equiv V \frac{k^3}{2\pi^2} \langle |\delta\varphi_{\mathbf{k}}|^2 \rangle = \left(\frac{H}{2\pi} \right)^2 . \tag{336}$$

[44] Liddle & Lyth [6], in Sec. 4.3.3, make this claim but do not provide a proof.

[45] We later give an upper limit to the inflation energy scale, i.e., $V(\varphi)$ at the time cosmological scales exited the horizon, $V^{1/4} < 1.9 \times 10^{16}$ GeV. From $H^2 = V(\varphi)/3M_{\rm Pl}^2$ we have $H < 9 \times 10^{13}$ GeV or for the Hubble length $H^{-1} > 2.3 \times 10^{-30}$ m. This is a lower limit to the horizon size, but it is not expected to be very many orders of magnitude larger.

After this we can ignore further quantum effects and treat the later evolution of the inflaton field, both the background and the perturbation, classically. The effect of the vacuum fluctuations was to produce "out of nothing" the perturbations $\delta\varphi_{\mathbf{k}}$. We can't predict their individual values; their production from quantum fluctuations is a random process. We can only calculate their statistical properties. Closer investigation reveals that this is a Gaussian random process. All $\delta\varphi_{\mathbf{k}}$ acquire their values as independent random variables (except for the reality condition $\delta\varphi_{-\mathbf{k}} = \delta\varphi^*_{\mathbf{k}}$) with a Gaussian probability distribution. Thus all statistical information is contained in the power spectrum $\mathcal{P}_\varphi(k)$.

The result (336) was obtained treating H as a constant. However, over long time scales, H does change. The main purpose of the preceding discussion was to follow the inflaton perturbations through the horizon exit. After the perturbation is well outside the horizon, we switch to other variables, namely the curvature perturbation $\mathcal{R}_{\mathbf{k}}$, which, unlike $\delta\varphi_{\mathbf{k}}$, remains constant outside the horizon, even though H changes. Therefore we have to use for each scale k a value of H which is representative for the evolution of that particular scale through the horizon. That is, we choose the value of H at horizon exit,[46] so that $aH = k$. Thus we write our power spectrum result as

$$\boxed{\mathcal{P}_\varphi(k) \;=\; V\frac{k^3}{2\pi^2}\langle|\delta\varphi_{\mathbf{k}}|^2\rangle \;=\; \left(\frac{H}{2\pi}\right)^2_{aH=k}} \,, \tag{337}$$

to signify that the value of H for each k is to be taken at horizon exit of that particular scale. Equation (337) is our main result from inflaton fluctuations.

8.6.4 Transfer functions

Since the inflaton fluctuations are assumed to be the origin of structure, all later perturbations are related to the inflaton perturbations $\delta\varphi_{\mathbf{k}}$. As long as all inhomogeneities are small ("perturbations"), the relationship is linear. We can express these linear relationships as *transfer functions* $T(t,k)$, e.g.,

$$g_{\mathbf{k}}(t) = T_{g\varphi}(t,k)\delta\varphi_{\mathbf{k}}(t_k)\,. \tag{338}$$

The linearity implies several things:

1. The Fourier coefficient $g_{\mathbf{k}}$ depends only on the Fourier coefficient of $\delta\varphi$ corresponding to the same wave vector $\mathbf{k}$, not on any other $\mathbf{k}'$.

2. The relationship is linear, so that if $\delta\varphi_{\mathbf{k}}$ were, e.g., twice as big, then so would $g_{\mathbf{k}}$ be.

3. The perturbations of g inherit the Gaussian statistics of $\delta\varphi$.

We could also define transfer functions relating perturbations at any two different times, t and t', and call them $T(t,t',k)$, but here we are referring to the inflaton perturbations at the time of horizon exit, t_k, which is different for different k. Actually, by $\delta\varphi_{\mathbf{k}}(t_k)$ we mean the constant value the perturbation approaches after horizon exit in the $H = const = H_k$ approximation.

That the transfer function depends only on the magnitude k results from the fact that physical laws are isotropic. The transfer function of Eq. (338) will then relate the power spectra of $\{g_{\mathbf{k}}(t)\}$ and $\{\delta\varphi_{\mathbf{k}}(t_k)\}$ as

$$\mathcal{P}_g(t,k) = T_{g\varphi}(t,k)^2\mathcal{P}_\varphi(k)\,. \tag{339}$$

The transfer functions thus incorporate all the physics that determines how structure evolves.

[46] One can do a more precise calculation, where one takes into account the evolution of $H(t)$. The result is that one gets a correction to the amplitude of $\mathcal{P}_\mathcal{R}(k)$, which is first order in slow-roll parameters, and a correction to its spectral index n which is second order in the slow-roll parameters.

For the largest scales, $k^{-1} \gg 10h^{-1}\,\mathrm{Mpc}$, the perturbations are still small today, and one needs not go beyond the transfer function. For smaller scales, corresponding to galaxies and galaxy clusters, the inhomogeneities have become large at late times, and the physics of structure growth has become nonlinear. This nonlinear evolution is typically studied using large numerical simulations. Fortunately, the relevant scales are small enough that Newtonian physics is usually sufficient.

We are now in position to put together all the results we obtained. From Eq. (316)

$$\mathcal{R}_{\mathbf{k}} = -H\frac{\delta\varphi_{\mathbf{k}}}{\dot{\bar{\varphi}}}\,, \tag{340}$$

so that

$$T_{\mathcal{R}\varphi}(k) = -\frac{H_k}{\dot{\bar{\varphi}}(t_k)} \tag{341}$$

and

$$\mathcal{P}_{\mathcal{R}}(k) = \left(\frac{H}{\dot{\bar{\varphi}}}\right)^2 \mathcal{P}_{\varphi}(k) = \left[\left(\frac{H}{\dot{\bar{\varphi}}}\right)\left(\frac{H}{2\pi}\right)\right]^2_{\mathcal{H}=k}\,, \tag{342}$$

where we used the result (337).

This primordial spectrum is the starting point for calculating structure formation (discussed already) and the CMB anisotropy (Chapter 9). Thus CMB and large-scale structure observations can be used to constrain $\mathcal{P}_{\mathcal{R}}$ together with other cosmological parameters.

8.6.5 Generation of primordial gravitational waves

The quantum fluctuations at subhorizon scales during inflation apply also to the spacetime itself. We do not yet have a complete theory of quantum gravity, so we do not know how spacetime behaves in the Planck era. At lower energy scales the spacetime fluctuations are smaller and for small perturbations around a FRW universe we can use the linearized equations for metric perturbations, for which quantization is straightforward. In fact, the proper treatment of the generation of inflaton perturbations, where we include the scalar metric perturbations in the inflaton perturbation equation (see Eq. 315), contains also the quantum treatment of scalar metric perturbations.

Likewise, we have quantum fluctuations of tensor metric perturbations during inflation. These do not couple to density perturbations, but they become classical gravitational waves after horizon exit. These *primordial gravitational waves* have an effect on CMB anisotropy and polarization.

In the quantum treatment, $(M_{\mathrm{Pl}}/\sqrt{2})h$ fluctuates like a scalar field, so that in inflation the gravitational wave amplitudes h acquire a spectrum

$$\mathcal{P}_h(k) \;\equiv\; 4\frac{V}{2\pi^2}k^3\langle|h_{\mathbf{k}}|^2\rangle \;=\; 4\frac{2}{M_{\mathrm{Pl}}^2}\left(\frac{H}{2\pi}\right)^2_{\mathcal{H}=k} \;=\; \frac{8}{M_{\mathrm{Pl}}^2}\left(\frac{H}{2\pi}\right)^2_{\mathcal{H}=k} \tag{343}$$

(the factor 4 in this customary definition is related to the way h appears in several places in the metric and to there being two modes for each $\mathbf{k}$).

The *tensor-to-scalar ratio* is the ratio of the two primordial spectra (343) and (342),

$$r \;\equiv\; \frac{\mathcal{P}_h(k)}{\mathcal{P}_{\mathcal{R}}(k)} \;=\; \frac{8}{M_{\mathrm{Pl}}^2}\left(\frac{\dot{\bar{\varphi}}}{H}\right)^2_{\mathcal{H}=k}\,. \tag{344}$$

8.7 The primordial spectrum

8.7.1 Primordial spectrum from slow-roll inflation

The final result of the previous section is thus that inflation generates primordial scalar perturbations $\mathcal{R}_{\mathbf{k}}$ with the power spectrum

$$\mathcal{P}_{\mathcal{R}}(k) \;=\; \left[\left(\frac{H}{\dot\varphi}\right)\left(\frac{H}{2\pi}\right)\right]^2_{\mathcal{H}=ak} \;=\; \frac{1}{4\pi^2}\left(\frac{H^2}{\dot\varphi}\right)^2_{t=t_k} . \tag{345}$$

and primordial tensor perturbations with the power spectrum

$$\mathcal{P}_h(k) \;=\; \frac{8}{M_{\rm Pl}^2}\left(\frac{H}{2\pi}\right)^2_{t=t_k} . \tag{346}$$

In this section φ and $\dot\varphi$ refer to the background values.

Applying the slow-roll equations

$$H^2 = \frac{V}{3M_{\rm Pl}^2} \quad \text{and} \quad 3H\dot\varphi = -V' \quad \Rightarrow \quad \frac{\dot\varphi}{H} = -M_{\rm Pl}^2\frac{V'}{V}$$

these become

$$\begin{aligned} \mathcal{P}_{\mathcal{R}}(k) \;&=\; \frac{1}{12\pi^2}\frac{1}{M_{\rm Pl}^6}\frac{V^3}{V'^2} \;=\; \frac{1}{24\pi^2}\frac{1}{M_{\rm Pl}^4}\frac{V}{\varepsilon} \\ \mathcal{P}_h(k) \;&=\; \frac{2}{3\pi^2}\frac{V}{M_{\rm Pl}^4}\,, \end{aligned} \tag{347}$$

where ε is the slow-roll parameter. The tensor-to-scalar ratio is thus

$$r \;\equiv\; \frac{\mathcal{P}_h(k)}{\mathcal{P}_{\mathcal{R}}(k)} \;=\; 16\varepsilon\,. \tag{348}$$

According to present observational CMB and large-scale structure data, the amplitude of the primordial power spectrum is about

$$\mathcal{P}_{\mathcal{R}}(k)^{1/2} \approx 5\times 10^{-5} \tag{349}$$

at cosmological scales. This gives a constraint on inflation

$$\left(\frac{V}{\varepsilon}\right)^{1/4} \;\approx\; 24^{1/4}\sqrt{\pi}\sqrt{5\times 10^{-5}}M_{\rm Pl} \;\approx\; 0.028M_{\rm Pl} \;=\; 6.8\times 10^{16}\,\text{GeV}\,. \tag{350}$$

The best chance of detecting primordial gravitational waves is based on their effect on CMB. They have not been observed so far and the present upper limit is about [8]

$$r < 0.07 \quad \Rightarrow \quad \mathcal{P}_h(k)^{1/2} < 1.3\times 10^{-5} \quad \text{and} \quad \varepsilon < 0.004\,. \tag{351}$$

This implies an upper limit to the inflation energy scale[47]

$$V^{1/4} \;\approx\; \varepsilon^{1/4}0.028M_{\rm Pl} \;<\; 0.007M_{\rm Pl} \;=\; 1.7\times 10^{16}\,\text{GeV}\,. \tag{352}$$

[47] For the epoch when perturbations at observable cosmological scales were generated. During earlier phases of inflation the energy scale was higher than in that epoch.

Since during inflation, V and V' change slowly while a wide range of scales k exit the horizon, $\mathcal{P}_\mathcal{R}(k)$ and $\mathcal{P}_h(k)$ should be slowly varying functions of k. We define the *spectral indices* n_s and n_t of the primordial spectra as

$$\begin{aligned} n_s(k) - 1 &\equiv \frac{d\ln \mathcal{P}_\mathcal{R}}{d\ln k} \\ n_t(k) &\equiv \frac{d\ln \mathcal{P}_h}{d\ln k}\,. \end{aligned} \tag{353}$$

(The -1 is in the definition of n_s for historical reasons, to match with the definition in terms of density perturbations, see Sec. 8.7.2.) If the spectral index is independent of k, we say that the spectrum is *scale free*. In this case the primordial spectra have the *power-law* form

$$\mathcal{P}_\mathcal{R}(k) = A_s^2 \left(\frac{k}{k_p}\right)^{n_s-1} \qquad \text{and} \qquad \mathcal{P}_h(k) = A_t^2 \left(\frac{k}{k_p}\right)^{n_t}\,, \tag{354}$$

where k_p is some chosen reference scale, "pivot scale", and A_s and A_t are the amplitudes at this pivot scale.

If the power spectrum is constant,

$$\mathcal{P} = \text{const.}\,, \tag{355}$$

corresponding to $n_s = 1$ and $n_t = 0$, we say that the spectrum is *scale invariant*. A scale-invariant scalar spectrum is also called the *Harrison–Zeldovich* spectrum.

If $n_s \neq 1$ or $n_t \neq 0$, the spectrum is called *tilted*. A tilted spectrum is called *red*, if $n_s < 1$ or $n_t < 0$ (more structure at large scales), and *blue* if $n_s > 1$ or $n_t > 0$ (more structure at small scales).

Using Eqs. (347) and (353) we can calculate the spectral index for slow-roll inflation.

Since $\mathcal{P}(k)$ is evaluated from Eqs. (345) and (346) or (347) when $k = aH$,

$$\frac{d\ln k}{dt} = \frac{d\ln(aH)}{dt} = \frac{\dot a}{a} + \frac{\dot H}{H} = (1-\varepsilon)H \approx H\,,$$

where we used $\dot H = -\varepsilon H^2$ (in the slow-roll approximation) in the last step. Thus

$$\frac{d}{d\ln k} = \frac{1}{1-\varepsilon}\frac{1}{H}\frac{d}{dt} = \frac{1}{1-\varepsilon}\frac{\dot\varphi}{H}\frac{d}{d\varphi} = -\frac{M_\mathrm{Pl}^2}{1-\varepsilon}\frac{V'}{V}\frac{d}{d\varphi} \approx -M_\mathrm{Pl}^2\frac{V'}{V}\frac{d}{d\varphi}\,. \tag{356}$$

Let us first calculate the scale dependence of the slow-roll parameters:

$$\frac{d\varepsilon}{d\ln k} = -M_\mathrm{Pl}^2\frac{V'}{V}\frac{d}{d\varphi}\left[\frac{M_\mathrm{Pl}^2}{2}\left(\frac{V'}{V}\right)^2\right] = M_\mathrm{Pl}^4\left[\left(\frac{V'}{V}\right)^4 - \left(\frac{V'}{V}\right)^2\frac{V''}{V}\right] = 4\varepsilon^2 - 2\varepsilon\eta \tag{357}$$

and, in a similar manner (**exercise**),

$$\frac{d\eta}{d\ln k} = \ldots = 2\varepsilon\eta - \xi\,, \tag{358}$$

where we have defined a third slow-roll parameter

$$\xi \equiv M_\mathrm{Pl}^4\frac{V'}{V^2}V'''\,. \tag{359}$$

The parameter ξ is typically second-order small in the sense that $\sqrt{|\xi|}$ is of the same order of magnitude as ε and η. (Therefore it is sometimes written as ξ^2, although nothing forces it to be positive.)

We are now ready to calculate the spectral indices:

$$\begin{aligned} n_s - 1 &= \frac{1}{\mathcal{P}_\mathcal{R}}\frac{d\mathcal{P}_\mathcal{R}}{d\ln k} = \frac{\varepsilon}{V}\frac{d}{d\ln k}\left(\frac{V}{\varepsilon}\right) = \frac{1}{V}\frac{dV}{d\ln k} - \frac{1}{\varepsilon}\frac{d\varepsilon}{d\ln k} \\ &= -M_{\rm Pl}^2\frac{V'}{V}\cdot\frac{1}{V}\frac{dV}{d\varphi} - 4\varepsilon + 2\eta = -6\varepsilon + 2\eta \\ n_t &= \frac{1}{\mathcal{P}_h}\frac{d\mathcal{P}_h}{d\ln k} = -M_{\rm Pl}^2\frac{V'}{V}\frac{1}{V}\frac{dV}{d\varphi} = -M_{\rm Pl}^2\left(\frac{V'}{V}\right)^2 = -2\varepsilon\,. \end{aligned} \tag{360}$$

Since $\varepsilon > 0$, the tensor spectrum is necessarily red. (This follows already from (346), since H is decreasing, or from (347) since V is decreasing.) Slow-roll requires $\varepsilon \ll 1$ and $|\eta| \ll 1$, so *both spectra are close to scale invariant*. For scalar perturbations this is verified by observation. Based on CMB anisotropy data from the Planck satellite, the Planck Collaboration [8] finds

$$n_s = 0.965 \pm 0.004\,. \tag{361}$$

If one were able to measure all three values n_s, r, and n_t from observations, one could solve from them the slow-roll parameters ε and η and moreover, check the *consistency condition*

$$n_t = -\frac{r}{8} \tag{362}$$

for single-field slow-roll inflation. This consistency condition is the only truly quantitative prediction of the inflation scenario (as opposed to some specific inflation model) – all the other predictions (Ω_k very small, n_s close to 1 and n_t close to 0, primordial perturbations Gaussian) are of qualitative nature, not a specific number not equal to 0 or 1.

Unfortunately, the existing upper limit to r already means that it will be difficult to ever determine the spectral index n_t with sufficient accuracy to distinguish between $n_t = -r/8$ and $n_t = 0$. The most sensitive probe to primordial gravitational waves is provided by polarization of CMB on which they will imprint a characteristic pattern (discussed briefly in the next chapter). The theoretical limit to detection is $r \sim 10^{-4}$ and there are proposals[48] for future CMB satellite missions that could reach $r \sim 10^{-3}$. If r is significantly larger than these detection limits, after detection one could still measure n_t accurately enough to distinguish, say, $n_t \approx -1$, $n_t \approx 0$ (which includes the case $n_t = -r/8$), and $n_t \approx 1$ from each other. There have been other proposals (other than inflation) for very-early-universe physics, which predict primordial tensor perturbations that deviate from scale invariance this much or more.

The Japanese space agency ISAS/JAXA selected in May 2019 the 3-year LiteBIRD mission[49] to be launched in the late 2020s. LiteBIRD is expected to measure r with accuracy $\Delta r \sim 10^{-3}$ (1 σ). A significant American and European participation in LiteBIRD is expected.

Detection of primordial gravitational waves, i.e., measurement of r, would be enough to determine ε and η and thus the inflation energy scale from Eq. (350).

One can also calculate the scale-dependence of the spectral index (**exercise**):

$$\frac{dn_s}{d\ln k} = 16\varepsilon\eta - 24\varepsilon^2 - 2\xi\,. \tag{363}$$

It is second order in slow-roll parameters, so it's expected to be even smaller than the deviation from scale invariance, $n_s - 1$. Planck Collaboration finds it consistent with zero to accuracy $\mathcal{O}(10^{-2})$, as expected.

Cosmologically observable scales have a range of about $\Delta\ln k \sim 10$. Planck measured the CMB anisotropy over a range $\Delta\ln k \sim 6$ (missing the shortest scales, where the CMB is expected

[48] See, e.g., `http://www.core-mission.org/`

[49] `http://litebird.jp/eng/`

to have negligible anisotropy). Some inflation models have $|n_s - 1|$, r, and $|dn_s/d\ln k|$ larger than the Planck results, while others do not. These observations already ruled out many inflation models.

Example: Consider the simple inflation model

$$V(\varphi) = \frac{1}{2}m^2\varphi^2 \,.$$

In Chapter 7 we already calculated the slow-roll parameters for this model:

$$\varepsilon = \eta = 2\frac{M_{\rm Pl}^2}{\varphi^2}$$

and we immediately see that $\xi = 0$. Thus

$$\begin{aligned} n_s &= 1 - 6\varepsilon + 2\eta \;=\; 1 - 8\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \\ \frac{dn_s}{d\ln k} &= 16\varepsilon\eta - 24\varepsilon^2 - 2\xi \;=\; -32\left(\frac{M_{\rm Pl}}{\varphi}\right)^4 \\ r &= 16\varepsilon \;=\; 32\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \\ n_t &= -2\varepsilon \;=\; -4\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 \end{aligned}$$

To get the numbers out, we need the values of φ when the relevant cosmological scales exited the horizon. The number of inflation e-foldings after that should be about $N \sim 50$. We have

$$N(\varphi) \;=\; \frac{1}{M_{\rm Pl}^2}\int_{\varphi_{\rm end}}^{\varphi}\frac{V}{V'}d\varphi \;=\; \frac{1}{M_{\rm Pl}^2}\int\frac{\varphi}{2} \;=\; \frac{1}{4M_{\rm Pl}^2}\left(\varphi^2 - {\varphi_{\rm end}}^2\right) ,$$

and we estimate $\varphi_{\rm end}$ from $\varepsilon(\varphi_{\rm end}) = 2M_{\rm Pl}^2/{\varphi_{\rm end}}^2 = 1 \quad \Rightarrow \quad \varphi_{\rm end} = \sqrt{2}M_{\rm Pl}$ to get

$$\varphi^2 = {\varphi_{\rm end}}^2 + 4M_{\rm Pl}^2 N = 2M_{\rm Pl}^2 + 4M_{\rm Pl}^2 N \approx 4M_{\rm Pl}^2 N \,.$$

Thus

$$\left(\frac{M_{\rm Pl}}{\varphi}\right)^2 = \frac{1}{4N}$$

and

$$\begin{aligned} n_s &= 1 - \frac{2}{N} \;\approx\; 0.96 \\ \frac{dn_s}{d\ln k} &= -\frac{2}{N^2} \;\approx\; -0.0008 \\ r &= \frac{8}{N} \;\approx\; 0.16 \\ n_t &= -\frac{1}{N} \;\approx\; -0.02 \end{aligned}$$

We see that this model is ruled out by the observed upper limit $r < 0.1$.[50]

---

[50] There was enormous excitement in early 2014, when the BICEP2 collaboration[12] claimed to have detected the effect of primordial gravitational waves with $r = 0.20^{+0.07}_{-0.05}$ in CMB polarization, consistent with this inflation model. However, it turned out that their data was contaminated by polarized emission from dust in our own galaxy.[13]

8.7.2 Scale invariance of the primordial power spectrum

Inflation predicts and observations give evidence for an almost scale invariant primordial power spectrum. Let us forget the "almost" for a moment and discuss what it means for the primordial spectrum to be scale invariant.

The primordial spectrum is something we have at superhorizon scales, where we have discussed it in terms of the comoving curvature perturbation $\mathcal{R}$, and we are calling it scale invariant, when

$$\mathcal{P}_{\mathcal{R}}(k) = A_s^2 = \text{const.} \tag{364}$$

We would like the spectrum in terms of more familiar concepts like the density perturbation, but at superhorizon scales the density perturbation is gauge dependent.

For small scales the perturbation spectrum gets modified when the scales enter the horizon, but for large scales $k \ll k_{\rm eq}$ the spectrum maintains its primordial shape, at least as long as the universe stays matter dominated. This allows the discussion of the primordial spectrum at subhorizon scales, where we can talk about the density perturbations without specifying a gauge.

From Eq. (259), the gravitational potential and density perturbation are related to the curvature perturbation as

$$\begin{aligned} \Phi_{\mathbf{k}} &= -\frac{3}{5}\mathcal{R}_{\mathbf{k}} \qquad \text{(mat.dom)} \\ \delta_{\mathbf{k}} &= -\frac{2}{3}\left(\frac{k}{\mathcal{H}}\right)^2 \Phi_{\mathbf{k}} \;=\; \frac{2}{5}\left(\frac{k}{\mathcal{H}}\right)^2 \mathcal{R}_{\mathbf{k}}\,, \end{aligned} \tag{365}$$

giving

$$\mathcal{P}_\Phi(k) = \frac{9}{25}\mathcal{P}_{\mathcal{R}}(k) = \frac{9}{25}A_s^2 = const \tag{366}$$

$$\begin{aligned} \mathcal{P}_\delta(t,k) &= \frac{4}{9}\left(\frac{k}{\mathcal{H}}\right)^4 \mathcal{P}_\Phi(k) = \frac{4}{25}\left(\frac{k}{\mathcal{H}}\right)^4 \mathcal{P}_{\mathcal{R}}(k) \\ &= \frac{4}{25}\left(\frac{k}{\mathcal{H}}\right)^4 A_s^2 \propto t^{4/3}k^4 \end{aligned} \tag{367}$$

Thus perturbations in the gravitational potential are scale invariant, but perturbations in density are not. Instead the density perturbation spectrum is steeply rising, meaning that there is much more structure at small scales than at large scales. Thus the scale invariance refers to the gravitational aspect of perturbations, which in the Newtonian treatment is described by the gravitational potential, and in the GR treatment by spacetime curvature.

The relation between density and gravitational potential perturbations reflects the nature of gravity: A 1% overdense region 100 Mpc across generates a much deeper potential well than a 1% overdense region 10 Mpc across, since the former has 1000 times more mass. Therefore we need much stronger density perturbations at smaller scales to have an equal contribution to Φ.

However, if we extrapolate Eq. (367) back to horizon entry, $k = \mathcal{H}$, we get

$$\delta_H^2(k) \equiv \text{“}\mathcal{P}_\delta(k,t_k)\text{”} \equiv \frac{4}{25}\mathcal{P}_{\mathcal{R}}(k) = \left(\frac{2}{5}A_s\right)^2 = const \tag{368}$$

Thus for scale-invariant primordial perturbations, *density perturbations of all scales enter the horizon with the same amplitude*, $\delta_H = (2/5)A_s \sim 2\times 10^{-5}$. Since the density perturbation at the horizon entry is actually a gauge-dependent quantity, and our extension of the above Newtonian relation up to the horizon scale is not really allowed, this statement should be taken just qualitatively (hence the quotation marks around the $\mathcal{P}_\delta$). As such, it applies also to the

smaller scales which enter during the radiation-dominated epoch, since the perturbations only begin to evolve after horizon entry.

What is the deep reason that inflation generates (almost) scale invariant perturbations? During inflation the universe is almost a de Sitter universe, which has the metric

$$ds^2 = -dt^2 + e^{2Ht}(dx^2 + dy^2 + dz^2)$$

with $H = const.$ In GR we learn that it is an example of a "maximally symmetric spacetime". In addition to being homogeneous (in the space directions), it also looks the same at all times. Therefore, as different scales exit at different times they all obtain the same kind of perturbations.

In terms of the other definition of the power spectrum, $P(k) \equiv (2\pi^2/k^3)\mathcal{P}(k)$, the relations (367) for scale-invariant perturbations give

$$\begin{aligned} P_{\mathcal{R}}(k) &\propto k^{-3}\mathcal{P}_R \propto k^{-3} \\ P_\delta(k) &\propto k^{-3}\mathcal{P}_\delta \propto k^4 P_{\mathcal{R}} \propto k\mathcal{P}_R \propto k \end{aligned} \tag{369}$$

For $\mathcal{P}_{\mathcal{R}}(k) \propto k^{n-1}$ we have $P_\delta(k) \propto k^n$. This is the reason for the -1 in the definition of the spectral index in terms of $\mathcal{P}_{\mathcal{R}}$—it was originally defined in terms of P_δ.

8.8 The power spectrum today

8.8.1 Density perturbations

From Eq. (265), the density perturbation spectrum at late times is

$$\mathcal{P}_\delta(k) = \frac{4}{25}\left(\frac{k}{\mathcal{H}}\right)^4 \left[\frac{D(a)}{a}\right]^2 T(k,t)^2 \mathcal{P}_{\mathcal{R}}(k) \tag{370}$$

where, from Eq. (267)

$$\begin{aligned} T(k) &= 1 && \text{for} \quad k \ll k_{\rm eq} \\ T(k) &\sim \left(\frac{k_{\rm eq}}{k}\right)^2 \ln k && \text{for} \quad k \gg k_{\rm eq}\,. \end{aligned}$$

(For the more precise form of $T(k)$ see Sec. 8.4.4, Eq. (270) and Fig. 11.) Thus the present-day density power spectrum rises steeply $\propto k^4$ at large scales, but turns at $\sim k_{\rm eq}$ to become less steep (growing $\sim \ln k$) at small scales. This is because the growth of density perturbations was inhibited while the perturbations were inside the horizon during the radiation-dominated epoch. The $\sim \ln k$ factor comes from the slow growth of CDM perturbations during this time.

Thus the structure in the universe appears stronger at smaller scales (higher k), until $k_{\rm eq}^{-1} \sim 100\,$Mpc. The $\sim 100\,$Mpc scale is indeed quite prominent in large scale structure surveys, like the 2dFGRS and SDSS galaxy distribution surveys. Towards smaller scales the structure keeps getting stronger, but now more slowly. However, perturbations are now so large that first-order perturbation theory begins to fail, and that limit is crossed at around $k^{-1} \sim k_{\rm nl}^{-1} \sim 10\,$Mpc. Nonlinear effects cause the density power spectrum to rise more steeply than calculated by perturbation theory at scales smaller than this.

The present-day density power spectrum $\mathcal{P}_\delta(k)$ can be determined observationally from the distribution of galaxies (Fig. 14). The quantity plotted is usually $P_\delta(k)$. It should go as

$$\begin{aligned} P_\delta(k) &\propto k^n && \text{for} \quad k \ll k_{\rm eq} \\ P_\delta(k) &\propto k^{n-4}\ln k && \text{for} \quad k \gg k_{\rm eq}\,. \end{aligned} \tag{371}$$

See Fig. 15.

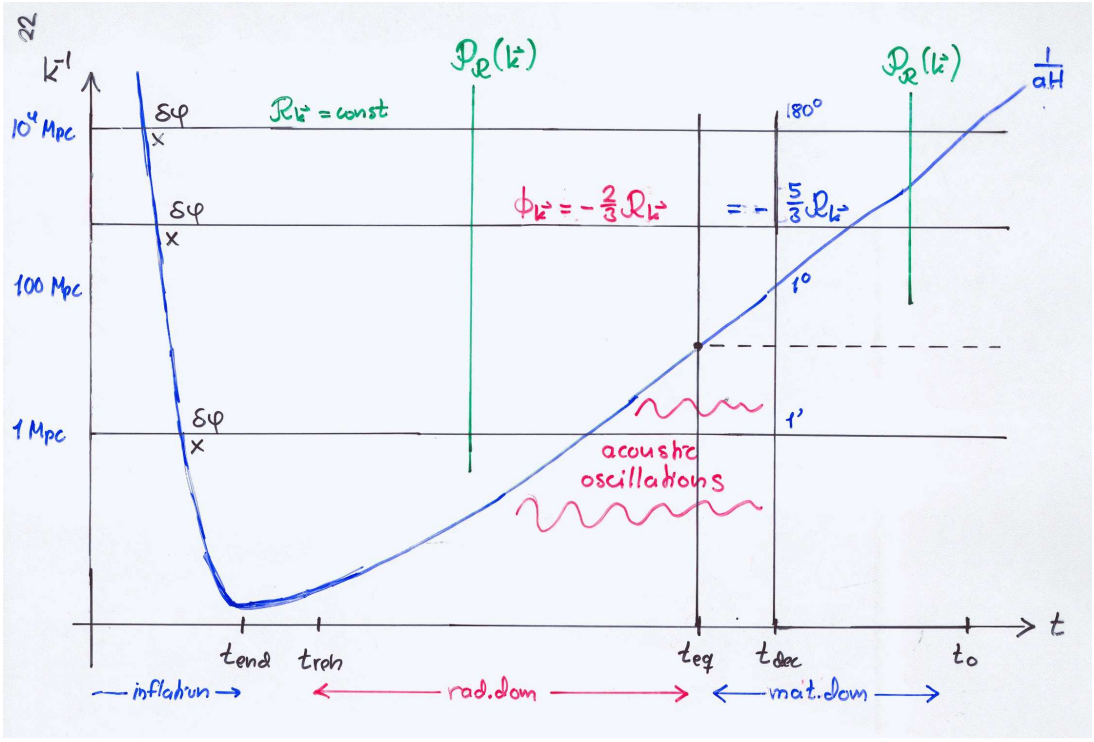


Figure 13: The whole picture of structure formation theory from quantum fluctuations during inflation to the present-day power spectrum at t_0.

Example: Errors on $P(k)$ estimation. The accuracy of power spectrum estimation from a galaxy survey is affected by *sample variance* and *shot noise*. Sample (or cosmic) variance was discussed earlier. Shot noise comes from the fact that instead of observing a continuous matter distribution, we are sampling it with a finite number density n of galaxies. To reduce sample variance one should increase the survey volume V_s. To reduce shot noise, one should increase n, by observing also smaller and fainter galaxies. We skip the math for shot noise (discussed in Galaxy Survey Cosmology), and give just the result that for Gaussian perturbations the combined effect of sample variance and shot noise is

$$\langle |\hat{P}(k) - P(k)|^2 \rangle \;=\; \frac{2}{N_k}\left[P(k) + \frac{1}{\bar{n}}\right]^2 , \tag{372}$$

where $\bar{n}$ is the mean number density of galaxies in the survey. (This is Eq. (335) modified by adding the effect of shot noise.)

Let us see how well this corresponds to the $P(k)$ estimate obtained in [15] for the SDSS survey (Fig. 15). Table I of [15] gives their estimate $\hat{P}(k)$ and their estimated uncertainty $\Delta P(k)$ for each of their 20 k-bins in numbers. We copy these numbers into our Table 1. The number of LRG galaxies in the survey was $N = 53\,860$. The effective sky area of the survey was $\Omega_s = 4259\,\mathrm{deg}^2 = 1.297\,\mathrm{sr}$, and the surveyed redshift range was $z = 0.155$–0.474. From Figs. 1 and 2 of [15] this corresponds to a comoving distance range $r = r_{\min} - r_{\max} = 450 - 1300\,h^{-1}\mathrm{Mpc}$, from which we get a survey volume and number density

$$\begin{aligned} V_s &= \Omega_s \left(\frac{r_{\max}^3}{3} - \frac{r_{\min}^3}{3} \right) = 0.9107\,h^{-3}\mathrm{Gpc}^3 , \\ \bar{n} &= \frac{N}{V_s} = 6.408 \times 10^{-5}\,h^3\mathrm{Mpc}^{-3} . \end{aligned} \tag{373}$$

If the survey volume were a cube, $V_s = L^3$, the side of the cube would be $L = 969.3\,h^{-1}\mathrm{Mpc}$, which means the components of $\mathbf{k}$ are integer multiples of $k_f = 2\pi/L = 0.006482\,h/\mathrm{Mpc}$. The volume of a k-bin

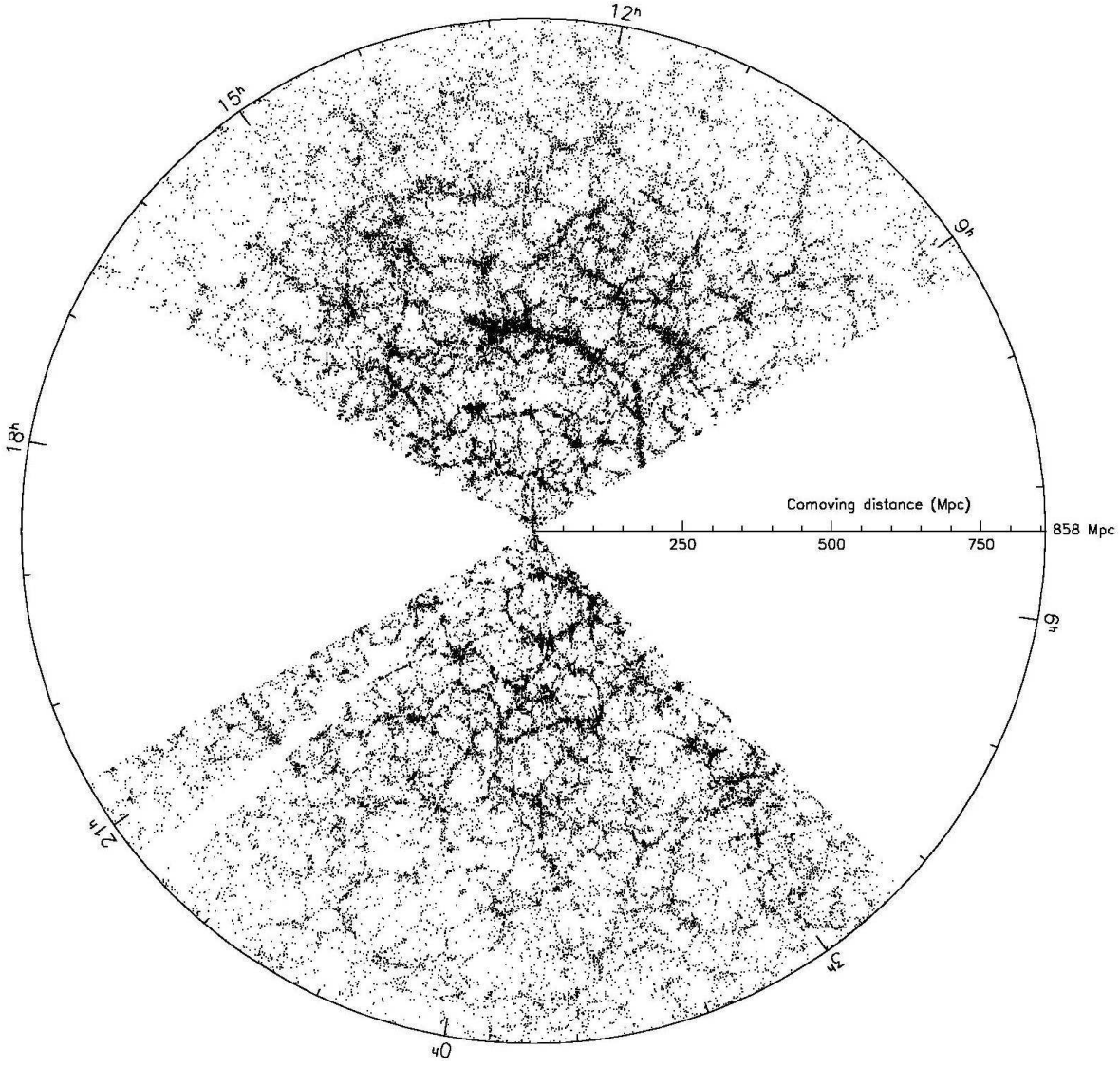


Figure 14: Distribution of galaxies according to the Sloan Digital Sky Survey (SDSS). This figure shows galaxies that are within 2° of the equator and closer than 858 Mpc (assuming $H_0 = 71$ km/s/Mpc). Figure from astro-ph/0310571[14].

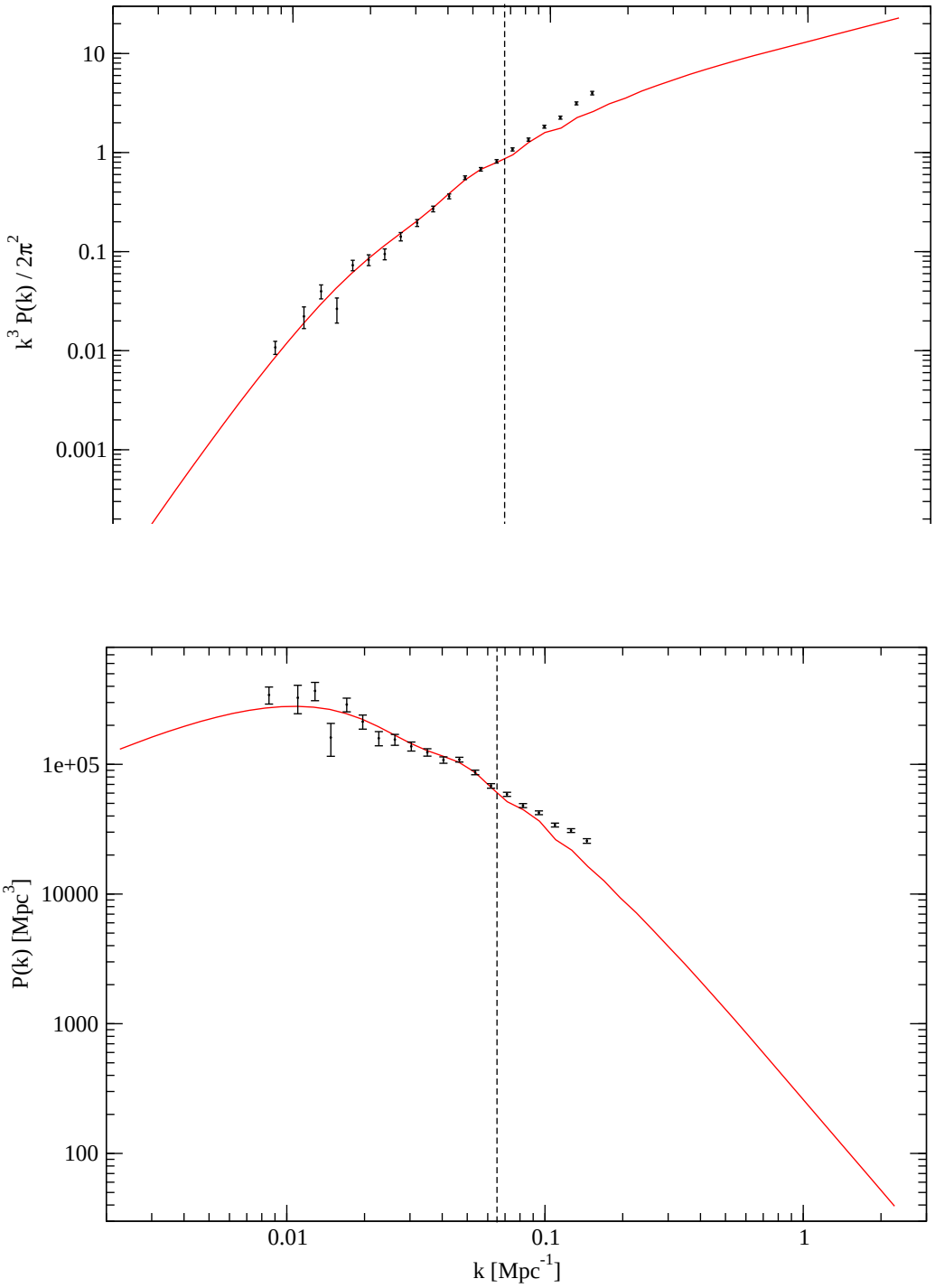


Figure 15: The power spectrum from the SDSS obtained using luminous red galaxies (LRG) [15]. Note that LRG are more strongly biased, $b \approx 1.9$, than galaxies on average, so to get the $P(k)$ of matter, one should divide by $b^2 \approx 3.6$. The top figure shows $\mathcal{P}_\delta(k)$ and the bottom figure $P_\delta(k)$. A Hubble constant value $H_0 = 71.4$ km/s/Mpc has been assumed for this figure. (These galaxy surveys only obtain the scales up to the Hubble constant, and therefore the observed $P_\delta(k)$ is usually shown in units of $h^{-3}\,\mathrm{Mpc}^3$ as a function of h Mpc^{-1}, so that no value for H_0 need to be assumed.) The black bars are the observations and the red curve is a theoretical fit, from linear perturbation theory, to the data. The bend in $P(k)$ at $k_{\mathrm{eq}} \sim 0.01$ Mpc^{-1} is clearly visible in the bottom figure. Linear perturbation theory fails when $\mathcal{P}(k) \gtrsim 1$, and therefore the data points do not follow the theoretical curve to the right of the dashed line (representing an estimate on how far linear theory can be trusted). Figure by R. Keskitalo.

is

$$V_k \;=\; \frac{4\pi}{3}\left(k_{\rm max}^3 - k_{\rm min}^3\right)\,. \tag{374}$$

and the number of $\mathbf{k}$ modes in the bin is $N_k = V_k/k_f^3$, which we have added as the third column in Table 1.

Using the estimated $\hat{P}(k)$ in place of $P(k)$ on the rhs of Eq. (372), we obtain from it our estimate for the relative uncertainty as

$$\frac{\sqrt{\langle|\hat{P}(k)-P(k)|^2\rangle}}{P(k)} \;\approx\; \frac{\sqrt{\frac{2}{N_k}}\left[\hat{P}(k)+\frac{1}{\bar{n}}\right]}{\hat{P}(k)}\,, \tag{375}$$

which we have added as the last column of Table 1.

Compared to the actual SDSS analysis the above is extremely naive. The facts that the survey volume is not a cube and that because of selection effects the mean number density of galaxies decreases as a function of z required a more sophisticated analysis. Also the k-bins SDSS used where not sharp $[k_{\rm min}, k_{\rm max}]$, but instead for each bin they used a smooth window function (of k) giving more weight to the modes near the center of the bin and some weight even to modes outside $[k_{\rm min}, k_{\rm max}]$. In addition, nonlinear effects cause deviations from Gaussianity, which typically increase the sample variance, and there are other effects contributing to the estimate uncertainty; so that we should expect (375) to be an underestimate. Nevertheless, comparing the last column of Table 1 (our naive estimate) to the second-to-last column (the SDSS uncertainty estimate) we see that we are on the right track.

SDSS LRG Power spectrum

$k_{\rm min}$	$k_{\rm max}$	N_k	$\hat{P}(k)$	$\Delta P(k)$	$\Delta P(k)/\hat{P}(k)$	Eq. (375)
0.008	0.017	68	124884	18775	0.150	0.193
0.013	0.018	56	118814	29400	0.247	0.214
0.016	0.022	101	134291	21638	0.161	0.157
0.018	0.025	151	58644	16647	0.284	0.146
0.021	0.028	195	105253	12736	0.121	0.116
0.025	0.033	312	77699	9666	0.124	0.096
0.029	0.037	404	57870	7264	0.126	0.089
0.033	0.043	670	56516	5466	0.097	0.070
0.037	0.051	1261	50125	3991	0.080	0.052
0.042	0.057	1708	45076	2956	0.066	0.046
0.050	0.066	2499	39339	2214	0.056	0.040
0.057	0.075	3640	39609	1679	0.042	0.033
0.066	0.086	5360	31566	1284	0.041	0.029
0.076	0.099	8171	24837	991	0.040	0.025
0.088	0.113	11710	21390	778	0.036	0.023
0.101	0.128	16407	17507	629	0.036	0.021
0.108	0.145	27511	15421	516	0.033	0.017
0.126	0.165	38320	12399	430	0.035	0.016
0.159	0.190	43665	11237	382	0.034	0.016
0.181	0.218	68135	9345	384	0.041	0.014

Table 1: The first and second column give roughly the extent of each k-bin used in units of $h\,\mathrm{Mpc}^{-1}$. (Note that the bins overlap.) N_k is the corresponding number of k modes in the bin. $\hat{P}(k)$ and $\Delta P(k)$ are the estimates from [15] in units of $(\,h^{-1}\mathrm{Mpc})^3$, and in the last two columns we compare their ratio to the uncertainty estimate we get from the survey characteristics.

8.8.2 σ_8

Instead of A_s, the amplitude of the primordial power spectrum at a pivot scale, astronomers like to use the parameter σ_8 to describe the amplitude of cosmological perturbations of a cosmological

model. σ_8 is defined as the top-hat window matter density variance $\sigma_T(R)$ at scale $R = 8\,h^{-1}$Mpc predicted by linear perturbation theory in the cosmological model.

We mentioned earlier (Sec. 8.1.7) that according to observations, $\sigma_{T,g}(R = 8\,h^{-1}\text{Mpc}) \approx 1$. This means that nonlinear effects are becoming important at this scale. The difference between $\sigma_{T,g}(R = 8\,h^{-1}\text{Mpc})$ and σ_8 is due to 1) galaxy bias and 2) nonlinear effects. Both should work in the direction of making $\sigma_{T,g}(R = 8\,h^{-1}\text{Mpc})$ larger than σ_8, so we expect that the correct cosmological model has $\sigma_8 < 1$, but not necessarily by very much.

To calculate σ_8 in a cosmological model specified by, say, A_s, n_s, Ω_m, Ω_Λ, ω_m, and ω_b, is not simple and requires numerical computation. This is done, e.g., by the CAMB code. The computation involves the transfer function $T(k)$, the growth function $D(a)$, and doing the integral (48a) for $\sigma_T^2(R)$.

For the Planck 2018 best-fit ΛCDM model, $\sigma_8 = 0.8210$.

Planck 2018 best-fit model σ_8. Let's see how well the results of this chapter allow us to calculate the Planck 2018 best-fit ΛCDM model σ_8. The model has parameter values (see Table I, `Plik` best-fit column in [8])

$$\begin{aligned} \ln(10^{10}A_s^2) &= 3.0448 \quad \Rightarrow \quad A_s^2 = 21.006\times 10^{-10} \\ n_s &= 0.96605 \\ H_0 &= 0.8120 \\ \Omega_m &= 0.3158 \\ \omega_b &= 0.022383\,, \end{aligned} \tag{376}$$

where A_s is for the pivot scale $k_p = 0.05\,\text{Mpc}^{-1}$.

The primordial power spectrum is

$$\mathcal{P}_\mathcal{R}(k) = A_s^2\left(\frac{k}{k_p}\right)^{n_s-1}. \tag{377}$$

We want to find the present-day linear power spectrum $\mathcal{P}_\delta(k)$. If the universe would have stayed matter dominated until today, $\delta_\mathbf{k}$ and $\mathcal{R}_\mathbf{k}$ would be related by

$$\tilde{\delta}_\mathbf{k} = \frac{2}{5}\left(\frac{k}{\tilde{H}_0}\right)^2 T(k)\mathcal{R}_\mathbf{k}\,, \tag{378}$$

where $\tilde{}$ denotes matter-dominated-model quantities, and "today" is defined by $a = 1$. As we noted in the footnote in Sec. 8.3.5, the comparison is to a matter-dominated model with the same matter density today, so it has a smaller total density and thus a smaller Hubble constant, $\tilde{H}_0 = \Omega_m^{1/2}H_0 = 37.83\,\text{km/s/Mpc}$. The true linear $\delta_\mathbf{k}$ differs from (378) due to the different growth function. For the matter-dominated model $D_m(a) = a$, so $D_m(1) = 1$, and

$$\begin{aligned} \delta_\mathbf{k}(t) &= \frac{2}{5}\frac{D(a)}{a}\left(\frac{k}{\tilde{\mathcal{H}}(a)}\right)^2 T(k)\mathcal{R}_\mathbf{k} \\ \delta_\mathbf{k}(t_0) &= \frac{2}{5}D(1)\left(\frac{k}{\tilde{H}_0}\right)^2 T(k)\mathcal{R}_\mathbf{k}\,, \end{aligned} \tag{379}$$

where $D(a)$ is the growth function of the ΛCDM model, given by Eq. (239). Thus the power spectra are related by

$$\mathcal{P}_\delta(k) = \frac{4}{25}D(1)^2\left(\frac{k}{\tilde{H}_0}\right)^4 T(k)^2\mathcal{P}_\mathcal{R}(k)\,. \tag{380}$$

We now calculate $\mathcal{P}_\delta(k)$ for $k = 1/(8\,h^{-1}\text{Mpc})$. For this k,

$$\begin{aligned} \frac{k}{k_p} &= \frac{1}{0.05\,\text{Mpc}^{-1}8\,h^{-1}\text{Mpc}} = \frac{h}{0.4} = 1.683 \\ \frac{k}{\tilde{H}_0} &= 666.9\,, \end{aligned} \tag{381}$$

so that

$$\mathcal{P}_\delta(k) \;=\; 65.30 D(1)^2 T(k)^2 \,. \tag{382}$$

The BBKS transfer function, using $k_{\rm eq}^{-1} = 13.7 \Omega_m^{-1} h^{-2}\,{\rm Mpc} = 64.44\, h^{-1}{\rm Mpc}$, gives for this k

$$T_{\rm BBKS}(k) = 0.1435\,, \qquad \text{with slope} \quad -1.17933\,, \tag{383}$$

so taht

$$\mathcal{P}_\delta(k) \;=\; 1.346 D(1)^2 \left(\frac{T(k)}{T_{\rm BBKS}(k)} \right)^2 . \tag{384}$$

We should now calculate $D(1)$ for the Planck best-fit ΛCDM model, and run CAMB to get the true $T(k)$ for this model, but I am lazy here, and use the results for our reference model ($\Omega_m = 0.3$, $h = 0.7$) from Sec. 8.3.5 and Fig. 11: $D(1) = 0.78$ and $T(k)/T_{\rm BBKS}(k) \approx 0.7$ to give

$$\mathcal{P}_\delta(k) \approx 0.40\,. \tag{385}$$

Finally we should calculate σ_T^2 for $R = 8\, h^{-1}{\rm Mpc}$, which is an integral of $\mathcal{P}_\delta(k)$ over k, but we try to get away with just using the value and slope at $k = 1/R$, i.e., we approximate $\mathcal{P}_\delta(k)$ with a power-law function. The slope is $n = 0.96605 - 2 \times 1.17933 = -1.39261$ (modifying the slope of $\mathcal{P}_\mathcal{R}(k)$ with the slope of $T_{\rm BBKS}(k)$) and, from (59),

$$\sigma_8^2 \;\approx\; \frac{9}{2^n}(n+1) \sin\frac{n\pi}{2}\; \frac{\Gamma(n-1)}{n-3} \mathcal{P}_\delta(k) \;=\; 1.939\, \mathcal{P}_\delta(k) \;\approx\; 0.776\,, \tag{386}$$

giving

$$\sigma_8 \approx 0.881 > 0.8210\,. \tag{387}$$

Using the power-law approximation should give an overestimate, since the true $\mathcal{P}_\delta(k)$ bends down compared to it on both sides of the chosen k value, and indeed we overestimated σ_8, but not badly.

8.8.3 Primordial gravitational waves

We found that outside the horizon tensor perturbations remain constant,

$$h_{\mathbf{k}}(t) = h_{\mathbf{k},\rm prim} = const\,, \tag{388}$$

whereas inside the horizon they become gravitational waves whose amplitude decays

$$|h_{\mathbf{k}}(t)| \propto a^{-1}\,. \tag{389}$$

Define the transfer function for gravitational waves

$$T_h(k) \equiv \frac{|h_{\mathbf{k}}(t_0)|}{h_{\mathbf{k},\rm prim}}\,, \tag{390}$$

so that the present-day power spectrum of primordial gravitational waves is

$$\mathcal{P}_{\rm grav}(k, t_0) = T_h(k)^2 \mathcal{P}_h(k)\,. \tag{391}$$

Make the approximation that the transition from (388) to (389) is instantaneous at horizon entry defined as

$$k = \mathcal{H} = aH\,. \tag{392}$$

Denote these values of a, H, and $\mathcal{H}$ by a_k, H_k, and $\mathcal{H}_k$. Then

$$T_h(k) \;=\; \frac{a_k}{a_0} \;=\; a_k\,. \tag{393}$$

The shape of the transfer function is determined by the rate at which different comoving scales k enter horizon as the universe expands. This is determined by the evolution of the comoving Hubble distance $\mathcal{H}^{-1}$.

In the matter-dominated universe

$$a \propto t^{2/3} \quad \text{and} \quad H = \frac{2}{3t} \propto a^{-3/2} \quad \Rightarrow \quad \mathcal{H} \propto a^{-1/2} \,. \tag{394}$$

Make first the approximation that the universe is still matter dominated. Then

$$T_h(k) \;=\; \frac{a_k}{a_0} \;=\; \left(\frac{\mathcal{H}_k}{\mathcal{H}_0}\right)^{-2} \;=\; \left(\frac{k}{a_0 H_0}\right)^{-2} \qquad (H_0 < k < k_{\rm eq}) \tag{395}$$

for scales that entered during the matter-dominated epoch.

To correct this result for the effect of dark energy at late times, we note that because of dark energy, the comoving Hubble distance $\mathcal{H}^{-1} = (aH)^{-1}$ stopped growing and began to shrink, so that the scale $k = H_0$ is actually exiting now, and it entered at an earlier time t_1 when the expansion was still (barely) matter dominated. Thus the above result for $T_h(k)$ should apply (roughly) at that earlier time:

$$T_h(t_1, k) \;=\; \left(\frac{k}{a_1 H_1}\right)^{-2} \;=\; \left(\frac{k}{a_0 H_0}\right)^{-2} \qquad (H_0 < k < k_{\rm eq}) \tag{396}$$

While the scale $k = H_0$ was inside the horizon, the universe expanded by about a factor of two, so the correct transfer function is about half of (395).

Exercise: Extend the result (395) to scales $k > k_{\rm eq}$. You can make the approximation where the transition from radiation-dominated expansion law to matter-dominated expansion law is instantaneous at $t_{\rm eq}$. (This approximation actually underestimates $T_h(k > k_{\rm eq})$ by a factor that roughly compensates the overestimation in (395) from ignoring dark energy at late times.)

Gravitational waves were detected for the first time on September 14, 2015 at the LIGO observatory. These were not primordial gravitational waves; they were caused by a collision of two black holes about 400 Mpc from here, and they were observed only for about 0.2 seconds. The peak amplitude was $h \approx 10^{-21}$. LIGO is sensitive to frequencies near 100 Hz, and with further refinements it is expected to reach a sensitivity of $h = 10^{-22}$. Assume the primordial tensor perturbations had amplitude $h = 10^{-5}$ (close to the upper limit from CMB observations). What is their amplitude today at the 100 Hz frequency?

ESA is planning to launch a space gravitational wave observatory (LISA) in 2034. It would have similar sensitivity as LIGO, but for frequencies lower by a factor 10^{-4}. What do you conclude about the prospect for observing primordial gravitational waves this way?

References

[1] C.M. Baugh, *The real-space correlation function measured from the APM Galaxy Survey*, Mon. Not. R. Astron. Soc. **280**, 267 (1996), astro-ph/9512011

[2] A.J. Benson, R.G. Bower, C.S. Frenk, C.G. Lacey, C.M. Baugh, and S. Cole, *What shapes the luminosity function of galaxies?*, Astrophys. J. **599**, 38 (2003), astro-ph/0302450

[3] A.G. Sánchez et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function*, Mon. Not. R. Astron. Soc. **425**, 415 (2012), arXiv:1203.6616

[4] E. Hawkins et al., *The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe*, Mon. Not. R. Astron. Soc. **346**, 78 (2003),astro-ph/0212375

[5] J.A. Peacock: Cosmological Physics (Cambridge University Press 1999), Chapter 16

[6] A.R. Liddle and D.H. Lyth: Cosmological Inflation and Large-Scale Structure (Cambridge University Press 2000)

[7] Planck Collaboration, Astronomy & Astrophysics **594**, A13 (2016), arXiv:1502.01589

[8] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, arXiv:1807.06209

[9] S. Dodelson: Modern Cosmology (Academic Press 2003), Chapter 7

[10] J.M. Bardeen, J.R. Bond, N. Kaiser, A.S. Szalay, *The statistics of peaks in Gaussian randomn fields*, Astrophys. J. **304**, 15 (1986), Appendix G

[11] E.W. Kolb and M.S. Turner: The Early Universe (Addison-Wesley 1990)

[12] P.A.R. Ade et al., Phys. Rev. Lett. **112**, 241101 (2014), arXiv:1403.3985

[13] Planck Collaboration, Astronomy & Astrophysics **586**, A133 (2016), arXiv:1409.5738

[14] J. Richard Gott III et al., *A Map of the Universe*, Astrophys. J. **624**, 463 (2005), astro-ph/0310571

[15] M. Tegmark et al., *Cosmological Constraints from the SDSS Luminous Red Galaxies*, Phys. Rev. **D74**, 123507 (2006), astro-ph/0608632

9 Cosmic Microwave Background Anisotropy

9.1 Introduction

The cosmic microwave background (CMB) is isotropic to a high degree. This tells us that the early universe was rather homogeneous at the time ($t = t_{\rm dec} \approx 370\,000$ years) the CMB was formed. However, with precise measurements we can detect a low-level anisotropy in the CMB (Fig. 1) which reflects the small perturbations in the early universe.

This anisotropy was first detected by the COBE (Cosmic Background Explorer) satellite in 1992, which mapped the whole sky in three microwave frequencies. The angular resolution of COBE was rather poor, 7°, meaning that only features larger than this were detected. Measurements with better resolution, but covering only small parts of the sky were then performed using instruments carried by balloons to the upper atmosphere, and ground-based detectors located at high altitudes. A significant improvement came with the WMAP (Wilkinson Microwave Anisotropy Probe) satellite, which made observations for nine years, from 2001 to 2010.

The best CMB anisotropy data to date, covering the whole sky, has been provided by the Planck satellite (Fig. 2). Planck was launched by the European Space Agency (ESA), on May 14th, 2009, to an orbit around the L2 point of the Sun-Earth system, 1.5 million kilometers from the Earth in the anti-Sun direction. Planck made observations for over four years, from August 12th, 2009 until October 23rd, 2013. The first major release of Planck results was in 2013 [1] and the second release in 2015 [2]. Final Planck results were released in 2018 and 2019.

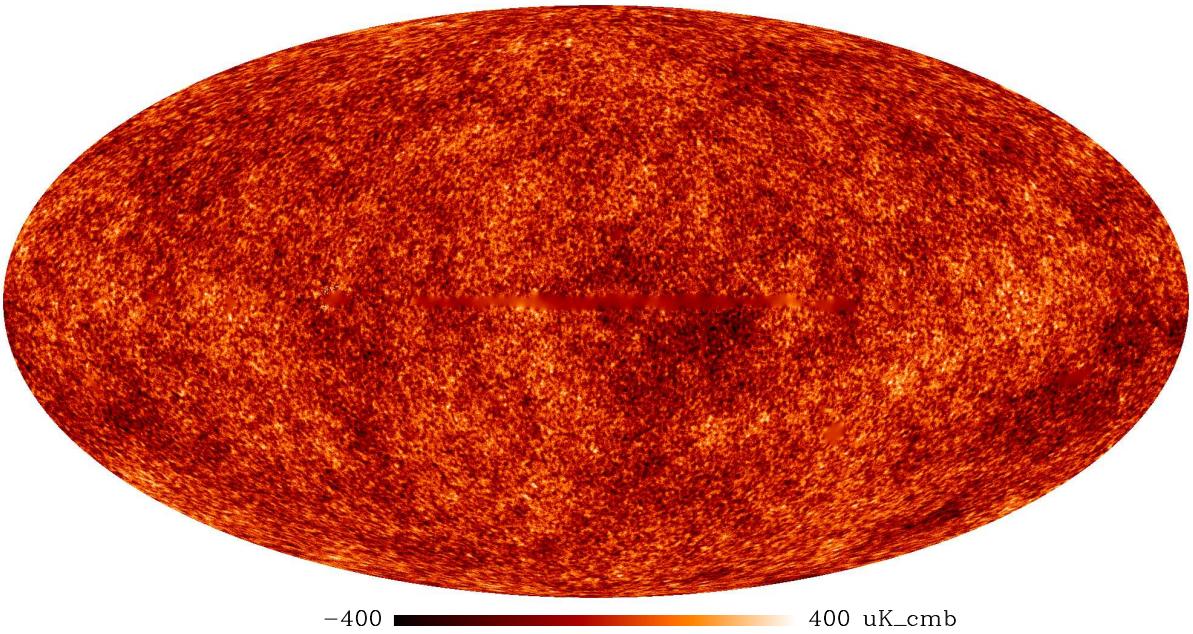


Figure 1: Cosmic microwave background. The figure shows temperature variations from $-400\,\mu$K to $+400\,\mu$K around the mean temperature (2.7255 K) over the whole sky, in galactic coordinates. The color is chosen to mimic the true color of CMB at the time it was formed, when it was visible orange-red light, but the brightness variation (the anisotropy) is hugely exaggerated by the choice of color scale. The fuzzy regions, notable especially in the galactic plane, are regions of the sky where microwave radiation from our own galaxy or nearby galaxies makes it difficult to separate out the CMB. (ESA/Planck data).

Planck observed the entire sky twice in a year. The satellite repeated these observations year after year, and the results become gradually more accurate, since the effects of instrument noise averaged out and various instrument-related systematic effects could be determined and corrected better with repeated observations.

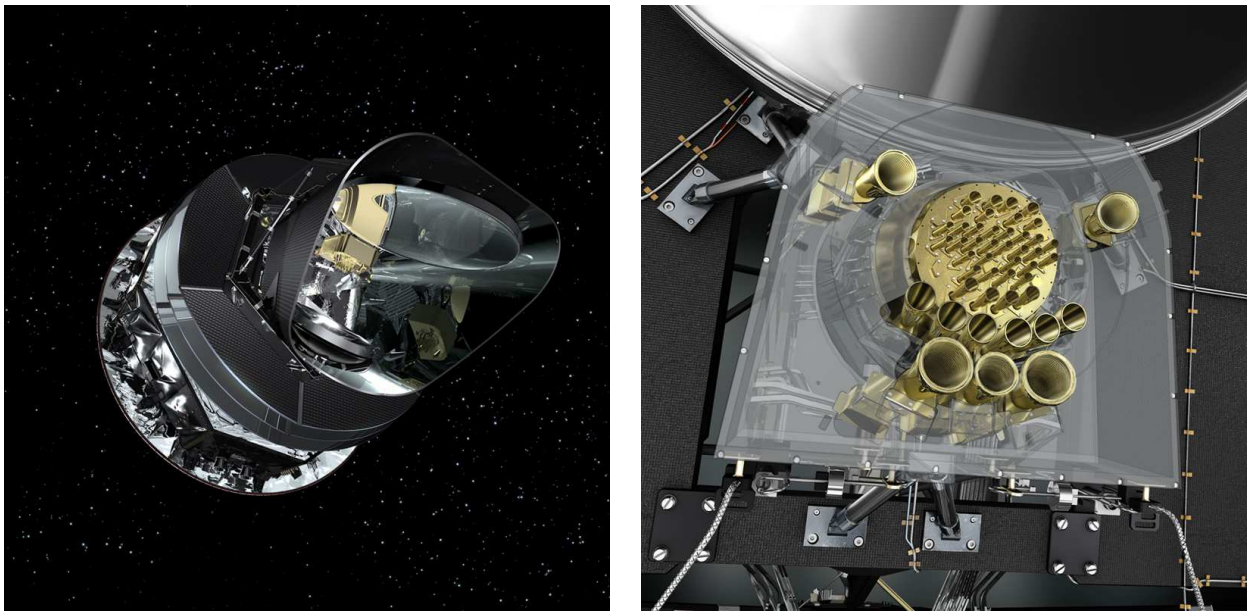


Figure 2: The Planck satellite and its microwave receivers. The larger horns are for receiving lower frequencies and the smaller horns for higher frequencies.

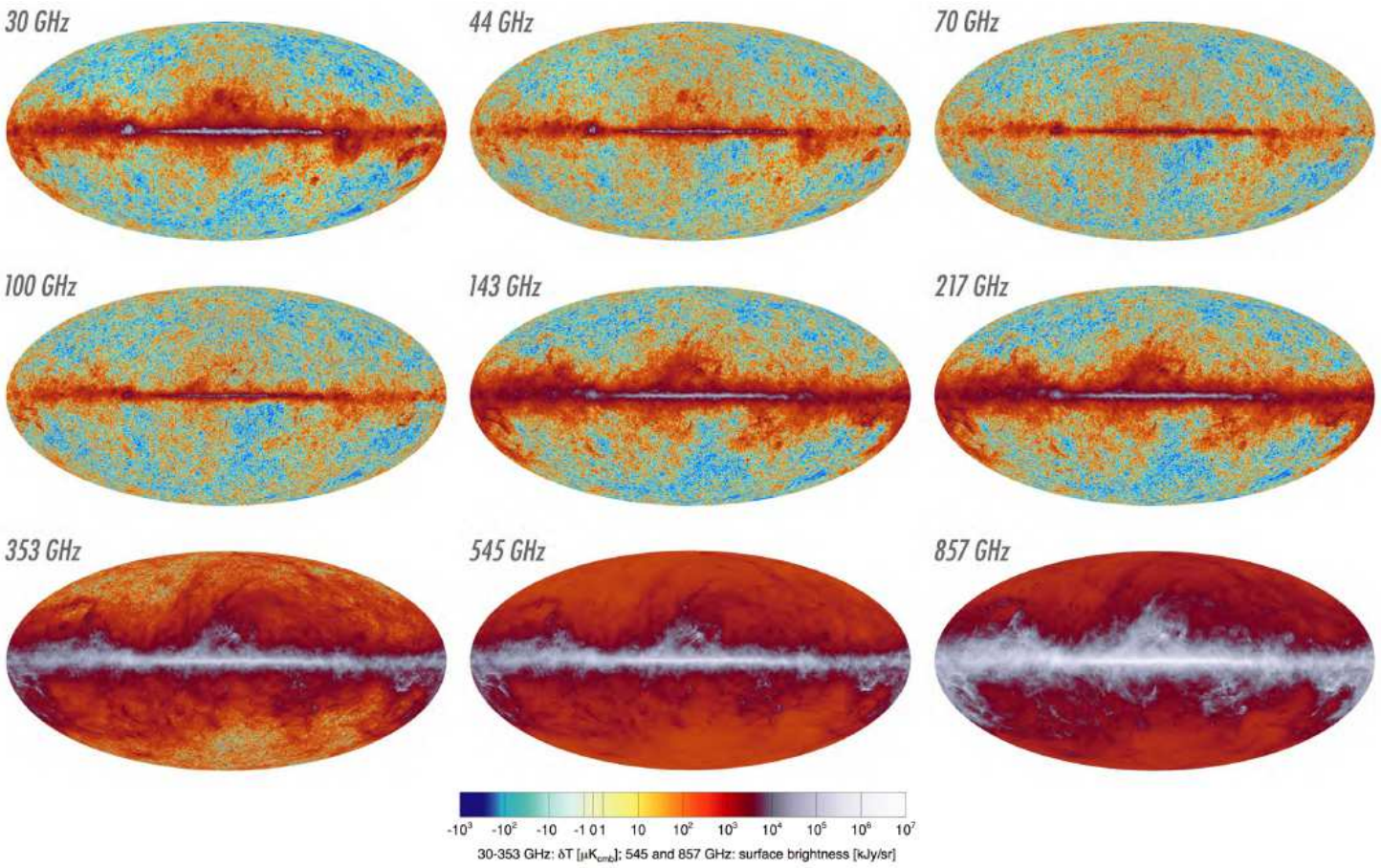


Figure 3: Brightness of the sky in the nine Planck frequency bands. These sky maps are in galactic coordinates so the Milky Way lies horizontally. From [2].

In addition to the CMB, there is microwave radiation from our own galaxy and other galaxies, called *foreground* by those who study CMB. This radiation can be separated from the CMB based on its different electromagnetic spectrum. To enable this *component separation*, Planck observed at 9 different frequency bands; the lowest one centered at 30 GHz and the highest at 857 GHz (Fig. 3). There were two different instruments on Planck, using different technologies to detect the variations in the microwave radiation. The Low Frequency Instrument (LFI) used radiometers for the 30, 44, and 70 GHz bands. The High Frequency Instrument (HFI) used bolometers for the bands from 100 GHz to 857 GHz. HFI is the barrel-shaped instrument at the center in Fig. 2 right panel and LFI was wrapped around it. With the additional help of WMAP and ground-based data 8 different foreground components could be distinguished (Fig. 4).

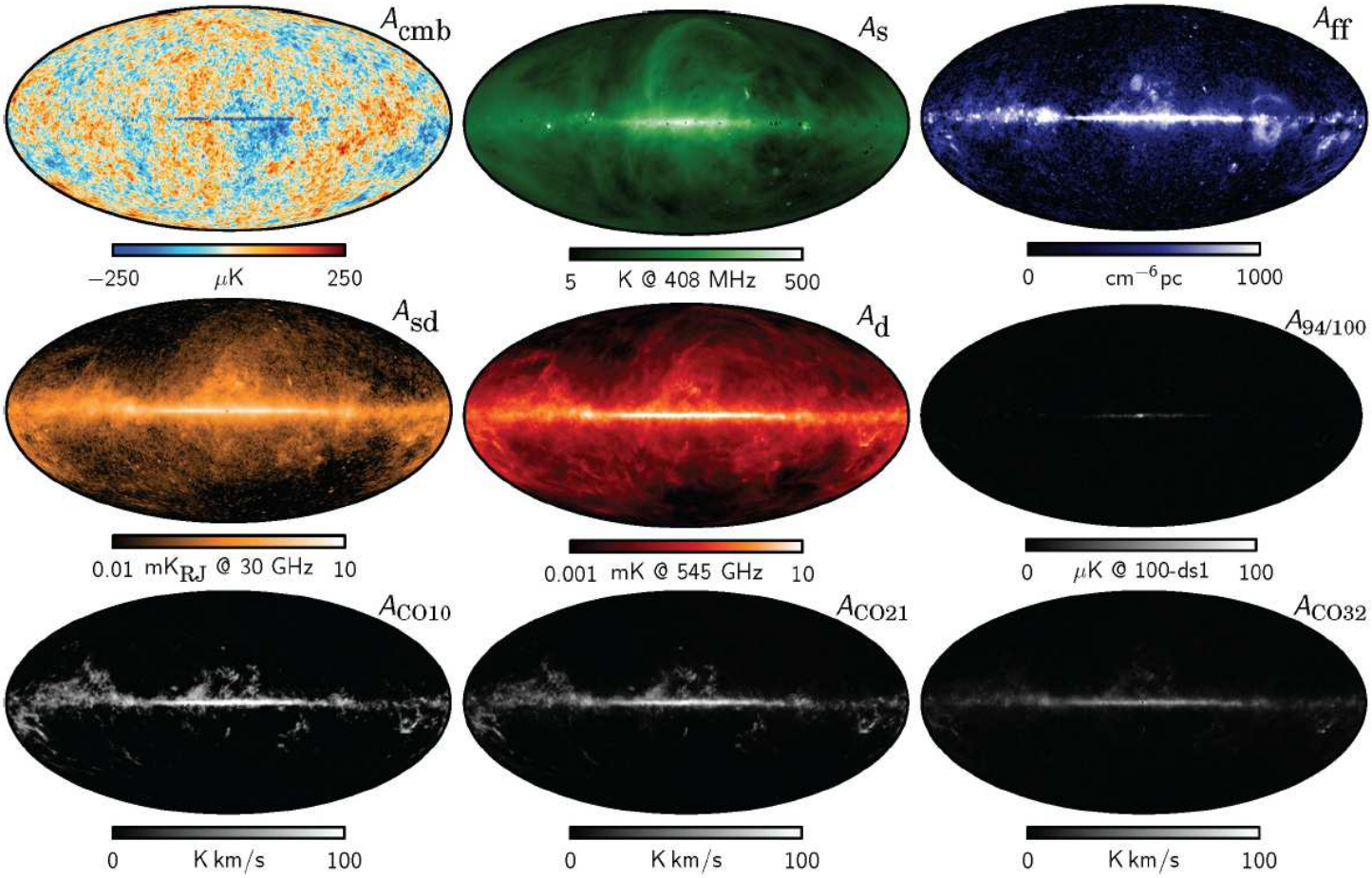


Figure 4: Result from Planck component separation. Also WMAP data and ground-based 408 MHz data was used. The extracted nine different components of the microwave radiation from top left to bottom right are: 1) CMB; 2) synchrotron radiation generated by relativistic cosmic-ray electrons accelerated by the galactic magnetic field; 3) "free-free emission" (bremsstrahlung) from electron-ion collisions; 4) emission from spinning galactic dust grains due to their electric dipole moment; 5) thermal emission from galactic dust (the typical dust temperatures are of order 20 K, so the dust thermal spectrum is peaked at much higher frequencies than CMB); 6) spectral line emission from HCN, CN, HCO, CS, and other molecules; 7) spectral line emission from the CO (carbon monoxide) $J = 1 \rightarrow 0$ transition; 8) CO $J = 2 \rightarrow 1$ line; 9) CO $J = 3 \rightarrow 2$ line (these emission lines from transitions between the four lowest rotation states of the CO molecule map the distribution of carbon monoxide in the Milky Way). From [2].

Figures 5–7 show the observed variation δT in the temperature of the CMB on the sky (red means hotter than average, blue means colder than average).

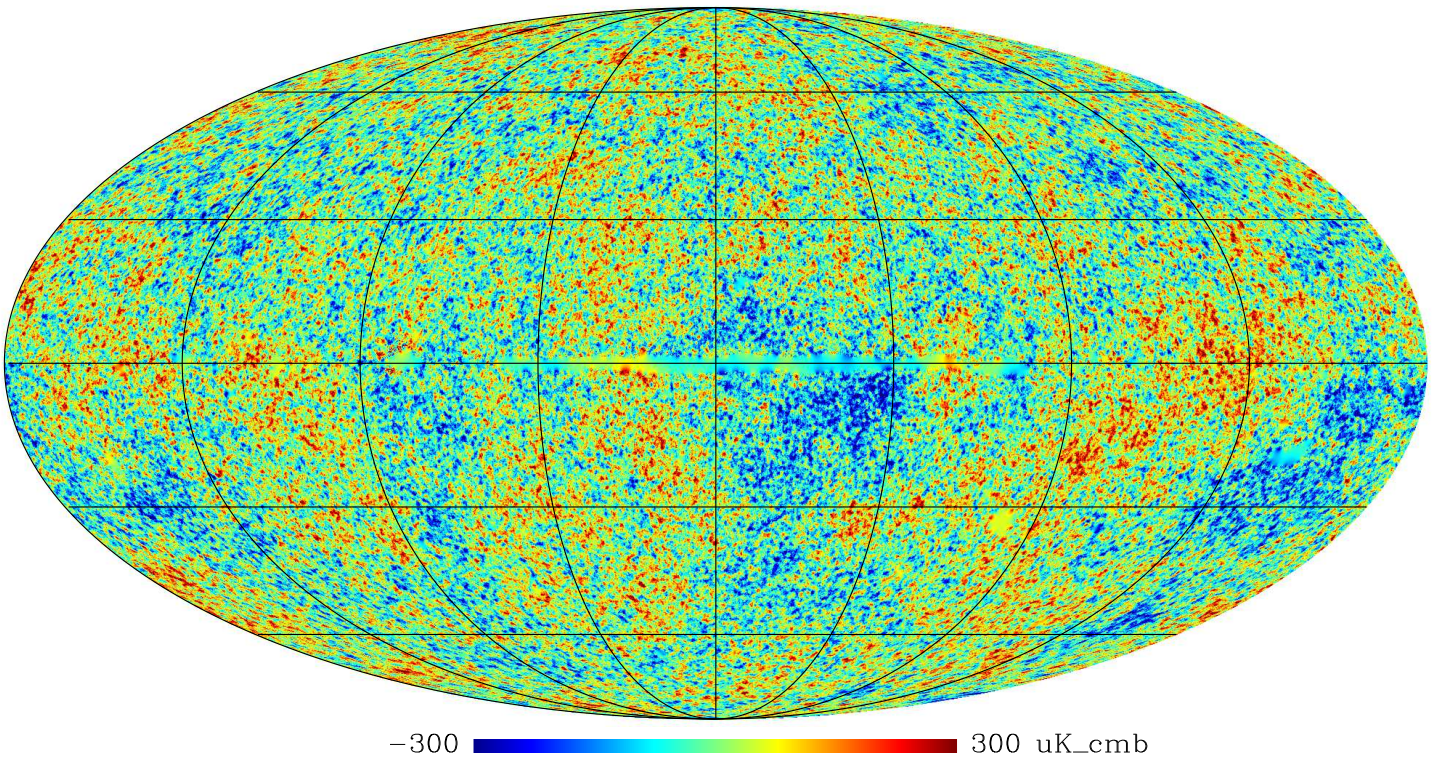


Figure 5: Cosmic microwave background: Fig. 1 reproduced in false color to bring out the patterns more clearly. The color range corresponds to CMB temperature variations from $-300\,\mu$K (blue) to $+300\,\mu$K (red) around the mean temperature. (ESA/Planck data).

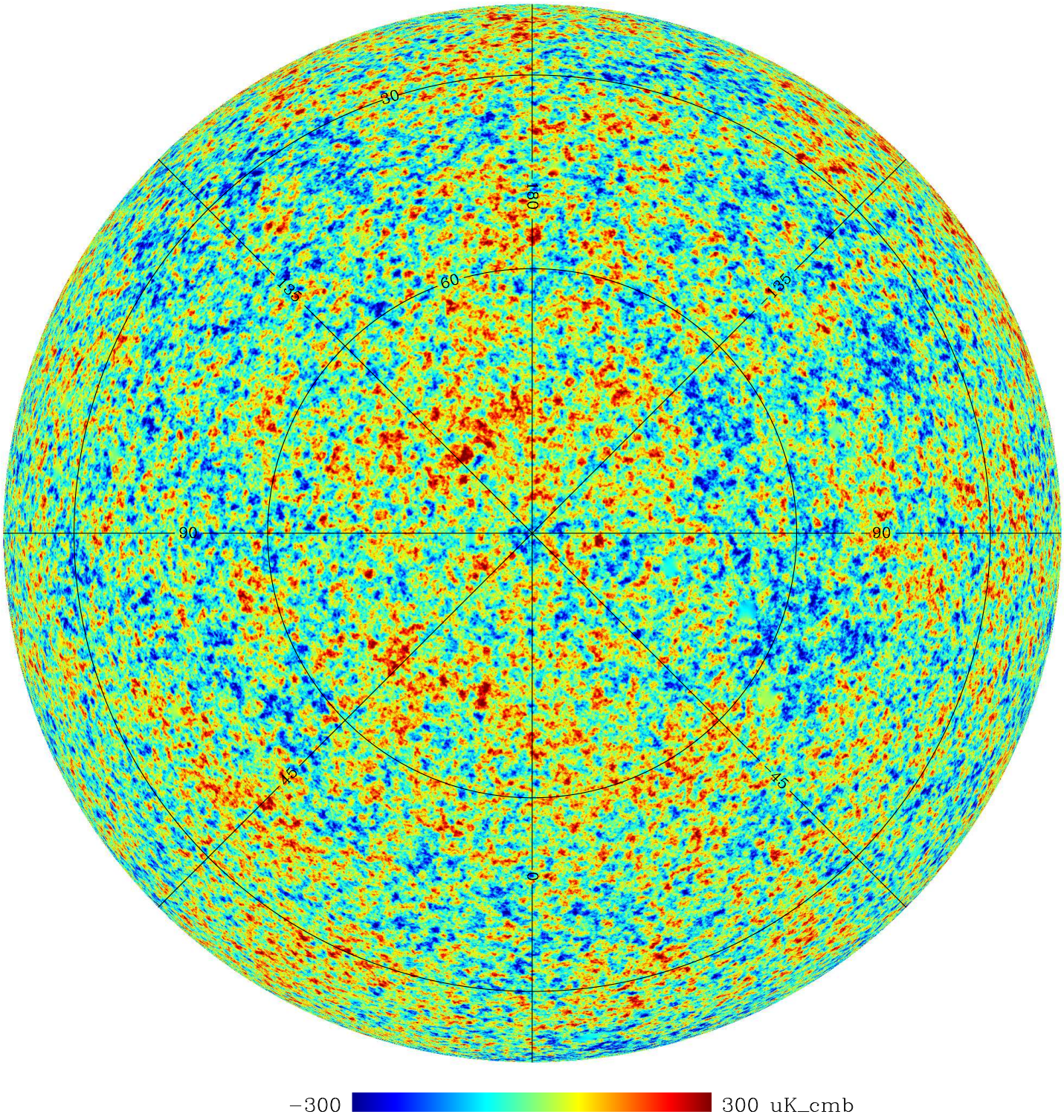


Figure 6: The northern galactic hemisphere of the CMB sky (ESA/Planck data).

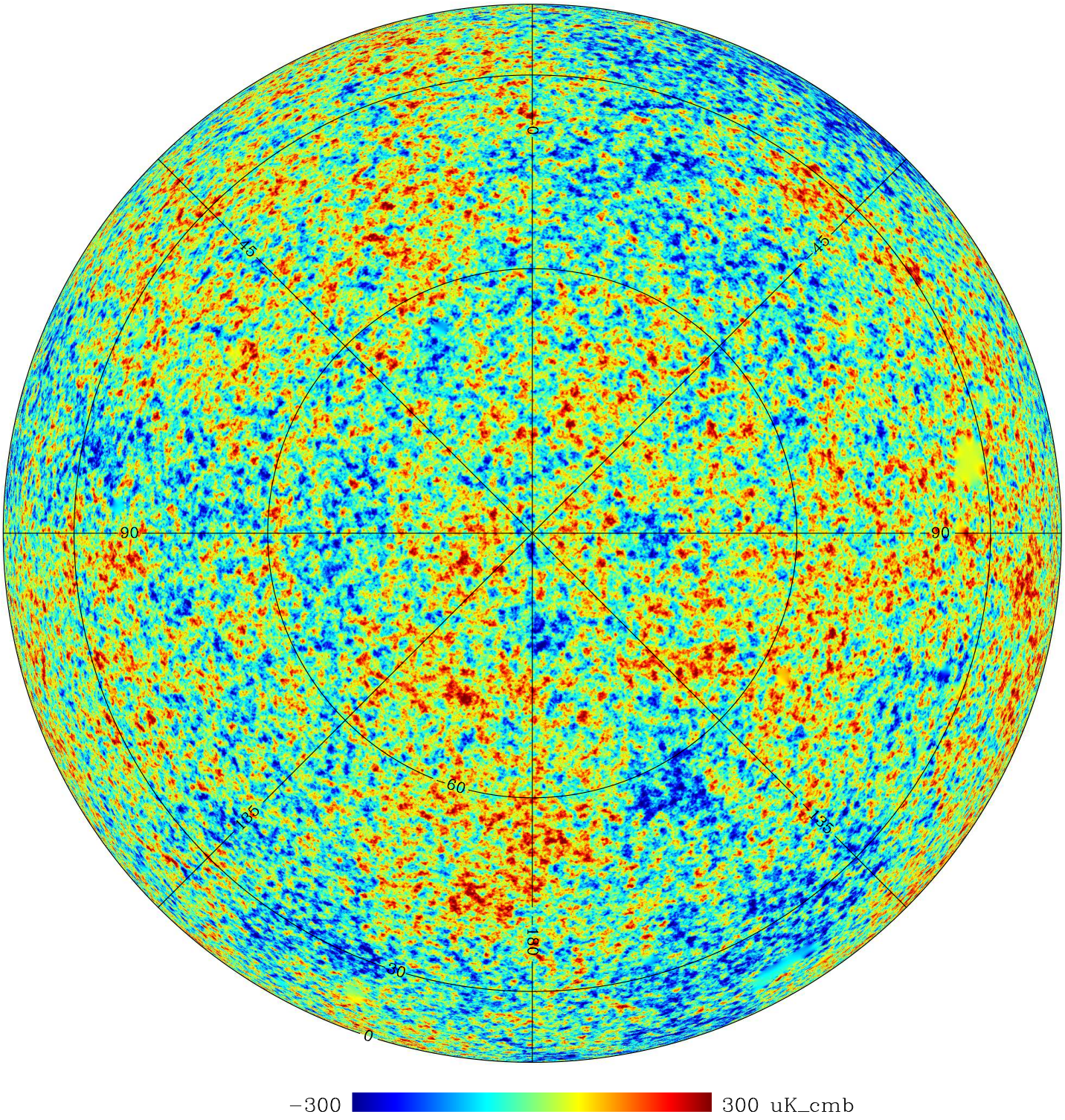


Figure 7: The southern galactic hemisphere of the CMB sky. The conspicuous cold region around $(-150°, -55°)$ is called the Cold Spot. The yellow smooth spot at $(-80°, -35°)$ in galactic coordinates is a region where the CMB is obscured by the Large Magellanic Cloud, and the light blue spot at $(-150°, -20°)$ is due to the Orion Nebula. (ESA/Planck data).

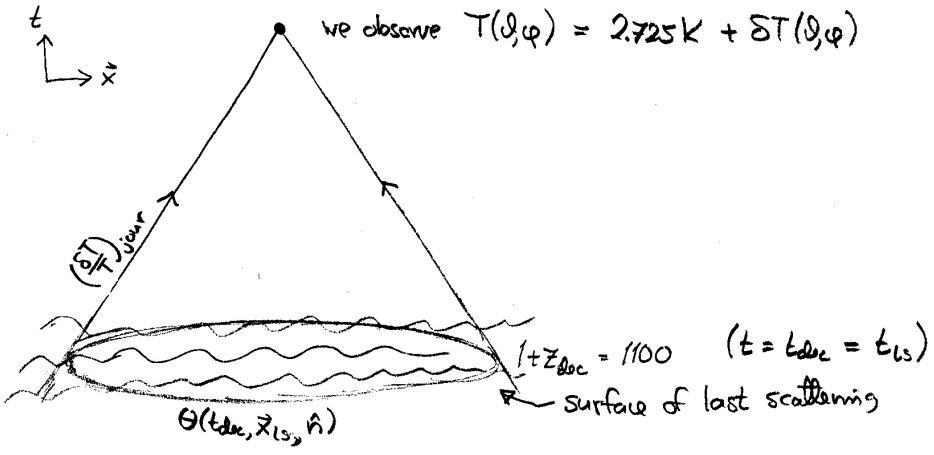


Figure 8: The observed CMB temperature anisotropy gets a contribution from the last scattering surface, $(\delta T/T)_{\rm intr} = \Theta(t_{\rm dec}, \mathbf{x}_{\rm ls}, \hat{\mathbf{n}})$ and from along the photon's journey to us, $(\delta T/T)_{\rm jour}$.

The photons we see as the CMB, have traveled to us from where our past light cone intersects the hypersurface corresponding to the time $t = t_{\rm dec}$ of photon decoupling. This intersection forms a sphere which we shall call the *last scattering surface.*[1] We are at the center of this sphere, except that timewise the sphere is located in the past.

The observed temperature anisotropy is due to two contributions, an *intrinsic* temperature variation at the surface of last scattering and a variation in the redshift the photons have suffered during their "journey" to us,

$$\left(\frac{\delta T}{T}\right)_{\rm obs} = \left(\frac{\delta T}{T}\right)_{\rm intr} + \left(\frac{\delta T}{T}\right)_{\rm jour} . \tag{1}$$

See Fig. 8.

The first term, $\left(\frac{\delta T}{T}\right)_{\rm intr}$ represents the temperature variation of the photon gas at $t = t_{\rm dec}$. We also include in it the Doppler effect from the motion of this photon gas. At that time the larger scales we see in the CMB sky were still outside the horizon, so we have to pay attention to the gauge choice. In fact, the separation of $\delta T/T$ into the two components in Eq. (1) is gauge-dependent. If the time slice $t = t_{\rm dec}$ dips further into the past in some location, it finds a higher temperature, but the photons from there also have a longer way to go and suffer a larger redshift, so that the two effects balance each other. We can calculate in any gauge we want, getting different results for $(\delta T/T)_{\rm intr}$ and $(\delta T/T)_{\rm jour}$ depending on the gauge, but their sum $(\delta T/T)_{\rm obs}$ is gauge independent. It has to be, being an observed quantity.

One might think that $(\delta T/T)_{\rm intr}$ should be equal to zero, since in our earlier discussion of recombination and decoupling we identified decoupling with a particular temperature $T_{\rm dec} \sim 3000\,{\rm K}$. This kind of thinking corresponds to a particular gauge choice where the $t = t_{\rm dec}$ time slice coincides with the $T = T_{\rm dec}$ hypersurface. In this gauge $(\delta T/T)_{\rm intr} = 0$, except for the Doppler effect (we are not going to use this gauge). Anyway, it is not true that all photons have their last scattering exactly when $T = T_{\rm dec}$. Rather they occur during a rather large temperature interval and time period. The zeroth-order (background) time evolution of the temperature of the photon distribution is the same before and after last scattering, $T \propto a^{-1}$, so it does not matter how we draw the artificial separation line, the time slice $t = t_{\rm dec}$ separating the fluid and free particle treatments of the photons. See Fig. 9.

[1] Or the *last scattering sphere.* "Last scattering surface" often refers to the entire $t = t_{\rm dec}$ time slice.

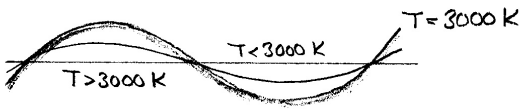


Figure 9: Depending on the gauge, the $T = T_{\rm dec}$ surface may, or (usually) may not coincide with the $t = t_{\rm dec}$ time slice.

9.2 Multipole analysis

The CMB temperature anisotropy is a function over a sphere (the celestial sphere, or the unit sphere of directions $\hat{\mathbf{n}}$). In analogy with the Fourier expansion in 3D space, we separate out the contributions of different angular scales by doing a multipole expansion,

$$\frac{\delta T}{T_0}(\theta,\phi) = \sum a_{\ell m} Y_{\ell m}(\theta,\phi) \tag{2}$$

where the sum runs over $\ell = 1, 2, \ldots \infty$ and $m = -\ell, \ldots, \ell$, giving $2\ell + 1$ values of m for each ℓ. The functions $Y_{\ell m}(\theta,\phi)$ are the *spherical harmonics* (see Fig. 10), which form an orthonormal set of functions over the sphere, so that we can calculate the multipole coefficients $a_{\ell m}$ from

$$a_{\ell m} = \int Y^*_{\ell m}(\theta,\phi)\frac{\delta T}{T_0}(\theta,\phi)d\Omega\,. \tag{3}$$

Definition (2) gives dimensionless $a_{\ell m}$. Often they are defined without the $T_0 = 2.7255\,\mathrm{K}$ in Eq. (2), and then they have the dimension of temperature and are usually given in units of μK. Here θ and ϕ are spherical coordinates, $d\Omega \equiv d\cos\theta d\phi$, θ ranges from 0 to π and ϕ ranges from 0 to 2π.[2]

The sum begins at $\ell = 1$, since $Y_{00} = $ const. and therefore we must have $a_{00} = 0$ for a quantity which represents a deviation from average. The dipole part, $\ell = 1$, is dominated by the Doppler effect due to the motion of the solar system with respect to the last scattering surface, and we cannot separate out from it the *cosmological dipole* caused by large scale perturbations. Therefore we are here interested only in the $\ell \geq 2$ part of the expansion.

Another notation for $Y_{\ell m}(\theta,\phi)$ is $Y_{\ell m}(\hat{\mathbf{n}})$, where $\hat{\mathbf{n}}$ is a unit vector whose direction is specified by the angles θ and ϕ.

9.2.1 Spherical harmonics

We list here some useful properties of the spherical harmonics.

They are orthonormal functions on the sphere, so that

$$\int d\Omega\; Y_{\ell m}(\theta,\phi)Y^*_{\ell' m'}(\theta,\phi) = \delta_{\ell\ell'}\delta_{mm'}\,. \tag{4}$$

They are elementary complex functions and are related to the *associated Legendre functions* $P^m_\ell(x)$ by

$$Y_{\ell m}(\theta,\phi) = (-1)^m\sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}P^m_\ell(\cos\theta)e^{im\phi}\,. \tag{5}$$

[2] They can also be given in degrees, the *colatitude* θ ranging from 0° (North) to 180° (South) and the *longitude* ϕ from 0° to 360°. There are a number of different astronomical coordinate systems (equatorial, ecliptic, galactic) in use, with their own historical conventions for the coordinate names, symbols, and units. Typically they involve the *latitude* $90° - \theta$ instead of the colatitude, so that North is at +90° and South at −90°, and the longitude is usually given between −180° and +180°, e.g., in Fig. 7.

Legendre polynomials
$P_0(x) = 1$
$P_1(x) = x$
$P_2(x) = \frac{1}{2}(3x^2 - 1)$
$P_3(x) = \frac{1}{2}(5x^3 - 3x)$
$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$
Associated Legendre functions $P_\ell^m(x) = P_\ell^m(\cos\theta)$
$P_1^1(x) = \sqrt{1-x^2} = \sin\theta$
$P_2^1(x) = 3x\sqrt{1-x^2} = 3\cos\theta\sin\theta$
$P_2^2(x) = 3(1-x^2) = 3\sin^2\theta$
Spherical harmonics
$Y_0^0(\theta,\phi) = \frac{1}{\sqrt{4\pi}}$
$Y_1^1(\theta,\phi) = -\sqrt{\frac{3}{8\pi}}\sin\theta e^{i\phi}$
$Y_1^0(\theta,\phi) = \sqrt{\frac{3}{4\pi}}\cos\theta$
$Y_2^2(\theta,\phi) = \sqrt{\frac{5}{96\pi}}3\sin^2\theta e^{i2\phi}$
$Y_2^1(\theta,\phi) = -\sqrt{\frac{5}{24\pi}}3\sin\theta\cos\theta e^{i\phi}$
$Y_2^0(\theta,\phi) = \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)$
Spherical Bessel functions
$j_0(x) = \dfrac{\sin x}{x}$
$j_1(x) = \dfrac{\sin x}{x^2} - \dfrac{\cos x}{x}$
$j_2(x) = \left(\dfrac{3}{x^3} - \dfrac{1}{x}\right)\sin x - \dfrac{3}{x^2}\cos x$

Table 1: Legendre functions, spherical harmonics, and spherical Bessel functions.

Thus the θ-dependence is in $P_\ell^m(\cos\theta)$ and the ϕ-dependence is in $e^{im\phi}$. The functions P_ℓ^m are real and

$$Y_{\ell,-m} = (-1)^m Y_{\ell m}^*\,, \tag{6}$$

so that

$$Y_{\ell 0} = \sqrt{\frac{2\ell+1}{4\pi}}P_\ell(\cos\theta) \qquad \text{is real.} \tag{7}$$

The functions $P_\ell \equiv P_\ell^0$ are called *Legendre polynomials*. See Table 9.2.1 for examples of these functions for $\ell \le 2$.

Summing over the m corresponding to the same multipole number ℓ gives the *addition theorem*

$$\sum_m Y_{\ell m}^*(\theta',\phi')Y_{\ell m}(\theta,\phi) = \frac{2\ell+1}{4\pi}P_\ell(\cos\vartheta)\,, \tag{8}$$

where ϑ is the angle between $\hat{\mathbf{n}} = (\theta, \phi)$ and $\hat{\mathbf{n}}' = (\theta', \phi')$, i.e., $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos\vartheta$. For $\hat{\mathbf{n}} = \hat{\mathbf{n}}'$ this becomes

$$\sum_m |Y_{\ell m}(\theta, \phi)|^2 = \frac{2\ell + 1}{4\pi} \tag{9}$$

(since $P_\ell(1) = 1$ always).

We shall also need the expansion of a plane wave in terms of spherical harmonics,

$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kx) Y_{\ell m}(\hat{\mathbf{x}}) Y^*_{\ell m}(\hat{\mathbf{k}}) \,. \tag{10}$$

Here $\hat{\mathbf{x}}$ and $\hat{\mathbf{k}}$ are the unit vectors in the directions of $\mathbf{x}$ and $\mathbf{k}$, and the j_ℓ are the spherical Bessel functions.

9.2.2 Theoretical angular power spectrum

The CMB anisotropy is due to primordial perturbations, and therefore it reflects their Gaussian nature. Because one gets the values of the $a_{\ell m}$ from the other perturbation quantities through linear equations (in first-order perturbation theory), the $a_{\ell m}$ are also (complex) Gaussian random variables. Since they represent a deviation from the average temperature, their expectation value is zero,

$$\langle a_{\ell m} \rangle = 0 \,. \tag{11}$$

From statistical isotropy follows that the $a_{\ell m}$ are independent random variables, except for the reality condition (13), so that

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = 0 \qquad \text{if } \ell \neq \ell' \text{ or } m \neq m'. \tag{12}$$

Since $\delta T/T_0$ is real,

$$a_{\ell,-m} = (-1)^m a^*_{\ell,m} \,. \tag{13}$$

Although thus $a_{\ell,-m}$ and $a_{\ell m}$ are not independent of each other, we still have $\langle a_{\ell m} a^*_{\ell,-m} \rangle = 0$ (**exercise**), so that (12) is satisfied even in this case. For each ℓ, there are $2\ell + 1$ independent real random variables: $a_{\ell 0}$ (which is always real), and $\mathrm{Re}\, a_{\ell m}$ and $\mathrm{Im}\, a_{\ell m}$ for $m = 1, \ldots, \ell$.

The quantity we want to calculate from theory is the variance $\langle |a_{\ell m}|^2 \rangle$ to get a prediction for the typical size of the $a_{\ell m}$. From statistical isotropy also follows that these expectation values depend only on ℓ not m. (The ℓ are related to the angular size of the anisotropy pattern, whereas the m are related to "orientation" or "pattern". See Fig. 10.) Since $\langle |a_{\ell m}|^2 \rangle$ is independent of m, we can define

$$C_\ell \equiv \langle |a_{\ell m}|^2 \rangle = \frac{1}{2\ell + 1} \sum_m \langle |a_{\ell m}|^2 \rangle , \tag{14}$$

and altogether we have

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell \,. \tag{15}$$

This function C_ℓ (of integers $\ell \geq 2$) is called the (theoretical) *angular power spectrum*. It is analogous to the power spectrum $\mathcal{P}(k)$ of density perturbations. For Gaussian perturbations, the C_ℓ contains all the statistical information about the CMB temperature anisotropy. And this is all we can predict from theory. Thus the analysis of the CMB anisotropy consists of calculating the angular power spectrum from the observed CMB (a map like Figure 5) and comparing it to the C_ℓ predicted by theory.[3]

[3] In addition to the temperature anisotropy, the CMB also has another property, its polarization. There are two additional power spectra related to the polarization, C_ℓ^{EE} and C_ℓ^{BB}, and one related to the correlation between temperature and polarization, C_ℓ^{TE}.

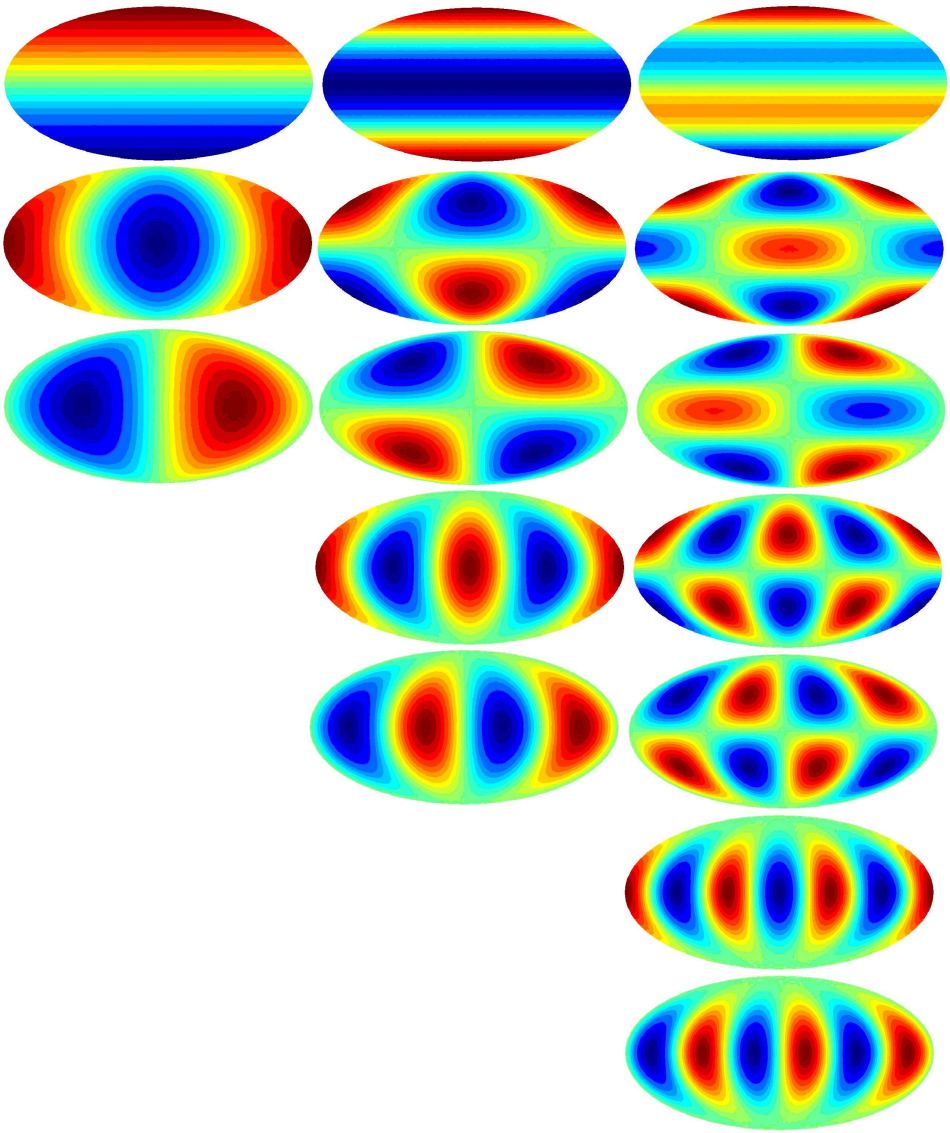


Figure 10: The three lowest multipoles $\ell = 1, 2, 3$ of spherical harmonics. Left column: Y_{10}, $\mathrm{Re}\, Y_{11}$, $\mathrm{Im}\, Y_{11}$. Middle column: Y_{20}, $\mathrm{Re}\, Y_{21}$, $\mathrm{Im}\, Y_{21}$, $\mathrm{Re}\, Y_{22}$, $\mathrm{Im}\, Y_{22}$. Right column: Y_{30}, $\mathrm{Re}\, Y_{31}$, $\mathrm{Im}\, Y_{31}$, $\mathrm{Re}\, Y_{32}$, $\mathrm{Im}\, Y_{32}$, $\mathrm{Re}\, Y_{33}$, $\mathrm{Im}\, Y_{33}$. Figure by Ville Heikkilä.

Just like the 3D density power spectrum $\mathcal{P}(k)$ gives the contribution of scale k to the density variance $\langle\delta(\mathbf{x})^2\rangle$, the angular power spectrum C_ℓ is related to the contribution of multipole ℓ to the temperature variance,

$$\begin{aligned}\left\langle\left(\frac{\delta T(\theta,\phi)}{T}\right)^2\right\rangle &= \left\langle\sum_{\ell m} a_{\ell m}Y_{\ell m}(\theta,\phi)\sum_{\ell' m'} a^*_{\ell' m'}Y^*_{\ell' m'}(\theta,\phi)\right\rangle \\ &= \sum_{\ell\ell'}\sum_{mm'} Y_{\ell m}(\theta,\phi)Y^*_{\ell' m'}(\theta,\phi)\langle a_{\ell m}a^*_{\ell' m'}\rangle \\ &= \sum_{\ell} C_\ell\sum_m |Y_{\ell m}(\theta,\phi)|^2 \;=\; \sum_\ell \frac{2\ell+1}{4\pi}C_\ell\,, \end{aligned} \tag{16}$$

where we used (15) and (9).

Thus, if we plot $(2\ell+1)C_\ell/4\pi$ on a linear ℓ scale, or $\ell(2\ell+1)C_\ell/4\pi$ on a logarithmic ℓ scale, the area under the curve gives the temperature variance, i.e., the expectation value for the squared deviation from the average temperature. It has become customary to plot the angular power spectrum as $\ell(\ell+1)C_\ell/2\pi$, which is neither of these, but for large ℓ approximates the second case. The reason for this custom is explained later.

Equation (16) represents the expectation value from theory and thus it is the same for all directions θ,ϕ. The actual, "realized", value of course varies from one direction θ,ϕ to another. We can imagine an ensemble of universes, otherwise like our own, but representing a different realization of the same random process of producing the primordial perturbations. Then $\langle\ \rangle$ represents the average over such an ensemble.

Equation(16) can be generalized to the angular correlation function (**exercise**)

$$C(\vartheta) \equiv \left\langle\frac{\delta T(\hat{\mathbf{n}})}{T}\frac{\delta T(\hat{\mathbf{n}}')}{T}\right\rangle \;=\; \frac{1}{4\pi}\sum_\ell (2\ell+1)C_\ell P_\ell(\cos\vartheta),, \tag{17}$$

where ϑ is the angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$.

9.2.3 Observed angular power spectrum

Theory predicts expectation values $\langle|a_{\ell m}|^2\rangle$ from the random process responsible for the CMB anisotropy, but we can observe only one realization of this random process, the set $\{a_{\ell m}\}$ of our CMB sky. We define the *observed* angular power spectrum as the average

$$\widehat{C}_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2 \tag{18}$$

of these observed values.

The variance of the observed temperature anisotropy is the average of $\left(\frac{\delta T(\theta,\phi)}{T}\right)^2$ over the celestial sphere,

$$\begin{aligned}\frac{1}{4\pi}\int\left[\frac{\delta T(\theta,\phi)}{T}\right]^2 d\Omega &= \frac{1}{4\pi}\int d\Omega\sum_{\ell m} a_{\ell m}Y_{\ell m}(\theta,\phi)\sum_{\ell' m'} a^*_{\ell' m'}Y^*_{\ell' m'}(\theta,\phi) \\ &= \frac{1}{4\pi}\sum_{\ell m}\sum_{\ell' m'} a_{\ell m}a^*_{\ell' m'}\underbrace{\int Y_{\ell m}(\theta,\phi)Y^*_{\ell' m'}(\theta,\phi)d\Omega}_{\delta_{\ell\ell'}\delta_{mm'}} \\ &= \frac{1}{4\pi}\sum_\ell\underbrace{\sum_m |a_{\ell m}|^2}_{(2\ell+1)\widehat{C}_\ell} \;=\; \sum_\ell\frac{2\ell+1}{4\pi}\widehat{C}_\ell\,. \end{aligned} \tag{19}$$

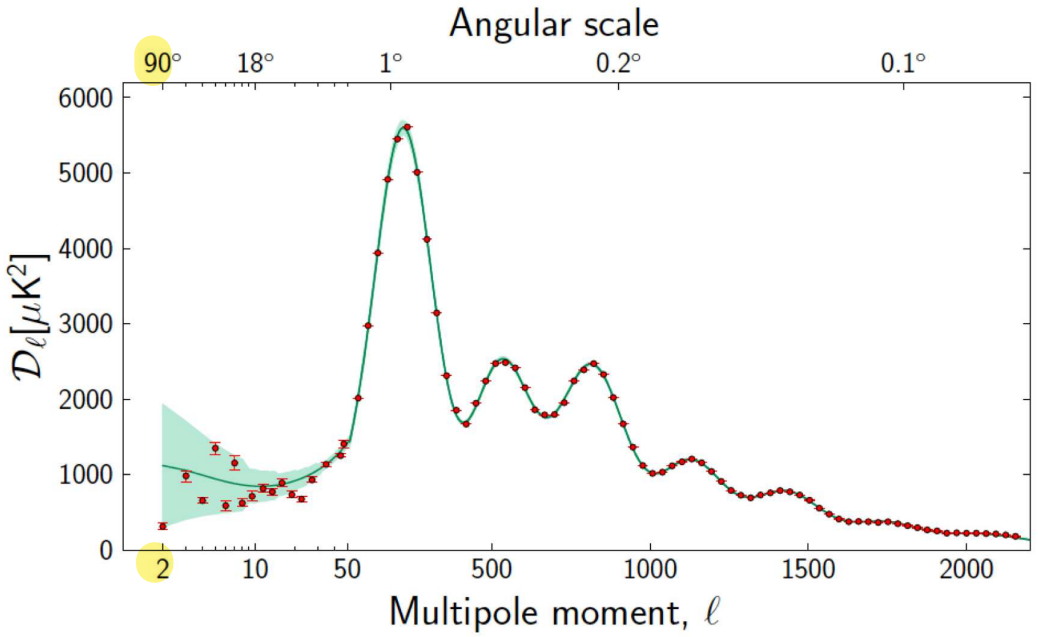


Figure 11: The angular power spectrum $\widehat{C}_\ell$ as observed by Planck. The observational results are the red data points with small error bars. The green curve is the theoretical C_ℓ from a best-fit model, and the light green band around it represents the cosmic variance corresponding to this C_ℓ. The quantity plotted is actually $\mathcal{D}_\ell \equiv T_0^2[\ell(\ell+1)/(2\pi)]C_\ell$. Note that the ℓ-axis is logarithmic until 50 and linear after that. (This is Fig. 21 of [1].)

Contrast this with (16), which gives the variance of $\delta T/T$ at an arbitrary location on the sky over different realizations of the random process which produced the primordial perturbations; whereas (19) gives the variance of $\delta T/T$ of our given sky over the celestial sphere.

9.2.4 Cosmic Variance

The expectation value of the observed spectrum $\widehat{C}_\ell$ is equal to C_ℓ, the *theoretical* spectrum of Eq. (14), i.e.,

$$\langle \widehat{C}_\ell \rangle = C_\ell \quad \Rightarrow \quad \langle \widehat{C}_\ell - C_\ell \rangle = 0\,, \tag{20}$$

but its actual, realized, value is not, although we expect it to be close. The expected squared difference between $\widehat{C}_\ell$ and C_ℓ is called the *cosmic variance* (of C_ℓ). We can calculate it using the properties of (complex) Gaussian random variables (**exercise**). The answer is

$$\langle (\widehat{C}_\ell - C_\ell)^2 \rangle = \frac{2}{2\ell+1} C_\ell^2\,. \tag{21}$$

We see that the expected relative difference between $\widehat{C}_\ell$ and C_ℓ is smaller for higher ℓ. This is because we have a larger (size $2\ell+1$) statistical sample of $a_{\ell m}$ available for calculating the $\widehat{C}_\ell$.

The cosmic variance limits the accuracy of comparison of CMB observations with theory, especially for large scales (low ℓ). See Fig. 11.

9.3 Multipoles and scales

9.3.1 Rough correspondence

The different multipole numbers ℓ correspond to different angular scales, low ℓ to large scales and high ℓ to small scales. Examination of the functions $Y_{\ell m}(\theta, \phi)$ reveals that they have an oscillatory pattern on the sphere, so that there are typically ℓ "wavelengths" of oscillation around a full great circle of the sphere. See Figs. 10 and 12.

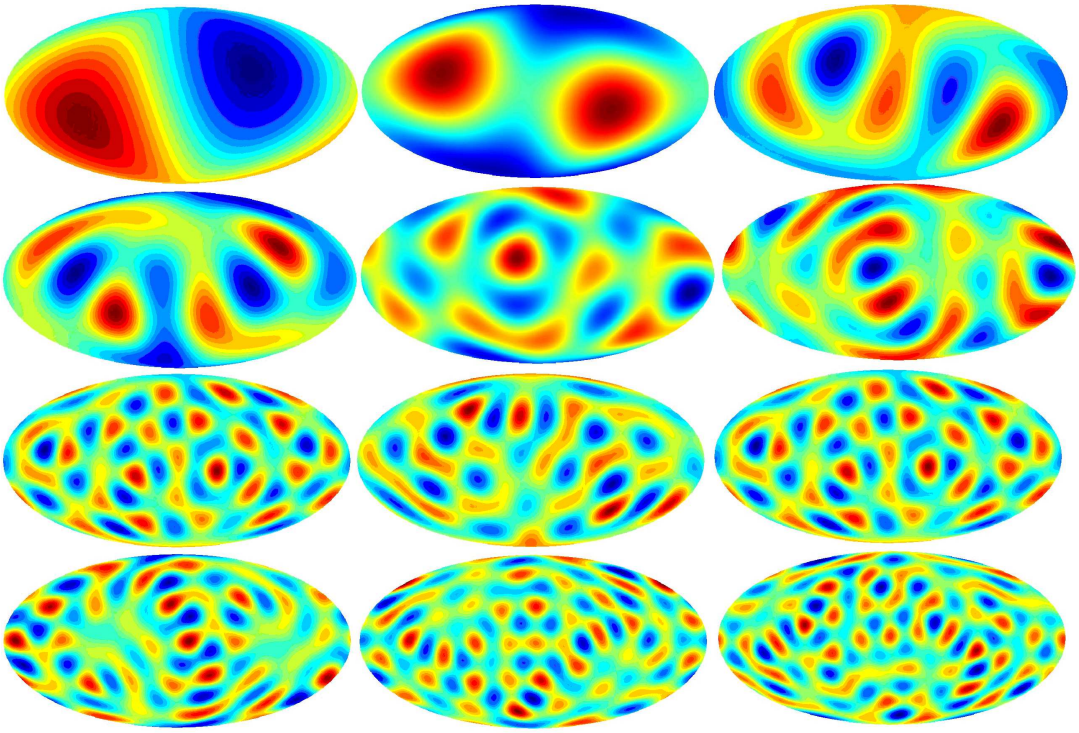


Figure 12: Randomly generated skies containing only a single multipole ℓ. Starting from top left: $\ell = 1$ (dipole only), 2 (quadrupole only), 3 (octupole only), 4, 5, 6, 7, 8, 9, 10, 11, 12. Figure by Ville Heikkilä.

Thus the angle corresponding to this wavelength is

$$\vartheta_\lambda = \frac{2\pi}{\ell} = \frac{360^\circ}{\ell}\,. \tag{22}$$

See Fig. 13. The angle corresponding to a "half-wavelength", i.e., the separation between a neighboring minimum and maximum is then

$$\vartheta_{\rm res} = \frac{\pi}{\ell} = \frac{180^\circ}{\ell}\,. \tag{23}$$

This is the angular resolution required of the microwave detector for it to be able to resolve the angular power spectrum up to this ℓ.

For example, COBE had an angular resolution of 7° allowing a measurement up to $\ell = 180/7 = 26$, WMAP had resolution 0.23° reaching to $\ell = 180/0.23 = 783$, and Planck had resolution $5'$, allowing the measurement of C_ℓ up to $\ell = 2160$.[4]

[4] In reality, there is no sharp cut-off at a particular ℓ, the observational error bars just blow up rapidly around this value of ℓ.

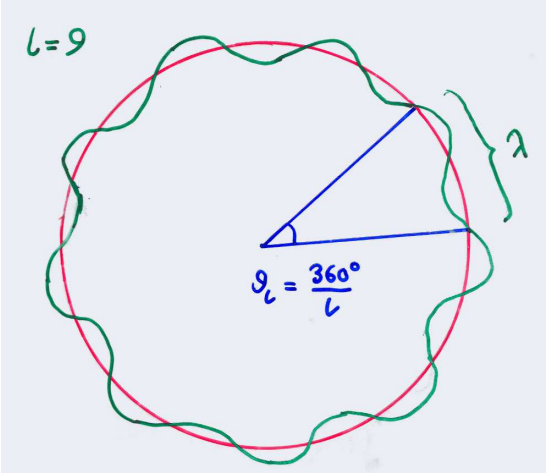


Figure 13: The rough correspondence between multipoles ℓ and angles.

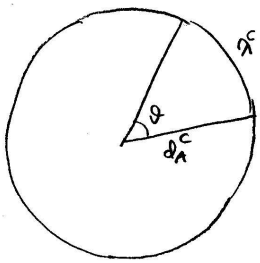


Figure 14: The comoving angular diameter distance relates the comoving size of an object and the angle in which we see it.

The angles on the sky are related to actual physical distances via the *angular diameter distance* d_A, defined as the ratio of the physical length (transverse to the line of sight) and the angle it covers (see Chapter 3),

$$d_A \equiv \frac{\lambda_{\rm phys}}{\vartheta}\,. \tag{24}$$

Likewise, we defined the *comoving angular diameter distance* d_A^c by

$$d_A^c \equiv \frac{\lambda^c}{\vartheta} \tag{25}$$

where $\lambda^c = a^{-1}\lambda_{\rm phys} = (1+z)\lambda_{\rm phys}$ is the corresponding comoving length. Thus $d_A^c = a^{-1}d_A = (1+z)d_A$. See Fig. 14.

Consider now the Fourier modes of our earlier perturbation theory discussion. A mode with comoving wavenumber k has comoving wavelength $\lambda^c = 2\pi/k$. Thus this mode should show up as a pattern on the CMB sky with angular size

$$\vartheta_\lambda = \frac{\lambda^c}{d_A^c} = \frac{2\pi}{kd_A^c} = \frac{2\pi}{\ell}\,. \tag{26}$$

For the last equality we used the relation (22). From it we get that the modes with wavenumber k contribute mostly to multipoles around

$$\ell = kd_A^c\,. \tag{27}$$

9.3.2 Exact treatment

The above matching of wavenumbers with multipoles was of course rather naive, for two reasons:

1. The description of a spherical harmonic $Y_{\ell m}$ having an "angular wavelength" of $2\pi/\ell$ is just a crude characterization. See Fig. 12.

2. The modes $\mathbf{k}$ are not wrapped around the sphere of last scattering, but the wave vector forms a different angle with the sphere at different places.

The following precise discussion applies only for the case of a flat universe ($K = 0$ FRW universe as the background), where one can Fourier expand functions on a time slice. We start from the expansion of the plane wave in terms of spherical harmonics, for which we have the result, Eq. (10),

$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kx) Y_{\ell m}(\hat{\mathbf{x}}) Y^*_{\ell m}(\hat{\mathbf{k}})\,, \tag{28}$$

where the j_ℓ are spherical Bessel functions.

Consider now some function

$$f(\mathbf{x}) = \sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \tag{29}$$

on the $t = t_{\rm dec}$ time slice. We want the multipole expansion of the values of this function on the last scattering sphere. See Fig. 15. These are the values $f(x\hat{\mathbf{x}})$, where $x \equiv |\mathbf{x}|$ has a constant value, the (comoving) radius of this sphere. Thus

$$\begin{aligned} a_{\ell m} &= \int d\Omega_x Y^*_{\ell m}(\hat{\mathbf{x}}) f(x\hat{\mathbf{x}}) \\ &= \sum_{\mathbf{k}} \int d\Omega_x Y^*_{\ell m}(\hat{\mathbf{x}}) f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \\ &= 4\pi \sum_{\mathbf{k}} \sum_{\ell' m'} \int d\Omega_x f_{\mathbf{k}} Y^*_{\ell m}(\hat{\mathbf{x}}) i^{\ell'} j_{\ell'}(kx) Y_{\ell' m'}(\hat{\mathbf{x}}) Y^*_{\ell' m'}(\hat{\mathbf{k}}) \\ &= 4\pi i^\ell \sum_{\mathbf{k}} f_{\mathbf{k}} j_\ell(kx) Y^*_{\ell m}(\hat{\mathbf{k}})\,, \end{aligned} \tag{30}$$

where we used the orthonormality of the spherical harmonics. The corresponding result for a Fourier transform $f(\mathbf{k})$ is

$$a_{\ell m} = \frac{4\pi i^\ell}{(2\pi)^3} \int d^3k f(\mathbf{k}) j_\ell(kx) Y^*_{\ell m}(\hat{\mathbf{k}})\,. \tag{31}$$

The j_ℓ are oscillating functions with decreasing amplitude. For large values of ℓ the position of the first (and largest) maximum is near $kx = \ell$ (see Fig. 16).

Thus the $a_{\ell m}$ pick a large contribution from those Fourier modes $\mathbf{k}$ where

$$kx \sim \ell\,. \tag{32}$$

In a flat universe the comoving distance x (from our location to the sphere of last scattering) and the comoving angular diameter distance d_A^c are equal, so we can write this result as

$$kd_A^c \sim \ell\,. \tag{33}$$

The conclusion is that a given multipole ℓ acquires a contribution from modes with a range of wavenumbers, but most of the contribution comes from near the value given by Eq. (27). This concentration is tighter for larger ℓ.

We shall use Eq. (27) for qualitative purposes in the following discussion.

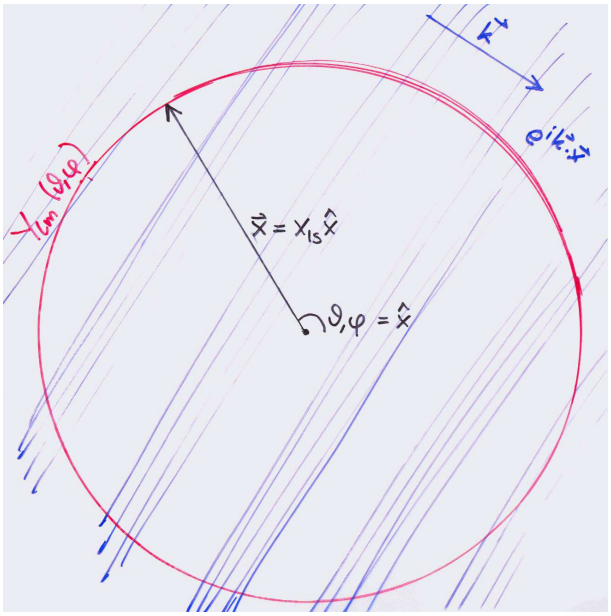


Figure 15: A plane wave intersecting the last scattering sphere.

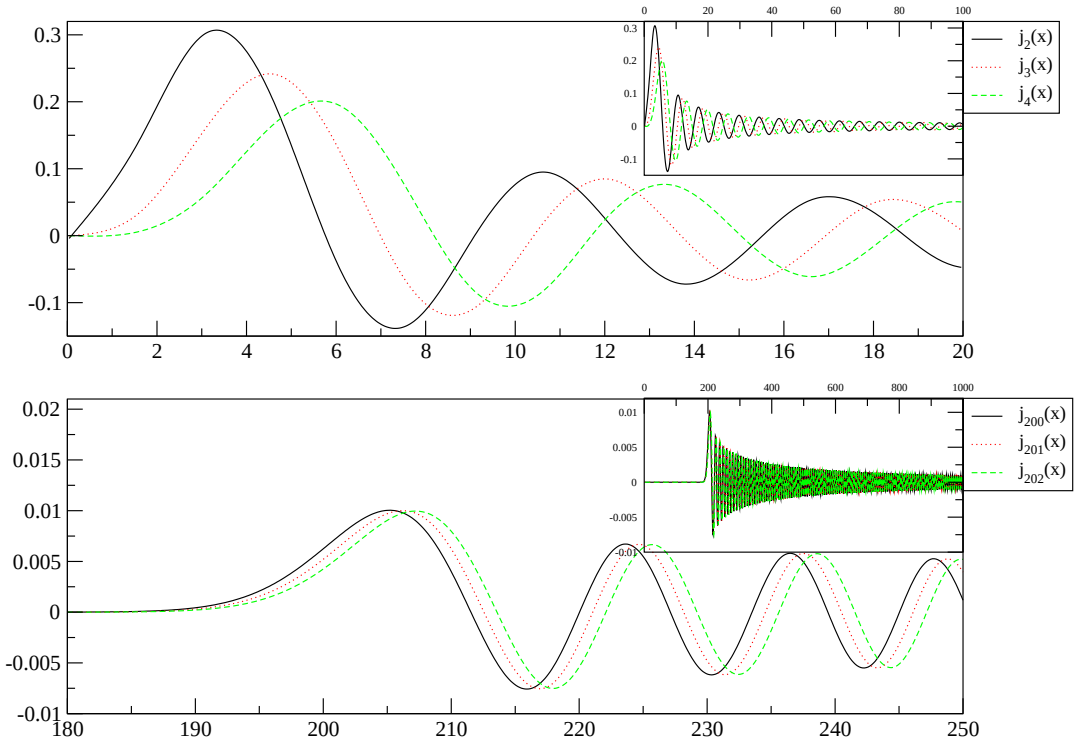


Figure 16: Spherical Bessel functions $j_\ell(x)$ for $\ell = 2$, 3, 4, 200, 201, and 202. Note how the first and largest peak is near $x = \ell$ (but to be precise, at a slightly larger value). Figure by R. Keskitalo.

9.4 Important distance scales on the last scattering surface

9.4.1 Angular diameter distance to last scattering

In Chapter 3 we derived the formula for the comoving distance to redshift z,

$$d^c(z) = H_0^{-1} \int_{\frac{1}{1+z}}^{1} \frac{da}{\sqrt{\Omega_0(a-a^2) - \Omega_\Lambda(a-a^4) + a^2}} \tag{34}$$

(where we have approximated $\Omega_0 \approx \Omega_m + \Omega_\Lambda$) and the corresponding comoving angular diameter distance

$$d_A^c(z) = f_K\left(d^c(z)\right) , \tag{35}$$

where

$$f_K(x) \equiv \begin{cases} K^{-1/2}\sin(K^{1/2}x) , & K > 0 \\ x , & K = 0 \\ |K|^{-1/2}\sinh(|K|^{1/2}x) , & K < 0 . \end{cases} \tag{36}$$

We also define

$$f_k(x) \equiv \begin{cases} \sin x , & k = 1 \\ x , & k = 0 \\ \sinh x , & k = -1 . \end{cases} \tag{37}$$

For the flat universe ($K = k = 0$, $\Omega_0 = 1$), the comoving angular diameter distance is equal to the comoving distance,

$$d_A^c(z) = d^c(z) \qquad (K = 0) . \tag{38}$$

For the open ($K < 0$, $\Omega_0 < 1$) and closed ($K > 0$, $\Omega_0 > 1$) cases we can write Eq. (35) as

$$\begin{aligned} d_A^c(z) &= \frac{H_0^{-1}}{\sqrt{|\Omega_k|}} f_k\left(\frac{\sqrt{|\Omega_k|}}{H_0^{-1}}\, d^c(z)\right) \\ &= H_0^{-1}\frac{1}{\sqrt{|\Omega_k|}} f_k\left(\sqrt{|\Omega_k|}\int_{\frac{1}{1+z}}^{1} \frac{da}{\sqrt{\Omega_0(a-a^2) - \Omega_\Lambda(a-a^4) + a^2}}\right) . \end{aligned} \tag{39}$$

Thus $d_A^c(z) \propto H_0^{-1}$, and has some more complicated dependence on Ω_0 and Ω_Λ (or on Ω_m and Ω_Λ).

We are now interested in the distance to the last scattering sphere, i.e., $d_A^c(z_{\rm dec})$, where $z_{\rm dec} \approx 1090$.

For the simplest case, $\Omega_\Lambda = 0$, $\Omega_m = 1$, the integral gives

$$d_A^c(z_{\rm dec}) = H_0^{-1}\int_{\frac{1}{1+z}}^{1} \frac{dx}{\sqrt{x}} = 2H_0^{-1}\left(1 - \frac{1}{\sqrt{1+z_{\rm dec}}}\right) = 1.94 H_0^{-1} \approx 2H_0^{-1} , \tag{40}$$

where the last approximation corresponds to ignoring the contribution from the lower limit.

We shall consider two more general cases, of which the above is a special case of both:

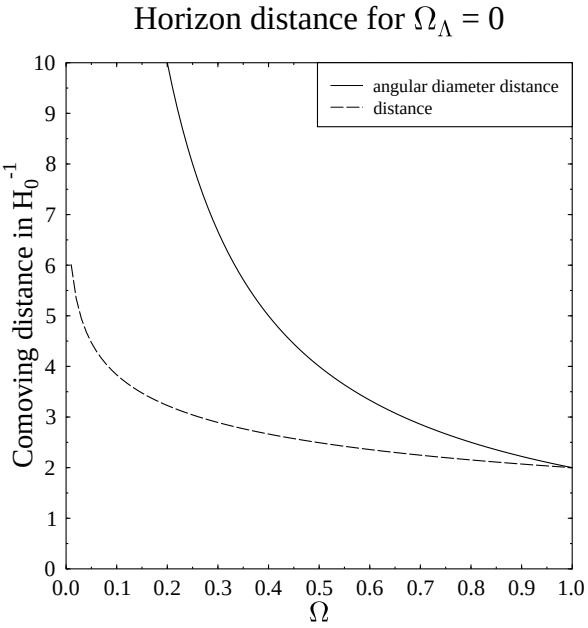


Figure 17: The comoving distance $d^c(z=\infty)$ (dashed) and the comoving angular diameter distance $d_A^c(z=\infty)$ (solid) to the horizon in matter-only open universe. The vertical axis is the distance in units of Hubble distance H_0^{-1} and the horizontal axis is the density parameter $\Omega_0 = \Omega_m$. The distances to last scattering, $d^c(z_{\rm dec})$ and $d_A^c(z_{\rm dec})$ are a few per cent less.

a) Open universe with no dark energy: $\Omega_\Lambda = 0$ and $\Omega_m = \Omega_0 < 1$. Now the integral gives

$$\begin{aligned} d_A^c(z_{\rm dec}) &= \frac{H_0^{-1}}{\sqrt{1-\Omega_m}} \sinh\left(\sqrt{1-\Omega_m}\int_{\frac{1}{1+z}}^1 \frac{dx}{\sqrt{(1-\Omega_m)x^2+\Omega_m x}}\right) \\ &= \frac{H_0^{-1}}{\sqrt{1-\Omega_m}} \sinh\left(\int_{\frac{1}{1+z}}^1 \frac{dx}{\sqrt{x^2+\frac{\Omega_m}{1-\Omega_m}x}}\right) \\ &= \frac{H_0^{-1}}{\sqrt{1-\Omega_m}} \sinh\left(2\,\mathrm{arsinh}\sqrt{\frac{1-\Omega_m}{\Omega_m}} - 2\,\mathrm{arsinh}\sqrt{\frac{1-\Omega_m}{\Omega_m}\frac{1}{1+z_{\rm dec}}}\right) \\ &\approx \frac{H_0^{-1}}{\sqrt{1-\Omega_m}} \sinh\left(2\,\mathrm{arsinh}\sqrt{\frac{1-\Omega_m}{\Omega_m}}\right) \;=\; 2\frac{H_0^{-1}}{\Omega_m}\,, \end{aligned} \tag{41}$$

where again the approximation ignores the contribution from the lower limit (i.e., it actually gives the angular diameter distance to the horizon, $d_A^c(z=\infty)$, in a model where we ignore the effect of other energy density components besides matter). In the last step we used $\sinh 2x = 2\sinh x\cosh x = 2\sinh x\sqrt{1+\sinh^2 x}$. We show this result (together with $d^c(z=\infty)$) in Fig. 17.

b) Flat universe with vacuum energy, $\Omega_\Lambda + \Omega_m = 1$. Here the integral does not give an elementary function, but a reasonable approximation, which we shall use in the following, is

$$d_A^c(z_{\rm dec}) = d^c(z_{\rm dec}) \approx \frac{2}{\Omega_m^{0.4}} H_0^{-1}\,. \tag{42}$$

The comoving distance $d_c(z_{\rm dec})$ depends on the expansion history of the universe. The longer it takes for the universe to cool from $T_{\rm dec}$ to T_0 (i.e., to expand by the factor $1+z_{\rm dec}$), the longer

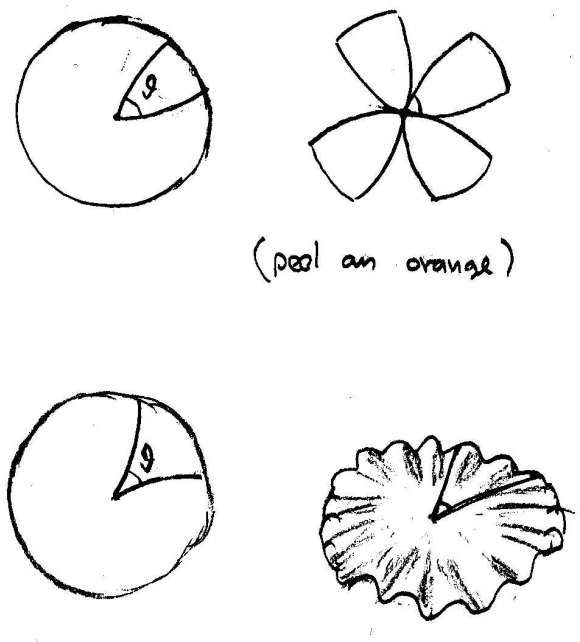


Figure 18: The geometry effect in a closed (top) or an open (bottom) universe affects the angle at which we see a structure of given size at the last scattering surface, and thus its angular diameter distance.

distance the photons have time to travel. When a larger part of this time is spent at small values of the scale factor, this distance gets a bigger boost from converting it to a comoving distance. For open/closed universes the angular diameter distance gets an additional effect from the geometry of the universe (the f_K), which acts like a "lens" to make the distant CMB pattern at the last scattering sphere to look smaller or larger (see Fig. 18).

9.4.2 Hubble scale and the matter-radiation equality scale

Subhorizon ($k \gg \mathcal{H}$) and superhorizon ($k \ll \mathcal{H}$) scales behave differently. Thus we want to know which of the structures we see on the last scattering surface are subhorizon and which are superhorizon. For that we need to know the comoving Hubble scale $\mathcal{H}$ at $t_{\rm dec}$. This was discussed in Sec. 8.3.1. At that time both matter and radiation are contributing to the energy density and the Hubble parameter. The scale which is just entering at $t = t_{\rm dec}$ is

$$\begin{aligned} k_{\rm dec}^{-1} \equiv \mathcal{H}_{\rm dec}^{-1} &= (1+z_{\rm dec})H_{\rm dec}^{-1} = (1+z_{\rm dec})^{-1/2} H_0^{-1}\Omega_m^{-1/2}\left[1+\frac{\Omega_r}{\Omega_m}(1+z_{\rm dec})\right]^{-1/2} \\ &= \Omega_m^{-1/2}(1+0.046\,\omega_m^{-1})^{-1/2}\,91\,h^{-1}{\rm Mpc} \end{aligned} \tag{43}$$

(using $z_{\rm dec} = 1090$; here $0.046\,\omega_m^{-1}$ is ρ_r/ρ_m at $t_{\rm dec}$) and the corresponding multipole number on the last scattering sphere is

$$\begin{aligned} \ell_H &\equiv k_{\rm dec} d_A^c \\ &= (1+z_{\rm dec})^{1/2}\Omega_m^{-1/2}\left[1+\frac{\Omega_r}{\Omega_m}(1+z_{\rm dec})\right]^{1/2} \times \begin{cases} 2/\Omega_m &= 66\,\Omega_m^{-0.5}\sqrt{1+0.046\omega_m^{-1}} \quad (\Omega_\Lambda = 0) \\ 2/\Omega_m^{0.4} &\approx 66\,\Omega_m^{0.1}\sqrt{1+0.046\omega_m^{-1}} \quad (\Omega_0 = 1) \end{cases} \end{aligned} \tag{44}$$

The angle subtended by a half-wavelength π/k of this mode on the last scattering sphere is

$$\vartheta_H \equiv \frac{\pi}{\ell_H} = \frac{180^\circ}{\ell_H} = \sqrt{1+0.046\,\omega_m^{-1}} \times \begin{cases} 2.7^\circ\Omega_m^{0.5} \\ 2.7^\circ\Omega_m^{-0.1} \end{cases} \tag{45}$$

For $\Omega_m \sim 0.3$, $\Omega_\Lambda \sim 0.7$, $h \sim 0.7$, $\ell_H \approx 67$ and $\vartheta_H \approx 3.5^\circ$ (the angle subtended by k^{-1} is 1.12°).

Another important scale is $k_{\rm eq}$, the scale which enters at the time of matter-radiation equality $t_{\rm eq}$, since the transfer function $T(k)$ is bent at that point. Perturbations for scales $k \ll k_{\rm eq}$ maintain essentially their primordial spectrum, whereas scales $k \gg k_{\rm eq}$ have lost relative power between their horizon entry and $t_{\rm eq}$. This scale is

$$k_{\rm eq}^{-1} \;=\; \mathcal{H}_{\rm eq}^{-1} \sim 13.7\,\Omega_m^{-1} h^{-2}\,{\rm Mpc} \;=\; 4.6\times 10^{-3}\,\Omega_m^{-1} h^{-1} H_0^{-1} \tag{46}$$

and the corresponding multipole number of these scales seen on the last scattering sphere is

$$l_{\rm eq} \;=\; k_{\rm eq} d_A^c \;=\; 219\,\Omega_m h \times \begin{cases} 2/\Omega_m & = \; 440\,h & (\Omega_\Lambda = 0) \\ 2/\Omega_m^{0.4} & \approx \; 440\,h\,\Omega_m^{0.6} & (\Omega_0 = 1) \end{cases} \tag{47}$$

Later we will introduce the *sound horizon* at photon decoupling, another important scale.

9.5 CMB anisotropy from perturbation theory

We began this chapter with the observation, Eq. (1), that the CMB temperature anisotropy is a sum of two parts,

$$\left(\frac{\delta T}{T}\right)_{\rm obs} = \left(\frac{\delta T}{T}\right)_{\rm intr} + \left(\frac{\delta T}{T}\right)_{\rm jour} , \tag{48}$$

and that this separation is gauge dependent. We shall consider this in the conformal-Newtonian gauge, since the second part, $\left(\frac{\delta T}{T}\right)_{\rm jour}$, the integrated redshift perturbation along the line of sight, is easiest to calculate in this gauge. (However, we won't do the calculation here.[5])

The result of this calculation is

$$\begin{aligned}\left(\frac{\delta T}{T}\right)_{\rm jour} &= -\int d\Phi + \int (\dot\Phi + \dot\Psi) dt + \mathbf{v}_{\rm obs}\cdot\hat{\mathbf{n}} \\ &= \Phi(t_{\rm dec},\mathbf{x}_{\rm ls}) - \Phi(t_0,\mathbf{0}) + \int (\dot\Phi + \dot\Psi) dt + \mathbf{v}_{\rm obs}\cdot\hat{\mathbf{n}} \\ &\overset{\Psi\approx\Phi}{\approx} \Phi(t_{\rm dec},\mathbf{x}_{\rm ls}) - \Phi(t_0,\mathbf{0}) + 2\int \dot\Phi dt + \mathbf{v}_{\rm obs}\cdot\hat{\mathbf{n}}\end{aligned} \tag{49}$$

where the integral is from $(t_{\rm dec},\mathbf{x}_{\rm ls})$ to $(t_0,\mathbf{0})$ along the path of the photon (a null geodesic), and $\dot{} \equiv \partial/\partial t$. The origin $\mathbf{0}$ is located where the observer is. The last term, $\mathbf{v}_{\rm obs}\cdot\hat{\mathbf{n}}$, is the Doppler effect from observer motion (assumed nonrelativistic), $\mathbf{v}_{\rm obs}$ being the observer velocity and $\hat{\mathbf{n}}$ the direction we are looking at. The $_{\rm ls}$ in $\mathbf{x}_{\rm ls}$ is just to remind us that $\mathbf{x}$ lies somewhere on the last scattering sphere. In the matter-dominated universe the Newtonian potential remains constant in time, $\dot\Phi = 0$, so we get a contribution from the integral only from epochs when radiation or dark energy contributions to the total energy density, or the effect of curvature, cannot be ignored. We can understand the above result as follows. If the potential is constant in time, the blueshift the photon acquires when falling into a potential well is canceled by the redshift from climbing up the well. Thus the net redshift/blueshift caused by gravitational potential perturbations is just the difference between the values of Φ at the beginning and in the end. However, if the potential is changing while the photon is traversing the well, this cancelation is not exact, and we get the integral term to account for this effect.

The value of the potential perturbation at the observing site, $\Phi(t_0,\mathbf{0})$ is the same for photons coming from all directions. Thus it does not contribute to the observed anisotropy. It just produces an overall shift in the observed average temperature. This is included in the observed value $T_0 = 2.7255 \pm 0.0006\,{\rm K}$, and so we just ignore this term, not attempting to separate it from the "correct" unperturbed value.[6] The observer motion $\mathbf{v}_{\rm obs}$ causes a dipole ($\ell = 1$) pattern

[5] It is done in my course on Cosmological Perturbation Theory, Sec. 25.

[6] I don't think the local value of the gravitational potential, $\Phi(t_0,\mathbf{0})$ is known very well. In a quick search I couldn't locate literature discussing it. We live within an overdensity (Solar System + Galaxy + Local Group + Local Supercluster) and thus in a potential well, so $\Phi(t_0,\mathbf{0})$ is negative, contributing a blueshift, and the true unperturbed value of T_0 is slightly lower than the observed value. The exact calculation of ω_γ from T_0 would require the use of this corrected value. While I don't have a good number it appears likely that the correction $\Phi(t_0,\mathbf{0})$ to T_0 is smaller than the uncertainty in the T_0 measurement. The order of magnitude of the Galactic contribution (which is much larger than the Solar contribution) is $\Phi \sim v^2$, where $v = 240\,{\rm km/s} = 8\times10^{-4}$ [4] is the local rotation velocity of the Galaxy. (For a spherically symmetric mass distribution, $\Phi = -v^2$, where v is the circular orbit velocity, is exact outside the mass distribution; if there is mass outside the orbit, it does not contribute to the gravitational field, but it does contribute to the potential. Most of the mass of the Galaxy is further away from the center than the Solar System, and therefore the Galactic contribution to the local potential $\Phi(t_0,\mathbf{0})$ is larger in absolute value than $-v^2 = -6.4\times10^{-7}$.) The large scale contribution could be estimated by cosmic flow velocities using linear perturbation theory, where velocity is related to the potential gradient; for the matter-dominated growing solution, $\mathbf{v} = -\frac{2}{3}\mathcal{H}^{-1}\nabla\Phi$. The motion of the Local Group in the CMB rest frame has $v = 620\,{\rm km/s} = 2.07\times10^{-3}$ [4], but we would need the relevant scale over which Φ varies to get an estimate of Φ from this. Anyway, this scale is much smaller than the Hubble scale, so the order of magnitude estimate of the linear contribution to $\Phi(t_0,\mathbf{0})$ is much less than this v.

in the CMB anisotropy, and likewise, we do not attempt to separate from it the cosmological dipole on the last scattering sphere. Therefore the dipole is usually removed from the CMB map before analyzing it for cosmological purposes. Accordingly, we shall ignore this term also, and our final result is

$$\left(\frac{\delta T}{T}\right)_{\rm jour} = \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) + 2\int \dot{\Phi} dt\,. \tag{50}$$

The other part, $\left(\frac{\delta T}{T}\right)_{\rm intr}$, comes from the local temperature perturbation at $t = t_{\rm dec}$ and the Doppler effect, $-\mathbf{v}\cdot\hat{\mathbf{n}}$, from the local (baryon+photon) fluid motion at that time. Since

$$\rho_\gamma = \frac{\pi^2}{15} T^4\,, \tag{51}$$

the local temperature perturbation is directly related to the relative perturbation in the photon energy density, and

$$\left(\frac{\delta T}{T}\right)_{\rm intr} = \frac{1}{4}\delta_\gamma - \mathbf{v}\cdot\hat{\mathbf{n}}\,. \tag{52}$$

We can now write the observed temperature anisotropy as

$$\boxed{\left(\frac{\delta T}{T}\right)_{\rm obs} = \frac{1}{4}\delta_\gamma^N - \mathbf{v}^N\cdot\hat{\mathbf{n}} + \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) + 2\int \dot{\Phi} dt\,.} \tag{53}$$

(note that both the density perturbation δ_γ and the fluid velocity $\mathbf{v}$ are gauge dependent).

To make further progress we now

1. consider adiabatic primordial perturbations only (like we did in Chapter 8), and
2. make the (crude) approximation that the universe is already matter dominated at $t = t_{\rm dec}$.

For adiabatic perturbations

$$\delta_b = \delta_c = \delta_m = \frac{3}{4}\delta_\gamma\,. \tag{54}$$

The perturbations stay adiabatic only at superhorizon scales. Once the perturbation has entered horizon, different physics begins to act on different matter components, so that the adiabatic relation between their density perturbations is broken. In particular, the baryon+photon perturbation is affected by photon pressure, which will damp their growth and cause them to oscillate, whereas the CDM perturbation is unaffected and keeps growing. Since the baryon and photon components see the same pressure they still evolve together and maintain their adiabatic relation until photon decoupling. Thus, after horizon entry, but before decoupling,

$$\delta_c \neq \delta_b = \frac{3}{4}\delta_\gamma\,. \tag{55}$$

At decoupling, the equality holds for scales larger than the photon mean free path at $t_{\rm dec}$.

After decoupling, this connection between the photons and baryons is broken, and the baryon density perturbation begins to approach the CDM density perturbation,

$$\delta_c \leftarrow \delta_b \not\equiv \frac{3}{4}\delta_\gamma\,. \tag{56}$$

We shall return to these issues as we discuss the shorter scales in Sections 9.7 and 9.8. But let us first discuss the scales which are still superhorizon at $t_{\rm dec}$, so that Eq. (54) still applies.

9.6 Large scales: Sachs–Wolfe part of the spectrum

Consider now the scales $k \ll k_{\rm dec}$, or $\ell \ll \ell_H$, which are still superhorizon at decoupling. We can now use the adiabatic condition (54), so that

$$\frac{1}{4}\delta_\gamma = \frac{1}{3}\delta_m \approx \frac{1}{3}\delta\,, \tag{57}$$

where the latter (approximate) equality comes from taking the universe to be matter dominated at $t_{\rm dec}$, so that we can identify $\delta \approx \delta_m$. For these scales the Doppler effect from fluid motion is subdominant, and we can ignore it (the fluid is set into motion by gradients in the pressure and gravitational potential, but the time scale of getting into motion is longer than the Hubble time for superhorizon scale gradients).

Thus Eq. (53) becomes

$$\left(\frac{\delta T}{T}\right)_{\rm obs} = \frac{1}{3}\delta^N + \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) + 2\int \dot\Phi dt\,. \tag{58}$$

The Newtonian relation

$$\delta = \frac{1}{4\pi G\bar\rho a^2}\nabla^2\Phi = \frac{2}{3}\left(\frac{1}{aH}\right)^2\nabla^2\Phi$$

(here ∇ is with respect to the comoving coordinates, hence the a^{-2}) or

$$\delta_{\mathbf{k}} = -\frac{2}{3}\left(\frac{k}{\mathcal{H}}\right)^2\Phi_{\mathbf{k}}$$

does not hold at superhorizon scales (where δ is gauge dependent). A GR calculation using the Newtonian gauge gives the result[7]

$$\boxed{\delta^N_{\mathbf{k}} = -\left[2 + \frac{2}{3}\left(\frac{k}{\mathcal{H}}\right)^2\right]\Phi_{\mathbf{k}}} \tag{59}$$

for perturbations in a matter-dominated universe. Thus for superhorizon scales we can approximate

$$\delta^N \approx -2\Phi \tag{60}$$

and Eq. (58) becomes

$$\begin{aligned}\left(\frac{\delta T}{T}\right)_{\rm obs} &= -\frac{2}{3}\Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) + \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) + 2\int \dot\Phi dt \\ &= \frac{1}{3}\Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) + 2\int \dot\Phi dt\,.\end{aligned} \tag{61}$$

This explains the "mysterious" factor 1/3 in this relation between the potential Φ and the temperature perturbation.

This result is called the *Sachs–Wolfe effect*. The first part, $\frac{1}{3}\Phi(t_{\rm dec}, \mathbf{x}_{\rm ls})$, is called the *ordinary* Sachs–Wolfe effect, and the second part, $2\int\dot\Phi dt$, the *integrated Sachs-Wolfe effect* (ISW), since it involves integrating along the line of sight. Note that the approximation of matter domination at $t = t_{\rm dec}$, making $\dot\Phi = 0$, does not eliminate the ISW, since it only applies to the "early part" of the integral. At times closer to t_0, dark energy becomes important, causing Φ to evolve again.

[7] Cosmological Perturbation Theory, Sec. 13.

This ISW caused by dark energy (or curvature of the background universe, if $k \neq 0$) is called the *late Sachs–Wolfe effect* (LSW) and it shows up as a rise in the smallest ℓ of the angular power spectrum C_ℓ. Correspondingly, the contribution to the ISW from the evolution of Φ near $t_{\rm dec}$ due to the radiation contribution to the expansion law (which we ignored in our approximation) is called the *early Sachs–Wolfe effect* (ESW). The ESW shows up as a rise in C_ℓ for larger ℓ, near ℓ_H.

We shall now forget for a while the ISW, which for $\ell \ll \ell_H$ is expected to be smaller than the ordinary Sachs–Wolfe effect.

9.6.1 Angular power spectrum from the ordinary Sachs–Wolfe effect

We now calculate the contribution from the ordinary Sachs–Wolfe effect,

$$\boxed{\left(\frac{\delta T}{T}\right)_{\rm SW} = \frac{1}{3}\Phi(t_{\rm dec}, \mathbf{x}_{\rm ls})\,,} \tag{62}$$

to the angular power spectrum C_ℓ. This is the dominant effect for $\ell \ll \ell_H$.

Since Φ is evaluated at the last scattering sphere, we have, from Eq. (30),

$$a_{\ell m} = 4\pi i^\ell \sum_{\mathbf{k}} \frac{1}{3}\Phi_{\mathbf{k}} j_\ell(kx) Y^*_{\ell m}(\hat{\mathbf{k}})\,, \tag{63}$$

In the matter-dominated epoch,

$$\Phi = -\frac{3}{5}\mathcal{R}\,, \tag{64}$$

so that

$$a_{\ell m} = -\frac{4\pi}{5} i^\ell \sum_{\mathbf{k}} \mathcal{R}_{\mathbf{k}} j_\ell(kx) Y^*_{\ell m}(\hat{\mathbf{k}})\,. \tag{65}$$

The coefficient $a_{\ell m}$ is thus a linear combination of the independent random variables $\mathcal{R}_{\mathbf{k}}$, i.e., it is of the form

$$\sum_{\mathbf{k}} b_{\mathbf{k}} \mathcal{R}_{\mathbf{k}}\,, \tag{66}$$

For any such linear combination, the expectation value of its absolute value squared is

$$\begin{aligned}\left\langle \left| \sum_{\mathbf{k}} b_{\mathbf{k}} \mathcal{R}_{\mathbf{k}} \right|^2 \right\rangle &= \sum_{\mathbf{k}} \sum_{\mathbf{k}'} b_{\mathbf{k}} b^*_{\mathbf{k}'} \langle \mathcal{R}_{\mathbf{k}} \mathcal{R}^*_{\mathbf{k}'} \rangle \\ &= \left(\frac{2\pi}{L}\right)^3 \sum_{\mathbf{k}} \frac{1}{4\pi k^3} \mathcal{P}_{\mathcal{R}}(k) \left|b_{\mathbf{k}}\right|^2\,,\end{aligned} \tag{67}$$

where we used

$$\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}^*_{\mathbf{k}'} \rangle = \delta_{\mathbf{k}\mathbf{k}'} \left(\frac{2\pi}{L}\right)^3 \frac{1}{4\pi k^3} \mathcal{P}_{\mathcal{R}}(k) \tag{68}$$

(the independence of the random variables $\mathcal{R}_{\mathbf{k}}$ and the definition of the power spectrum $\mathcal{P}(k)$).

Thus

$$\begin{aligned} C_\ell &\equiv \frac{1}{2\ell+1} \sum_m \langle |a_{\ell m}|^2 \rangle \\ &= \frac{16\pi^2}{25} \frac{1}{2\ell+1} \sum_m \left(\frac{2\pi}{L}\right)^3 \sum_{\mathbf{k}} \frac{1}{4\pi k^3} \mathcal{P}_{\mathcal{R}}(k) j_\ell(kx)^2 \left| Y^*_{\ell m}(\hat{\mathbf{k}}) \right|^2 \\ &= \frac{1}{25} \left(\frac{2\pi}{L}\right)^3 \sum_{\mathbf{k}} \frac{1}{k^3} \mathcal{P}_{\mathcal{R}}(k) j_\ell(kx)^2\,. \end{aligned} \tag{69}$$

(Although all $\langle|a_{\ell m}|^2\rangle$ are equal for the same ℓ, we used the sum over m, so that we could use Eq. (9).) Replacing the sum with an integral, we get

$$\begin{aligned} C_\ell &= \frac{1}{25}\int \frac{d^3k}{k^3}\mathcal{P}_\mathcal{R}(k) j_\ell(kx)^2 \\ &= \frac{4\pi}{25}\int_0^\infty \frac{dk}{k}\mathcal{P}_\mathcal{R}(k) j_\ell(kx)^2\,, \end{aligned} \tag{70}$$

where $x = d_A^c(z_{\rm dec})$, the final result for an arbitrary primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$.

The integral can be done for a power-law power spectrum, $\mathcal{P}_\mathcal{R}(k) = A_s^2(k/k_p)^{n-1}$. In particular, for a scale-invariant ($n = 1$) primordial power spectrum,

$$\mathcal{P}_\mathcal{R}(k) = \text{const.} = A_s^2\,, \tag{71}$$

we have

$$C_\ell = A_s^2\frac{4\pi}{25}\int_0^\infty \frac{dk}{k} j_\ell(kx)^2 = \frac{A_s^2}{25}\frac{2\pi}{\ell(\ell+1)}\,, \tag{72}$$

since

$$\int_0^\infty \frac{dk}{k} j_\ell(kx)^2 = \frac{1}{2\ell(\ell+1)}\,. \tag{73}$$

We can write this as

$$\frac{\ell(\ell+1)}{2\pi}C_\ell = \frac{A_s^2}{25} = \text{const. (independent of } \ell) \tag{74}$$

This is the reason why the angular power spectrum is customarily plotted as $\ell(\ell+1)C_\ell/2\pi$; it makes the ordinary Sachs–Wolfe part of the C_ℓ flat for a scale-invariant primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$.

Observations are consistent with an almost scale-invariant primordial power spectrum (they favor a small red tilt, $n < 1$). The constant A_s can be determined from the ordinary Sachs–Wolfe part of the observed $\widehat{C}_\ell$. From Fig. 11 we see that at low ℓ

$$\frac{\ell(\ell+1)}{2\pi}\widehat{C}_\ell \sim \frac{800\,\mu\text{K}^2}{(2.725\,K)^2} \sim 10^{-10} \tag{75}$$

on the average. This gives the amplitude of the primordial power spectrum as

$$\mathcal{P}_\mathcal{R}(k) = A_s^2 \sim 25\times 10^{-10} = \left(5\times 10^{-5}\right)^2\,. \tag{76}$$

We already used this result in Chapter 8 as a constraint on the energy scale of inflation.

Exercise: Find the C_ℓ of the ordinary Sachs-Wolfe effect due to a power-law power spectrum $\mathcal{P}_\mathcal{R}(k) = A_s^2(k/k_p)^{n-1}$. Help:

$$\int_0^\infty dx\, x^{n-2} j_\ell^2(x) = 2^{n-4}\pi\frac{\Gamma(\ell+\frac{n}{2}-\frac{1}{2})\Gamma(3-n)}{\Gamma(\ell+\frac{5}{2}-\frac{n}{2})\Gamma(2-\frac{n}{2})^2}\,. \tag{77}$$

Take $A_s = 4.58\times 10^{-5}$, for a pivot scale $k_p = 0.05\,\text{Mpc}^{-1}$, and $n = 0.965$ (Planck 2018 central values). Give the numerical values for C_2 and C_{20}. Use $d_A^c(z_{\rm dec}) = 2\Omega_m^{-0.4}H_0^{-1}$, with $\Omega_m = 0.315$ and $H_0 = 67.36\,\text{km/s/Mpc}$ (Planck 2018 central values). Give also $\mathcal{D}_2$ and $\mathcal{D}_{20}$, where $\mathcal{D}_\ell \equiv T_0^2[\ell(\ell+1)/(2\pi)]C_\ell$, and compare to Fig. 11. What explains the difference?

9.7 Acoustic oscillations

Consider now the scales $k \gg k_{\rm dec}$, or $\ell \gg \ell_H$, which are subhorizon at decoupling. The observed temperature anisotropy is, from Eq. (53)

$$\left(\frac{\delta T}{T}\right)_{\rm obs} = \frac{1}{4}\delta_\gamma(t_{\rm dec}, \mathbf{x}_{\rm ls}) + \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls}) - \mathbf{v}_\gamma \cdot \hat{\mathbf{n}}(t_{\rm dec}, \mathbf{x}_{\rm ls}) + 2\int \dot{\Phi} dt\,. \tag{78}$$

Since we are considering subhorizon scales, we dropped the reference to the Newtonian gauge. We shall concentrate on the three first terms, which correspond to the situation at the point $(t_{\rm dec}, \mathbf{x}_{\rm ls})$ we are looking at on the last scattering sphere.

Before decoupling photons are coupled to baryons. Perturbations in the baryon-photon fluid are oscillating, whereas CDM perturbations grow (slowly during the radiation-dominated epoch, and then faster during the matter-dominated epoch). Therefore CDM perturbations begin to dominate the total density perturbation $\delta\rho$ and thus also Φ already before the universe becomes matter dominated and CDM begins to dominate the background energy density. Thus we make the approximation that Φ is given by the CDM perturbation. The baryon-photon fluid oscillates in these potential wells caused by the CDM. The potential Φ evolves at first but then becomes constant as the universe becomes matter dominated.

We shall not attempt an exact calculation of the $\delta_{b\gamma}$ oscillations in the expanding universe. One reason is that $\rho_{b\gamma}$ is a relativistic fluid, and we have derived the perturbation equations for a nonrelativistic fluid only. From Sec. 8.2.7 we have that the nonrelativistic perturbation equation for a fluid component i is

$$\ddot{\delta}_{\mathbf{k}i} + 2H\dot{\delta}_{\mathbf{k}i} = -\frac{k^2}{a^2}\left(\frac{\delta p_{\mathbf{k}i}}{\bar{\rho}_i} + \Phi_{\mathbf{k}}\right). \tag{79}$$

The generalization of the (subhorizon) perturbation equations to the case of a relativistic fluid is considerably easier if we ignore the expansion of the universe. Then Eq. (79) becomes

$$\ddot{\delta}_{\mathbf{k}i} + k^2\left(\frac{\delta p_{\mathbf{k}i}}{\bar{\rho}_i} + \Phi_{\mathbf{k}}\right) = 0\,. \tag{80}$$

According to GR, the density of "passive gravitational mass" is $\rho + p = (1+w)\rho$, not just ρ as in Newtonian gravity. Therefore the force on a fluid element of the fluid component i is proportional to $(\rho_i + p_i)\nabla\Phi = (1+w_i)\rho_i\nabla\Phi$ instead of just $\rho_i\nabla\Phi$, and Eq. (80) generalizes to the case of a relativistic fluid as[8]

$$\ddot{\delta}_{\mathbf{k}i} + k^2\left[\frac{\delta p_{\mathbf{k}i}}{\bar{\rho}_i} + (1+w_i)\Phi_{\mathbf{k}}\right] = 0\,. \tag{81}$$

In the present application the fluid component ρ_i is the baryon-photon fluid $\rho_{b\gamma}$ and the gravitational potential Φ is caused by the CDM. Before decoupling, the adiabatic relation $\delta_b = \frac{3}{4}\delta_\gamma$ still holds between photons and baryons, and we have the adiabatic relation between pressure and density perturbations,

$$\delta p_{b\gamma} = c_s^2 \delta\rho_{b\gamma}\,. \tag{82}$$

Thus we have

$$\ddot{\delta}_{b\gamma\mathbf{k}} + k^2\left[c_s^2\delta_{b\gamma\mathbf{k}} + (1+w_{b\gamma})\Phi_{\mathbf{k}}\right] = 0\,. \tag{83}$$

Here

$$c_s^2 = \frac{\delta p_{b\gamma}}{\delta\rho_{b\gamma}} \approx \frac{\delta p_\gamma}{\delta\rho_{b\gamma}} = \frac{1}{3}\frac{\delta\rho_\gamma}{\delta\rho_\gamma + \delta\rho_b} = \frac{1}{3}\frac{\bar{\rho}_\gamma\delta_\gamma}{\bar{\rho}_\gamma\delta_\gamma + \bar{\rho}_b\delta_b} = \frac{1}{3}\frac{1}{1+\frac{3}{4}\frac{\bar{\rho}_b}{\bar{\rho}_\gamma}} \equiv \frac{1}{3}\frac{1}{1+R} \tag{84}$$

[8] Actually the derivation is more complicated, since also the density of "inertial mass" is $\rho_i + p_i$ and the energy continuity equation is modified by a work-done-by-pressure term. The more detailed derivation of Eq. (81) was given in Sec. 8.2.8.

gives the speed of sound c_s of the baryon-photon fluid. We defined

$$R \equiv \frac{3}{4}\frac{\bar{\rho}_b}{\bar{\rho}_\gamma}\,. \tag{85}$$

The equation-of-state parameter for the baryon-photon fluid is

$$w_{b\gamma} \equiv \frac{\bar{p}_{b\gamma}}{\bar{\rho}_{b\gamma}} = \frac{\frac{1}{3}\bar{\rho}_\gamma}{\bar{\rho}_\gamma + \bar{\rho}_b} = \frac{1}{3}\frac{1}{1+\frac{4}{3}R}\,, \tag{86}$$

so that

$$1 + w_{b\gamma} = \frac{\frac{4}{3}(1+R)}{1+\frac{4}{3}R} \tag{87}$$

and we can write Eq. (83) as

$$\ddot{\delta}_{b\gamma\mathbf{k}} + k^2\left[\frac{1}{3}\frac{1}{1+R}\delta_{b\gamma\mathbf{k}} + \frac{\frac{4}{3}(1+R)}{1+\frac{4}{3}R}\Phi_\mathbf{k}\right] = 0\,. \tag{88}$$

For the CMB anisotropy we are interested in[9]

$$\Theta_0 \equiv \frac{1}{4}\delta_\gamma\,, \tag{89}$$

which gives the local temperature perturbation, not in $\delta_{b\gamma}$. These two are related by

$$\delta_{b\gamma} = \frac{\delta\rho_{b\gamma}}{\bar{\rho}_{b\gamma}} = \frac{\delta\rho_\gamma + \delta\rho_b}{\bar{\rho}_\gamma + \bar{\rho}_b} = \frac{\bar{\rho}_\gamma\delta_\gamma + \bar{\rho}_b\delta_b}{\bar{\rho}_\gamma + \bar{\rho}_b} = \frac{1+R}{1+\frac{4}{3}R}\delta_\gamma\,. \tag{90}$$

Thus we can write Eq. (83) as (we are ignoring the expansion of the universe, so R is constant)

$$\ddot{\delta}_{\gamma\mathbf{k}} + k^2\left[\frac{1}{3}\frac{1}{1+R}\delta_{\gamma\mathbf{k}} + \frac{4}{3}\Phi_\mathbf{k}\right] = 0\,, \tag{91}$$

or

$$\ddot{\Theta}_{0\mathbf{k}} + k^2\left[\frac{1}{3}\frac{1}{1+R}\Theta_{0\mathbf{k}} + \frac{1}{3}\Phi_\mathbf{k}\right] = 0\,, \tag{92}$$

or

$$\ddot{\Theta}_{0\mathbf{k}} + c_s^2k^2\left[\Theta_{0\mathbf{k}} + (1+R)\Phi_\mathbf{k}\right] = 0\,, \tag{93}$$

If we now take R and Φ_k to be constant, this is the harmonic oscillator equation for the quantity $\Theta_{0\mathbf{k}} + (1+R)\Phi_\mathbf{k}$ with the general solution

$$\Theta_{0\mathbf{k}} + (1+R)\Phi_\mathbf{k} = A_\mathbf{k}\cos c_skt + B_\mathbf{k}\sin c_skt\,, \tag{94}$$

or

$$\Theta_{0\mathbf{k}} + \Phi_\mathbf{k} = -R\Phi_\mathbf{k} + A_\mathbf{k}\cos c_skt + B_\mathbf{k}\sin c_skt\,, \tag{95}$$

or

$$\Theta_{0\mathbf{k}} = -(1+R)\Phi_\mathbf{k} + A_\mathbf{k}\cos c_skt + B_\mathbf{k}\sin c_skt\,. \tag{96}$$

We are interested in the quantity $\Theta_0 + \Phi = \frac{1}{4}\delta_\gamma + \Phi$, called the *effective temperature perturbation*, since this combination appears in Eq. (78). It is the local temperature perturbation minus the redshift photons suffer when climbing from the potential well of the perturbation (negative Φ

[9]The subscript 0 refers to the monopole ($\ell = 0$) of the *local* photon distribution. Likewise, the dipole ($\ell = 1$) of the local photon distribution corresponds to the velocity of the photon fluid, $\Theta_1 \equiv v_\gamma/3$.

for a CDM overdensity). We see that this quantity oscillates in time, and the *effect of baryons* (via R) is to *shift the equilibrium point* of the oscillation by $-R\Phi_{\mathbf{k}}$.

In the preceding we ignored the effect of the expansion of the universe. The expansion affects the preceding in a number of ways. For example, c_s, $w_{b\gamma}$ and R change with time. The potential Φ also evolves, especially at the earlier times when radiation dominates the expansion law. However, the qualitative result of an oscillation of $\Theta_0 + \Phi$, and the shift of its equilibrium point by baryons, remains. The time t in the solution (95) gets replaced by conformal time η, and since c_s changes with time, $c_s\eta$ is replaced by

$$r_s(t) \equiv \int_0^\eta c_s d\eta = \int_0^t \frac{c_s(t)}{a(t)} dt\,. \tag{97}$$

We call this quantity $r_s(t)$ the *sound horizon* at time t, since it represents the comoving distance sound has traveled by time t.

The relative weight of the cosine and sine solutions (i.e., the constants $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ in Eq. (94) depends on the initial conditions. Since the perturbations are initially at superhorizon scales, the initial conditions are determined there, and the present discussion does not really apply. However, using the Newtonian gauge superhorizon initial conditions gives the correct qualitative result for the phase of the oscillation.

We had that for adiabatic primordial perturbations, initially $\Phi = -\frac{3}{5}\mathcal{R}$ and $\frac{1}{4}\delta_\gamma^N = -\frac{2}{3}\Phi = \frac{2}{5}\mathcal{R}$, giving us an initial condition $\Theta_0 + \Phi = \frac{1}{3}\Phi = -\frac{1}{5}\mathcal{R} = \text{const.}$ (At these early times $R \ll 1$, so we don't write the $1 + R$.) Thus adiabatic primordial perturbations correspond essentially to the cosine solution.[10] (There are effects at the horizon scale which affect the amplitude of the oscillations—the main effect being the decay of Φ as it enters the horizon—so we can't use the preceding discussion to determine the amplitude, but we get the right result about the initial phase of the $\Theta_0 + \Phi$ oscillations.)

Thus we have that, qualitatively, the effective temperature behaves at subhorizon scales as

$$\Theta_{0\mathbf{k}} + (1+R)\Phi_{\mathbf{k}} \;\propto\; \cos kr_s(t)\,, \tag{98}$$

Consider a region which corresponds to a positive primordial curvature perturbation $\mathcal{R}$. It begins with an initial overdensity (of all components, photons, baryons, CDM and neutrinos), and a negative gravitational potential Φ. For the scales of interest for CMB anisotropy, the potential stays negative, since the CDM begins to dominate the potential early enough and the CDM perturbations do not oscillate, they just grow. The effective temperature perturbation $\Theta_0 + \Phi$, which is the oscillating quantity, begins with a negative value. After half an oscillation period it is at its positive extreme value. This increase of $\Theta_0 + \Phi$ corresponds to an increase in δ_γ; from its initial positive value it has grown to a larger positive value. Thus the oscillation begins by the, already initially overdense, baryon-photon fluid falling deeper into the potential well, and reaching its maximum compression after half a period. After this maximum compression the photon pressure pushes the baryon-photon fluid out from the potential well, and after a full period, the fluid reaches its maximum decompression in the potential well. Since the potential Φ has meanwhile decayed (horizon entry and the resulting potential decay always happens during the first oscillation period, since the sound horizon and the Hubble length are close to each other, as the sound speed is close to the speed of light), the decompression does not bring the $\delta_{b\gamma}$ back to its initial value (which was overdense), but the photon-baryon fluid actually becomes underdense in the potential well (and overdense in the neighboring potential "hill"). And so the oscillation goes on until photon decoupling.

These are standing waves and they are called *acoustic oscillations*. See Fig. 19. Because of the potential decay at horizon entry, the amplitude of the oscillation is larger than Φ, and thus also Θ_0 changes sign in the oscillation.

[10] The sine solution corresponds to what are called *isocurvature* primordial perturbations.

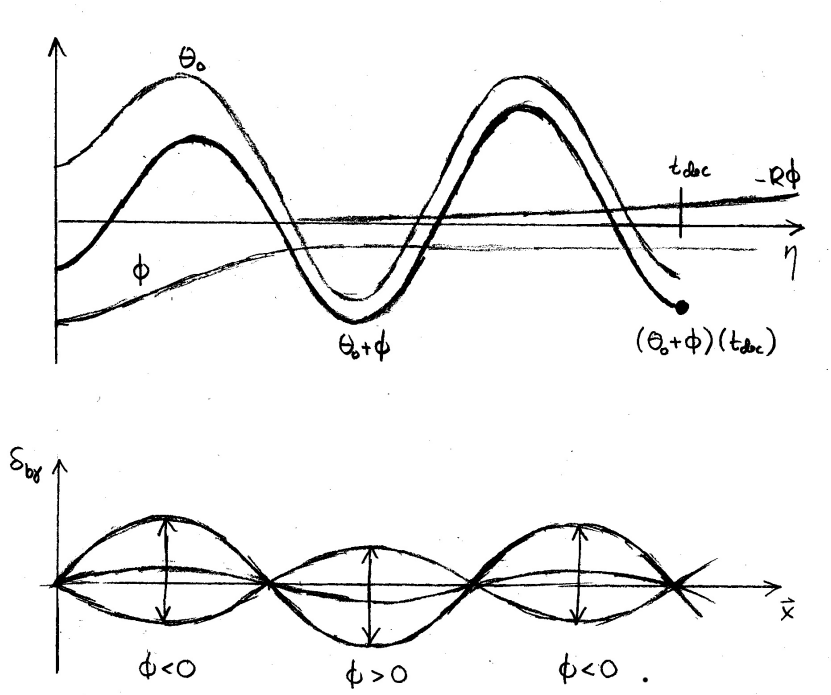


Figure 19: Acoustic oscillations. The top panel shows the time evolution of the Fourier amplitudes $\Theta_{0\mathbf{k}}$, $\Phi_{\mathbf{k}}$, and the effective temperature $\Theta_{0\mathbf{k}} + \Phi_{\mathbf{k}}$. The Fourier mode shown corresponds to the fourth acoustic peak of the C_ℓ spectrum. The bottom panel shows $\delta_{b\gamma}(\mathbf{x})$ for one Fourier mode as a function of position at various times (maximum compression, equilibrium level, and maximum decompression).

These oscillations end at photon decoupling, when the photons are liberated. The CMB shows these standing waves as a snapshot[11] at their final moment $t = t_{\rm dec}$.

At photon decoupling we have

$$\Theta_{0\mathbf{k}} + (1+R)\Phi_{\mathbf{k}} \propto \cos kr_s(t_{\rm dec})\,. \tag{99}$$

At this moment oscillations for scales k which have

$$kr_s(t_{\rm dec}) = m\pi \tag{100}$$

($m = 1, 2, 3, \ldots$) are at their extreme values (maximum compression or maximum decompression). Therefore we see strong structure in the CMB anisotropy at the multipoles

$$\ell = kd_A^c(t_{\rm dec}) = m\pi\frac{d_A^c(t_{\rm dec})}{r_s(t_{\rm dec})} \equiv m\ell_A \tag{101}$$

corresponding to these scales. Here

$$\ell_A \equiv \pi\frac{d_A^c(t_{\rm dec})}{r_s(t_{\rm dec})} \equiv \frac{\pi}{\vartheta_s} \tag{102}$$

is the *acoustic scale* in multipole space and

$$\vartheta_s \equiv \frac{r_s(t_{\rm dec})}{d_A^c(t_{\rm dec})} \tag{103}$$

[11] Actually, photon decoupling takes quite a long time. Therefore this "snapshot" has a rather long "exposure time" causing it to be "blurred". This prevents us from seeing very small scales in the CMB anisotropy.

is the *sound horizon angle*, i.e., the angle at which we see the sound horizon on the last scattering surface.

Because of these acoustic oscillations, the CMB angular power spectrum C_ℓ has a structure of *acoustic peaks* at subhorizon scales. The centers of these peaks are located approximately at $\ell_m \approx m\ell_A$. An exact calculation shows that they will actually lie at somewhat smaller ℓ due to a number of effects. The separation of neighboring peaks is closer to ℓ_A than the positions of the peaks are to $m\ell_A$.

These acoustic oscillations involve motion of the baryon-photon fluid. When the oscillation of one Fourier mode is at its extreme, e.g., at the maximal compression in the potential well, the fluid is momentarily at rest, but then it begins flowing out of the well until the other extreme, the maximal decompression, is reached. Therefore those Fourier modes $\mathbf{k}$ which have the maximum effect on the CMB anisotropy via the $\frac{1}{4}\delta_\gamma(t_{\rm dec}, \mathbf{x}_{\rm ls}) + \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls})$ term (the effective temperature effect) in Eq. (78) have the minimum effect via the $-\mathbf{v}\cdot\hat{\mathbf{n}}(t_{\rm dec}, \mathbf{x}_{\rm ls})$ term (the Doppler effect) and vice versa. Therefore the Doppler effect also contributes a peak structure to the C_ℓ spectrum, but the peaks are in the locations where the effective temperature contribution has troughs.

The Doppler effect is subdominant to the effective temperature effect, and therefore the peak positions in the C_ℓ spectrum are determined by the effective temperature effect, according to Eq. (101). The Doppler effect just partially fills the troughs between the peaks, weakening the peak structure of C_ℓ. See Fig. 22.

Fig. 20 shows the values of the effective temperature perturbation $\Theta_0 + \Phi$ (as well as Θ_0 and Φ separately) and the magnitude of the velocity perturbation ($\Theta_1 \sim v/3$) at $t_{\rm dec}$ as a function of the scale k. This is a result of a numerical calculation which includes the effect of the expansion of the universe, but not diffusion damping (Sec. 9.8).

9.8 Diffusion damping

For small enough scales the effect of photon diffusion and the finite thickness of the last scattering surface ($\sim$ the photon mean free path just before last scattering) smooth out the photon distribution and the CMB anisotropy.

This effect can be characterized by the damping scale $k_D^{-1} \sim$ photon diffusion length $\sim$ geometric mean of the Hubble time and photon mean free path λ_γ. Actually λ_γ is increasing rapidly during recombination, so a calculation of the diffusion scale involves an integral over time which includes this effect.

A calculation, that we shall not do here,[12] gives that photon density and velocity perturbations at scale k are damped at $t_{\rm dec}$ by

$$e^{-k^2/k_D^2}\,, \tag{104}$$

where the diffusion scale is

$$k_D^{-1} \sim \frac{1}{\text{few}}\frac{1}{a}\sqrt{\frac{\lambda_\gamma(t_{\rm dec})}{H_{\rm dec}}}\,. \tag{105}$$

Accordingly, the C_ℓ spectrum is also damped as

$$e^{-\ell^2/\ell_D^2} \tag{106}$$

where

$$\ell_D \sim k_D d_A^c(t_{\rm dec})\,. \tag{107}$$

For typical values of cosmological parameters $\ell_D \sim 1500$. See Fig. 21 for a result of a numerical calculation with and without diffusion damping.

[12] See, e.g., Dodelson [9], Chapter 8.

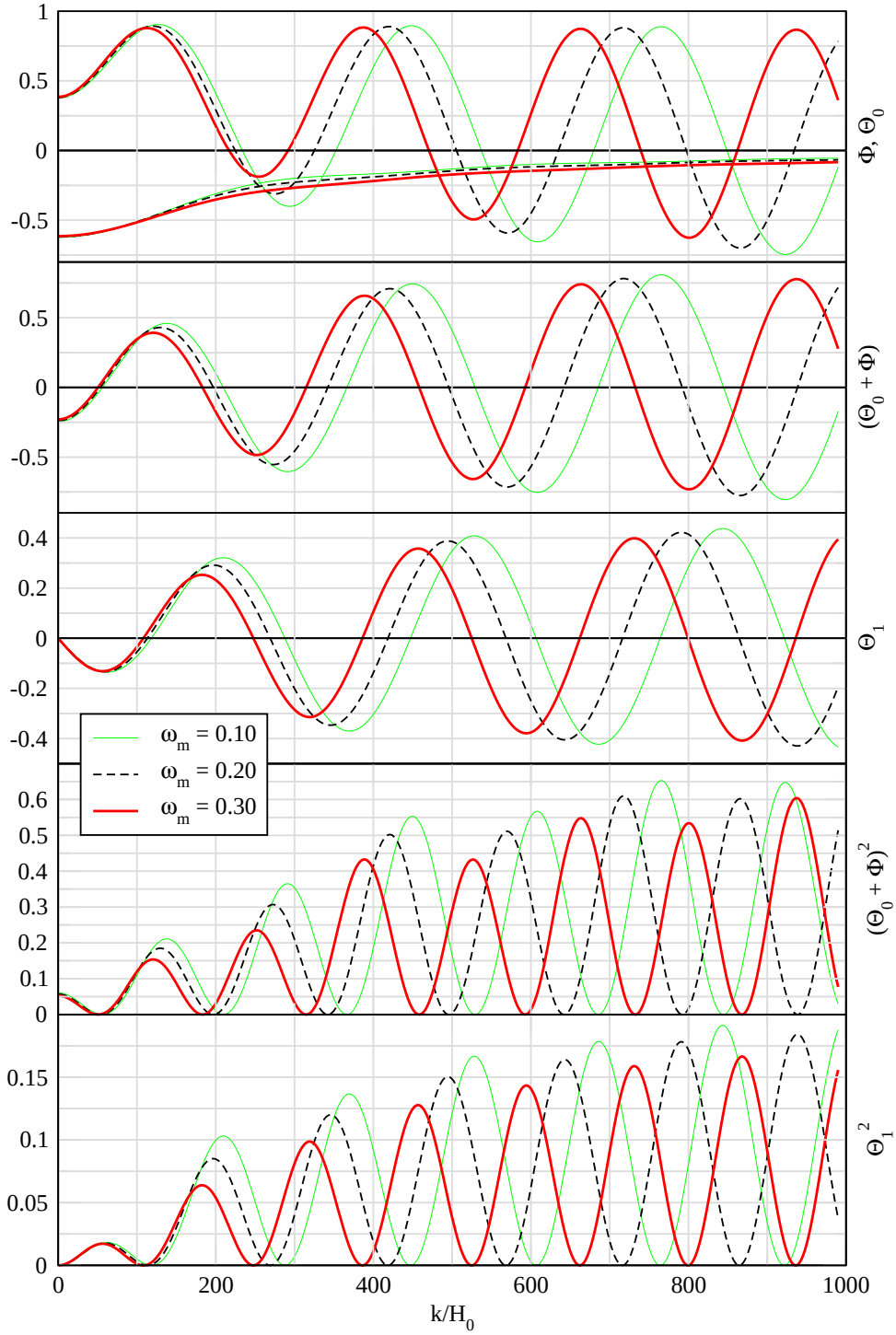


Figure 20: Values of oscillating quantities (normalized to an initial value $\mathcal{R}_{\mathbf{k}} = 1$) at the time of decoupling as a function of the scale k, for three different values of ω_m, and for $\omega_b = 0.01$. Θ_1 represents the velocity perturbation. The effect of diffusion damping is neglected. Figure and calculation by R. Keskitalo.

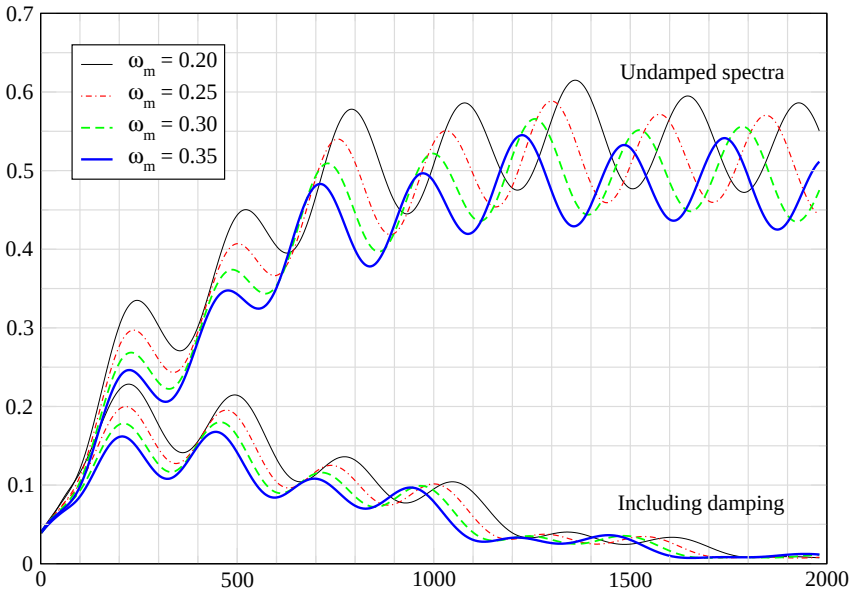


Figure 21: The angular power spectrum C_ℓ, calculated both with and without the effect of diffusion damping. The spectrum is given for four different values of ω_m, with $\omega_b = 0.01$. (This is a rather low value of ω_b, so $\ell_D < 1500$ and damping is quite strong.) Figure and calculation by R. Keskitalo.

Of the cosmological parameters, the damping scale is the most strongly dependent on ω_b, since increasing the baryon density shortens the photon mean free path before decoupling. Thus for larger ω_b the damping moves to shorter scales, i.e., ℓ_D becomes higher (there is less damping).

(Of course, decoupling only happens as the photon mean free path becomes comparable to the Hubble length, so one might think that λ_γ at $t_{\rm dec}$ should be independent of ω_b. However there is a distinction here between whether a photon will not scatter again after a particular scattering and what was the mean free path between the second-to-last and the last scattering. And k_D depends on an integral over the past history of the photon mean free path, not just the last one. The factor 1/few in Eq. (105) comes from that integration, and actually depends on ω_b. For small ω_b the λ_γ has already become quite large through the slow dilution of the baryon density by the expansion of the universe, and relies less on the fast reduction of free electron density due to recombination. Thus the time evolution of λ_γ before decoupling is different for different ω_b and we get a different diffusion scale.)

9.9 The complete C_ℓ spectrum

As we have discussed the CMB anisotropy has three contributions (see Eq. 78), the effective temperature effect,

$$\frac{1}{4}\delta_\gamma(t_{\rm dec}, \mathbf{x}_{\rm ls}) + \Phi(t_{\rm dec}, \mathbf{x}_{\rm ls})\,, \tag{108}$$

the Doppler effect,

$$-\mathbf{v}\cdot\hat{\mathbf{n}}(t_{\rm dec}, \mathbf{x}_{\rm ls})\,, \tag{109}$$

and the integrated Sachs–Wolfe effect,

$$2\int_{t_{\rm dec}}^{t_0} \dot{\Phi}(t, \mathbf{x}(t))dt\,. \tag{110}$$

Since the C_ℓ is a quadratic quantity, it also includes cross terms between these three effects.

The calculation of the full C_ℓ proceeds much as the calculation of just the ordinary Sachs–Wolfe part (which the effective temperature effect becomes at superhorizon scales) in Sec. 9.6.1, but now with the full $\delta T/T$. Since all perturbations are proportional to the primordial perturbations, the C_ℓ spectrum is proportional to the primordial perturbation spectrum $\mathcal{P}_\mathcal{R}(k)$ (with integrals over the spherical Bessel functions $j_\ell(kx)$, like in Eq. (70), to get from k to ℓ).

The difference is that instead of the constant proportionality factor $(\delta T/T)_{SW} = -(1/5)\mathcal{R}$, we have a k-dependent proportionality resulting from the evolution (including, e.g., the acoustic oscillations) of the perturbations.

In Fig. 22 we show the full C_ℓ spectrum and the different contributions to it.

Because the Doppler effect and the effective temperature effect are almost completely off-phase, their cross term gives a negligible contribution.

Since the ISW effect is relatively weak, it contributes more via its cross terms with the Doppler effect and effective temperature than directly. The cosmological model used for Fig. 22 has $\Omega_\Lambda = 0$, so there is no late ISW effect (which would contribute at the very lowest ℓ), and the ISW effect shown is the early ISW effect due to radiation contribution to the expansion law. This effect contributes mainly to the first peak and to the left of it, explaining why the first peak is so much higher than the other peaks. It also shifts the first peak position slightly to the left and changes its shape.

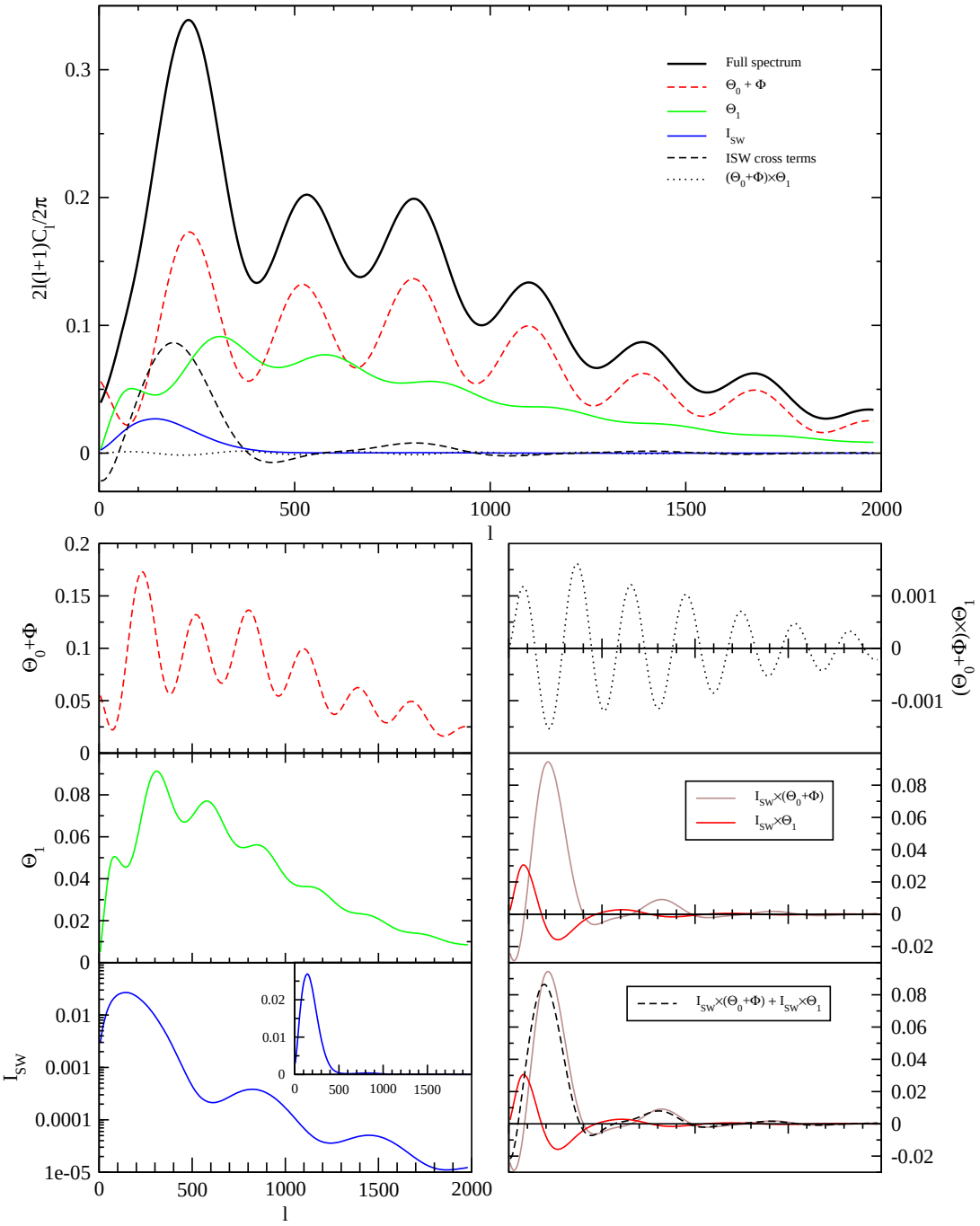


Figure 22: The full C_ℓ spectrum calculated for the cosmological model $\Omega_0 = 1$, $\Omega_\Lambda = 0$, $\omega_m = 0.2$, $\omega_b = 0.03$, $A_s = 1$, $n_s = 1$, and the different contributions to it. (The calculation involves some approximations which allow the description of C_ℓ as just a sum of these contributions and is not as accurate as a CMBFAST or CAMB calculation.) Here Θ_1 denotes the Doppler effect. Figure and calculation by R. Keskitalo.

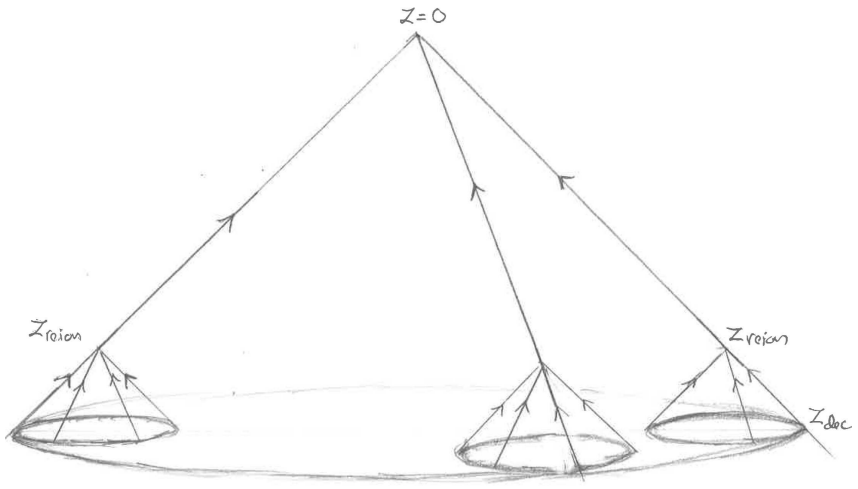


Figure 23: Spacetime diagram of the rescattering of CMB photons due to free electrons released in reionization. The rescattering continues after reionization, but most of it happens relatively soon after it, since the n_e is diluted by expansion.

9.10 Reionization and optical depth

When radiation from the first stars reionizes the intergalactic gas, CMB photons may scatter from the resulting free electrons. The optical depth τ due to reionization is the expectation number of such scatterings per CMB photon. It is expected to be less than 0.1, i.e., most CMB photons do not scatter at all. This rescattering causes additional polarization[13] of the CMB, and CMB polarization measurements are actually the best way to determine τ. Most of this scattering happens relatively soon after the reionization, since the number density of free electrons is diluted by the expansion of the Universe.

The optical depth is thus directly related to the reionization redshift $z_{\rm reion}$. A smaller τ corresponds to later reionization and thus means that the first stars formed later.

Because of this scattering, not all the CMB photons come from the location on the last scattering surface they seem to come from. The effect of the rescattered photons is to mix up signals from different directions and therefore to reduce the CMB anisotropy. The reduction factor on $\delta T/T$ is $e^{-\tau}$ and on the C_ℓ spectrum $e^{-2\tau}$. However, this does not affect the largest scales, scales larger than the area from which the rescattered photons, reaching us from a certain direction, originally came from. Such a large-scale anisotropy has affected all such photons the same way, and thus is not lost in the mixing. See Fig. 23.

[13] Due to time constraints, CMB polarization is not discussed in these lectures.

9.11 Cosmological parameters and CMB anisotropy

Let us finally consider the total effect of the various cosmological parameters on the C_ℓ spectrum. The C_ℓ provides the most important single observational data set for determining (or constraining) cosmological parameters, since it has a rich structure which we can measure with an accuracy that other cosmological observations cannot match, and because it depends on so many different cosmological parameters in many ways. The latter is both a strength and weakness: the number of cosmological parameters we can determine is large, but on the other hand, some feature in C_ℓ may depend on more than one parameter, so that we may only be able to constrain some combination of such parameters, not the parameters individually. We say that such parameters are *degenerate* in the CMB data. Other cosmological observations are then needed to break such degeneracies.

We shall consider 7 "standard parameters":

- Ω_0 total density parameter
- Ω_Λ cosmological constant (or vacuum energy) density parameter
- A_s amplitude of primordial scalar perturbations (at some pivot scale k_p)
- n_s spectral index of primordial scalar perturbations
- τ optical depth due to reionization (discussed in Sec. 9.11.6)
- $\omega_b \equiv \Omega_b h^2$ "physical" baryon density parameter
- $\omega_m \equiv \Omega_m h^2$ "physical" matter density parameter

There are other possible cosmological parameters ("additional parameters") which might affect the C_ℓ spectrum, e.g.,

- m_{ν_i} neutrino masses
- w dark energy equation-of-state parameter
- $\frac{dn_s}{d\ln k}$ scale dependence of the spectral index
- r, n_T relative amplitude and spectral index of tensor perturbations
- B, $n_{\rm iso}$ amplitudes and spectral indices of primordial isocurvature perturbations
- $A_{\rm cor}$, $n_{\rm cor}$ and their correlation with primordial curvature perturbations

We assume here that these additional parameters have no impact, i.e., they have the "standard" values

$$r = \frac{dn}{d\ln k} = B = A_{\rm cor} = 0\,, \quad w = -1\,, \quad \text{and} \quad \sum m_{\nu_i} = 0.06\,\text{eV}\,, \tag{111}$$

to the accuracy which matters for C_ℓ observations. This is both observationally and theoretically reasonable. There is no sign in the present-day CMB data for deviations from these values. On the other hand, significant deviations can be consistent with the current data, and may be discovered by more accurate future observations. The primordial isocurvature perturbations refer to the possibility that the primordial scalar perturbations are not adiabatic, and therefore are not completely determined by the comoving curvature perturbation $\mathcal{R}$.

The assumption that these additional parameters have no impact, leads to a determination of the standard parameters with an accuracy that may be too optimistic, since the standard parameters may have degeneracies with the additional parameters.

9.11.1 Independent vs. dependent parameters

The above is our choice of independent cosmological parameters. Ω_m, Ω_b and H_0 (or h) are then dependent (or "derived") parameters, since they are determined by

$$\Omega_0 = \Omega_m + \Omega_\Lambda \quad \Rightarrow \quad \Omega_m = \Omega_0 - \Omega_\Lambda \tag{112}$$

$$h = \sqrt{\frac{\omega_m}{\Omega_m}} = \sqrt{\frac{\omega_m}{\Omega_0 - \Omega_\Lambda}} \tag{113}$$

$$\Omega_b = \frac{\omega_b}{h^2} = \frac{\omega_b}{\omega_m}(\Omega_0 - \Omega_\Lambda) \tag{114}$$

Note that the Hubble constant $H_0 \equiv h \cdot 100$ km/s/Mpc is now a dependent parameter! We cannot vary it independently, but rather the varying of ω_m, Ω_0, or Ω_Λ also causes H_0 to change.

Different choices of independent parameters are possible within our 7-dimensional parameter space (e.g., we could have chosen H_0 to be an independent parameter and let Ω_Λ to be a dependent parameter instead). They can be thought of as different coordinate systems[14] in this 7D space. *It is not meaningful to discuss the effect of one parameter without specifying what is your set of independent parameters!*

Some choices of independent parameters are better than others. The above choice represents standard practice in cosmology today.[15] The independent parameters have been chosen so that they correspond as directly as possible to physics affecting the C_ℓ spectrum and thus to observable features in it. We want the effects of our independent parameters on the observables to be as different ("orthogonal") as possible in order to avoid parameter degeneracy.

In particular,

- ω_m (not Ω_m) determines $z_{\rm eq}$ and $k_{\rm eq}$, and thus, e.g., the magnitude of the early ISW effect and which scales enter during matter- or radiation-dominated epochs.
- ω_b (not Ω_b) determines the baryon/photon ratio and thus, e.g., the relative heights of the odd and even peaks.
- Ω_Λ (not $\Omega_\Lambda h^2$) determines the late ISW effect.

There are many effects on the C_ℓ spectrum, and parameters act on them in different combinations. Thus there is no perfectly "clean" way of choosing independent parameters. Especially having the Hubble constant as a dependent parameter takes some getting used to.

In the following CAMB[16] plots we see the effect of these parameters on C_ℓ by varying one parameter at a time around a *reference model*, whose parameters have the following values.

Independent parameters:

$$\begin{aligned} \Omega_0 &= 1 & \qquad \Omega_\Lambda &= 0.7 \\ A_s &= 1 & \qquad \omega_m &= 0.147 \\ n_s &= 1 & \qquad \omega_b &= 0.022 \\ \tau &= 0.1 \end{aligned}$$

[14] The situation is analogous to the choice of independent thermodynamic variables in thermodynamics.

[15] There are other choices in use, which are even more geared to minimizing parameter degeneracy. For example, the sound horizon angle ϑ_s may be used instead of Ω_Λ as an independent parameter, since it is directly determined by the acoustic peak separation. However, since the determination of the dependent parameters from it is complicated, such use is more directed towards technical data analysis than pedagogical discussion.

[16] CAMB is a publicly available code for precise calculation of the C_ℓ spectrum. See http://camb.info

which give for the dependent parameters

$$\begin{aligned} \Omega_m &= 0.3 & h &= 0.7 \\ \Omega_c &= 0.2551 & \omega_c &= 0.125 \\ \Omega_b &= 0.0449 \end{aligned}$$

The meaning of setting $A_s = 1$ is just that the resulting C_ℓ still need to be multiplied by the true value of A_s^2. (In this model the true value should be about $A_s = 5 \times 10^{-5}$ to agree with observations.) If we really had $A_s = 1$, perturbation theory of course would not be valid! This is a relatively common practice, since the effect of changing A_s is so trivial that it makes not much sense to plot C_ℓ separately for different values of A_s.

9.11.2 Sound horizon angle

The positions of the acoustic peaks of the C_ℓ spectrum provide us with a measurement of the sound horizon angle

$$\vartheta_s \equiv \frac{r_s(t_{\rm dec})}{d_A^c(t_{\rm dec})}$$

We can use this in the determination of the values of the cosmological parameters, once we have calculated how this angle depends on those parameters. It is the ratio of two quantities, the sound horizon at photon decoupling, $r_s(t_{\rm dec})$, and the angular diameter distance to the last scattering, $d_A^c(t_{\rm dec})$.

Angular diameter distance to last scattering

The angular diameter distance $d_A^c(t_{\rm dec})$ to the last scattering surface we have already calculated and it is given by Eq. (39) as

$$d_A^c(t_{\rm dec}) = H_0^{-1} \frac{1}{\sqrt{|\Omega_0 - 1|}} f_k \left(\sqrt{|\Omega_0 - 1|} \int_{\frac{1}{1+z_{\rm dec}}}^1 \frac{da}{\sqrt{\Omega_0(a - a^2) - \Omega_\Lambda(a - a^4) + a^2}} \right) , \quad (115)$$

from which we see that it depends on the three cosmological parameters H_0, Ω_0 and Ω_Λ. Here $\Omega_0 = \Omega_m + \Omega_\Lambda$, so we could also say that it depends on H_0, Ω_m, and Ω_Λ, but it is easier to discuss the effects of these different parameters if we keep Ω_0 as an independent parameter, instead of Ω_m, since the "geometry effect" of the curvature of space, which determines the relation between the comoving angular diameter distance d_A^c and the comoving distance d^c, is determined by Ω_0.

1. The comoving angular diameter distance is inversely proportional to H_0 (directly proportional to the Hubble distance H_0^{-1}).

2. Increasing Ω_0 decreases $d_A^c(t_{\rm dec})$ in relation to $d^c(t_{\rm dec})$ because of the geometry effect.

3. With a fixed Ω_Λ, increasing Ω_0 decreases $d^c(t_{\rm dec})$, since it means increasing Ω_m, which has a decelerating effect on the expansion. With a fixed present expansion rate H_0, deceleration means that expansion was faster earlier $\Rightarrow$ universe is younger $\Rightarrow$ there is less time for photons to travel as the universe cools from $T_{\rm dec}$ to T_0 $\Rightarrow$ last scattering surface is closer to us.

4. Increasing Ω_Λ (with a fixed Ω_0) increases $d^c(t_{\rm dec})$, since it means a larger part of the energy density is in dark energy, which has an accelerating effect on the expansion. With fixed H_0, this means that expansion was slower in the past $\Rightarrow$ universe is older $\Rightarrow$ more time for photons $\Rightarrow$ last scattering surface is further out. $\therefore$ Ω_Λ increases $d_A^c(t_{\rm dec})$.

Here 2 and 3 work in the same direction: increasing Ω_0 decreases $d_A^c(t_{\rm dec})$, but the geometry effect (2) is stronger. See Fig. 17 for the case $\Omega_\Lambda = 0$, where the dashed line (the comoving distance) shows effect (3) and the solid line (the comoving angular diameter distance) the combined effect (2) and (3).

However, now we have to take into account that, in our chosen parametrization, H_0 is not an independent parameter, but

$$H_0^{-1} \propto \sqrt{\frac{\Omega_0 - \Omega_\Lambda}{\omega_m}}\,,$$

so that via H_0^{-1}, Ω_0 increases and Ω_Λ decreases $d_A^c(t_{\rm dec})$, which are the opposite effects to those discussed above. For Ω_Λ this opposite effect wins. See Fig. 26.

Sound horizon

To calculate the sound horizon,

$$r_s(t_{\rm dec}) \;=\; \int_0^{t_{\rm dec}} \frac{c_s(t)}{a(t)} dt \;=\; \int_0^{a_{\rm dec}} \frac{da}{a\cdot(da/dt)} c_s(a)\,, \tag{116}$$

we need the speed of sound, from Eq. (84),

$$c_s^2(x) = \frac{1}{3}\frac{1}{1+\frac{3}{4}\frac{\bar\rho_b}{\bar\rho_\gamma}} = \frac{1}{3}\frac{1}{1+\frac{3}{4}\frac{\omega_b}{\omega_\gamma}a}\,, \tag{117}$$

where the upper limit of the integral is $a_{\rm dec} = 1/(1+z_{\rm dec})$.

The other element in the integrand of Eq. (116) is the expansion law $a(t)$ before decoupling. From Chapter 3 we have that

$$a\frac{da}{dt} = H_0\sqrt{\Omega_r + \Omega_m a + (1-\Omega_0)a^2 + \Omega_\Lambda a^4}\,. \tag{118}$$

In the integral (115) we dropped the Ω_r, since it is important only at early times, and the integral from $a_{\rm dec}$ to 1 is dominated by late times. Integral (116), on the other hand, includes only early times, and now we can instead drop the Ω_Λ and $1-\Omega_0$ terms (i.e., we can ignore the effect of curvature and dark energy in the early universe, before photon decoupling), so that

$$a\frac{da}{dt} \approx H_0\sqrt{\Omega_m a + \Omega_r} = H_{100}\sqrt{\omega_m a + \omega_r} = \frac{\sqrt{\omega_m a + \omega_r}}{2998\,{\rm Mpc}}\,, \tag{119}$$

where we have written

$$H_0 \equiv h\cdot 100\frac{\rm km/s}{\rm Mpc} \equiv h\cdot H_{100} = \frac{h}{2997.92\,{\rm Mpc}}\,. \tag{120}$$

Thus the sound horizon is given by

$$\begin{aligned} r_s(a) &= 2998\,{\rm Mpc}\int_0^a \frac{c_s(x)dx}{\sqrt{\omega_m x + \omega_r}} \\ &= 2998\,{\rm Mpc}\cdot\frac{1}{\sqrt{3\omega_r}}\int_0^a \frac{dx}{\sqrt{\left(1+\frac{\omega_m}{\omega_r}x\right)\left(1+\frac{3}{4}\frac{\omega_b}{\omega_\gamma}x\right)}}\,. \end{aligned} \tag{121}$$

Here

$$\omega_\gamma = 2.473\times 10^{-5} \qquad \text{and} \tag{122}$$

$$\omega_r = \left[1+\frac{7}{8}N_{\rm eff}\left(\frac{4}{11}\right)^{4/3}\right]\omega_\gamma = 1.692\,\omega_\gamma = 4.184\times 10^{-5} \tag{123}$$

are accurately known from the CMB temperature $T_0 = 2.7255\,\mathrm{K}$ (and therefore we do not consider them as cosmological parameters in the sense of something to be determined from the C_ℓ spectrum).

Thus the sound horizon depends on the two cosmological parameters ω_m and ω_b,

$$r_s(t_{\rm dec}) = r_s(\omega_m, \omega_b)$$

From Eq. (121) we see that increasing either ω_m or ω_b *makes the sound horizon at decoupling,* $r_s(a_{\rm dec})$*, shorter*:

- ω_b slows the sound down
- ω_m speeds up the expansion at a given temperature, so the universe cools to $T_{\rm dec}$ in less time.

The integral (121) can be done and it gives

$$r_s(t_{\rm dec}) = \frac{2998\,\mathrm{Mpc}}{\sqrt{1+z_{\rm dec}}}\frac{2}{\sqrt{3\omega_m R_*}}\ln\frac{\sqrt{1+R_*}+\sqrt{R_*+r_*R_*}}{1+\sqrt{r_*R_*}}\,, \tag{124}$$

where

$$r_* \equiv \frac{\bar\rho_r(t_{\rm dec})}{\bar\rho_m(t_{\rm dec})} = \frac{\omega_r}{\omega_m}(1+z_{\rm dec}) = 0.0456\frac{1}{\omega_m}\left(\frac{1+z_{\rm dec}}{1091}\right) \tag{125}$$

$$R_* \equiv \frac{3\bar\rho_b(t_{\rm dec})}{4\bar\rho_\gamma(t_{\rm dec})} = \frac{3\omega_b}{4\omega_\gamma}\frac{1}{1+z_{\rm dec}} = 27.8\,\omega_b\left(\frac{1091}{1+z_{\rm dec}}\right)\,. \tag{126}$$

For our reference values $\omega_m = 0.147$, $\omega_b = 0.022$, and $1 + z_{\rm dec} = 1091$[17] we get $r_* = 0.310$ and $R_* = 0.614$ and $r_s(t_{\rm dec}) = 144\,\mathrm{Mpc}$ for the sound horizon at decoupling.

Summary

The angular diameter distance $d_A^c(t_{\rm dec})$ is the most naturally discussed in terms of H_0, Ω_0, and Ω_Λ, but since these are not the most convenient choice of independent parameters for other purposes, we shall trade H_0 for ω_m according to Eq. (113). Thus we have that the sound horizon angle depends on 4 parameters,

$$\vartheta_s \equiv \frac{r_s(\omega_m,\omega_b)}{d_A^c(\Omega_0,\Omega_\Lambda,\omega_m)} = \vartheta_s(\Omega_0,\Omega_\Lambda,\omega_m,\omega_b) \tag{127}$$

9.11.3 Acoustic peak heights

There are a number of effects affecting the heights of the acoustic peaks:

1. **The early ISW effect.** The early ISW effect raises the first peak. It is caused by the evolution of Φ because of the effect of the radiation contribution on the expansion law after $t_{\rm dec}$. This depends on the radiation-matter ratio at that time; decreasing ω_m makes the early ISW effect stronger.
2. **Shift of oscillation equilibrium by baryons.** (Baryon drag.) This makes the odd peaks (which correspond to compression of the baryon-photon fluid in the potential wells, decompression on potential hills) higher, and the even peaks (decompression at potential wells, compression on top of potential hills) lower.

[17] Photon decoupling temperature, and thus $1+z_{\rm dec}$, depends somewhat on ω_b, but since this dependence is not easy to calculate (recombination and photon decoupling were discussed in Chapter 4), we have mostly ignored this dependence and used the fixed value $1 + z_{\rm dec} = 1091$.

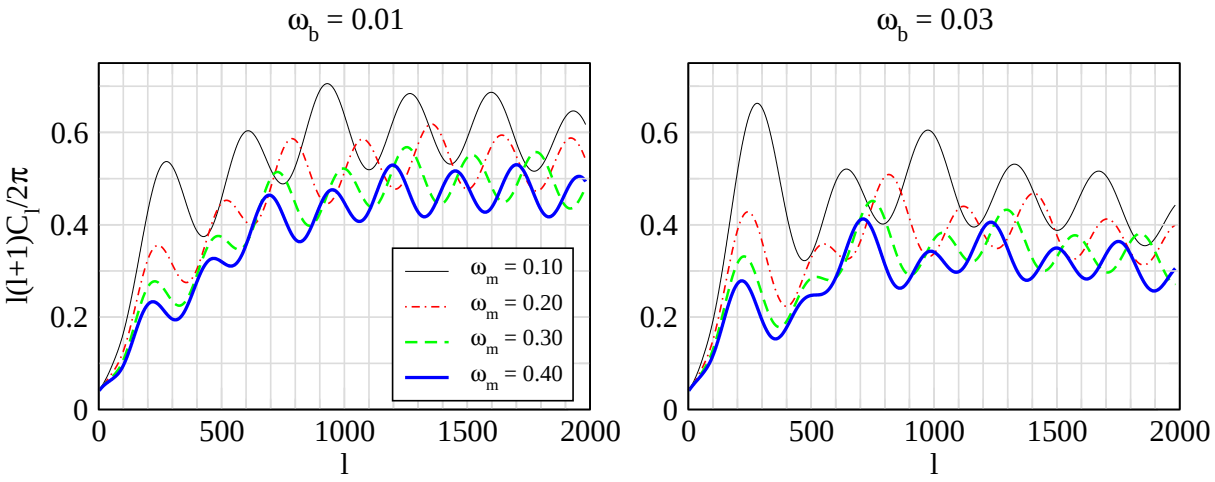


Figure 24: The effect of ω_m. The angular power spectrum C_ℓ is here calculated without the effect of diffusion damping, so that the other effects on peak heights could be seen more clearly. Notice how reducing ω_m raises all peaks, but the effect on the first few peaks is stronger in relative terms, as the radiation driving effect is extended towards larger scales (smaller ℓ). The first peak is raised mainly because the ISW effect becomes stronger. Figure and calculation by R. Keskitalo.

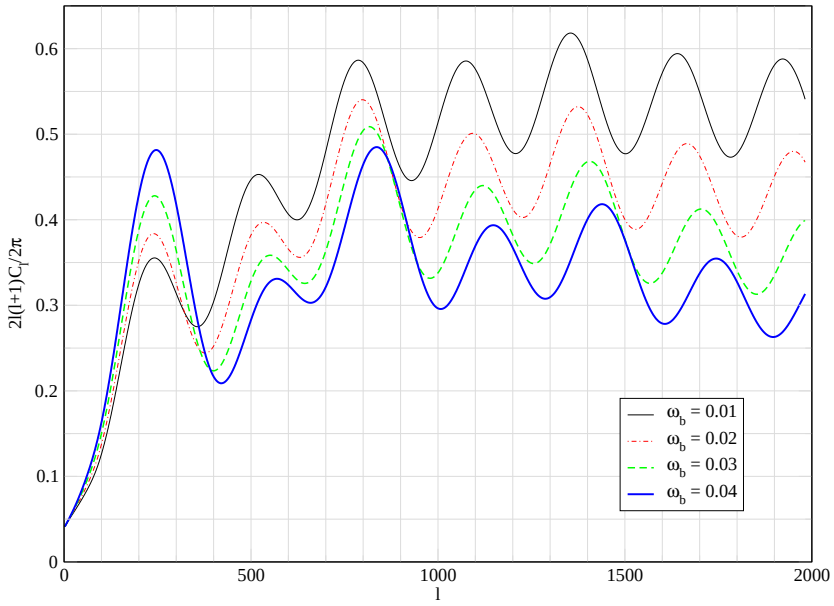


Figure 25: The effect of ω_b. The angular power spectrum C_ℓ is here calculated without the effect of diffusion damping, so that the other effects on peak heights could be seen more clearly. Notice how increasing ω_b raises odd peaks relative to the even peaks. Because of baryon damping there is a general trend downwards with increasing ω_b. This figure is for $\omega_m = 0.20$. Figure and calculation by R. Keskitalo.

3. **Baryon damping.** The time evolution of $R \equiv 3\bar{\rho}_b/4\bar{\rho}_\gamma$ causes the amplitude of the acoustic oscillations to be damped in time roughly as $(1+R)^{-1/4}$. This reduces the amplitudes of all peaks.

4. **Radiation driving.**[18] This is an effect related to horizon scale physics that we have not tried to properly calculate. For scales k which enter during the radiation-dominated epoch, or near matter-radiation equality, the potential Φ decays around the time when the scale enters. The potential keeps changing as long as the radiation contribution is important, but the largest change in Φ is around horizon entry. Because the sound horizon and Hubble length are comparable, horizon entry and the corresponding potential decay always happen during the first oscillation period. This means that the baryon-photon fluid is falling into a deep potential well, and therefore is compressed by gravity by a large factor, before the resulting overpressure is able to push it out. Meanwhile the potential has decayed, so it is less able to resist the decompression phase, and the overpressure is able to kick the fluid further out of the well. This increases the amplitude of the acoustic oscillations. The effect is stronger for the smaller scales which enter when the universe is more radiation-dominated, and therefore raises the peaks with a larger peak number m more. Reducing ω_m makes the universe more radiation dominated, making this effect stronger and extending it towards the peaks with lower peak number m.

5. **Diffusion damping.** Diffusion damping lowers the heights of the peaks. It acts in the opposite direction than the radiation driving effect, lowering the peaks with a larger peak number m more. Because the diffusion damping effect is exponential in ℓ, it wins for large ℓ.

Effects 1 and 4 depend on ω_m, effects 2, 3, and 5 on ω_b. See Figs. 24 and 25 for the effects of ω_m and ω_b on peak heights.

[18] This is also called gravitational driving, which is perhaps more appropriate, since the effect is due to the change in the gravitational potential.

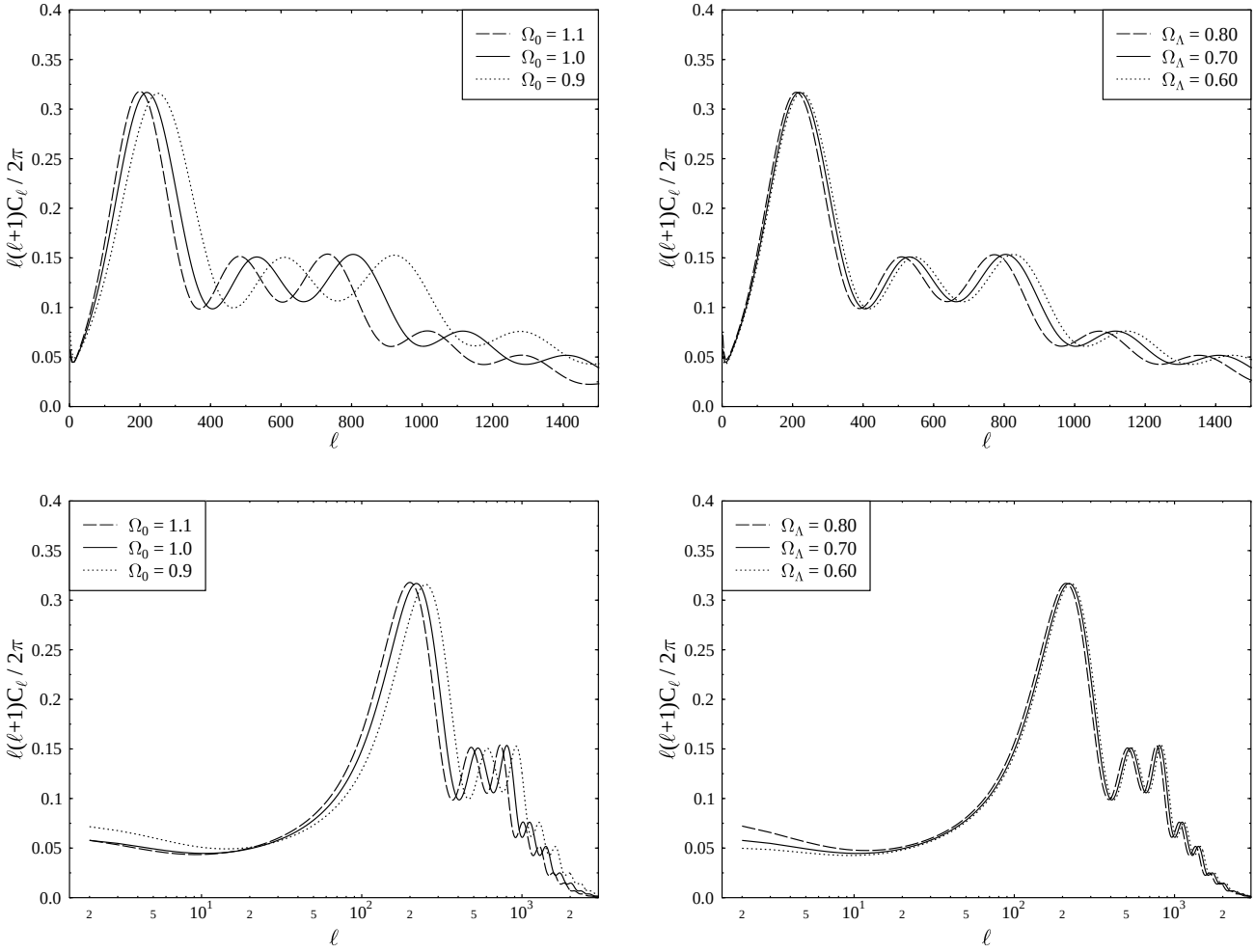


Figure 26: The effect of changing Ω_0 or Ω_Λ from their reference values $\Omega_0 = 1$ and $\Omega_\Lambda = 0.7$. The top panels shows the C_ℓ spectrum with a linear ℓ scale so that details at larger ℓ where cosmic variance effects are smaller can be better seen. The bottom plot has a logarithmic ℓ scale so that the integrated Sachs-Wolfe effect at small ℓ can be better seen. The logarithmic scale also makes clear that the effect of the change in sound horizon angle is to stretch the spectrum by a constant factor in ℓ space.

9.11.4 Effect of Ω_0 and Ω_Λ

These two parameters have only two effects:

1. they affect the sound horizon angle and thus the positions of the acoustic peaks
2. they affect the late ISW effect

See Fig. 26. Since the late ISW effect is in the region of the C_ℓ spectrum where the cosmic variance is large, it is difficult to detect. Thus we can in practice only use ϑ_s to determine Ω_0 and Ω_Λ. Since ω_b and ω_m can be determined quite accurately from C_ℓ acoustic peak heights, peak separation, i.e., ϑ_s, can then indeed be used for the determination of Ω_0 and Ω_Λ. Since one number cannot be used to determine two, the parameters Ω_0 and Ω_Λ are degenerate. CMB observations alone cannot be used to determine them both. Other cosmological observations (like the power spectrum $P_\delta(k)$ from large scale structure, or the SNIa redshift-distance relationship) are needed to break this degeneracy.

A fixed ϑ_s together with fixed ω_b and ω_m determine a line on the (Ω_0, Ω_Λ) -plane. See Fig. 27. Derived parameters, e.g., h, vary along that line. As you can see from Fig. 26, changing Ω_0 (around the reference model) affects ϑ_s much more strongly than changing Ω_Λ. This means

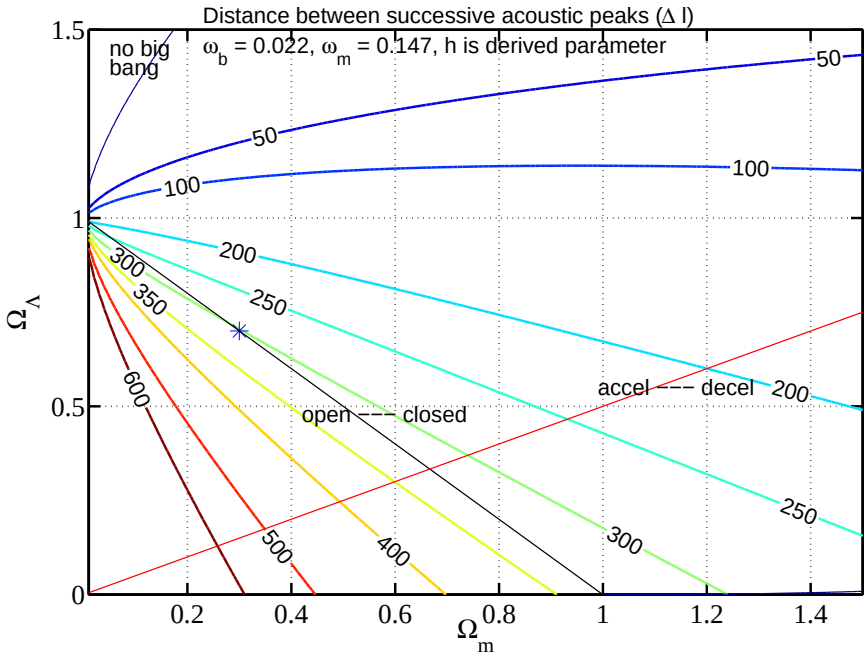


Figure 27: The lines of constant sound horizon angle ϑ_s on the $(\Omega_m,\Omega_\Lambda)$ plane for fixed ω_b and ω_m. The numbers on the lines refer to the corresponding acoustic scale $\ell_A \equiv \pi/\vartheta_s$ ($\sim$ peak separation) in multipole space. Figure by J. Väliviita. See his PhD thesis[10], p.70, for an improved version including the HST constraint on h.

that the orientation of the line is such that Ω_Λ varies more rapidly along that line than Ω_0. Therefore using additional constraints from other cosmological observations, e.g., the Hubble Space Telescope determination of h based on the distance ladder, which select a short section from that line, gives us a fairly good determination of Ω_0, leaving the allowed range for Ω_Λ still quite large.

Therefore it is often said that CMB measurements have determined that $\Omega_0 \sim 1$. But as explained above, this determination necessary requires the use of some auxiliary cosmological data besides the CMB.

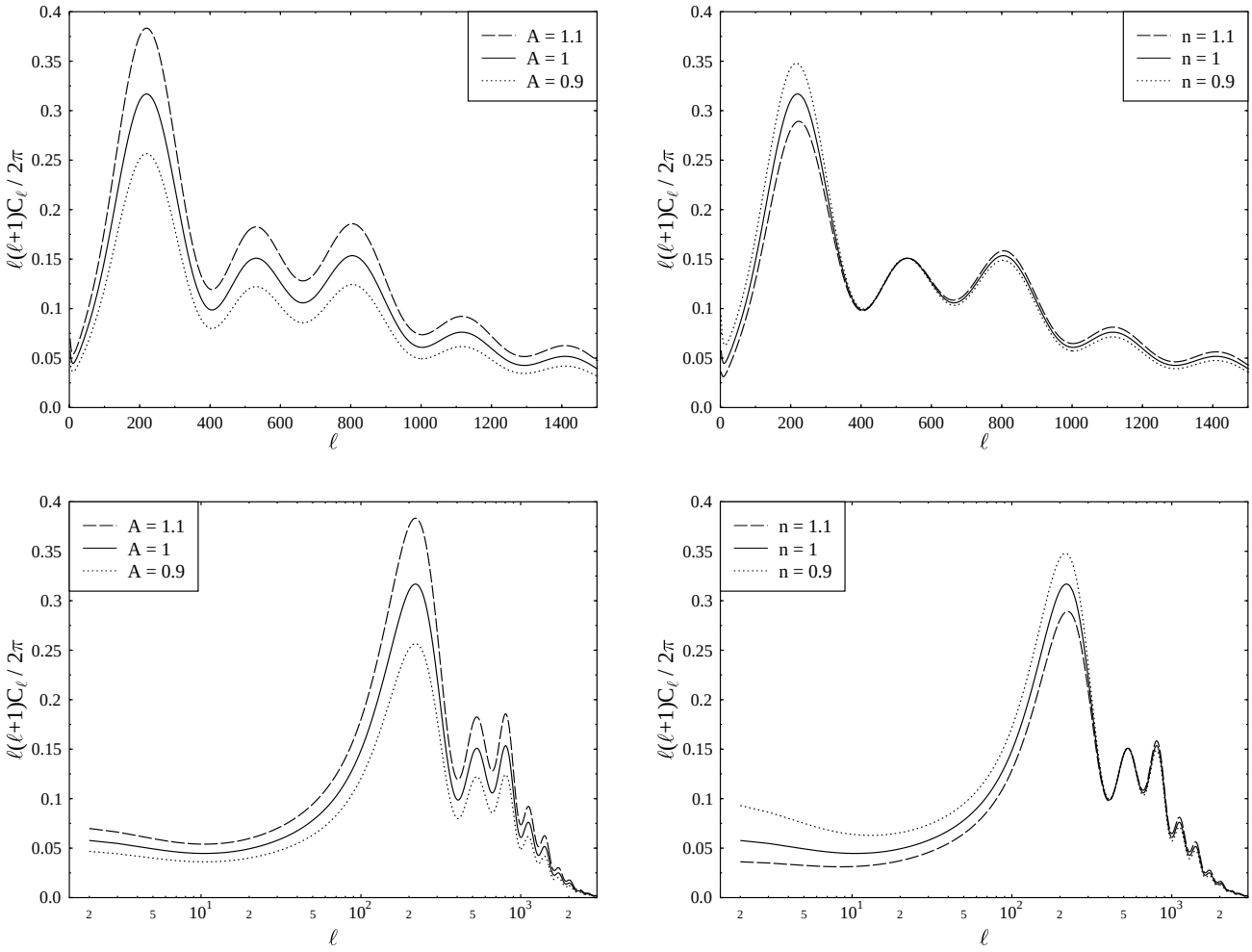


Figure 28: The effect of changing the primordial amplitude and spectral index from their reference values $A_s = 1$ and $n_s = 1$.

9.11.5 Effect of the primordial spectrum

The effect of the primordial spectrum is simple: raising the amplitude A_s raises the C_ℓ also, and tilting the primordial spectrum tilts the C_ℓ also. See Fig. 28.

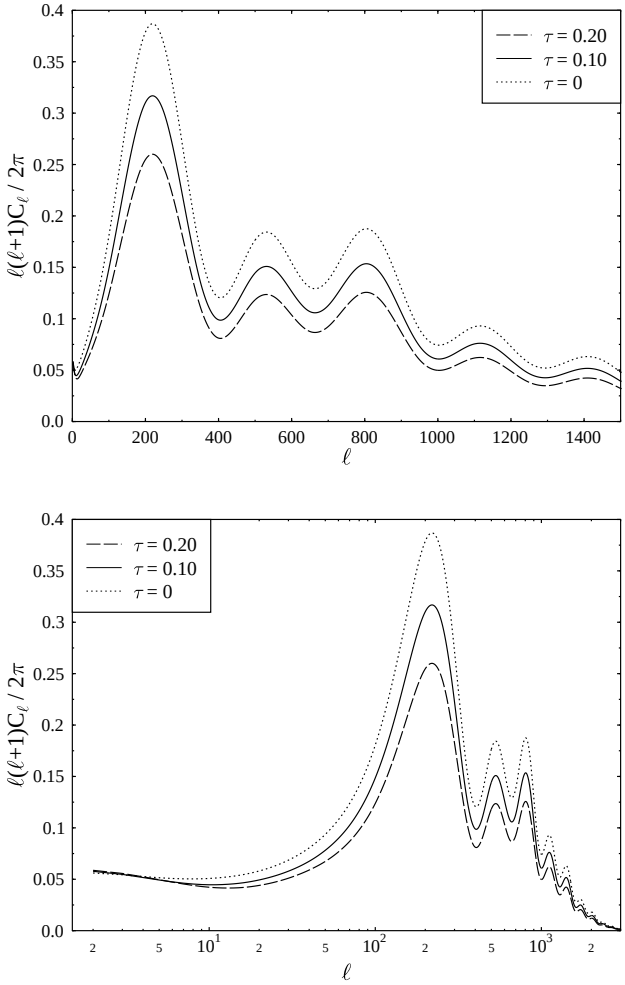


Figure 29: The effect of changing the optical depth from its reference value $\tau = 0.1$.

9.11.6 Optical depth due to reionization

The optical depth τ due to reionization was discussed in Sec. 9.10. See Fig. 29 for its effect on C_ℓ.

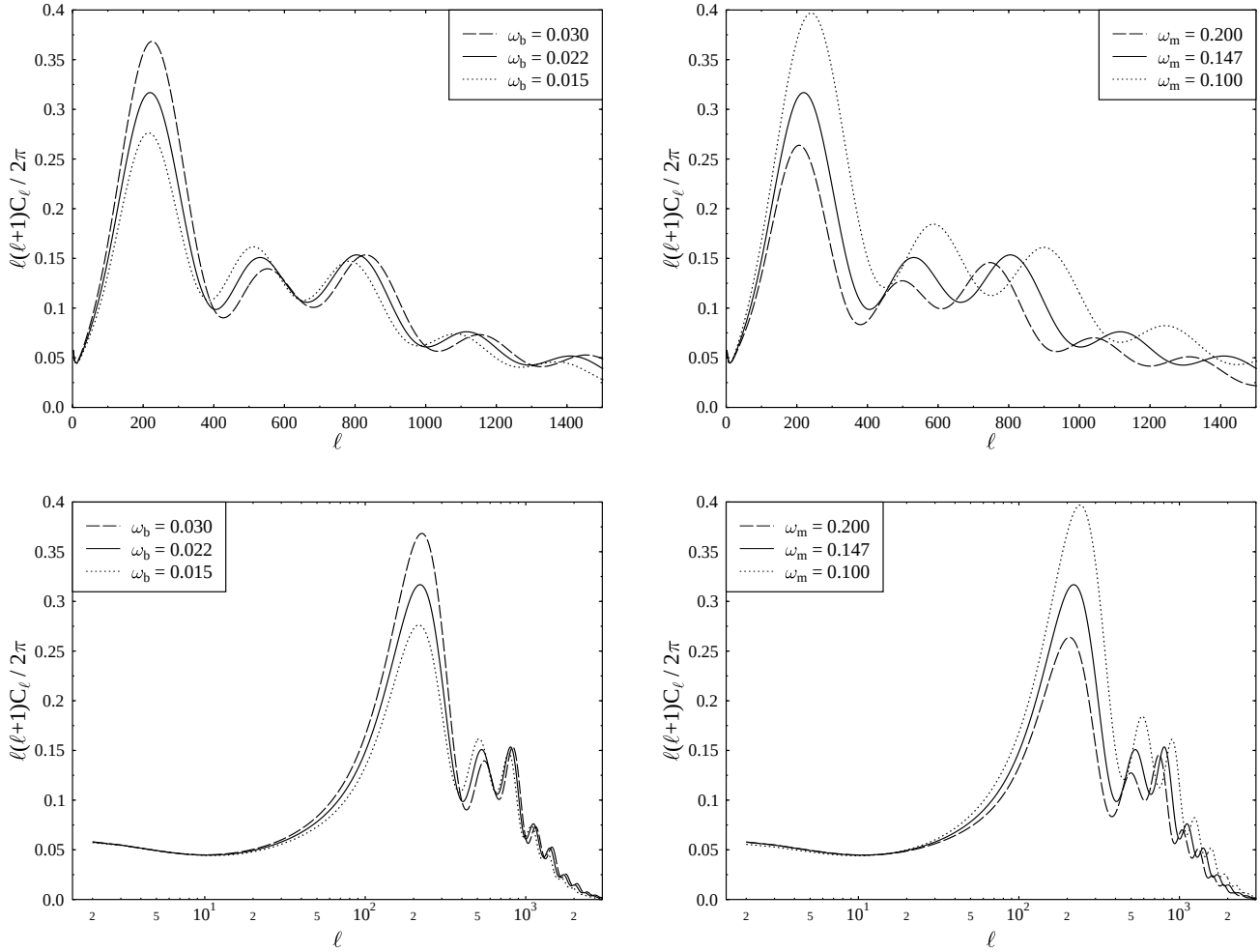


Figure 30: The effect of changing the physical baryon density and matter density parameters from their reference values $\omega_b = 0.022$ and $\omega_m = 0.147$.

9.11.7 Effect of ω_b and ω_m

These parameters affect both the positions of the acoustic peaks (through ϑ_s) and the heights of the different peaks. The latter effect is the more important, since both parameters have their own signature on the peak heights, allowing an accurate determination of these parameters, whereas the effect on ϑ_s is degenerate with Ω_0 and Ω_Λ.

Especially ω_b has a characteristic effect on peak heights: Increasing ω_b raises the odd peaks and reduces the even peaks, because it shifts the balance of the acoustic oscillations (the $-R\Phi$ effect). This shows the most clearly at the first and second peaks. Raising ω_b also shortens the damping scale k_D^{-1} due to photon diffusion, moving the corresponding damping scale ℓ_D of the C_ℓ spectrum towards larger ℓ. This has the effect of raising C_ℓ at large ℓ. See Fig. 30.

There is also an overall "baryon damping effect" on the acoustic oscillations which we have not calculated. It is due to the time dependence of $R \equiv 3\bar{\rho}_b/4\bar{\rho}_m$, which reduces the amplitude of the oscillation by about $(1+R)^{-1/4}$. This explains why the third peak in Fig. 30 is no higher for $\omega_b = 0.030$ than it is for $\omega_b = 0.022$.

Increasing ω_m makes the universe more matter dominated at $t_{\rm dec}$ and therefore it reduces the early ISW effect, making the first peak lower. This also affects the shape of the first peak.

The "radiation driving" effect is most clear at the second to fourth peaks. Reducing ω_m makes these peaks higher by making the universe more radiation-dominated at the time the corresponding scales enter, strengthening this radiation driving. The fifth and further peaks

Parameters for the ΛCDM model

	Planck 2018	best fit
ω_b	0.02237 ± 0.00015	0.022383
ω_m	0.1424 ± 0.0012	0.14249
Ω_Λ	0.685 ± 0.007	0.6841
τ	0.054 ± 0.007	0.0543
A_s	$4.58 \pm 0.04 \times 10^{-5}$	4.5832×10^{-5}
n_s	0.965 ± 0.004	0.96605
H_0	67.36 ± 0.54 km/s/Mpc	67.32 km/s/Mpc
ω_c	0.1200 ± 0.0012	0.12011
Ω_m	0.315 ± 0.007	0.3158
$z_{\rm eq}$	3402 ± 26	
$k_{\rm eq}^{-1}$	96.3 ± 0.8 Mpc	
$z_{\rm dec}$	1089.92 ± 0.25	
k_D^{-1}	7.10 ± 0.02 Mpc	
$z_{\rm reion}$	7.7 ± 0.7	7.68
ϑ_s	$0.5965^\circ \pm 0.0002^\circ$	0.59651°
t_0	$13.797 \pm 0.023 \times 10^9$ a	13.7971×10^9 a

Table 2: These parameter values are based on the CMB temperature power spectrum C_ℓ, CMB polarization, and gravitational lensing of the CMB, as observed by the Planck satellite [5]. The first six parameters, above the line, are independent parameters, and the parameters below the line are quantities that can be derived from them in the ΛCDM model. The error estimates are 68% confidence limits. The best-fit column gives a representative model that is an excellent fit to the data; nearby models in the 6-parameter space may be practically equally good fits. Note that here Ω_m includes the contribution from neutrinos with $\sum m_\nu = 0.06$ meV ($\Omega_\nu = 0.0014$) whereas ω_m does not.

correspond to scales that have anyway essentially the full effect, and for the first peak this effect is anyway weak. (We see instead the ISW effect in the first peak.) See Fig. 30.

9.12 Current best estimates for the cosmological parameters

9.12.1 Planck values for ΛCDM parameters

The most important data set for determining cosmological parameters is the Planck data [5] on the CMB anisotropy. We give the parameter values determined by Planck for the ΛCDM model in Table 2. Note that all independent parameters of the model are fit simultaneously to the same data. The determination is based on the assumption that the model, here ΛCDM, is correct. One can judge this assumption based on how well the model fits the data. In the case of Planck and ΛCDM the fit is good; adding more parameters to the model does not improve the fit significantly.

This model agrees reasonable well with most of the other available cosmological data, with the exception of the distance-ladder determination of the Hubble constant, based on Cepheids and Type Ia supernovae, which gives $H_0 = 73.5 \pm 1.6$ km/s/Mpc [11, 12]. This is called the *local* measurement of H_0, since these measurements are from nearby parts of the Universe, in contrast to the *global* determination from the CMB, where the CMB has traversed the entire observable Universe. This discrepancy has been evident in the data for some time, but it has gradually become more serious as the error bars on H_0 from both CMB and local measurements have become tighter without the central values changing much. One may suspect systematic errors in the distance ladder data or that the ΛCDM model is a too simple model for the universe.

Constraints for extended models

	ΛCDM	Planck 2018	Planck+ext
Ω_0	1.0	1.011 ± 0.013	0.9993 ± 0.0037
r	0	< 0.101	< 0.065
$dn_s/d\ln k$	0	-0.005 ± 0.013	-0.004 ± 0.013
w	-1	-1.6 ± 0.5	-1.04 ± 0.10
$\sum m_\nu$	0.06 eV	< 0.241 eV	< 0.120 eV
$N_{\rm eff}$	3.046	2.89 ± 0.38	2.99 ± 0.34

Table 3: Each row is a different model and we show limits only to the "additional" parameter. As is customary with limits, the ranges are given as 95% confidence limits. $N_{\rm eff}$, the "effective number of neutrino species", refers to relativistic energy density (in addition to photons) near the time of photon decoupling.

9.12.2 Extended models and external data

In the ΛCDM model the universe is flat, $\Omega_0 = 1$. We can also fit *extended models*, with additional independent parameters. Such 7-parameter models, with one extra parameter in addition to the ΛCDM parameters, are fit to Planck data in Table 3. Since the ΛCDM model is a good fit, the estimates for these extra parameters are consistent with their values in the ΛCDM model. Instead of the central value, we therefore concentrate on the estimated probable range, i.e., *limits* to the deviation from the ΛCDM model. Note that in these extended models the ranges for the 6 ΛCDM parameters will be different from Table 2; they will be wider and the central values will be slightly different. One could of course consider models with more independent parameters, e.g., the 12-parameter model, where all the 6 parameters of Table 3 were added to ΛCDM. In such a model the allowed ranges for all these parameters would be wider than in Tables 2 and 3. The argument against such a model is *Occam's razor*: if there are many models that fit the data, one should prefer the simplest one; a corollary to this is that the models one should consider next are those that are almost as simple. Of course, there is no guarantee against all these parameters having a significant effect on the CMB. These one-parameter extensions to ΛCDM do not relieve the tension with the local determination of H_0 much, but by adding sufficiently many additional parameters one can get rid of the tension.

Dark radiation. The parameter $N_{\rm eff}$ corresponds to making ω_r a free parameter. From the discussion in Sec. 9.11.2 we see that we are constraining relativistic energy density at or before $t_{\rm dec}$. Additional relativistic particle species, in addition to photons and neutrinos, would raise $N_{\rm eff}$ above the Particle Physics Standard Model value 3.046. The 95% confidence upper limit for the contribution from such extra species is $\Delta N_{\rm eff} \equiv N_{\rm eff} - 3.046 < 0.3$. This rules out any new (currently unknown) particles that would decouple after QCD transition and stay relativistic until photon decoupling, see Fig. 31.

External data. Because of degeneracies of cosmological parameters in the CMB data, most importantly the *geometrical degeneracy* between parameters, like Ω_0, Ω_Λ, and the dark energy equation-of-state parameter w, whose main effect on CMB is via their effect on the angular diameter distance to the last scattering sphere, some parameters of these extended models are only weakly constrained by Planck data. To break these degeneracies, additional cosmological data (BAO and BICEP2/Keck, see below) has been used in the fourth column of Table 3 (ext = external to Planck). The impressive accuracy in this column is, however, mainly due to the accuracy of Planck. The parameter values allowed by Planck form a narrow but long region in the 7-parameter space, and the external data allows a region that is wider, but oriented differently; the intersection of these regions is then a shorter segment of the region allowed by Planck alone, see Fig. 32.

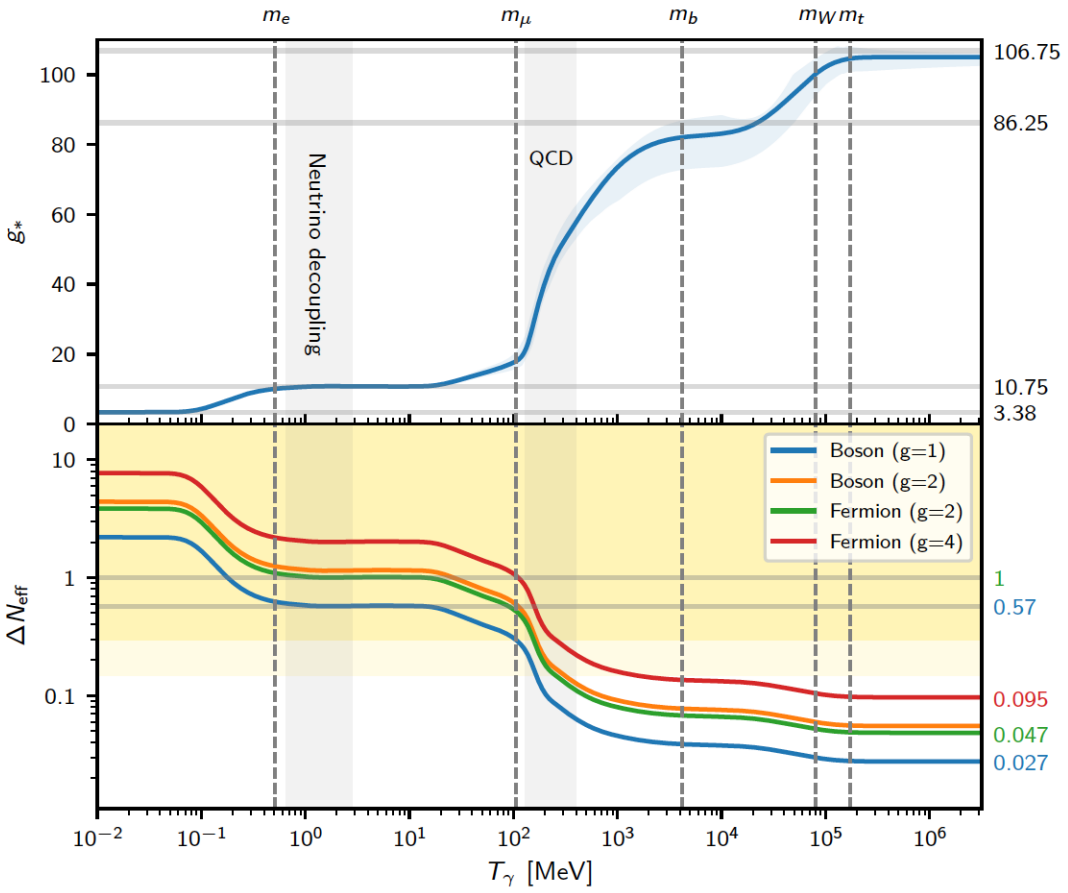


Figure 31: Upper panel: The effective number $g_*(T)$ of degrees of freedom in the Standard Model of particle physics. Note that the drop in $g_*(T)$ due to the QCD transition is not sharp, since this is not a phase transition (taking place at a fixed critical temperature T_c), but is a cross-over transition (happening gradually over a temperature range). Bottom panel: The colored curves show contributions to $\Delta N_{\rm eff}$ from different types of light (relativistic at photon decoupling) thermal relics as a function of their decoupling temperature. The darker yellow region is ruled out by the Planck upper limit $\Delta N_{\rm eff} < 0.3$. (The lighter yellow region corresponds to the 68% confidence upper limit $\Delta N_{\rm eff} < 0.13$.) From [5].

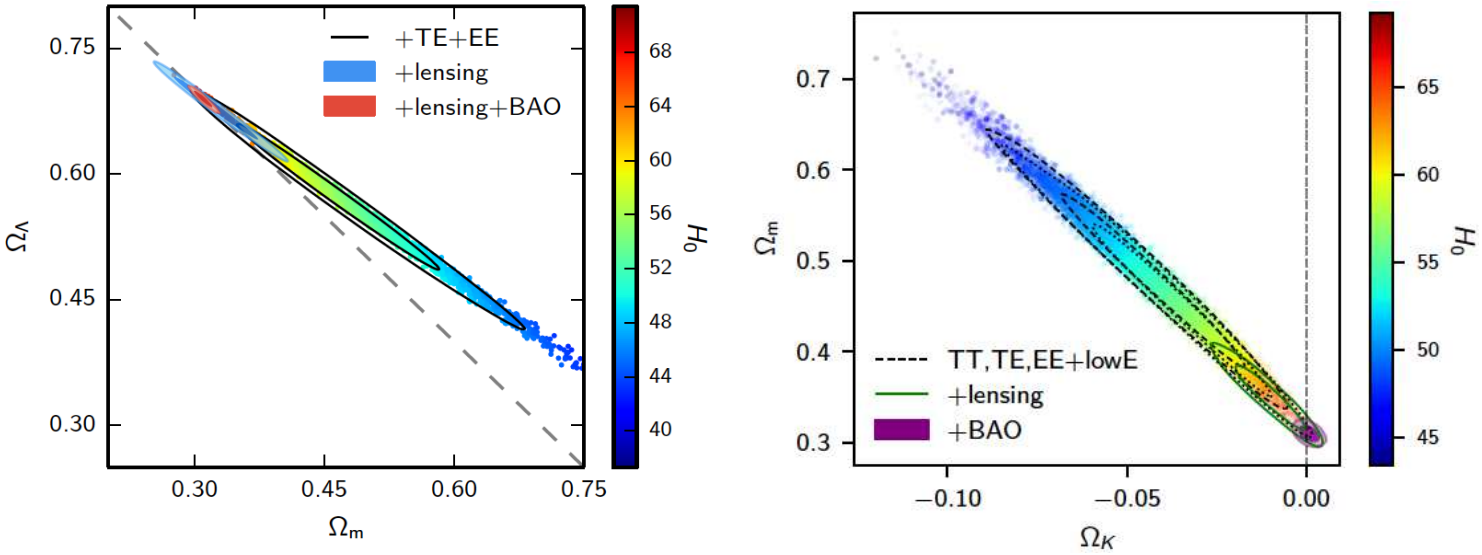


Figure 32: Left: Constraints on Ω_Λ and Ω_m (or $\Omega_0 = \Omega_\Lambda + \Omega_m$) in the ΛCDM+Ω_0 model from Planck 2015 and BAO data. The colored dots represent parameter values that fit Planck temperature C_ℓ and large scale polarization data, the color giving the value of H_0 required for the fit. The black contours (inner 68% and outer 95% confidence limits) give the models that remain allowed when Planck small-scale polarization data is also used; and blue contours when Planck CMB lensing data is used instead. The red contours show the effect of adding BAO data to break the Ω_0-Ω_Λ degeneracy. From the colors one can see that also independent H_0 data could be used to break the degeneracy. The dashed line corresponds to a flat universe. From [3]. Right: The same from Planck 2018 data, except shown for (Ω_k,Ω_m) instead of (Ω_m,Ω_Λ). From [5].

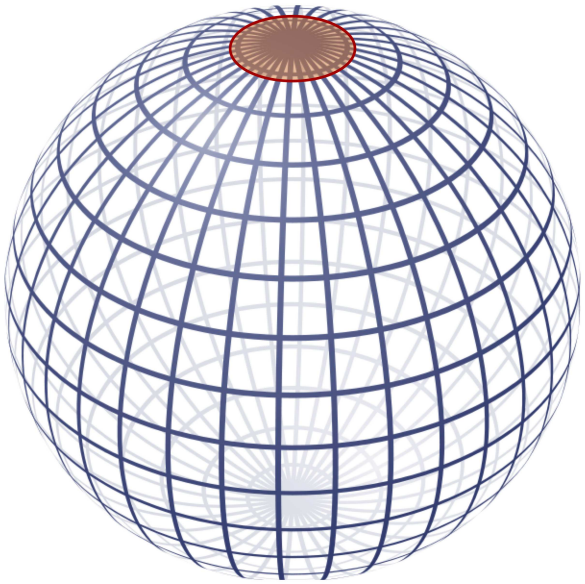


Figure 33: The upper limit $\Omega_0 < 1.003$ means that if we live in a closed universe, its curvature radius $R_{\rm curv} = H^{-1}/\sqrt{|\Omega_k|} > H^{-1}/\sqrt{0.003} = 18.3H^{-1} = 5.9d^c(z = 1090)$ is more than 5 times larger than the distance we can see (to the last scattering sphere, corresponding to the red circle, which is actually drawn too large here, since this figure was drawn when the limit was weaker).

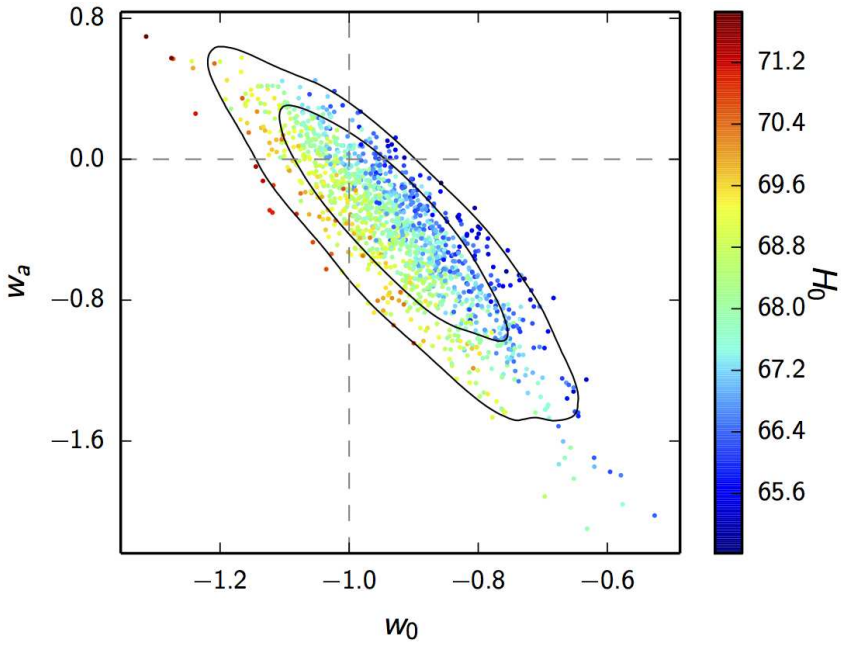


Figure 34: Constraints on the dark energy equation-of-state parameters w_0 and w_a (see text) in the 8-parameter $(w_0 + w_a)$CDM model from Planck, BAO, and SNIa data. From [3].

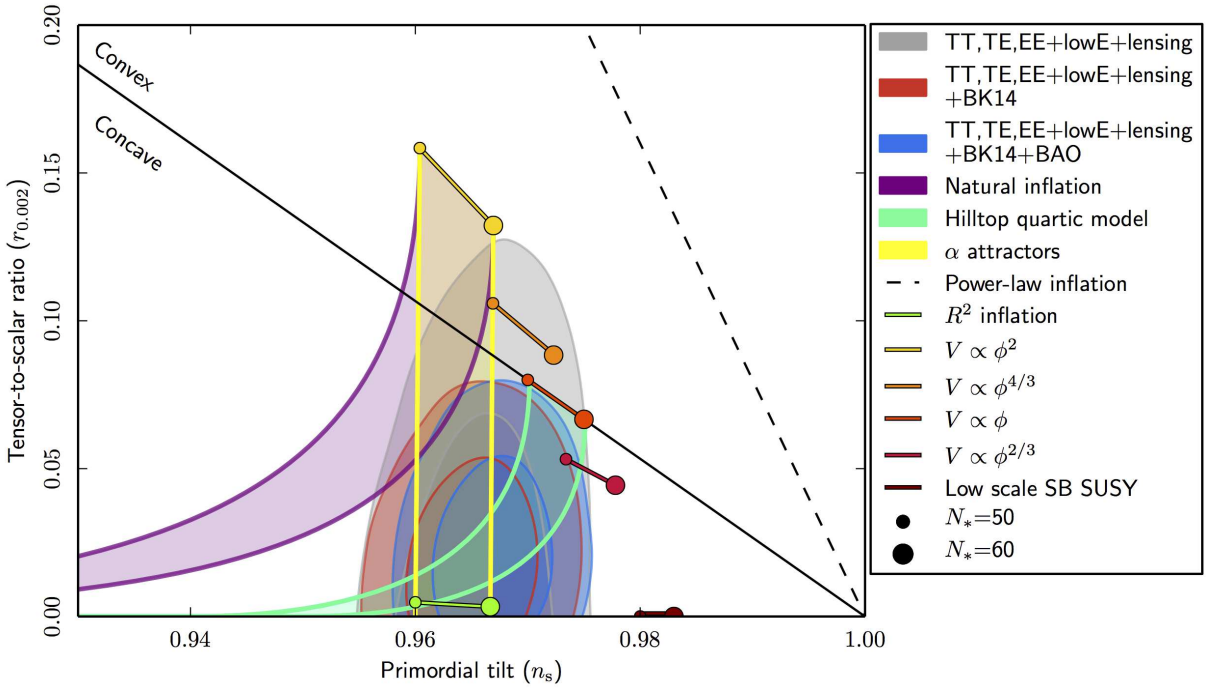


Figure 35: Constraints on the parameters n_s and r, which constrain inflation models, in the 7-parameter ΛCDM+r model from Planck data. Gray contours are based on Planck data only; red and blue contours include external data. Predictions from a selection of inflation models are marked on the plot. From [7].

Large scale structure surveys, i.e., the measurement of the 3-dimensional matter power spectrum $P_\delta(k)$ from the distribution of galaxies, mainly measure the combination $\Omega_m h$, since this determines where $P_\delta(k)$ turns down. Actually it turns down at $k_{\rm eq}$ which is proportional to $\omega_m \equiv \Omega_m h^2$, but since in these surveys the distances to galaxies are deduced from their redshifts (these surveys are also called galaxy redshift surveys), which give the distances only up to the Hubble constant H_0, these surveys determine $h^{-1}k_{\rm eq}$ instead of $k_{\rm eq}$. This cancels one power of h. Having $\Omega_m h^2$ from CMB and $\Omega_m h$ from the galaxy surveys, gives us both h and $\Omega_m = \Omega_0 - \Omega_\Lambda$, which breaks the Ω_0-Ω_Λ degeneracy.

BAO. Measurements of $P_\delta(k)$ are now so accurate that the small residual effect from the acoustic oscillations before photon decoupling can be seen as a weak wavy pattern [13]. This is the same structure which we see in the C_ℓ but now much fainter, since now the baryons have fallen into the CDM potential wells, and the CDM was only mildly affected by these oscillations in the baryon-photon fluid. In this context these are called *baryon acoustic oscillations* (BAO). The half-wavelength of this pattern, however, corresponds to the same sound horizon distance $r_s(t_{\rm dec})$ in both cases.[19] But now the angular scale on the sky is related to it by the angular diameter distance $d_A^c(z)$ to the much smaller redshifts z of the galaxy survey. This $d_A^c(z)$ has then a different relation to Ω_0, Ω_Λ, and ω_m. Comparing CMB data to galaxy surveys gives us the ratio $d_A^c(z)/d_A^c(t_{\rm dec})$, which gives us independent information on these parameters. The large scale structure surveys used for the BAO measurements to supplement Planck 2018 data were the 6dF Galaxy Survey (6dFGS) [14] and the Sloan Digital Sky Survey (SDSS) [15, 16].

Curvature. Because of the geometrical degeneracy, the CMB angular power spectra alone are not good for constraining Ω_k. The peak structure gives a precise measurement of the angular diameter distance to last scattering $d_A^c(t_{\rm dec})$. In the ΛCDM+Ω_k model his translates into a curve on the (Ω_m, Ω_Λ) plane (see Fig. 27). The late ISW effect due to Ω_Λ would break this degeneracy, but since this affects only the lowest multipoles it is lost in the cosmic variance. More significant is a higher-order (beyond linear perturbation theory) effect on the C_ℓ; that of gravitational lensing of the CMB due to large-scale structure. This smooths the acoustic peaks and the effect is proportional to Ω_m. Although the effect is small, it occurs at high ℓ where cosmic variance is small and provides some degeneracy breaking power between Ω_m and Ω_Λ, or equivalently, between Ω_m and Ω_k. The resulting constraint, from Planck C_ℓ only, on Ω_k is $\Omega_k = -0.044^{+0.018}_{-0.015}$ (68% CL), which favors a closed universe at well over 2 σ. The best-fit such models have $\Omega_m > 0.45$ and $H_0 < 55\,$km/s/Mpc, which are ruled out by other cosmological data. The problematic feature in the data is that the acoustic peaks are slightly lower than predicted by the ΛCDM model fit to the data, as if there was too much lensing (requiring higher Ω_m, which then leads to lower Ω_Λ and negative Ω_k as we move along the degeneracy line).

From the Planck data one can also measure the gravitational lensing of the CMB more directly from the effect it has on higher-order correlations (higher than the 2-point correlation measured by C_ℓ) of the CMB. This measurement of CMB lensing agrees with the ΛCDM prediction and thus with a flat universe, giving the constraint $\Omega_k = -0.0106 \pm 0.00065$ (when combined with the Planck C_ℓ). When one adds also BAO data to break the geometric degeneracy, one arrives at the final result given in the Planck+ext column of Table 3,

$$\Omega_k = 0.0007 \pm 0.0019 \qquad (68\%\text{CL})\,. \tag{128}$$

(Table 3 gives 95% confidence limits, which are twice as wide.) The 95% upper limit $\Omega_0 < 1.003$, or $\Omega_k > -0.003$ gives a minimum size to the Universe, see Fig. 33.

Supernovae. Another way to break the geometric degeneracy, is to use the redshift-distance relationship from Supernova Type Ia (SNIa) surveys [17], or simply the distance-ladder determination of H_0, where Cepheids and Supernovae are the last two steps of the ladder. These were

[19] To be accurate, the $t_{\rm dec}$ value to represent the effect in $P_\delta(k)$, is not exactly the same as for C_ℓ, since photon decoupling was not instantaneous, and in one we are looking at the effect on matter and in the other on photons.

not used in the Planck 2018 analysis of 6- and 7-parameter models, because of the discrepancy with the local H_0 determination[20], and since the SNIa data adds little statistical power to the CMB+BAO combination, but the SNIa data was used for the following 8-parameter model.

Dark energy. To constrain properties of dark energy, the 7-parameter wCDM model is probably too simplistic, since it assumes that the equation-of-state parameter w stays constant during the epoch when dark energy has a significant effect on the expansion. To stay at a phenomenological level, i.e., not assuming a particular dark-energy model, but just attempting to constrain its equation of state, the next step is a two-parameter equation of state $w(a) = w_0 + w_a(1-a)$, i.e., a first-order Taylor expansion with w_0 the current value of w, and w_a related to its first derivative with respect to the scale factor, leading to an 8-parameter model. From Fig. 34 you can see that the best fits are near the ΛCDM values $w_0 = -1$, $w_a = 0$, but that the equation of state is poorly constrained.

Neutrino masses. Neutrino masses, i.e., the amount of hot dark matter, have a larger effect on large-scale structure than CMB; the CMB data is mainly needed to determine the other parameters after which the large-scale structure power spectrum $P_\delta(k)$ can be used to determine the sum of the neutrino masses. The value 0.06 eV used for the ΛCDM model is the minimum allowed by neutrino oscillation data.

Tensor perturbations. The polarization pattern of the CMB on the sky can be divided into what are called E and B modes. This is analogous to the division of a vector field into irrotational (curl-free) and rotational (divergence-free) parts. To first (i.e. linear) order in perturbation theory, only tensor perturbations produce B-mode polarization in the CMB. Only E-mode polarization has so far been detected in the CMB. Upper limits to CMB B-mode polarization provide upper limits to the tensor-scalar ratio r. Planck was not optimized for polarization measurements, so its B-mode measurement is noisy and suffers from instrument systematics. Thus the Planck upper limit to r from B modes is weak, $r < 0.41$, and the Planck constraint $r < 0.101$ comes from the effect of tensor perturbations on the CMB temperature C_ℓ. The ground-based BICEP2/Keck Array [18] at the South Pole can measure polarization more accurately, but it has limited sky coverage and needs to be combined with Planck data to separate the CMB from the foreground. This combination leads to the B-mode upper limit $r < 0.065$.

Inflation models. Since inflation produces tensor perturbations, and many inflation models predict that they should be strong enough to have an observable effect on the CMB, the simplest way to constrain inflation is to fit the ΛCDM+r model to CMB data. From Fig. 35 you can see that the $V(\varphi) = \frac{1}{2}m^2\varphi^2$ inflation model, is ruled out by Planck data alone at a 95% confidence level (assuming that ΛCDM+r is the correct model for the universe).

[20] One should not combine discrepant data in parameter fitting. This would lead to artificially tight parameter values with poor fits to both data sets.

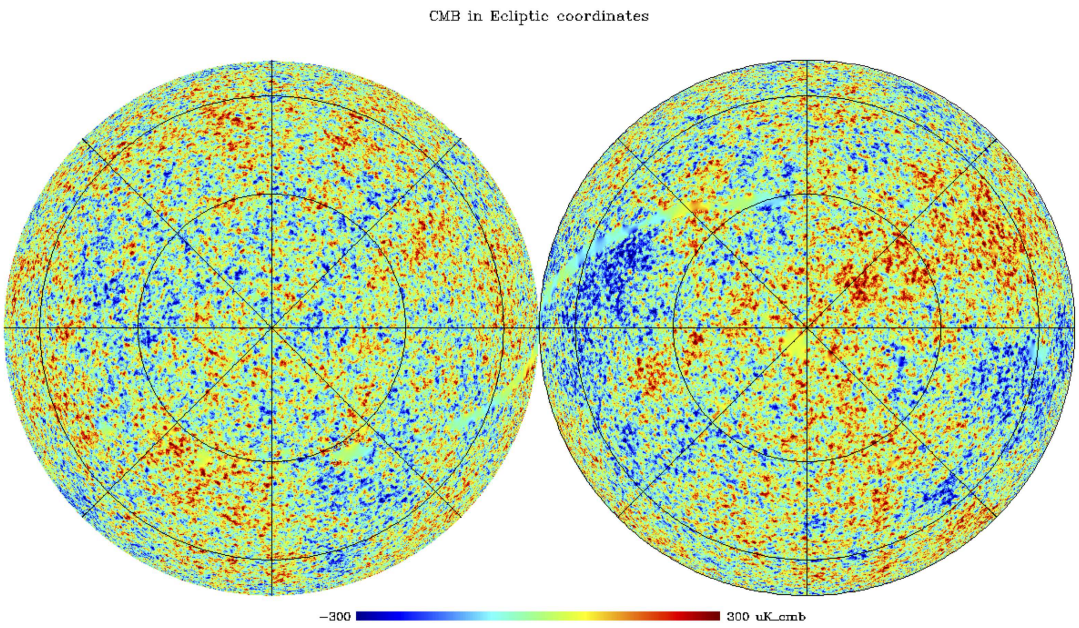


Figure 36: CMB temperature anisotropy in ecliptic coordinates.

9.13 Issues with CMB data

While the agreement of the CMB observations with the predictions of the ΛCDM model is impressive (see Fig. 11), it is not perfect. Also, while combining other cosmological data with CMB data adds more support for the ΛCDM model, there are some discrepancies. There are at least three issues:

Large scale anomalies. Comparing the northern (Fig. 6) and southern (Fig. 7) galactic hemispheres, one may notice that the southern hemisphere has stronger large-scale CMB anisotropies. The difference is more clear between the ecliptic hemispheres, see Fig. 36. This is not what we would expect from statistical isotropy. Also the quadrupole and octupole have planar shape ($m = \ell$ dominates, see the Y_{22} and Y_{33} in Fig. 10) and are aligned with each other (and the quadrupole is rather weak). These large-scale anomalies are seen in both WMAP and Planck data, so they are real. They may be just a statistical fluke, or a sign that the Universe deviates from standard ΛCDM at the very largest observable scales. Because of cosmic variance it is difficult to tell.

"Lensing smoothing". While the scatter of data points around the theoretical prediction (see Fig. 37) is mainly as expected (the error bars are 68% CL, so we expect 32% of the data points deviate from the prediction by more than the error bar), there are some features. Data for the low multipoles $\ell < 30$ are mostly below prediction. This is due to the lack of large-scale power in the northern ecliptic hemisphere, and thus related to the large-scale anomalies. In the range $\ell = 1100$–2000 the data residuals seem to oscillate in opposite phase to the acoustic peak pattern, i.e., the acoustic peaks in the data are slightly smoothed compared to theoretical prediction. Gravitational lensing of the CMB by the large-scale structure causes such smoothing, and it is as if this smoothing effect is some 10–20% larger than predicted by ΛCDM. These features can be fit better with the ΛCDM+Ω_k model with negative Ω_k (closed universe), but the Planck direct measurement of CMB lensing (from higher-order correlations) and other cosmological data disagrees with such a model.

Local vs global Hubble constant. The Planck ΛCDM value for the Hubble constant, $H_0 = 67.36 \pm 0.54\,\mathrm{km/s/Mpc}$ disagrees with distance-ladder measurements, which give, e.g., $H_0 = 74.0 \pm 1.4\,\mathrm{km/s/Mpc}$ [19]. One may think of many possible causes for this discrepancy. 1)

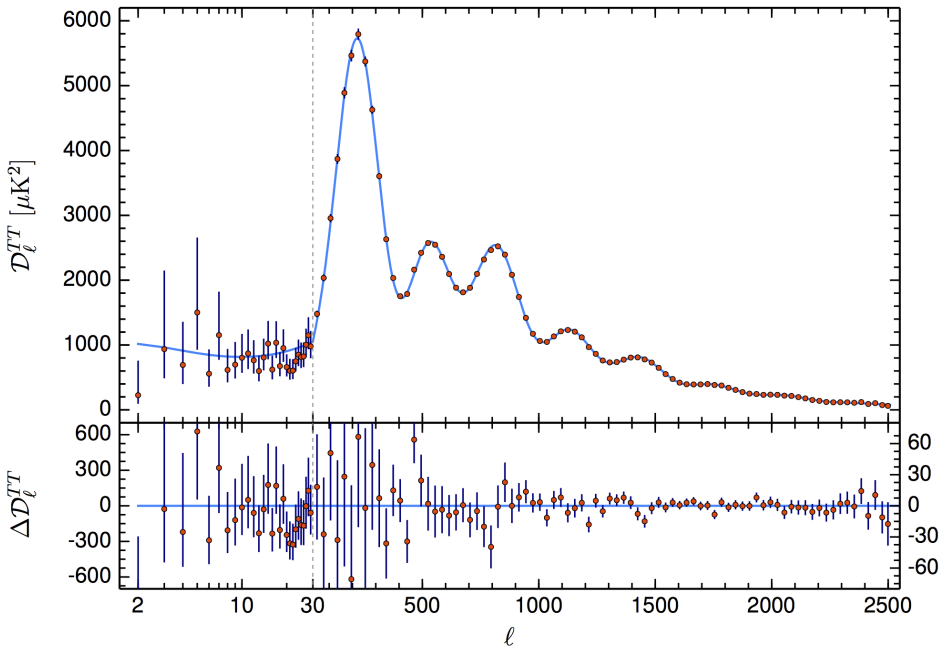


Figure 37: The temperature C_ℓ from Planck 2018 data. Unlike in Fig. 11, here the cosmic variance is included in the error bars, so the data can be more directly compared with theory. The blue curve is the ΛCDM prediction, and the bottom panel shows the difference (data residuals) between data and prediction. Note the different scale (on the right) for residuals at $\ell > 30$, where also the horizontal axis changes from logarithmic to linear. From [5].

There may be underestimated systematic errors in the distance-ladder measurements. 2) Note that the Planck value is not a "measurement" of H_0. It is a result of a six-parameter fit to the data, assuming the ΛCDM model, where H_0 is one of the six parameters. So the discrepancy could point to ΛCDM not being the correct model. One could alleviate the discrepancy by adding extra parameters to the model. (However, e.g., in the ΛCDM+Ω_k model the discrepancy becomes worse.) 3) The distance-ladder measurements are from a "local" part of the universe, $z < 1$, so it represents a local measurement of H_0; whereas the result from the CMB is related to the distance from here to the last scattering surface, so it corresponds to a value of H_0 which is representative of the entire observable Universe. The explanation of the discrepancy could thus be an unexpectedly large inhomogeneity: we live in a large underdense region, which thus expands faster than the Universe on average.

What should one make of such discrepancies? Similar issues are common in the progress of science. More often than not, they go away with improved data; but sometimes they point to something new, which improved data later confirms. For the large scale anomalies, "improved data" will be difficult to get, since here we are limited by cosmic variance. More accurate data on CMB polarization would help: are there similar large-scale anomalies in the polarization or not? The Planck data on polarization was inconclusive on this question [6], since the Planck design was not optimized for polarization measurements, and therefore there are residual systematic effects in the polarization data at large scales, which limits the accuracy.

References

[1] Planck Collaboration, *Planck 2013 results. I. Overview of products and scientific results*, arXiv:1303.5062, Astronomy & Astrophysics **571**, A1 (2014)

B Quantum Fluctuations during Inflation

Subhorizon scales during inflation are microscopic and therefore quantum effects are important. Thus we should study the behavior of the inflaton field using quantum field theory. To warm up we first consider quantum field theory of a scalar field in Minkowski space.

B.1 Vacuum fluctuations in Minkowski space

The field equation for a massive free (i.e. $V(\varphi) = \frac{1}{2}m^2\varphi^2$) real scalar field in Minkowski space is

$$\ddot{\varphi} - \nabla^2\varphi + m^2\varphi = 0\,, \tag{1}$$

or

$$\ddot{\varphi}_{\mathbf{k}} + E_k^2\varphi_{\mathbf{k}} = 0\,, \tag{2}$$

where $E_k^2 = k^2 + m^2$, for Fourier components. We recognize Eq. (2) as the equation for a harmonic oscillator. Thus each Fourier component of the field behaves as an independent harmonic oscillator.

In the quantum mechanical treatment of the harmonic oscillator one introduces the creation and annihilation operators, which raise and lower the energy state of the system. We can do the same here.

Now we have a different pair of creation and annihilation operators $\hat{a}^\dagger_{\mathbf{k}}$, $\hat{a}_{\mathbf{k}}$ for each Fourier mode $\mathbf{k}$. We denote the ground state of the system by $|0\rangle$, and call it the *vacuum*. As discussed earlier, *particles* are quanta of the oscillations of the field. The vacuum is a state with no particles. Operating on the vacuum with the creation operator $\hat{a}^\dagger_{\mathbf{k}}$, we add one quantum with momentum $\mathbf{k}$ and energy E_k to the system, i.e., we create one particle. We denote this state with one particle, whose momentum is $\mathbf{k}$ by $|1_{\mathbf{k}}\rangle$. Thus

$$\hat{a}^\dagger_{\mathbf{k}}|0\rangle = |1_{\mathbf{k}}\rangle\,. \tag{3}$$

This particle has a well-defined momentum $\mathbf{k}$, and therefore it is completely unlocalized (Heisenberg's uncertainty principle). The annihilation operator acting on the vacuum gives zero, i.e., not the vacuum state but the zero element of Hilbert space (the space of all quantum states),

$$\hat{a}_{\mathbf{k}}|0\rangle = 0\,. \tag{4}$$

We denote the hermitian conjugate of the vacuum state by $\langle 0|$. Thus

$$\langle 0|\hat{a}_{\mathbf{k}} = \langle 1_{\mathbf{k}}| \qquad \text{and} \qquad \langle 0|\hat{a}^\dagger_{\mathbf{k}} = 0\,. \tag{5}$$

The commutation relations of the creation and annihilation operators are

$$[\hat{a}^\dagger_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = 0, \qquad [\hat{a}_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = \delta_{\mathbf{k}\mathbf{k}'}\,. \tag{6}$$

When going from classical physics to quantum physics, classical observables are replaced by operators. One can then calculate expectation values for these observables using the operators. Here the classical observable

$$\varphi(t, \mathbf{x}) = \sum \varphi_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}} \tag{7}$$

is replaced by the *field operator*

$$\hat{\varphi}(t, \mathbf{x}) = \sum \hat{\varphi}_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}} \tag{8}$$

where[1]

$$\hat{\varphi}_{\mathbf{k}}(t) = w_k(t)\hat{a}_{\mathbf{k}} + w_k^*(t)\hat{a}^\dagger_{-\mathbf{k}} \tag{9}$$

and

$$w_k(t) = V^{-1/2}\frac{1}{\sqrt{2E_k}}e^{-iE_kt} \tag{10}$$

is the mode function, a normalized solution of the field equation (2). We are using the Heisenberg picture, i.e. we have time-dependent operators; the quantum states are time-independent.

Classically the ground state would be one where $\varphi = \text{const.} = 0$, but we know from the quantum mechanics of a harmonic oscillator that there are oscillations even in the ground state. Likewise, there are fluctuations of the scalar field, *vacuum fluctuations*, even in the vacuum state.

We shall now calculate the *power spectrum* of these vacuum fluctuations. The power spectrum is defined as the expectation value

$$\mathcal{P}_\varphi(k) = V\frac{k^3}{2\pi^2}\langle|\varphi_{\mathbf{k}}|^2\rangle \tag{11}$$

and it gives the variance of $\varphi(\mathbf{x})$ as

$$\langle\varphi(\mathbf{x})^2\rangle = \int_0^\infty \frac{dk}{k}\mathcal{P}_\varphi(k)\,. \tag{12}$$

For the vacuum state $|0\rangle$ the expectation value of $|\varphi_{\mathbf{k}}|^2$ is

$$\begin{aligned}\langle 0|\hat{\varphi}_{\mathbf{k}}\hat{\varphi}^\dagger_{\mathbf{k}}|0\rangle &= |w_k|^2\langle 0|\hat{a}_{\mathbf{k}}\hat{a}^\dagger_{\mathbf{k}}|0\rangle + w_k^2\langle 0|\hat{a}_{\mathbf{k}}\hat{a}_{-\mathbf{k}}|0\rangle + (w_k^*)^2\langle 0|\hat{a}^\dagger_{-\mathbf{k}}\hat{a}^\dagger_{\mathbf{k}}|0\rangle + |w_k|^2\langle 0|\hat{a}^\dagger_{-\mathbf{k}}\hat{a}_{-\mathbf{k}}|0\rangle \\ &= |w_k|^2\langle 1_{\mathbf{k}}|1_{\mathbf{k}}\rangle \;=\; |w_k|^2\,,\end{aligned} \tag{13}$$

since all but the first term give 0, and our states are normalized so that $\langle 1_{\mathbf{k}}|1_{\mathbf{k}'}\rangle = \delta_{\mathbf{k}\mathbf{k}'}$. From Eq. (10) we have $|w_k|^2 = 1/(2VE_k)$. Our main result is that

$$\mathcal{P}_\varphi(k) = V\frac{k^3}{2\pi^2}|w_k|^2 \tag{14}$$

for vacuum fluctuations, which we shall now apply to inflation, where the mode functions $w_k(t)$ are different.

B.2 Vacuum fluctuations during inflation

During inflation the field equation (for inflaton perturbations) is, from Sec. 8.6,

$$\delta\ddot{\varphi}_{\mathbf{k}} + 3H\delta\dot{\varphi}_{\mathbf{k}} + \left[\left(\frac{k}{a}\right)^2 + V''(\bar{\varphi})\right]\delta\varphi_{\mathbf{k}} = 0\,. \tag{15}$$

There are oscillations only in the perturbation $\delta\varphi$, the background $\bar{\varphi}$ is homogeneous and evolving slowly in time. For the particle point of view, the background solution represents the vacuum,[2] i.e., particles are quanta of oscillations around that value.

[1] We skip the detailed derivation of the field operator, which belongs to a course of quantum field theory. See, e.g., Peskin & Schroeder, section 2.3 (note different normalizations of operators and states, related to doing Fourier integrals rather than sums and considerations of Lorentz invariance.)

[2] This is not the vacuum state in the sense of being the ground state of the system. The true ground state has $\bar{\varphi}$ at the minimum of the potential. However there are no particles related to the background evolution $\bar{\varphi}(t)$.

After making the approximations $H = \text{const.}$ and

$$\frac{V''}{H^2} = 3\eta \approx 0 \tag{16}$$

we found that the two independent solutions for $\delta\varphi_{\mathbf{k}}(t)$ are

$$w_k(t) \;=\; V^{-1/2}\frac{H}{\sqrt{2k^3}}\left(i + \frac{k}{aH}\right)\exp\left(\frac{ik}{aH}\right) \tag{17}$$

and its complex conjugate $w_k^*(t)$, where the time dependence is in $a = a(t) \propto e^{Ht}$. The factor $V^{-1/2}H/\sqrt{2k^3}$ is here for normalization purposes ($V = L^3$ being the reference volume, not the inflaton potential).

When the scale k is well inside the horizon, $k \gg aH$, $\delta\varphi_{\mathbf{k}}(t)$ oscillates rapidly compared to the Hubble time H^{-1}. If we consider distance and time scales much smaller than the Hubble scale, spacetime curvature does not matter and things should behave like in Minkowski space. Considering Eq. (17) in this limit, one finds (**exercise**) that $w_k(t)$ indeed becomes equal to the Minkowski space mode function, Eq. (10). (We cleverly chose the normalization in Eq. (17) so that the normalizations would agree.) Therefore the $w_k(t)$ of Eq. (17) is our mode function. We can use it to follow the evolution of the mode functions as the scale approaches and exits the horizon.

The field operator for the inflaton perturbations is

$$\delta\hat{\varphi}_{\mathbf{k}}(t) = w_k(t)\hat{a}_{\mathbf{k}} + w_k^*(t)\hat{a}^\dagger_{-\mathbf{k}}\,, \tag{18}$$

and the power spectrum of inflaton fluctuations is

$$\mathcal{P}_\varphi(k) = V\frac{k^3}{2\pi^2}|w_k|^2\,. \tag{19}$$

Well before horizon exit, $k \gg aH$, observed during timescales $\ll H^{-1}$, the field operator $\delta\hat{\varphi}_{\mathbf{k}}(t)$ becomes the Minkowski space field operator and we have standard vacuum fluctuations in $\delta\varphi$.

Well after horizon exit, $k \ll aH$, the mode function becomes a constant

$$w_k(t) \;\to\; V^{-1/2}\frac{iH}{\sqrt{2k^3}}\,, \tag{20}$$

the vacuum fluctuations "freeze", and the power spectrum acquires the constant value

$$\mathcal{P}_\varphi(k) \;=\; V\frac{k^3}{2\pi^2}|w_k|^2 \;=\; \left(\frac{H}{2\pi}\right)^2\,. \tag{21}$$