Lecture notes: Cosmology

Luca Amendola
University of Heidelberg
l.amendola@thphys.uni-heidelberg.de
http://www.thphys.uni-heidelberg.de/~amendola/teaching.html
v. 5.11

January 31, 2023

Contents

[a] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.
[c] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.
[d] This section follows closely the treatment of Prof. Bartelmann's lecture notes.
[e] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.
[f] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

Acknowledgments and credits

This course is addressed to master students; there are no special pre-requisites although often we will make use of concepts from General Relativity and some basic astronomy. All the concepts will be introduced in a self-consistent way but clearly the student will benefit a lot by reading the relevant chapters in the following texts and in astrophysics textbooks.

Suggested readings:
S. Dodelson, *Modern Cosmology*, Academic Press (my favourite)
L. Amendola & S. Tsujikawa, *Dark Energy. Theory and Observations*, CUP (more advanced material)
D. Baumann, *Cosmology* (cmb.wintherscoming.no/pdfs/baumann.pdf)
O. Piattella, *Lecture Notes in Cosmology*, Springer (recent and complete)
M. Bartelmann, *Observing the Big Bang*, Lecture notes
M. Bartelmann, *Cosmology*, Lecture notes
For a quick (24 pages) but complete introduction to GR:
S. Carroll, *A no-nonsense introduction to GR*, preposterousuniverse.com/wp-content/uploads/2015/08/grtinypdf.pdf
More specialized texts:
N. Sugiyama, *Introduction to temperature anisotropies of CMB*, PTEP,2014, 06B101
D. Weinberg et al., *Observational probes of cosmic acceleration*, arXiv:1201.2434

Part I

The homogeneous Universe

Chapter 1

A short history of cosmology

- In 1917, Einstein publishes the first cosmological model, based on the introduction of the cosmological constant and on assuming homogeneity and isotropy. Einstein's model was static (later it was shown however that this static model is unstable).
- in 1918, de Sitter shows that a Universe dominated by the cosmological constant would be expanding.
- in 1922, Alexander Friedmann solves Einstein's cosmological equations with and without the cosmological constant, showing they generically allow for a dynamic Universe (i.e. expanding or contracting) obeying what was later called Hubble's law. In 1927, Lemaitre discovers independently this general cosmological model and makes it well known among the community. Lemaitre was the first to actually formulate Hubble law explicitly and evaluate the Hubble constant from the then available data.
- in 1929 Edwin Hubble discovers the cosmic expansion obeying Hubble's law, after several years of pioneristic work by himself and by several other scientists: Milton Humason, Henrietta Leavitt, Vesto Slipher, and others.
- Hubble constant, whose inverse gives the time scale for the expansion, was found to be around 600 km/sec, almost ten times larger than the currently accepted value. With this constant, the Universe would be a couple billion years old, too short to allow for star evolution.
- In the 30s, Fritz Zwicky postulates the existence of a large component of dark matter to explain the velocities of the galaxis within the Coma cluster.
- After the second WW, Gamow and collaborators investigate the physics of a hot big-bang Universe and formulate definite predictions about primordial nucleosynthesis and the cosmic microwave background.
- In 1965, Penzias and Wilson discover the 3K cosmic microwave background radiation (CMB), interpreted by R. Dicke and collaborators as the relic of the hot primordial phase. This practically ruled out the alternative "steady-state" cosmological model proposed by Hoyle and colaborators in the 50s.
- In the same years, the first precise calculations of the abundance of light nuclei formed during the first minutes after the big bang are found to be consistent with observations, lending further strong support to the big bang model.
- During the 70s, strong evidence for the existence of dark matter assembled in extended spherical halos around galaxies begins to build up, after the work by Rubin, Ford, Bosma, and several others.
- During the 80s, this dark matter component becomes explained in terms of elementary particles rather than as "not yet seen" stars or gas. DM candidates should be stable, neutral, abundant. Neutrinos are the first candidates, but they are soon ruled out because too hot and too light.
- Supersymmetry, a general theory of elementary particles elaborated in the 70s, opens up the possibility of many new unseen particles and the hypothesis was then advanced that DM is the lightest supersymmetric partner. This is still today one of the main models of DM particles, often refered to as WIMP, weakly

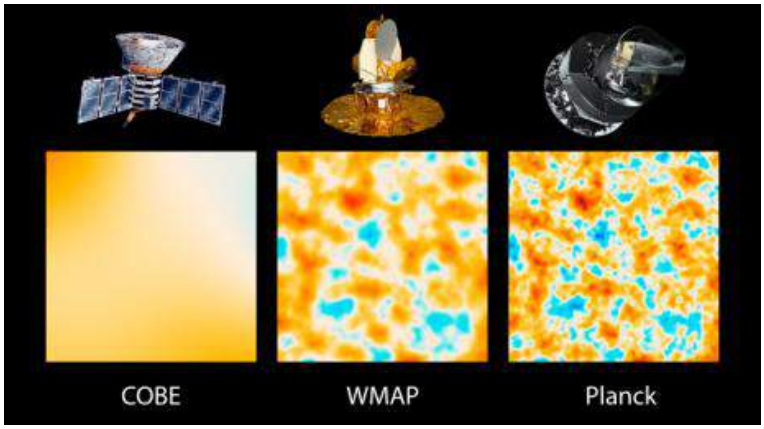


Figure 1.0.1: A patch of 10 square degrees on the CMB sky as seen by COBE. WMAP and Planck (left to right). (NASA/JPL-Caltech/ESA)

interacting massive particles. If the WIMPs interact only weakly, then their abundance is predicted to be close to the observed values if their mass is around 100 GeV (the so-called WIMP-miracle). Unfortunately, so far no particle of this type has been detected.

- In 1981, Alan Guth, after similar work by A. Starobinsky and other precursors, proposes that an epoch of accelerated expansion took place in the very early Universe, the so-called inflationary model. This model predicted a spatially flat Universe.
- Invented to solve the paradoxes of the horizon and of the flatness, the inflationary universe is rapidly found to contain a quantum mechanism to generate initial fluctuations at all scales.
- In 1992, the COBE satellite finds the anisotropies of the CMB. They are in agreement with the existence of dark matter and with the inflationary paradigm.
- In 1999, two groups lead by Perlmutter, Schmidt and Riess, discover the acceleration of the cosmic expansion, by studying distant supernovae. They explain it by re-introducing Einstein's cosmological constant.
- In 2000, the Boomerang ballon experiment finds the first acoustic peak in CMB temperatre anisotropies. Its position measures the spatial curvature of the universe, and find it in agreement with inflationary predictions. The satellite WMAP first, and Planck later, confirm and extend spectacularly the agreement of the CMB spectrum with the so-called standard model of cosmology, ΛCDM plus inflation.

Chapter 2

Introduction to Relativity

Quick summary

- This chapter recalls concepts of Special and General Relativity. The readers might skip it, at the cost of accepting as a given the cosmological Friedmann equations introduced in the next chapter, and the equations of perturbations that will be discussed later on.
- Special relativity is based on a generalization of the concept of distance to four dimension (three spatial plus one time dimension). This generalized distance between two events is called *Minkowski metric* and describes a flat space-time geometry.
- General relativity further generalizes the Minkowski metric to describe intervals between events in a curved space-time.
- Particles propagate along lines (geodesics) that extremize the space-time interval.
- If we assume the Universe to be homogeneous and isotropic, we find that the metric has a simple form, called Friedmann-Robertson-Lemaitre-Walker (FLRW) metric. The FLRW metric depends on a function of time $a(t)$ called scale factor and on a parameter k that, after a rescaling of coordinates, can be taken to be 0 or ± 1.
- These three values define the only three possible three-dimensional homogeneous and isotropic spatial geometries, namely flat space, spherical space ($k = 1$) and hyperbolic space ($k = -1$).
- The metric obeys the Einstein GR equations. These differential equations depend on the metric and on the energy-momentum tensor that describes the properties of matter.
- Once we solve the Einstein equations for the FLRW metric, we obtain the cosmological Friedmann equations that govern the dynamics of the space-time expansion, to be discussed in the next chapter.

2.1 Special relativity

Special relativity is based on the assumption (experimentally tested with great precision) that the space-time interval

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 \tag{2.1.1}$$

is invariant under Lorentz transformations, which generalize the inertial transformations of Galileo. These transformations are defined by the general laws ($\mathbf{y} = (ct', x', y', z')$ = new coordinates; $\mathbf{x} = (ct, x, y, z)$ = old coordinates)

$$y^\alpha = \Lambda^\alpha_\beta x^\beta + a^\alpha \tag{2.1.2}$$

where Λ^α_β and a^α are *constant*s. Taking differentials, we obtain

$$dy^\alpha = \Lambda^\alpha_\beta dx^\beta \tag{2.1.3}$$

Greek indices run over $0, 1, 2, 3$; the Latin indices i, j, k, over the space coordinates $1, 2, 3$; repeated indices imply sum, i.e. $\Lambda^\alpha_\beta x^\beta \equiv \sum_\beta \Lambda^\alpha_\beta x^\beta$. The Kronecker symbol δ^β_α indicates the identity matrix.

In order for the ds^2 to be invariant, the matrix Λ^α_β must be subject to the relation

$$\Lambda^\alpha_\gamma \Lambda^\beta_\delta \eta_{\alpha\beta} = \eta_{\gamma\delta} \tag{2.1.4}$$

where we have introduced the *Minkowski metric*

$$\eta_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

In fact, the interval (2.1.1) can also be written as

$$ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta \tag{2.1.5}$$

or

$$ds^2 = (cdt, dx, dy, dz) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} cdt \\ dx \\ dy \\ dz \end{pmatrix} \tag{2.1.6}$$

Replacing (2.1.3) in $ds^2 = \eta_{\alpha\beta} dy^\alpha dy^\beta$ and using (2.1.4), one sees immediately that ds^2 does not change.

Evaluating the determinant of eq. 2.1.4 we see that $(\det \Lambda)^2 = 1$. We restrict ourselves now to the subgroup of Lorentz

$$\Lambda^0_0 \geq 0 \tag{2.1.7}$$

$$\det \Lambda = +1 \tag{2.1.8}$$

This subgroup, called *proper*, contains the identity transformation and can therefore be generated through continuous transformations from an initial state. Another subgroup consists of the rotations, $\Lambda^\alpha_\beta = 0$ except $\Lambda^i_j = R_{ij}$, where R is an orthonormal matrix) and of space-time translations, $y^\alpha = x^\alpha + a^\alpha$. These roto-translations do not differ from Galilean transformations and are of no interest for relativity. The relativistic transformations are those that involve the speed of an observer respect to another (*boosts*).

We use now units such that $c = 1$, i.e. we measure the speed in units of the speed of light. To an observer B moving along the x axis with speed v_1 with respect to an observer A we have

$$dx' = \Lambda^1_0 dt + \Lambda^1_1 dx \tag{2.1.9}$$

$$dt' = \Lambda^0_0 dt + \Lambda^0_1 dx \tag{2.1.10}$$

while we assume that the coordinates y, z remain unvaried (and therefore $\Lambda^2_2 = \Lambda^3_3 = 1$ and all other component vanish). Let us define the velocity v_1 as the one measured by A when B is just passing it (and therefore $-v_1$ is A's speed measured by B). The origin of the B frame has equation $x' = 0$ so $dx' = 0$. In the A frame, the trajectory of B has equation $x = v_1 t$ so $v_1 = dx/dt$; together with $dx' = 0$ we have $\Lambda^1_1 dx = -\Lambda^1_0 dt$ and therefore $v_1 = dx/dt = -\Lambda^1_0/\Lambda^1_1$. Inverting the roles of A and B, we find analogously that when $dx = 0$, B will measure $v_1 = -dx'/dt' = -\Lambda^1_0/\Lambda^0_0$. If we put $\Lambda^0_0 \equiv \gamma$ we can then write the general Lorentz transformation from A to B as

$$\Lambda^\alpha_\beta = \begin{pmatrix} \gamma & -\gamma v_1 & 0 & 0 \\ -\gamma v_1 & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.1.11}$$

The unknown γ can now be determined by requiring that $\det \Lambda = 1$, from which

$$\gamma = (1 - v^2)^{-1/2} \tag{2.1.12}$$

The generalization to an observer with velocity $\mathbf{v} = (v_1, v_2, v_3)$ is

$$\Lambda^\alpha_\beta = \begin{pmatrix} \gamma & -\gamma v_1 & -\gamma v_2 & -\gamma v_3 \\ -\gamma v_1 & & & \\ -\gamma v_2 & & \Lambda^i_j & \\ -\gamma v_3 & & & \end{pmatrix} \tag{2.1.13}$$

where

$$\Lambda^i_j = \delta_{ij} + \frac{v_i v_j}{v^2}(\gamma - 1) \tag{2.1.14}$$

The most famous result of Lorentz transformations is the time dilation. Consider an observer at rest measuring the ticking of a clock also at rest with respect to A. She will measure an interval $ds = dt$ (remember $c = 1$). A second observer$B$ moving with velocity v along the x axis will instead measure for the same clock the interval $ds^2 = dt'^2 - dx'^2 = dt'^2(1 - v^2)$. But by Lorentz invariance the two intervals must be equal. We have therefore

$$\gamma dt = dt'$$

that is, B will see the ticking at intervals greater than A (γ is greater than 1).

Relativistic mechanics can be deduced from the action. For a free particle we have

$$A = -mc \int ds \tag{2.1.15}$$

To first order ($v \ll c$ and putting for simplicity dr^2 in place of $dx^2 + dy^2 + dz^2$, that is by considering a radial motion) we have

$$ds = cdt\sqrt{1 - \frac{dr^2}{c^2 dt^2}} = cdt\sqrt{1 - \frac{v^2}{c^2}} \approx cdt(1 - \frac{1}{2}\frac{v^2}{c^2}) \tag{2.1.16}$$

and we obtain the non-relativistic action

$$A = -mc \int cdt\sqrt{1 - \frac{dr^2}{c^2 dt^2}} = -mc^2 \int dt + \int \frac{1}{2} m v^2 dt \tag{2.1.17}$$

Light follows the path $ds = 0$, i.e. propagates with constant velocity dr/dt=c. Since ds is invariant, the light has the same speed c with respect to all Lorentz observers.

The metric $\eta_{\alpha\beta}$ is clearly symmetric, $\eta_{\alpha\beta} = \eta_{\beta\alpha}$, since ds^2 does not change by inverting the indices α, β. Moreover we have

$$\eta_{\alpha\beta}\eta^{\beta\gamma} = \delta^\gamma_\alpha$$

where δ^γ_α, the Kronecker symbol, is the identity matrix, $\delta^\beta_\alpha =$ diag$(1, 1, 1, 1)$.

The interval ds, being an invariant under Lorentz transformations, is a *scalar*. From its invariance with respect to Lorentz transformations it follows that

$$\eta_{\alpha\beta} dx^\alpha dx^\beta = \eta'_{\mu\nu} dy^\mu dy^\nu \tag{2.1.18}$$

Let us define now *vectors and tensors*. The most important vector is the differential dx^α. Under a coordinate change $y^\mu = y^\mu(x^\nu)$ we have clearly

$$dy^\alpha = \frac{\partial y^\alpha}{\partial x^\beta} dx^\beta \tag{2.1.19}$$

This transformation law is called *contravariant* . A contravariant vector (upper indices) is any quantity that transforms in this way. We can also define a *covariant* vector (lower indices):

$$dy_\alpha \equiv \eta_{\alpha\beta} dy^\beta \tag{2.1.20}$$

We see then that $dy_\alpha dy^\alpha = dx_\alpha dx^\alpha$ is a scalar, that is it does not change under a general transformation, and therefore dy_α transforms in the opposite way compared to contravariant vectors (note that we have used the identity $\frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial x^\nu}{\partial y^\alpha} = \delta^\nu_\mu$). We say that the metric $\eta_{\alpha\beta}$ "lowers the indices". Similarly, the contravariant metric $\eta^{\alpha\beta}$ can be used to raise the indices. Another fundamental vector is the 4-velocity

$$u^\mu \equiv \frac{dx^\mu}{ds}$$

In the limit of negligible velocities we have from (2.1.16) $ds = cdt$ and thus $u^\mu = (1, 0, 0, 0)$.

The fundamental tensor is obviously the metric . From the invariance of ds^2 it follows the transformation law $\eta_{\alpha\beta}$

$$\eta'_{\alpha\beta} = \eta_{\mu\nu}\frac{\partial x^\mu \partial x^\nu}{\partial y^\alpha \partial y^\beta} \tag{2.1.21}$$

The general rule is then that a tensor with n lower indices and m upper indices transforms through n terms of type $\partial x^\mu/\partial y^\nu$ and m of type $\partial y^\mu/\partial x^\nu$ (where y are the "new" coordinates and x the "old" ones).

The importance of the tensor notation is that it makes readily apparent the fundamental property of relativistic equations: the invariance under Lorentz transformations. It is sufficient to write equations with equal indices left and right to make them automatically Lorentz-covariant. For example the equation

$$ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta$$

is Lorentz-invariant, as it is also $ds = 0$ (obviously a zero is a Lorentz invariant) or $u_\nu u^\nu = 1$.

2.2 Metric and gravitation

The Lorentz transformations are a very small group. Their generalization is the basis of general relativity.

Consider the Minkowski metric

$$ds^2 = \eta_{\mu\nu} dy^\mu dy^\nu$$

and perform a general coordinate transformation

$$y^\alpha = y^\alpha(x^\mu) \tag{2.2.1}$$

We obtain

$$ds^2 = \left[\eta_{\mu\nu}\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\nu}{\partial x^\beta}\right] dx^\alpha dx^\beta \equiv g_{\alpha\beta} dx^\alpha dx^\beta \tag{2.2.2}$$

If the new reference system is non-inertial (e.g., accelerated), then $\partial y^\alpha/\partial x^\beta \neq const$ and the new metric $g_{\alpha\beta}$ is different from the original. The equivalence principle says that every gravitational field can be described, locally, by a metric obtained by a transformation to a non-inertial reference. This reflects the famous elevator *gedanken* experiment: in an elevator freely falling on Earth, the dynamics of bodies is the same as for inertial observers, i.e. as if no gravitational force were present. That is, gravity is indistinguishable, locally, from a general trasformation of coordinates (the accelerated elevator). General Relativity is based on the assumption that any gravitational field can be described, overall, by a general metric $g_{\mu\nu}$. Since a metric is described by 10 independent functions, while the non-inertial transformations are only 4, it is clear that in general a gravitational field can not be described in a global manner by a non-inertial transformation.

Let's make an example. The action of a special-relativistic particle is $A = -mc\int ds$ where $ds = cdt(1 - v^2/c^2)^{1/2}$. In presence of a gravitational potential it becomes

$$A = -m\int c^2 dt\sqrt{1 - \frac{v^2}{c^2}} - m\int \Phi dt \tag{2.2.3}$$

and its equation of motion for $v \ll c$ is

$$\dot{v} = -\nabla\Phi \tag{2.2.4}$$

(Note that the Newtonian potential generated by a massive point is negative, $\Phi = -GM/r$, so that $\dot{v} < 0$, as in falling motion) .

Exercise: evaluate Φ *on Earth* ($M = 6{\cdot}10^{24}kg$, $R \approx 6{\cdot}10^{3}km$) *and on the Sun* ($M = 2{\cdot}10^{30}kg$, $R = 7{\cdot}10^{5}km$). *Observe that* $\Phi \ll 1$ *in both cases* ($G = 6.7 \cdot 10^{-11} m^3 kg^{-1} s^{-2}$).

We can rewrite the action as

$$\begin{aligned} -m\int c^2 dt\sqrt{1-\frac{v^2}{c^2}} - m\int \Phi dt &= -mc\int cdt(\sqrt{1-\frac{v^2}{c^2}} + \frac{\Phi}{c^2}) \\ &\approx -mc\int cdt(1-\frac{v^2}{c^2}+\frac{2\Phi}{c^2})^{1/2} \\ &= -mc\int (c^2dt^2(1+\frac{2\Phi}{c^2}) - dr^2)^{1/2} = -mc\int ds' \end{aligned} \tag{2.2.5}$$

where now

$$ds'^2 = c^2dt^2(1+\frac{2\Phi}{c^2}) - dr^2 \tag{2.2.6}$$

is the space-time interval of a non-Minkowskian metric. The force was then absorbed in the definition of new metrics.

Now we consider again (2.2.1). From SR we know that the equation of motion of a inertial particle with coordinates $y^\mu = (ct, x, y, z)$ in a Minkowskian reference frame $ds^2 = \eta_{\mu\nu}dy^\mu dy^\nu = g_{\alpha\beta}dx^\alpha dx^\beta$ is

$$\frac{d^2y^\mu}{ds^2} = 0$$

Under a general transformation we have

$$dy^\mu = \frac{\partial y^\mu}{\partial x^\nu}dx^\nu$$

or, replacing,

$$\frac{d^2y^\mu}{ds^2} = \frac{d}{ds}\frac{dy^\mu}{ds} = \frac{d}{ds}\left(\frac{\partial y^\mu}{\partial x^\nu}\frac{dx^\nu}{ds}\right) = \frac{dx^\nu}{ds}\left(\frac{d}{ds}\frac{\partial y^\mu}{\partial x^\nu}\right) + \frac{\partial y^\mu}{\partial x^\nu}\frac{d^2x^\nu}{ds^2} = 0 \tag{2.2.7}$$

The first term is

$$\frac{dx^\nu}{ds}\left(\frac{d}{ds}\frac{\partial y^\mu}{\partial x^\nu}\right) = \frac{dx^\nu dx^\beta}{ds^2}\frac{\partial}{\partial x^\beta}\frac{\partial y^\mu}{\partial x^\nu} = \frac{dx^\nu dx^\beta}{ds^2}\frac{\partial^2 y^\mu}{\partial x^\beta \partial x^\nu} \tag{2.2.8}$$

from which we can multiply the last equation of (2.2.7) by $\partial x^\tau/\partial y^\mu$

$$\frac{d^2x^\tau}{ds^2} + \frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}\Gamma^\tau_{\alpha\beta} = 0 \tag{2.2.9}$$

where we defined the Christoffel symbols

$$\Gamma^\tau_{\alpha\beta} = \frac{\partial^2 y^\mu}{\partial x^\alpha \partial x^\beta}\frac{\partial x^\tau}{\partial y^\mu} \tag{2.2.10}$$

Eq. (2.2.9) is the motion equation in the transformed system. Since GR interprets each transformation as a non-inertial gravitational field, this equation tells us how a particle moves in a field described by the general metric $g_{\mu\nu}$. Here and in other similar calculations, it is useful to note the two identities

$$\frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial y^\mu}{\partial x^\beta} = \delta^\alpha_\beta \tag{2.2.11}$$

$$\frac{\partial x^\alpha}{\partial x^\beta} = \delta^\alpha_\beta \tag{2.2.12}$$

Many of the properties already described for the Minkowskian metric also apply to the general metric $g_{\mu\nu}$. We have in fact that $g_{\mu\nu}$ is symmetric and $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$. In addition, the metric also has the function of " contracting" indices: given a tensor $T^{\mu\nu}$ one has

$$g_{\mu\nu}T^{\mu\nu} = T^{\mu}_{\mu} = T$$

i.e. the trace of $T^{\mu\nu}$. The inverse of the metric is the contravariant metric, $g^{\mu\nu} = (g_{\mu\nu})^{-1}$. In fact, $dx_\alpha = \delta^\beta_\alpha dx_\beta$ but also, by definition, $dx_\alpha = g_{\alpha\nu}g^{\nu\beta}dx_\beta$, from which we see that

$$g^{\beta\nu}g_{\nu\alpha} = \delta^\beta_\alpha \tag{2.2.13}$$

Then the transformation law (2.1.21) becomes

$$g'_{\mu\nu}(x) = g_{\alpha\beta}(y)\frac{\partial y^\alpha \partial y^\beta}{\partial x^\mu \partial x^\nu} \tag{2.2.14}$$

Let us differentiate now $g'_{\mu\nu}$ with respect to x^λ. We obtain ($g_{\alpha\beta}$ does not depend on x)

$$g'_{\mu\nu,\lambda} \equiv \frac{\partial g'_{\mu\nu}}{\partial x^\lambda} = g_{\alpha\beta}\frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial}{\partial x^\lambda}\frac{\partial y^\beta}{\partial x^\nu} + g_{\alpha\beta}\frac{\partial y^\beta}{\partial x^\nu}\frac{\partial}{\partial x^\lambda}\frac{\partial y^\alpha}{\partial x^\mu} \tag{2.2.15}$$

where we have introduced the comma notation to indicate the derivation. Substituting again the metric with the transformed one (and by removing the apex) one obtains

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \Gamma^\alpha_{\lambda\mu}g_{\alpha\nu} + \Gamma^\beta_{\lambda\nu}g_{\beta\mu} \tag{2.2.16}$$

Rewriting the equation (2.2.16) and exchanging first λ, μ and then λ, ν, and then multiplying by $g^{\sigma\mu}$, and finally by combining the three equations we can see that (Exercise: prove by replacement!)

$$\Gamma^\gamma_{\alpha\beta} = \frac{1}{2}g^{\gamma\eta}\left(g_{\alpha\eta,\beta} + g_{\beta\eta,\alpha} - g_{\alpha\beta,\eta}\right) \tag{2.2.17}$$

Then, the metric completely determines, through the Christoffel symbols, the geometric and dynamic properties of spacetime. This statement is the essence of General Relativity.

Completing the example above, we now see that Eq. of motion (2.2.4) is precisely of the form (2.2.9) in the metric (2.2.6). In fact we have that the only nonzero term is $\Gamma^i_{00} = \frac{1}{2}\nabla_i g_{00}$ and therefore (for small velocities, i.e. putting $ds \approx dt$)

$$\ddot{x} = -\frac{c^2}{2}\nabla g_{00} = -\nabla\Phi \tag{2.2.18}$$

2.3 Covariant derivative

We have seen that ds is invariant under general coordinate transformations, and therefore is a scalar. We introduce now GR vectors and tensors.

As before, we define the four-velocity $u^\mu = dx^\mu/ds$. As already seen, its transformation law is the same as for the coordinates,

$$\begin{aligned} dy^\mu &= \frac{\partial y^\mu}{\partial x^\nu}dx^\nu \\ u'^\mu &= \frac{\partial y^\mu}{\partial x^\nu}u^\nu \end{aligned}$$

You can see that $dy_\mu \equiv g_{\mu\nu}dy^\nu$ transforms in the opposite way. The metric can therefore be used to lower and raise indices. Since by definition the scalar product of the four-velocity is a scalar (ie invariant)

$$u^\mu u_\mu = 1 \tag{2.3.1}$$

it follows that the four-velocity u^μ is a vector (contravariant).

The eq. (2.2.9) can be written also as

$$\frac{d}{ds}u^\mu + \Gamma^\mu_{\alpha\nu}u^\alpha u^\nu = u^\nu(\frac{\partial}{\partial x^\nu}u^\mu + \Gamma^\mu_{\alpha\nu}u^\alpha) = u^\nu u^\mu_{;\nu} = 0 \tag{2.3.2}$$

where we defined

$$u^\mu_{;\nu} \equiv u^\mu_{,\nu} + \Gamma^\mu_{\alpha\nu}u^\alpha \tag{2.3.3}$$

This equation is valid in all frames of reference, because the transformation that we performed to obtain it is quite general. But u^ν is a vector. Therefore, $u^\mu_{;\nu}$ must be a tensor, i.e. it must transform in such a way to make the whole combination $u^\nu u^\mu_{;\nu}$ a vector. The "semicolon" derivative defines the *covariant derivative*, i.e. the proper way to take derivatives of a vector and generate a tensor. Intuitively, the extra piece in the covariant derivative is necessary because when we differentiate vectors in a curved space, we need to take into account both the change in the vector coordinates, and the change in the frame, or equivalently in the vector basis.

The metric $g_{\mu\nu}$ is obviously a tensor, since it obeys the invariant law $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$. The covariant derivative of a tensor can be obtained by differentiating a generic tensor product of two vectors

$$T^{\mu\nu}_{;\alpha} = (V^\mu U^\nu)_{;\alpha} = V^\mu U^\nu_{;\alpha} + V^\mu_{;\alpha}U^\nu = T^{\mu\nu}_{,\alpha} + \Gamma^\mu_{\beta\alpha}T^{\beta\nu} + \Gamma^\nu_{\beta\alpha}T^{\mu\beta} \tag{2.3.4}$$

and similarly

$$T_{\mu\nu;\alpha} = T_{\mu\nu,\alpha} - \Gamma^\beta_{\alpha\mu}T_{\beta\nu} - \Gamma^\beta_{\alpha\nu}T_{\mu\beta} \tag{2.3.5}$$

From (2.2.16) the fundamental rule follows

$$g^{\mu\nu}_{;\lambda} = 0$$

Another very useful rule is the derivative of the determinant $g \equiv \det g_{\mu\nu}$. The inverse of the metric tensor can be written as $g^{\mu\nu} = M^{(\mu\nu)}/g$ where g is the determinant $g \equiv \det g_{\mu\nu}$ and $M^{(\mu\nu)}$ is the cofactor (determinant of the matrix $g_{\mu\nu}$ obtained by removing the row and column μ, ν, times $(-1)^{\mu+\nu}$). Therefore we have (notice that $M^{(\mu\nu)}$ does not depend on $g_{\mu\nu}$)

$$\frac{dg^{\mu\nu}}{dg_{\mu\nu}} = -M^{(\mu\nu)}\frac{dg}{g^2 dg_{\mu\nu}} = -g^{\mu\nu}\frac{dg}{g dg_{\mu\nu}} \tag{2.3.6}$$

and therefore

$$dg = -gg_{\mu\nu}dg^{\mu\nu} = gg^{\mu\nu}dg_{\mu\nu} \tag{2.3.7}$$

(the last step can be obtained by starting with $g_{\mu\nu} = M_{(\mu\nu)}/g^{-1}$, where g^{-1} is the determinant of $g^{\mu\nu}$) . Now we can derive $\partial g/\partial x^\mu$ and show that

$$\Gamma^\alpha_{\beta\alpha} = [\log\sqrt{(-g)}]_{,\beta} \tag{2.3.8}$$

Since only equations formed by tensors of the same rank and position indices on both sides are valid in all frames of reference, it follows that all the equations of general relativity must be generally covariant. Since they also have to be reduced to the special relativity when the metric is Minkowskian, the simplest generalization to GR consist in replacing ordinary derivatives with covariant derivatives in all equations of dynamics.

2.4 The FLRW metric

We derive now the metric of a homogeneous and isotropic space. The most general metric can be described as follows

$$ds^2 = g_{00}dt^2 + 2g_{0i}dx^i dt - \sigma_{ij}dx^i dx^j \tag{2.4.1}$$

We impose now some simple assumptions:

1) isotropy (note that the g_{0i} is a space vector, i.e. transforms as a vector under transformations of spatial coordinates: it should therefore be zero, otherwise it would introduce a privileged direction)

$$g_{0i} = 0$$

2) redefinition of time (synchronization)

$$d\tau = \sqrt{g_{00}}dt \rightarrow g_{00} = 1$$

We have now (employing t instead of τ)

$$ds^2 = dt^2 - \sigma_{ij}dx^i dx^j$$

Now we seek the metric that describes a three-dimensional spherical hypersurface immersed in a four-dimensional Euclidean space, analogous to the bidimensional surface of a sphere embedded in a 3D world. The properties of this hypersurface will obviously be the same for every point belonging to it. Therefore we require that the 3D space satisfies the condition of three-dimensional "sphericalness"

$$a^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 \tag{2.4.2}$$

We introduce the 4-dimensional spherical coordinates ($\theta = 0$ defines the equatorial plane)

$$x_1 = a\cos\chi\sin\theta\sin\phi \tag{2.4.3}$$
$$x_2 = a\cos\chi\cos\theta \tag{2.4.4}$$
$$x_3 = a\cos\chi\sin\theta\cos\phi \tag{2.4.5}$$
$$x_4 = a\sin\chi \tag{2.4.6}$$

Differentiating (2.4.2) we have

$$x_4 dx_4 = -(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)$$

from which

$$\begin{aligned} ds^2 &= dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 \\ &= dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)^2}{x_4^2} \\ &= a^2(d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)) \end{aligned} \tag{2.4.7}$$

We define now the coordinate r by $\sin\chi = r$ and therefore $d\chi = \lambda dr$, with

$$\lambda = \frac{1}{\sqrt{1-r^2}} \tag{2.4.8}$$

We can now generalize this to a more general line element (whose homogeneity however is not as obvious as in the spherical case)

$$a^2 = x_1^2 + x_2^2 + x_3^2 + kx_4^2 \tag{2.4.9}$$

We obtain then

$$ds_3^2 = a^2(d\chi^2 + F(\chi)(d\theta^2 + \sin^2\theta d\phi^2)) \tag{2.4.10}$$

where

$$F(\chi) = \begin{array}{cc} \sin\chi & k = 1 \\ \chi & k = 0 \\ \sinh\chi & k = -1 \end{array} \tag{2.4.11}$$

and

$$\lambda = \frac{1}{\sqrt{1-kr^2}} \tag{2.4.12}$$

The three values of k produce the only three homogenous and isotropic 3D spatial metrics. They are is called the Friedmann-Lemaître- Robertson-Walker metric(s)

$$ds^2 = dt^2 - a^2(t)[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)] \tag{2.4.13}$$

The constant k can take any value, but we can actually absorb $|k|$ in a redefinition of r, so from now on we can consider only three separate cases $k = 0, \pm 1$. The same metric can be written in Cartesian form as

$$ds^2 = dt^2 - \frac{a^2(t)}{(1+kr^2/4)^2}[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)] = dt^2 - \frac{a^2(t)}{(1+kr^2/4)^2}[dx^2 + dy^2 + dz^2] \tag{2.4.14}$$

very convenient for analytical work, especially in the case $k = 0$. The overall sign of the metric is arbitrary, and often one uses the form or "signature" denoted as $-+++$,i.e.

$$ds^2 = -dt^2 + a^2(t)[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)] \tag{2.4.15}$$

Once fixed, the signature and the value of k cannot change during the cosmological evolution.

The FLRW so obtained is "seen" by an observer at rest the center of the coordinate frame, so is in a sense the simplest form of the metric. Different observers, e.g. moving on a spacecraft in some direction, will derive a boosted form of the FLRW. The important point is that all possible ways to write down a FLRW are equivalent, in the sense that there is a coordinate transformation that brings one form into the other.

2.4.1 Hubble law

It's clear from the form of the FRW metric that if we assign the coordinates r, θ, ϕ at a given time t_0, the function $a(t)$ acts as a overall factor in the expansion or contraction. The physical distance measured along a null geodesic $ds = 0$ (ie along a light beam) is, for small propagation distances and for a radial $d\theta = d\phi = 0$, simply $D = cdt \approx a(t)dr$. We have then Hubble's Law (or Lemaître-Hubble Law)

$$\dot{D} = \dot{a}dr = HD \tag{2.4.16}$$

where

$$H = \frac{\dot{a}}{a} \tag{2.4.17}$$

is the Hubble constant, or the rate of expansion of space (at the time of observation). Hubble's law applies to any system that expands (or contracts) in a homogeneous and isotropic way.

The coordinate distances

$$r = \int \frac{dr'}{\sqrt{1-kr'^2}} \tag{2.4.18}$$

are fixed on the points moving with the general expansion (the so-called Hubble flow). Since they "move" with the expansion itself, they are called comoving distances. The physical distances $D = a(t)r$ vary instead with the expansion. For convenience, we often define the present distances such that $D = r$, ie $a(t = 0) = 1$. In this way, the astronomical distances measured at the present epoch, for example, the distance between the Milky Way and the Virgo Cluster, are also comoving distances, which are fixed forever. In other words, the comoving distance of the Virgo cluster is 15 Mpc at every epoch, while the physical distance increases with time.

2.4.2 Redshift

Consider a wave source at rest. The interval between two crests is $\lambda_{em} = cdt$, whereλ_0 is the wavelength and c is their speed. If now in the same dt the source moves away from the observer with velocity $-v$, it is clear that the interval between two crests stretches by the distance traveled by the source, that is vdt, and therefore (for non-relativistic speeds) one observes a wavelength $\lambda_{obs} = cdt + vdt$. Thus there is a Doppler shift between the emitted wave (subscript em) and the observed one (obs):

$$\frac{d\lambda}{\lambda} = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{v}{c} \tag{2.4.19}$$

The *redshift* is defined therefore as

$$z \equiv \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{2.4.20}$$

If we now imagine that the signal was emitted from a source moving away according to Hubble's law (eg a galaxy) we get $v = HD$, and then we obtain a relationship between wavelength shift and scale factor:

$$\frac{d\lambda}{\lambda} = \frac{v}{c} = \frac{HD}{c} = -Hdt = -\frac{da}{a} \tag{2.4.21}$$

where we have considered a negative $dt = t_{emission} - t_{observation}$. Therefore, by integrating $d\lambda/\lambda = -da/a$ and normalizing the scale factor such that at the present epoch $a = a_0 = 1$, we obtain that the observed wavelength λ_{obs} of a source that has emitted the signal at epoch a_e is $\lambda_{obs} = \lambda_{em}/a_{em}$. The relation between redshift and scale factor at the emission epoch is then:

$$1 + z = a^{-1} \tag{2.4.22}$$

This relation is of the utmost importance, because it ties an easily observed quantity, z, with the main function of cosmological scale factor $a(t)$. The interpretation of redshift as a Doppler effect is valid only at short distances, at long distances to the relation $\delta\lambda/\lambda = v/c$ should be modified because of the relativistic effects. However eq. (2.4.22) remains valid, as it can be shown by considering the two propagations from the same receding source along $ds = 0$ in a FRW metric, in which r remains constant:

$$r = \int_0^t \frac{dt}{a} = \int_{\Delta t_0}^{t+\Delta t_1} \frac{dt}{a} \tag{2.4.23}$$

2.5 General relativity equations

2.5.1 Energy-momentum tensor

Consider the conservation laws of a perfect fluid, homogeneous and isotropic in the frame at rest relative to the center of mass:

$$\begin{aligned} \dot{\rho} &= 0 \\ \nabla p &= 0 \end{aligned} \tag{2.5.1}$$

where the energy density is $\rho = nmc^2$ (n being the density of particles of mass m) and the pressure in the direction i is i is $p_i = nmv_i^2$ (that is $p = F/A$ where the force acting on a surface of area A is $F = \frac{mv}{dt} n(vdt)A$). We can then define the matrix

$$T^{\mu\nu} = \text{ diag}(\rho, p_x, p_y, p_z) = \text{ diag}(\rho, p, p, p) \tag{2.5.2}$$

(the last step requires isotropy) that is also $T^{\mu}_{\nu} = \text{ diag}(\rho, -p, -p, -p)$. We see then that the laws (2.5.1) amount to

$$T^{\mu\nu}_{,\mu} = 0$$

Let us now find a tensor that reduces to (2.5.2) in the special-relativity limit. We could in fact make a Lorentz transformation on $T^{\mu\nu}$, but we can also notice that only two tensors can be part of the result, $u^\mu u^\nu$ e $g^{\mu\nu}$. The only expression linear in the two tensors and function of ρ, p that reduces to (2.5.2) in the Minkowski limit is

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu - pg^{\mu\nu} \tag{2.5.3}$$

(this becomes $T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}$ if the metric has opposite signature).

If the reference system is at rest relative to the matter, one has $u^\mu = (1, 0, 0, 0)$ and so in this case the components of the tensor are:

$$T^{00} = \rho, \quad T^{ii} = \frac{p}{a^2}, \quad T \equiv T^\mu_\mu = \rho - 3p \tag{2.5.4}$$

The covariant generalization of the conservation equation is now immediate (see eq. 2.3.4)

$$T^{\mu\nu}_{;\mu} = T^{\mu\nu}_{,\mu} + \Gamma^\mu_{\beta\mu} T^{\beta\nu} + \Gamma^\nu_{\beta\mu} T^{\mu\beta} = 0 \tag{2.5.5}$$

Exercise: explicit form of (2.5.5) in FRW when $\nu = 0$. Result:

$$\dot\rho + 3H(\rho + p) = 0 \tag{2.5.6}$$

2.5.2 The curvature tensor

We have so far seen how the metric determines the equation of motion of bodies, but still we have no equation that determines the metric itself in the presence of matter. Since the properties of matter are described fully by the tensor $T_{\mu\nu}$, it is now necessary to formulate a general equation that links $g_{\mu\nu}$ to the energy tensor . We require the following properties:

1) generally covariant equations
2) equations which are covariantly conserved, i.e. obey (2.5.5)
3) and that reduce to the Poisson equation

$$\nabla^2\Phi = 4\pi G\rho \tag{2.5.7}$$

in the weak field, small velocities limit.

Now, one can prove that (up to a constant term, see later) there is only a tensor $G_{\mu\nu}$ with second order derivatives in $g_{\mu\nu}$ such that $G^\nu_{\mu;\nu} = 0$:

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

where

$$R_{\alpha\beta} = \Gamma^\mu_{\alpha\beta,\mu} - \Gamma^\mu_{\alpha\mu,\beta} + \Gamma^\mu_{\sigma\mu}\Gamma^\sigma_{\alpha\beta} - \Gamma^\mu_{\sigma\beta}\Gamma^\sigma_{\mu\alpha} \tag{2.5.8}$$

This is the Ricci tensor, obtained as a contraction of the Rieman tensor $R^\mu_{\alpha\nu\beta}$, which describes the properties of curvature of space-time. The trace of $R = g^{\alpha\beta}R_{\alpha\beta}$ is the curvature scalar. The Einstein equations are therefore of the form

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \kappa^2 T_{\mu\nu} \tag{2.5.9}$$

The trace of this equation is

$$R = -\kappa^2 T \tag{2.5.10}$$

Now we determine the constant κ^2 by comparison with the Poisson equation. We take the metric (2.2.6) that describes a weak gravitational field and write the trace of Einstein's equation in the limit $\Phi \ll 1$. To further simplify we assume a static gravitational field, $\dot\Phi = 0$. The only non-zero Christoffel terms are (here we assume $c = 1$).

$$\Gamma^i_{00} = \frac{1}{2}\nabla^i g_{00} = \Phi^{,i} \tag{2.5.11}$$

$$\Gamma^0_{i0} = \Gamma^0_{0i} = \Phi_{,i} \tag{2.5.12}$$

(note that $\Phi_{,i} = -\Phi^{,i}$). It follows then, by neglecting the quadratic terms of type $\Gamma\Gamma$,

$$R = -g^{ii}\Gamma^0_{i0,i} + g^{00}\Gamma^i_{00,i} = -2\nabla^2\Phi \tag{2.5.13}$$

Note that we adopted the metric signature such that

$$\nabla^2\Phi \equiv \Phi_{,ii} = -\Phi^{,i}_{,i} \tag{2.5.14}$$

In the non-relativistic limit $p \ll \rho$, so that we can put $T = \rho - 3p \approx \rho$, eq. (2.5.10) becomes

$$R = -2\nabla^2\Phi = -\kappa^2\rho \tag{2.5.15}$$

Comparing with the Poisson eq. (2.5.7) we find

$$\kappa^2 = 8\pi G \tag{2.5.16}$$

(putting back c we get $\kappa^2 = 8\pi G/c^4$).

2.6 Hilbert-Einstein Lagrangian

Einstein's equations *in vacuo* can also be obtained by varying a gravitational action, called Hilbert-Einstein action

$$A = \int \sqrt{-g} R d^4x \tag{2.6.1}$$

In fact, we note that from eq. (2.2.14), the metric determinant transforms as

$$g' = gJ^{-2}$$

where $J_{\mu\nu} \equiv \partial y_\mu/\partial x_\nu$ is the Jacobian of the general transformation that brings us from g to g'. It is clear then that $\sqrt{-g'}d^4x = \sqrt{-gJ^{-2}}|J|d^4y = \sqrt{-g}d^4y$ is invariant under general transformations: this explains the factor $\sqrt{-g}$ in the action. By varying A with respect to the metric and using the relation

$$\frac{\partial R}{\partial g_{\mu\nu}} = \frac{\partial(g_{\alpha\beta}R^{\alpha\beta})}{\partial g_{\mu\nu}} = R^{\mu\nu} + g_{\alpha\beta}\frac{\partial R^{\alpha\beta}}{\partial g_{\mu\nu}} \tag{2.6.2}$$

and also

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}(\delta g^{\mu\nu})g_{\mu\nu} \tag{2.6.3}$$

we obtain

$$\delta A = \int \sqrt{-g}d^4x[-\frac{1}{2}g_{\mu\nu}R + R_{\mu\nu} + g^{\alpha\beta}\frac{\partial R_{\alpha\beta}}{\partial g_{\mu\nu}}]\delta g^{\mu\nu} = 0 \tag{2.6.4}$$

We can now show that the term

$$\delta A = \int \sqrt{-g}d^4x[g^{\alpha\beta}\frac{\partial R_{\alpha\beta}}{\partial g_{\mu\nu}}]\delta g^{\mu\nu} = 0 \tag{2.6.5}$$

is a total differential (i.e. $\delta A = 0$ is an identity) and is therefore irrelevant for as concerns the equation of motion. In fact one can write

$$R_{\mu\nu} = \Gamma^\alpha_{\mu\nu;\alpha} - \Gamma^\beta_{\beta\mu;\nu} \tag{2.6.6}$$

(where the covariant derivative is to be meant only wrt the upper index of the Christoffel symbols) and

$$\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} = \sqrt{-g}(g^{\mu\nu}\delta\Gamma^{\alpha}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^{\beta}_{\beta\mu})_{;\alpha} \tag{2.6.7}$$

The term inside parentheses is the covariant derivative of the vector $V^{\alpha} \equiv g^{\mu\nu}\delta\Gamma^{\alpha}_{\mu\nu} - g^{\mu\alpha}\Gamma^{\beta}_{\beta\mu}$ and can therefore be written as

$$(V^{\alpha}\sqrt{-g})_{,\alpha} \tag{2.6.8}$$

(notice now the derivative is the ordinary one))i.e. as a total derivative.

Then the Einstein equations in vacuum follow

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0 \tag{2.6.9}$$

2.7 Spatial curvature of FRW

The unknown function $\lambda(r)$ of the FRW metric defined in Sect. (2.4) can be evaluated also by requiring that the space has a constant spatial curvature P, defined as

$$P = \sigma^{ij}P_{ij} = \sigma^{ij}P^{m}{}_{imj} \tag{2.7.1}$$

where σ_{ij} is the spatial metric defined in (2.4).

The curvature scalar P is (**obtain it** using eq. 2.5.8)

$$P = 2\frac{(-\lambda + \lambda^3 + 2r\lambda')}{r^2a^2\lambda^3} \tag{2.7.2}$$

We can now solve the equation $P = constant = k/a(t)^2$, (that is, a constant independent of the spatial coordinates), and finally we find

$$\lambda = \frac{1}{1 - kr^2} \tag{2.7.3}$$

2.8 Natural units and Planck units

Often we use in this course *the natural units and the Planck units*. These are defined from the fundamental constants $c, G, \hbar$. The Planck length is:

$$L_P = \left(\frac{G\hbar}{c^3}\right)^{1/2} = 1.61 \cdot 10^{-33} cm\,. \tag{2.8.1}$$

while Planck mass, time and energy are:

$$M_P = \left(\frac{c\hbar}{G}\right)^{1/2} = 2.17 \cdot 10^{-5} gr\,, \tag{2.8.2}$$

$$t_P = \left(\frac{G\hbar}{c^5}\right)^{1/2} = 5.39 \cdot 10^{-44} sec\,, \tag{2.8.3}$$

$$E_P = \left(\frac{c^5\hbar}{G}\right)^{1/2} = 1.22 \cdot 10^{19} GeV\,. \tag{2.8.4}$$

We can also define the Planck temperature, $T = 1.4 \cdot 10^{32} K$. Natural units are defined putting $c = \hbar = 1$ (and also $k_B = 1$ to express temperature in energy units). Then we see that in natural units

$$L_P = t_P = M_P^{-1} = E_P^{-1} \tag{2.8.5}$$

In this way, we can express everything in terms of energy. For example, the energy density has dimensions $energy/length^3 = energy^4$.

These units arise when trying to tie together quantum mechanics and general relativity. For instance, if we consider that black holes have mass-radius relation (we skip all factors of order unity) $GM/c^2 = R$, and that the time it takes for light to cross a radius R is $\Delta t = R/c$, and that Heisenberg relation says that $\Delta t \cdot \Delta E \geq \hbar$, where $\Delta E = Mc^2$, then one gets immediately that the value of M such that Heisenberg relation is minimally fulfilled is given my M_P above. These consideration are only qualitative and we do no yet know how to handle such kind of phenomena.

As a quick application, let us convert $1g/cm^3$ in units of energy GeV^4. We can proceed this way:

$$1g/cm^3 = 10^5 M_p/(10^{33})^3 L_p^3 = 10^{-94} E_p^4 = 10^{-18} GeV^4 \tag{2.8.6}$$

Or, to convert the Stefan-Boltzmann constant

$$\sigma = \frac{2\pi^5 k_B^4}{15c^2h^3} = \frac{\pi^2 k_B^4}{60c^2\hbar^3} = 5.67 \cdot 10^{-8} Js^{-1}m^{-2}K^{-4} \tag{2.8.7}$$

we put

$$\sigma = 5.67 \cdot 10^{-8} 10^{-8} M_P 10^{-42} t_P^{-1} 10^{-70} L_P^{-2} K^{-4} = 5.67 \cdot 10^{-128} E_p^4 K^{-4} \tag{2.8.8}$$

So, for instance, the energy density of a black body at 1K is $\rho_\gamma = \sigma T^4 = 5.67 \cdot 10^{-128} E_P^4$ equal to roughly $10^{-36} g/cm^3$. At 3K, the value is 100 times higher and

$$T_{2.7K} = 2.3 \cdot 10^{-13} GeV \tag{2.8.9}$$

The critical density today is

$$\rho_c = 2 \cdot 10^{-29} h^2 g/cm^3 = 2 \cdot 10^{-47} h^2 GeV^4 \tag{2.8.10}$$

Often in this text we'll use approximate Planck units, ie take into account only the orders of magnitude. This simplifies the treatment but now and then the quantitative values reported here differ from other texts by order of unity factors.

Chapter 3

The expanding Universe

Quick summary

- We introduce and solve the Friedmann equations, valid for a homogeneous and isotropic Universe
- We consider matter in form of non-relativistic particles, of relativistic particles, and with a general equation of state
- We find the general behavior of the scale factor and of the cosmic age
- We also introduce the cosmological constant
- Finally we see how measurements of distances can test the Friedmann equations.

3.1 Friedmann equations

Let us now write down the metric equations in a homogeneous and isotropic space, i.e. in the FRW metric:

$$ds^2 = dt^2 - a^2(t)[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)] \tag{3.1.1}$$

For $k = 0$ the Christoffel symbols are all vanishing except (it is easier here to perform the calculations using the Cartesian form 2.4.14)

$$\Gamma^i_{j0} = \Gamma^i_{0j} = H\delta^i_j \quad , \quad \Gamma^0_{ij} = a\dot{a}\delta_{ij} \tag{3.1.2}$$

We have then

$$R_{00} = \Gamma^\mu_{00,\mu} - \Gamma^\mu_{0\mu,0} + \Gamma^\mu_{\sigma\mu}\Gamma^\sigma_{00} - \Gamma^\mu_{\sigma 0}\Gamma^\sigma_{\mu 0} = -3\dot{H} - H^2\delta^i_j\delta^j_i = -3(\dot{H} + H^2) = -3\frac{\ddot{a}}{a} \tag{3.1.3}$$

and the trace

$$R = -\frac{6}{a^2}(\dot{a}^2 + a\ddot{a} + k) = -6\dot{H} - 12H^2 - 6ka^{-2} \tag{3.1.4}$$

Let us now consider the $(0,0)$ component and the trace component of the Einstein equations:

$$R_{00} - \frac{1}{2}g_{00}R = 8\pi T_{00} \tag{3.1.5}$$

$$R = -8\pi T \tag{3.1.6}$$

From the first equation and by combining the two we obtain the two *Friedmann equations*:

$$H^2 = \frac{8\pi}{3}\rho - \frac{k}{a^2} \tag{3.1.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}(\rho + 3p) \tag{3.1.8}$$

to be complemented by the conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0 \tag{3.1.9}$$

The Friedmann equations and the conservation equations are however not independent. By differentiating eq. (3.1.7) and inserting eq. (3.1.9) we obtain the other Friedmann equation. Let us define now the critical density

$$\rho_c = \frac{3H^2}{8\pi G} \tag{3.1.10}$$

and the density parameter

$$\Omega = \frac{\rho}{\rho_c} \tag{3.1.11}$$

so that eq. (3.1.7) becomes

$$1 = \Omega - \frac{k}{a^2 H^2} \tag{3.1.12}$$

This shows that $k = 0$ corresponds to a universe with density equal to the critical one, that is $\Omega = 1$. Spaces with $k = +1$ correspond instead to $\Omega > 1$, spaces with $k = -1$ to $\Omega < 1$. We can also define a curvature component

$$\Omega_k \equiv -\frac{k}{a^2 H^2} \tag{3.1.13}$$

(which implies the definition $\rho_k = -3k/8\pi a^2$). At every epoch we have then

$$1 = \Omega(a) + \Omega_k(a) \tag{3.1.14}$$

As we will see, this relation extends directly to models with several components.

3.2 Non relativistic component

Let us consider now a fluid with zero pressure

$$p = 0 \tag{3.2.1}$$

Such a fluid approximates "dust" matter (like e.g. galaxies) or a gas composed by non-interacting particles with non-relativistic velocities (like e.g. cold dark matter). In fact, the pressure of a free-particle fluid with mean square velocity v is $p = nmv^2$, much smaller than $\rho = nmc^2$ for non relativistic speeds. Then we have from (3.1.9) that

$$\dot{\rho}/\rho = -3\dot{a}/a \tag{3.2.2}$$

or

$$\rho \sim a^{-3} \tag{3.2.3}$$

Every time we write a relation of this kind we mean a power law normalized to an arbitrary instant a_0 (here assumed to be the present epoch). We mean then

$$\rho = \rho_0 \left(\frac{a_0}{a}\right)^3 \tag{3.2.4}$$

As a function of redshift we have (the subscript NR or m denotes the pressureless non-relativistic component)

$$\rho = \rho_0 (1+z)^3 = \rho_c \Omega_{m,0} (1+z)^3 \tag{3.2.5}$$

where from now on, except where otherwise denoted, Ω_i represents the *present* value for the species i and ρ_c is the *present* critical density. When needed, a subscript 0 will be added.

Let us assume now a flat space $k = 0$. The present density ρ_0 is linked to the Hubble constant by the relation

$$H_0^2 = \frac{8\pi}{3}\rho_0 \tag{3.2.6}$$

Then we have (defining distances such that $a_0 = 1$)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\rho_0 a_0^3 a^{-3} = H_0^2 a^{-3} \tag{3.2.7}$$

from which, integrating and putting the epoch of Big Bang at $t = 0$,i.e. $a(t = 0) = 0$, we find

$$a(t) = (\frac{3}{2}H_0 t)^{2/3}$$

or simply

$$a \sim t^{2/3} \tag{3.2.8}$$

Since we measure a present Hubble constant

$$H_0 = 100h\,\mathrm{km/sec/Mpc} \tag{3.2.9}$$

where $h = 0.70 \pm 0.04$, (according to the recent estimates), and since $1\mathrm{Mpc} = 3 \cdot 10^{19}\mathrm{km}$ and $G = 6.67 \cdot 10^{-8}\mathrm{cm}^3 g^{-1}\mathrm{sec}^{-2}$ we have the present critical density

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} \approx 2 \cdot 10^{-29} h^2 \mathrm{gcm}^{-3} \tag{3.2.10}$$

The matter density currently measured is indeed close to $\rho_{c,0}$.

3.3 Relativistic component

A photon gas distributed as a black body has a pressure equal to

$$p = \frac{1}{3}\rho \tag{3.3.1}$$

(notice that for radiation the energy-momentum trace vanishes, $T = \mathrm{Trace}(T^\mu_\nu) = \rho - 3p = 0$); this can be seen also from the form of the electromagnetic tensor

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\lambda}F^\nu_\lambda - \frac{1}{4}g^{\mu\nu}F^{\lambda\sigma}F_{\lambda\sigma}\right) \tag{3.3.2}$$

whose trace vanishes). Then we have from (3.1.9) that

$$\dot{\rho} = -4H\rho \tag{3.3.3}$$

from which (the subscript r or γ denotes the relativistic or radiation component)

$$\rho_R \sim a^{-4} = \rho_c \Omega_{r,0}(1 + z)^4 \tag{3.3.4}$$

The radiation density dilutes as a^{-3} because of the volume expansion and as a^{-1} because of the energy redshift.

To evaluate the present radiation density we'll remind that a photon gas in equilibrium with matter (black body) has energy density ($\hbar = c = 1$)

$$\rho_\gamma = \frac{g}{2\pi^2}\int \frac{E^3 dE}{e^{E/T} + 1} = \frac{g\pi^2}{30}T^4 \tag{3.3.5}$$

where T is expressed in energy units and g are the degrees of freedom of the relativistic particles ($g = 2$ for the photons, $g \approx 3.36$ including 3 massless neutrino species, see Sec. 4.2). Notice that since $\rho_\gamma \sim a^{-4}$ the radiation temperature scales as

$$T \sim \frac{1}{a} \tag{3.3.6}$$

Since today we measure $T \approx 3K \approx 10^{-13}GeV = 1.38 \cdot 10^{-23}JK^{-1}$, we have

$$\rho_\gamma = g \cdot 2.3 \cdot 10^{-34}\text{gcm}^{-3} \tag{3.3.7}$$

which is much smaller than the present matter density. The present epoch is denoted therefore *matter dominated epoch* (MDE).

(Notice: in Planck units $T_{3K} = 10^{-32}E_P$; so that, approximately, $T_{3K}^4 = 10^{-128}E_P^4 = 10^{-128}M_P L_P^{-3} = 10^{-128}10^{-5}10^{99}$ g/cm^3. Thus $T_{3K}^4 \approx 10^{-34}$g/cm^3.)

From the matter and radiation trends

$$\rho_m = \rho_{m,0}a^{-3} \tag{3.3.8}$$
$$\rho_\gamma = \rho_{\gamma,0}a^{-4} \tag{3.3.9}$$

we can define the equivalence epoch a_e for which $\rho_\gamma = \rho_m$:

$$a_e = \frac{\rho_{\gamma,0}}{\rho_{m,0}} = \frac{\Omega_\gamma}{\Omega_m} \tag{3.3.10}$$

Since we have (including neutrinos)

$$\Omega_\gamma = \frac{\rho_\gamma}{\rho_{crit}} \simeq 4.15 \cdot 10^{-5}h^{-2} \tag{3.3.11}$$

it follows that the equivalence occurred at a redshift

$$1 + z_e = a_e^{-1} = \left(4.15 \cdot 10^{-5}\right)^{-1}\Omega_m h^2 = 24,000\Omega_m h^2 \tag{3.3.12}$$

Putting $\Omega_c = 0.3$ and $h = 0.7$ we obtain $z_e \approx 3500$.

3.4 General component

It is clear now that any fluid with equation of state

$$p = w\rho \tag{3.4.1}$$

goes like

$$\rho \sim a^{-3(1+w)} \tag{3.4.2}$$

In the case $k = 0$ and if the fluid is the dominating component in the Friedmann equation, the scale factor grows like

$$a \sim t^{2/3(1+w)} \tag{3.4.3}$$

3.5 General Friedmann equation

We can now write the Friedmann equation as

$$H^2 = \frac{8\pi}{3}(\rho_m a^{-3} + \rho_\gamma a^{-4} + \rho_k a^{-2}) = H_0^2(\Omega_m a^{-3} + \Omega_\gamma a^{-4} + \Omega_k a^{-2}) \tag{3.5.1}$$

where as already noted Ω_i denotes the present density of species i, so that $\sum_i \Omega_i = 1$. Every other hypothetical component can be added to this Friedmann equation when its behavior wih a is known.

3.6 Qualitative trends

In all the cases seen so far we always had $\rho + 3p > 0$. Then from (3.1.8) it follows $\ddot{a} < 0$, that is, a decelerated trend at all times. From this it follows that 1) the scale factor must have been zero at some time t_{sing} in the past; and 2) the trajectory with $\dot{a} = const, \ddot{a} = 0$ is the one with minimal velocity in the past (among the decelerated ones) . From $\dot{a} = const$ it follows the law

$$a(t) = a_0 + \dot{a}_0(t - t_0) \tag{3.6.1}$$

and one can derive that the time

$$T = t_0 = a_0/\dot{a}_0 = H_0^{-1} \tag{3.6.2}$$

it takes for the expansion to go from $a = 0$ (when $t = 0$) to $a = a_0$ is the maximal one. H_0^{-1} is then the maximal age of the universe for all models with $\rho + 3p > 0$. Note that

$$H_0^{-1} = \frac{\text{sec} \cdot \text{Mpc}}{100h\,\text{km}} = 9.78\text{Gyr}/h \tag{3.6.3}$$

This extremal model is called Milne's universe and can be obtained from 3.5.1 for

$$\Omega_m = \Omega_\gamma = 0 \tag{3.6.4}$$

so that $\Omega_k = 1$ (hyperbolic space). Then we have $H^2 = H_0^2 a^{-2}$ and thus $\dot{a} = H_0$.

For a general case with non vanishing matter we have instead, by integrating the Friedmann equation,

$$H = H_0(\Omega_m a^{-3} + \Omega_k a^{-2})^{1/2} \equiv H_0 E(a) \tag{3.6.5}$$

that

$$H_0 T = \int_0^{a_0} \frac{da}{aE(a)} = \int_0^{a_0} \frac{da}{\sqrt{\Omega_m a^{-1} + 1 - \Omega_m}} \tag{3.6.6}$$

Equivalently, one can write this (and similar) integrals by replacing $da/a = -dz/(1 + z)$. For $\Omega_m = 1$ (called Einstein-de Sitter Universe) we get

$$T = \frac{2}{3H_0} = 6.7h^{-1}\text{Gyr} \tag{3.6.7}$$

an age too short to accommodate the oldest stars in our Galaxy, unless h is smaller than 0.5. The age corresponding to a given redshift z can be obtained by integrating from $a = 0$ to $a = (1 + z)^{-1}$, or (again in a Einstein - de Sitter Universe)

$$t = \frac{2}{3} H_0^{-1}(1 + z)^{-3/2} \tag{3.6.8}$$

For instance, $z_{dec} = 1100$ corresponds to an age $t = 200,000h^{-1}$yr.

Finally, in the case $k = 1$, i.e. a closed spherical geometry, we can see that H vanishes for $\rho = 3/8\pi a^2$ or (when only matter is present and obviously for $\Omega_m > 1$) when

$$a_{max} = \frac{\Omega_{m,0}}{\Omega_{m,0} - 1} \tag{3.6.9}$$

At this epoch, expansion stops and a contraction phase with $H < 0$ begins. This phase will end in a *big crunch* after an interval equal to the one needed to reach the maximum a_{max}.

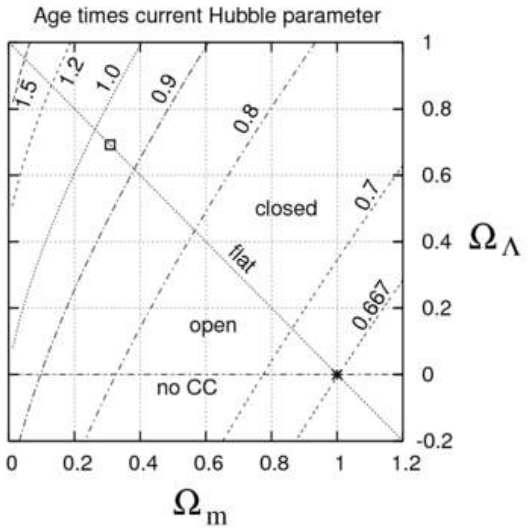


Figure 3.6.1: Age of the Universe as a function of matter and cosmological constant fractions. Notice how for constant Ω_m the age increases with Ω_Λ (from WikiCommons, author Panos84).

3.7 Cosmological constant

To obtain a cosmic age larger than H_0^{-1} it is necessary to violate the so-called "strong energy condition" $\rho + 3p > 0$. The most important example of this case is the vacuum energy or cosmological constant.

Let us consider the energy-momentum tensor (2.5.3). This holds for observers which are comoving with the expansion. Every other observer will see a different content of energy/pressure. There exists however a case in which every observer sees exactly the same energy-momentum tensor, regardless of the 4-velocity u_μ: this occurs when $\rho = -p$: in such a case in fact $T_{\mu\nu} = \rho g_{\mu\nu}$. The conservation condition then implies $\rho_{,\mu} = 0$ or $\rho = const$. It follows that the tensor

$$T_{\mu\nu} = \frac{\Lambda}{8\pi} g_{\mu\nu} \tag{3.7.1}$$

where Λ, the *cosmological constant*, is independent of the observer motion. This condition indeed characterizes an empty space, i.e. a space without real particles. $T_{\mu\nu(\Lambda)}$ is denoted then *vacuum energy*. Conparing with $T^\mu_\nu = \mathrm{diag}(\rho, -p, -p, -p)$, we see that

$$\rho_\Lambda = -p_\Lambda = \frac{\Lambda}{8\pi} \tag{3.7.2}$$

which corresponds to the equation of state $w = p/\rho = -1$.

In the Einstein equations, the cosmological constant appears therefore simply as an additional term that is also covariantly conserved:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \tag{3.7.3}$$

3.8 Cosmological observations

Let us see now how we can connect the cosmological definitions to the astrophysical observables. Let us define first of all the magnitude as a function of the luminosity L (energy output per second) of a source

$$M = -2.5 \log_{10} L + const \tag{3.8.1}$$

(all logarithms in base 10 in this section) The constant is chosen arbitrarily and depends on the observed waveband. For instance, $M_{sun,B} = 5.48$ (B is the blue band at 4400 A) and $L_{sun} \simeq 4 \times 10^{33}$erg s^{-1}.

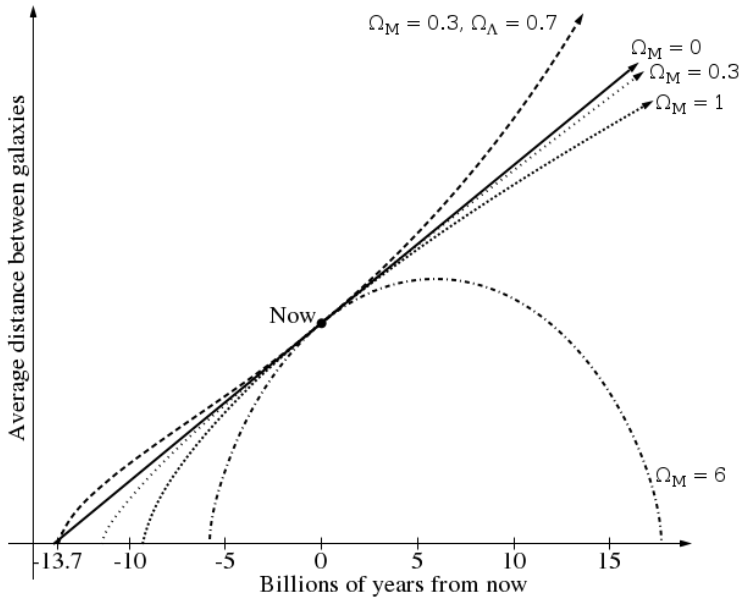


Figure 3.8.1: Evolution of the scale factor in various cosmological models (from WikiCommons, author BenRG, public domain).

The relation between flux received at distance d in a non-expanding Euclidean geometry,

$$f = \frac{L}{4\pi d^2} \tag{3.8.2}$$

and apparent magnitude m is

$$m = -2.5 \log f + const \tag{3.8.3}$$

and the constant is such that $m = 0$ for an object with $f = 2.5 \times 10^{-5}$ erg cm^{-2}s^{-1}. It follows than

$$m = M + 25 + 5 \log d \tag{3.8.4}$$

if d is measured in Mpc (1 Mpc$\simeq 3 \cdot 10^{24}$ cm); notice that $m = M$ for an object at 10pc$= 10^{-5}$Mpc. The difference $m - M$, proportional to $\log d$, is called the *distance modulus* and is in practice a measure of distance. We have then

$$L = 3.02 \times 10^{35-0.4M_b} \text{erg s}^{-1} \tag{3.8.5}$$

$$f = 2.52 \times 10^{-5-0.4m_b} \text{erg cm}^{-2}\text{s}^{-1} \tag{3.8.6}$$

A typical galaxy contains 10^{10} stars, and therefore it has a magnitude difference with respect to the Sun (assumed to be a typical star) equal to $-2.5 \times \log 10^{10} \approx -25$; therefore, it has absolute magnitude near $M_* = -20$. Then, at a distance of 100 Mpc it shows an *apparent magnitude*

$$m \approx M_* + 25 + 5 \log d = 15 \tag{3.8.7}$$

3.9 Luminosity distance

Let us now work out the relativistic version of the flux-luminosity relation. If an object is at a comoving distance r, the relation is

$$f = \frac{L}{4\pi r^2 (1+z)^2} \tag{3.9.1}$$

Let us define then the *luminosity distance*

$$d(z) = r(z)(1+z) \tag{3.9.2}$$

so that the Euclidean relation (3.8.2) is formally unchanged. This is by definition the distance that occurs in the distance modulus relation $(m-M)$ in eq. (3.8.4). The two extra factors $(1+z)$ in (3.9.1) arise because the emitted energy is redshifted away and because the time interval during which is received, dt_0,is a_0/a_1 times the emission interval dt_1. The coordinate distance $r(z)$ is the distance along the null geodesic

$$ds^2 = c^2dt^2 - a^2dr^2 = 0 \tag{3.9.3}$$

In a flat universe ($k=0$) we have

$$\int_0^r dr' = \int_{t_1}^{t_0} \frac{cdt}{a(t)} \tag{3.9.4}$$

That is

$$r = \int_0^r dr' = c\int_{t_1}^{t_0} \frac{dt}{a(t)} = c\int_{a_1}^{a_0} \frac{dt}{da}\frac{da}{a} = c\int \frac{da}{\dot{a}a} = c\int_{a_1}^{a_0} \frac{da}{Ha^2} \tag{3.9.5}$$

Using the redshift z we obtain

$$dz = -da/a^2 \tag{3.9.6}$$

so that

$$r = c\int_0^z \frac{dz'}{H(z')} \tag{3.9.7}$$

For a non-flat space we have instead

$$\int_0^r \frac{dr'}{\sqrt{1-kr^2}} = \int_0^z \frac{dz'}{H(z')} \tag{3.9.8}$$

or, putting $H = H_0E(z)$, $y = rH_0$ and $\Omega_k = -k/H_0^2$

$$\int_0^r \frac{dy}{\sqrt{1+\Omega_k y^2}} = \int_0^z \frac{dz'}{E(z')} \tag{3.9.9}$$

This can be integrated to give

$$r = \frac{1}{H_0\sqrt{|\Omega_k|}} S[\sqrt{|\Omega_k|}\int_0^z \frac{dz'}{E(z')}] \tag{3.9.10}$$

where

$$S(x) = \begin{cases} \sin(x) & if\ k=+1 \\ x & if\ k=0 \\ \sinh(x) & if\ k=-1 \end{cases} \tag{3.9.11}$$

This is quite a general formula. Given any cosmological model (i.e., fixing the parameters $\Omega_r, \Omega_m, \Omega_\Lambda$ etc) we can obtain $r(z)$, then $d(z)$ and finally predict the magnitude $m(z)$ that a source of given absolute magnitude M should have. For instance, in a flat universe with pure matter

$$H^2 = H_0^2a^{-3} = H_0^2(1+z)^3 \tag{3.9.12}$$

from which

$$r(z) = cH_0^{-1}\int_0^z dz(1+z)^{-3/2} = \frac{2c}{H_0}\left[1-(1+z)^{-1/2}\right] \tag{3.9.13}$$

and the luminosity distance is

$$d(z) = \frac{2c}{H_0}\left[(1+z) - (1+z)^{1/2}\right] \tag{3.9.14}$$

Suppose now we have a supernova at $z = 1$. If its magnitude is M we have that

$$m(z=1) = M + 25 + 5\log d(z=1) = M + 25 + 5\log 3514 = M + 42.7$$

For a supernova type Ia we have $M \approx -19.5$, so that for a flat universe we predict $m \approx 23.2$. If we evaluate $m(z)$ for a model with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ we obtain instead $m = 23.8$. The difference $\Delta m = 0.6$ is distinguishable with the present data and the best model contains in fact a fraction of cosmological constant around 70% of the critical one.

We observe finally that $r(\infty) = 2c/H_0$ (the big bang distance) and that

$$\frac{c}{H_0} = \frac{300,000\,\mathrm{km/sec}}{100h\,\mathrm{km/sec/Mpc}} = 3,000\mathrm{Mpc}/h \tag{3.9.15}$$

so that $r(\infty) = 6000\ \mathrm{Mpc}h^{-1}$.

Chapter 4

Thermal processes

Quick summary

- As the temperature goes down in the past we expect that the various species of particles in the universe gradually break out of thermal equilibrium.
- In other words, the mean free path from a given interaction between the particles becomes larger the horizon scale H^{-1} .
- This phenomenon leads to the annihilation of electrons from radiation at $z \approx 10^{10}$ ($T \approx 1$MeV), the formation of light atomic nuclei at $z \approx 10^{8}$ ($T \approx 0.1$MeV) and the decoupling of baryons from photons, $z \approx 10^{3}$ ($T \approx 1$eV).

4.1 Preliminaries

Along the universe cosmic history we can identify a number of phases, the first four still hypothetical while the latter two rather well established:

- Quantum gravity. $T \sim 10^{19} GeV$ (Planck energy), $t \sim 10^{-43} sec$.
- Baryogenesis. $10^{17} GeV < T < 10^{2} GeV$ [perhaps 10^{15} GeV, $t \sim 10^{-35} sec$].
- Electroweak transition. $T \sim 10^{3} GeV$ (mass of weak bosons, $Z^0, W^{\pm}$).
- Quark-Hadron transition. $T \sim 1 GeV$ (nucleon mass).
- Nucleosynthesis. $T \sim 1 MeV$, $t \sim 3$min (nuclear levels).
- Recombination. $T \sim 10 eV$, $t \sim 10^{6} yr$ (atomic levels).

The energy scales here indicated are very approximated, sometimes even by an order of magnitude larger that the real values, due to the very high number of photons per particle (so that a sufficient number of hot photons still remain even at low temperatures), as we will see in due course. All these phases occur because the temperature dropped enough to break reversibility, i.e. to allow some process while forbidding others. For instance, electrons remain in thermodynamic equilibrium with radiation as long as the electron-positron annihilation is compensated by photon pair production, i.e. $e^- + e^+ \leftrightarrow \gamma + \gamma$. But when the average energy of the photons decreases below the 0.5 MeV needed for the electron mass, the pair production becomes forbidden and the residual pair annihilation raises the photon temperature. More in general, denoting with Γ the interaction rate (probability of events per unit time), we say that there is a *freezing* for that particular interaction when $\Gamma \ll H$ or, in other words, when the interaction rate is smaller than the expansion rate, or again, equivalently, when the mean free path for the interaction is much larger than the cosmic horizon cH^{-1}. We discuss now in some detail three thermal events: neutrino abundance, primordial nucleosynthesis, and recombination.

In the following we need some integrals of the thermal equilibrium distribution of a particle A with g_A internal degrees of freedom as a function of energy $E = \sqrt{k^2 + m^2}$, momentum k, mass m and temperature T (we assume $\hbar = c = k_B = 1$), given by

$$f(\mathbf{k}, t)d^3k = \frac{g_A}{(2\pi)^3}[e^{\frac{E-\mu_A}{T(t)}} \pm 1]^{-1}d^3k \tag{4.1.1}$$

where μ_A is the chemical potential of that species (defined as the change in total energy when varying the number of particles keeping constant entropy, volume and number of other particles). The distribution f is normalized in such a way that the integral over all momenta is the number density. It's important to note that in every reaction, the total chemical potential is preserved. The sign $\pm$ refers to bosons ($-$, Bose-Einstein) or fermions ($+$, Fermi-Dirac). From this we derive the number density distribution n, the energy density ρ, the pressure p as a function of T:

$$\begin{aligned} n &= \int f(k)d^3k = \frac{g}{2\pi^2}\int \frac{(E^2-m^2)^{1/2}EdE}{e^{\frac{E-\mu_A}{T}} \pm 1} \\ \rho &= \int Ef(k)d^3k = \frac{g}{2\pi^2}\int \frac{(E^2-m^2)^{1/2}E^2dE}{e^{\frac{E-\mu_A}{T}} \pm 1} \\ p &= \int \frac{k^2}{3E}f(k)d^3k = \frac{g}{6\pi^2}\int \frac{(E^2-m^2)^{3/2}dE}{e^{\frac{E-\mu_A}{T}} \pm 1} \end{aligned} \tag{4.1.2}$$

(notice that d^3k integrated over the angles gives $4\pi k^2 dk = 4\pi(E^2-m^2)^{1/2}EdE$).

The third equation has been obtained as follows. In standard thermodynamics, the relation between pressure and momentum for particle with momentum in d^3k is $p_k = \frac{1}{3}n_k(mv)v$, where the term in parenthesis is the momentum and n_k is the number of particles with momentum in d^3k, i.e. $n_k = f(k)d^3k$. In Special Relativity, this becomes simply $p_k = \frac{1}{3}n_k kv$. Moreover, since $k = \gamma mv$ and $E = \gamma m$ (γ being the relativistic factor), one has the general relation $v = k/E$, from which $p_k = n_k k^2/3E$. Therefore, integrating over d^3k, we have indeed

$$p = \int \frac{k^2}{3E}f(k)d^3k \tag{4.1.3}$$

Applying the first equation to the photons, for which $g = 2$ and $m = \mu = 0$, we obtain the numerical density

$$n_\gamma = \frac{2\zeta(3)}{\pi^2}T^3 \tag{4.1.4}$$

(where the Riemann function ζ is defined as $\zeta(x) = \Gamma(x)^{-1}\int_0^\infty \frac{u^{x-1}}{e^u-1}du$ and $\zeta(3) \approx 1.2$). The same result, with a factor 3/4, is valid for a relativistic fermionic component, e.g. neutrinos. To recover standard units, one should multiply by $(k_B/\hbar c)^3$. The energy density is

$$\rho_\gamma = g_\gamma \frac{\pi^2}{30}T^4 \tag{4.1.5}$$

where $g_\gamma = 2$ for the photons (and a factor $k_B^4/(\hbar c)^3$ to recover standard units). Eq. (4.1.5) holds also for every relativistic components ($T \gg m$), with the factor 7/8 for fermions. From the pressure relation, we find $p = \rho/3$ in the relativistic regime.

4.2 The abundance of cosmic neutrinos

(In this section I followed the approach in Piattella, *Lecture Notes in Cosmology*, Springer.) Since the Universe is a closed system, the total entropy S is constant. The specific entropy $s = S/V$ decreases therefore as a^{-3}, so that sa^3 is constant. Let us now find an expression for the specific entropy. We begin with the first law of thermodynamics

$$TdS = pdV + dU \tag{4.2.1}$$

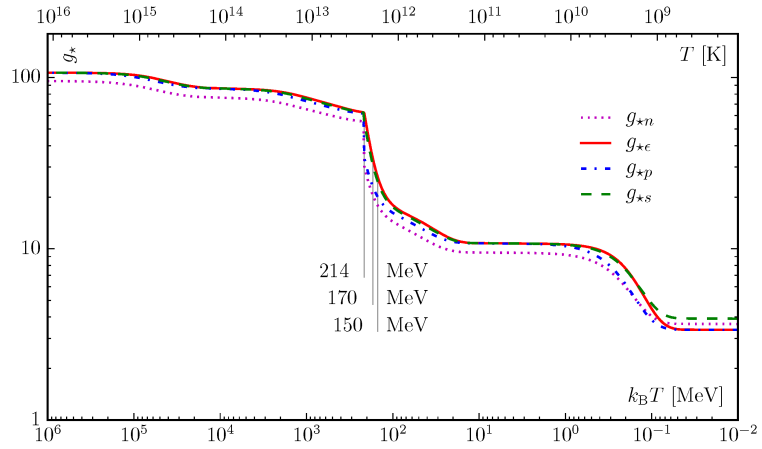


Figure 4.2.1: Relativistic degrees of freedom in the cosmic plasma as a function of temperature. The various lines refer to slightly different definitions of "degree of freedom". The big drop around 150-200 MeV is the transition from quarks to hadrons. The smaller drop around 0.5 MeV is the electron pair annihilation. The final value, around 3.3, is evaluated in the text. From Husdal, *Galaxies* 2016, 4, 78.

Putting $U = \rho V$ and constant entropy, $dS = 0$, this becomes

$$TdS = (\rho + p)dV + Vd\rho = 0 \tag{4.2.2}$$

i.e. the continuity equation, since $V \sim a^3$. In a thermal equilibrium or relativistic particles, as we have seen in the previous section, ρ depends only on T and therefore we can write

$$dS = \frac{(\rho + p)}{T}dV + \frac{V}{T}\frac{d\rho}{dT}dT \tag{4.2.3}$$

Since S under these conditions is a function of state that depends only on V, T, the total differential $dS = \partial_V S dV + \partial_T S dT$ implies

$$\frac{\partial S}{\partial V} = \frac{\rho + p}{T}, \quad \frac{\partial S}{\partial T} = \frac{V}{T}\frac{d\rho}{dT} \tag{4.2.4}$$

from which, differentiating again wrt T and V, respectively, and equating,

$$\frac{d(\rho + p)}{TdT} - \frac{\rho + p}{T^2} = \frac{1}{T}\frac{d\rho}{dT} \tag{4.2.5}$$

from which

$$\frac{dp}{dT} = \frac{\rho + p}{T} \tag{4.2.6}$$

Now we can insert this expression in

$$TdS = (\rho + p)dV + Vd\rho = d[(\rho + p)V] - Vdp = d[(\rho + p)V] - V\frac{\rho + p}{T}dT \tag{4.2.7}$$

to obtain

$$dS = d\left[\frac{(\rho + p)V}{T}\right] \tag{4.2.8}$$

and finally

$$s = \frac{S}{V} = \frac{\rho + p}{T} \tag{4.2.9}$$

up to a constant that we can set to zero. For relativistic particles, $p = \rho/3$, and $s = 4\rho/3T$. Notice that this calculation only applies to relativistic particles, because otherwise ρ would not be just a function of T and the entropy would depend also on the number of particles, N, beside V, T.

Neutrinos and photons do not interact, so if they start with the same temperature at very early times when $T \gg 1$MeV (as expected because the equilibrium is realized through weak interactions), they should maintain it forever. However, photons are heated by electron pair annihilation, so we expect a temperature mismatch below 0.5 MeV, right before nucleosynthesis. To determine this, we can evaluate s for relativistic bosons (photons) and fermions (neutrinos and electrons). From (4.1.5) we have

$$s_\nu = \frac{4\rho}{3T} = \frac{7}{8} g_\nu \frac{4\pi^2}{3 \cdot 30} T^3 = \frac{7}{8} g_\nu \frac{2\pi^2}{45} T^3 \tag{4.2.10}$$

$$s_\gamma = g_\gamma \frac{2\pi^2}{45} T^3 \tag{4.2.11}$$

where the degrees of freedom are $g_\gamma = 2$ for photons (and also for electrons and positrons), and $2g_\nu N_\nu$ for N_ν species of neutrinos (i.e. N_ν times g_ν for neutrinos and g_ν for antineutrinos). In the standard particle physics model, each neutrino type has only one dof instead of two, since right-handed neutrinos, if they exist at all, interact only with gravity. At $T \gg 0.5$ MeV, i.e. *before* pair annihilation, all species have the same temperature, so one has for the relativistic dofs (photons plus electrons/positrons plus neutrinos/antineutrinos)

$$s_b = \frac{2\pi^2}{45} T_b^3 (2 + \frac{7}{8} \cdot 4 + \frac{7}{8} \cdot 2 N_\nu g_\nu) \tag{4.2.12}$$

After pair annihilation stops, i.e. at $T \ll 0.5$ MeV, the temperatures of the surviving species, photons and neutrinos, will be different, so we have instead

$$s_a = \frac{2\pi^2}{45} (2T_\gamma^3 + \frac{7}{8} \cdot 2 g_\nu N_\nu T_\nu^3) \tag{4.2.13}$$

Since sa^3 is constant, we have

$$(a_b T_b)^3 [2 + \frac{7}{8} \cdot 4 + \frac{7}{8} \cdot 2 g_\nu N_\nu] = (a_a T_\nu)^3 [2 \left(\frac{T_\gamma}{T_\nu}\right)^3 + \frac{7}{8} \cdot 2 g_\nu N_\nu] \tag{4.2.14}$$

Now the neutrino temperature goes always as $1/a$ since they are not heated by electron pair annihilation, so $a_b T_b = a_a T_\nu$ and finally

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{1/3} \approx 0.714 \tag{4.2.15}$$

This ratio is maintained forever after. Since today $T_\gamma \approx 2.7$K, the unobservable T_ν should be around 1.9 K.

As a consequence, since $\rho_\gamma, \rho_\nu \sim T^4$, we have

$$\rho_{\nu_{tot}} = \rho_{\nu+\bar{\nu}} = \frac{7}{8} N_\nu g_\nu \left(\frac{4}{11}\right)^{4/3} \rho_\gamma \tag{4.2.16}$$

A recent estimation of $N_\nu g_\nu$ via CMB gives roughly $N_\nu g_\nu \approx 3 \pm 0.3$, consistent with the three families in the standard particle physics model and $g_\nu = 1$, so that $\rho_{\nu_{tot}} \approx 0.68 \rho_\gamma$. The total relativistic energy density today is therefore $\rho_{\nu_{tot}} + \rho_\gamma \approx 1.68 \rho_\gamma$; in practice, one can simply say that instead of $g_\gamma = 2$, today's relativistic degrees of freedom are $g_\gamma^* = 2 \cdot 1.68 = 3.36$.

Similarly, since there are $2N_\nu$ species of neutrinos plus antineutrinos (from now on, ν refers to the sum of $\nu, \bar{\nu}$), and writing $n_\gamma(T) = AT^3$, their number density is

$$n_\nu = \frac{3}{4}(2N_\nu) n_\gamma(T_\nu) = \frac{3}{2} N_\nu A T_\nu^3 = \frac{3}{2} N_\nu A T_\gamma^3 \left(\frac{T_\nu}{T_\gamma}\right)^{1/3} = \frac{3}{2} N_\nu \left(\frac{4}{11}\right) A T_\gamma^3 = \frac{6}{11} N_\nu n_\gamma(T_\gamma) \tag{4.2.17}$$

This gives roughly 340 neutrinos per cm^3. If the neutrinos have a small mass $m_\nu < 1$eV, as it appears from the current constraints from solar neutrino oscillations and cosmological observations, they have become non-relativistic just recently, so the description above remains substantially intact. However, their energy density now would be $n_\nu \bar{m}_\nu$ where the average mass is $\bar{m}_\nu = \sum m_\nu / N_\nu$. This gives

$$\rho_{\nu,0} = \frac{6}{11} n_\gamma \sum m_\nu \tag{4.2.18}$$

and their cosmic fraction

$$\Omega_{\nu,0} = \frac{8\pi G}{3H_0^2} \frac{6}{11} n_\gamma \sum m_\nu \approx \frac{\sum m_\nu}{93 h^2 eV} \tag{4.2.19}$$

(here h is the dimensionless Hubble constant). With $h \approx 0.7$, three neutrinos of 5 eV each would compose all of the observed dark matter. Unfortunately, this simple explanation for the dark matter is today ruled out on several grounds. Massive neutrinos (at least in the standard scenario) can compose less than one percent of the total energy content. More details on possible candidates of dark matter will be discussed later on.

4.3 Primordial nucleosynthesis

The binding energy of nuclei is at the MeV scale. Therefore, at temperatures much higher than 1 MeV, protons and neutrons cannot combine to form heavier nuclei like deuterium, helium etc. because the hot thermal bath would immediately reionize them. Since protons are slightly lighter than neutrons, in an equilibrium Boltzmann distribution with temperature $T \approx m_n - m_p \approx 1.3\text{MeV} \approx 1.5 \cdot 10^{10}$ K there will be more protons than neutrons.

Already at $T_D \approx 0.7$ MeV neutrons and protons are no longer in equilibrium under reactions like $n + \nu_e \leftrightarrow p + e^-$ and their numerical ratio n/p freezes at

$$\frac{n_n}{n_p} = e^{-\frac{\Delta m}{T_D}} \approx \frac{1}{6} \tag{4.3.1}$$

Today's temperature 3K corresponds to $2.4 \cdot 10^{-4}$eV, so at 0.7 MeV the scale factor was $a_D = T_{3K}/T_{0.7MeV} \approx 3.4 \cdot 10^{-10}$. Since we are deep into the radiation era $H = H_0 \sqrt{\Omega_\gamma} a^{-2}$, so this corresponds to a time after big bang

$$t_D = \frac{1}{H_0 \sqrt{\Omega_\gamma}} \int_{z_D}^{\infty} \frac{dz}{(1+z)^3} = \frac{1}{2H_0 \sqrt{\Omega_\gamma}(1+z_D)^2} \approx 10s \tag{4.3.2}$$

In the mean time, a fraction of neutrons decayed spontaneously (lifetime 900 sec) so that one can estimate $n/p = 1/7$ at around 0.1 MeV, when most of the nucleosynthesis process can be considered completed, as will see in the next section. If in a first approximation we assume that all the neutrons n_n end up into in $n_n/2$ nuclei of ^4He (which is a very stable nucleus), we can estimate the mass fraction in Helium as

$$Y = \frac{4(n_n/2)}{n_n + n_p} \approx 0.25$$

a value which is very close to the observed one.

In reality, one should consider a network of reactions, the most important of which are (Fig. 4.3.1) ($D = ^2H =$ deuterium, $T = ^3H =$ tritium)

$$\begin{aligned}
p + n &\rightarrow D + \gamma \\
D + p &\rightarrow {}^3He + \gamma \\
D + D &\rightarrow {}^3He + n \\
D + D &\rightarrow T + p \\
{}^3He + n &\rightarrow T + p \\
{}^3He + D &\rightarrow {}^4He + p \\
T + D &\rightarrow {}^4He + n
\end{aligned}$$

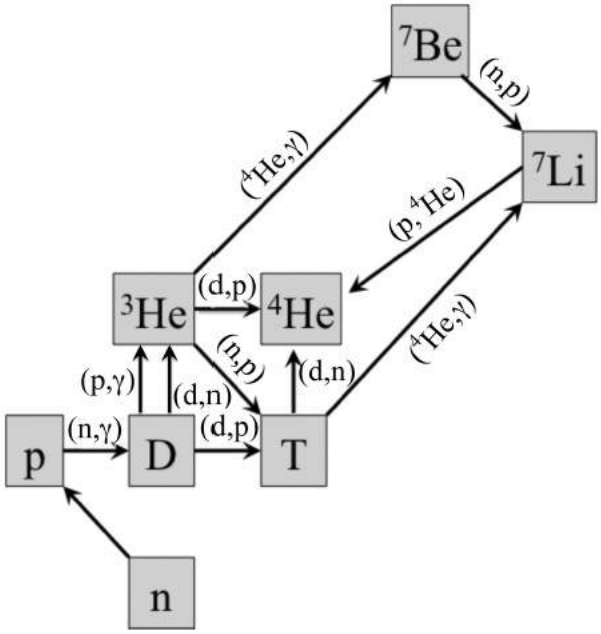


Figure 4.3.1: Big bang nucleosynthesis main reactions (from WikiCommons, author Pamputt).

All the other higher mass nuclei have much lower abundances, especially after atomic mass 5 and 8, for which the nuclei are unstable. Reactions like ${}^4He + T \rightarrow^7 Li + \gamma$ or ${}^3He +^4 He \rightarrow^7 Be + \gamma$ do occur but with much reduced probability. So hydrogen, helium, lithium and beryllium are essentially the only elements primordially produced.

The nuclei abundance depends critically on the baryon/photon ratio $\eta \approx 10^{-8}\Omega_b h^2$ (see next section) because if there are more photons then the time at which nuclei can form will be delayed due to the high energy tail of the photon distribution (see Eq (4.4.12) below). The quantity η is constant since both the number of baryons and photons scale as a^{-3}. The number of photons today can be directly measured from the CMB black body temperature $T = 2.7$K. Therefore from η we can obtain the density of baryons in the universe, i.e. $\Omega_b h^2$. Comparing with the theoretical predictions with the observed element abundances in the oldest stars (Fig. 4.3.2) we obtain

$$\Omega_b h^2 = 0.022 \pm 0.002$$

In combination with the estimates of $h \approx 0.7$ one obtains $\Omega_b \approx 0.05$. This shows that the cosmological baryon density is much smaller than the critical one.

4.4 Primordial nucleosynthesis, more details

Since the baryonic density decreases as T^3 and it is now (we put $m_B \approx 1GeV$)

$$n_B = \frac{\rho_B}{m_B c^2} = \Omega_b \frac{\rho_c}{m_B c^2} = (\Omega_b h^2) 10^{-47} GeV^3 \tag{4.4.1}$$

there is a constant ratio baryons/photons equal to

$$\eta = \frac{n_B}{n_\gamma} = 2.68 \cdot 10^{-8} (\Omega_b h^2) \tag{4.4.2}$$

(where Ω_b is the present value). We need also the low temperature limit $T \ll m$ of eqs. (4.1.2)

$$n = g(\frac{mT}{2\pi})^{3/2} e^{-\frac{m-\mu}{T}} \tag{4.4.3}$$

$$\rho = nm \tag{4.4.4}$$

$$p = nT \ll \rho \tag{4.4.5}$$

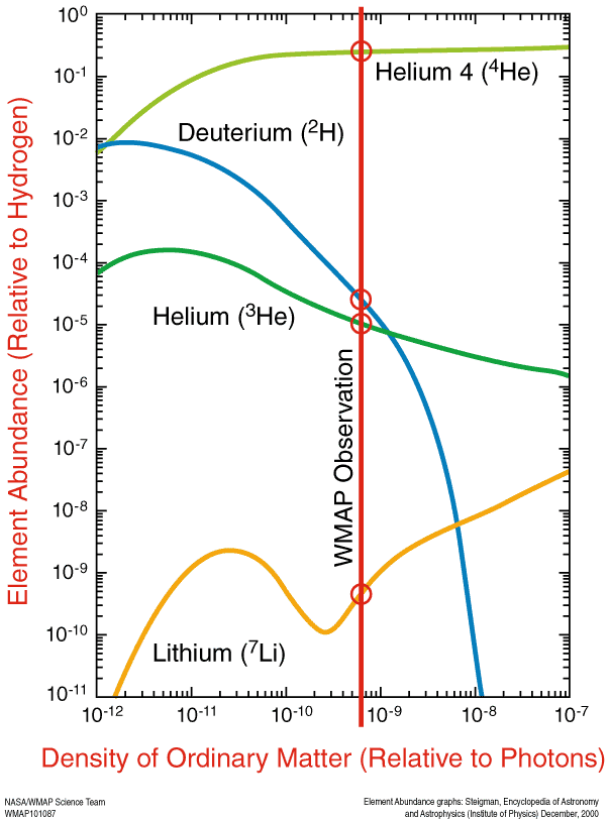


Figure 4.3.2: Big bang nucleosynthesis yields (NASA/WMAP Science team)

(one has to expand $\sqrt{m^2+k^2} \approx m(1+k^2/2m^2)$ in the exponential). Consider now the universe at a temperature $T \sim 1$ MeV. Since the binding energy of light nuclei as ${}^2H, {}^3H, {}^3He$ is 2.22, 6.92, 7.72 MeV, respectively, at $T \sim 1$ Mev the nuclei should be able to form, out of the neutron-proton plasma: as we show now, however, the actual temperature of formation is quite lower, due to the high photon-to-baryon ratio. In a thermal equilibrium their distributions are ($g = 2$)

$$n_n = 2(\frac{m_B T}{2\pi})^{3/2} e^{-\frac{m_n - \mu_n}{T}} \tag{4.4.6}$$

$$n_p = 2(\frac{m_B T}{2\pi})^{3/2} e^{-\frac{m_p - \mu_p}{T}} \tag{4.4.7}$$

where $m_B \approx m_n \approx m_p$. Let us consider now a reaction in which a nucleus ${}^A N_Z$ with Z protons and $A - Z$ neutrons forms. The distribution of such a nucleus in thermal equilibrium is, similarly, (we use the subscript A for the nucleus)

$$n_A = g_A(\frac{m_A T}{2\pi})^{3/2} e^{-\frac{m_A - \mu_A}{T}} \tag{4.4.8}$$

For the preservation of chemical potentials (i.e. conservation of energy) we have

$$\mu_A = Z\mu_p + (A - Z)\mu_n \tag{4.4.9}$$

This allows us to write the number density n_A of the nuclei as a function of $m_{n,p}, n_{n,p}$:

$$n_A = g_A 2^{-A} A^{3/2} (\frac{m_B T}{2\pi})^{3(1-A)/2} n_p^Z n_n^{A-Z} e^{\frac{B_A}{T}} \tag{4.4.10}$$

$$B_A \equiv Z m_p + (A - Z)m_n - m_A \tag{4.4.11}$$

where B_A is the (positive) binding energy of the nucleus, which as already mentioned is of order 1 MeV or larger. Let us define the mass fraction of nucleus N with respect to the baryonic density $n_B = \eta n_\gamma = \eta 2\zeta(3)T^3/\pi^2$:

$$X_A = \frac{A n_A}{n_B}$$

and similarly $X_{n,p} = n_{n,p}/n_B$. Then we obtain finally,

$$X_A = F(A)(\frac{T}{m_B})^{3(A-1)/2} \eta^{A-1} X_p^Z X_n^{A-Z} e^{B_A/T} \tag{4.4.12}$$

where

$$F(A) = g_A A^{5/2} \zeta(3)^{A-1} \pi^{(1-A)/2} 2^{(3A-5)/2} \tag{4.4.13}$$

is a factor of order unity. From (4.4.12) we see that the production of the nucleus A, defined as the epoch at which the value of $X_A \sim 1$ is reached, can only occur at temperatures T_A much lower than B_A, because of the small factor η^{A-1}, i.e. because of the large number of photons per baryon present (high entropy). In other words, one needs to assume a very small η in order to obtain small temperatures T_A, which in turn agree with the observed nuclei abundance. Neglecting all factors of order unity, and also putting $X_{n,p} \approx 1$, we can write numerically

$$T_A \approx \frac{B_A/(A-1)}{\log \eta^{-1} + 1.5 \log(m_B/T)} \tag{4.4.14}$$

We obtain then $T_A \approx 0.07, 0.11, 0.28$MeV for ${}^2H, {}^3He, {}^4He$, i.e. values much lower than their respective binding energies. Temperatures of the order of 0.1 MeV imply $a = T_0/T_A \approx 10^{-8}$ and therefore $z_{nucl} \approx 10^8$. You can also show that $t_{nucl} \approx 10^2 sec$.

4.5 Matter-radiation decoupling

Let us apply now similar arguments at a lower temperature, $T \approx 10$ eV, at which we expect that hydrogen atoms, whose binding energy is 13.6 eV, can form. At higher temperatures, the Compton and bremsstrahlung interactions keep the thermal equilibrium between matter and radiation. In this case, just as before, we have $T \ll m$ and we can write

$$n_i = 2(\frac{m_i T}{2\pi})^{3/2} e^{-\frac{m_i - \mu_i}{T}}$$

where i =electrons, protons and hydrogen atoms. In the reaction $p + e \to n_H + \gamma$ the chemical equilibrium requires $\mu_p + \mu_e = \mu_H$ and therefore, as in the previous case, we obtain

$$n_H = \frac{g_H}{g_p g_e} n_p n_e \left(\frac{m_e T}{2\pi}\right)^{-3/2} e^{B/T}$$

where $g_{p,e} = 2$ and $g_H = 4$, and $B = m_p + m_e - m_H = 13.6$ eV. We define now the ionization fraction for each species, $x_i = n_i/n_B$. Since $n_p = n_e$, and $n_p + n_H = n_B$ we have $x_p = x_e$ and $x_H = (n_H/n_B) = 1 - x_e$. Therefore we obtain a relation for x_e (Saha equation):

$$\frac{1 - x_e}{x_e^2} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \eta \left(\frac{T}{m_e}\right)^{3/2} e^{B/T} \approx 3.84\eta \left(\frac{T}{m_e}\right)^{3/2} e^{B/T}$$

From this equation we can track the evolution of the ionization fraction $x_e(T)$. The lower is x_e, the more advanced the recombination process is. For instance, we'll have an almost total recombination $x_e = 0.1$ at the temperature $\tau \equiv T/1eV$ where (by iteration)

$$\tau^{-1} = 3.084 - 0.00735 \log(\Omega_b h^2)$$

therefore since $z = (T/T_{2.7K}) = \tau(1eV/T_{2.7K}) \approx 4200\tau$ we have the *recombination redshift*

$$z \approx 1367(1 - 0.024 \log \Omega_b h^2)^{-1}$$

Then we see that the recombination of electrons and protons into hydrogen atoms can be said to be completed at $T \approx 0.3$eV, i.e. at a temperature quite smaller then the typical binding energy, analogously to the nucleosynthesis case.

When neutral atoms form, radiation cannot keep thermal equilibrium with matter. At some temperature T_{dec} the Thomson interaction becomes smaller than the expansion rate H (or in other words the photon mean free path becomes of the order of the Hubble radius) and matter decouples. This happens roughly at

$$z_{dec} \approx 1100$$

This is therefore the instant at which the photons last scatter off the electrons and protons and begin their free travel through space, some of them ultimately to be captured as CMB photons by our detectors.

Chapter 5

The distance ladder

Quick summary

- Distances are measured with a variety of ways, valid for different ranges and for different type of sources.
- The basic method, on which practically all the others rely, is the parallax method, valid up to 50 kpc, i.e. the distance of nearby satellite galaxies.
- One of the most important method to extend the distance range is based on the period-luminosity-color relation of the Cepheid variable stars, extending to 20 or so Mpc.
- Beside this distance one has to rely either on the Tully-Fisher and Fundamental Plane relation for spiral and elliptical galaxies, respectively, or on the standardized supernovae Ia, or on a number of alternative but less precise techniques.
- Using a combination of all these methods, the Hubble constant has been recently measured to be $H_0 \approx 72 \pm 3$ km/sec/Mpc.

5.1 The parallax method

The basic method to determine distance is the parallax method. All the other methods build on this first step. The idea (first successfully used in 1838 by Friedrich Bessel) is to find the change in direction to a nearby star, with respect to the Sun (or, better, to very distant stars), when the Earth is at opposite locations on the ecliptic (i.e. 2A.U.= 300 million kilometers apart). If there is no change, the star is at infinity. Simple trigonometry shows that if the parallax is 1 arcsec, then the star is roughly $2 \cdot 10^{13} km$ away, i.e. (by definition) 1 parsec= 3.26 light years. The distance in general is then

$$d = \frac{1\text{pc}}{\theta[\text{arcsec}]} \tag{5.1.1}$$

The satellites Hipparcos (1989) and Gaia (2013) measured parallaxes down to $10\,\mu$arcsec=10^{-5}arcsec, reaching therefore distances of the order of 100 kpc, much beyond the limits of the Milky Way (which has a size of 10kpc roughly). In particular the distance to the Large Magellanic Cloud (LMC) has been found to be

$$D_{LMC} \approx (50 \pm 1)\text{kpc} \tag{5.1.2}$$

This is a very important value since all the other methods give only relative measurements (i.e. are calibrated ultimately through the parallax method).

5.2 Cepheids

The second most important method is based on the variable stars known as Cepheids (from the constellation where the first example has been observed). These variable stars are red giant near the end of their life. The He

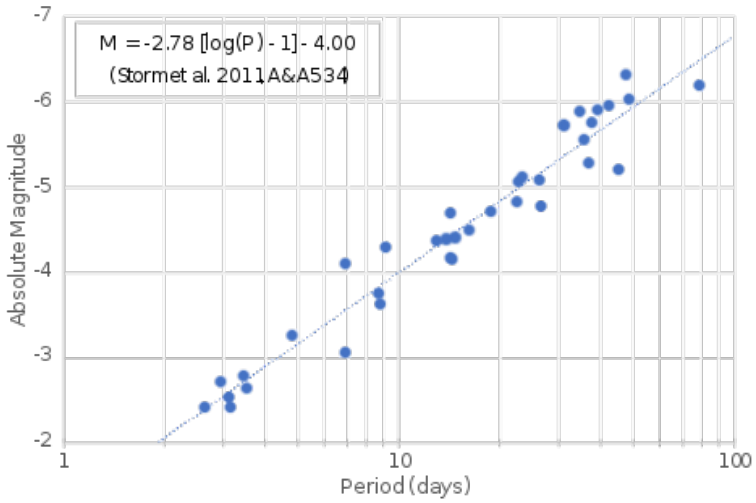


Figure 5.2.1: Period-Luminosity relation for Magellanic Cloud Cepheid variable stars; data from Storm et al. 2011, A&A, 534. (Wikicommons, author: Dbenford)

envelope is ionized when the temperature is high; a ionized He is highly opaque to radiation coming from the core which therefore is trapped and expand the envelope, making the star brighter. When the envelope expands, the temperature drops and the helium recombines (de-ionizes) so the envelope becomes more transparent. The radiation flows out, the pressure decreases, the envelope contracts, and the star becomes dimmer. The contraction however increases the temperature again and the cycle restarts, lasting from days to several weeks. The series of oscillations is very stable for some thousand years. Henrietta Leavitt around 1910 found that the period is larger is the peak absolute luminosity is larger according to roughly

$$L \sim P^{1.3} \tag{5.2.1}$$

In order to fix the proportionality constant in this relation (and to ensure that is really universal, i.e. it applies to all Cepheids), one needs to calibrate it, i.e. to measure the luminosity and the period for a number of nearby Cepheids for which we know the distance (and therefore the absolute magnitude) with the parallax method. This crucial step is necessary for all the methods we are going to discuss.

A recent determination from the Hubble Space Telescope on nearby Cepheids found (P is in days)

$$M_V = (-3.43 \pm 0.11)(\log_{10}(P) - 1) - (5.93 \pm 0.03) \tag{5.2.2}$$

This can used to determine distances directly as

$$5 \log d[Mpc] + 25 = V + 3.746(\log_{10}(P) - 1) - 2.523(V - I) + 5.959 \tag{5.2.3}$$

where V, I are the apparent magnitudes in two different wavebands, and a color correction due to the influence of metallicity on the period-luminosity relation, proportional to $V - I$, is included (Fig. 5.2.1)

Recently, Cepheids have been identified in galaxies of the Virgo cluster (15 Mpc away) and, still in progress, in the Coma cluster (65 Mpc/h). The Cepheids in the Virgo cluster have been fundamental in reducing drastically the error in the measurement of H_0 (Fig. 5.2.2)

5.3 Planetary nebulae

Planetary nebulae (PN) are stars with an extended envelope of hot gas ionized by the star itself. They appeared to early astronomers as "planets" because of their disk-like extended form. They are relatively easy to identify because most of the emission is in some specific line, eg. OIII. Their luminosity function (i.e. the function Φ_{PN} that gives how many PN one has in a luminosity range dL, $dN_{PN} = \Phi_{PN} dL$) has a sharp cut-off at high luminosity (see Fig. 5.3.1). Essentially, no PN is found with absolute magnitude brighter than -5. If one can collect many PN in a galaxy so as to determine this cutoff with good precision, then the value of the brightest PN can be used as a distance indicator.

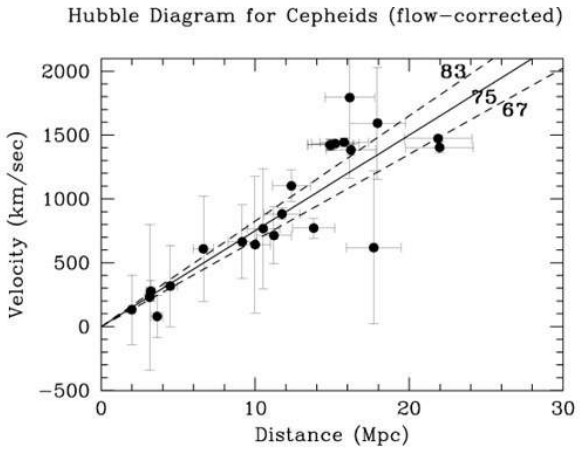


Figure 5.2.2: Hubble constant from the Cepheid method (Freedman et al., Ap. J., Volume 553, Issue 1, pp. 47-72, 2001. © *AAS. Reproduced with permission*).

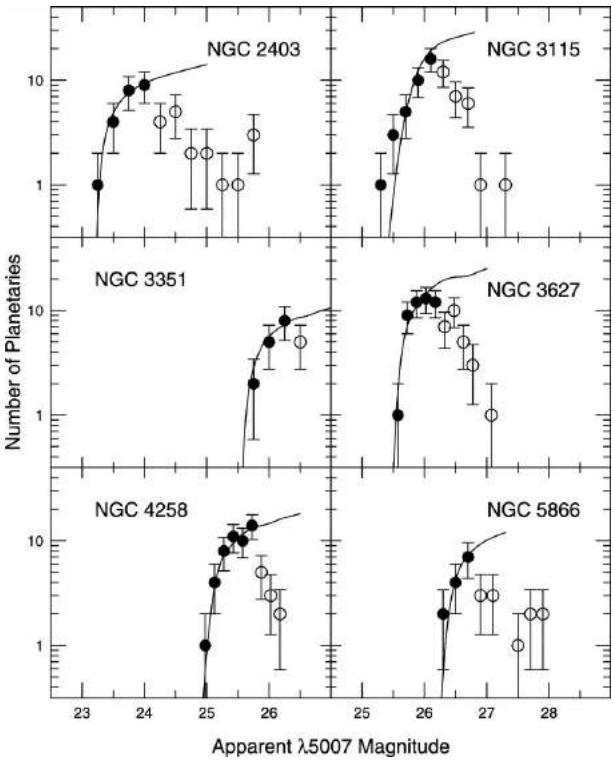


Figure 5.3.1: Luminosity functions of planetary nebulae in a sample of galaxies. Notice the sharp cut-off at high luminosity. (From Ciardullo et al. 2002, astro-ph/0206177, Astrophys.J. 577 (2002) 31-50. © *AAS. Reproduced with permission.*)

5.4 Surface Brightness Fluctuations

Another method exploits the fact that in a random process the fluctuations have a variance proportional to the number of trials. For instance, if we throw randomly balls inside a grid, without aiming at any cell in particular, and if $\bar{N}$ is the average number of balls per cell, the typical fluctuation (i.e. the square root of the variance) of the number of balls is $\sqrt{\bar{N}}$. More exactly, the variance

$$\sigma^2 = \frac{1}{n_c}\sum_i^{n_c}(N_i - \bar{N})^2 \approx \bar{N} \tag{5.4.1}$$

where n_c is the number of cells in the grid and N_i the count in the i-th cell. The relation is exact in the limit of an infinite number of cells.

So in a galaxy one can divide the image into many little pixels and estimate the fluctuations in the surface brightness (SFB) (i.e. the flux divided by area of the pixel). The variance σ^2_{SBF} in SBF goes like the average number of stars in the pixel N, and the fluctuation as $\sqrt{N}$. The total luminosity of the galaxy however goes obviously like N. So we have

$$L = \alpha\sigma^2_{SBF} \tag{5.4.2}$$

and we expect α to be a universal quantity since the random events that distributed stars in a galaxy should not depend on the galaxy distance.

Once again, if one estimates α with a sample of known distance and therefore known luminosity, one can use this relation for distant galaxies. Of course the calibration will be tighter if one uses galaxies of the same type

5.5 Tully-Fisher relation and the Fundamental Plane

Spiral galaxy stars have little random motion and large ($\approx 200 - 300$ km/sec) rotational velocity; in contrast, elliptical galaxy stars have little rotational velocities (less then 100 km/sec) but large random motion (several hundreds of km/sec). In both cases, the velocities seem correlated with the absolute luminosity: brighter galaxies (therefore very likely more massive) contain faster stars. This is entirely to be expected because of the virial theorem that links (twice) the average kinetic energy to the average gravitational potential energy

$$\langle v^2\rangle = \langle\Phi\rangle = \frac{GM}{R_{vir}} \tag{5.5.1}$$

The velocity of stars with respect to the galaxy center is estimated by measuring the width of emission lines: a line emitted by several sources with some velocity dispersion will be a superposition of several lines with slightly displaced wavelength due to Doppler effect: the result is a wider line (for ellipticals) or a double-peaked line (for spirals, where stars are either moving relatively towards us or away from us), see Fig. 5.5.1.

The Tully-Fisher relation valid for spiral galaxies read

$$L \sim \sigma^\alpha_{rot} \tag{5.5.2}$$

with $\alpha \approx 3.5 - 4$, depending on the band at which L is measured (Fig. 5.5.2).

For elliptical galaxies one finds empirically a number of relations among four quantities: the effective radius R_e, the surface brightness within that radius I_e, the luminosity L and the velocity dispersion σ_V. In particular one has the Faber-Jackson relation

$$L \sim \sigma^\alpha_V \tag{5.5.3}$$

$$\alpha \approx 3 - 4 \tag{5.5.4}$$

and also

$$R_e \sim I_e^{-0.83\pm0.08} \tag{5.5.5}$$

Moreover, by definition, $L_e = \pi R_e^2 I_e$. One finds also $I_e \sim L^{-3/2}$ (brighter ellipticals have smaller surface brightness) and the fundamental plane for ellipticals, i.e. a relation among three of the four variables, for instance

$$R = c\cdot\sigma^a_V I^b \tag{5.5.6}$$

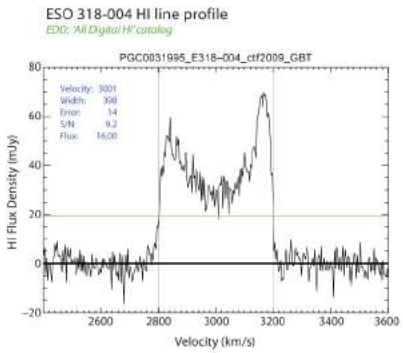


Figure 5.5.1: Double-peaked line from a spiral galaxy. (From the website https://www.ipnl.in2p3.fr/projet/cosmicflows/).

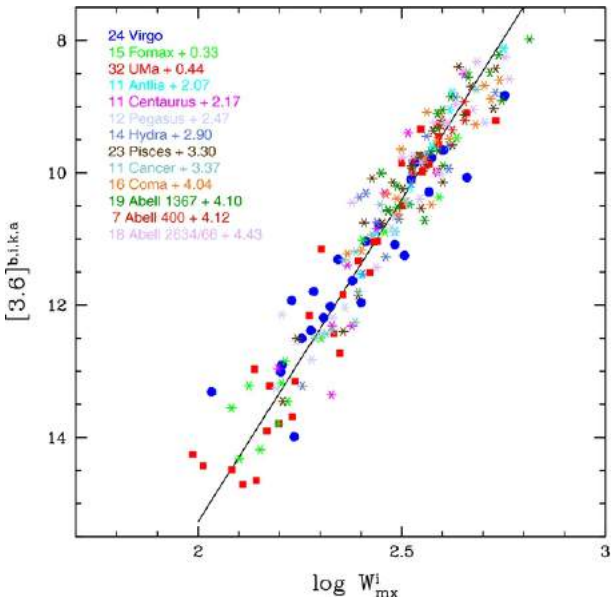


Figure 5.5.2: Tully-Fisher relation in the IR band around 3.6μm (Sorce et al., 2013ApJ...765...94S, © *AAS. Reproduced with permission*) for galaxies up to roughly 200 Mpc.

where σ_V is in km/sec and R is in kpc, a, b, c are constants, or equivalently (redefining b, c)

$$\log R_e = a \log \sigma_V + b\mu + c \tag{5.5.7}$$

where $\mu = -2.5 \log I$ (measured in mag arcsec^{-2}, analogously to the definition of magnitudes). A recent calibration through 1430 nearby galaxies gives $a \approx 1.5$, $b \approx 0.3$ $c \approx -9$ (La Barbera et al. ApJ 2008).

Once again, from the absolute luminosity and the measured flux, or from the absolute radius and the measured angular size, one can obtain the luminosity or the angular diameter distance, respectively.

5.6 Supernovae Ia

In 1998 Riess *et al.* [High-redshift Supernova Search Team (HSST)] and Perlmutter *et al.* [Supernova Cosmology Project (SCP)] independently reported the late-time cosmic acceleration by observing distant supernovae of type Ia (SN Ia). Up to 1998 Riess *et al.* had discovered 16 high-redshift SN Ia together with 34 nearby supernovae, while Perlmutter *et al.* had found 42 supernovae in the redshift range $z = 0.18$-0.83.

The explosion of supernovae is extremely luminous and causes a burst of radiation. The supernovae can be classified according to the absorption lines of chemical elements (Fig. 5.6.1). If the spectrum of a supernova includes a spectral line of hydrogen, it is classified Type II. Otherwise it is called Type I. If a supernova contains an absorption line of singly-ionized silicon, it is further classified Type Ia (note that Type Ib contains a line of helium, whereas Type Ic lacks the lines of both silicon and helium). The explosion of Type Ia occurs when the mass of a white dwarf in a binary system exceeds the Chandrasekhar limit by absorbing gas from a companion star (Fig. 5.6.2). A white dwarf is a relatively simple object which undergoes passive cooling in equilibrium between gravity and the electron degeneracy pressure, without nuclear fusion. When the mass exceeds the limit, the star collapses but contrary to normal stars the degeneracy pressure cannot increase because it is independent of temperature. The temperature increases so fast that the heavy nuclei (mostly C, O) fuse (that is why these supernovae are also called thermonuclear SNa) and the thermal motion of the particles exceeds the escape speed and the star explodes. Ultimately the typical form of the SN Ia lightcurve depends on the subsequent radioactive decay of the Ni, Fe elements from the exploding envelope, which takes almost a month before fading off. The Chandrasekhar limit depends only on fundamental constants

$$M_{Ch} = c \cdot \left(\frac{\hbar c}{G}\right)^{3/2} \frac{1}{(\mu_e m_H)^2} \approx \frac{M_P^3}{m_p^2} \approx 1.4 M_{\odot} \tag{5.6.1}$$

where M_P is the Planck mass, c is a constant of order unity, m_p is the proton mass, μ_e is the average molecular weight per electron (it depends then on the chemical composition of the star). One expects therefore that the energy emitted by the disruption of such an object is an almost universal constant. To test this hypothesis is necessary as usual a calibration, i.e. a sample of nearby SNIa of which we know the distance with other means, and from which we can obtain therefore their absolute magnitude M_{SN}. Then the SN Ia become a kind of "standard candle" by which luminosity distance can be measured observationally.

In reality things are more complicated than this simple view. The intrinsic spread in absolute magnitudes is actually too large to produce stringent cosmological constraints. However, at the end of the 1990s, a high quality sample of "local" (i.e. $z \ll 1$) supernovae allowed to correlate the absolute magnitude with the width of the light curve: brighter supernovae have a broader light curve, roughly $L \sim \tau^{1.7}$ if τ is a measure of the width of the lightcurve (Fig. 5.6.3). By measuring at the same time the apparent magnitude and the light curve it is possible therefore to predict the absolute magnitude. Although in the following we refer to a universal SN Ia absolute magnitude, we always mean the magnitude corrected for the light curve width.

Since the (corrected) peak absolute magnitude M_{SN} is the same for any SN Ia under the assumption of standard candles, the luminosity distance $d_L(z)$ is obtained from measuring m and the relation

$$5 \log d_L = m - M_{SN} - 25 \tag{5.6.2}$$

At small redshifts, $d_L = cz/H_0$ and we can therefore infer H_0. At large redshifts SNIa can be employed to measure $d_L(z)$ as a function of cosmological parameters, as we will see in another chapter.

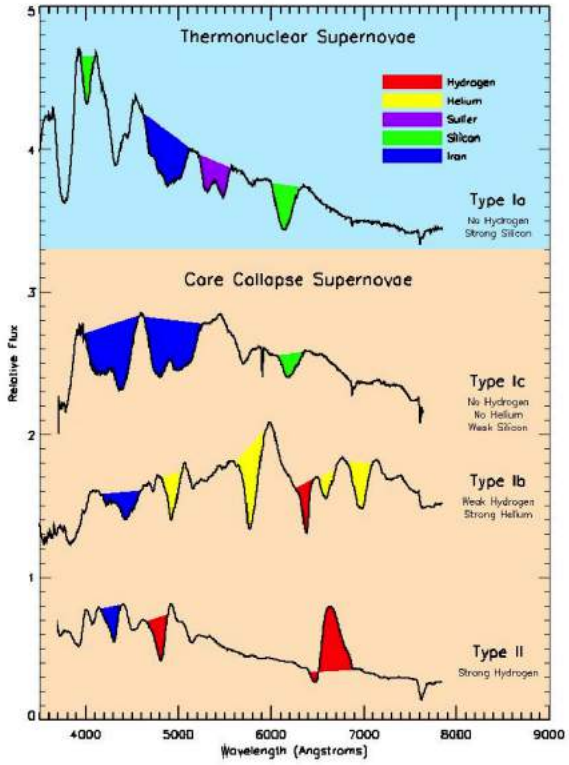


Figure 5.6.1: Classification of SN types. The different spectra make possible a reliable identification of SN Ia (from `supernova.lbl.gov`).

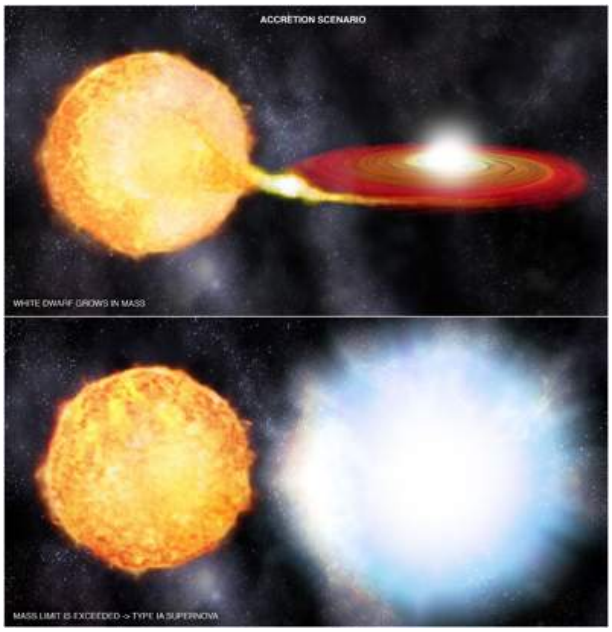


Figure 5.6.2: Pictorial representation of a white dwarf accreting matter from a red giant companion and then exploding as a SNIa (Image: NASA/CXC/M Weiss).

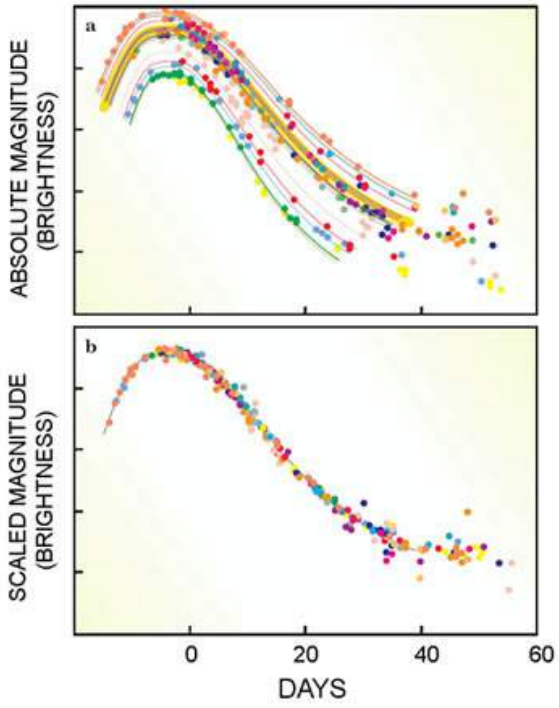


Figure 5.6.3: Calibration of nearby SNIa. Above: several slightly different lightcurves: brighter SN have a more stretched lightcurve than dimmer ones. Below: "standardized" lightcurve obtained by a stretch factor applied to the lightcurves (Hamuy et al. 1996, from `supernova.lbl.gov`).

Chapter 6

Accelerated expansion

Quick summary

- Supernovae Ia are probably the thermonuclear explosion of white dwarfs accreting matter near or beyond the Chandrasekhar mass from a companion star
- At the luminosity peak, SNIa are as bright as an entire galaxy; the light curve fades away in a month roughly
- Since the Chandrasekhar mass depends only on fundamental constants, one could reasonably expect that the energy emitted during the explosion is roughly constant
- This expectation is confirmed by observations of local SNIa; a correction of the peak luminosity with the light curve width makes the SNIa excellent standard candles
- Since they can be seen at very large distances, SNIa can be used as distance indicator to $z \approx 1$
- In 1998, the Hubble diagram of SNIa showed for the first time an accelerated expansion of the universe, favoring the existence of a large component of a cosmological constant or dark energy; this discovery was awarded the Nobel Prize in 2011
- There are many models of dark energy that expand upon the idea of a large, weakly clustered, unseen, energy component with negative pressure

6.1 SNIa at high redshifts[a]

In 1998 Riess *et al.* and Perlmutter *et al.* released observational data of the apparent luminosity of high-redshift Type Ia supernovae ($0.2 \lesssim z \lesssim 0.8$). The data of low-redshift regions ($z < 0.1$) reported previously was also used in their analysis. Let us pick up a few examples of data to understand how the luminosity distance is known observationally. First, consider two data of the apparent magnitudes in the low redshift region of SN Ia: (i) 1990O: $m = 16.26$ ($z = 0.03$) and (ii) 1992bg: $m = 16.66$ ($z = 0.036$). Since the luminosity distance in the region $z \ll 1$ is well approximated by $d_L \simeq cz/H_0$, the absolute magnitude M is known. We take the value $h = 0.7$ for the Hubble constant. We then obtain $M = -19.29$ and $M = -19.28$ for 1990O and 1992bg, respectively. This shows that the absolute luminosity of SN Ia is nearly constant ($M \simeq -19$), as we already mentioned.

Let us next use the high-redshift data reported by Perlmutter *et al.* . Consider the two SN Ia data of the apparent magnitudes: (a) 1997R: $m = 23.83$ ($z = 0.657$), (b) 1995ck: $m = 23.57$ ($z = 0.656$). Employing the value $M = -19.15$ for the absolute magnitude, we find that the luminosity distance is given by $H_0 d_L/c = 0.920$ for 1997R and $H_0 d_L/c = 0.817$ for 1995ck. Notice that the approximation $d_L \simeq cz/H_0$ is no longer valid in the

[a]Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

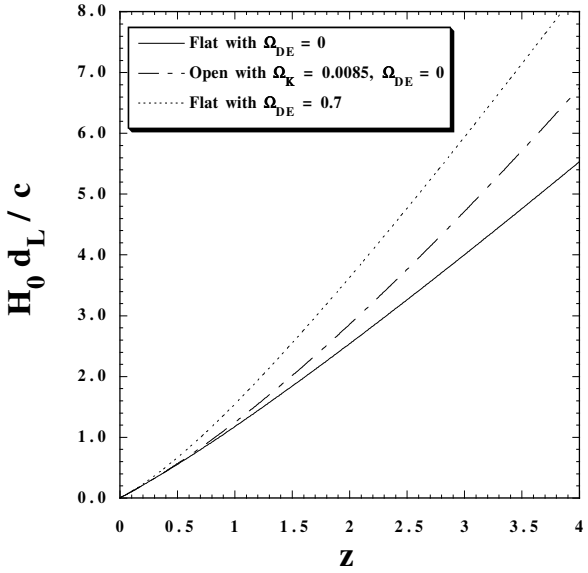


Figure 6.0.1: The luminosity distance d_L versus the redshift z for three cases: (a) a flat universe without dark energy, (b) an open universe ($\Omega_K^{(0)} = 0.0085$) without dark energy, and (c) a flat universe with the cosmological constant ($\Omega_{\rm DE}^{(0)} = 0.7$ and $w_{\rm DE} = -1$). The presence of dark energy leads to a larger luminosity distance relative to the case without it. In the open universe the luminosity distance also gets larger than that in the flat universe.

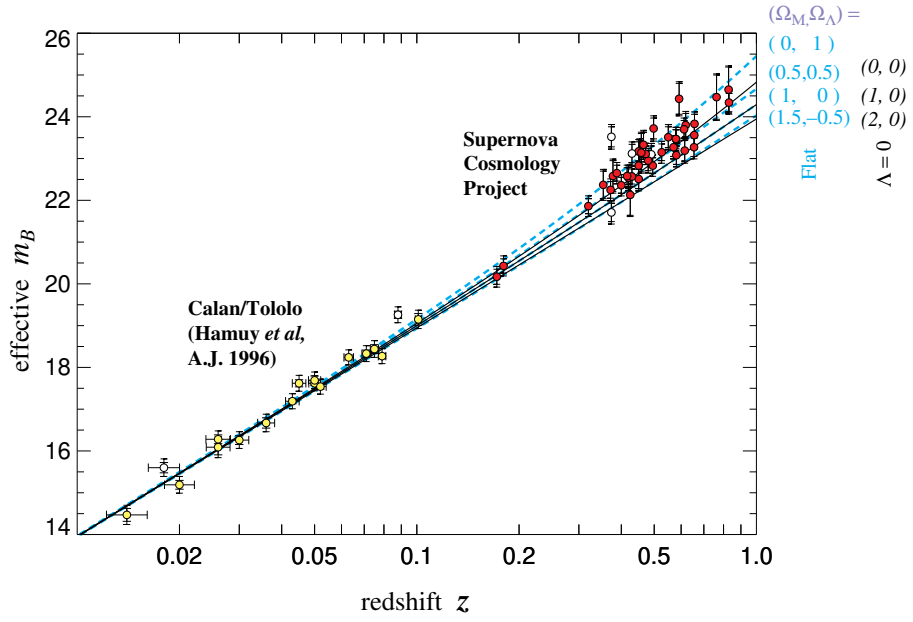


Figure 6.1.1: The effective apparent luminosity m_B versus the redshift z for 42 high-redshift SN Ia from the SCP and 18 low-redshift SN Ia from the Calan/Tololo Supernova Survey. The solid curves are the theoretical prediction for m_B for a number of cosmological models without the cosmological constant: $(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = (0, 0)$ (top), $(1, 0)$ (middle), and $(2, 0)$ (bottom). The dashed curves correspond to a number of flat cosmological models: $(\Omega_m^{(0)}, \Omega_\Lambda^{(0)}) = (0, 1)$ (top), $(0.5, 0.5)$ (second from top), $(1, 0)$ (third from top), and $(1.5, -0.5)$ (bottom). From Perlmutter et al. 1998.

high-redshift regime. Let us consider a flat universe with a dark energy equation of state $w_{\rm DE} = -1$ (i.e. the cosmological constant). Since $E(z) = [\Omega_m^{(0)}(1+z)^3 + \Omega_{\rm DE}^{(0)}]^{1/2}$ in this case, the luminosity distance reads

$$d_L(z) = \frac{c(1+z)}{H_0}\int_0^z \frac{{\rm d}\tilde{z}}{[(1-\Omega_{\rm DE}^{(0)})(1+\tilde{z})^3 + \Omega_{\rm DE}^{(0)}]^{1/2}}\,, \tag{6.1.1}$$

which can be evaluated numerically for given $\Omega_{\rm DE}^{(0)}$. In order to satisfy the observational data $(H_0/c)d_L(z = 0.657) = 0.920$ for 1997R, we require that $\Omega_{\rm DE}^{(0)} = 0.70$. Similarly we get $\Omega_{\rm DE}^{(0)} = 0.38$ from the 1995ck data. Both data indicate the existence of dark energy.

Since observational data are prone to statistical and systematic errors, a few data points are not enough to conclude that the present universe is accelerating. Using 42 high-redshift SN Ia at redshifts between 0.18 and 0.83 together with 18 low-redshift SN Ia data from the Calan/Tololo Supernova Survey, Perlmutter *et al.* (1998) showed that the cosmological constant is present at the 99 % confidence level. They also found that the open universe without the cosmological constant does not fit the data well. From Eq. (3.8.4) the apparent luminosity m gets larger for increasing luminosity distance d_L. Figure 6.1.1 shows that the observational data in the high redshift regime favor the luminosity distance larger than the one predicted by the CDM model ($\Omega_m^{(0)} = 1$ and $\Omega_\Lambda^{(0)} = 0$). This means that the SN are on average dimmer than expected in pure flat CDM. Dimmer sources imply, in turn, a larger luminosity distance and, therefore, a smaller value of H in the past: the conclusion is then that the Universe is accelerating.

From a full statistical analysis of the SN Ia data accumulated by the year 1998, Perlmutter *et al.* found that the density parameter of non-relativistic matter is constrained to be $\Omega_m^{(0)} = 0.28^{+0.09}_{-0.08}$ (1σ statistical) in the flat universe with the cosmological constant. After 1998 more SN Ia data have been collected by a number of high-redshift surveys–including SuperNova Legacy Survey (SNLS), Hubble Space Telescope (HST), and "Equation of State: SupErNovae trace Cosmic Expansion" (ESSENCE) survey. The SNLS project, which is based on the Canada-France-Hawaii Telescope, consists of two components: (i) a large imaging survey to detect about 2000 supernovae and monitor their light curves, and (ii) a large spectroscopic survey to obtain supernovae identification and redshift. The HST survey is based on the image subtraction to search the SN Ia data in the high redshift region $z > 1$ by including search depth, efficiency, timing and false-positive discrimination. The ESSENCE project is a ground-based survey designed to detect about 200 SN Ia in the redshift range $z = 0.2$-0.8 to measure the equation of state of dark energy to better than 10 %. In Fig. 6.1.3 the observational contours on $(\Omega_m^{(0)}, w_{\rm DE})$ are plotted from the Union2 catalog of SN Ia by Amanullah *et al.* . Note that the equation of state of dark energy is assumed to be constant. While the SN Ia data alone are not yet sufficient to place tight bounds on $w_{\rm DE}$, Fig. 6.1.3 clearly shows the presence of dark energy responsible for the late-time cosmic acceleration ($w_{\rm DE} < -1/3$). A modern publicly available catalog that includes most of the known supernovae is called JLA (Betoule et al. 2014, arXiv:1401.4064).

When combined with the results from CMB and from galaxy clustering on the plane Ω_Λ, Ω_m, the three probes intersect in a point $\Omega_m = 0.3, \Omega_\Lambda = 0.7$, which is denoted "concordance cosmology" (Fig. 6.1.4) or ΛCDM ($\Lambda-$Cold Dark Matter) model.

If the equation of state of dark energy varies in time, we need to parametrize $w_{\rm DE}$ as a function of the redshift z. A popular parametrization is the Chevalier-Polarski-Linder parametrization

$$w_{DE} = w_0 + w_a(1-a) = w_0 + w_a\frac{z}{1+z} \tag{6.1.2}$$

which can be thought as a Taylor expansion of an unknown $w_{DE}(a)$ around $a = 1$. The present value is then w_0 and the asymptotic future one is $w_0 + w_a$. Now from the conservation equation of ρ_{DE} we have

$$\dot{\rho}_{DE} + 3H\rho_{DE}(1 + w_{DE}(a)) = 0 \tag{6.1.3}$$

which can be integrated as

$$\int \frac{d\rho_{DE}}{\rho_{DE}} = -3\int \frac{da}{a}(1+w_{DE}) = -3\int \frac{da}{a}(1+w_0+w_a(1-a)) = -3(1+w_0+w_a)\log a + 3w_a(a-1) \tag{6.1.4}$$

so that

$$\rho_{DE} \sim a^{-3(1+w_0+w_a)}e^{3w_a(a-1)} \tag{6.1.5}$$

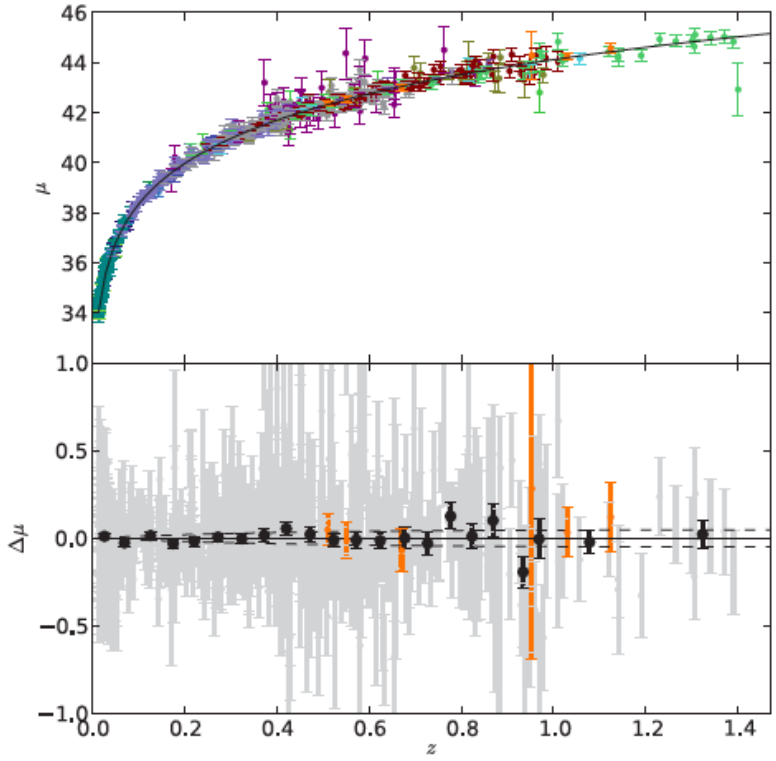


Figure 6.1.2: Union2 combined catalog of 557 SNIa. Here $\mu = 5 \log d_L + const$. From Amanullah et al. 2010 ApJ...716..712A *(© AAS. Reproduced with permission).* Bottom panel: residuals.

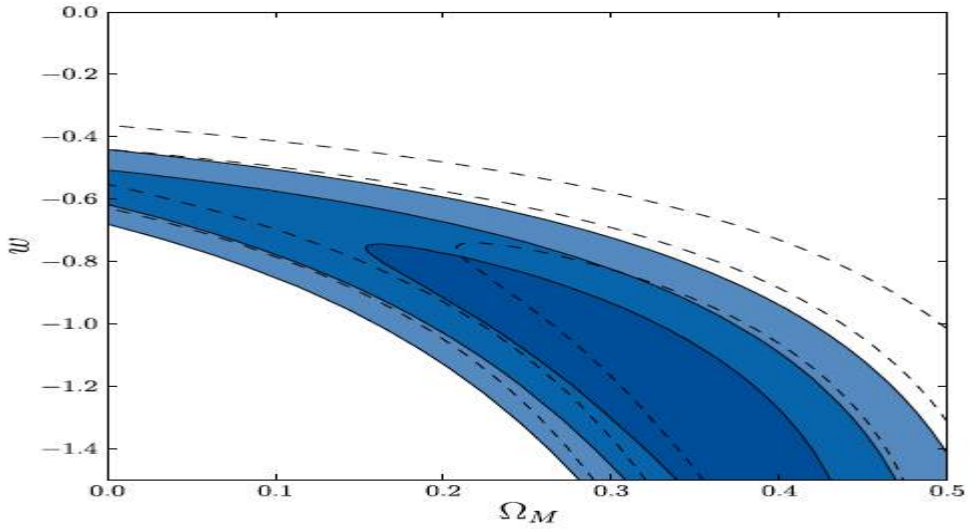


Figure 6.1.3: 68.3 %, 95.4 % and 99.7 % confidence level contours on $(\Omega_m^{(0)}, w_{\rm DE})$ from the SN Ia observations (denoted as Ω_M and w in the figure) compiled in Amanullah et al. 2010ApJ...716..712A *(© AAS. Reproduced with permission).*

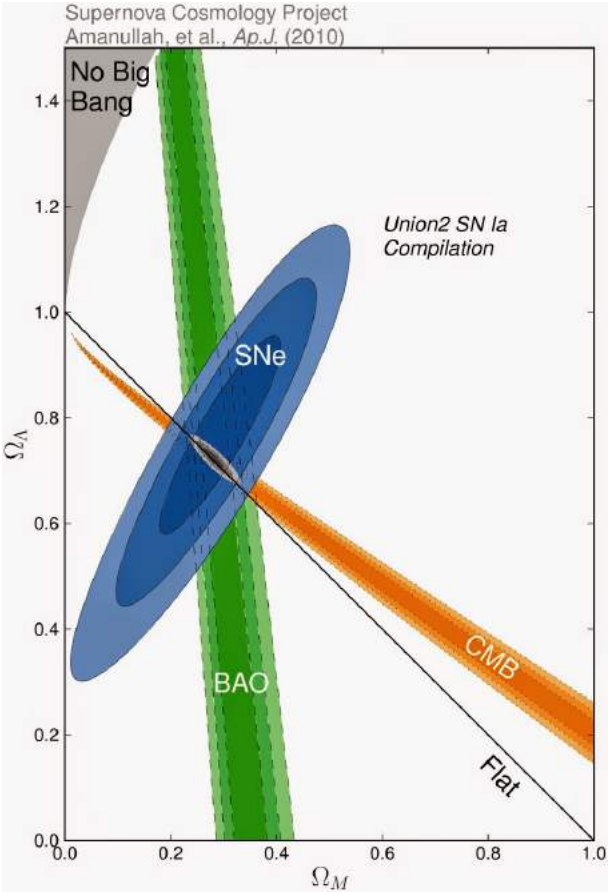


Figure 6.1.4: Combined results from supernovae, CMB and clustering of galaxies (BAO). The point where the three probes intersect is called "Concordance cosmology". (From Amanullah et al. 2010ApJ...716..712A, © *AAS. Reproduced with permission*).

If the SN Ia data are accurate enough to measure the luminosity distance $d_L(z)$ in terms of z, it is possible to determine the evolution of any function $w_{DE}(z)$ by using

$$H^2 = H_0^2(\Omega_{m0}(1+z)^3 + \Omega_{DE,0}f(z) + (1-\Omega_{m0}-\Omega_{DE,0})(1+z)^2) \tag{6.1.6}$$

where, in general

$$f(z) = \exp\int_0^z \frac{3(1+w_{DE}(z))}{1+z}dz \tag{6.1.7}$$

However there is very little information on $w_{\rm DE}$ at $z > 0.5$ so $w_{DE}(z)$ cannot be well reconstructed beyond this. Any specific model of dark energy will in general provide a particular form of $w_{DE}(z)$ that will depend on the theoretical parameters.

6.2 Models of dark energy

So far we have assumed that dark energy only changes the background expansion, i.e. $H(z)$. However, in general any new component added to the cosmic fluid will change the full set of equations of general relativity and will also respond to the gravitational fluctuations like all other components, with the only exception of the pure cosmological constant. Although a pure Λ is a very good fit to practically all observational data, it is still possible and interesting to see if the data can tell us more about possible alternatives.

The simplest possibility to add a component that does not violate the high degree of homogeneity and isotropy of the universe is a scalar field, which has no direction. Such a field must be thought as a sort of new "gravitational" potential extending everywhere in space and evolving in time. Since the matter clustered in galaxies and clusters seem to exhaust all the pressureless matter in the universe, whatever is needed to reach $\Omega_{tot} \approx 1$ and to accelerate the expansion must be distributed almost homogeneously and have an effective equation of state close to -1. Below the scale corresponding to its Compton wavelength the field will not cluster. If the mass of the field is very small, the Compton scale might be larger than the galaxy cluster scale. The Compton wavelength

$$\lambda = \frac{\hbar}{mc} \tag{6.2.1}$$

corresponding to the size of a cluster of galaxies (1 Mpc) is roughly obtained with a mass $m \approx 10^{-28}$eV. Such an incredibly small mass cannot be observed in an accelerator (even if as large as a galaxy!) and can only be observed via astrophysical experiments.

A scalar field with self-interaction potential $V(\phi)$ can be described in FRW as a perfect fluid with energy density and pressure

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{6.2.2}$$

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{6.2.3}$$

The equation of state $w = p/\rho$ is therefore a function of time through $\phi(t)$; this can be obtained by solving the conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0 \tag{6.2.4}$$

i.e. a Klein-Gordon equation in an expanding universe:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0 \tag{6.2.5}$$

The potential $V(\phi)$ is of course unknown and can be partially reconstructed from observations of w. The scalar field mass is simply $m^2 = d^2V/d\phi^2$, itself in general a function of time.

Cosmologists have proposed an infinity of scalar field models, inspired by new physics (supersymmetry, extra dimensions) or just by trial and error hypotheses. Just like the Higgs field, a scalar field can couple to any other matter component, so one expects that the dark energy field might carry a new force (fifth force). In fact, every model of scalar fields is characterized by its self-interaction potential and by coupling functions to standard matter particles. The new force will be unscreened and will extend to very large distances, so it will appear as a modification of gravity. In fact, one can write a modified Poisson equation

$$\nabla^2\Phi = -4\pi G Y(a,k) a^2 \bar{\rho}\delta \tag{6.2.6}$$

that replaces the standard Poisson equation, where $Y(a,k)$ is a new unknown function of time and space. Then the linear growth equation (see Sec. 8.7) becomes

$$\delta_m'' + \frac{1}{2}(1 - 3w_{eff})\delta_m' - \frac{3}{2}Y(a,k)\Omega_m\delta_m = 0 \tag{6.2.7}$$

It will be therefore possible to constrain dark energy also by measuring gravity at large scales through the growth of linear perturbations.

Chapter 7

Cosmic inflation

Quick summary

- The Universe history can be classified into a number of epochs according to the photon temperature
- The standard decelerated model filled by matter and radiation has two important shortcomings: the problem of flatness and the problem of the horizon
- An epoch of primordial accelerated expansion, known as inflation, solves these problem
- Inflation predicts a total amount of matter such that the present Universe is spatially flat
- Inflation is based on the existence of a phase transition or a slow-rolling of a scalar field
- As we will see in the following, it also predicts the formation of initial inhomogeneities
- In this chapter we denote
  – with subscript '0' quantities at the present epoch
  – with subscript 'i' a fixed initial epoch
  – with subscript 'eq' the matter-radiation equivalence
  – with subscript 'dec' the decoupling epoch
  – with subscript 'e' the end of inflation

7.1 A short history of the Universe

The history of Universe evolution is foremost a thermal history, since the physical processes are dramatically different depending on the mean photon energy. Since the universe is expanding, the temperature T decreases. As long as the photon gas remains a black body we have that $\rho_\gamma \sim T^4$ and on the other hand $\rho_\gamma \sim a^{-4}$, so that $T \sim a(t)^{-1}$. Naturally this relation is violated whenever some process injects energy in the photon gas, e.g. when pairs $e^- e^+$ annihilates out of equilibrium.

Along the universe cosmic history we can identify a number of phases, the first four still hypothetical while the latter two rather established:

- Quantum gravity. $T \sim 10^{19} GeV$ (Planck energy), $t \sim 10^{-43} sec$.
- Baryogenesis. $10^{17} GeV < T < 10^2 GeV$ [perhaps 10^{15} GeV, $t \sim 10^{-35} sec$].
- Electroweak transition. $T \sim 10^3 GeV$ (mass of weak bosons, $Z^0, W^\pm$).
- Quark-Hadron transition. $T \sim 1 GeV$ (nucleon mass).

- Nucleosynthesis. $T \sim 1MeV$, $t \sim 3$min (nuclear levels).
- Recombination. $T \sim 10eV$, $t \sim 10^6 yr$ (atomic levels).

As well known we have currently $T = 2.75K \sim 1$meV.

The baryogenesis is the epoch in which the universe becomes asymmetric in the baryonic number. We observe today in our galaxy neighborhood an almost absolute absence of antimatter, for instance in the cosmic rays particles. If one starts with an equal number of particles and antiparticles at the density of the early universe and complete symmetry of interaction rates, after just a few seconds the annihilations would have produced a matter density equal to the present one and no matter would be left over for us today. At $t = 10$sec the physical horizon was about equal to 10 light-seconds; the subsequent expansion $a(t_{now})/a(10sec) \approx 10^8$ assuming a radiation epoch would have stretched out this distance to a few tens of parsecs, too short a distance to hide the antimatter. It is then necessary to invoke processes that violated the conservation of the baryonic number. These processes probably took place at an energy of 10^{15}GeV.

7.2 The problems of the standard model

Among the motivations that led A. Guth in 1981 to propose the first inflationary model two are fundamental for the standard cosmological model so far exposed. These are the so-called *flatness problem* and *horizon problem*.

In the Friedmann equations, the solution with zero spatial curvature, $\Omega = 1, k = 0$, is a highly unstable trajectory. A small deviation from it leads to a rapid departure, either towards a collapse ($\Omega \to \infty$) or towards complete emptying ($\Omega \to 0$). Yet we know that today $\Omega_0 \sim 1$. This is the *flatness problem*.

Let us see this more in detail. As we have seen, the Einstein equation in the FRW metric are (in Planck units)

$$(0,0) \quad : \qquad H^2 = \frac{8\pi}{3}\rho - \frac{k}{a^2}\,, \tag{7.2.1}$$

$$\text{Trace} \quad : \qquad \dot{H} + 2H^2 = \frac{4\pi}{3}(\rho - 3p) - \frac{k}{a^2}\,. \tag{7.2.2}$$

Subtracting the first from the second one it follows also

$$\ddot{a}/a = -\frac{4\pi}{3}(\rho + 3p)\,. \tag{7.2.3}$$

Therefore, every form of matter with $\rho + 3p > 0$ will follow a decelerated evolution. By putting $p = w\rho$, there will be deceleration for $w > -1/3$. From the conservation equation it follows then that

$$\dot{\rho} + 3H(\rho + p) = 0\,, \tag{7.2.4}$$

from which

$$\rho \sim a^{-3(1+w)}\,. \tag{7.2.5}$$

In the case of a flat space, from (7.2.1) we find

$$\frac{\dot{a}}{a} = \sqrt{\frac{8\pi}{3}} a^{-3(1+w)/2}$$

from which, integrating we obtain

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

where we observe that an accelerated behavior t^n with $n > 1$ requires indeed $w < -1/3$.

By defining $\rho_c = 3H_0^2/8\pi$ (present critical density) and $\Omega_m = \rho_{m,0}/\rho_c$ (present density parameter), and leaving w as a free parameter, Eq. (7.2.1) can be rewritten as

$$H^2 = H_0^2[\Omega_m a^{-3(w+1)} + \Omega_k a^{-2}]$$

where the present curvature parameter is $\Omega_k \equiv -k(H_0 a_0)^{-2} = 1 - \Omega_m$. Now we can define the curvature or the "deviation from flatness" $C(a)$ as

$$C(a) = \frac{|\Omega_m - 1|a^{-2}}{\Omega_m a^{-3(w+1)}} = C_0 a^{1+3w}\,, \tag{7.2.6}$$

where as usual C_0 is the present value. We see then that in the past, for $a \ll 1$, C tends to zero. That is, space in the past was much flatter than today. Now, since $a \sim T^{-1}$, it follows, as a function of temperature, that the curvature at the epoch i was

$$C_i = C_0 (T_0/T_i)^{1+3w}\,. \tag{7.2.7}$$

We know that today $T_0 \sim 3K \sim 10^{-12} GeV$, while for T_i we can assume the Planck value $T_i = 10^{19} GeV$. If today $|\Omega_m - 1| < 0.1$, as indeed it appears from the recent cosmic background data, we have that $C_i < 0.1 \cdot 10^{-31(1+3w)}$. Now, if $w = 1/3$ (radiation), it follows that Ω_i at Planck time was close to unity to within a part over 10^{-63}! If it was just a bit larger than this, e.g. one part over say 10^{-59}, today we would have $C_0 = 10^4$ and therefore $\Omega \to 0$, i.e. an almost empty universe. A similar result can be obtained also for dust, $w = 0$, and for every equation of state $w > -1/3$. In other words, even a small displacement from the initial flatness would lead to an immediate collapse of the Universe or to a fast cooling; finding today a value for Ω close to unity can be explained then only with values of w different from the Friedmannian ones. From (7.2.7) it is clear that, *at least for some time,* it is necessary $w < -1/3$: solutions of Einstein equations with $w < -1/3$, or in general with an accelerated behavior, are called *inflationary.* A solution with $w < -1/3$ implies $\ddot{a} > 0$.

Let us consider now in particular $w = -1$. From (7.2.7) it follows

$$C_0 = C_i (T_0/T_i)^2 = C_i 10^{-62}\,. \tag{7.2.8}$$

Now the situation is opposite. Even beginning with a very large C_i, e.g. 10^{50}, we'll have at the present $\Omega = 1 \pm 10^{-12}$, i.e. a Universe spatially flat within any possible observation. The curvature problem is then solved: in an accelerated expansion the trajectory $\Omega = 1$ is stable.

Assuming $a(t) = a_i \exp H_i t$, corresponding to $w = -1$, and defining the expansion ratio

$$\log a(t)/a_i = \Delta N\,, \tag{7.2.9}$$

also called *e-folding* time, we have $(T_e/T_i)^2 = e^{-2\Delta N}$, where any quantity with an index e is meant to be evaluated at the end of inflation. At the end of the inflationary stage , after ΔN e-foldings, if C_i was of the order of unity, we had $C_e = e^{-2\Delta N}$. If after t_e standard cosmology begins, with $w = 1/3$ and with a temperature *again* of the order of $T_{GUT} \sim 10^{15}$ GeV (we'll see later on that at the end of inflation the universe goes back to high temperatures), we can apply (7.2.7) with the initial condition at epoch t_e, and arrive therefore at a present curvature $C = e^{-2\Delta N} 10^{54} = 10^{54 - 0.87\Delta N}$. Then, to produce a present curvature in agreement with observations, i.e. near unity, inflation must have lasted for *at least* $\Delta N = 62$ e-folding. This is evidently just an estimate, but the condition $\Delta N \geq 60$ is considered the minimal condition for every acceptable model.

Let us switch now to the second problem of the standard model, the *horizon problem* . The causal horizon L_h is defined by the physical distance that a light ray (null geodesics) travels from $t = 0$ to t. Putting $ds = 0$ in a flat FRW metric we have

$$L_h = a(t) \int_0^t \frac{dt'}{a'(t)} = a(t) \int_0^a \frac{da'}{H a'^2}$$

Approximating the cosmic expansion with a single law $a(t) \sim t^n$, we have $H = H_0 a^{-1/n}$ and therefore

$$L_h = \frac{n a^{1/n}}{(1-n) H_0} = \frac{n}{1-n} H^{-1} \tag{7.2.10}$$

which, today and assuming $n = 2/3$ (dust matter) gives as already seen (eq. 3.9.13 for $z \to \infty$)

$$L_{oss} = 2/H_0 = 6000 Mpc/h\,.$$

If $t = t_0$, present epoch, L_h is also the size of the *observable universe.* In a Friedmannian universe $L_h \sim 1/H$ and therefore we can practically identify the causal horizon with H^{-1}. However this is not so in general: for instance if a expands exponentially, the causal horizon grows also exponentially while $H = const.$

Now, even if L_h is the maximal distance traveled by a photon, the relevant scale for the astrophysical processes, for instance the growth of perturbations is given simply by $R_h = H^{-1}$. This distance is called *effective horizon* or *Hubble radius*. Its present value is $R_0 = H_0^{-1} = 3000h^{-1}$Mpc. We say therefore that a perturbation of comoving scale λ_c is *inside the horizon* (or Hubble radius) at a given time t if $a\lambda_c < H^{-1}$, and *outside the horizon* if $a\lambda_c > H^{-1}$. Introducing also the comoving wavenumber $k = 2\pi/\lambda_c$, the equivalent conditions

$$\lambda_c = 1/aH, \qquad k = aH\,, \tag{7.2.11}$$

will denote the horizon crossing of a perturbation (note that these equalities are order of unity approximations) .

Which is the scale that, at the instant in which the scale factor is a, is in horizon crossing? All distances will always be evaluated at the present epoch, that is we will determine the distances reached today by the scales in horizon crossing when the scale factor was a. This amounts to assigning as comoving coordinates the present astronomical distances λ_0, that is to assume $a_0 = a(t_0) = 1$. Let us then compare the expansion law of a perturbation of physical scale $\lambda = \lambda_0(a/a_0)$ with the evolution of the Hubble radius. We have then $R_h = H^{-1} = H_0^{-1}(a/a_0)^{3(1+w)/2} = R_0(a/a_0)^{3(1+w)/2}$. Notice that $R_h = H^{-1}$ grows faster than the scale factor only if $n < 1$. The condition of horizon crossing can be found by putting $\lambda = R_h$, i.e. solving the system

$$\begin{aligned} R_h &= R_0(a/a_0)^{3(1+w)/2}\,, \\ \lambda &= \lambda_0(a/a_0)\,. \end{aligned} \tag{7.2.12}$$

We have then that the comoving scale in horizon crossing when the scale factor was a is given by $\lambda_0 = R_0(a/a_0)^{3(1+w)/2-1}$. Then for instance the scale in horizon crossing at decoupling, when $a/a_0 = 1/z_{dec}$ in a MDE expansion with $w = 0$ is

$$\lambda_{dec} = 3000 \cdot 1000^{-1/2} = 95h^{-1}\mathrm{Mpc}\,. \tag{7.2.13}$$

This scale gives then the measure of the horizon at the decoupling epoch. Analogously, at the equivalence epoch, when $z_{eq} \approx 2.3 \cdot 10^4(\Omega_m h^2)$, one has

$$\lambda_{eq} = 3000 \cdot (2.3 \cdot 10^4 \Omega_m h^2)^{-1/2} = 19h^{-1}\mathrm{Mpc}(\Omega_m h^2)^{-1/2}\,. \tag{7.2.14}$$

The value λ_{eq} is just an approximation, since in reality we should use both radiation and matter expansion laws.

Let us consider now the scale that was in horizon crossing at decoupling, i.e. roughly $\sim$100 h^{-1}Mpc. At the distance of the last scattering surface, $6000h^{-1}$Mpc, 100 Mpc are subtended by only $100/6000 \cdot 180/\pi = 0.8°$: two points A and B on the microwave background with angular distance larger than this are therefore causally disconnected. How can it happens then that the background temperature is almost the same over the whole sky? This is the *horizon problem*.

Here again, the solution is to modify the Friedmannian kinematics, i.e. w. Let us suppose that before some epoch t_e (end of inflation) the scale factor evolved according to $w_i < -1/3$, and later on with a behavior $w_f > -1/3$ (e.g. $1/3$). Let us rewrite the system (7.2.12)

$$R_h = R_0(a/a_e)^{m_i}(a_e/a_0)^{m_f}\,, \tag{7.2.15}$$

$$\lambda = \lambda_0(a/a_0) = \lambda_0(a/a_e)(a_e/a_0)\,, \tag{7.2.16}$$

where we have put

$$m \equiv 3(1+w)/2 \tag{7.2.17}$$

valid for $a < a_e$ (notice that in the first one we can write $R_0(a_e/a_0)^{m_f} = R_e$, the horizon at t_e and in the second one $\lambda_0(a_e/a_0) = \lambda_e$, i.e. the scale in *horizon-crossing* at t_e), and let us find the value of a, when the present horizon scale , R_0, was in *horizon-crossing*). This condition is sufficient to solve the horizon problem. We'll have in fact that the whole universe has been inside the Hubble radius at some time in the past, before t_e. Consequently, every smaller scale has been in horizon crossing before t_e. In other words, we have two epochs of *horizon-crossing*: in the first, the inflationary one, $t < t_e$, the effective horizon grows more slowly than the comoving expansion and the scales leave the horizon (the smaller the later): in the second epoch, the Friedmannian one, the horizon grows faster and the scale reenter the horizon: *last out, first in* (see Fig. (7.2.1)).

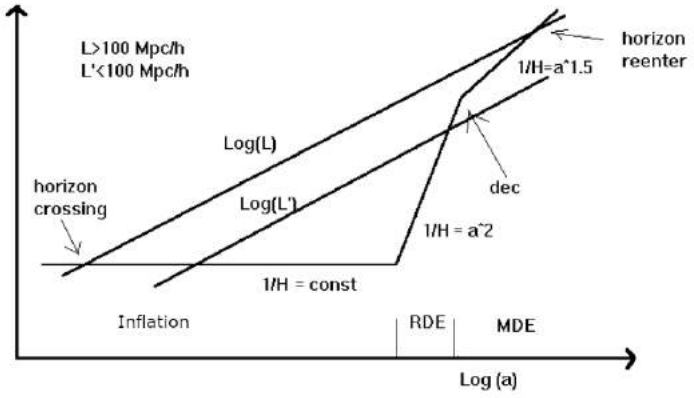


Figure 7.2.1: Horizon H^{-1} and perturbation scales versus $\log a$.

Back to (7.2.15) we have, by defining the number of$e$-foldings $N = \log(a_e/a)$ (the quantity ΔN defined above is $\Delta N = N_T - N$, where N_T is the total number of e-foldings during inflation) and putting $\lambda_0 = R_0$,

$$e^{-N+Nm_i} = (a_0/a_e)^{1-m_f} . \tag{7.2.18}$$

We have a valid solution only if $N > 0$, i.e. if $a < a_e$, because it is in this range that (7.2.15) are defined. It follows then

$$N_T = \left(\frac{m_f - 1}{m_i - 1}\right) \log(T_0/T_e) \tag{7.2.19}$$

(because $a_0/a_e = T_e/T_0$). This is the general solution, with $m_i < 1$ and $m_f > 1$: we have then the number of e-foldings needed so that the scale R_0 had been inside the horizon during inflation . $N > N_T$ is then the minimal duration of inflation. The ratio T_e/T_0 depends on the temperature needed, at the end of inflation, for the standard processes of baryogenesis, nucleosynthesis. One assumes usually $T_e \approx 10^{15}$GeV to realize baryogenesis but even putting $T_e = 10^2$GeV the result would be qualitatively similar. We assume finally for simplicity $m_f = 2$ (RDE). Finally, if $T_0 = 3K \approx 10^{-12}$GeV we obtain

$$N_T \approx 62 \left(\frac{1}{1 - m_i}\right) + \log\left(\frac{T_e}{10^{15} GeV}\right) , \tag{7.2.20}$$

which for $m_i = 0$ (i.e. $w = -1$) gives$N_T \approx 60$. Eq. (7.2.20) is then the necessary condition to produce a present universe inside the horizon before t_e and coincide with the previous condition for the curvature problem.

The same formulas give us another useful relation, between the comoving scale that exited the horizon N e-foldings before the end of inflation. From (7.2.12) it follows

$$\lambda_0 = R_0 e^{(N-N_T)(1-m_i)} \tag{7.2.21}$$

Approximating $m_i \to 0$ one gets the simple and useful relation $\lambda_0 = R_0 e^{N-N_T}$: the scale $(3000e^{-\Delta N})h^{-1}$Mpc crossed out the horizon after ΔN e-foldings from the beginning of inflation, if we state that the beginning of inflation was the epoch at which R_0 was in horizon crossing. Assuming $N_T = 60$, we can say then that at $N \approx 54$ a scale of $7h^{-1}$Mpc (comoving, i.e. evaluated at the present time) was in horizon crossing, at $N = 1$ the scale was $10^{-23}h^{-1}$Mpc ≈ 10 cm, and so on. It is clear that the true beginning of inflation could have been much more remote than $N = 60$; every scale larger than the horizon is however stretched beyond observation today. All events at $N \gg N_T$ are indeed unobservable.

From the point of view of the causal horizon things are a bit different. Evaluating the integral $L_h = a \int dt/a$ we obtain indeed, in the case of an exponential expansion, that every perturbation that is born inside the

horizon remains always inside it, while every scale larger than the horizon it will cross inside sooner or later. In other words two points in causal contact remain always in causal contact. Moreover for any expansion power law $n \geq 1$, the causal horizon strictly speaking does not exist since the integral does not converge for $t \to 0$. In any case, as already mentioned, the scales of astrophysical interest are those of the perturbations and in this case the effective horizon is the Hubble radius.

Guth's original work was also motivated by the problem of the topological defects: magnetic monopoles, cosmic strings and domain walls. These are induced by different orientation of the Higgs field in distant regions of space. The monopoles in particular are extremely energetic structures and the experimental limits on their presence are very strong. In the Friedmannian model in which in the past there were many regions causally disconnected one should find many magnetic monopoles around. The inflationary model solves this problem as well since now the whole observable universe is contained within one causal horizon.

To summarize: inflation explains the observation that $\Omega \sim 1$. Moreover it predicts exactly $\Omega = 1$ if the number of e-foldings is large than 60. In several models one has indeed $N \gg 60$, and therefore it is assumed generally that inflation is confirmed only if $\Omega = 1$. Since the luminous matter in stars and gas or dust gives at most $\Omega \sim 0.005$, while the nucleosynthesis abundance requires a baryonic density $\Omega_b \sim 0.02$, it follows that inflation requires a large contribution of dark components.

7.3 Old inflation and scalar field dynamics

We have seen how an accelerated expansion turns out to be very useful in cosmology. On the other hand, this corresponds to an equation of state $w < -1/3$. In particular, the exponential acceleration, or de Sitter model, $a = a_i \exp H_i t$ with $H_i \equiv \dot{a}/a = const$, corresponds to $w = -1$ (indeed $H = cost.$ implies $\rho = cost.$ and from (7.2.4), $\rho + p = 0$). Which form of matter has this kind of equation of state? And how is possible after 60 e-foldings to go back to the ordinary expansion with $w \geq 0$? The original answer in Guth's paper, and in most other models of inflation, has been the scalar field dynamics.

Before discussing the scalar field we first observe that de Sitter solution is a maximally symmetric solution of Einstein equations. In every space-time with maximal symmetry, one has

$$R_{\mu\nu} = \frac{1}{4} g_{\mu\nu} R \,. \tag{7.3.1}$$

Inserted in Bianchi equations this gives $R_{;\mu} = const.$ The simplest solution in a FRW space-time is de Sitter's solution

$$H = H_i = const, \to a = a_i \exp H_i t \,, \tag{7.3.2}$$

if $k = 0$, and

$$a = H_i^{-1} \sinh H_i t \,, \qquad a = H_i^{-1} \cosh H_i t \,, \tag{7.3.3}$$

for$k = -1$ and $k = 1$ respectively. Note that for $t \to \infty$ all solutions tend to the flat exponential one; this is another way to see that de Sitter flat space solution is stable. Form now on we restrict therefore to the case $k = 0$. Einstein equations with de Sitter solution give us then $\rho = const$ and $p = -\rho$ (and therefore $w = -1$, as we were expecting). A simple form of matter that has these properties is in reality already known: the cosmological constant Λ. The energy-momentum tensor of the cosmological constant (or vacuum energy) is $T_{\mu\nu}(\Lambda) = (\Lambda/8\pi G) g_{\mu\nu}$ and therefore $p_\Lambda = -(\Lambda/8\pi G)$ and $\rho_\Lambda = (\Lambda/8\pi G)$. Today the cosmological constant is roughly zero; in Planck units is actually $\rho_\Lambda < 10^{-120} M_P^4$ (obtained putting $\rho_\Lambda < \rho_c$). We should then look for something else.

Let us consider the Lagrangian of a scalar field:

$$L = \sqrt{-g} \left[-R + 16\pi [\frac{1}{2} g^{\mu\nu} \phi_{,\mu} \phi_{,\nu} - V(\phi)] \right] \,, \tag{7.3.4}$$

The first term is the Einstein-Hilbert Lagrangian (2.6.1) that we already know. The other terms generalize to a curved space-time the action of a scalar field that in Minkowsky space gives the Klein-Gordon equation if

$V = \frac{1}{2}m^2\phi^2$. The potential $V(\phi)$ should take into account possible terms of autointeraction of the scalar field and the mass term $\frac{1}{2}m^2\phi^2$. To perform the variation of L with respect to $g_{\mu\nu}$ we should remember that

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}(\delta g^{\mu\nu})g_{\mu\nu}$$

The variation with respect to $g_{\mu\nu}$ of the part that depends on ϕ gives:

$$\delta L_\phi = -\frac{1}{2}\left[g_{\mu\nu}L_\phi - \phi_{;\mu}\phi_{;\nu}\right]\sqrt{-g}\delta g^{\mu\nu}\,. \tag{7.3.5}$$

Together with the Einstein part we obtain

$$G_{\mu\nu} = 8\pi[\phi_{;\mu}\phi_{;\nu} - \frac{1}{2}\phi_{;\alpha}\phi^{;\alpha}g_{\mu\nu} + g_{\mu\nu}V] \equiv 8\pi T_{\mu\nu}\,. \tag{7.3.6}$$

The equation we get by varying the Action with respect to ϕ is the KG equation in a general space-time (signature $+---$):

$$\Box\phi + V' = 0 \tag{7.3.7}$$

where $V' \equiv dV/d\phi$ and the d'Alambertian is

$$\Box\phi \equiv \phi^{;\mu}_{;\mu} = (-g)^{-1/2}\partial_\mu(-g)^{1/2}g^{\mu\nu}\partial_\nu\phi$$

From $T_{\mu\nu}$ we obtain immediately in a FRW metric

$$T^0_0 = \rho = \frac{1}{2}\dot\phi^2 + V(\phi)\,, \qquad -T^1_1 = -T^2_2 = -T^3_3 = p = \frac{1}{2}\dot\phi^2 - V(\phi)\,. \tag{7.3.8}$$

A simple way to get the KG equation in a FRW metric is to replace these ρ, p in the conservation equation

$$\dot\rho + 3H(\rho + p) = 0$$

The equation of state is then characterized b $w(t) = p/\rho \neq const.$ What we need is a kinematic such that $w(t)$ is -1 in a first phase and ≥ 0 later on. It is clear that if $\dot\phi \sim 0$ one realizes the first condition $p = -\rho$.

The model proposed by Guth is inspired by the Higgs potential. Such a model presents two vacuum states, a stable one at $\phi = \phi_0$ and a metastable one at $\phi = 0$, and a correction $f(\phi, T)$ due to the interaction with all other field at temperature T :

$$V = -a\phi^2 + b\phi^3 + c\phi^4 + V_0 + f(\phi, T) \qquad a, b, c > 0\,. \tag{7.3.9}$$

The function $f(\phi, T)$ is such that $f(\phi, 0) = 0$ and $f(\phi, T \to \infty) \sim \phi^2T^2$. To provide for a phase transition we need that V has at $T = 0$ a minimum at u $\phi = 0$ with finite energy, $V(0) = V_{FV} > 0$ (*fals*e vacuum, FV) and another global minimum at $\phi = \phi_0$ such that $V(\phi_0) = 0$ (true vacuum, TV). The constant V_0 is chosen so to have in fact $V(\phi_0) = 0$. The energy scale was expected to correspond to the grand unification scale $V_{FV} \sim M^4_{GUT}$.

At high temperatures the potential (7.3.9) is symmetric: $V(-\phi) = V(\phi)$, since the term $f(\phi, T)$ dominates and the vacuum energy value of the field, $< \phi >$, vanishes. The system is therefore in thermal equilibrium. When T decreases the potential develops a secondary minimum at ϕ_0. For $T \sim T_c \sim T_{GUT}$ the two minima are degenerate: $V(\phi_0) = V(0)$. If in this moment there is no barrier between FV and TV, the transition can proceed with continuity. If on the contrary there is anyway a barrier between the two vacua the transition can take place via quantum tunneling or thermal fluctuations. The theory we are considering now, called *Old Inflation*, is first order. After a further cooling, the false vacuum at $\phi = 0$ becomes metastable, while $V(\phi_0)$ goes to zero and the average value of ϕ vanishes: $< \phi >= \phi_0$. This is a spontaneous breaking of symmetry. During a certain phase called *supercooling*, the field remains trapped in the FV state. Then we have now $V' = 0$, and from (7.3.7) $\phi = 0$, from which $\dot\phi = 0$ and finally $-p = \rho \sim M^4_{GUT}$. From the Friedmann Eq. (7.2.1) at $k = 0$ it follows, reminding that $G = M_P^{-2}$,

$$H \sim M^2_{GUT}/M_P \sim 10^{-8}\,, \tag{7.3.10}$$

in Planck units. Later on we will see that the energy acquired during the supercooling is delivered at the end of the inflationary era, reheating the Universe back to $T \sim T_{GUT}$ and starting the ordinary thermal processes of the standard model.

The crucial 60 e-foldings are reached after $\log(a/a_i) = Ht = 60$, i.e. at $t = 60/H \sim 10^{10} t_P \sim 10^{-33} sec$ from the beginning of the inflationary phase. Before this phase the Universe evolved probably as in a radiation epoch (so $a \sim t^{1/2}$) and to cool from $T_P = 10^{19} GeV$ down to $T_{GUT} = 10^{15} GeV$ one needs $t = (T_P/T_{GUT})^2 = 10^8$ Planck times, i.e. $10^{-35} sec$. In short, the various epochs can be summarized as follows:

- before $T = 10^{19} GeV$, $t = 10^{-43} sec$: Planck era
- down to $T = 10^{15} GeV$, $t = 10^{-35} sec$: first radiation era;
- down to $T \sim 0$, $t = 10^{-33}$ sec: inflationary era, *supercooling*;
- T rises back to $\sim 10^{15} GeV$ after bubble coalescence: reheating;
- until now: standard cosmology.

7.4 Slow rolling

The only really necessary ingredient to realize a period of inflation is that $p/\rho < -1/3$ during some epoch. This does not require necessarily a model with false and true vacua. Indeed, it can be easily realized in any scalar field theory in which the kinetic energy is much smaller than the potential energy, $\dot{\phi}^2 \ll V(\phi)$.

The first model to eliminate the phase transition has been the so-called *new inflation* (Albrecht and Steinhardt 1982, and Linde 1982): here there is actually a phase transition but it is a smooth one, without any tunneling between vacua. At any given temperature there is only one vacuum. At large T the field lies in a symmetrical vacuum. When $T < T_c$ the field begins to move towards the new vacuum at $\phi = \phi_0$, without a barrier to overcome, or perhaps with a very small one. The main point here is that the field slow rolls toward the global minimum.

In 1983 Linde proposed an even simpler model, the so-called *chaotic inflation*. Now there is no trace left of the original transition, and the range of potential one can choose is enlarged considerably.

The Klein-Gordon equation and the $(0,0)$ component of the Einstein equations are

$$H^2 = \frac{8\pi}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right] \tag{7.4.1}$$

$$\ddot{\phi} + 3H\dot{\phi} + V' = 0\,, \tag{7.4.2}$$

and this is all we need. Around its global minimum, say at $\phi = 0$, the potential $V(\phi)$ can in general be approximated by a parabola, and the motion of ϕ will appear like a ball reaching the equilibrium state at the bottom of the potential after a number of oscillations quenched by the cosmological friction $3H\dot{\phi}$. Here we search for a *slow-rolling*, solution, in which the following conditions hold:

$$\dot{\phi}^2 \ll V(\phi)\,, \quad \ddot{\phi} \ll 3H\dot{\phi}, V'\,. \tag{7.4.3}$$

If these are applied, Eqs. (7.4.2) simplify:

$$H^2 = \frac{8\pi}{3}V(\phi)\,,$$
$$3H\dot{\phi} + V' = 0\,. \tag{7.4.4}$$

We'll make often reference to (7.4.4) as *slow-rolling* equations. We will consider for simplicity the case of a free massive field, i.e. with $V(\phi) = m^2\phi^2/2$. Substituting the first of the (7.4.4) in the second one and integrating one finds immediately

$$\phi = \phi_i - \frac{m}{\sqrt{12\pi}}t = \phi_i(1 - t/\tau)\,, \tag{7.4.5}$$

where $\tau^{-1} = m/(\phi_i\sqrt{12\pi})$ is the time constant and ϕ_i is obviously the initial value of the field at $t = 0$. We can say that the epoch of *slow-rolling* ends when ϕ first crosses the potential minimum at $\phi = 0$, before the

oscillating phase begins. In this moment in fact all the potential energy is converted in kinetic energy and the conditions (7.4.3) cannot be satisfied. The inflationary phase lasts then approximately τ.

It is now again convenient to introduce the variable $\alpha = \Delta N = \log(a/a_i)$ as a time variable. From the definition $H = \dot{a}/a$ it follows in all generality $a = a_i \exp \int H dt$, from which $\alpha = \int H dt$ e $d\alpha = H dt$. Employing α in (7.4.4) we obtain

$$3H^2\phi' = -m^2\phi\,, \tag{7.4.6}$$

where $\phi' = d\phi/d\alpha$. From the integration it follows

$$(\phi^2 - \phi_i^2) = -\alpha/2\pi = -\frac{1}{2\pi}\log(a/a_i)\,, \tag{7.4.7}$$

from which finally

$$a = a_i \exp 2\pi(\phi_i^2 - \phi^2)\,. \tag{7.4.8}$$

Since at the end of inflation $\phi \to 0$ we have that the number of e -foldings has been

$$\Delta N = \log(a/a_i) = 2\pi\phi_i^2\,. \tag{7.4.9}$$

The condition $\Delta N > 60$ is then satisfied if $\phi_i \geq 3M_P$. (Notice that ϕ has the units of mass). We can imagine that the initial conditions are distributed in a chaotic way, i.e. different in each region. It has been sufficient then that in a region sufficiently large a shift $\phi_i > 3M_P$ when from the equilibrium value has been generated to produce, in that region, a sufficient inflation. If inflation started when $V(\phi_i) \approx 1$, one has $\phi_i^2 \sim m^{-2}$ for the quadratic potential, and therefore a duration of $\Delta N \sim m^{-2}$ e-foldings. Since we will see that $m \leq 10^{-6}M_P$, we have $\Delta N > 10^{10^{12}}$, an enormous number of e-foldings.

We can now easily verify that the conditions of *slow-rolling* (7.4.3) are satisfied by our solution. In particular, we have $\ddot{\phi} = 0$, while the kinetic energy condition can be written in the form $m \ll H$. This latter condition is called "flat potential": the chaotic model works if the mass (defined as $m^2 = d^2V/d\phi^2$) evaluated at the minimum, is much smaller than H. In the present case the condition of small kinetic energy implies $\phi^2 \gg 1/12\pi$. Only if ϕ is larger than this are the approximations we have employed valid; when ϕ goes below this threshold the inflation is essentially over. During the oscillation stage one has $< V > \approx < (\dot{\phi})^2/2 >$. It follows then $p \approx 0$, and therefore $a(t)$ begins a phase that is kinematically similar to the dust phase, $a \sim t^{2/3}$, and therefore decelerated. The reheating process takes place in this epoch.

It is not difficult to generalize to potentials such as $V = \lambda\phi^{2n}/2n$. One finds then

$$a = a_i \exp 2\pi(\phi_i^2 - \phi^2)/n\,, \tag{7.4.10}$$

and the condition of sufficient expansion is $\phi_i^2 > 60n/2\pi$.

Part II

The perturbed Universe

Chapter 8

Linear perturbations

Quick summary

- The evolution of linear matter perturbations can be studied in Newtonian gravity in a simplified case
- In GR, Einstein equations can be written in a perturbed metric with linear fluctuations
- The fluctuations can be divided into scalar, vector and tensor. We discuss only the scalar ones because are relevant for structure formation
- Choosing a gauge, we are left with just two independent functions of space and time, that generalize Netwon's potential
- After Fourier transformation, the equations can be solved analytically in various regimes: sub- and super-horizon, matter and radiation epoch
- Fluctuations grow when they are smaller than the horizon but larger than the Jeans length

8.1 The Newtonian equations

Most of the physics of cosmological perturbations is essentially Newtonian, at least as long as we consider only linear (i.e., small) pressureless perturbations at sub-horizon scales and we assume that the background evolution has already been obtained by solving the unperturbed Einstein equations. In this section we follow in part Peacock, *Cosmological Physics*. Let us consider a density field $\rho(x)$, associated to a pressure $p(x)$ and a velocity field $\mathbf{v}(x)$, under the action of a gravitational potential $\Phi(x)$ generated by the fluid itself (where x denotes the dependence on space-time coordinates). Then we have three relevant equations: a conservation equation (the fluid moves but neither disappears nor is created), the Euler equation (the fluid is accelerated under the action of pressure gradient and of gravity), and the Poisson equation (the gravitational potential depends on the fluid's density). The conservation equation in one dimension, for instance, says that the number of particles with velocity v entering in a small region dx at the border 1, equal to $(\rho dx)_1 = (\rho v dt)_1$, minus the number leaving the region at the border 2, $(\rho v dt)_2$, must be equal to the change of the number of particle in the same region, $d\rho dx$, so that $d\rho dx = [(\rho v)_1 - (\rho v)_2]dt$, or

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho v)}{\partial_{\lrcorner} x} \tag{8.1.1}$$

This generalizes to 3D as

$$\frac{\partial \rho}{\partial t} = -\nabla(\rho \mathbf{v}) = -\rho \nabla \cdot \mathbf{v} - \mathbf{v} \cdot \nabla \rho \tag{8.1.2}$$

where $\nabla = \{\partial_x, \partial_y, \partial_z\}$ in Cartesian coordinates. Therefore, the three corresponding equations are (as usual, $G = 1$)

$$\dot{\rho} + \mathbf{v} \cdot \nabla\rho = -\rho\nabla \cdot \mathbf{v} \qquad \text{conservation} \tag{8.1.3}$$

$$\rho(\dot{\mathbf{v}} + \mathbf{v} \cdot \nabla\mathbf{v}) = -\nabla p - \rho\nabla\Phi \qquad \text{Euler} \tag{8.1.4}$$

$$\nabla^2\Phi = 4\pi\rho \qquad \text{Poisson} \tag{8.1.5}$$

We now need to do three operations: expand the equations to first order, employ the background cosmological equations, and adopt comoving coordinates.

Let us expand the conservation equation. We put $\rho = \rho_0 + \delta\rho$, $\mathbf{v} = \mathbf{v}_0 + \delta\mathbf{v}$, where ρ_0 depends only on time (so $\nabla\rho_0 = 0$), and $\mathbf{v}_0 = H\mathbf{x}$ is the cosmological expansion. The perturbed terms $\delta\rho, \delta\mathbf{v}$ are supposed to be much smaller than the background terms. Then we have from (8.1.3)

$$\dot{\rho}_0 + \dot{\delta\rho} + (\mathbf{v}_0 + \delta\mathbf{v}) \cdot \nabla\delta\rho = -(\rho_0 + \delta\rho)\nabla \cdot (\mathbf{v}_0 + \delta\mathbf{v}) \tag{8.1.6}$$

At zero-th order this gives the expected result for a pressureless fluid

$$\dot{\rho}_0 = -\rho_0 H\nabla \cdot \mathbf{x} = -3H\rho_0 \tag{8.1.7}$$

We can use this equation now to simplify the perturbed one,

$$\dot{\delta\rho} + \mathbf{v}_0 \cdot \nabla\delta\rho = -\rho_0\nabla \cdot \delta\mathbf{v} - 3H\delta\rho \tag{8.1.8}$$

where we also discarded two higher order terms, $\delta\mathbf{v} \cdot \nabla\delta\rho$ and $\delta\rho\nabla \cdot \delta\mathbf{v}$. Next, we introduce the density contrast

$$\delta \equiv \frac{\delta\rho}{\rho_0} \tag{8.1.9}$$

and we notice that $\rho_0\dot{\delta} = \dot{\delta\rho} - \delta\rho\dot{\rho}_0/\rho_0 = \dot{\delta\rho} + 3H\delta\rho$. Then

$$\dot{\delta} + \mathbf{v}_0 \cdot \nabla\delta = -\nabla \cdot \delta\mathbf{v} \tag{8.1.10}$$

We notice that on the lhs we have a total derivative, $d\delta/dt \equiv \dot{\delta} + \mathbf{v}_0 \cdot \nabla\delta$.

The same operations can be applied to the Euler and Poisson equations, and we find

$$\frac{d\delta\mathbf{v}}{dt} = \dot{\delta\mathbf{v}} + \mathbf{v}_0 \cdot \nabla\delta\mathbf{v} = -\frac{\nabla\delta p}{\rho_0} - \nabla\delta\Phi - (\delta\mathbf{v} \cdot \nabla)\mathbf{v}_0 \tag{8.1.11}$$

$$\nabla^2\delta\Phi = 4\pi\rho_0\delta \tag{8.1.12}$$

The term $(\delta\mathbf{v} \cdot \nabla)\mathbf{v}_0$ can be written in explicit component form (sum over repeated indexes) as $(\delta\mathbf{v})_i\nabla_i v_{0j} = (\delta\mathbf{v})_i H\nabla_i x_j = (\delta\mathbf{v})_j H$. We have now a linearized set of equations, but we still want to put them in comoving coordinates $\mathbf{r}$,

$$\mathbf{x}(t) = a(t)\mathbf{r}(\mathbf{t}) \tag{8.1.13}$$

Then we have $\nabla = a^{-1}\nabla_r$. We now use simply $\mathbf{v}$ for $\delta\mathbf{v}$ and ϕ for $\delta\Phi$. It follows

$$a\frac{d\delta}{dt} = -\nabla_r \cdot \mathbf{v} \tag{8.1.14}$$

$$\frac{d\mathbf{v}}{dt} = -\frac{\nabla_r\delta p}{a\rho_0} - \frac{\nabla_r\phi}{a} - H\mathbf{v} \tag{8.1.15}$$

$$\nabla_r^2\phi = 4\pi a^2\rho_0\delta \tag{8.1.16}$$

This set of equations is closed by introducing the sound speed $c_s^2 \equiv \delta p/\delta\rho$. Now, with respect to the comoving observers, the background velocity v_0 is zero, and δv is the peculiar velocity, so the total derivative d/dt is identical to the partial derivative $\partial/\partial t$. More formally, moving from $\mathbf{x}$ to $a\mathbf{r}$, one has (let's consider only the

$x = aX$ coordinate for simplicity) $dx = [\partial(aX)/\partial t]dt + [\partial(aX)/\partial X]dX = \dot{a}Xdt + adX$, so for any function $f(t,x)$

$$df(t,x) = \left(\frac{\partial f}{\partial t}\right)_x dt + \left(\frac{\partial f}{\partial x}\right)_t dx = \left(\frac{\partial f}{\partial t}\right)_x dt + \left(\frac{\partial f}{\partial x}\right)_t (\dot{a}Xdt + adX) \tag{8.1.17}$$

$$= \left[\left(\frac{\partial f}{\partial t}\right)_x + \left(\frac{\partial f}{\partial x}\right)_t Hx\right] dt + \left(\frac{\partial f}{\partial x}\right)_t adX = \left(\frac{\partial f}{\partial t}\right)_X dt + \left(\frac{\partial f}{\partial X}\right)_t dX \tag{8.1.18}$$

(since $\dot{a}X = Hx$) from which we see that

$$\left(\frac{\partial}{\partial t}\right)_X = \left(\frac{\partial}{\partial t}\right)_x + \left(\frac{\partial}{\partial x}\right)_t Hx \tag{8.1.19}$$

and, generalizing to 3D and applying it to δ,

$$\left(\frac{\partial \delta}{\partial t}\right)_{\mathbf{r}} \equiv \dot{\delta} + \mathbf{v}_0 \cdot \nabla\delta = \left(\frac{d\delta}{dt}\right)_x \tag{8.1.20}$$

which means that, in comoving coordinates, the total derivative wrt time is actually a partial derivative. Finally, adopting the conformal time $d\tau = dt/a$ (see next section) we obtain (the dot is now $\partial/\partial\tau$)

$$\dot{\delta} = -\nabla_r \cdot \mathbf{v} \tag{8.1.21}$$

$$\dot{\mathbf{v}} = -\nabla_r c_s^2 \delta - \nabla_r \phi - Ha\mathbf{v} \tag{8.1.22}$$

$$\nabla_r^2 \phi = 4\pi a^2 \rho_0 \delta \tag{8.1.23}$$

We will show now how to recover and generalize these equations in a full GR treatment.

8.2 Introduction to the relativistic treatment

This section provides a summary of the relevant GR equations for a homogeneous universe composed of matter, radiation and a scalar field. We will sometimes convert between Kelvins and units of length and units of energy:

$$1eV = 10^4\text{K}, \quad T_0 = 2.73K = 0.23 \cdot 10^{-12}\text{GeV}, \quad 1\text{ Mpc} = 2 \cdot 10^{38}\text{GeV}^{-1} \tag{8.2.1}$$

Accordingly,

$$H_0 = (3000\text{Mpc}/h)^{-1} = 0.15 \cdot 10^{-41} h\text{GeV} \tag{8.2.2}$$

We also need to express the gravitational constant in units of energy

$$G = M_P^{-2} = 10^{-38}\text{GeV} \tag{8.2.3}$$

We take the *conformal* flat FRW metric

$$ds^2 = a^2\left[d\tau^2 - dr^2 - r^2(d\theta^2 + \sin\theta^2 d\phi)\right] \tag{8.2.4}$$

where we introduced the conformal time τ. The relation with the usual time is

$$\frac{dt}{a(t)} = d\tau \tag{8.2.5}$$

so that for instance

$$\frac{d\phi}{dt} = \frac{d\tau}{dt}\frac{d\phi}{d\tau} = \frac{d\phi}{ad\tau} \tag{8.2.6}$$

$$\frac{d^2\phi}{dt^2} = \frac{d}{ad\tau}\frac{d\phi}{ad\tau} = \frac{1}{a^2}\frac{d^2\phi}{d\tau^2} - \frac{da}{a^3 d\tau}\frac{d\phi}{d\tau} = \frac{1}{a^2}\left(\ddot{\phi} - H\dot{\phi}\right) \tag{8.2.7}$$

In this Chapter, the dot means derivative wrt conformal time τ. To avoid confusion, notice that we define the conformal Hubble function

$$\mathcal{H} = \frac{1}{a}\frac{da}{d\tau} \tag{8.2.8}$$

which corresponds to $aH(t)$ in the time t. The conformal time τ is essentially the comoving size of the horizon at time t, if we normalize the scalar factor as $a(t_{now}) = 1$:

$$\tau = \int^t \frac{dt'}{a(t')}$$

The energy-momentum tensor for a perfect fluid tensor is

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - p g_{\mu\nu} \tag{8.2.9}$$

(with this sign convention, $T^\nu_\mu = diag(\rho, -p, -p, -p)$) where the equation-of-state for each component is

$$p = w\rho \tag{8.2.10}$$

The equation of state of pressureless matter (or dust) is clearly $w = 0$, while for a relativistic component we have $p = \rho/3$, that is $w = 1/3$.

The EM tensor for a scalar field is

$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \frac{1}{2}\phi_{,\alpha}\phi^{,\alpha} g_{\mu\nu} + U(\phi) g_{\mu\nu} \tag{8.2.11}$$

Notice that $T^0_0 = \rho_\phi$, where, in a homogeneous universe, $\rho_\phi = \frac{1}{2}\frac{\dot\phi^2}{a^2} + U$, while $-T^1_1 = p_\phi = \frac{1}{2}\frac{\dot\phi^2}{a^2} - U$. The effective equation of state for a scalar field is

$$w = \frac{p}{\rho} = \frac{\frac{1}{2}\frac{\dot\phi^2}{a^2} - U}{\frac{1}{2}\frac{\dot\phi^2}{a^2} + U} \tag{8.2.12}$$

Then we see that w is not constant in general, and it depends on the specific solution. There are however three limiting behavior worth discussing even without a precise knowledge of the potential. First, consider a situation in which the field ϕ sits in a non-zero energy minimum of the potential. The field is then static, its kinetic energy vanishes and we get $w = -1$, which, as we will see below, is the same equation of state of the cosmological constant. This in fact corresponds to a state in which the field has a nonzero vacuum energy; it is this constant energy density that drives inflation. Next, consider the case in which the field oscillates harmonically around a zero-energy potential minimum. Here the kinetic energy and the potential energy are equal when averaging over several cycles. Then we get $w \simeq 0$, just as pressureless matter. Therefore the scalar field can effectively resemble as a cosmological constant or as dust, depending on its specific solution.

Here the four-velocity is

$$u^\alpha = \frac{dx^\alpha}{ds} = \{\frac{d\tau}{a d\tau}, 0\} = \{\frac{1}{a}, 0\} \tag{8.2.13}$$

assuming the fluid to be at rest on comoving coordinates (i.e., free falling). Notice that $u_\alpha u^\alpha = 1$. The GR equations with the cosmological constant are

$$T^\mu_{(m)\nu;\mu} = T^\mu_{(\gamma)\nu;\mu} = T^\mu_{(\phi)\nu;\mu} = 0 \tag{8.2.14}$$

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} + \Lambda g_{\mu\nu} \tag{8.2.15}$$

The cosmological constant can be described as an additional component with EM tensor (8.2.9) with

$$\rho_\Lambda = -p_\Lambda = \frac{\Lambda}{8\pi} \tag{8.2.16}$$

that is, with equation of state $w = -1$. In the following we always put $\Lambda = 0$. Let us notice here that the Christoffel symbols

$$\Gamma^{\gamma}_{\alpha\beta} = \frac{1}{2} g^{\gamma\eta} \left(g_{\alpha\eta,\beta} + g_{\beta\eta,\alpha} - g_{\alpha\beta,\eta} \right) \tag{8.2.17}$$

are all zero except

$$\Gamma^{\beta}_{0a} = \mathcal{H}\delta^{\beta}_{\alpha}, \quad \Gamma^{0}_{\alpha\beta} = \mathcal{H}\delta_{\alpha\beta} \tag{8.2.18}$$

The Ricci tensor is

$$R_{\alpha\beta} = \Gamma^{\gamma}_{\alpha\beta,\gamma} - \Gamma^{\gamma}_{\alpha\gamma,\beta} + \Gamma^{\sigma}_{\gamma\sigma}\Gamma^{\gamma}_{\alpha\beta} - \Gamma^{\sigma}_{\gamma\alpha}\Gamma^{\gamma}_{\sigma\beta} \tag{8.2.19}$$

and the Ricci scalar and the $(0,0)$ components are

$$R = -6a^{-2}(\dot{\mathcal{H}} + \mathcal{H}^2) \tag{8.2.20}$$

$$R^0_0 = -3\dot{\mathcal{H}} \tag{8.2.21}$$

(remember that we define $\dot{\mathcal{H}} = d(da/ad\tau)/d\tau$). The equations for two components, matter (subscript c), and radiation (subscript γ) reduce to

$$\dot{\rho}_c + 3\mathcal{H}(w_c + 1)\rho_c = 0 \tag{8.2.22}$$

$$\dot{\rho}_\gamma + 3\mathcal{H}(w_\gamma + 1)\rho_\gamma = 0 \tag{8.2.23}$$

while the (0,0) and the Trace component of the Einstein equations are

$$\mathcal{H}^2 = \frac{8\pi a^2}{3} (\rho_c + \rho_\gamma) \tag{8.2.24}$$

$$\dot{\mathcal{H}} = -\frac{4\pi}{3} a^2 (\rho_{tot} + 3p_{tot}) \tag{8.2.25}$$

In the unperturbed FRW metric with a conformal time, the D'Alambertian operator $\Box$ is (notice that $\sqrt{-g} = a^4$ in the conformal metric)

$$\begin{aligned} \Box\phi &= \frac{1}{\sqrt{-g}} \partial_\mu \sqrt{-g} g^{\mu\nu} \partial_\nu \phi \\ &= a^{-4} \partial_\mu (a^4 a^{-2} \eta^{\mu\nu} \partial_\nu \phi) \\ &= a^{-2}(-\ddot{\phi} - 2H\dot{\phi} + \phi^{;i}_{;i}) \end{aligned} \tag{8.2.26}$$

so that the Klein-Gordon equation in conformal time becomes

$$\ddot{\phi} + 2\mathcal{H}\dot{\phi} + a^2 V' = 0 \tag{8.2.27}$$

The scalar field will be considered only in Appendix C: in the following it will be neglected. It is sometimes useful to write the equation using $\alpha = \log a$ as independent variable, using the relation

$$\frac{d}{d\tau} = \mathcal{H}\frac{d}{d\alpha}$$

For instance one has that for a single fluid with equation of state w

$$\frac{\mathcal{H}'}{\mathcal{H}} = -\frac{1}{2} - \frac{3}{2} w \tag{8.2.28}$$

where the prime is with respect to $\alpha \equiv log$a . We need also the behavior of ρ as a function of the conformal time τ. The relation with the usual time is

$$\tau \sim t^{1/2} \text{in RDE}, \ \sim t^{1/3} (\text{MDE}) \tag{8.2.29}$$

so that we have, as well-known,

$$a \sim t^{1/2}, \mathcal{H} \sim a^{-2} \tag{8.2.30}$$

$$a \sim t^{2/3}, \mathcal{H} \sim a^{-3/2} \tag{8.2.31}$$

It is easy to see that in τ we have

$$a(\tau) = a_1 \quad (\tau/\tau_1), \quad \mathcal{H}(\tau) = \tau^{-1} (\text{RDE}),$$

$$a(\tau) = a_0 \quad (\tau/\tau_0)^2, \quad \mathcal{H}(\tau) = 2\tau^{-1} (\text{MDE})$$

8.3 The fluctuation equations

We discuss now analytically the perturbation equations by reducing the system to three components, CDM (that is, a pressureless component), scalar field, and radiation.

To perturb the equations one must first of all perturb the metric, writing at first order

$$g_{\mu\nu} = g^{(0)}_{\mu\nu} + g^{(1)}_{\mu\nu}$$

where the perturbed metric $g^{(1)}_{\mu\nu}$ has to be small with respect to the zero-th order part. Now, the GR equations are invariant with respect to a general coordinate change. This means that the separation between a background metric and a perturbed one is not unique: since $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ has to remain constant, changing dx^μ induces changes in the metric coefficients. However, although it is often a great simplification to choose some special coordinate frame, it would be very confusing if we change in the process also the unperturbed metric (or *background*): we would like for instance to remain in the FRW background whenever we make a general transformation. Therefore, we select a class of transformation that leaves $g^{(0)}_{\mu\nu}$ as it is, and only changes the coefficient of $g^{(1)}_{\mu\nu}$. These transformation are called *gauge transformations*.

In the unperturbed universe, we defined comoving coordinates in such a way that the matter particles expanding with the universe remain at fixed coordinates. When perturbations are added, we can either use the same coordinates, or set up a new set of coordinates that free-fall with the particles in the perturbed gravitational field, or even adopt a totally different frame not related to matter particles. That is, barring the latter case, we can choose to attach the observers to the unperturbed matter particles or to the perturbed particles. In the former case, to be called the Newtonian or longitudinal gauge, the observers will detect a velocity field of particles falling into the clumps of matter, and, in the Newtonian limit, will measure a gravitational potential. This choice is in fact the most intuitive one, and reduces easily to the Newtonian case. On the other hand, when the perturbations are larger than the horizon, to attach observers to an invisible background is not a logic choice. In the second case, called the synchronous gauge, the observers are attached to the free falling particles, so they do not see any velocity field (unless there are other non-gravitational forces, like pressure gradients), and, being always free falling, do not measure a gravitational potential. This gauge, therefore, does not have a proper Newtonian limit, but is convenient for perturbations larger than the horizon, essentially because all the observers share the same time. In the following we choose the longitudinal gauge for all our calculations. For a detailed discussion on gauge choice, see Bertschinger (1995).

8.4 The Newtonian gauge

Let us start then with the Newtonian gauge. Here we consider only a single CDM (i.e., pressureless, non-relativistic and uncoupled) component, because we are interested in deriving the Newtonian limit in the MDE era. The most general perturbed metric can be written schematically as $g_{\mu\nu} = g^{(0)}_{\mu\nu} + a^2 g^{(1)}_{\mu\nu}$, where

$$g^{(1)}_{\mu\nu} = \begin{pmatrix} 2\psi & w_i \\ w_i & 2\phi\delta_{ij} + h_{ij} \end{pmatrix} \tag{8.4.1}$$

where ψ, ϕ are space scalars, w_i is a 3-vector and h_{ij} is a traceless 3-tensor. For instance, it is easy to show that g_{00} is a spatial scalar: the general transformation law for a tensor is

$$g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta} \tag{8.4.2}$$

If we perform a purely spatial transformation, $x'^0 = x^0, x'^i = f(x^i)$ then we have immediately that $g'_{00} = g_{00}$, as requested for a spatial scalar.

If we write the perturbed metric as $g_{\alpha\beta} = g^{(0)}_{\alpha\beta} + h_{\alpha\beta}$ we see that the condition that $g_{\alpha\gamma}g^{\gamma\beta} = \delta^\beta_\alpha$ imposes at first order

$$h^{\mu\nu} = -h_{\alpha\beta}g^{(0)\alpha\mu}g^{(0)\beta\nu} \tag{8.4.3}$$

That is, the inverse of the perturbed metric is minus the perturbed metric with indices raised by the unperturbed metric.

A decomposition analogous to $g_{\alpha\beta}$ can be done for the energy momentum tensor. Now, let us decompose the vector w_i into a longitudinal and a transverse component

$$w = w^{\parallel} + w^{\perp} \tag{8.4.4}$$

where by construction

$$\nabla \cdot w^{\perp} = \nabla \times w^{\parallel} = 0 \tag{8.4.5}$$

The longitudinal component, $w^{\parallel}$, being curl-free, is the gradient of a scalar, $w^{\parallel} = \nabla w_s$. When we derive the Einstein equations for the $0i$ components, we will have therefore terms longitudinal and transverse, both in G_{0i} and in T_{0i} . Taking the curl of the equations, we will be left with only the transverse equations; taking the divergence, we will be left with the longitudinal ones. Therefore, the two components completely decouple from each other, and can be treated separately. Just the same situation occurs for the (ij) components, except that now we will have also the pure tensorial equations (which turn out to be the gravitational wave equation). Since the density perturbation δ is a scalar quantity, only the longitudinal terms, which can be derived from a scalar quantity, couple to the density perturbations. The terms which are intrinsically vectorial couple to pure rotational modes, while tensorial terms represent gravitational waves, coupled to matter only for anisotropic perturbations. Furthermore, it can be shown that if initially the rotational, or vorticity, modes, are zero, they remain zero throughout (unless there are entropy gradients). If they are present initially, they decay as a^{-1}.

A similar argument holds for the traceless spatial part h_{ij} . This tensor can be written in general as a sum of three traceless terms:

$$h_{ij} = h^{\parallel}_{ij} + h^{\perp}_{ij} + h^{T}_{ij} \tag{8.4.6}$$

where the divergences $\partial^i h^{\parallel}_{ij}, \partial^i h^{\perp}_{ij}$ (which are vectors) are longitudinal and transverse, respectively, and h^T_{ij} is transverse, that is

$$\epsilon_{ijk}\partial_i\partial_k h^{\parallel}_{\ell j} = 0, \quad \partial_i\partial_j h^{\perp}_{ij} = 0, \quad \partial_i h^{T}_{ij} = 0 \tag{8.4.7}$$

(that is, the curl of $\partial_i h^{\parallel}_{ij}$ and the div of $\partial_j h^{\perp}_{ij}$ vanish, and the div of h^T_{ij}. also vanishes). Here, ϵ_{ijk} is the totally antisymmetric tensor, which vanishes whenever two indexes are equal, and is $(-1)^p$ for the other cases, where p is the order of the permutations of i, j, k. Now, since $\partial_i h^{\parallel}_{ij}$ is curl-free, it can be written in terms of a scalar function B, and it is easy to verify that $\epsilon_{ijk}\partial_i\partial_k h^{\parallel}_{ij} = 0$ if

$$h^{\parallel}_{ij} = \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right) B \tag{8.4.8}$$

On the contrary, the perturbations $h^{\perp}_{ij}, h^T_{ij}$ cannot be derived from a scalar function. The first one is a vector, and the second one is a tensor: they give rise to rotational velocity perturbations and gravitational waves, and decouple completely from the scalar terms. They are therefore to be treated separately, as for the Newtonian gauge.

Therefore, we need to take into account only the part of w_i and h_{ij} which can be derived from scalars. This may be done introducing two new scalar functions, E and B, that produce the vector $E_{,i}$ and the tensor $D_{ij}B \equiv B_{,ij} - B^{,k}_{,k}\delta_{ij}/3$ (this combination is chosen to give a traceless tensor), in analogy to the electromagnetic forces

$$g^{(1)}_{\mu\nu} = \begin{pmatrix} 2\psi & E_{,i} \\ E_{,i} & 2\phi\delta_{ij} + D_{ij}B \end{pmatrix} \tag{8.4.9}$$

Out of the four scalar functions ψ, ϕ, E, B, one can construct gauge-invariant quantities, that is, combinations that remain invariant under a general coordinate transformation $x^{\mu\prime} = x^{\mu} + \xi^{\mu}$. The gauge-invariant formalism, however, is algebraically very complicate. Things can be much simplified if one works in a specific gauge. This can be done imposing up to four conditions on the metric, corresponding to the four coordinate transformations: here we choose them to be $w_i = 0$ (from which $E = 0$) and $B = 0$. This leaves the perturbed metric in the Newtonian or longitudinal gauge:

$$ds^2 = a^2\left[(1+2\psi)d\tau^2 - (1-2\phi)dx^i dx_i\right] \tag{8.4.10}$$

Here the perturbed four-velocity (from now on all expressions are first order in the perturbations) is

$$\begin{aligned} u^\alpha &= \frac{dx^\alpha}{ds} = \{\frac{d\tau}{a(1+\psi)d\tau}, \frac{dx^i}{ad\tau}\} = \{\frac{1}{a}\left(1-\psi\right), \frac{v^i}{a}\} \\ u_\alpha &= g_{\alpha\beta}u^\beta = \{a\left(1+\psi\right), -av_i\} \\ u_\alpha u^\alpha &= 1 \end{aligned} \tag{8.4.11}$$

where $v^i = dx^i/d\tau = adx^i/dt$ is the matter peculiar velocity (physical distance divided time interval) with respect to the general expansion. The notation for the perturbed quantities is

$$\delta = \delta\rho/\rho, \qquad \nabla^i v_i = \theta. \tag{8.4.12}$$

where $\delta\rho/\rho \equiv (\rho(x) - \hat{\rho})/\hat{\rho}$ if $\rho(x)$ is the density field and $\hat{\rho}$ the average .Also, θ is the velocity divergence. In general, there will be several pairs δ, θ, one for each fluid composing the universe. We start now with the single fluid model. We have to perturb the following equations:

$$T^\mu_{\nu;\mu} = 0 \tag{8.4.13}$$

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} \tag{8.4.14}$$

The perturbed energy-momentum tensor for a component of equation-of-state $p = w\rho$ with constant w is

$$\delta T^0_0 = \delta\rho \tag{8.4.15}$$

$$\delta T^1_1 = \delta T^2_2 = \delta T^3_3 = -c_s^2\delta\rho \tag{8.4.16}$$

$$\delta T^i_0 = (1+w)\rho v^i \tag{8.4.17}$$

(notice that $\delta T^i_0 = -\delta T^0_i$). Here we introduced the sound velocity, $c_s^2 \equiv \frac{\delta p}{\delta\rho}$. Let us recall the operation of covariant divergence of a tensor T^μ_ν :

$$T^\mu_{\nu;\mu} = T^\mu_{\nu,\mu} - \Gamma^\alpha_{\nu\beta}T^\beta_\alpha + \Gamma^\alpha_{\beta\alpha}T^\beta_\nu \tag{8.4.18}$$

The perturbed Christoffel symbols are

$$\begin{aligned} \delta\Gamma^0_{ij} &= -\delta_{ij}\left[2\mathcal{H}\left(\phi+\psi\right) + \dot{\phi}\right] \\ \delta\Gamma^0_{00} &= \dot{\psi} \\ \delta\Gamma^0_{0i} &= \delta\Gamma^i_{00} = \psi_{,i} \\ \delta\Gamma^j_{i0} &= -\delta^j_i\dot{\phi} \qquad (8.4.19) \\ \delta\Gamma^i_{jl} &= -\delta^i_j\partial_l\phi - \delta^i_l\partial_j\phi + \delta_{jl}\partial_i\phi \qquad (8.4.20) \end{aligned}$$

all the other being zero. Let us then calculate explicitly the first equation, $\delta T^\mu_{0;\mu} = 0$ for $w = 0$ and $c_s^2 = 0$. Notice that we have here $T^0_0 = \rho$ and $T^j_i = 0$. We have then

$$\delta T^\mu_{0,\mu} - \delta\Gamma^\alpha_{0\beta}T^\beta_\alpha - \Gamma^\alpha_{0\beta}\delta T^\beta_\alpha + \delta\Gamma^\alpha_{0\alpha}T^0_0 + \Gamma^\alpha_{\beta\alpha}\delta T^\beta_0 = 0 \tag{8.4.21}$$

where

$$\begin{aligned} \delta T^\mu_{0,\mu} &= (\dot{\delta\rho}) + \rho v^i_{,i}, \quad \delta\Gamma^\alpha_{0\beta}T^\beta_\alpha = \rho\delta\Gamma^0_{00} = \rho\dot{\psi} \\ \Gamma^\alpha_{0\beta}\delta T^\beta_\alpha &= \mathcal{H}\delta\rho \\ \delta\Gamma^\alpha_{0\alpha}T^0_0 &= \rho(\delta\Gamma^0_{00} + \delta\Gamma^i_{i0}) = \rho(\dot{\psi} - 3\dot{\phi}) \\ \Gamma^\alpha_{\beta\alpha}\delta T^\beta_0 &= 4\mathcal{H}\delta\rho \end{aligned} \tag{8.4.22}$$

and finally

$$(\dot{\delta\rho}) + \rho v^i_{,i} - \rho\dot{\psi} - \mathcal{H}\delta\rho + \rho(\dot{\psi} - 3\dot{\phi}) + 4\mathcal{H}\delta\rho = 0 \rightarrow$$

$$(\dot{\delta\rho}) + \rho\theta - 3\rho\dot{\phi} + 3\mathcal{H}\delta\rho = 0 \rightarrow \tag{8.4.23}$$

$$\frac{\dot{\delta\rho}}{\rho} + 3\mathcal{H}\frac{\delta\rho}{\rho} = -\theta + 3\dot{\phi} \tag{8.4.24}$$

Finally, using the unperturbed conservation equation, $\dot{\rho} + 3\mathcal{H}\rho = 0$, and defining the density contrast $\delta = \delta\rho/\rho$, we derive the relation $(\dot{\delta\rho}) = \rho\dot{\delta} + \delta\dot{\rho}$ and we obtain the perturbed energy conservation equation

$$\dot{\delta} = -\theta + 3\dot{\phi} \tag{8.4.25}$$

This is called *continuity equation*.

The equation for $\nu = i$ is (here we put back c_s^2)

$$\dot{v}_i = -\mathcal{H}v_i - \nabla_i\psi - \nabla_i c_s^2\delta \tag{8.4.26}$$

We can take the divergence ∇^i of this equation and obtain

$$\dot{\theta} = -\mathcal{H}\theta - \nabla^2\psi - \nabla^2 c_s^2\delta \tag{8.4.27}$$

where we define $\nabla^2 \equiv \sum_i \nabla_i^2$ just as in Minkowski space , i.e. without the use of the spatial metric g_{ij}.

Eq. (8.4.27) is the *Euler equation*. Notice that in Minkowski space ($a = const$), the left-hand-side of Eq. (8.4.26) reduces to $\dot{v}_i = \frac{d^2x_i}{dt^2}$: this shows explicitly that this equation is the force equation for a fluid: the acceleration equals the sum of gradient of the potential and the pressure gradient (since $c_s^2\delta = \frac{dp}{d\rho}\frac{\delta\rho}{\rho} = \frac{\delta p}{\rho}$).

We go now to *Fourier space*. This means that all perturbation quantities will be Fourier expanded:

$$\phi = \int e^{ikr}\phi_k d^3k, \quad \psi = \int e^{ikr}\psi_k d^3k \tag{8.4.28}$$

$$\delta = \int e^{ikr}\delta_k d^3k, \quad \theta = \int e^{ikr}\theta_k d^3k \tag{8.4.29}$$

but the subscript k will be dropped in the following (the Fourier normalization factors here play no role, since the equations are all in Fourier space; when needed, they will be put back, adopting the convention of section 8.7). Here and in what follows I mean $kr \equiv \mathbf{k}\cdot\mathbf{r}$. In other words, we assume that the perturbation variables $\delta, \theta, \psi, \phi$ etc. are the sum of plane waves $\delta_k e^{ikr}$; since the equations are linear, each plane wave obey the same equations with a different k. In practice, each perturbation quantity and its derivatives can be substituted as follows

$$\begin{aligned} \phi(x,\tau) &\rightarrow e^{ikr}\phi_k(\tau) \\ \nabla\phi(x,\tau) &\rightarrow ie^{ikr}\mathbf{k}\phi_k(\tau) \\ \nabla^2\phi(x,\tau) \equiv \nabla_i\nabla_i\phi(x,\tau) &\rightarrow -e^{ikr}k^2\phi_k(\tau) \end{aligned} \tag{8.4.30}$$

Furthermore, the Fourier modes e^{ikr} can be simply dropped out, since the equations are linear and therefore decoupled in the modes. Eq. (8.4.27) becomes then

$$\dot{\theta} = -\mathcal{H}\theta + k^2\psi + k^2 c_s^2\delta \tag{8.4.31}$$

In Fourier space, from the Einstein equations we obtain finally

$$k^2\phi + 3\mathcal{H}(\dot{\phi} + \mathcal{H}\psi) = -4\pi a^2\rho\delta \tag{8.4.32}$$

$$k^2(\dot{\phi} + \mathcal{H}\psi) = 4\pi a^2\theta\rho \tag{8.4.33}$$

$$\phi = \psi \tag{8.4.34}$$

Notice that the last equation holds only if there is no anisotropic stress $\Sigma_i^j \equiv \delta T_i^j$ in the energy-momentum tensor, that is, only for perfect fluid. For a generic component of *constant* parameter of state w the equations are modified as follows

$$\begin{aligned} \dot{\delta} &= -(w+1)(\theta - 3\dot{\phi}) + 3\mathcal{H}(w - c_s^2)\delta \\ \dot{\theta} &= \mathcal{H}\theta(3w-1) + k^2\left(\frac{c_s^2}{w+1}\delta + \psi\right) \end{aligned} \tag{8.4.35}$$

$$\begin{aligned} k^2(\dot{\phi} + \mathcal{H}\psi) &= 4\pi(w+1)a^2\theta\rho \\ k^2\phi + 3\mathcal{H}(\dot{\phi} + \mathcal{H}\psi) &= -4\pi a^2\rho\delta \end{aligned} \tag{8.4.36}$$

This is our general set of perturbation equations. In the next sections we will solve it in several regimes: during the epochs of inflation, radiation and matter, and both at large and small scales.

Finally, one can write the last two equations in a quasi-Poissonian form

$$k^2\phi = -4\pi a^2\rho[\delta + 3\mathcal{H}(w+1)\theta/k^2] = -4\pi a^2\rho\delta^* \tag{8.4.37}$$

by defining the *total matter variable*

$$\delta^* \equiv \delta + 3\mathcal{H}(w+1)\theta/k^2$$

.

8.5 Scales larger than the horizon

In the limit $k \ll \mathcal{H}$, i.e. for scales larger than the horizon, the system (8.4.35) when $w = c_s^2 = \text{const}$ (which is valid both for matter and for radiation) becomes

$$\begin{aligned} \dot{\delta} &= -(w+1)(\theta - 3\dot{\psi}) \\ \dot{\theta} &= \mathcal{H}\theta(3w-1) \\ 3\mathcal{H}(\dot{\psi} + \mathcal{H}\psi) &= -4\pi a^2\rho\delta \end{aligned} \tag{8.5.1}$$

It is easy to show that $\dot{\psi} = 0$ is a solution. In fact, assuming $\dot{\psi} = 0$, deriving the first equation and inserting the second, we obtain

$$\ddot{\delta} = \mathcal{H}\dot{\delta}(3w-1)$$

which is solved by $\delta = \text{const}$ (plus a decaying solution that can be neglected). In turn, for the third equation we get $3\mathcal{H}^2\psi = -4\pi a^2\rho\delta$ or, by using the Friedmann equation,

$$\delta = -2\psi \approx \text{const} \tag{8.5.2}$$

Therefore $\delta = const$ implies $\psi = const$, as assumed. It turns out that this is the dominating solution: then, we have shown that the gravitational potential remains constant for scales outside the horizon.

8.6 Newtonian limit & the Jeans length

Let us now go back to the general perturbation equations and derive the sub-horizon (or Newtonian) limit. This is found for small scales

$$k \gg \mathcal{H} \tag{8.6.1}$$

The fluctuation of a pressureless fluid grows indefinitely because there is no counteracting force. In general, however, the pressure of the fluid will resist gravity, and stop the collapse. It is then instructive to derive the equations for a fluid which is pressureless when unperturbed (i.e. $w = 0$), but has a finite sound velocity or velocity dispersion (much smaller than the velocity of light, otherwise the Newtonian approximation will break)

$$c_s^2 = \frac{\delta p}{\delta\rho} \ll 1 \tag{8.6.2}$$

For large k, in Eq. (8.4.32) we can then neglect $3\mathcal{H}(\dot{\phi} + \mathcal{H}\psi)$ with respect to $k^2\phi$. Then Eq. (8.4.32) is the Fourier transformed Poisson equation

$$k^2\phi = -4\pi a^2\rho\delta = -\frac{3}{2}\mathcal{H}^2\delta \tag{8.6.3}$$

where we used the unperturbed Friedmann equation. This shows that δ is of the same order as $k^2\phi$, so in Eq. (8.4.25) we can neglect the $\dot{\phi}$ term and find

$$\dot{\delta} \simeq -\theta \tag{8.6.4}$$

This equation reduces then to the energy conservation equation in the Newtonian limit. (Note that terms like $\dot{\phi}$ can be written as $\mathcal{H}d\phi/d\log a = -(3/2)k^{-2}\mathcal{H}d(\mathcal{H}^2\delta)/d\log a$, which is indeed much smaller than $\dot{\delta} \sim \mathcal{H}d\delta/d\log a$ when $k/\mathcal{H}$ is large, unless $\mathcal{H}, \delta$ vary with a in an exceedingly fast way: this approximation is called *quasi-static limit*).

The perturbation equations in the Newtonian limit become then

$$\begin{aligned} \dot{\delta} &= -\theta \\ \dot{\theta} &= -\mathcal{H}\theta + c_s^2 k^2 \delta + k^2\phi \\ k^2\phi &= -\frac{3}{2}\mathcal{H}^2\delta \end{aligned} \tag{8.6.5}$$

i.e. the continuity, the Euler and the Poisson equations, which are as expected the Fourier transform of the equations (8.1.23) already seen in Sect. (8.1). Deriving the first equation we obtain

$$\ddot{\delta} + \mathcal{H}\dot{\delta} + \left(k^2 c_s^2 - \frac{3}{2}\mathcal{H}^2\right)\delta = 0 \tag{8.6.6}$$

(In the Minkowski limit, $H = 0$, this equation reduces to the fluid wave equation $\ddot{\delta} - c_s^2\nabla^2\delta = 0$, and shows that c_s is indeed the sound velocity.) This shows at once that the perturbation does not grow if

$$k^2 c_s^2 - \frac{3}{2}\mathcal{H}^2 > 0 \tag{8.6.7}$$

i.e., if the perturbation scale $\lambda = 2\pi a/k$ is smaller than the Jeans length,

$$\lambda_J = c_s\sqrt{\frac{\pi}{\rho}} \tag{8.6.8}$$

For scales smaller than λ_J the perturbations undergo damped oscillations. For the CDM particles the velocity dispersion is always negligible. For the photons is $c/\sqrt{3}$, so that the physical scale

$$\lambda_J \simeq H^{-1} \tag{8.6.9}$$

and growth is prevented on all scales smaller than the horizon. For the baryons, finally, the sound velocity is comparable to the photon velocity before decoupling, so that baryon perturbations are damped out, but drops rapidly to a comoving scale of less than 1Mpc just after decoupling. Then the baryons are free to fall inside the dark matter potential wells, and their perturbation spectrum catches the dark matter one.

8.7 Perturbation evolution

For $kc_s \ll \mathcal{H}$ the perturbations grow freely: this is the phenomenon of gravitational instability. We can rewrite the Newtonian equation as

$$\ddot{\delta} + \mathcal{H}\dot{\delta} - \frac{3}{2}\mathcal{H}^2\delta = 0 \tag{8.7.1}$$

Adopting now the time variable $\alpha = \log a$ we obtain

$$\delta'' + (\frac{\mathcal{H}'}{\mathcal{H}} + 1)\delta' - \frac{3}{2}\delta = 0 \tag{8.7.2}$$

Putting (show this as an exercise)

$$\frac{\mathcal{H}'}{\mathcal{H}} = -\frac{1}{2} - \frac{3}{2}w \tag{8.7.3}$$

we obtain for $w = 0$

$$\delta'' + \frac{1}{2}\delta' - \frac{3}{2}\delta = 0 \tag{8.7.4}$$

which is simply solved by

$$\delta = Ae^{m\alpha} = Aa^m$$

with the two solutions

$$m_\pm = 1, -3/2 \tag{8.7.5}$$

Therefore the growth functions are

$$\delta_+ = Aa^1, \quad \delta_- = Ba^{-3/2}$$

(*growing* and *decaying* modes, respectively). The second solution becomes rapidly negligible with respect to the first one and normally is neglected. With respect to conformal time we have $\delta_+ \sim \tau^2$. Obviously the constants A, B must be fixed by the observations.

From δ_+ an important consequence follows. From the Poisson equation

$$k^2\phi = -\frac{3}{2}\mathcal{H}^2\delta$$

we see that during the matter dominated era $\mathcal{H}^2\delta_+ = a^2H^2\delta_+ = a^2a^{-3}a = cost.$ and therefore the gravitational potential remains constant.

8.8 Two-fluids solution

The solution just found holds true for matter fluctuations during MDE. To generalize we need to add radiation. Let us restart from (8.4.35) and introduce the matter perturbations δ_m, θ_m and the radiation variables $\delta_\gamma, \theta_\gamma$. We'll have then two conservation equation pairs plus the two-fluid Einstein equations

$$\begin{aligned}
\dot{\delta}_m &= -(\theta_m - 3\dot{\phi}) \\
\dot{\theta}_m &= -\mathcal{H}\theta_m + k^2\psi \\
\dot{\delta}_\gamma &= -\frac{4}{3}(\theta_\gamma - 3\dot{\phi}) \\
\dot{\theta}_\gamma &= k^2(\frac{3c_s^2}{4}\delta_\gamma + \psi) \\
k^2(\dot{\phi} + \mathcal{H}\psi) &= 4\pi(w+1)a^2\theta_t\rho_t \\
k^2\phi + 3\mathcal{H}(\dot{\phi} + \mathcal{H}\psi) &= -4\pi a^2\rho_t\delta_t
\end{aligned} \tag{8.8.1}$$

where t is for the total quantities and

$$\begin{aligned}
\rho_t &= \rho_m + \rho_\gamma \\
\rho_t\delta_t &= (\delta\rho_t) = \rho_m\delta_m + \rho_\gamma\delta_\gamma
\end{aligned}$$

In the sub-horizon limit we have then (note that $w = c_s^2 = 1/3$ for radiation)

$$\begin{aligned}
\dot{\delta}_m &= -\theta_m \\
\dot{\theta}_m &= -\mathcal{H}\theta_m + k^2\psi \\
\dot{\delta}_\gamma &= -\frac{4}{3}\theta_\gamma \\
\dot{\theta}_\gamma &= \frac{1}{4}k^2\delta_\gamma \\
k^2\phi &= -4\pi a^2(\rho_m\delta_m + \rho_\gamma\delta_\gamma) = -\frac{3}{2}\mathcal{H}^2(\Omega_m\delta_m + \Omega_\gamma\delta_\gamma)
\end{aligned} \tag{8.8.2}$$

(note that here Ω_i denotes the fractional density at any time, not today's value). As before, by differentiating the equations for $\dot{\delta}$ we obtain the two coupled equations

$$\begin{aligned}
\ddot{\delta}_m + \mathcal{H}\dot{\delta}_m - \frac{3}{2}\mathcal{H}^2(\Omega_m\delta_m + \Omega_\gamma\delta_\gamma) &= 0 \\
\ddot{\delta}_\gamma + \frac{k^2}{3}\delta_\gamma &= 0
\end{aligned} \tag{8.8.3}$$

In RDE $\Omega_m \approx 0$ and $\Omega_\gamma \approx 1$; moreover, the second equation shows that radiation oscillates rapidly around zero (since $k > \mathcal{H}$). As long as the baryons are electromagnetically coupled to the photons we can approximate the photon-baryon fluid as a single fluid end therefore δ_γ represents the perturbations of the coupled plasma. The oscillations of this fluid produce the acoustic effects on the microwave background (see Chap. 11).

Averaging over the oscillations we can put $< \delta_\gamma > \approx 0$. It follows $\Omega_m \delta_m + \Omega_\gamma \delta_\gamma \approx 0$ and finally

$$\ddot{\delta}_m + \mathcal{H}\dot{\delta}_m = 0$$

which is solved by $\delta_m = \text{const}$ and $\delta_m \sim \tau^{-1}$. That is, matter perturbations do not grow when they are sub-horizon during RDE.

We have now the perturbation evolution in both RDE and MDE and both for sub- and super-horizon perturbations. More accurate treatment require numerical integrations of the whole set of coupled equations.

We should now establish the initial conditions, based on the inflationary model. When the inflaton field decades in particles through reheating we can assume that the numerical ratio among the various species is the same everywhere. Then the ratio photon/(matter particle) is constant, i.e. $\delta(n_\gamma/n_m) = 0$. Then, since $n_\gamma \sim \rho_\gamma^{3/4}$ while $n_m \sim \rho_m$, we have

$$\delta \log\left(\frac{n_\gamma}{n_m}\right) = \frac{\delta n_\gamma}{n_\gamma} - \frac{\delta n_m}{n_m} = \frac{3}{4}\delta_\gamma - \delta_m = 0$$

from which

$$\delta_\gamma = \frac{4}{3}\delta_m. \tag{8.8.4}$$

This is often called *adiabatic condition* since it implies that the entropy-per-particle is constant (the entropy density is proportional to the photon number).

8.9 Growth rate and growth function in ΛCDM.

Let us observe now that the equation

$$\ddot{\delta}_m + \mathcal{H}\dot{\delta}_m - \frac{3}{2}\mathcal{H}^2(\Omega_m \delta_m + \Omega_\gamma \delta_\gamma) = 0$$

can be immediately generalized. If in fact we assume the two fluids are matter and a cosmological constant (instead of radiation), we have obviously

$$\ddot{\delta}_m + \mathcal{H}\dot{\delta}_m - \frac{3}{2}\mathcal{H}^2(\Omega_m \delta_m + \Omega_\Lambda \delta_\Lambda) = 0$$

Since ρ_Λ is constant, $\delta_\Lambda = 0$ and we obtain

$$\ddot{\delta}_m + \mathcal{H}\dot{\delta}_m - \frac{3}{2}\mathcal{H}^2\Omega_m \delta_m = 0$$

Assuming $\Omega_m = cost.$ (a rough approximation indeed) and using $\alpha = \log a$ we immediately find the law $\delta_m \sim a^m$ with the generalized growth exponents

$$m_\pm = \frac{1}{4}\left(-1 \pm \sqrt{1 + 24\Omega_m}\right). \tag{8.9.1}$$

Then we see that the cosmological constant slows down the perturbation growth: when $\Omega_m \to 0$, $m \to 0$. A better approximation for a time-varying Ω_m is however given in terms of the *growth rate* f, defined as

$$f \equiv \frac{d\log\delta_m}{d\log a} \approx \Omega_m^\gamma(z) \tag{8.9.2}$$

with $\gamma \approx 0.55$ (called the *growth index*) and for ΛCDM

$$\Omega_m(z) = \frac{\rho_m}{\rho_{crit}} = \frac{\Omega_{m0}a^{-3}}{\Omega_{m0}a^{-3} + 1 - \Omega_m} \tag{8.9.3}$$

The *growth function* (normalized to unity today) is given by

$$G(z) \equiv \frac{\delta_m(z)}{\delta_m(0)} = \exp \int_1^a f(z) d\log a \approx \exp \int_1^a \Omega_m^\gamma(z) d\log a \tag{8.9.4}$$

In this case is no longer true that $\mathcal{H}^2\delta$ is constant and therefore the gravitational potential is not a constant.

Chapter 9

Correlation function and power spectrum

Quick summary

- Here we define several measures of clustering of a distribution of points
- This chapter deal only with the mathematical properties of these statistical descriptors. In the next one we study the physical properties.
- The correlation function describes the clustering of a distribution of points in space.
- The angular correlation function describes the correlation as a function of angular separation only
- The power spectrum is the Fouriwr conjugate of the correlation function
- Correlation function and power spectrum are two-point descriptors. One can generalize them to n-point descriptors.
- Moments are an integral measure of clustering. The second order moment can be estimated by integrating the correlation function or the power spectrum.

9.1 Why we need correlation functions, power spectra and all that

All the perturbation variables we have studied so far, δ, Ψ, θ etc, and their Fourier transforms, are random variables. That is, we cannot know if in a given point in space-time, δ is zero, -0.01 or any other value. All we have found with GR equations is how this value will evolve in time, e.g. as $\delta \sim a$ as in MDE. If in a location δ is initially zero, it will remain so. If it is initialy negative (an underdensity) it will become more and more underdense, until the region is empty. If it is an overdensity, it will grow until it exits the regime of linearity. The perturbation variables are random fields in space and as such must be studied statistically. In practice, this means that instead of studying the field $\delta(x, y, z)$, we study its moments, in particular mean and variance. In real space, the variance $\langle \delta(x)\delta(x+r) \rangle$ as a function of separation r is called correlation function. In Fourier space, the variance $\langle \delta^2 \rangle$ as a function of wavenumber k is called power spectrum. However, the mean is trivial: $\langle \delta \rangle = 0$, since the density contrast is exactly defined as the fractional difference of the density minus its mean. Since all other perturbation variables are proportional to δ, the same applies to all of them: $\langle \Psi \rangle = \langle \Phi \rangle = \langle \theta \rangle = 0$ etc.. It is then clear that only the variance contains useful information. In principle, also higher order moments can be employed, and often they are. However, a fundamental assumption (based on the initial conditions set up by inflation) is that the initial fluctuations are Gaussian distributed, and for such a distribution only the variance is needed, since all higher order moments are either zero (the odd ones) or are function of the variance: for instance $\langle \delta^4 \rangle = 3\langle \delta^2 \rangle^2$. This is why most cosmology needs only to focus on the quadratic statistics, namely variance, correlation function, and power spectrum. However, when fluctuations grow and start becoming non-linear, their Gaussian nature will in general be lost. This is obvious if only one realizes that if in a given region δ is initially negative and keeps growing, it will at some point become a void, $\delta = -1$, and will stop there; but if it is positive, δ can grow without limit (that is, the region collapses into a structure, e.g. a galaxy, a star or

even a black hole), and the final distribution of underdensities and overdensities will obviously be asymmetric (skewed) and no longer Gaussian. The exact non-linear distribution is very difficult to obtain and generally has to be found numerically by N-body simulations.

9.2 Average, variance, moments.

Here we recall a few basic facts about random variables and statistics. If x is a random variable (e.g. δ at a point r in our case), and if $f(x)dx$ is the probability of finding x in the interval $x, x+dx$, then (integral extended in the entire domain of definition of x)

$$\int f(x)dx = 1 \tag{9.2.1}$$

$$\langle x \rangle = \int x f(x)dx \tag{9.2.2}$$

$$\langle x^2 \rangle = \int x^2 f(x)dx \tag{9.2.3}$$

and in general $\langle g(x) \rangle = \int g(x)f(x)dx$. It follows that, if a, b are constants

$$\langle a \rangle = a \tag{9.2.4}$$

$$\langle ax \rangle = a\langle x \rangle \tag{9.2.5}$$

$$\langle ag(x) + bp(x) \rangle = a\langle g(x) \rangle + b\langle p(x) \rangle \tag{9.2.6}$$

We define n-th order moments and n-th order central moments as, respectively,

$$\bar{M}_n = \int x^n f(x)dx \tag{9.2.7}$$

$$M_n = \int (x - \langle x \rangle)^n f(x)dx \tag{9.2.8}$$

The central moment M_2 is called *variance*. If x_i is a vector of several random variables, one defines multivariate moments, in particular the covariance

$$c_{ij} = \langle x_i x_i \rangle \tag{9.2.9}$$

9.3 Definition of the correlation function

Other common statistical descriptors are the $n-$point correlation functions. Let $\langle n \rangle = \rho_0 dV$ be the average number of particles in an infinitesimal volume dV, being ρ_0 the average number density. If $dN_{ab} = \langle n_a n_b \rangle$ is the average number of *pairs* in the volumes dV_a and dV_b (i.e., the product of the number of particles in one volume with the number in the other volume), separated by r_{ab}, then the 2-point correlation function $\xi(r_{ab})$ is defined as

$$dN_{ab} = \langle n_a n_b \rangle = \rho_0^2 dV_a dV_b (1 + \xi(r_{ab})) \tag{9.3.1}$$

If the distribution is uniform, then the average number of pairs is exactly equal to the product of the average number of particles in the two volumes, and the correlation ξ vanishes; if there is correlation among the volumes, on the other hand, then the correlation is different from zero. The correlation function is also defined, equivalently, as *the spatial average of the product of the density contrast* $\delta(r_a) = n_a/(\rho_0 dV) - 1$ at two different points

$$\xi(r_{ab}) = \frac{dN_{ab}}{\rho_0^2 dV_a dV_b} - 1 = \frac{\langle n_a n_b \rangle}{\rho_0^2 dV_a dV_b} - 1 = \langle (\delta_a + 1)(\delta_b + 1) \rangle - 1 = \langle \delta(r_a)\delta(r_b) \rangle \tag{9.3.2}$$

because $\langle \delta_{a,b} \rangle = 0$.

In practice it is easier to derive the correlation function as the average density of particles at a distance r from another particle. This is a *conditional* density, that is the density of particles at distance r given that there is a particle at $r = 0$. The number of pairs is then the number of particles in both volumes divided by the number of particles $dN_a = ndV_a$ in the volume dV_a at $r = 0$:

$$dN_b = dN_{ab}/dN_a = \rho_0^2 dV_a dV_b(1 + \xi(r_{ab}))/dN_a = \rho_0 dV_b(1 + \xi(r_b)) \tag{9.3.3}$$

The correlation function can then be defined as

$$\xi(r) = \frac{dN_c(r)}{\rho_0 dV} - 1 = \frac{\langle \rho_c \rangle}{\rho_0} - 1 \tag{9.3.4}$$

(where c stands for conditional) i.e. as the average number of particles at distance r from any given particle (or number of neighbors), divided by the expected number of particles at the same distance in a uniform distribution, minus 1, or *conditional density contrast.* If the correlation is positive, there are then more particles than in a uniform distribution: the distribution is then said to be positively clustered. This definition is purely radial, and does not distinguish between isotropic and anisotropic distributions. One could generalize this definition by introducing the anisotropic correlation function as the number of pairs in volumes at distance r and a given longitude and latitude. This is useful whenever there is some reason to suspect that the distribution is indeed anisotropic, as when there is a significant distortion along the line-of-sight due to the redshift.

If the average density of particles is estimated from the sample itself, i.e. $\rho_0 = N/V$, it is clear that the integral of $dN_c(r)$ must converge to the number of particles in the sample :

$$\int_0^R dN_c(r) = \int \rho(r) dV = N \tag{9.3.5}$$

In this case the correlation function is a sample quantity, and it is subject to the integral constraint (Peebles 1980)

$$\int_0^R \xi_s(r) dV = N/\rho_0 - V = 0 \tag{9.3.6}$$

Assuming spatial isotropy this is

$$4\pi \int_0^R \xi_s(r) r^2 dr = 0 \tag{9.3.7}$$

If the sample density is different from the true density of the whole distribution, we must expect that the $\xi_s(r)$ estimated in the sample differs from the true correlation function. From Eq. (9.3.4), we see that $g(r) = 1 + \xi(r)$ scales as ρ_0^{-1}. Only if we can identify the sample density ρ_0 with the true density the estimate of $\xi(r)$ is correct. In general, the density is estimated in a survey centered on ourselves, so that what we obtain is in reality a conditional density.

The conditional density at distance r from a particle, averaged over the particles in the survey, is often denoted in the statistical literature as $\Gamma(r)$; we have therefore from Eq. (9.3.4)

$$\Gamma(r) \equiv \langle \rho_c \rangle = \rho_0(1 + \xi) \tag{9.3.8}$$

The average in spherical cells of radius R and volume V of this quantity is denoted as

$$\Gamma^*(R) \equiv \langle \rho_c \rangle_{sph} = \rho_0(1 + \widehat{\xi}) \tag{9.3.9}$$

where

$$\widehat{\xi} = V^{-1} \int \xi dV \tag{9.3.10}$$

To evaluate $\Gamma^*(R)$ one finds the average of the number of neighbors inside a distance R from any particle contained in the sample.

9.4 Measuring the correlation function in real catalog

Consider now the estimator (9.3.4). It requires the estimation of the density ρ_c inside a shell of thickness dr at distance r from every particle. In other words, it requires the estimation of the volume of every shell. In practice, a direct estimation of the shell volume is difficult because of the complicate boundary that a real survey often has. Moreover, if we are working on a magnitude-limited sample, the expected density ρ_0 must take into account the selection function. The simplest way to measure ξ is to compare the real catalog to a MonteCarlo catalog with exactly the same number of particles, the same boundaries and the same selection function. Then, the estimator can be written as

$$\xi = \frac{DD}{DR} - 1 \tag{9.4.1}$$

where DD means we center on a real galaxy (data D), count the number DD of galaxies at distance r, and divide by the number of galaxies DR at the same distance but in the MonteCarlo catalog (label R). In other words, instead of calculating the volume of the shell, which is a difficult task in realistic cases, we estimate it by counting the galaxies in the Poissonian MonteCarlo realization. In this way, all possible boundaries and selection function can be easily mimicked in the Poisson catalog, and will affect DD and DR in the same way (statistically). To reduce the effect of the Poisson noise in the MC catalog, we can in fact use a higher number of artificial particles, say α times the real data, and then multiply DD/DR by α.

9.5 Correlation function: examples

Let us estimate now the CF of a *planar distribution*. Consider two large spherical volumes of radius R_s. Let us distribute in one N particles uniformly on a plane passing through the center, and in the other the same N particles but now uniformly in the all volume. This latter is our MonteCarlo artificial catalog. We have to estimate DD in a spherical shell at distance r from in the planar distribution, and DR in the artificial one. In the planar world, the spherical shell cuts a circular ring of radius $r \ll R_s$ and thickness dr, so we have, on average

$$DD = \text{superf. density} \times 2\pi r dr = \frac{N}{\pi R_s^2} 2\pi r dr \tag{9.5.1}$$

In the uniform world we have

$$DR = \text{density} \times 4\pi r^2 dr = \frac{3N}{4\pi R_s^3} 4\pi r^2 dr \tag{9.5.2}$$

Then we get

$$\xi = \frac{2R_s}{3} r^{-1} - 1 \tag{9.5.3}$$

This is the CF of a planar distribution. As we can see, $1+\xi$ goes as r^{-1}, and its amplitude depends on the size of the "universe" R_s. It is clear that, in this case, the amplitude of the correlation function is not a measure of the amount of inhomogeneity of the content, but rather a measure of the geometry of the container.

Notice that the constraint 9.3.7 is satisfied:

$$\int_0^{R_s} \xi r^2 dr = \frac{2R_s}{3} \int^{R_s} r dr - \int^{R_s} r^2 dr = \frac{R_s^3}{3} - \frac{R_s^3}{3} = 0$$

Consider now another example, the *cluster model*. There are m clumps of N particles each uniformly distributed inside cubes of side R_c, and the balls are distributed uniformly in the universe. The total number of particles is mN . The total volume is mD^3, if $D \gg R_c$ is the mean intercluster distance. For $r \ll R_c$, each particle sees around it a uniform distribution with density $\rho_c = N/R_c^3$, while the global mean density is $\rho = mN/(mD^3) = N/D^3$. It follows

$$\xi(r \ll R_c) = \frac{N}{R_c^3} \frac{D^3}{N} - 1 = \left(\frac{D}{R_c}\right)^3 - 1 \tag{9.5.4}$$

On the other hand, for $r \gg D$, the distribution of particles is essentially random, and

$$\xi(r \gg D) = 0 \tag{9.5.5}$$

There are therefore three regimes: at very small scales, the CF is constant and positive; at large scales, the model is homogeneous, and at intermediate scales it decreases from one plateau to the other. Notice however that, in order to verify the integral constraint, the CF must become negative at some intermediate scale. This corresponds to the fact that outside the clusters there are less particles that in a uniform distribution. Notice also that now the CF amplitude does not depend on the universe volume, but only on the fixed parameters D and R_c.

9.6 The angular correlation function

Because the angular position of the galaxies is so much easier to determine than their distance, the angular correlation function has been often employed in astronomy. Here we write down the relation between the two correlations, that is the Limber equation, in order to show some properties.

Let $\Phi(r; m_{\rm lim})$ denote the radial selection function,

$$\Phi(r; m_{\rm lim}) = \int_{-\infty}^{M(r,m_{\rm lim})} \phi(M) dM \tag{9.6.1}$$

that is the density of galaxies at distance r in a magnitude-limited, such that $\int \Phi dV = N$ is the total number of sources selected. Since the density at distance r is $\Phi(r)$, instead of the constant ρ_0, the number of pairs in volumes dV_1, dV_2 at positions $\mathbf{r}_1, \mathbf{r}_2$ is now modified as follows

$$dN_{12} = dV_1 dV_2 (1 + \xi(r_{12}))\Phi(r_1)\Phi(r_2) \tag{9.6.2}$$

where

$$r_{12} = |\mathbf{r}_1 - \mathbf{r}_2| = \left(r_1^2 + r_2^2 - 2 r_1 r_2 \cos\theta\right)^{1/2}$$

Now, the number of pairs dN_θ which appear to be separated by an angle θ in the sky is clearly the integral of dN_{12} over all positions $\mathbf{r}_1, \mathbf{r}_2$ provided that their angular separation θ is constant. Then we have

$$dN_\theta = \int dN_{12} = \int dV_1 dV_2 (1 + \xi(r_{12}))\Phi(r_1)\Phi(r_2) \tag{9.6.3}$$

The angular correlation function is defined, in analogy to the spatial correlation

$$w(\theta) = \frac{dN_\theta}{\rho_s^2 dA_1 dA_2} - 1 \tag{9.6.4}$$

where ρ_s is the surface density, and $\rho_s dA_1 = \left(\int_{V_1} \Phi dV\right)$ is the expected number of particles in the area dA that subtends the volume V_1 (e.g., dA_1 is a circular patch of angular radius α and V_1 is the line-of-sight cone of beam size α). Then we obtain the relation between spatial and angular correlation functions:

$$w(\theta) = \frac{\int dV_1 dV_2 \xi(r_{12})\Phi(r_1)\Phi(r_2)}{\left(\int \Phi dV\right)^2} \tag{9.6.5}$$

In the limit of small separations, this equation can be simplified. If $\xi(r_{12})$ declines rapidly for large separations, we might assume that the integral is important only if $r_1 \simeq r_2 \simeq r$; if we also take a small θ we have

$$r_{12}^2 = (r_1 - r_2)^2 + r^2\theta^2 = u^2 + r^2\theta^2 \tag{9.6.6}$$

where $u = r_1 - r_2$. Passing from r_1, r_2 to u, r in the integral, and integrating out the angular variables, we get the Limber equation

$$w(\theta) = \frac{\int_0^\infty r^4 \Phi(r)^2 dr \int_{-\infty}^\infty du \xi(x)}{\left(\int r^2 \Phi dr\right)^2} \tag{9.6.7}$$

where $x^2 = u^2 + r^2\theta^2$. A simple use of this equation is when a power law approximation holds, $\xi = Ar^{-\gamma}$. Then we can define a variable z such that $u = \theta r z$, and we obtain

$$\begin{aligned} w(\theta) &= \frac{\int_0^\infty r^4\Phi(r)^2 dr \int_{-\infty}^\infty \theta r dz \left[(\theta r z)^2 + (r\theta)^2\right]^{-\gamma/2}}{\left(\int r^2\Phi dr\right)^2} \\ &= \frac{\int_0^\infty r^4\Phi(r)^2 dr \int_{-\infty}^\infty \theta^{1-\gamma} r^{1-\gamma} dz \left[z^2+1\right]^{-\gamma/2}}{\left(\int r^2\Phi dr\right)^2} = B\theta^{1-\gamma} \end{aligned} \tag{9.6.8}$$

where the coefficient B is just some number that depends on $m_{\rm lim}$ and γ

$$B = \frac{\int_0^\infty r^{5-\gamma}\Phi(r)^2 dr \int_{-\infty}^\infty dz \left[z^2+1\right]^{-\gamma/2}}{\left(\int r^2\Phi dr\right)^2} \tag{9.6.9}$$

Eq. (9.6.8) reveals that, in the limit of small angular scales and negligible correlations at large distances, the angular power law is $1-\gamma$. This is in fact roughly confirmed in several angular catalogues, although in a limited range of angular scales.

9.7 The n-point correlation function and the scaling hierarchy

The correlation function can be generalized to more than two points. The 3-point function for instance is defined as

$$\varsigma(r_a, r_b, r_c) = \langle \delta(r_a)\delta(r_b)\delta(r_c)\rangle \tag{9.7.1}$$

In terms of the counts in infinitesimal cells we can write

$$\begin{aligned} \varsigma(r_a, r_b, r_c) &= \langle \left(\frac{n_a}{\rho_0 dV_a} - 1\right)\left(\frac{n_b}{\rho_0 dV_b} - 1\right)\left(\frac{n_c}{\rho_0 dV_c} - 1\right)\rangle \\ &= \frac{\langle n_a n_b n_c\rangle}{\rho_0^3 dV_a dV_b dV_c} - \xi_{ab} - \xi_{bc} - \xi_{ac} - 1 \end{aligned} \tag{9.7.2}$$

so that we obtain the useful relation

$$\langle n_a n_b n_c\rangle = \rho_0^3 dV_a dV_b dV_c (1 + \xi_{ab} + \xi_{bc} + \xi_{ac} + \varsigma_{abc}) \tag{9.7.3}$$

In some simple and interesting cases, the moments M_p of the counts obey the following relation *for any box size* in a certain range of scales

$$S_p = \frac{M_p}{M_2^{p-1}} = const$$

Theoretical motivations for this *scaling* relation include the BBGKY equations in the strongly non-linear gravitational regime (Peebles 1980) and the second-order perturbative expansion of the gravitational evolution of the fluctuations (Fry 1984; Juszkiewicz, Bouchet & Colombi 1993). Moreover, the scaling relation is expected for a generic random variable, as the counts in cells, that are a linear combination of n independent random variables (the $k-$modes of the linear density field), expanding the PDF in powers of $n^{-1/2}$.

The scaling relation has been observed up to several tens of megaparsecs in many surveys (e.g. Gaztanaga 1994). Since the moments are the volume integrals of the correlation function (see e.g. Eq. (9.9.7) below) we expect that a similar scaling relation holds for the correlation functions themselves. The n-point correlation function is then a linear combination of products of $(n-1)$ two-points correlation functions. For instance, we can assume that

$$\varsigma_{ijk} = Q[\xi_{ij}\xi_{jk} + \xi_{ij}\xi_{ik} + \xi_{ik}\xi_{jk}] \tag{9.7.4}$$

where Q is independent of the spatial coordinates. We will often make use of such scaling relations.

9.8 The power spectrum

One of the most employed statistical estimator for density fields is the power spectrum. In recent years it has been used to quantify the clustering properties in many galaxy surveys. The main reason is that almost all theories of structure formation predict a specific shape of the spectrum, because the plane waves evolve independently in the linear approximation of the gravitational equations.

Unless otherwise specified, the conventions for the 3D Fourier transforms are

$$\begin{aligned} f(x) &= \frac{V}{(2\pi)^3}\int f_k e^{ikx} d^3k \\ f_k &= \frac{1}{V}\int f(x) e^{-ikx} d^3x \end{aligned} \tag{9.8.1}$$

and it is always understood that $ikx = i\mathbf{k}\cdot\mathbf{x}$ (we'll use the bold face for vectors when there is risk of confusion). With this conventions, $f(x)$ and f_k have the same dimensions. Since $f(x)$ is real, $f^*_{\mathbf{k}} = f_{-\mathbf{k}}$.

If we consider the Fourier transform of $f(x) \equiv 1$ we have

$$1 = \frac{V}{(2\pi)^3}\int f_k e^{ikx} d^3k \tag{9.8.2}$$

with

$$f_k = \frac{1}{V}\int e^{-ikx} d^3x \tag{9.8.3}$$

This shows that in the limit $V \to \infty$ we can take as a representation of the Dirac delta the function

$$\delta_D(k) = \frac{1}{(2\pi)^3}\int e^{-ikx} d^3x \tag{9.8.4}$$

(and similarly for $k \leftrightarrow x$) since indeed the oscillations cancel each other in an infinite volume and therefore $\delta_D(k) = 0$ for every $k \neq 0$, while $\delta_D(k)$ diverges for $k = 0$ and

$$\int \delta_D(k) d^3k = \tfrac{V}{(2\pi)^3}\int f_k d^3k = f(x=0) = 1 \tag{9.8.5}$$

Let now $\delta(x)$ be the density contrast of a density field in a survey of volume V and

$$\delta_k = \frac{1}{V}\int \delta(x) e^{-ikx} dV \tag{9.8.6}$$

its Fourier transform. The power spectrum is defined as

$$P(k) = V\delta_k\delta_k^* \tag{9.8.7}$$

Notice that the power spectrum has the dimension of a volume. It follows

$$P(k) = \frac{1}{V}\int \delta(x)\delta(y) e^{-ik(x-y)} dV_x dV_y \tag{9.8.8}$$

Now, putting $r = x - y$, and taking the *volume average*

$$\xi(r) = \langle \delta(y+r)\delta(y)\rangle_V = \frac{1}{V}\int \delta(y+r)\delta(y) dV_y \tag{9.8.9}$$

then,

$$P(\mathbf{k}) = \int \xi(r) e^{-i\mathbf{kr}} dV \tag{9.8.10}$$

Therefore, the power spectrum is the Fourier transform of the correlation function (Wiener-Khintchin theorem). The converse property is

$$\xi(\mathbf{r}) = (2\pi)^{-3}\int P(k) e^{i\mathbf{kr}} d^3k \tag{9.8.11}$$

(notice that here, following most literature, the Fourier volume factor is not included). Finally, assuming spatial isotropy, i.e. that the correlation function depends only on the modulus $|r|$, we obtain

$$P(k) = 4\pi \int \xi(r) \frac{\sin kr}{kr} r^2 dr \tag{9.8.12}$$

A more general definition of power spectrum can also be given, but this time we have to think in terms of *ensemble averages,* rather than volume averages. Consider in fact the ensemble average of $V\delta_k\delta^*_{k'}$:

$$V\langle \delta_k \delta^*_{k'} \rangle = \frac{1}{V} \int \langle \delta(y+r)\delta(y) \rangle e^{i(k-k')y + ikr} dV_r dV_y \tag{9.8.13}$$

Performing ensemble averages, one has to think of fixing a positions and making the average over the ensemble of realizations. Then the average can enter the integration, and average only over the random variables δ. Then we obtain

$$V\langle \delta_k \delta^*_{k'} \rangle = \frac{1}{V} \int \xi(r) e^{i(k-k')y + ikr} dV_r dV_y = \frac{(2\pi)^3}{V} P(k) \delta_D(k-k') \tag{9.8.14}$$

The definition (9.8.14) states simply that modes at different wavelengths are uncorrelated if the field is statistically homogeneous (that is, if ξ does not depend on the position in which is calculated but only on the distance r). This will often be useful later.

These definitions refer to infinite samples and to a continuous field. In reality, we always have a finite sample and a discrete realization of the field, i.e.. a finite number of particles. Therefore, we have to take into account the effects of both finiteness and discreteness.

To investigate the discreteness, we assume as field a collection of N particles of dimensionless masses m_i expressed in units of the average mass m_0 at positions x_i, in a volume V . In the following we will make use of the window function $W(x)$, a function which expresses the way in which the particles are selected. A typical selection procedure is to take all particles within a given region, and no particles elsewhere. In this case, the function will be a constant inside the survey, and zero outside. We will always consider such a kind of window function in the following, and normalize it so that

$$\int W(x) dV = 1 \tag{9.8.15}$$

With this normalization, $W(x) = 1/V$ inside the survey. The density contrast field we have in a specific sample is therefore the universal field times the window function (times the sample volume V because of the way we normalized W)

$$\delta_s = \delta(x) V W(x) \tag{9.8.16}$$

Let us now express the field as a sum of Dirac functions

$$\delta(x) = \left(\frac{\rho(x)}{\rho_0} - 1 \right) VW(x) = \frac{V}{N} \sum_i m_i w_i \delta_i(x - x_i) - VW(x) \tag{9.8.17}$$

where $w_i = VW(x_i)$. The Fourier transform is

$$\delta_k = \frac{1}{V} \int \left(\frac{V}{N} \sum_i m_i w_i \delta_i - VW(x) \right) e^{ikx} dV = \frac{1}{N} \sum_i m_i w_i e^{ikx_i} - W_k \tag{9.8.18}$$

where we introduced the $k-$space window function

$$W_k = \int W(x) e^{ikx} dV \tag{9.8.19}$$

normalized so that $W_0 = 1$. The most commonly used window function is the so-called top-hat function, which is the FT of the simple selection rule

$$\begin{aligned} W(x) &= 1/V \text{ inside a spherical volume } V \text{ of radius } R \\ W(x) &= 0 \text{ outside} \end{aligned} \tag{9.8.20}$$

We have then

$$\begin{aligned} W_k &= \int W(x)e^{i\mathbf{kx}}dV = V^{-1}\int e^{i\mathbf{kx}}dV \\ &= \frac{3}{4\pi}R^{-3}\int^R r^2dr\int_{-\pi}^{\pi} e^{ikr\cos\theta}d\cos\theta d\phi \\ &= \frac{3}{2}R^{-3}\int^R \left(e^{ikr}-e^{-ikr}\right)\frac{r^2}{ikr}dr \\ &= 3R^{-3}\int^R \frac{r\sin kr}{k}dr = 3\frac{\sin kR - kR\cos kR}{(kR)^3} \end{aligned} \tag{9.8.21}$$

Notice that $W_0 = 1$, and that the WF declines rapidly as $k \to \pi/R$. Now, the expected value of the power spectrum is

$$P(k) = V\langle\delta_k^2\rangle \tag{9.8.22}$$

that is

$$P(k) = \frac{V}{N^2}\sum_{ij} m_i m_j w_i w_j e^{ik(x_i-x_j)} - VW_k^2 \tag{9.8.23}$$

We used the relation

$$\langle\frac{1}{N}\sum_i m_i w_i e^{ikx_i}\rangle = \frac{1}{N}\sum_i m_i\int W(x)e^{ikx}dV = W_k \tag{9.8.24}$$

Finally, if the positions x_i and x_j are uncorrelated, we can pick up only the terms with $i = j$, so that, neglecting the window function, which is important only for $k \to 0$, we obtain the pure noise spectrum

$$P_n(k) = \frac{V}{N^2}\sum_i m_i^2 w_i^2 = V/N \tag{9.8.25}$$

where the last equality holds only if $m_i = 1$ for all particles and w_i equals 0 or 1. The noise spectrum is negligible only for large densities, $\rho_0 = N/V \to \infty$. Since the galaxy distributions are often sparse, the noise is not always negligible and has to be subtracted from the estimate. For the power spectrum applies the same consideration expressed for the moments: the power spectrum does not characterize completely a distribution, unless we know the distribution has some specific property, e.g. is Gaussian, or Poisson, etc.

9.9 From the power spectrum to the moments

The power spectrum is often the basic outcome of the structure formation theories, and it is convenient to express all the other quantities in terms of it. Here we find the relation between the power spectrum and the moments of the counts in random cells.

Consider a finite cell. Divide it into infinitesimal cells with counts n_i either zero or unity. We have by definition of ξ

$$\langle n_i n_j\rangle = \rho_0^2 dV_i dV_j[1+\xi_{ij}] \tag{9.9.1}$$

The count in the cell is $N = \sum n_i$. The variance is then

$$M_2 = \langle\frac{(N-N_0)^2}{N_0^2}\rangle = \frac{\langle N^2 + N_0^2 - 2NN_0\rangle}{N_0^2} = \frac{\langle N^2\rangle + N_0^2 - 2\langle N\rangle N_0}{N_0^2} = \frac{\langle N^2\rangle - N_0^2}{N_0^2} \tag{9.9.2}$$

where

$$\begin{aligned} \langle N^2\rangle &= \langle\sum n_i\sum n_j\rangle = \sum\langle n_i^2\rangle + \sum\langle n_i n_j\rangle = \\ &\quad N_0 + N_0^2\int dV_i dV_j W_i W_j[1+\xi_{ij}] \end{aligned} \tag{9.9.3}$$

where $N_0 = \rho_0 V$ is the count average, and $\xi_{ij} \equiv \xi(|\mathbf{r}_i - \mathbf{r}_j|)$. Let us simplify the notation by putting

$$W_i dV_i = dV_i^*$$

We define the integral (by definition $\int W dV = \int dV^* = 1$ for any window function)

$$\sigma^2 = \int dV_1^* dV_2^* \xi_{12} \tag{9.9.4}$$

Inserting the power spectrum we have

$$\sigma^2 = (2\pi)^{-3} \int P(k) e^{i\mathbf{k}(\mathbf{r}_1 - \mathbf{r}_2)} W_1 W_2 d^3k d^3r_1 d^3r_2 \tag{9.9.5}$$

This becomes, for *spherical cells*,

$$\sigma^2 = (2\pi^2)^{-1} \int P(k) W^2(k) k^2 dk \tag{9.9.6}$$

so $\langle N^2 \rangle = N_0 + N_0^2 \sigma^2$. Finally we obtain the relation between the power spectrum (or the correlation function) and the second-order moment of the counts:

$$M_2 = N_0^{-2}(\langle N_i^2 \rangle - N_0^2) = N_0^{-1} + \sigma^2 \tag{9.9.7}$$

The first term is the so called shot-noise, the second term is the count variance in the continuous limit.

For the third order moment we proceed in a similar fashion:

$$\langle N^3 \rangle = \langle \sum n_i \sum n_j \sum n_k \rangle = \sum \langle n_i^2 \rangle + 3 \sum \langle n_i^2 \rangle \sum n_i + \sum \langle n_i n_j n_k \rangle = \tag{9.9.8}$$

$$N_0 + 3N_0^2 + N_0^3 \int dV_i^* dV_j^* dV_k^* [1 + \xi_{ij} + \xi_{ik} + \xi_{jk} + \varsigma_{ijk}] \tag{9.9.9}$$

where in the last equality we used the definition of the three point correlation given in Eq. (9.7.3)

$$\langle n_i n_j n_k \rangle = \rho_0^3 dV_i dV_j dV_k [1 + \xi_{ij} + \xi_{ik} + \xi_{jk} + \varsigma_{ijk}] \tag{9.9.10}$$

The third order moment is then

$$M_3 = N_0^{-3} \langle (\Delta N)^3 \rangle = N_0^{-2} + \int dV_i^* dV_j^* dV_k^* \varsigma_{ijk} \tag{9.9.11}$$

where $\Delta N = N - N_0$. If we can assume the scaling relation $\varsigma_{ijk} = Q[\xi_{ij}\xi_{jk} + \xi_{ij}\xi_{ik} + \xi_{ik}\xi_{jk}]$ then we can express M_3 in terms of $P(k)$ and of the new parameter Q. In the limit of large N_0, a Gaussian field ($M_3 = 0$) has $Q = 0$.

Appendix: Bias in a Gaussian field

Three ways to normalize the *mass* power spectrum have been employed so far: via the velocity field, via the cosmic microwave fluctuations, and via the cluster abundance method. The third method makes use of the fact that the spectrum at a given wavenumber k gives the variance of the fluctuations at a scale $\lambda = 2\pi/k$. If this variance is high at a given epoch, many fluctuations will have a density contrast higher than unity, so that they will decouple from the Hubble expansion, and begin to recollapse and to form structures. Then, the amplitude of the power spectrum can be directly related to the abundance of structure at any given epoch. In particular, a high normalization means a large number of clusters. It appears from N-body simulations and from theoretical expectations (Press-Schechter model) that the abundance of clusters depends strongly on the spectrum normalization. Using the different power spectra of CDM with various values of Ω_0, the following relation has been found (White, Efstathiou and Frenk 1993)

$$\sigma_8 \simeq 0.5 \Omega_0^{-0.5} \tag{9.9.12}$$

with a negligible dependence on the cosmological constant. The normalization via the velocity field seems to provide a similar behavior. The power spectrum of the galaxies, however, has in general a different normalization, because the galaxies might be a biased tracer of mass. Let us now find an approximate model of the biasing factor.

Consider a Gaussian density field with correlation $\xi(r)$. By definition, we have that the fluctuation density contrast field $\delta = \delta\rho/\rho$ obey the rules

$$\langle \delta_1 \delta_2 \rangle = \xi(r) \tag{9.9.13}$$

$$\langle \delta_1^2 \rangle = \xi(0) = \sigma^2 \tag{9.9.14}$$

The density δ at each point is distributed then as

$$P(\delta) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{\delta^2}{2\sigma^2}\right]$$

where by definition $\sigma^2 = \int \delta_1^2 P(\delta) d\delta$. The probability that the fluctuation field is above a certain threshold $\nu\sigma$, where σ is the field variance, is

$$P_1 = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{\nu\sigma} \exp\left[-\frac{\delta^2}{2\sigma^2}\right] d\delta \tag{9.9.15}$$

Now, the joint probability that the density at one point is δ_1 and the density at another is δ_2,

$$P(\delta_1, \delta_2) = \left[(2\pi)^2 \det M\right]^{-1/2} \exp\left[-\frac{1}{2}\delta^i \delta^j M_{ij}\right] \tag{9.9.16}$$

where $\delta^i = \{\delta_1, \delta_2\}$ and where r is the distance between the two points. The covariance matrix is

$$M^{-1} = \begin{matrix} \sigma^2 & \xi(r) \\ \xi(r) & \sigma^2 \end{matrix} \tag{9.9.17}$$

We can write then the probability that the x field is above the threshold ν at both location as

$$\begin{aligned} P_2 &= \left[(2\pi)^2 \det M\right]^{-1/2} \int_{\nu\sigma}\int_{\nu\sigma} \exp\left[-\frac{1}{2}\delta^i \delta^j M_{ij}\right] d\delta_i d\delta_2 \\ &= \left[(2\pi)^2 (\sigma^4 - \xi^2)\right]^{-1/2} \int_{\nu\sigma}\int_{\nu\sigma} \exp\left[-\frac{\sigma^2\delta_1^2 + \sigma^2\delta_2^2 - 2\xi\delta_1\delta_2}{2\left[\sigma^4 - \xi^2\right]}\right] d\delta_1 d\delta_2 \end{aligned} \tag{9.9.18}$$

Now, suppose there are N particles in the field; the number of particles in regions above threshold is $N_1 = P_1 N$, while the number of pairs in regions above threshold is $N_2 = P_2 N^2$. The correlation function of the regions above threshold is, by definition of correlation function

$$1 + \xi_\nu = \frac{N_2}{N_1^2} = \frac{P_2}{P_1^2} \tag{9.9.19}$$

The integral can be done easily numerically, but an interesting approximation is to take the limit for $\xi(r) \ll 1$ and $\nu \gg 1$, i.e. at large scales and for high peaks. Using the large ν approximation (Abramovitz-Stegun 7.1.23)

$$P_1 = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{\nu\sigma} \exp\left[-\frac{\delta^2}{2\sigma^2}\right] d\delta \simeq \frac{1}{(2\pi\nu^2)^{1/2}} e^{-\nu^2/2} \tag{9.9.20}$$

and expanding

$$\begin{aligned} \exp\left[-\frac{\sigma^2\delta_1^2 + \sigma^2\delta_2^2 - 2\xi\delta_1\delta_2}{2\left[\sigma^4 - \xi^2\right]}\right] &\simeq \exp\left[-\frac{\delta_1^2 + \delta_2^2}{2\sigma^2}\right] \exp\left[\frac{\xi\delta_1\delta_2}{\sigma^4}\right] \\ &\simeq \exp\left[-\frac{\delta_1^2 + \delta_2^2}{2\sigma^2}\right] \left(1 + \frac{\xi\delta_1\delta_2}{\sigma^4}\right) \end{aligned} \tag{9.9.21}$$

we get

$$\frac{P_2}{P_1^2} \simeq 1 + \frac{\nu^2}{\sigma^2}\xi e^{-\nu^2} \int_{\nu\sigma} e^{-\frac{\delta_1^2+\delta_2^2}{2\sigma^2}} \delta_1\delta_2 \frac{d\delta_1 d\delta_2}{\sigma^4}$$

and finally (Kaiser 1984)

$$\xi_\nu \simeq \frac{\nu^2}{\sigma^2}\xi \tag{9.9.22}$$

This shows that peaks are more correlated than the background density field. This is the same effects one observes on mountain ranges: near a peak there is very likely another peak. Eq. (9.9.22) gives some credibility to the approximation usually made that the galaxy density field is a scale-independent-biased version of the mass density field, but it should be noticed that this is expected only for large ν and at large scales, and that the whole mechanism relies on the assumption that there is only one object per threshold region.

Eq. (9.9.22) can be applied to clusters. Suppose first we smooth the field on scales of, say, 5 Mpc$/h$, so that the variance σ on such scales is of order unity. It is found observationally that the cluster correlation function is roughly ten times larger than the galaxy correlation. This would imply a $\nu \simeq 3$, which is not unreasonable. Notice that some level of biasing is necessary: collapsed object form only where $\delta > 1$.

This concludes the study of the principal estimators of clustering. We can summarize it all by saying that all the information we have on the large scale distribution can be expressed by the power spectrum $P(k)$ and related quantities, and by the higher order parameters like S_3, S_4. Usually the power spectrum itself is parametrized by an amplitude σ_8 and a shape parameter, Γ, so that we can embody our knowledge in four number, $\sigma_8, \Gamma, S_3, S_4$, and a biasing factor if we want to relate observations of different objects to the dark matter distribution. Higher order moments, or more refined parametrization of the power spectrum are rarely employed, given the relatively small amount of data we still have. Of course, we have to be sure that our sample is big enough to include all relevant wavelengths, otherwise we will find different results in different samples.

Chapter 10

Origin of inflationary perturbations

Quick summary

- In general, as we have seen, the perturbations in GR are of three types, scalar, vector and tensor.
- The inflationary quantum mechanism that we are going to discuss will in general excite all three types. However, here again we decide to neglect the vector perturbations since they decay rapidly away.
- The scalar perturbations are complicated by the fact that one needs to perturb at the same time the scalar field and the metric.
- The tensor perturbations are somehow the easiest to deal with, so we begin with this case.
- We derive the initial spectrum induced by inflation as a function of the slow-rolling parameters.
- Most of the discussion that follows is adapted from Dodelson, Modern Cosmology, Chapter 6.

10.1 From a harmonic oscillator to field quantization

Let us consider a simple harmonic oscillator:

$$\ddot{x} + \omega^2 x = 0 \tag{10.1.1}$$

After quantization, the solution can be formed as a sum of creation/annihilation operators

$$\hat{x} = v(\omega, t)\hat{a} + v^*(\omega, t)\hat{a}^\dagger \tag{10.1.2}$$

where $v, v^*$are a solution of (10.1.1) and its complex conjugate. Clearly

$$v(\omega, t) \sim e^{-i\omega t} \tag{10.1.3}$$

The operators $\hat{a}, \hat{a}^\dagger$ obey the commutation relations

$$[\hat{a}, \hat{a}^\dagger] \equiv \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = 1 \tag{10.1.4}$$

$$[\hat{a}, \hat{a}] = [\hat{a}^\dagger, \hat{a}^\dagger] = 0 \tag{10.1.5}$$

and of course the operator $\hat{a}$ annihilates the vacuum state:

$$\hat{a}|0\rangle = 0 \tag{10.1.6}$$

These relations imply

$$[\hat{x},\hat{p}] = i \tag{10.1.7}$$

if $\hat{p} = d\hat{x}/dt = -i\omega v\hat{a} + i\omega v^*\hat{a}^\dagger$ is the momentum operator and if v is normalized such that

$$v(\omega, t) = \frac{e^{-i\omega t}}{\sqrt{2\omega}} \tag{10.1.8}$$

The variance of $\hat{x}$ can then be evaluated as

$$\langle|\hat{x}|^2\rangle = \langle 0|\hat{x}^\dagger\hat{x}|0\rangle = |v^2|\langle 0|[\hat{a}, \hat{a}^\dagger] + \hat{a}^\dagger\hat{a}|0\rangle = |v^2| \tag{10.1.9}$$

In GR a generic tensor perturbation has two degress of freedom, corresponding to two polarization states, h_+, h_X (massless spin 2 field). At first order they both obey an equation which is very smilar to the harmonic one:

$$\ddot{h} + 2\frac{\dot{a}}{a}h + k^2 h = 0 \tag{10.1.10}$$

where as in the previous Chapter the dot represents derivative wrt to conformal time τ. If we define

$$\tilde{h} = \frac{ah}{\sqrt{16\pi G}} \tag{10.1.11}$$

then we obtain

$$\ddot{\tilde{h}} + (k^2 - \frac{\ddot{a}}{a})\tilde{h} = 0 \tag{10.1.12}$$

and as before the quantization of this oscillator requires the introduction of the operator

$$\hat{\tilde{h}} = v(k, \tau)\hat{a} + v^*(k, \tau)\hat{a}^\dagger \tag{10.1.13}$$

where v satisfies Eq. (10.1.12). However now $\tilde{h}$ is a field with a continuous distribution of momenta $\mathbf{k}$; its commutation relation is therefore

$$[\hat{\tilde{h}}^\dagger, \hat{\tilde{h}}] = |v(k, \tau)|^2(2\pi)^3\delta_D(\mathbf{k} - \mathbf{k}') \tag{10.1.14}$$

where δ_D is a Dirac delta. Then the variance of $\hat{h}$ is

$$\langle\hat{h}^\dagger\hat{h}\rangle = \frac{16\pi G}{a^2}v^2(2\pi)^3\delta_D(\mathbf{k} - \mathbf{k}') \tag{10.1.15}$$

$$= P_h(k)(2\pi)^3\delta_D(\mathbf{k} - \mathbf{k}') \tag{10.1.16}$$

where we introduced the power spectrum

$$P_h = 16\pi G\frac{|v(k, \tau)|^2}{a^2} \tag{10.1.17}$$

(in this definition of power spectrum, the factors of volumes are neglected by convention). Now during inflation, and assuming H approximately constant, the conformal time is

$$\tau = \int_{a_e}^{a}\frac{dt}{a} = -\frac{1}{aH} + \frac{1}{a_e H} \approx -\frac{1}{aH} \tag{10.1.18}$$

if we consider an epoch of inflation much before the end, i.e. $a_e \gg a$. The negative sign is necessary in order to have a growing time. Then we have ($H = \dot{a}/a^2$ since "dot" here is derivative wrt τ)

$$\dot{a} = -\frac{a}{\tau}, \quad \frac{\ddot{a}}{a} = \frac{2}{\tau^2} \tag{10.1.19}$$

so that replacing into Eq. (10.1.12) we see that v obeys the equation

$$\ddot{v} + (k^2 - \frac{2}{\tau^2})v = 0 \tag{10.1.20}$$

This equation can be solved exactly in terms of Bessel functions. With the initial conditions such that initially one has only an outgoing wave, i.e. $v \sim e^{-ik\tau}$, one gets the exact solution

$$v = \frac{e^{-ik\tau}}{\sqrt{2k}}(1 - \frac{i}{k\tau}) \tag{10.1.21}$$

where we employed the same normalization $\sqrt{2k}$ we have used for the harmonic oscillator. When the waves exit the horizon during inflation, ie for $k|\tau| \ll 1$, one gets

$$v = -\frac{e^{-ik\tau}}{\sqrt{2k}}\frac{i}{k\tau} \tag{10.1.22}$$

$$P_h = 8\pi G\frac{H^2}{k^3} \tag{10.1.23}$$

where we also employed the inflationary solution (10.1.18) $a\tau = -1/H$. This is the expression for the power spectrum when the waves exit the horizon during inflation. We may expect that the spectrum amplitude does not grow as long as the waves remains super-horizon sized, since gravity is effectively frozen on such scales. This spectrum will therefore be the same for large scales when it is finally observed imprinted on the cosmic microwave fluctuations.

On the CMB, the observed fluctuations are smaller than $\Delta T/T \approx 10^{-4}$. The variance in real space one expects from a spectrum P_h is (see eq. 9.9.4)

$$\sigma_R^2 \approx \int P_h W^2(k) d^3k \approx 8\pi G H^2 \int_{k_{min}}^{\infty} k^{-3} W^2(k,R) d^3k \tag{10.1.24}$$

The last integral is of order unity (it actually has a small logarithmic dependence on the cut-off k_{min}), so $\sigma^2 \approx GH^2$, up to order-of-unity factors. The fluctuations due to the GW therefore must be smaller or equal to the observed signal, which means $GH^2 = G^2V = M_P^{-4}E_{infl}^4 \leq 10^{-8}$. This gives $E_{infl} \leq 10^{-2}M_P$, well below the Planck scale. If we could measure directly the tensor perturbations on the CMB we could fix the inflationary scale.

The factor $16\pi G$ that we inserted in Eq. (10.1.11) can be justified only by writing down the Lagrangian to second order in h. Then one sees that in order to obtain a canonical kinetic term, such a factor becomes necessary. In turn, a canonical kinetic term is needed to adopt the standard commutation relations.

The slope of the spectrum is defined as

$$n_T - 3 = \frac{d\log P_h}{d\log k} \tag{10.1.25}$$

so that

$$n_T = 3 + \frac{d\log(H^2k^{-3})}{d\log k} = 3 + \frac{d\log k^{-3}}{d\log k} + \frac{d\log H^2}{d\log k} = 2\frac{d\log H}{d\log k} \tag{10.1.26}$$

Now we have

$$\frac{d\log H}{d\log k} = \frac{k}{H}\frac{dH}{dk} = \frac{k}{H}\frac{dH}{d\tau}\frac{d\tau}{dk}|_{k=aH} \tag{10.1.27}$$

The slope depends therefore on how H changes with time, ie on $\dot{H}$. Before we assumed that H is approximately constant during inflation, but now this assumption would give zero for n_T, so we must work out the first order correction. We define two dimensionless slow rolling parameters:

$$\varepsilon = \frac{dH^{-1}}{dt} = -\frac{\dot{H}}{aH^2} \tag{10.1.28}$$

$$\delta = \frac{1}{H}\frac{d^2\phi/dt^2}{d\phi/dt} = -\frac{a}{H\dot{\phi}}(3H\frac{\dot{\phi}}{a} + V') \tag{10.1.29}$$

where the prime is $d/d\phi$ (Other choices for the slow-rolling parameters have been introduced, for instance $\varepsilon_1 = \varepsilon$ and the sequence $\varepsilon_{n+1} = \frac{d\ \ln|\varepsilon_n|}{dN}$. Then one has $\varepsilon_2 = 2\delta + 2\varepsilon$.) Notice that $\delta = 0$ in the $m^2\phi^2$ case we have seen before, because we have $d^2\phi/dt^2 = 0$.

The two parameters can be written also directly in terms of the inflationary potential V . Since we are interested only in the first order corrections, we assume now $3H^2 \approx 8\pi GV$ and $3H(d\phi/dt) + V' = 0$; differentiating the second relation and using the first, we obtain $d^2\phi/dt^2 = -d\phi/dt(V'' + 3dH/dt)/3H$. Then one obtains

$$\varepsilon = \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2 \tag{10.1.30}$$

$$\delta = \varepsilon - \frac{1}{8\pi G}\frac{V''}{V} \tag{10.1.31}$$

Then we see from Eq. (10.1.28) that

$$\dot{H} = -aH^2\varepsilon \tag{10.1.32}$$

Moreover, we have $\tau = -1/aH$ and therefore

$$\frac{d\tau}{dk}|_{k=aH=|\tau^{-1}|} = -\frac{d(aH)^{-1}}{dk} = -\frac{d(1/k)}{dk} = \frac{1}{k^2} \tag{10.1.33}$$

Finally, we obtain

$$n_T = 2\frac{d\log H}{d\log k} = 2\frac{k}{H}\frac{dH}{d\tau}\frac{d\tau}{dk}|_{k=aH} = 2\left[\frac{k}{H}(-aH^2\varepsilon)\frac{1}{k^2}\right]|_{k=aH} = 2\left[-\varepsilon aH\frac{1}{k}\right]|_{k=aH} = -2\varepsilon \tag{10.1.34}$$

This important relation shows that measuring n_T is equivalent to measuring ε, i.e. the rate of change of H.

10.2 Scalar perturbations

The same technique can be applied to the scalar perturbations. Here however one has to first perturb the field ϕ and then relate it to the metric perturbations Φ, Ψ. A detailed treatment can be found in Dodelson, Modern Cosmology, Chapter 6. Here we just state the final result. The power spectrum for the potentials Ψ, Φ is

$$P_\Psi = P_\Phi = \frac{8\pi G}{9k^3}\frac{H^2}{|\varepsilon|}|_{k=aH} \tag{10.2.1}$$

The most important difference is that now there is the small number ε at the denominator. This means that we expect

$$P_S \gg P_T \tag{10.2.2}$$

where S, T denote scalar, tensor perturbations.

The slope of the scalar modes is defined as

$$n_S - 4 = \frac{d\log P_\Psi}{d\log k} \tag{10.2.3}$$

This definition implies that the power spectrum of the density contrast δ_m, which through the Poisson equation is proportional to $k^2\Psi$ for sub-horizon modes, is

$$P_{\delta_m} \sim k^4 P_\Psi \tag{10.2.4}$$

and therefore the slope of P_{δ_m}will be exactly n_S.

To estimate n_S we do as before. We have then

$$n_S - 1 = \frac{d}{d\log k}[\log H^2 - \log\varepsilon] \tag{10.2.5}$$

The first term gives -2ε from (10.1.34). The second one can be obtained noting that $d/d\log k \approx d/d\log a$ since $k = aH$ and the variation of H is negligible wrt the variation of a. Then we have from Eq. (10.1.30) and from $d\phi/d\log a = -V'/3H^2 = -V'/(8\pi GV)$ that

$$\frac{d\log\varepsilon}{d\log k} \approx \frac{d\log\varepsilon}{d\log a} = \frac{d\log\varepsilon}{d\phi}\frac{d\phi}{d\log a} = -\frac{1}{4\pi G}(\frac{V''}{V'} - \frac{V'}{V})\frac{V'}{V} = 2(\varepsilon+\delta) \tag{10.2.6}$$

so that finally

$$n_S = 1 - 4\varepsilon - 2\delta \tag{10.2.7}$$

Both n_S and n_T are in general a function of time and therefore, through the relation $k = aH$, of the scale k. The current estimated value of n_s is 0.96 ± 0.01 (Planck 2015). In many models δ is quite smaller than ϵ so, since ϵ is positive definite, we xpect indeed n_S slightly smaller than unit, as observed.

The ratio of tensor to scalar spectra can be defined as

$$r_P = \frac{P_h}{P_\Psi} = |9\varepsilon| \tag{10.2.8}$$

but actually it is normally defined as the ratio of the observed spectra, rather than the spectra at horizon exit as here. The observed ratio r turns out to be in simple models of inflation equal to roughly 10ε, but the exact value is model- and k-dependent. One very common way to represent the constraints from CMB observations is to employ the parameter plane τ, n_S, as in Fig. (10.2.1).

In summary, while the tensor spectrum amplitude measures directly the inflationary energy scale $H^2 \approx 8\pi GV/3$, the scalar spectrum and the spectra slopes measure also the potential derivatives V', V'' through ε and δ.

It is interesting to notice that from the slow roll relation $3H^2\phi' + V' = 0$ (the prime is d/dN) one has

$$\frac{V'}{V} = \frac{3H^2}{V}\frac{d\phi}{dN} = 8\pi G\frac{\Delta\phi}{\Delta N} \tag{10.2.9}$$

and therefore, from (10.1.30),

$$8\pi G\frac{\Delta\phi}{\Delta N} = \sqrt{16\pi G\varepsilon} = \sqrt{16\pi G\frac{r}{10}} \tag{10.2.10}$$

If, as the experiment BICEP2 (2014) reported, $r \approx 0.2$, then

$$\Delta\phi \approx 0.14 M_P \Delta N \tag{10.2.11}$$

where $M_P = (8\pi G)^{-2}$ is the reduced Planck mass. Here ΔN is the number of e-folding elapsed while ϕ moves by $\Delta\phi$. The observational range observed by BICEP2 is of a few e-foldings, eg $\Delta N \approx 4$, so finally we see that the field ϕ has to move by half a Planck mass or so to produce a value of r close to the observed one. Although is not difficult to design by hand models that satisfy this condition, many models that have some theoretical support predict an extremely flat potential and consequently a very small change in ϕ. In any case, the BICEP results have been found to be erroneous so at the moment there is no detection of r but only upper limits, $r < 0.1$ roughly.

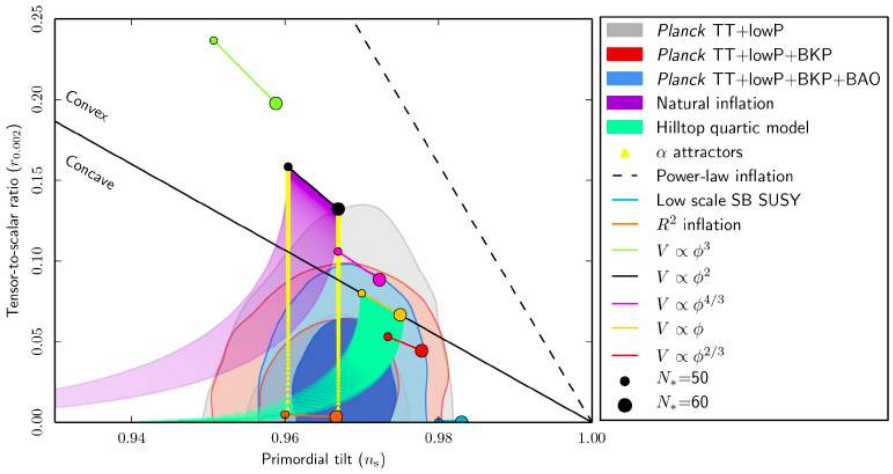


Figure 10.2.1: Tensor-to-scalar ratio r and spectral slope n_S estimated by the Planck satellite (2015). From arXiv:1502.02114, Planck 2015 (XX) Ade et al..

Chapter 11

The Cosmic Microwave Background

The cosmic microwave background (CMB) is the oldest cosmological signal we can currently observe. Mapping its fluctuations gives a wealth of accurate information on processes before, during and after the recombination. The recent WMAP and Planck satellites have measured several parameter to percent accuracy by comparing the temperature and polarization anisotropies to predictions.

Quick summary

- The CMB photons have been first captured in 1965 by Penzias and Wilson, definitely confirming the hot big bang scenario
- The primordial fluctuations created during inflation left their imprint on the photons last scattered at recombination, i.e. at a redshift of around 1000
- These temperature anisotropies, first observed by the COBE satellite, are due to several physical phenomena at recombination time: Sachs-Wolfe effect, baryon oscillation, Silk damping
- Other phenomena are produced along the line of sight: integrated Sachs-Wolfe, lensing, reionization
- Polarization adds another dimension to observations; primordial gravitational waves leave a characteristic imprint on the B-mode polarization
- One of the main problem of CMB observations is the removal of foregrounds due to various non-cosmological effects

11.1 A short history of the CMB research

In the '50 scientists started taking into serious consideration the effects of a hot big bang as implied by the discovery of the cosmic expansion by Hubble. Alpher, Herman, Gamow and others realized that the photons that initially dominated the energy density should have been still around after they last scattered when neutral atoms first formed at a temperature of a few thousand degrees and at an epoch roughly 350,000 years after the big bang. They predicted then that a isotropic black-body bath of radiation at around 5K could be detected. This indeed occurred in 1965 when Penzias and Wilson discovered by chance a unexplained isotropic 3K radiation, correctly interpreted by Dicke, Peebles and collaborators as the sought-for big bang remnant.

After the full relativistic treatment of linear perturbations developed by Lifshitz and Khalatnikov, and Sakharov in the early 60s, Sachs and Wolfe (1966) showed then that the existence of today's inhomogeneity should imply some level of inhomogeneities also on the CMB due to the gravitational red- or blueshift induced by matter perturbations of CMB photons. In subsequent years, Silk, Peebles, Yu and others found out how the interactions between photon and baryons before and during the recombination process could have imprinted small angular scale perturbations due to the propagation of sound waves in the coupled plasma. In the early 80s these calculations were extended to take into account dark matter (Silk, Vittorio, Efstathiou, Bond).

Table 4. Parameter 68 % confidence limits for the base ΛCDM model from *Planck* CMB power spectra, in combination with lensing reconstruction ("lensing") and external data ("ext," BAO+JLA+H_0). Nuisance parameters are not listed for brevity (they can be found in the PLA tables), but the last three parameters give a summary measure of the total foreground amplitude (in μK^2) at $\ell = 2000$ for the three high-ℓ temperature spectra used by the likelihood. In all cases the helium mass fraction used is predicted by BBN (posterior mean $Y_P \approx 0.2453$, with uncertainty due to *Planck*'s error on $\Omega_b h^2$ being smaller than the rate uncertainty and difference between BBN codes).

Parameter	TT+lowP 68 % limits	TT+lowP+lensing 68 % limits	TT+lowP+lensing+ext 68 % limits	TT,TE,EE+lowP 68 % limits	TT,TE,EE+lowP+lensing 68 % limits	TT,TE,EE+lowP+lensing+ext 68 % limits
$\Omega_b h^2$	0.02222 ± 0.00023	0.02226 ± 0.00023	0.02227 ± 0.00020	0.02225 ± 0.00016	0.02226 ± 0.00016	0.02230 ± 0.00014
$\Omega_c h^2$	0.1197 ± 0.0022	0.1186 ± 0.0020	0.1184 ± 0.0012	0.1198 ± 0.0015	0.1193 ± 0.0014	0.1188 ± 0.0010
$100\theta_{MC}$	1.04085 ± 0.00047	1.04103 ± 0.00046	1.04106 ± 0.00041	1.04077 ± 0.00032	1.04087 ± 0.00032	1.04093 ± 0.00030
τ	0.078 ± 0.019	0.066 ± 0.016	0.067 ± 0.013	0.079 ± 0.017	0.063 ± 0.014	0.066 ± 0.012
$\ln(10^{10}A_s)$	3.089 ± 0.036	3.062 ± 0.029	3.064 ± 0.024	3.094 ± 0.034	3.059 ± 0.025	3.064 ± 0.023
n_s	0.9655 ± 0.0062	0.9677 ± 0.0060	0.9681 ± 0.0044	0.9645 ± 0.0049	0.9653 ± 0.0048	0.9667 ± 0.0040
H_0	67.31 ± 0.96	67.81 ± 0.92	67.90 ± 0.55	67.27 ± 0.66	67.51 ± 0.64	67.74 ± 0.46
Ω_Λ	0.685 ± 0.013	0.692 ± 0.012	0.6935 ± 0.0072	0.6844 ± 0.0091	0.6879 ± 0.0087	0.6911 ± 0.0062
Ω_m	0.315 ± 0.013	0.308 ± 0.012	0.3065 ± 0.0072	0.3156 ± 0.0091	0.3121 ± 0.0087	0.3089 ± 0.0062
$\Omega_m h^2$	0.1426 ± 0.0020	0.1415 ± 0.0019	0.1413 ± 0.0011	0.1427 ± 0.0014	0.1422 ± 0.0013	0.14170 ± 0.00097
$\Omega_m h^3$	0.09597 ± 0.00045	0.09591 ± 0.00045	0.09593 ± 0.00045	0.09601 ± 0.00029	0.09596 ± 0.00030	0.09598 ± 0.00029
σ_8	0.829 ± 0.014	0.8149 ± 0.0093	0.8154 ± 0.0090	0.831 ± 0.013	0.8150 ± 0.0087	0.8159 ± 0.0086
$\sigma_8\Omega_m^{0.5}$	0.466 ± 0.013	0.4521 ± 0.0088	0.4514 ± 0.0066	0.4668 ± 0.0098	0.4553 ± 0.0068	0.4535 ± 0.0059
$\sigma_8\Omega_m^{0.25}$	0.621 ± 0.013	0.6069 ± 0.0076	0.6066 ± 0.0070	0.623 ± 0.011	0.6091 ± 0.0067	0.6083 ± 0.0066
z_{re}	$9.9^{+1.8}_{-1.6}$	$8.8^{+1.7}_{-1.4}$	$8.9^{+1.3}_{-1.2}$	$10.0^{+1.7}_{-1.5}$	$8.5^{+1.4}_{-1.2}$	$8.8^{+1.2}_{-1.1}$
$10^9 A_s$	$2.198^{+0.076}_{-0.085}$	2.139 ± 0.063	2.143 ± 0.051	2.207 ± 0.074	2.130 ± 0.053	2.142 ± 0.049
$10^9 A_s e^{-2\tau}$	1.880 ± 0.014	1.874 ± 0.013	1.873 ± 0.011	1.882 ± 0.012	1.878 ± 0.011	1.876 ± 0.011
Age/Gyr	13.813 ± 0.038	13.799 ± 0.038	13.796 ± 0.029	13.813 ± 0.026	13.807 ± 0.026	13.799 ± 0.021
z_*	1090.09 ± 0.42	1089.94 ± 0.42	1089.90 ± 0.30	1090.06 ± 0.30	1090.00 ± 0.29	1089.90 ± 0.23
r_*	144.61 ± 0.49	144.89 ± 0.44	144.93 ± 0.30	144.57 ± 0.32	144.71 ± 0.31	144.81 ± 0.24
$100\theta_*$	1.04105 ± 0.00046	1.04122 ± 0.00045	1.04126 ± 0.00041	1.04096 ± 0.00032	1.04106 ± 0.00031	1.04112 ± 0.00029
z_{drag}	1059.57 ± 0.46	1059.57 ± 0.47	1059.60 ± 0.44	1059.65 ± 0.31	1059.62 ± 0.31	1059.68 ± 0.29
r_{drag}	147.33 ± 0.49	147.60 ± 0.43	147.63 ± 0.32	147.27 ± 0.31	147.41 ± 0.30	147.50 ± 0.24
k_D	0.14050 ± 0.00052	0.14024 ± 0.00047	0.14022 ± 0.00042	0.14059 ± 0.00032	0.14044 ± 0.00032	0.14038 ± 0.00029
z_{eq}	3393 ± 49	3365 ± 44	3361 ± 27	3395 ± 33	3382 ± 32	3371 ± 23
k_{eq}	0.01035 ± 0.00015	0.01027 ± 0.00014	0.010258 ± 0.000083	0.01036 ± 0.00010	0.010322 ± 0.000096	0.010288 ± 0.000071
$100\theta_{s,eq}$	0.4502 ± 0.0047	0.4529 ± 0.0044	0.4533 ± 0.0026	0.4499 ± 0.0032	0.4512 ± 0.0031	0.4523 ± 0.0023
f_{2000}^{143}	29.9 ± 2.9	30.4 ± 2.9	30.3 ± 2.8	29.5 ± 2.7	30.2 ± 2.7	30.0 ± 2.7
$f_{2000}^{143\times217}$	32.4 ± 2.1	32.8 ± 2.1	32.7 ± 2.0	32.2 ± 1.9	32.8 ± 1.9	32.6 ± 1.9
f_{2000}^{217}	106.0 ± 2.0	106.3 ± 2.0	106.2 ± 2.0	105.8 ± 1.9	106.2 ± 1.9	106.1 ± 1.8

Table 5. Constraints on 1-parameter extensions to the base ΛCDM model for combinations of *Planck* power spectra, *Planck* lensing, and external data (BAO+JLA+H_0,"ext"). Note that we quote 95 % limits here.

Parameter	TT	TT+lensing	TT+lensing+ext	TT, TE, EE	TT, TE, EE+lensing	TT, TE, EE+lensing+ext
Ω_K	$-0.052^{+0.049}_{-0.055}$	$-0.005^{+0.016}_{-0.017}$	$-0.0001^{+0.0054}_{-0.0052}$	$-0.040^{+0.038}_{-0.041}$	$-0.004^{+0.015}_{-0.015}$	$0.0008^{+0.0040}_{-0.0039}$
Σm_ν [eV]	< 0.715	< 0.675	< 0.234	< 0.492	< 0.589	< 0.194
$N_{\rm eff}$	$3.13^{+0.64}_{-0.63}$	$3.13^{+0.62}_{-0.61}$	$3.15^{+0.41}_{-0.40}$	$2.99^{+0.41}_{-0.39}$	$2.94^{+0.38}_{-0.38}$	$3.04^{+0.33}_{-0.33}$
Y_P	$0.252^{+0.041}_{-0.042}$	$0.251^{+0.040}_{-0.039}$	$0.251^{+0.035}_{-0.036}$	$0.250^{+0.026}_{-0.027}$	$0.247^{+0.026}_{-0.027}$	$0.249^{+0.025}_{-0.026}$
$dn_s/d\ln k$	$-0.008^{+0.016}_{-0.016}$	$-0.003^{+0.015}_{-0.015}$	$-0.003^{+0.015}_{-0.014}$	$-0.006^{+0.014}_{-0.014}$	$-0.002^{+0.013}_{-0.013}$	$-0.002^{+0.013}_{-0.013}$
$r_{0.002}$	< 0.103	< 0.114	< 0.114	< 0.0987	< 0.112	< 0.113
w	$-1.54^{+0.62}_{-0.50}$	$-1.41^{+0.64}_{-0.56}$	$-1.006^{+0.085}_{-0.091}$	$-1.55^{+0.58}_{-0.48}$	$-1.42^{+0.62}_{-0.56}$	$-1.019^{+0.075}_{-0.080}$

Figure 11.0.1: Cosmological parameters estimated by the Planck satellite (2015). From irsa.ipac.caltech.edu/data/Planck/release_2/ancillary-data/

Nowadays there are several publicly available numerical codes that in a very short time perform the full set of calculations in the linear (or even non-linear) regime, including dark matter, baryons, photons, neutrinos, dark energy, curvature and all possible physical effects, finally producing the predicted CMB and matter clustering spectra. Among these codes we mention CMBFAST, CMBeasy, CAMB, CLASS.

In the mean time, observations of the CMB failed to find anisotropies until the COBE satellite launched in 1989 by NASA confirmed the almost perfect black body nature of the radiation spectrum and finally found small fluctuations at angular scales larger than 7 degrees. In the following years many other ground based observations tried to increase the resolution to detect the baryon acoustic oscillations. This was first achieved by the Boomerang ballon experiment (de Bernardis), and other similar experiments. The detection of the baryon peak allowed for the first time to measure the total energy density Ω_{tot}, which turned out to be close to unity as predicted by inflation. In 2003 the high-resolution WMAP NASA satellite extended the resolution to 10 primes roughly, while Planck (launch 2009), a ESA satellite increased it to roughly 5 primes, also adding more channels to improve foreground subtraction, and included for the first time sensitive polarization measurements. Planck measured several cosmological parameters to a percent accuracy (see Fig. 11.0.1).

11.2 Anisotropies on the cosmic microwave background

At a temperature of roughly 3,000 K, protons combine with electrons to form neutral hydrogen atoms, and decouple from the radiation. The background radiation we receive today has been emitted therefore roughly when the scale factor was $a_d = 3/3,000 = 0.001$, if today's temperature is 3K.A black body at this temperature has a wavelength peak at $\lambda_{max} \approx (2.3 \cdot 10^{-3} K \cdot m) T^{-1} \approx 1mm$. The redshift of decoupling is $z_{dec} \simeq 1000$. The distribution of the matter at that epoch is therefore imprinted on the background radiation. There are three main effects arising from the decoupling surface:

a) The Sachs-Wolfe effect. This occurs because when the photons are emitted from inside a overdensity, they have to climb out the gravitational potential. The SW effect dominates at scales much larger than the horizon, corresponding to roughly 1^0

b) The acoustic effect. This occurs because the baryons and the photons perturbations oscillate acoustically when they are smaller than their combined Jeans length. We see then fluctuations on the scale of the Jeans length and higher, corresponding to scales around 1^0 and smaller.

c) The adiabatic effect. This occurs because if the initial conditions are adiabatic then

$$\delta_\gamma = \frac{4}{3}\delta_c. \tag{11.2.1}$$

All these effects are averaged out on scales much smaller than the last scattering thickness, which is roughly 7 Mpc/h. In terms of angular scales, this is

$$\theta_{th} = 7/6000(180 \cdot 60/\pi) = 4' \tag{11.2.2}$$

Two effects arise instead along the radiation geodesic to us:

d) The integrated SW effect. This occurs whenever the gravitational potential of perturbations is not constant. Suppose in fact it increases with time. Then, a photon enter at a time when the potential is smaller than when it exits. Therefore, the blueshift acquired when falling is smaller than the redshift suffered when exiting, and there is a net redshift contribution. This happens both when the universe is not flat, or when the perturbations do not grow as a, or when the fluctuations become non-linear (Rees-Sciama effect). In the latter case the effect is on very small scales (ten arcminutes or so).

e) The Sunyaiev-Zeldovich effect. This occurs when the CMB photon crosses a cluster which contains hot plasma. The electrons in the plasma confer energy to the CMB photons, thereby increasing their temperature (inverse Compton scattering). This effect occurs on very small scales, of the order of some arcminutes.

Let us now focus on the first three effects. We can write that the temperature of the photons coming from the last scattering surface varies by a $\Delta T/T$ due to the sum of these three effects

$$\frac{\Delta T}{T} = \text{grav. redshift} + \text{adiab. pert.} + \text{acoustic oscill.} \tag{11.2.3}$$

We'll consider these three effects in turn.

11.3 The CMB power spectrum

A general field $f(\theta,\varphi)$ over a sphere can always be expanded in multipoles

$$f(\theta,\varphi)=\sum_{ml} a_\ell^m Y_{m\ell}(\theta,\varphi) \tag{11.3.1}$$

where $Y_{m\ell}$ are the spherical harmonics (see e.g. Abramovitz-Stegun, ch. 8 or Arfken & Weber, ch. 12),

$$Y_{m\ell}(\theta,\varphi) = Ne^{im\varphi}P_\ell^m(\cos\theta) \tag{11.3.2}$$

$$N = (-1)^m\sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} \tag{11.3.3}$$

and P_ℓ^m are the associated Legendre functions. A multipole ℓ corresponds to fluctuations of angular scale approximately equal to $180/\ell$. For instance, $\ell=1$, the dipole, gives the temperature fluctuation averaged over hemispheres, and $\ell=2$, the quadrupole term, corresponds to features that extends over 90^0. The angular scale of the decoupling surface thickness, 4', corresponds to a multipole $\ell=180\cdot 60/4=2700$. Beyond this multipole, therefore, the intrinsic fluctuations will be smeared out. The acoustic peak at 1^0 will appear at a multipole of $\ell\simeq 180$.

The constant N ensures the orthonormality of the spherical harmonics

$$\int Y_{m\ell}Y^*_{m'\ell'}d\Omega=\delta_{\ell\ell' mm'} \tag{11.3.4}$$

where here and in the following, $d\Omega=d\cos\theta d\varphi$. We also use, interchangeably, the notation d^2q for the angular area element, and $\mathbf{q}$ for the unit vector specifying a direction. The expansion coefficients a_ℓ^m can be found by inversion,

$$a_\ell^m=\int f(\theta,\varphi)Y^*_{m\ell}(\theta,\varphi)d\Omega \tag{11.3.5}$$

We apply now this formalism to CMB anisotropy. In any direction of the sky we observe a primordial blackbody spectrum characterized by a temperature T. We call temperature anisotropies the differences $\Delta T/T$ that see in different directions. When we confront the predicted temperature anisotropies with CMB observations, it is convenient to expand the fluctuation $\Delta T/T$ in terms of spherical harmonics:

$$\frac{\Delta T}{T}(\hat n)=\sum_{\ell=1}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}Y_{\ell m}(\hat n)\,, \tag{11.3.6}$$

where the subscripts ℓ and m are conjugate to a real space unit vector $\hat n$ representing the direction of the incoming photons, i.e. the angular position in the sky. The coefficients $a_{\ell m}$ in Eq. (11.3.6) are assumed to be statistically independent. This means that the mean value of $a_{\ell m}$'s is zero ($\langle a_{\ell m}\rangle=0$) with a nonzero variance defined by

$$\hat C_\ell\equiv\langle|a_{\ell m}|^2\rangle\,. \tag{11.3.7}$$

What is usually is denoted CMB power spectrum is (Fig. 11.3.2)

$$C_\ell=\frac{\ell(\ell+1)\hat C_\ell}{2\pi} \tag{11.3.8}$$

11.4 The Sachs-Wolfe effect

If photons are last scattered towards our telescopes from inside an overdensity, they are gravitationally redshifted because they lose energy to climb up the potential; viceversa they are blueshifted from inside an underdensity.

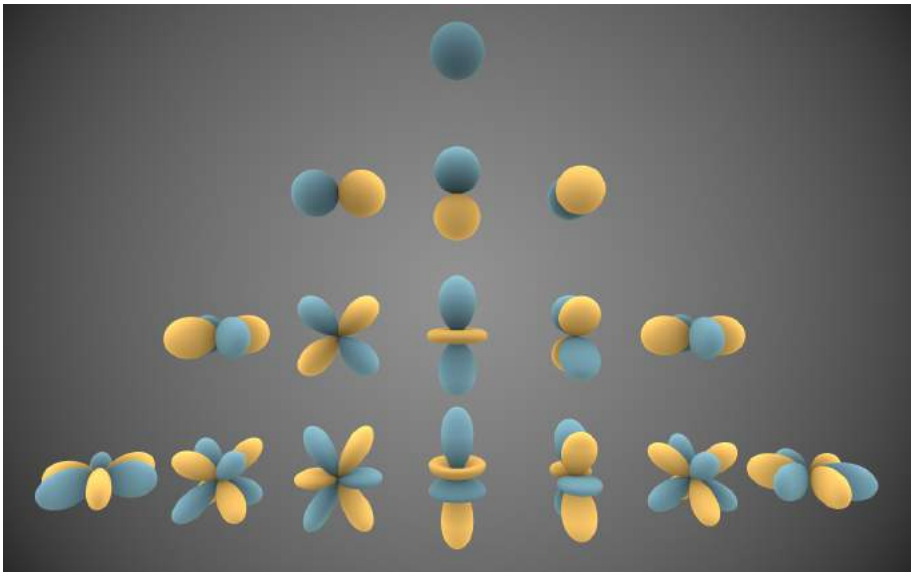


Figure 11.3.1: First spherical harmonics ($\ell = 0, 1, 2, 3$). Yellow regions are negative, blue are positive. The distance to the origin is proportional to the value of $Y_{\ell m}$ in that direction (By Inigo.quilez - own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=32782753).

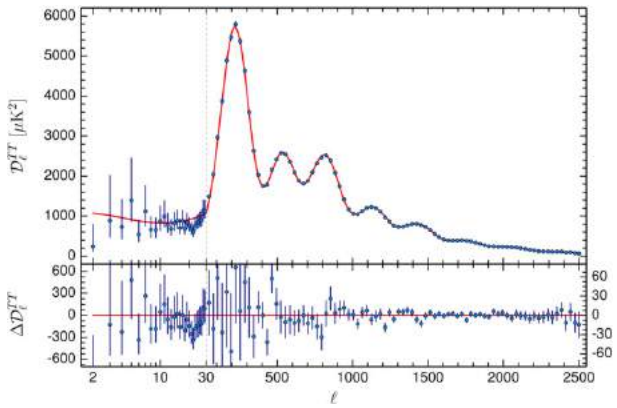


Figure 11.3.2: Planck (2015) temperature anisotropy spectrum (Ade et al. A&A 594, A13, 2016; ESA and the Planck Collaboration)

Since we are in a linear regime, the effect must be proportional to the gravitational potential Ψ. It turns out that the temperature anisotropy is (more details in the Appendix to this chapter)

$$\frac{\Delta T}{T} = \frac{\Psi}{3} \tag{11.4.1}$$

(for adiabatic perturbations, i.e. when matter and radiation are equally perturbed). Compared to the other effects, the SW effect is maximal for large perturbations that are still superhorizon at last scattering. For an initial inflationary power spectrum $P(k) \sim k^1$ one finds $C_\ell \sim 1/\ell(\ell+1)$ so that conventionally the power spectrum is multiplied by $\ell(\ell+1)$ to render it flat for the SW scales (SW plateau).

The comoving horizon scale at redshift z in a spatially flat (which is a good approximation at large z) matter-dominated universe (neglecting the fact that for $z > 3500$ the Universe was radiation dominated) is

$$d_H^c = \int_z^\infty \frac{dz'}{H(z')} = \frac{2}{H_0\sqrt{\Omega_m^{(0)}(1+z)}} \tag{11.4.2}$$

At recombination,r $z_{rec} = 1100$, we obtain

$$d_H^c(z_{rec}) \approx 180(\Omega_m h^2)^{-1/2} Mpc \tag{11.4.3}$$

The comoving angular diameter distance $d_A^{(c)}(z_{\rm rec})$ to recombination is expressed as (we now include curvature)

$$d_A^c(z_{\rm rec}) = \int_0^{z_{rec}} \frac{dz'}{H(z')} = \frac{c}{H_0}\sqrt{\frac{1}{\Omega_K^{(0)}}}\sinh\left(\sqrt{\Omega_K^{(0)}}\int_0^{z_{\rm dec}}\frac{{\rm d}z}{E(z)}\right) = \frac{c}{H_0}\frac{1}{\sqrt{\Omega_m^{(0)}}}\mathcal{R}\,, \tag{11.4.4}$$

where $\mathcal{R}$ is the so-called CMB shift parameter defined by

$$\mathcal{R} = \sqrt{\frac{\Omega_m^{(0)}}{\Omega_K^{(0)}}}\sinh\left(\sqrt{\Omega_K^{(0)}}\int_0^{z_{\rm dec}}\frac{{\rm d}z}{E(z)}\right)\,. \tag{11.4.5}$$

and in ΛCDM we have $E^2(z) = \Omega_m^{(0)}(1+z)^3 + \Omega_K^{(0)}(1+z)^2 + (1-\Omega_m^{(0)} - \Omega_K^{(0)})$. The angular size of the horizon at recombination is therefore

$$\theta_H \approx \frac{d_H^c(z_{rec})}{d_A^c(z_{rec})} \approx (0.0345 - 0.05\Omega_K^{(0)}){\rm rad} \approx 2\,{\rm deg} \tag{11.4.6}$$

where the approximation is valid for $\Omega_m^{(0)} = 0.3$ and for $|\Omega_K^{(0)}| < 0.05$. corresponding to $\ell \approx 90$. One can see that if $\Omega_K^{(0)}$ increases, θ_H decreases, and ℓ increases (the peak moves to the right, as in Fig. 11.9.1). At larger scales, where the SW dominates, we expect therefore C_ℓ to have an almost flat trend.

It is important to remark that the SW effect depends on the total matter content, not just the baryonic one. In fact, if there were no dark matter, the SW effect would be much smaller than observed or, more precisely, the SW plateau temperature anisotropy would be much lower, with respect to the baryonic peaks, than it is observed.

Photons can undergo a gravitational shift also during their way to us if they pass through a perturbation that is evolving with time, so that the opposite shifts when entering and when leaving do not cancel out. This effect is called integrated Sachs-Wolfe (ISW) effect if linear and Rees-Sciama effect if non-linear. The gravitational potential is normally constant during the perturbation evolution but not during cosmological transitions, i.e. from radiation to matter ($z \approx 3000$) and during any dark energy dominated epoch ($z < 1$). We expect therefore two kinds of ISW, an early and a late one. Since the late effect involves perturbations that are very close to us, and therefore occupy large angular scales, it should affect CMB mostly at very large scales (small ℓs).

If at the last scattering epoch there are also gravitational waves (generated along with ordinary scalar perturbations during inflation) then these induce also a kind of tensorial SW effect since they act as a oscillating gravitational potential. Their effect should be much subdominant on the temperature spectrum but might be visible on the polarization spectrum.

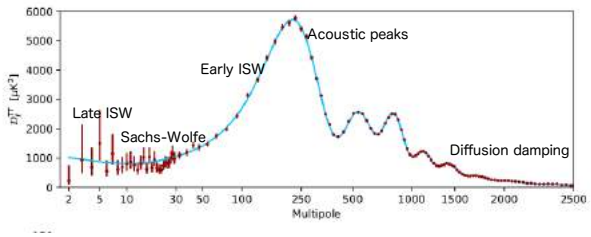


Figure 11.4.1: The different contributions to the CMB spectrum (Planck 2018, ESA and the Planck Collaboration).

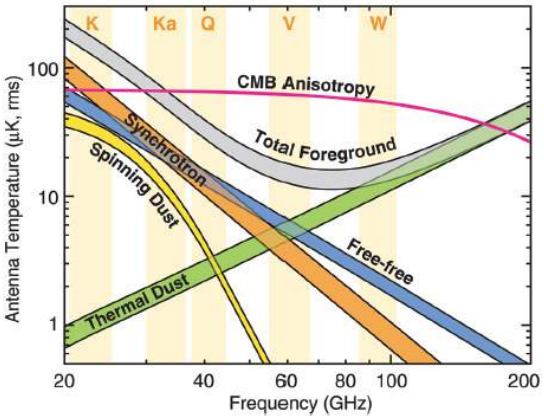


Figure 11.4.2: Foregrounds on the CMB spectrum as a function of wavelength (Bennett, C.L., et.al., 2013, ApJS., 208, 20B, WMAP Science Team).

11.5 The baryon acoustic peaks

The next important physical effect on CMB is due to the pressure oscillations in the coupled photon-baryon plasma before last scattering. The sound speed c_s is close to the relativistic sound speed $c_s^2 = dp/d\rho = 1/3$. Due to the presence of baryons this speed is however modified as

$$c_s^2 = \frac{dp_\gamma}{d(\rho_\gamma + \rho_b)} = \frac{\dot{p}_\gamma}{\dot{\rho}_\gamma + \dot{\rho}_b} = \frac{\frac{4}{3}\rho_\gamma}{4\rho_\gamma + 3\rho_b} = \frac{1}{3}\frac{1}{1 + \frac{3\rho_b}{4\rho_\gamma}} = \frac{1}{3}\frac{1}{1 + \frac{3\Omega_{b0}}{4(1+z)\Omega_{\gamma 0}}} \tag{11.5.1}$$

and becomes time-dependent. For instance, at $z_{rec} = 1100$, since we have $\Omega_{b0} = 0.022h^{-2}, \Omega_{\gamma 0} = 2.5 \cdot 10^{-5}h^{-2}$, one has

$$c_{s,rec} \approx 0.45 \tag{11.5.2}$$

With this approximation, the sound speed angular size $\theta_s \approx c_s\theta_H$ and therefore $\ell_s \approx 200$. The oscillations travel with this speed from big bang (or more exactly from the inflationary epoch) to recombination and therefore add a peak of anisotropies at a fixed scale $c_s t_{rec} \approx 200,000$ly. An exact measurement of this value can find a combination of c_s and $\Omega_K^{(0)}$, as in the approximation (11.4.6). This was first obtained by the Boomerang balloon CMB experiment in 2001. Planck (2015) obtained $\Omega_K^{(0)} = 1 - \Omega_m^{(0)} - \Omega_\Lambda^{(0)} = 0.01 \pm 0.02$.

11.6 The small angular scales

On smaller scales, the anisotropies are smoothed out because of the finite size of the last scattering "surface". That is, the photons that escape do not come all from exactly the same redshift but from a region of thickness $\Delta z \approx 80$. Small fluctuations that fit into this thickness scatter photons with a mixture of red- and blue-shifts, which therefore tend to compensate each other in a poissonian way. One can estimate that the diffusion scale, or Silk damping scale, is

$$\theta_d \approx 6.4 \left(\frac{\Omega_m h^2}{0.15}\right)^{1/4} \left(\frac{\Omega_b h^2}{0.02}\right)^{-1/2} arcmin \tag{11.6.1}$$

The CMB power spectrum will be therefore suppressed below this value, corresponding to $\ell \approx 2000$.

11.7 Reionization and other line-of-sight effects.

After recombination, the universe is mostly composed of neutral hydrogen and helium atoms. However when the first stars become active, at a redshift of a few, the intergalactic medium reionizes due to the star's UV emission (see Chap. 17). The free electrons can now re-scatter the CMB photons, effectively smoothing out the CMB angular spectrum. The Thomson optical depth is defined as

$$\tau = c \int dt n_e \sigma_T \tag{11.7.1}$$

and is an additional parameter that enters the prediction of the CMB spectrum. Current estimates put the reionization epoch at $z \approx 10$.

Photons interact with matter along the line of sight in two other ways. First, they get deviated by the gravitational potentials due to the lensing effect. Second, when they encounter the hot intracluster gas of galaxy clusters they are up-scattered by the electrons (Sunyaev-Zel-dovich effect); this distorts the black-body spectrum and induce an effective additional blue- or red-shift according to the waveband in which it is observed.

11.8 Foregrounds

The CMB photons get mixed with several other components along their ride to Earth. For instance: emission from dust in the solar system (zodiacal light) and from the Milky Way; galaxy bremmstrahlung; synchrotron and free-free emission. Fortunately, all these components can be subtracted off the signal because they either come from well-known regions of sky or because they have a peculiar frequency spectrum. Observing the microwave sky at different wavebands allow then both to minimize their impact and to subtract it (Fig. 11.4.2).

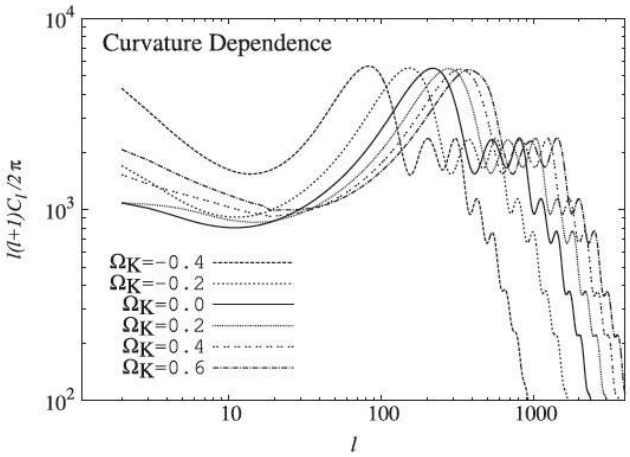


Figure 11.9.1: Effect of the space curvature on the CMB spectrum (from Sugiyama, PTEP 2014)

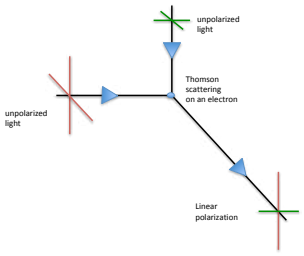


Figure 11.9.2: Generation of polarization on the CMB (inspired to a drawing by Wayne Hu, background.uchicago.edu/~whu/polar/webversion/polar.html).

11.9 Polarization

The scattering of photons through electrons and protons introduce also a polarization. If the photons were homogeneously distributed around the scatterer, the overall polarization would be zero, but since the distribution is not homogeneous, we expect also a net non zero polarization (see Fig. 11.9.2). This has been in fact observed by Planck. The "electric" component of the polarization is tightly related to the temperature anisotropies and is mostly a cross-check of the latter (although additional polarization is induced by lensing). The "magnetic" component can instead be attributed entirely (on large scales) to a very interesting phenomenon, the possible existence of gravitational waves on the last scattering surface, generated during inflation. In 2014 the BICEP experiment announced the detection of this kind of polarization at multipole around $\ell \approx 100$ but subsequent joint BICEP-Planck analysis has shown that so far we can only speak of an upper limit to the tensor-to-scalar ration $r \leq 0.15$ (see definition of r in Ch. 10).

11.10 Boltzmann codes

When the baryons are inserted into the fluctuation equations, the fluid approximation breaks over. The photons and the baryons in fact are coupled via the Thomson cross section; on scales larger than the photon mean free path the full distribution in space and momenta needs to be taken into account (see e.g. Ma & Bertschinger 1995). Publicly available codes (e.g. CMBFAST by Seljak & Zaldarriaga (1995), CAMB by Lewis and collaborators,

CLASS by Lesgourgues and collaborators), do all the numerical computation to a great precision and with high speed. At present, the programs can deal with both flat and curved models, and include as components photons, baryons, cold dark matter, and massless and massive neutrinos, some form of dark energy. The main input parameters are

$$H_0, \Omega_0, \Omega_b, \Omega_\Lambda, n, h, T_{cmb}, N_\nu \tag{11.10.1}$$

A reionization epoch can also be inserted. The codes include corrections for the gravitational lensing and for reionization. The primordial power spectrum can be parametrized by the slope alone, n. Otherwise, a full functional form of the spectrum can be input. The main output is

$$P(k), C_{\ell,s}, C_{l,t} \tag{11.10.2}$$

where the subscript t denotes the tensorial perturbations (gravitational waves). Moreover, the power spectrum can be output at any epoch. The results can be compared to the observations via publicly available MonteCarlo routines.

Appendix: The Sachs-Wolfe effect

Let us now considerin more detail the Sachs-Wolfe effect. A photon emitted from inside a fluctuation has to climb out the gravitational potential, therefore it loses energy and undergoes a gravitational redshift. From GR we know that the change in wavelength is given by

$$\frac{\lambda_{obs}}{\lambda_{em}} = \frac{1}{\sqrt{|g_{00}|}} \tag{11.10.3}$$

In linear perturbation theory, we know that $g_{00} = 1 + 2\Psi$ where $\Psi \ll 1$ and therefore

$$\frac{\lambda_{obs}}{\lambda_{em}} = \frac{1}{\sqrt{|g_{00}|}} \approx \frac{1}{1+\psi} \approx 1 - \Psi \tag{11.10.4}$$

and finally

$$\frac{\Delta\lambda}{\lambda} = -\Psi \tag{11.10.5}$$

Now, if in a black-body distribution the photons experience a small change in frequency $\nu \sim 1/\lambda$, the new black-body distribution will have a change in temperature

$$\frac{\Delta T}{T} = \frac{\Delta\nu}{\nu} = -\frac{\Delta\lambda}{\lambda} = \Psi \tag{11.10.6}$$

(an increase in frequency produces a proportional increase in temperature, as we see for instance from Wien's law, $\nu_{max} = 5.8 \cdot 10^{\frac{GHz}{K}} \cdot T \approx 170$ GHz). Since Ψ is negative for a mass concentration, the temperature is sligthly decreased.

At the same time, since there is more matter in this direction, there will be in general also more radiation. In fact, we can assume the adiabatic relation (8.8.4)

$$\delta_\gamma = \frac{4}{3}\delta_m$$

so that, since $\rho_\gamma \sim T^4$,

$$\frac{\Delta T}{T} = \frac{1}{4}\delta_\gamma = \frac{1}{3}\delta_m$$

Now, at large scale we have seen that δ_m and ψ are roughly constant and linked by the relation (8.5.2) $\delta_m = -2\psi$. Therefore the total *Sachs-Wolfe* effect is

$$\frac{\Delta T}{T} = \Psi + \frac{1}{3}\delta_m = \Psi - \frac{2}{3}\Psi = \frac{1}{3}\Psi$$

We have seen in Sec.3.3 that the gravitational potential for a wavemode k can be written in the Poisson form (we need to use the Poisson-like expression 8.4.37 which is valid at super-horizon scales, so δ here is in fact the combination δ^*)

$$\Psi_k = -\frac{3\mathcal{H}^2}{2}\frac{\delta_k}{k^2} \tag{11.10.7}$$

This gives finally the total Sachs-Wolfe effect

$$\frac{\Delta T_k}{T} = \frac{\psi_k}{3} = -\frac{\mathcal{H}^2}{2}\frac{\delta_k}{k^2} = -\frac{H_0^2}{2}\frac{\delta_{k,0}}{k^2} \tag{11.10.8}$$

In the last equality we used the fact that δ_k grows as $a \sim \tau^2 \sim H^{-2}$, i.e. in MDE: as we already derived, in this epoch the gravitational potential is *constant*. Now we proceed to derive the multipole coefficient of the Sachs-Wolfe fluctuations. Suppose we have a temperature Sachs-Wolfe fluctuation field $\Delta T/T$, and we expand it in plane waves,

$$\frac{\Delta T}{T} = V\int \frac{d^3k}{(2\pi)^3}\frac{\Psi_k}{3}e^{-ikr} = -\frac{H_0^2 V}{2}\int \frac{d^3k}{(2\pi)^3}\frac{\delta_k}{k^2}e^{-ikr} \tag{11.10.9}$$

Remember that in 3D Fourier transforms we always imply $e^{ikr} \equiv e^{ikr\cos\theta_{kr}}$. Here $r = r_{em}$ is the position at which the radiation is emitted. Because we are considering only the SW effect on the last scattering surface, r_{em} is constant in magnitude, and depends only on θ, φ. In fact, the last scattering distance equals the conformal time

$$r_{em} = \int_{t_{em}}^{t_0}\frac{dt}{a(t)} \simeq \int_0^{t_0}\frac{dt}{a(t)} = \tau \tag{11.10.10}$$

The fluctuation field can also be expanded in spherical harmonics, with coefficients

$$a_\ell^m = \int \frac{\Delta T}{T}Y^*_{m\ell}d^2q = -\frac{H_0^2 V}{2}\int \frac{d^3k}{(2\pi)^3}\frac{\delta}{k^2}e^{-ikr}Y^*_{m\ell}(q)d^2q \tag{11.10.11}$$

Since the temperature fluctuation field has zero average (because the mean is estimated from the field itself, and subtracted off), we have that

$$\langle a_\ell^m\rangle = 0$$

while the variance of the multipole coefficient is

$$\begin{aligned} C_\ell &= \langle (a_\ell^m)^2\rangle = \frac{H_0^4V^2}{4}\int \frac{d^3k}{(2\pi)^3}\frac{d^3k'}{(2\pi)^3}\frac{\delta_k\delta_{k'}}{k^2k^{2\prime}}e^{-ikr+ik'r'}Y^*_{m\ell}(q)Y_{m\ell}(q')d^2qd^2q' \\ &= \frac{H_0^4}{4}\int \frac{d^3k}{(2\pi)^3}\frac{P(k)}{k^4}e^{-ikr}e^{ikr'}Y^*_{m\ell}(q)Y_{m\ell}(q')d^2qd^2q' \end{aligned} \tag{11.10.12}$$

where we used the definition of power spectrum given in Eq. (12.4.14)

$$V\langle\delta_k\delta_{k'}\rangle = P(k)\frac{(2\pi)^3}{V}\delta_D(k-k') \tag{11.10.13}$$

Now, the plane waves can be expanded in terms of the spherical harmonics and the spherical Bessel functions as (Abramovitz-Stegun 10.1.47)

$$e^{ikr\cos\theta} = \sum_0^\infty (2\ell+1)i^\ell P_\ell(\cos\theta)j_\ell(kr) \tag{11.10.14}$$

There is an important addition theorem on spherical harmonics (e.g. Arfken & Weber, Eq. 8.189):

$$P_\ell(\cos\theta) = \frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell} Y_{m\ell}\left(\mathbf{q}\right) Y^*_{m\ell}\left(\mathbf{q}'\right) \tag{11.10.15}$$

where $\mathbf{q}\cdot\mathbf{q}'$=cos$\theta$ ($P_\ell = P^0_\ell$ are Legendre polynomials). Moreover, from (11.3.4) and (11.10.15) we derive the useful rule

$$\int P_{\ell'}(\mathbf{q}\cdot\mathbf{q}')Y^*_{m\ell}(\mathbf{q})d^2q = \frac{4\pi}{2\ell+1}Y^*_{m\ell}(\mathbf{q}')\delta_{\ell\ell'} \tag{11.10.16}$$

Therefore, by the rule (11.10.16), we get

$$\int e^{ikr\cos\theta}Y^*_{m\ell}(q)d^2q = 4\pi i^\ell j_\ell(kr)Y^*_{m\ell}(q_0) \tag{11.10.17}$$

where q_0 is here the unit vector $\mathbf{k}/k$.This now can be used as follows (Peebles 1973):

$$\begin{aligned} C_\ell &= \langle(a^m_\ell)^2\rangle = \frac{H_0^4}{4}(4\pi)^2\int\frac{d^3k}{(2\pi)^3}\frac{P(k)}{k^4}j^2_\ell(kr)Y^*_{m\ell}(q_0)Y_{m\ell}(q_0) \\ &= \frac{H_0^4}{4}(4\pi)^2\int\frac{k^2dk}{(2\pi)^3}\frac{P(k)}{k^4}j^2_\ell(kr)Y^*_{m\ell}(q_0)Y_{m\ell}(q_0)d\Omega_k \qquad (11.10.18) \\ &= \frac{H_0^4}{2\pi}\int\frac{P(k)}{k^2}j^2_\ell(kr)dk \qquad (11.10.19) \end{aligned}$$

where we remember that $r=\tau$. In a matter-dominated flat universe, in which $t_0 = 2/3H_0$, we have

$$\tau = \int_0^{t_0}\frac{dt}{a(t)} = t_0\int_0^1 y^{-2/3}dy = 3t_0 = 2/H_0 \tag{11.10.20}$$

Now, the integral can be done analytically for power law $P(k) = Ak^n$, (notice that A has the units of volume$\cdot k^{-n}$, that is of length^{3+n}) where (Abramovitz-Stegun 11.4.41)

$$\begin{aligned} C_\ell &= \frac{H_0^4}{2\pi}\int Ak^{n-2}j^2_\ell(k\tau)dk \\ &= \frac{A2^{-n}H_0^{3+n}}{\pi}\int y^{n-2}j^2_\ell(y)dy \\ &= \frac{AH_0^{n+3}}{16}\frac{\Gamma(3-n)}{\Gamma^2\left[(4-n)/2\right]}\frac{\Gamma\left[(2\ell+n-1)/2\right]}{\Gamma\left[(2\ell+5-n)/2\right]} \qquad (11.10.21) \end{aligned}$$

where $\Gamma(x)$ is the Gamma function. A fit to observations at small ℓ gives therefore directly A and n. For the latter, the COBE data give roughly $n = 1.2\pm0.3$. For the typical inflationary case, $n=1$, we get

$$C_\ell = \frac{AH_0^4}{16}\frac{\Gamma(2)}{\Gamma^2\left[3/2\right]}\frac{\Gamma\left[\ell\right]}{\Gamma\left[\ell+2\right]} = \frac{AH_0^4}{4\pi\ell\left(\ell+1\right)} \tag{11.10.22}$$

Notice then that $\ell\left(\ell+1\right)C_\ell$ is constant in this case. It is customary to express then the results of CMB experiments in terms of $\ell\left(\ell+1\right)C_\ell$. This will show at once the SW behavior at small multipoles. Often is employed the quantity

$$\left(C^*_\ell\right)^{1/2} \equiv T_{CMB}\left[\ell\left(\ell+1\right)C_\ell/2\pi\right]^{1/2}$$

expressed in μK. To make an estimate, let us remember that the peak of the spectrum at $k_{pk}\simeq 0.05$ is roughly P_{pk} =2 $\cdot 10^4\left(Mpc/h\right)^3$ so that $A = P_{pk}/k_{pk}$. This gives a typical value for the quadrupole fluctuations if $n=1$ of

$$\left(C^*_2\right)^{1/2} = 2\cdot10^{-5}\mathrm{K} = 20\mu\mathrm{K} \tag{11.10.23}$$

which is in fact close to the observed values.

Chapter 12

The galaxy power spectrum

In Ch. (8) we have seen how to derive the linear matter power spectrum as a function of scale and redshift. What we observe however is the power spectrum of the linear and non-linear galaxy distribution when distances are measured with the redshift. Its relation to the theoretical prediction is not a straightforward one, as we will see in this Chapter.

Quick summary

- The linear galaxy power spectrum expresses the clustering of galaxies in redshift space
- It is related to theoretical prediction of the matter power spectrum in real space through several corrections (the bias, the redshift distortion) and depends on the growth function
- On small scales, the non-linear effects should also be taken into account
- The baryon oscillations that are prominently visible on the CMB spectrum are also visible as small wiggles on the galaxy power spectrum
- The comparison to observations can constrain several cosmological parameters, from the primordial slope n_s to Ω_m, Ω_b, to the dark energy equation of state.
- The Euclid satellite (launch 2021) will measure the power spectrum up to redshift 2 to great precision.

12.1 Large scale structure

Let us see now how to link the primordial inflationary spectrum with the present observations of the galaxy distribution. Instead of the spectrum evaluated at horizon reenter t_H, which is different for every scale, we prefer to evaluate it at a fixed epoch t_F, for instance the decoupling epoch $z \approx 1000$. The difference is that at a fixed epoch, the perturbations that are already inside the horizon had the time to grow, contrary to those still outside. Let us consider a perturbation that reenter in MDE when the scale factor was a, and $H = (2/3)t^{-1} = a^{-3/2}H_0$ (as usual we assume the present value $a_0 = 1$) ; at that epoch, $k = aH = aa^{-3/2}H_0 = a^{-1/2}H_0$. Thus a perturbation with wavenumber k reenters when the scale factor was $a = (k/H_0)^{-2}$. This perturbation grows between a and an arbitrary instant, say a_F, as $\delta_k \sim (a_F/a) = (k/k_F)^2$ (since we are in MDE), if k_F is the scale that reenters at a_F. Therefore smaller scales reenter the horizon earlier than larger ones and have therefore more time to grow in amplitude. To conclude, the amplitude of a perturbation of size k^{-1} has the time to grow between its reenter and a_F by a factor $(k/k_F)^2$. Then the relation between δ at t_F and at t_H is

$$(\delta\rho/\rho)_k(t_F) \quad = \quad (k/k_F)^2(\delta\rho/\rho)_k(t_H)\,, \tag{12.1.1}$$

and consequently

$$P_k(t_F) = (k/k_F)^4 P_k(t_H)\,.$$

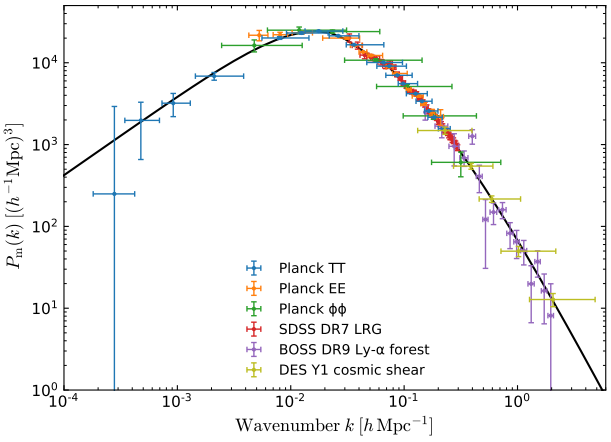


Figure 12.1.1: Power spectrum of the large scale structure (at small scales) and of the cosmic microwave background (at large scales). (ESA and the Planck Collaboration).

Assuming the inflationary spectrum k^{-3} of Eq. (10.2.3) with $n_s \approx 1$ it follows that for the scales reentering after equivalence

$$P_k(t_F) = Ak\,. \qquad \text{Harrison} - \text{Zeldovich spectrum} \tag{12.1.2}$$

This scale-invariant spectrum is a remarkable prediction of inflation. During the subsequent evolution, this initial spectrum will be modified in a way that depends on the detailed components, e.g. the amount of dark matter and baryons. Note that we obtained (12.1.2) assuming exponential inflation. If inflation is a power law, the resulting spectrum becomes $P_k = Ak^{n_s}$, with $n_s \leq 1$. The Planck data of the cosmic background found $n_s = 0.96 \pm 0.01$, in complete agreement with inflation.

We can now predict the observational quantities: the power spectrum and (in the next section) the peculiar velocity field. Notice that the average of the density contrast vanishes, $< \delta >= 0$, by definition of average density. The simplest non-trivial average quantity that describe the fluctuation field is therefore the dispersion δ, that is the variance $< \delta^2 >$. In the Fourier space, the corresponding quantity is the spectrum $P(k) = \delta_k^2$. We have seen that the spectrum at horizon crossing goes like k^{-3}, modified in k^1 for the perturbations that reenter in MDE. However, those that are smaller than λ_{eq} =13 Mpc h^{-1}, reentering in RDE, do not suffer this correction since as we have seen they do not grow during RDE. In other words, for scales smaller than λ_{eq}, the spectrum at horizon crossing coincides with the one at equivalence. For these scales then the behavior remains k^{-3}. All this means that the inflationary spectrum at a fixed t, $P(k) = Ak^{n_s}$ (with n_s close to 1) is modified during RDE and MDE by a function $T^2(k)$ such that $T \sim 1$ at large scales and $T \sim k^{-4}$ at small ones. For the final spectrum we have then

$$P(k) = Ak^{n_s}T^2(k) \tag{12.1.3}$$

where $T(k)$ is called *transfer function*. This spectrum remains unvaried in slope from equivalence to now: during this epoch, in fact, perturbations grow independently of k. A simple form of T^2 that produces the requested

behavior is $T^2 = 1/[1 + (k/k_{eq})^4]$. The exact transfer function can be evaluated only numerically, integrating for every k the perturbation equations. A popular approximation is (Efstathiou, Bond White 1992)

$$P(k) = Ak^{n_s}T(k)^2 \tag{12.1.4}$$

$$T(k) = \left[1 + \left[ak + (bk)^{1.5} + (ck)^2\right]^\nu\right]^{-1/\nu} \tag{12.1.5}$$

$$(a, b, c) = (6.4, 3.0, 1.7)\,\Gamma^{-1}Mpc/h, \nu = 1.13 \tag{12.1.6}$$

$$\Gamma = \Omega_{nr}h \tag{12.1.7}$$

where $T^2(k)$ has the correct limit k^{-4} . The scale at which this limit is approached is $k_t = \Gamma = \Omega_{nr}h$. In a flat model with $h = 0.5$ this is roughly $2\pi/0.5$ Mpc$/h = 12.6$ Mpc$/h$, close to the horizon scale at equivalence, as it should be.

The whole linear treatment fail naturally at very small scales, those that are non-linear and collapsed into galaxies and clusters. A rough approximation to the linearity scale is 10 Mpch^{-1}.

12.2 The bias factor

The power spectrum we observe is obtained by mapping the position of the galaxies and then performing a Fourier transform of the density contrast. However the power spectrum that is predicted by the linear perturbation theory is the mass power spectrum, not the galaxy one. The density contrast of the galaxies can be well different from the matter density contrast: for instance, galaxies could form only when the matter density is above a certain threshold. The simplest possibility is to assume that the two density contrast are proportional to each other

$$\delta_g = b\delta_m \tag{12.2.1}$$

where b is called linear *bias* factor. In this case obviously

$$P_g = b^2 P_m \tag{12.2.2}$$

Of course the real bias could be a non-linear function of δ_m, or an average of δ_m at various locations. In general, the bias factor might depend on space (i.e. on k after Fourier transformation) and time, and also on galaxy type and luminosity. In fact it is often reported that brighter galaxies have a larger bias, perhaps because they form only on density peaks. The values of b that have been reported so far are always around unity.

12.3 Normalization of the power spectrum

Several ways to normalize the *mass* power spectrum have been employed so far: the cosmic microwave fluctuations, the cluster abundance, weak lensing.

As we have seen the CMB measures at large scales the Sachs-Wolfe effect, i.e. temperature anisotropies proportional to the gravitational potential. The amount of anisotropies is then a direct measurement of the power spectrum at large scales. If we know the shape of the transfer function, measuring the spectrum at large scales fixes its amplitude at all scales and we can estimate the normalization (see Sect. 9.8)

$$\sigma_8^2 = (2\pi^2)^{-1} \int P(k) W_{8Mpc/h}^2(k) k^2 dk \tag{12.3.1}$$

Planck (2015) result assuming ΛCDM is

$$\sigma_8 = 0.834 \pm 0.03 \tag{12.3.2}$$

Cluster abundance, along with X-ray mass-temperature relation, is another tool for estimating σ_8. In this case σ_8 is strongly degenerate with Ω_m but using also the cluster correlation function one can reduce the level of degeneracy and measure separately the two parameters. The X-ray survey REFLEX obtained in 2002 the value

$$\sigma_8 = 0.711 \pm 0.04 \tag{12.3.3}$$

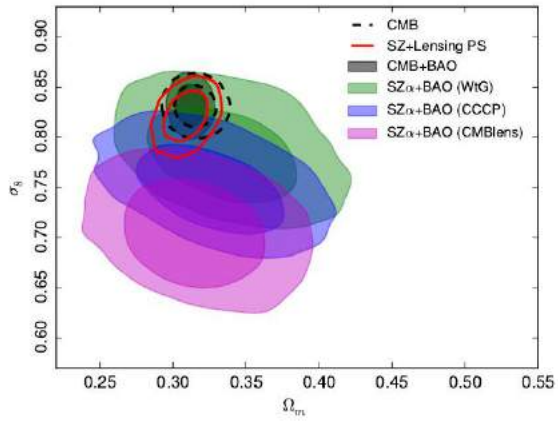


Figure 12.3.1: Planck (2015) constraints on Ω_m, σ_8 from SZ clusters and from CMB (Ade et al. A&A 594, A13, 2016; ESA and the Planck Collaboration).

Clusters detected with the SZ effect by Planck can also give an estimate of σ_8 (see Fig. 12.3.1).

Finally, in the next chapter we will see that weak lensing can also estimate the amount of clustering regardless of the galaxy bias, and recent results provide

$$\sigma_8(\Omega_m/0.3)^{0.7} = 0.64 \pm 0.1 \tag{12.3.4}$$

12.4 The peculiar velocity field

The mass power spectrum can be studied also by analyzing the peculiar motion of the galaxies. It is intuitive, in fact, that a more clustered distribution of matter will induce stronger peculiar velocities. The importance of this is that the velocity field depends on the *total* mass distribution, including any unseen component. Let us start from Eq. (8.6.4). In terms of the time t it becomes

$$a\frac{d\delta}{dt} = -ik^i v_i \tag{12.4.1}$$

Consider now Eq. (8.4.26) for $c_s = 0$:

$$\dot{v}^i = -\mathcal{H}v^i - a^2 ik^i \psi \tag{12.4.2}$$

Expressing the velocity vector by a component parallel and a component orthogonal to the potential gradient (the peculiar acceleration), we see that the orthogonal component (subscript t) obeys the equation

$$\dot{v}_t^i = -\mathcal{H}v_t^i \tag{12.4.3}$$

so that it will decay as a^{-1}. Neglecting therefore this purely rotational component, we can look for solutions of (12.4.1) in the form $v^i = F(k)k^i$. This gives immediately from (12.4.1) the relation between the peculiar velocity field and the density fluctuations in linear perturbation theory, in the Newtonian regime:

$$v^i = iHa\delta f\frac{k^i}{k^2} \tag{12.4.4}$$

where

$$f = \frac{a}{\delta}\frac{d\delta}{da} \tag{12.4.5}$$

is a function that expresses the growth rate of the fluctuations; for a flat universe in MDE we have seen that $\delta \sim a$ and $f = 1$. A good approximation for a model with matter parameter $\Omega_m(t)$, as we have seen in Sec. 8.8 , is

$$f = \Omega_m^{\gamma} \tag{12.4.6}$$

with $\gamma \approx 0.55$ and, for ΛCDM,

$$\Omega_m(z) = \frac{\Omega_{m0}(1+z)^3}{\Omega_{m0}(1+z)^3 + 1 - \Omega_{m0}} \tag{12.4.7}$$

Let us consider now the present epoch. We put then $a = a_0 = 1$ and $H = H_0$,

$$\mathbf{v} = iH_0 f \delta_k \frac{\mathbf{k}}{k^2} \tag{12.4.8}$$

The peculiar velocity at location x is

$$\mathbf{v}(x) = iH_0 f \frac{V}{(2\pi)^3} \int \delta_k \frac{\mathbf{k}}{k^2} e^{ikr} d^3k \tag{12.4.9}$$

and the average in a volume of radius R is

$$\mathbf{v}_R = \frac{1}{V_R} \int \mathbf{v} d^3r = iH_0 f \frac{V}{(2\pi)^3 V_R} \int \delta_k \frac{\mathbf{k}}{k^2} e^{ikr} W(r) d^3k d^3r = iH_0 f \frac{V}{(2\pi)^3} \int \delta_k \frac{\mathbf{k}}{k^2} W(kR) d^3k \tag{12.4.10}$$

where $W(kR)$ is the Fourier transform of the *window function*, defined as

$$W(kR) = \frac{1}{V_R} \int W(x) e^{ikx} d^3x$$

Therefore, the average of the square of the velocity is

$$\begin{aligned}
\langle v^2 \rangle_R &= H_0^2 f^2 \frac{V^2}{(2\pi)^6} \int \delta_k \delta_{k'} \frac{\mathbf{k}}{k^2} \frac{\mathbf{k}'}{k'^2} W(k'R) W(kR) d^3k d^3k' && (12.4.11)\\
&= \frac{H_0^2 f^2}{(2\pi)^3} \int P(k) \delta_D(k-k') \frac{\mathbf{k}}{k^2} \frac{\mathbf{k}'}{k'^2} W(k'R) W(kR) d^3k d^3k' && (12.4.12)\\
&= \frac{H_0^2 f^2}{2\pi^2} \int P(k) W^2(kR) dk && (12.4.13)
\end{aligned}$$

where in the last integral we integrated over the solid angle 4π and we used the definitions (see 9.8)

$$\begin{aligned}
V < \delta_k \delta_{k'} > &= \frac{(2\pi)^3}{V} P(k) \delta_D(k-k')\\
\int \delta_D(k-k') d^3k &= 1 && (12.4.14)
\end{aligned}$$

The square root of $< v^2 >_R$ is called *bulk flow*, that is the magnitude of the peculiar flow on the scale R. Estimates of the bulk flow can be used then to constrain or normalize the mass power spectrum (see e.g. Lauer & Postman 1996). On the other hand, independent measures of $P(k)$ and $\langle v^2 \rangle_R$ can give $f^2 \simeq \Omega_0^{1.2}$ and thus Ω_0. However, since what we can measure is actually the galaxy power spectrum, assuming a simple biasing $P_g(k) = b^2 P(k)$ we can constrain only the combination $\beta = \Omega_m^{0.6}/b$.

12.5 The redshift distortion

The galaxy distances we measure are mostly obtained through their redshift. The redshift however includes the peculiar velocity of the galaxies themselves, so that there is an error in the distances we assign to galaxies. On very small scales the peculiar velocity of a galaxy is more or less randomly oriented, so that the error in

the distance is statistical, and can be taken into account along the experimental errors. On redshift maps, the small scale peculiar velocities cause the finger-of-god effect: galaxies in a cluster acquire an additional random redshift that distorts the cluster distribution, making it appear elongated along the line of sight.

On large scales, however, the galaxies tend to fall toward concentrations, so that the velocity field is coupled to the density field. This correction is systematic, and can be accounted for in the following way.

A source at distance r with peculiar velocity along the line of sight u

$$u = \mathbf{v} \cdot \frac{\mathbf{r}}{r} \tag{12.5.1}$$

will be assigned a distance $s = r + u(r) - u(0)$ if distances are expressed in km/sec through multiplication for the Hubble constant (so that the Hubble law is simply $v = r$). Consider then a coordinate transformation from real space (subscript r) to redshift space (subscript s):

$$\mathbf{s} = \mathbf{r}\left[1 + \frac{u(r) - u(0)}{r}\right] \tag{12.5.2}$$

Then, if dV_s and dV_r are the volume elements in the two coordinates, we can write

$$n(r)dV_r = n(s)dV_s \tag{12.5.3}$$

where the volume element dV_s can be written in terms of the r coordinates as

$$dV_s = s^2 ds d\cos\theta d\phi = r^2\left(1 + \frac{\Delta u(r)}{r}\right)^2 |J| dr d\cos\theta d\phi = \left(1 + \frac{\Delta u(r)}{r}\right)^2 |J| dV_r \tag{12.5.4}$$

where the Jacobian is

$$|J| = \left|\frac{\partial s}{\partial r}\right| = 1 + \frac{du}{dr} \tag{12.5.5}$$

Then we have that the density contrast in s-space is

$$\delta_s = \frac{n(s)dV_s}{n_0 dV_s} - 1 = \frac{n(r)dV_r}{n_0 dV_r\left(1 + \frac{\Delta \mathbf{u}(r)}{r}\right)^2 |J|} - 1 \tag{12.5.6}$$

where n_0 is the average density. To first order, this is

$$\begin{aligned} \delta_s &= \frac{n(r)}{n_0}(1 - 2\frac{\Delta u(r)}{r} - \frac{du}{dr}) - 1 \\ &= \left[\frac{n(r)}{n_0} - 1\right] - \frac{n(r)}{n_0}\left[2\frac{\Delta u(r)}{r} + \frac{du}{dr}\right] \\ &= \delta_r - 2\frac{\Delta u(r)}{r} - \frac{du}{dr} \end{aligned} \tag{12.5.7}$$

where in the last line we used the fact that to first order we can approximate $n(r)$ with n_0. Therefore, we see that the density contrast will be different in the two spaces. As a consequence, the correlation function and the power spectrum measured in redshift space will have to be corrected to be expressed in real space. To do so, we have to take the velocity field from the linear perturbation theory.

Eq. (12.4.9) is clearly in real space:

$$\mathbf{v} = H_0 f i \int \delta_k e^{i\mathbf{k}\cdot\mathbf{r}} \frac{\mathbf{k}}{k^2} d^3k^* \tag{12.5.8}$$

where, to simplify notation, the Fourier factor $V/(2\pi)^3$ is included in the differential d^3k^*. Its line-of-sight component, measuring distances in units of H_0, is

$$u(r) = \frac{\mathbf{r}}{r} \cdot \mathbf{v} = f i \int \delta_k e^{i\mathbf{k}\cdot\mathbf{r}} \frac{\mathbf{kr}}{k^2 r} d^3k^* \tag{12.5.9}$$

while its derivative is (notice that $\mathbf{k}\cdot\mathbf{r}/(kr) = \mu$ does not depend on r)

$$\frac{du}{dr} = -f\int \delta_k e^{i\mathbf{k}\cdot\mathbf{r}}\left(\frac{\mathbf{kr}}{kr}\right)^2 d^3k^* \tag{12.5.10}$$

where we used the relation

$$\frac{d}{dr}e^{i\mathbf{k}\cdot\mathbf{r}} = i\frac{\mathbf{k}\cdot\mathbf{r}}{r}e^{ikr} \tag{12.5.11}$$

Finally, we have

$$\delta_s = \delta_r - \frac{du}{dr} = \delta_r + f\int \delta_k e^{i\mathbf{kr}}\left(\frac{\mathbf{kr}}{kr}\right)^2 d^3k^* = \delta_r + f\int \delta_k e^{i\mathbf{kr}}\mu^2 d^3k^* \tag{12.5.12}$$

where $\mu = \mathbf{kr}/(kr)$ and we neglected the second term in (12.5.7) because it is negligible for large r. It is useful now to notice that the density fluctuation in the second term r.h.s. is the *mass* density fluctuation, responsible for the velocity field, while the other fluctuation terms can refer to the number density of any class of sources, e.g. galaxies. If the mass fluctuations are b times smaller that the galaxy fluctuations, we could write inside the integral at r.h.s. δ_k/b instead of δ_k. Then, the relation holds for any class of objects provided we use $\beta = f/b$ in place of f in the final result. The Fourier transform of this relation is

$$\delta_{sk} = \delta_{rk} + \beta\int \delta_{rk'} I(k,k')d^3k' \tag{12.5.13}$$

$$I(k,k') = (2\pi)^{-3}\int e^{i(k-k')r}\mu^2 d^3r \tag{12.5.14}$$

The redshift distortion then introduces a mode-mode coupling. This coupling can be broken if we can assume that the cosine

$$\mu = \frac{\mathbf{kr}}{kr} \tag{12.5.15}$$

varies much slower with r than the fast oscillating $e^{i(k-k')r}$ term, i.e. if the survey covers a small angular size and the direction of line-of-sight $\mathbf{r}/r$ is almost constant. In this case the μ^2 term goes out of the integral and we have then $I = \mu^2\delta_D(k-k')$, so finally

$$\delta_{sk} = \delta_{rk}(1+\beta\mu^2) \tag{12.5.16}$$

Notice that these are the three-dimensional Fourier coefficients. The power spectrum now is

$$P_s(k) = V\delta_{rk}^2(1+\beta\mu^2)^2 = P_r(k)(1+\beta\mu^2)^2 \tag{12.5.17}$$

If we average it over angles we get

$$P_s(k) = P_r(k)(1+2\beta\langle\mu^2\rangle + \beta^2\langle\mu^4\rangle) \tag{12.5.18}$$

where

$$\langle\mu^2\rangle = \frac{1}{2}\int_{-1}^{1}\cos^2\theta' d\cos\theta' = 1/3$$

$$\langle\mu^4\rangle = \frac{1}{2}\int\cos^4\theta' d\cos\theta' = 1/5 \tag{12.5.19}$$

Finally we obtain for the μ-averaged spectrum

$$P_s(k) = P_r(k)(1+2\beta/3+\beta^2/5) \tag{12.5.20}$$

The power spectrum is then boosted in redshift space, because the velocity field is directed toward mass concentrations: as a result, galaxies seem more concentrated when seen in redshift space.

At very small scales, on the other hand, the velocity orientation can be assumed to be random. The variance of **s** will be larger than the variance of **r** along the line of sight and unchanged across it: the sphere will appear pointing towards us. This is the non-linear redshift distortion, also called "fingers-of-God". The power spectrum is therefore decreased. Empirical studies have shown that a sufficiently good approximation is given by an exponential damping factor,

$$P_s(k,\mu) \quad = \quad P_r(k)(1+\beta\mu^2)^2 e^{-k^2\mu^2\sigma_v^2} \tag{12.5.21}$$

where σ_v is the cloud velocity dispersion along the line of sight in units of H_0; one has typically $\sigma_v \approx (300km/sec)/H_0 \approx 3Mpc/h$.

The final result taking into account redshift distortion (RSD), bias, growth $G(z)$ (see Eq. 8.9.4) and velocity dispersion is then rather simple:

$$P_s(k,\mu,z) = (1+\beta\mu^2)^2 b^2 G^2 P_r(k) e^{-k^2\mu^2\sigma_v^2} \tag{12.5.22}$$

where f, the growth rate and b, G all depend on z and perhaps on k, and $P_r(k)$ is today's matter power spectrum. By measuring the anisotropy combination

$$\sqrt{\frac{P_s(k,\mu=1,z)}{P_s(k,\mu=0,z)}} - 1 = \frac{f}{b} \tag{12.5.23}$$

we can then obtain information on the combination $\beta = f/b$. By measuring instead

$$\sqrt{\frac{P_s(k,1,z) - P_s(k,0,z)}{P_s(k,1,z') - P_s(k,0,z')}} = \frac{f(z)G(z)}{f(z')G(z')} \tag{12.5.24}$$

we obtain information on the combination

$$fG = \frac{\delta'}{\delta(0)} \tag{12.5.25}$$

Often the combination $f\sigma_8(z) \equiv f\sigma_8 G$ is said to be directly measurable via the redshift distortion, but this assumes one has a specific model for $P_r(k)$, e.g. ΛCDM.

12.6 Baryon acoustic oscillations

The spectrum of fluctuations at linear scales we observe today is a combination of the fluctuations in the dark matter and in the baryonic content. As we have seen in the CMB chapter, the baryonic component was tightly coupled to radiation before recombination and underwent Jeans oscillations on scales below the sound horizon, roughly 100 Mpc/h in comoving distance (i.e. the size they will have been expanded to by today). When the baryons finally decouple from radiation, the oscillations remain imprinted on their distribution and, in the linear regime, show up as an additional feature on the smooth power spectrum proportional to the amount of baryons, called *baryon acoustic oscillations* (BAO). Since the amount of baryons and radiation are well known, the BAO scale is essentially fixed: it constitutes then a *standard rod*.

In the correlation function, we expect therefore a local peak at roughly the same comoving scale as found in Sec. 11.5

$$d_s^c = \frac{c_s}{c} d_H^c \approx 144\text{Mpc} \tag{12.6.1}$$

(as measured by Planck in 2015).

The correlation function is on average a isotropic function of distance. When plotted as a function of the distance along and across the line of sight the baryonic peak should then appear as a circular ring around the origin with radius d_s^c. The correlation function for galaxies in a shell around redshift z_1 is measured in redshift and angular space and the peak will therefore appear as a peak in the z, θ plane. The radius in the θ direction gives the comoving angular diameter distance $d_A^c = (1+z)d_A$ of that shell

$$d_A^c(z) = \theta d_s^c \tag{12.6.2}$$

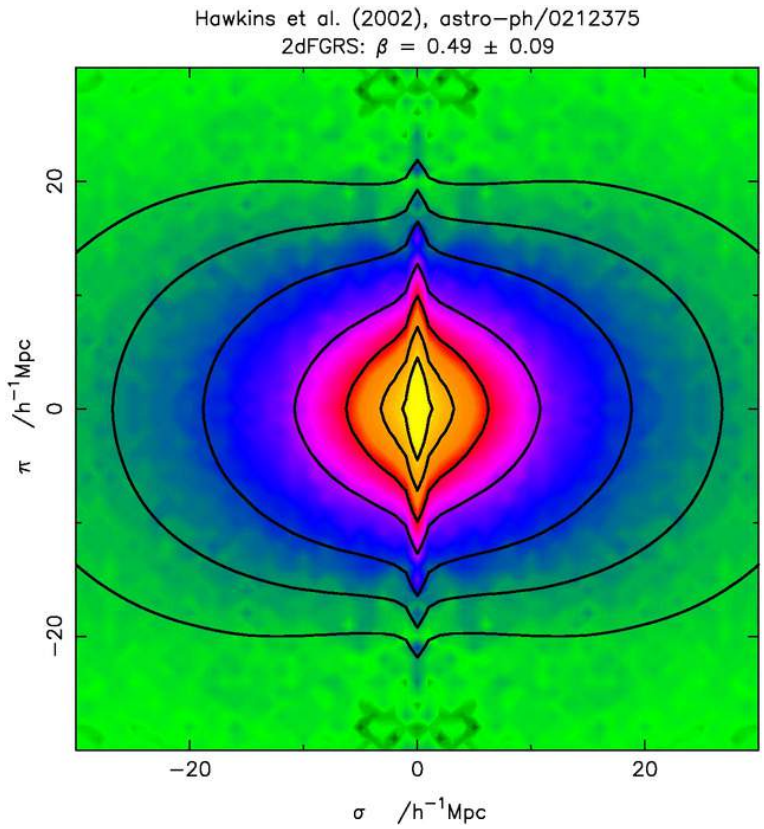


Figure 12.5.1: Correlation function versus radial (line of sight) and longitudinal (orthogonal to line of sight) coordinates π, σ, respectively. Notice the elongated "finger-of-God" feature along the radial coordinate and the squashed form along the longitudinal one (linear redshift distortion). (Hawkins et al. 2003MNRAS.346...78H).

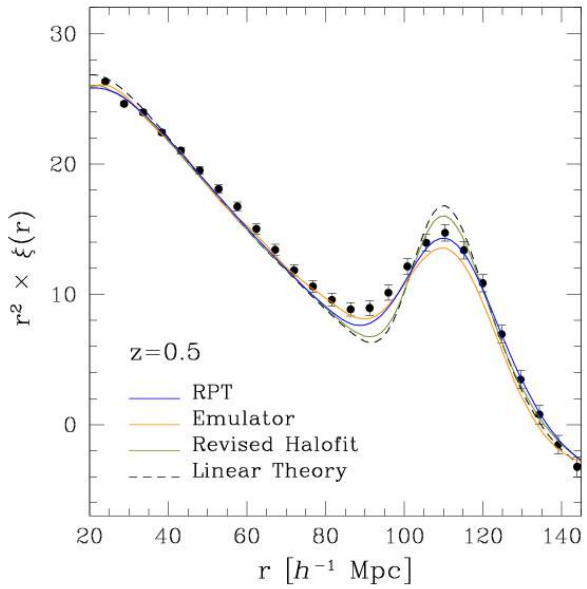


Figure 12.6.1: Baryon acoustic peak in a correlation function obtained from MICE, a very large N-body simulation. From Fosalba et al. MNRAS 448 (2015), 2987-3000 arXiv:1312.1707.

By estimating the redshift difference Δz between the two opposite points in the ring one measures instead the difference of comoving distances along the line of sight:

$$\Delta d_A^c = \Delta r = \int_0^{z_1+\Delta z} \frac{dz}{H(z)} - \int_0^{z_1} \frac{dz}{H(z)} \approx \frac{\Delta z}{H(z)}|_{z_1} \tag{12.6.3}$$

In practice of course the entire ring is fitted by varying the cosmological parameters. Since there are two directions orthogonal to the line of sight and one along it, one can define a combined distance

$$D_V(z) \equiv [\frac{(d_A^c)^2}{H(z)}\Delta z]^{1/3} \tag{12.6.4}$$

and express the results in terms of D_V. Current measurements are however good enough to measure separately $d_A^c(z)$ and $1/H(z)$.

The peak in the correlation function manifests itself also in the power spectrum as the Fourier transform of a peak, namely oscillations at a wavenumber $k_{BAO} \approx 2\pi/d_A^c$ and multiples, appropriately called baryon acoustic oscillations (BAO), overimposed to the smooth spectrum, with an amplitude proportional to Ω_b.

The baryonic oscillations have been detected for the first time in 2005 with SDSS and 2dFGRS data, obtaining

$$\frac{r_s}{D_V(z=0.35)} \approx 0.1 \pm 0.003 \tag{12.6.5}$$

Subsequent analyses measured the BAO scale to a precision of 2-4%, with results always consistent with concordance cosmology.

To go below this precision one needs to take into account small effects that tend to "distort the ring" and smear out the oscillations: the redshift measurement error, the peculiar velocity redshift, the non-linear correction to the spectrum. Of course when estimating the cosmological parameters from the power spectrum the entire shape of the spectrum has to be taken into account, not just the BAO wiggles.

12.7 Non-linear correction

All we have seen so far is only valid at linear scales where $\delta_m \ll 1$, i.e. at scales larger than 10 Mpc or so (so in fact the k^{-3} slope is never exactly reached). To measure smaller scales one has two ways: either estimate the spectrum at high redshift or find the non-linear correction.

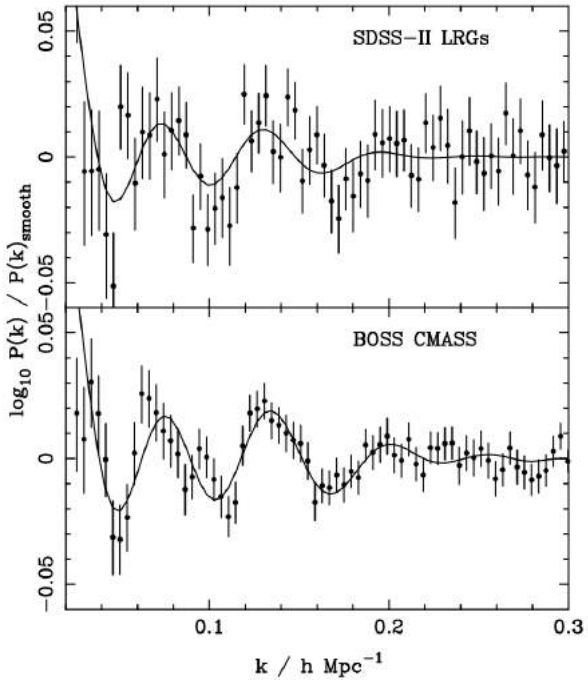


Figure 12.6.2: Baryon acoustic peak observed in the power spectrum of SDSS and BOSS galaxies. The smooth part of the spectrum has been subtracted. From Anderson et al. 2012, arXiv:1203.6594, Mon. Not. R. Astron. Soc. 427, 3435–3467 (2012).

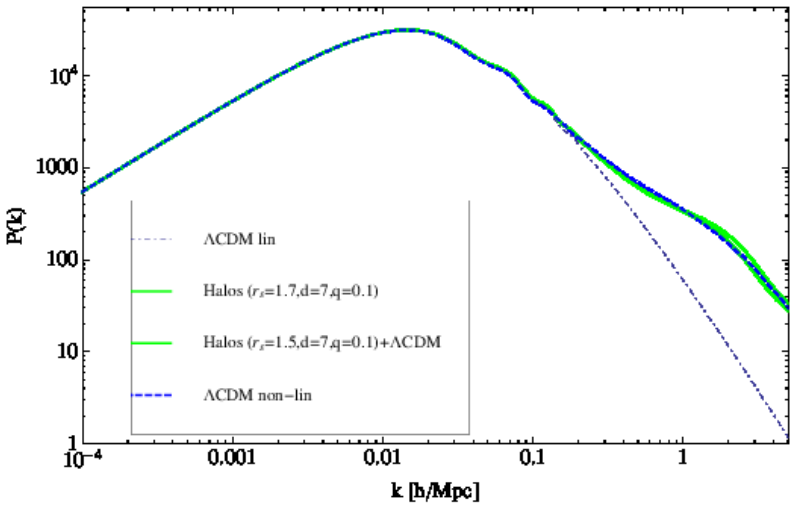


Figure 12.6.3: Matter power spectrum in linear ΛCDM (black dotted line) and including a simplified halo non-linear correction ($q = r_s/R$ and $n = d^3$, green lines), compared with a numerical non-linear spectrum (thick blue dashed line).

At redshift zero the ΛCDM spectrum normalization σ equals 0.8 when averaged over spherical cells of radius $R \approx 8$ Mpc/h. At higher redshifts the value of σ_8 decreases proportionally to the growth function G. Since $f = d\log G/d\log a$, and putting $f \approx \Omega_m(a)^\gamma$one has $\sigma_8(z) = G(z)\sigma_8(0)$ where

$$G(z) = \exp\int_1^a f(a')\frac{da'}{a'} = \exp\int_1^a \Omega_m^\gamma(a')\frac{da'}{a'} \tag{12.7.1}$$

For instance, assuming $\Omega_{m0} = 0.3$, one has $\sigma_8(1) \approx 0.6$ and $\sigma_8(3) \approx 0.3$, and therefore the scale of non-linearity moves to smaller and smaller scales. Intergalactic lumps of neutral hydrogen called Lyman$-\alpha$ clouds along the line of sight of distant quasars absorb part of the quasar radiation and due to different redshifts appear as a "forest" of Lyman$-\alpha$ lines on the quasar's spectrum (Fig. 12.7.1). Their power spectrum can be calculated up to redshifts of order 4 or even more and provide a window on the linear high$-k$ tail that is otherwise hidden in non-linearities for nearby galaxies.

On smaller scales one should consider non-linear corrections that generally push the power up by even one order of magnitude. One rough way of estimating the non-linear correction is to imagine that at small scales the universe can be seen as a random collection of identical spherical halos with, for instance, the Navarro-Frenk-White (NFW) profile

$$\rho_{NFW} = \frac{\rho_0}{\frac{r}{r_s}(1+\frac{r}{r_s})^2} \tag{12.7.2}$$

where r_s is a free scale parameter that has to be fit to each halo. This functional form has been found to be a very good fit to the profiles in N-body simulations. Then one should evaluate the Fourier transform of a single halo density contrast in a radius R and volume $V = 4\pi R^3/3$,

$$\delta_1 = \frac{4\pi}{V}\int_0^R \left(\frac{\rho_{NFW}}{\bar\rho} - 1\right)\frac{\sin(kr)}{(kr)}r^2 dr \tag{12.7.3}$$

$$= \frac{1}{V}[\frac{4\pi}{\bar\rho}\int_0^R \rho_{NFW}\frac{\sin(kr)}{(kr)}r^2 dr - \frac{4\pi}{\bar\rho}\int_0^R \frac{\sin(kr)}{(kr)}r^2 dr] \tag{12.7.4}$$

$$= \frac{1}{V}(W_{NFW} - W_{TH}) \tag{12.7.5}$$

where $\bar\rho$ is the average density

$$\bar\rho = \frac{4\pi}{V}\int_0^R \rho_{NFW} r^2 dr \tag{12.7.6}$$

and then obtain the halo power spectrum as the sum of many random (uncorrelated) halos

$$P_h(n, R, r_s) = NV\delta_1^2 = n(W_{NFW} - W_{TH})^2 \tag{12.7.7}$$

where n is the halo number density. This can be considered a correction to be added to the linear spectrum

$$P_{NL} = P_{LIN} + P_h \tag{12.7.8}$$

Some results for P_{NL} are in Fig. (12.6.3). This particular halo correction is very naive and depends on three free parameters r_s, R, n; better non-linear schemes are based on fitting N-body simulations or on higher-order perturbation theory. We discuss them again in Chap. (14).

12.8 The Euclid satellite

In 2021 the European Space Agency (with participation from NASA) will launch Euclid, a satellite dedicated to the study of large scale structure through, primarily, galaxy clustering and weak lensing. The Euclid collaboration involves 15 countries and over a thousand scientists. The satellite should stay in orbit for five years and scan roughly 15,000 square degrees of sky, that is almost all the sky safely far from the Milky Way disk.

Euclid will have on board a 1,2m telescope with two detectors: a imager and a spectrograph. The imager will collect the images of two billion galaxies up to redshift 3, while the spectrograph will find the redshift of roughly

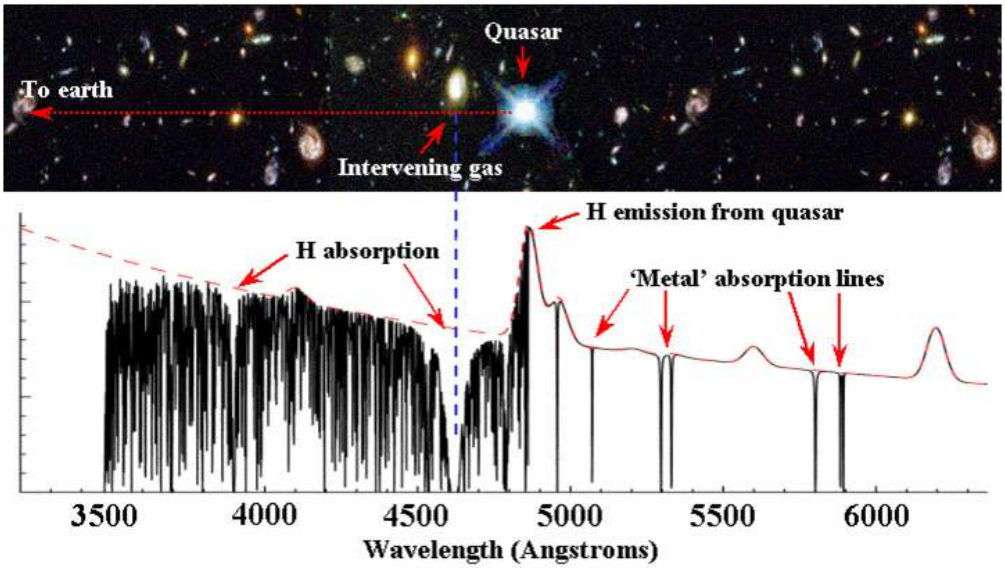


Figure 12.7.1: Schematic representation of the spectrum of a distant quasar. On the left of the H emission from the quasar itself one sees the forest of Lyman-α absorption lines from clouds on the line of sight at various redshifts. (From `www.nat.vu.nl/~wimu/FundConst-Notes.html`, courtesy of Wim Ubachs).

50 million galaxies. The images will be scanned to extract the ellipticity and estimate the shear weak lensing power spectrum. The galaxy redshifts, along with their angular position in the sky, will provide an accurate estimation of the power spectrum in a volume of $80h^{-3}$ cubic Gigaparsecs. This will allow to reconstruct the power spectrum with an error below 1% over several redshift slices down to $z \approx 2$. The cosmological parameters like n_s, Ω_m, w_{DE} or even the mass of neutrinos will be estimated to a precision of one per cent or better.

Beside cosmology, Euclid will create immense catalogs of sources (stars, galaxies, clusters, supernovae) that will have an impact in almost every field of astronomy. More info on `www.euclid-ec.org`.

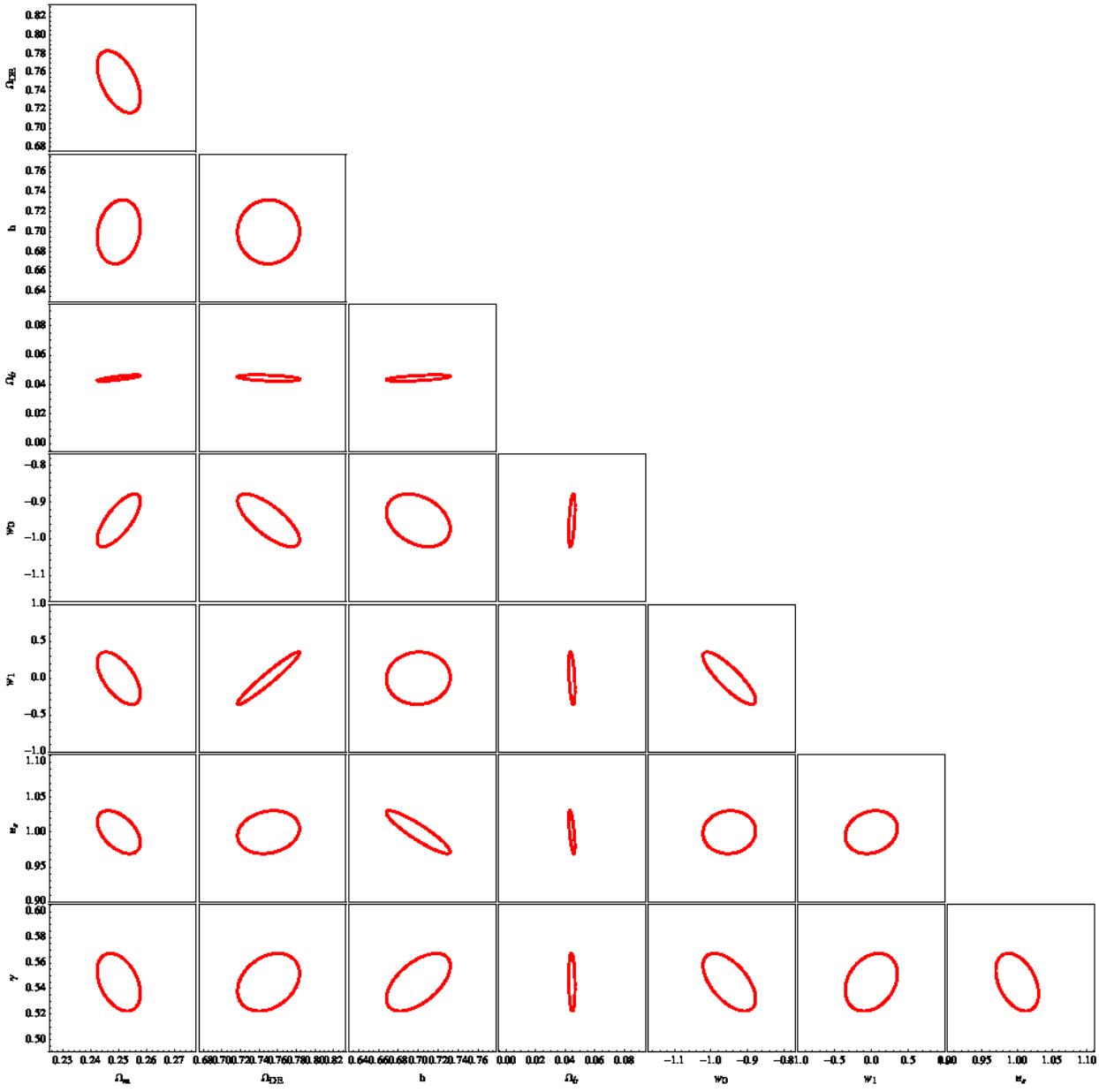


Figure 12.8.1: Euclid prediction for various cosmological parameters, using only galaxy clustering. The ellipses represent the confidence regions to 1σ for pairs of parameters, obtained with the Fisher matrix approximation.

Chapter 13

Weak lensing

General relativity shows that a light ray passing at a distance r from a point-like mass M is deflected by an angle

$$\theta = \frac{4GM}{rc^2} \tag{13.0.1}$$

A similar naive calculation in Newtonian physics gives half this result. The light deflection of stars near the limb of the Sun during an eclipse, found by Eddington in 1919, was one of the first evidence in support of GR against Newtonian gravity. In this chapter we extend this concept to random small fluctuations of the gravitational potential.

Quick summary

- The image of distant galaxies is magnified and distorted (sheared) by the gravitational potential of structures along the line of sight
- The magnification is however difficult to detect since galaxies have a very large intrinsic luminosity scatter
- In linear approximation, the distortion is proportional to the second derivative of the potential orthogonal to the trajectory and can be entirely described by a 2×2 distortion matrix
- The shear correlation or power spectrum is then proportional to the correlation function or power spectrum of matter, no bias has to be taken into account
- Detection of ellipticity has to be extremely precise in order not to introduce artificial anisotropies
- Wide area surveys are necessary to use shear lensing to constrain cosmological parameters
- More details and derivation of several formulae in Dodelson, *Modern Cosmology*, or Amendola-Tsujikawa, *Dark Energy.*

13.1 Convergence and shear

Eq. (13.0.1) can be generalized to give the deflection along a trajectory that passes through a continuous gravitational potential Ψ in the limit of small deviation (weak lensing)

$$\theta_i = \hat{\theta}_i + 2\int_0^r dr'(1 - \frac{r'}{r})[\nabla_{\perp i}\Psi(r')] \tag{13.1.1}$$

(units $c = 1$) where $\hat{\theta}_i$ are the angular coordinates of the source (i.e., the unperturbed position) and $\nabla_{\perp i}$ is the gradient orthogonal to the line of sight at position r. The index i runs over $1, 2$ to express the two angular

dimensions on the sky. If the deviation is large then we have some form of strong lensing, for instance double images or arcs. We will not discuss these here.

The angle of deflection itself however cannot be observed since we do not know the "true" direction of the incoming photon, contrary to the case of stars during an eclipse. We can only observe the average statistical distortion in nearby light rays, i.e. the derivative

$$\frac{\partial\theta_i}{\partial\hat{\theta}_j} = \delta_{ij} + 2\int_0^r dr'(1-\frac{r'}{r})r'\Psi_{,ij}(r') \equiv \delta_{ij} + D_{ij} \tag{13.1.2}$$

where here $\hat{\theta}_i$ are the components of the unperturbed direction of the incoming photons and now we have second orthogonal derivatives of Ψ (always to be evaluated orthogonally to the line of sight). This equation gives therefore a mapping from the source plane $\hat{\theta}$ to the image plane θ. This defines a distortion matrix for a source (a galaxy) at distance r:

$$D_{ij}(r) = 2\int_0^r dr'(1-\frac{r'}{r})r'\Psi_{,ij}(r') \tag{13.1.3}$$

This matrix can be written as the sum of a diagonal matrix and a trace-free one,

$$D = \begin{pmatrix} -\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & -\kappa+\gamma_1 \end{pmatrix} = -\kappa I + \begin{pmatrix} -\gamma_1 & -\gamma_2 \\ -\gamma_2 & \gamma_1 \end{pmatrix} \tag{13.1.4}$$

where I is the identity matrix and

$$\kappa = -\int_0^r dr'(1-\frac{r'}{r})r'(\Psi_{,11}+\Psi_{,22}) \tag{13.1.5}$$

$$\gamma_1 = -\int_0^r dr'(1-\frac{r'}{r})r'(\Psi_{,11}-\Psi_{,22}) \tag{13.1.6}$$

$$\gamma_2 = -2\int_0^r dr'(1-\frac{r'}{r})r'\Psi_{,12} \tag{13.1.7}$$

Here κ is called *convergence* and $\gamma_{1,2}$ *shear*. If one has a small circular image on the sky, with equation $\theta_x^2+\theta_y^2 = r^2$ and "distorts" it by the transformation $\theta_i' \to \theta_i + D_{ij}\theta_j$ (where $i,j = x,y$) one obtains at first order in $\kappa, \gamma_{1,2}$

$$\theta_x^2(1-2\kappa-2\gamma_1) + \theta_y^2(1-2\kappa+2\gamma_1) - 4\gamma_2\theta_x\theta_y = r^2 \tag{13.1.8}$$

or dividing by $1-2\kappa$ (and always linearizing, i.e. for $\kappa, \gamma_{1,2} \ll 1$)

$$\theta_x^2 + \theta_y^2 - 2\gamma_1(\theta_x^2-\theta_y^2) - 4\gamma_2\theta_x\theta_y = r^2(1+2\kappa) \tag{13.1.9}$$

i.e. the equation of an ellipse that can be decomposed into a new circle of radius $r(1+\kappa)$ and an ellipsoidal shear with tilt angle equal to $\gamma_2/2\gamma_1$ and difference between semimajor axes equal to $2\gamma_1$ (again for small ellipticity). The parameters $\gamma_{1,2}$ can indeed be directly used as a measure of the ellipticity of the distortion.

If the galaxy were intrinsically spherical and with a known intrinsic luminosity, then its lensed image would show a change in radius proportional to κ, leading to flux magnification ($\kappa > 0$) or demagnification ($\kappa < 0$) equal to $1+2\kappa$, and a distortion given by $\gamma_{1,2}$ and we could directly measure the effect of the potential. Since however galaxies are not spherical nor standard candles, we need to use statistical methods.

Along any given observation beam of finite angular size we see of course not just one but a number of galaxies, distributed according to some function $n(r)$. Assuming this distribution is normalized to unity, i.e. $\int_0^\infty n(r)dr = 1$, the total distortion matrix can be formed as an average over all the matrices along the line of sight,

$$D_{ij} = 2\int_0^\infty n(r')dr' \int_0^{r'} dr(1-\frac{r}{r'})r\Psi_{,ij}(r) \tag{13.1.10}$$

(we have interchanged r, r'). Now the integration boundaries can be manipulated and the integral can be written as

$$D_{ij} = \int_0^\infty w(r)\Psi_{,ij} \tag{13.1.11}$$

$$w(r) = \int_r^\infty dr'(1-\frac{r}{r'})rn(r') \tag{13.1.12}$$

Similarly, one has an average convergence κ and an average shear $\gamma_{1,2}$ at any position in the sky, so one could create a map of convergence and shear for the entire extragalactic sky. Weak lensing methods seek to compare this map to theoretical predictions, just like one does for CMB temperature and polarization.

13.2 Ellipticities and systematics

Weak lensing induces very small distortions: the typical shear is of order 1% at $z \approx 1$ over angular scales around 1deg. We need therefore to measure the ellipticity very precisely. The simplest method is the estimation of the quadrupole of an image, given as a intensity function $I(\theta)$ (assumed centered on the origin) on the plane θ_1, θ_2:

$$Q_{ij} = \int I(\theta)\theta_i\theta_j d^2\theta \tag{13.2.1}$$

The relation between Q_{ij} and the shear parameters is then

$$\gamma_1 = \frac{1}{2}\frac{Q_{11}-Q_{22}}{Q_{11}+Q_{22}}, \quad \gamma_2 = \frac{Q_{12}}{Q_{11}+Q_{22}} \tag{13.2.2}$$

However the integral converges very badly and is influenced by noise at large angular separation. In practice is much better to fit the images with some model (exponential or de Vaucouleurs profile) and evaluate Q_{ij} integrating the best fit model. Other methods fit the images with a series of basis functions.

A common problem is the deconvolution of the effects of the point spread function (PSF) i.e. the spreading of the incoming light due to the telescope optics and to atmosphere. The PSF can easily introduce minuscule anisotropies in otherwise perfectly circular images and lead to spurious lensing detection. Also, the PSF might vary with time during a survey due to changing conditions and to the action of gravity or of thermal effects on the telescope mirrors. Observations from space have the great advantage of a much stabler environment.

A systematics that instead is impossible to remove but has to be modeled is the fact that galaxy ellipticities are actually correlated even without lensing because nearby galaxies formed probably together from a larger perturbation and their shape will be not completely random (intrinsic-intrinsic correlation). For instance, galaxies formed out of a flattened structure will tend to be flattened on the same plane. Moreover, the same gravitational fluctuation that lenses some background galaxies will also contribute to their intrinsic correlation, so lensing and intrinsic ellipticity will also be correlated at some level (galaxy-intrinsic correlation). All these effects ultimately depend on the matter fluctuation field so can in principle be estimated and taken into account.

13.3 The shear power spectrum

The convergence induces a positive or negative magnification of the source. In principle this effect could be observed as an additional luminosity fluctuation of a population of standard candles, eg supernovae Ia. So far however this "weak magnification" has not been measured well enough to be used in cosmology because of the large scatter in the absolute luminosity of even the best standard candles we have available. We focus therefore now on the shear effect.

As we mentioned, galaxies do have their own large intrinsic ellipticity so it is not possible to measure directly the additional ellipticity induced by weak lensing. However the intrinsic ellipticity arises out of the local formation process so two distant galaxies should be totally uncorrelated; therefore the correlation function, or power spectrum, of the ellipticity parameters, should be vanishing above some angular separation. For instance, two galaxies at distance 1000 Mpc separated by 10 Mpc, so at an angle of roughly 10^{-2}rad≈ 0.6 deg, should have intrinsic ellipticity completely independent of each other. In contrast, a gravitational potential fluctuation

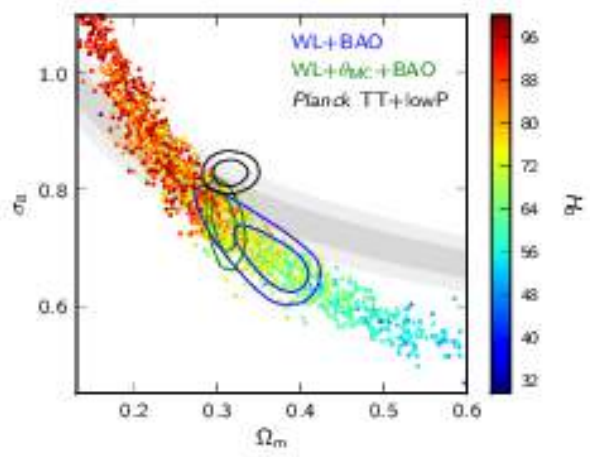


Figure 13.3.1: Estimates of σ_8 and Ω_{m0} with lensing, BAO and Planck (closed contours). Notice the hint of a tension between the weak lensing estimates and Planck. From Planck collab., 2015 (Ade et al. A&A 594, A13, 2016); ESA and the Planck Collaboration).

somewhere close to us along the line of sight will lens both galaxies in the same way, inducing a correlation in $\kappa, \gamma_{1,2}$ between different directions.

The calculation of the angular correlation can be performed relatively easily for small angular distances, where the sky can be assumed flat, by applying the Limber projection. At the end of the day, it is clear that since $\kappa, \gamma_{1,2}$ are linear functions of Ψ, and Ψ is a linear function of the matter density contrast δ_m, the final result must be that the power spectrum of κ or $\gamma_{1,2}$ can be written as a linear function of the matter power spectrum $P_{\delta_m}(k)$. In fact one finds (in flat space)

$$P_\kappa(\ell) = \frac{9}{4}\int_0^\infty dz \frac{W(z)^2 H(z)^3 \Omega_m(z)^2}{(1+z)^4} P_{\delta_m}(\frac{\ell}{r(z)}) \tag{13.3.1}$$

$$= \frac{9}{4}\Omega_{m0}^2 H_0^4 \int_0^\infty \frac{dz}{H(z)} W(z)^2 (1+z)^2 P_{\delta_m}(\frac{\ell}{r(z)}) \tag{13.3.2}$$

where $\ell = 180\,\mathrm{deg}\,/\theta$ is the multipole and

$$W(z) = \int_z^\infty \frac{dz'}{H(z')}\left(1 - \frac{r(z)}{r(z')}\right) n(z') \tag{13.3.3}$$

The galaxy "selection function" $n(z)$ models the number of galaxies $n(z)dz$ at any given redshift z; certainly for small z it must go as z^2 since $dn = n(z)dz \approx 4\pi n_0 z^2 dz$, but at larger z the galaxies will begin to be too faint to be detected and evolutionary effects will start to take place, so a typical form of $n(z)$ is

$$n(z) \approx z^2 \exp\left[-(z/z_0)^\alpha\right] \tag{13.3.4}$$

(to be normalized to unity). Very conveniently, one finds that the power spectra of $\gamma_{1,2}$ are proportional to P_κ and do not give additional information. One also finds that the power spectrum of the combination $B = -\sin(2\phi)\gamma_1 + \cos(2\phi)\gamma_2$ vanishes for any ϕ as long as the perturbations are induced only by the scalar part of the metric (i.e. do not include gravitational waves and rotational flows): this part of the weak lensing shear, called magnetic part or B-mode, can then be employed to cross-check the observations. Notice that P_κ depends at small z on $\Omega_{m0}^2\sigma_8^2$; this is the combination to which weak lensing is most sensitive to and often the observational constraints are reported on the σ_8, Ω_{m0} plane.

Finally, one can improve upon Eq. (13.3.1) by considering the shear correlation only among galaxies in a redshift bin $z_1, z_1 + \Delta z$, or among two population of galaxies, one in a bin centered at z_i and the other at z_j: this is called *tomographic* weak lensing, because it employs also radial information, not just angular correlations. Then given a matter power spectrum $P_{\delta_m}(k)$ one has

$$P_{\kappa,ij}(\ell) = \frac{9}{4}\int_0^\infty dz \frac{W_i(z) W_j(z) H(z)^3 \Omega_m(z)^2}{(1+z)^4} P_{\delta_m}(\frac{\ell}{r(z)}) \tag{13.3.5}$$

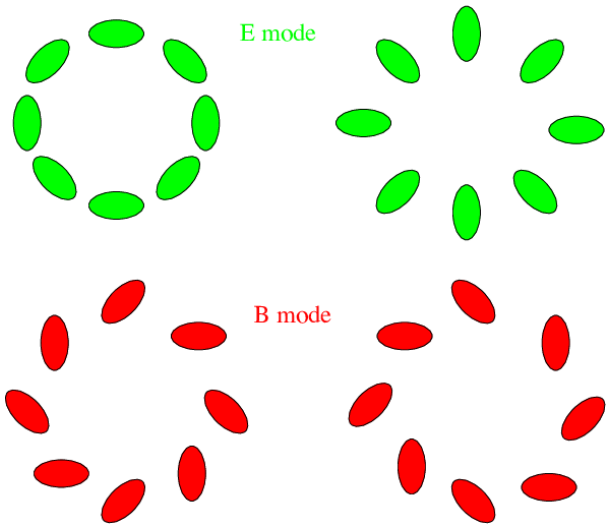


Figure 13.3.2: Patterns of lensing E- and B-modes. (From van Waerbeke L., Mellier Y., *Gravitational Lensing by Large Scale Structures: A Review.* Proceedings of Aussois Winter School, astroph/0305089 (2003))

where now the window functions W_i include the selection function n_i only of the galaxies in the i-th bin.

It is important to realize that contrary to the matter power spectrum, here the observed convergence spectrum does not need a bias correction, since it depends on the total matter content. Shear weak lensing can give therefore an absolute normalization of $P_\delta(k)$.

13.4 Current results

Shear lensing has first been detected around year 2000 but it took several more years before accurate results with systematics under control were produced. In 2011 Lin et al. using SSDS data on 275 square degrees obtained the result

$$\sigma_8(\Omega_m/0.3)^{0.7} = 0.64 \pm 0.1 \tag{13.4.1}$$

A summary of current results is in Table 3 of Weinberg et al. 2012. A comparison with CMB Planck results is in Fig. (13.3.1). Shear weak lensing will be measured on more than 1/3 of the sky by the Euclid satellite, as we have seen in Sect. (12.8).

Part III

Galaxies and Clusters

Chapter 14

Non-linear perturbations: simplified approaches

Quick summary

- Strongly non-linear fluctuations are difficult to handle and normally one has to employ powerful numerical simulations
- We first introduce the Zel'dovich approximation, which allows to follow in an almost analytical way the initial stages of structure formation beyond linearity
- Some more analytical results can be obtained assuming a spherical collapse. On scales of galaxies and clusters, Newtonian physics is sufficient
- Spherical collapse gives a simple but surprisingly accurate expression for the density of collapse and of virialization
- Using the so-called Press-Schechter formalism, one can approximately predict the number density of collapsed object as a function of their mass, to be compared to real data or simulations
- In this entire chapter we can safey use Newtonian gravity since we deal with scales well smaller than the horizon.

14.1 The Zel'dovich approximation

So far we have only investigated linear perturbations, except for a brief comment on Sect. 12.7. Stars, galaxies and clusters, however, are certainly not linear objects. For instance, the density contrast of a typical cluster of galaxies can be $\delta > 200$. Going from the linear treatment to the non-linear one is however generally very difficult. Even if some important step forward can be achieved by going to higher order in perturbation theory, ultimately one needs large N-body simulations.

A popular way to make progress in non-linear evolution before ressorting to numerical methods is to adopt the Zel'dovich approximation. The idea is to follow the movement of particles under the action of gravity until they hit each other and create a (fomally) infinite density. This should approximate the behavior of particles in a N-body simulation at least at some early time. Consider two sets of *comoving* coordinates. One, x_0, represents the coordinates of particles in an unperturbed Universe. Since they are comoving, they do not depend on time. The other, $\mathbf{x}(t)$, in a perturbed one. Initially, we perturb the position of each particle by a vector field $\mathbf{s}$, called displacement. Then we assume that at some later time t the position of the particles is given by

$$\mathbf{x}(t) = \mathbf{x}_0 + g(t)\mathbf{s}(\mathbf{x}_0) \tag{14.1.1}$$

This means we assume the position at time t only depends on the initial displacement through a time (and not space) dependent function, still to be defined. The density of the particle at any given time is $\rho(x, t)$ in the

perturbed Universe and $\rho_0(t)$ in the unperturbed one (which just follow the cosmic expansion, $\rho_0 \sim a^{-3}$). Since the particle number density $dn = \rho dV$ must be conserved, we have

$$\rho(x,t)d^3x = \rho_0(t)d^3x_0 \tag{14.1.2}$$

which implies

$$\rho(x,t) = \rho_0(t)\left|\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right|^{-1} \tag{14.1.3}$$

Let us assume now, without loss of generality, that the coordinates have been chosen along the direction of the eigenvectors of the deformation tensor

$$d_{ij} \equiv -\frac{\partial s_i}{\partial x_{0,j}} \tag{14.1.4}$$

In this case d_{ij} is diagonal and therefore

$$\left|\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right| = \left|I + g(t)\frac{\partial \mathbf{s}(\mathbf{x}_0)}{\partial \mathbf{x}_0}\right| = |\delta_{ij} - g(t)d_{ij}| = (1-g\lambda_1)(1-g\lambda_2)(1-g\lambda_3) \tag{14.1.5}$$

where I and δ_{ij} represent the identity matrix and λ_i are the three eigenvalues of d_{ij} (we show below that the eigenvalues are real). This means

$$\rho(x,t) = \frac{\rho_0(t)}{(1-g\lambda_1)(1-g\lambda_2)(1-g\lambda_3)} \tag{14.1.6}$$

Before we comment on this important expression, let us understand the meaning of g and s. Expanding (14.1.6) for small $g\lambda_i$, we find

$$\rho(x,t) \approx \rho_0(t)(1 + g(t)(\lambda_1+\lambda_2+\lambda_3) = \rho_0(t)(1 + g(t)\mathrm{Tr}(d_{ij})) \tag{14.1.7}$$

and therefore

$$\delta(t) \equiv \frac{\rho(x,t)-\rho_0(t)}{\rho_0(t)} = -g(t)\frac{\partial s_i}{\partial x_{0,i}} = -g(t)\nabla_{x_0}\mathbf{s}(x_0) \tag{14.1.8}$$

This expression, being a linearized one, must coincide with the growth law, $\delta(t) = G(t)\delta_0$, where $G(t)$ is the growth function we have evaluated for various cases in Chap. (8). Then we see that we should identify $g(t)$ with $G(t)$ and

$$-\nabla_{x_0}\mathbf{s}(x_0) = \delta_0 \tag{14.1.9}$$

Now, from the Poisson equation and the Friedmann equation we have

$$\nabla^2\Psi = 4\pi\rho_m\delta = \frac{3}{2}a^2H^2\Omega_m G(t)\delta_0 \tag{14.1.10}$$

where the factor a^2 arises because we are adopting comoving, rather than physical, coordinates. Then we see that

$$\delta_0 = \frac{2}{3a^2H^2\Omega_m G}\nabla^2\Psi \tag{14.1.11}$$

and therefore

$$\mathbf{s}(x_0) = -\frac{2}{3a^2H^2\Omega_m G}\nabla\Psi \tag{14.1.12}$$

With this identification of $\mathbf{s}$, the deformation tensor d_{ij} is symmetric and therefore its three eigenvalues are real. Therefore, we have completely specified the prescription (14.1.1): $g(t)$ is the growth factor, and the initial

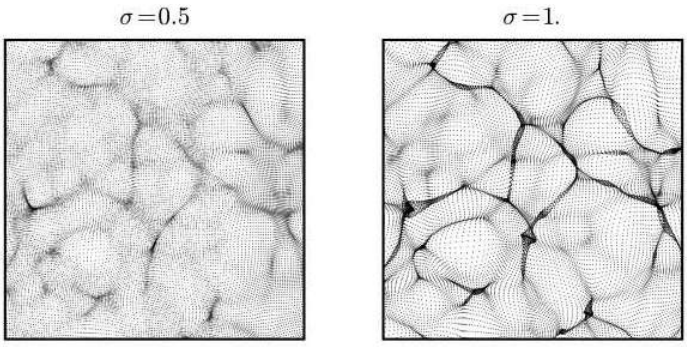


Figure 14.1.1: Formation of pancakes in a simulation based on the Zel'dovich approximation (from S. Shandarin, arXiv:0912.4520).

displacement field $\mathbf{s}$ is essentially the gradient of the gravitational potential, i.e. the force acting on the particles. In this way, one can run a very cheap N-body simulation: first, take the *linear* power spectrum at some early epoch for the model you want to simulate; second, convert the power spectrum for δ into a power sectrum for Ψ using Poisson equation in Fourier space; third, create a real space realization of this spectrum by overimposing sinusoidal oscillations with amplitude given by the spectrum and random phases; fourth, put particles on a regular grid; fifth, evaluate the displacement field by evaluating at every grid point (14.1.12); finally, move the particles out of their initial grid point by using (14.1.1).

To appreciate strenghts and limits of this technique, let us now come back to Eq. (14.1.6). Since $g(t)$ is a growing function (we discard the decaying mode, if any), $\rho(x,t)$ will develop a singularity as soon as one the largest λ_i is positive. This means that the particle will move primarily along the eigenvector associate to $\max\lambda_i$ and form regions of high density on the plane orthogonal to this direction: in other words, particle will tend to form planar structures, called pancakes (or *blinis* in the original Russian) by Zel'dovich, clearly visible in Fig. (14.1.1). After this singularity, the approximation will no longer be valid. In reality, is already quite surprising that the prescription (14.1.1) holds quite well beyond the linear regime!

Once the pancakes have been reached, one might assume that the particles "stick" onto, or oscillate around, the planar regions by friction or some hydrodynamic mechanism, and then continue flowing along the planes reaching the edges (called filaments) and finally slide along the filaments towards halos from which, in turn, galaxies and clusters will form. This is indeed qualitatively what is seen in full N-body simulations. Most of the current codes actually exploit the Zel'dovich approximation to speed up the calculations during the earliest stages at $z \gg 1$.

14.2 Spherical collapse[a]

After the formation of pancakes, the Zel'dovich approximation is no longer viable, although it can be extended through second-order schemes or *ad hoc* prescriptions. There is however a way to get, on a first approximation which however turns out to be surprisingly accurate, an estimate of an important observable, namely how many objects form (i.e., collapse into a virialized structure) for a given mass. This aproximation relies on sphericality and Gaussianity. The idea is first to find the value of the density contrast in the linear approximation at which a spherical perturbation collapse and virializes and then, find the fraction of the Gaussian distribution of perturbations that are above this δ collapse value. This fraction corresponds to the fraction of perturbations that form structures for a given mass.

For scales at which Newtonian thery applies, a shell of matter at distance R from the center of a spherical overdensity with uniform density ρ moves according to the Newtonian force law

$$\frac{\mathrm{d}^2R}{\mathrm{d}t^2} = -\frac{GM(R)}{R^2} = -\frac{4}{3}\pi G\rho R\,, \tag{14.2.1}$$

where $M(R) = 4\pi\rho R^3/3$ is the *constant* mass inside the shell. Since for pressureless matter the background

[a]Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

density scales as $\rho_0 = (3M(R_0)/4\pi)(R_0 a(t))^{-3}$, where R_0 is the initial size of the perturbation, we can define the density contrast as

$$\delta = \left(\frac{a(t)R_0}{R}\right)^3 - 1\,, \tag{14.2.2}$$

inside the shell and $\delta = 0$ outside. The crucial assumption here is that δ is a step, or *top-hat*, function, which allows in fact to cancel all spatial derivatives. Replacing R with δ, the equation for δ in our time variable N is then:

$$\delta'' + \left(1 + \frac{\mathcal{H}'}{\mathcal{H}}\right)\delta' - \frac{3}{2}\Omega_m\delta = \frac{4}{3}\frac{\delta'^2}{1+\delta} + \frac{3}{2}\Omega_m\delta^2\,. \tag{14.2.3}$$

Multyiplying Eq. (14.2.1) on both sides by $2\mathrm{d}R/\mathrm{d}t$ the equation can be integrated once as

$$\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 = \frac{2GM}{R} - C\,, \tag{14.2.4}$$

where C is an integration constant. This is the cycloid equation, whose solution for $C > 0$ can be given parametrically as $R = GM(1-\cos\tau)/C$ and $t = GM(\tau - \sin\tau)/C^{3/2}$ where $\tau \in (0, 2\pi)$. Substituting in δ and putting $a(t) = a_0(t/t_0)^{2/3}$ we obtain in the Einstein-de Sitter case:

$$\delta = \frac{9}{2}\frac{(\tau-\sin\tau)^2}{(1-\cos\tau)^3} - 1\,, \tag{14.2.5}$$

$$\delta_L = \frac{3}{5}\left[\frac{3}{4}(\tau-\sin\tau)\right]^{2/3}\,, \tag{14.2.6}$$

where δ_L (> 0) is the solution of the linearized equation, i.e. the left-hand-side of Eq. (14.2.3). Note that $\delta(\tau = 0) = 0$. It is convenient to use δ_L as a bookkeeping device: we express the behavior of δ as a function of δ_L instead of the parameter τ. A similar solution exists for an underdensity $\delta_L < 0$. We have assumed a constant mass $M(R)$: this implies that our analysis is valid only until shell-crossing occurs. As one expects, the radius R first increases (a small perturbation expands with the cosmological expansion), reaches a turnaround point and then decreases to zero (the perturbation collapses under its own gravity). The final singular phase is of course unphysical because the dust assumption will fail at some high density, non-radial fluctuations will develop and even the dark matter collisionless component will undergo the so-called "violent relaxation" mechanism and will set into virial equilibrium.

The main result we get from this model is the *critical* or *collapse* value $\delta_{\rm coll}$ of the *linear* fluctuation δ_L that is reached at the time of collapse. This quantity is of cosmological relevance because it is used in the Press-Schechter theory [**?**, **?**] as a first approximation to the epoch of galaxy formation and to calculate the abundance of collapsed objects, as we will discuss below. It can be seen from Eq. (14.2.6) that when $\tau = 0$ the perturbations are zero, then δ reaches a turnaround at $\tau = \pi$ (for which $\delta_T \equiv \delta(\pi) = (3\pi/4)^2 - 1 \approx 4.6$ and $\delta_L \approx 1.063$) and finally for $\tau = 2\pi$ the overdensity δ (but of course not δ_L) becomes singular. This singularity occurs when

$$\delta_L = \delta_{\rm coll} = (3/5)(3\pi/2)^{2/3} \approx 1.686\,, \tag{14.2.7}$$

and it takes exactly twice as much time as for the turnaround. Notice that this value is independent of time: a spherical perturbation in the Einstein-de Sitter universe collapses to a singularity whenever the linear density contrast equals 1.686. For other models, however, $\delta_{\rm coll}$ depends on time. An approximation for dark energy with constant $w_{\rm DE}$ in flat space is (Weinberg and Kamionkowski, MNRAS 341, 2003, 251)

$$\delta_{\rm coll}(z) = 1.686\,[1 + \alpha(w_{\rm DE})\log_{10}\Omega_m(z)]\,, \tag{14.2.8}$$

$$\alpha(w_{\rm DE}) = 0.353 w_{\rm DE}^4 + 1.044 w_{\rm DE}^3 + 1.128 w_{\rm DE}^2 + 0.555 w_{\rm DE} + 0.131\,. \tag{14.2.9}$$

One can define other phenomenologically interesting epochs that are sometimes used: the epoch of non-linearity ($\delta = 1$, corresponding to $\delta_L \approx 0.57$) and the epoch of *expected* virialization. The latter is *defined* to

correspond to the instant in which the kinetic energy K is related to the gravitational potential energy U by the condition

$$K = \frac{R}{2}\frac{\partial U}{\partial R}\,. \tag{14.2.10}$$

However, it is by no means obvious that this condition is enough to realize virialization, especially when dark energy is present. For an inverse-power potential ($U \propto -1/R$), the virialization implies $K = -U/2$. The radius and the density of the perturbation at virialization can be calculated by assuming conservation of energy at turnaround (when the kinetic energy vanishes; subscript T) and at a virialization epoch t_V when the kinetic energy satisfies $K_V = -U_V/2$, i.e.

$$U_T = U_V + K_V = U_V/2\,. \tag{14.2.11}$$

Since for a uniform sphere $U = -3GM/5R$ (and remembering once again we are assuming $M =$ constant), we obtain the relation $R_V = R_T/2$. Hence the virialized radius is half the turnaround radius. The density inside this radius turns out to be $\delta_V \approx 178$ and the epoch of this occurrence is very close to the final collapse time. A numerical fit for $w_{\rm DE} =$ constant models in flat space gives (Weinberg and Kamionkowski, MNRAS 341, 2003, 251)

$$\delta_V \approx 178[1 + b_1\theta^{b_2}(z)]\,, \tag{14.2.12}$$

$$\theta = \frac{1-\Omega_m(z)}{\Omega_m(z)}\,, \tag{14.2.13}$$

$$b_1 = 0.399 - 1.309(|w_{\rm DE}|^{0.426} - 1)\,, \tag{14.2.14}$$

$$b_2 = 0.941 - 0.205(|w_{\rm DE}|^{0.938} - 1)\,, \tag{14.2.15}$$

if z is the collapse redshift.

It is difficult to go much beyond this kind of phenomenological parametrization. A full understanding of non-linear physics in dark energy would require extensive N-body simulations coupled to lattice simulations of scalar fields, a technical feat which is still largely to be explored.

14.3 The mass function of collapsed objects[b]

The main reason why it is worthwhile to discuss the abstract phenomenon as a "spherical collapse" is that the critical value $\delta_{\rm coll}$ and the virial radius R_V (or rather the mass contained within that radius) enter the Press-Schecther (PS) formula for the abundance of virialized objects. The main idea behind the PS formula is that we can estimate the number of collapsed objects formed in a random Gaussian field by simply counting at any given time how many regions have an overdensity above the collapse threshold given by $\delta_{\rm coll}$.

Suppose at some redshift z we smooth a random Gaussian field of density fluctuations over cells of radius R, each containing on average the mass $M = 4\pi R^3\rho/3$ with $\rho(z)$ the background density. Since the smoothing is a linear operation, if the field is Gaussian then also the density contrast δ in the cells will be distributed as a Gaussian probability distribution function with variance $\sigma_M^2(z)$. Suppose that *all* the cells with $\delta > \delta_{\rm coll}$ undergo collapse and virialization. The fraction of collapsed regions (i.e. the fraction of space containing objects of mass *larger* than M) will be then

$$p(M,z)|_{\delta>\delta_{\rm coll}} = \frac{1}{\sigma_M(z)\sqrt{2\pi}}\int_{\delta_{\rm coll}}^{\infty}\exp\left(-\frac{\delta_M^2}{2\sigma_M^2(z)}\right)\,{\rm d}\delta_M = \frac{1}{2}\,{\rm erfc}\left(\frac{\delta_{\rm coll}}{\sqrt{2}\sigma_M(z)}\right)\,, \tag{14.3.1}$$

where ${\rm erfc}(x)$ is the error function. The fraction containing objects of mass within the range $[M, M+{\rm d}M]$ is given by

$${\rm d}p(M,z) = \left|\frac{\partial p(M,z)|_{\delta>\delta_{\rm coll}}}{\partial M}\right|{\rm d}M\,. \tag{14.3.2}$$

Remember that in general the threshold $\delta_{\rm coll}$ depends on z. Although the boxes with $\delta > \delta_{\rm coll}$ are certainly not in the linear regime, the idea is to use the linear regime to estimate the fraction of collapsed regions. We

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

are then implicitly assuming that the variance $\sigma_M(z)$ is in the linear regime ($\sigma_M \ll 1$) and therefore that it can be calculated from Eq. (9.9.6) with the linear spectrum at any redshift. By using the growth function $D(z)$ we have $\sigma_M(z) = D(z)\sigma_M(0)$.

Now, suppose in a volume V we find N collapsed objects, each occupying a volume $V_M = M/\rho$. Then by definition the volume occupied collectively by the N objects is the fraction $\mathrm{d}p$ of V, i.e.

$$NV_M = V\mathrm{d}p\,, \tag{14.3.3}$$

and therefore the number density $\mathrm{d}n$ of collapsed halos with mass in the $\mathrm{d}M$ range (the *mass function*) will be

$$\mathrm{d}n = \frac{N}{V} = \frac{\mathrm{d}p}{V_M} = \frac{\rho}{M}\left|\frac{\partial p(M,z)|_{\delta>\delta_{\rm coll}}}{\partial M}\right|\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\rho}{M^2}\frac{\delta_{\rm coll}}{\sigma_M}\left|\frac{\mathrm{d}\,\ln\sigma_M}{\mathrm{d}\,\ln M}\right| e^{-\delta^2_{\rm coll}/(2\sigma^2_M)}\mathrm{d}M\,. \tag{14.3.4}$$

The extra factor of two that we have inserted in the last step is required because we want all the masses to end up in some object, so that we impose the condition

$$V\int_0^\infty \left(\frac{\mathrm{d}n}{\mathrm{d}M}\right)\mathrm{d}M = 1\,. \tag{14.3.5}$$

This factor-of-2 adjustment can be justified with a random walk analysis of fluctuations. In any case, one finds it necessary to fit N-body simulations. Sometimes the number density $n(M,z)$ is taken to be the comoving number density (i.e. is multiplied by a^3): in this case also ρ should be identified with the comoving background density.

Equivalently, Eq. (14.3.4) is sometimes written as

$$\frac{M}{\rho}\left|\frac{\mathrm{d}n}{\mathrm{d}\,\ln\sigma_M}\right| = f(\sigma_M, z)\,, \tag{14.3.6}$$

where all the cosmological information is contained in the function

$$f(\sigma_M, z) = \sqrt{\frac{2}{\pi}}\frac{\delta_{\rm coll}}{\sigma_M}e^{-\delta^2_{\rm coll}/(2\sigma^2_M)}\,. \tag{14.3.7}$$

The number density $\mathrm{d}n(M,z)$ can then be "directly" confronted with the observed densities of objects (clusters, galaxies, quasars) at any redshift. The mass M is often taken to be the virial mass of that class of objects. Because of the exponential dependence on $\delta_{\rm coll}/\sigma_M$, the PS formula is quite sensitive to the cosmological model (see Fig. 14.3.1).

The simplicity of the PS approach must not hide the fact that it relies on a dangerous extrapolation of the linear theory, on the critical assumption of spherical collapse with top-hat filter, on a dubious definition of virialization, and on the absence of processes like merging, dissipation, shell crossing. Surprisingly, this shaky foundation did not prevent the PS formula to prove itself a valuable first approximation to the abundances obtained through numerical simulations. Not surprisingly, many works have been dedicated to improving the original PS formula by including corrections due to departure from sphericity or merging or by directly fitting to large N-body simulations. A remarkably successful fit is given by (Jenkins et al. MNRAS 321 (2001) 372)

$$f(\sigma_M, z) = 0.315\exp(-|0.61 - \ \ln\sigma_M(z)|^{3.8})\,. \tag{14.3.8}$$

This fit has been found to hold for a large range of masses, redshifts, and cosmological parameters, including dark energy with constant or varying $w_{\rm DE}$.

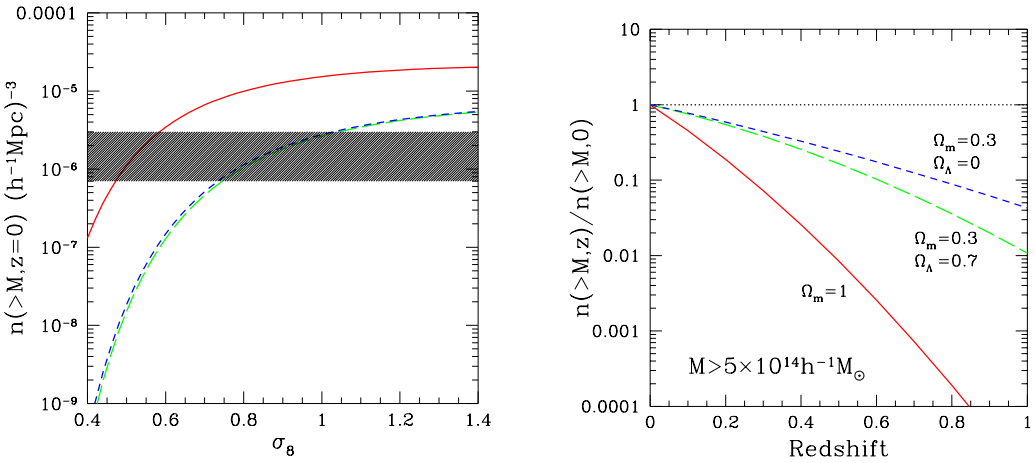


Figure 14.3.1: The sensitivity of the cluster mass function to cosmological models. Left panel: The cumulative mass function at $z = 0$ for $M > 5 \times 10^{14}\, h^{-1} M_\odot$ ($M_\odot$ is the solar mass) for three cosmologies, as a function of σ_8; solid line: $\Omega_m^{(0)} = 1$; short–dashed line: $\Omega_m^{(0)} = 0.3$, $\Omega_\Lambda^{(0)} = 0.7$; long–dashed line: $\Omega_m^{(0)} = 0.3, \Omega_\Lambda^{(0)} = 0$. The shaded area indicates the observational uncertainty in the determination of the local cluster space density. Right panel: Evolution of $n\,(> M, z)$ for the same cosmologies and the same mass limit, with $\sigma_8 = 0.5$ for the $\Omega_m^{(0)} = 1$ case and $\sigma_8 = 0.8$ for the low-density models. From Rosati, Borgani, Norman, ARAA 40 (2002) 539.

Chapter 15

Measuring mass in stars and galaxies

The matter content of the Universe is measured by first finding the mass contained in stars, then by weighing the mass of the galaxies by dynamical means. The comparison between the mass in the visible component, stars, dust and gas and the mass obtained by gravitational methods show that most of the mass should be in a component that is not directly visible, dubbed dark matter. In the next chapter we extend these ideas to cluster of galaxies.

Quick summary

- The measurement of masses and densities in the universe is a key task of cosmology since energy density in one of the main ingredients of GR equations
- The basic measure of star masses is through Kepler's third law applied to binary stars
- Using models of stellar evolution, the mass of a star can be approximately deduced by its luminosity and color
- Masses of galaxies can be inferred indirectly using the virial theorem: it turns out that the total gravitational mass is larger that the stellar and gas content of many galaxies, pointing to the existence of a large amount of dark matter in the halos.

15.1 Mass of stars[a]

The fundamental way to measure star mass is through application of Kepler's third law to binary stars. If we measure the orbital period P in years, the semi-major axis a around the center of mass in A.U. and the masses m_1, m_2 in solar masses, Kepler's law becomes very simple

$$(m_1 + m_2)P^2 = a^3 \tag{15.1.1}$$

Of course to measure the semi-major axis we must know the distance to the stars. Let us consider first the simplest case, a visual binary system in which we can directly see both stars. If the stars exhibit very small Doppler shifts in their spectra, it means that they orbit face-on. In this case, we can measure separately the two semi-major axes of the two orbits around the common center of mass. Then we can use the simple relation

$$m_1 a_1 = m_2 a_2 \tag{15.1.2}$$

along with (15.1.1) to measure both m_1, m_2, since $a = a_1 + a_2$. If we assume circular orbits, often a very good approximation, then we don't need to assume face-on orbits.

The second case is a spectroscopic binary, in which the two stars are too close to be resolved and we see only the varying spectra. If the star eclipse each other it means they orbit edge-on with respect to us. In this case,

[a] This section follows closely the treatment of Prof. Bartelmann's lecture notes.

we can measure by Doppler shift their highest velocity when they are at highest elongation: with this geometry the star velocity is directed entirely towards or away from us and we can measure distances from the center of mass and periods (using the eclipse epochs). So again we have two equations, Kepler 3rd law and the velocity relation

$$m_1 v_1 = m_2 v_2 \tag{15.1.3}$$

and we can close the system.

In more general cases, the orbits can take place on a plane with inclination i with respect to us and we can see only the velocity component along the line of sight, $v \sin i$. The masses will then be known only up to this factor.

In this way we can calibrate a number of scaling relations between the main properties of stars, namely mass, luminosity, temperature, density, pressure, radius, age. These relation are obtained by the fundamental equations of star equilibrium. These are: hydrostatic equilibrium between the pressure gradient and the force of gravity

$$\frac{1}{\rho}\frac{dP}{dr} = -\frac{GM}{r^2} \tag{15.1.4}$$

the conservation of mass

$$\frac{dM}{dr} = 4\pi r^2 \rho \tag{15.1.5}$$

the conservation of energy

$$\frac{dL}{dr} = 4\pi r^2 \rho \epsilon \tag{15.1.6}$$

where ϵ is the energy production rate per mass. Then there is a fourth relation which expresses how the temperature changes with r according to the energy transport in the star interior. Assuming pure radiative transfer (i.e propagation of photons) as in main sequence stars, this equation reads

$$\frac{dT}{dr} = -\frac{3L\kappa\rho}{64\pi r^2 \sigma T^3} \tag{15.1.7}$$

where

$$\sigma = \frac{2\pi^5 k_B^4}{15} h^3 c^2 = 5.67 \cdot 10^{-5} \frac{erg}{cm^2 K^4} \tag{15.1.8}$$

is Stefan-Boltzmann constant and κ is the opacity of the star. Basing on these equations the evolution of stars can in principle be followed until their death through explosion or passive cooling (see Fig. 15.1.1).

Putting $\rho \sim M/R^3$ the hydrostatic equation gives immediately (here and below "$\sim$" means proportional to)

$$P \sim \frac{GM^2}{R^4} \tag{15.1.9}$$

and inserting the equation of state of an ideal gas $PV = nRT$ or $T \sim P/\rho$, we find

$$T \sim \frac{M}{R} \tag{15.1.10}$$

Assuming κ independent of temperature, Eq. (15.1.7) gives

$$T^3 dT \sim LMr^{-5} dr \tag{15.1.11}$$

i.e. $L \sim T^4 R^4/M$ or the important relation

$$L \sim M^3 \tag{15.1.12}$$

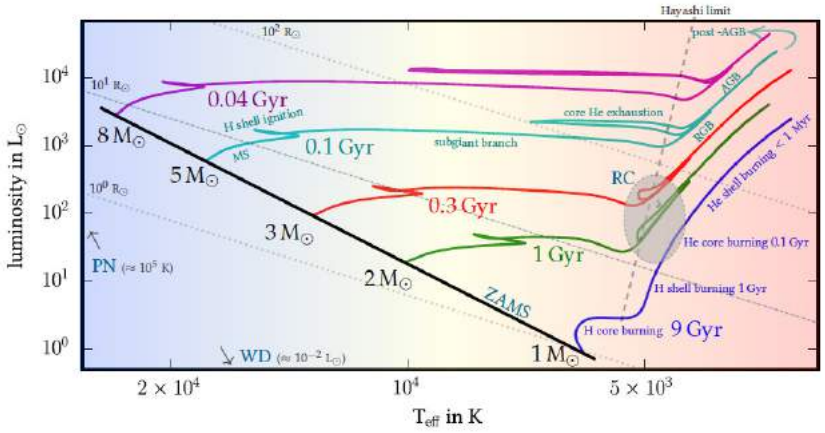


Figure 15.1.1: Evolutionary tracks of stars on the Hertzsprung-Russell diagram (courtesy of Jan Rybizki).

when combined with (15.1.10). A power $M^{3.5}$ gives however a more accurate approximation. Now the lifetime of a star on the main sequence times the luminosity should be proportional to the star fuel and therefore to its mass, so

$$L\tau_{MS} \sim M \tag{15.1.13}$$

and therefore

$$\tau_{MS} \sim M^{-2} \sim L^{-2/3} \tag{15.1.14}$$

The total luminosity of a star must also be proportional to its surface area and to T^4, i.e. $L \sim R^2T^4$. Then by combining with $T \sim M/R$ and with $L \sim M^3$ one also obtains

$$T \sim R, \qquad M \sim R^2, \qquad \tau_{MS} \sim T^{-1} \tag{15.1.15}$$

In practice, one evolves a numerical code to estimate precise relations among the various parameters. The mass-luminosity relation

$$\frac{L}{L_{\odot}} = \beta \left(\frac{M}{M_{\odot}} \right)^{\alpha} \tag{15.1.16}$$

shows an index α that varies from 2 to 4 depending on the star mass range and type and a constant β that is around unity for masses $< 20M_{\odot}$. Calibrating these relations with local samples one can estimate the masses of stars by their luminosity.

15.2 Mass of galaxies

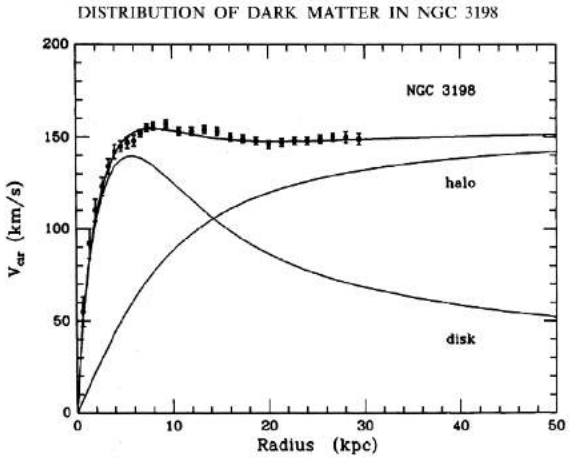
If a star moves in a circular orbit around the center of a spherical galaxy at distance R, then by the equilibrium of centripetal force and gravity one has

$$v^2 = \frac{GM}{R} \tag{15.2.1}$$

and the mass would be very easily estimated. For the Sun, moving at $V \approx 220$ km/sec at a distance of 8kpc$\approx 2.6 \cdot 10^{20}m$, one would get $M \approx 2 \cdot 10^{41}kg$ i.e. roughly 10^{11} solar masses, which is a very reasonable estimate. However galaxies do not look spherical and many stars do not have circular orbits.

If one assumes that most of the mass of spiral galaxies is concentrated in the roughly spherical central part (the bulge) then stars outside this region should have as in the previous case a Keplerian velocity

$$v \sim r^{-1/2} \tag{15.2.2}$$



31

Figure 15.1.2: Rotation curve in NGC3198 and the disk and halo components.

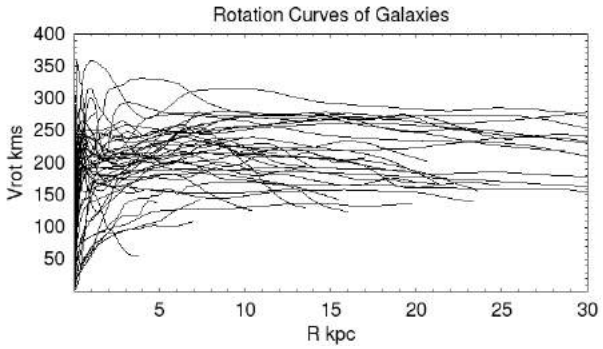


Figure 15.2.1: Rotation curves for many spiral galaxies (Sofue et al. 1999 ApJ, 523, 136S).

However we observe that the rotation velocity of stars and gas in spirals does not decay as $r^{-1/2}$ but rather stays approximately constant up to large distances, eg 30 kpc (see Fig. 15.2.1). The simple sun-like geometry of a central concentration cannot explain this behavior. Spiral galaxies are then modeled as the sum of a small spherical bulge, an extended flattened disk and a large spherical halo mostly composed of invisible dark matter (Fig. 15.2.2). If the halo density profile goes like $\rho \sim r^{-2}$ then one finds $v = const$. This density profile however does not converge to a finite mass so it must be truncated or steepened at some large radius.

The comparison of these galaxy models to observations can be summarized in the mass-to-light ratio

$$\mathcal{Y} \equiv \frac{M}{M_{\odot}} \cdot \frac{L_{\odot}}{L} \tag{15.2.3}$$

Current results vary from $\mathcal{Y} \approx 1 - 30$ for galaxies to $\mathcal{Y} \approx 100$ for clusters. This is a strong evidence in favor of the existence of a large component of dark matter. The ratio depends also on galaxy type, on the radius within which is measured, on the waveband and on various other characteristics of the objects. The value in the solar neighborhood is 2 ± 1. In a sample of elliptical and lenticulars it has been found

$$\mathcal{Y} = (3.8 \pm 0.14)(\frac{\sigma_e}{200km/sec})^{0.84\pm0.07} \tag{15.2.4}$$

where σ_e is the line-of-sight velocity dispersion, within an effective radius R_e measured as the radius that contains half of the total galaxy light.

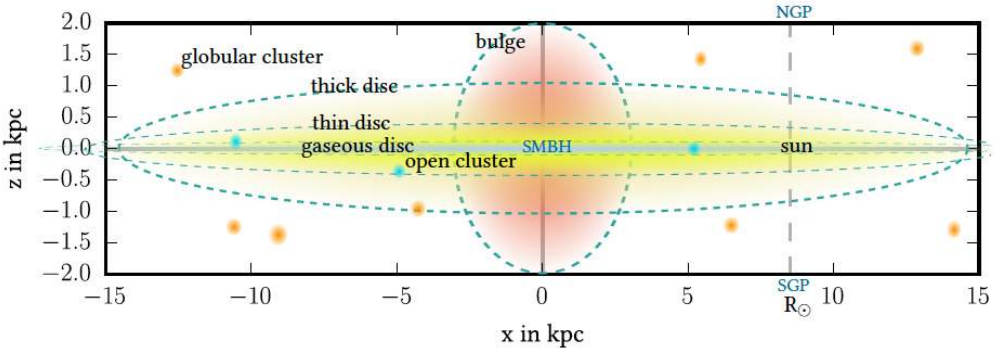


Figure 15.2.2: Structure of the Milky Way (courtesy of Jan Rybizki).

15.3 Halo profiles

An important outcome of N-body simulations is the observation that the density $\rho(r)$ of dark matter in halos has a more or less universal form. A lot of discussion about the validity of the cosmological models, and of the simulations themselves, centers around the determination of the best fitting halo profiles, and on their comparison to real galaxies, in particular to their rotation curves. All the halo profiles assume a spherical distribution (although halos can have significant deviations from sphericity). It is to be remarked that most proposed halo profiles are purely phenomenological, i.e. fits to N-body simulation, and have very little theoretical support. An extensive collection of results concerning halo profile can be found in Coe astro-ph/1005.0411.

The most well-known is the Navarro-Frenk-White (NFW) profile

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2} \tag{15.3.1}$$

which can be applied to dark matter halos from galaxy to clusters. Since it is singular at $r = 0$, this profile is called a *cuspy* profile, in contrast with *cored* profiles for which $\rho(r \to 0) = \rho_c$. At $r = r_s$, one can see that $\rho \sim r^{-2}$, i.e.

$$s = \frac{d\log\rho}{d\log r}|_{r=r_s} = -2 \tag{15.3.2}$$

At these scales, therefore, we expect flat rotation curves.

The mass inside r is

$$M(r) = 4\pi \int \rho(r) r^2 dr = 4\pi\rho_s r_s^3 (\ \ln(1+x) - \frac{x}{1+x}) \tag{15.3.3}$$

where $x = r/r_s$. Notice that the mass vanishes at $r \to 0$ but slowly diverges at large distances: a cut-off is then necessary. As we have seen in the section on spherical collapse (14.2), objects virialize when the non-linear density contrast is around 178, with a weak dependence on cosmology. This number is often approximated as $\Delta_c = 200$. This means the object contains within the virialization radius a density $\Delta_c \rho_{crit}$, where $\rho_{crit} = 3H^2/8\pi G$ is the critical density at that epoch. The mass inside the radius of virialization r_{vir} is then

$$M_{vir} = \frac{4}{3}\pi\Delta_c\rho_{crit} r_{vir}^3 \tag{15.3.4}$$

(also called M_{200} to specify the exact value chosen for Δ_c) So we can introduce a new combination of parameters, the so-called concentration c

$$c = \frac{r_{vir}}{r_s} \tag{15.3.5}$$

More in general, the concentration is defined as $c = r_{vir}/r_{-2}$, where r_{-2} is the radius at which the slope of $\rho(r)$ is -2, so for NFW $r_{-2} = r_s$. Then by combining (15.3.3) and (15.3.4) one has

$$\frac{\rho_s}{\rho_{crit}} = \frac{\Delta_c}{3} \frac{c^3}{\ln(1+c) - c/(1+c)} \tag{15.3.6}$$

One could then equivalently characterize the NFW profile either by ρ_s, r_s or by M_{vit}, c. The value of c can however be linked to M_{200}, although with a large scatter. From N-body simulation it appears in fact that there exists a concentration relation, for instance (Duffy et al. 2008)

$$c_{200} \approx 5.7 \left(\frac{M_{200}}{2 \times 10^{12} h^{-1} M_\odot} \right)^{-0.097} \tag{15.3.7}$$

Assuming this relation, the NFW profile depends on a single parameter. For instance, for a cluster of $M \approx 2 \cdot 10^{14} M_\odot/h$, one has $r_{vir} \approx 1$ Mpc$/h$ and $r_s = r_{vir}/c \approx 0.3$ Mpc$/h$. The concentration relation can be different for more relaxed clusters, and also depends on redshift and on the underlying cosmology, especially on the level of fluctuations σ_8. A recent accurate determination is in Dutton and Maccio' (2014). When applied to real data (through rotation curves), the concentration relation looks quite less tight than in simulations.

The gravitational potential from a mass profile $M(r)$ is

$$\Psi = G \int_{r_0}^{r} \frac{M(x)}{x^2} dx \tag{15.3.8}$$

and the velocity of stable circular orbits is

$$v^2 = r \frac{d\Psi}{dr} = \frac{M(r)}{r} \tag{15.3.9}$$

which as expected decreases much slower than the Keplerian law $v \sim 1/r$ for $r > r_s$ (se Fig. 15.3.1).

The observed velocities are however to be obtained by the sum of the gravitational potential of dark matter and baryons. Usually the baryons are modeled as the combination of a roughly spherical bulge, a gaseous component and an extended disk:

$$\Psi = \Psi_{bulge} + \Psi_{gas} + \Psi_{disk} + \Psi_{DM} \tag{15.3.10}$$

We can infer the baryonic mass directly from observations, up to unknown multiplicative constants: the mass-to-light ratios for disk and bulge, $Y_{b,d}$ (the gas component is supposed to have a universal mass-to-light ratio that can be inferred from fundamental principles). These ratios can be estimated from stars near the Sun, but remain quite uncertain when applied to other galaxies, so are better left as free parameters. The $Y_{b,d}$ parameters are $O(1)$ dimensionless numbers that quantify how much bigger or smaller the mass associated to the integrated light from the bulge or the disk is. For instance, $Y_b = 0.5$ means that for every erg/s of light coming from the bulge of a galaxy, one should associate half the mass it would associate in our local environment. So the baryonic component in the bulge, for instance, contributes by Y_b times the gravitational potential we would have inferred had we used the Milky Way local mass-to–light ratio. The total velocity we expect to observe can then be decomposed into a sum

$$v^2_{obs} = Y_b v^2_{bulge} + v^2_{gas} + Y_d v^2_{disk} + v^2_{DM} \tag{15.3.11}$$

A galaxy rotation curve can then be fitted quite generally with four free parameters (Y_b, Y_d, M_{200}, c), or three if we adopt the concentration relation. If the best fit gives $M_{200} > 0$, as it occurs for almost all galaxies we currently observe, then we can say we need dark matter.

Many other profiles have been proposed, for instance generalized NFW profiles

$$\rho(r) = \frac{\rho_s}{(r/r_s)^\gamma (1 + r/r_s)^{3-\gamma}} \tag{15.3.12}$$

which has the same large r behavior $\rho \sim r^{-3}$ but a flexible inner profile, $\rho(r \to 0) \sim r^{-\gamma}$. Another one is the Einasto profile

$$\rho(r) = \rho_{-2} \exp\left(-\frac{2}{\alpha} [(\frac{r}{r_{-2}})^\alpha - 1] \right) \tag{15.3.13}$$

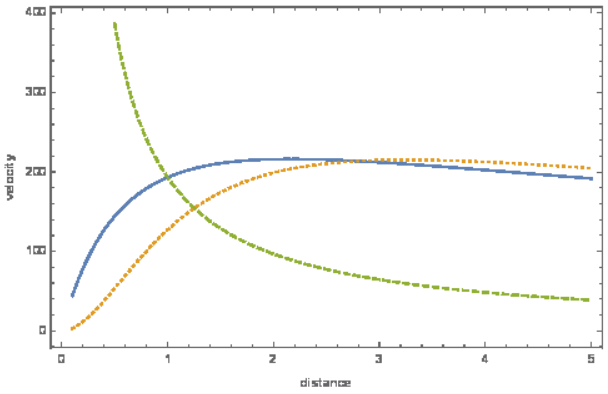


Figure 15.3.1: Circular velocities from the NFW profile (blue continuous line), from the Burkert profile (yello dotted line) and the Keplerian law (dashed green line). Here $r_s = r_0 = 1$.

where ρ_{-2} and r_{-2} are the density and radius at which $\rho(r) \sim r^{-2}$. The concentration is now defined as $c = r_{vir}/r_{-2}$. The parameter α has been found to be around 0.15-0.3.

The Burkert (1995) profile was adapted to the observed rotation curves of galaxies, in particular dwarf spiral galaxies that are supposed to be everywhere dominated by DM, rather than to N-body simulations. It is given by

$$\rho(r) = \frac{\rho_0}{(1 + r/r_0)(1 + (r/r_0)^2)} \tag{15.3.14}$$

and has the same large r properties as the NFW one, so as to match the dark matter halo at large distances, but a cored inner part, $\rho(r \to 0) = \rho_0$, closer to observations.

Finally, one can also notice that a small number of astronomers claim that one needs no dark matter at all, at least for as concerns galaxies, and that one should rather modify Eq. (15.3.8), the so-called MOdified Newtonian Dynamic (MOND) theories or their generalization and variants.

15.4 Galaxy luminosity function

The galaxies span a wide range of luminosities. The number density of galaxies in a range dL can be well approximated by the Schechter function

$$dn = \frac{\Phi^*}{L^*}\left(\frac{L}{L^*}\right)^{-\alpha} \exp(-\frac{L}{L^*})dL \tag{15.4.1}$$

where $L^* \approx (1 \div 2) \cdot 10^{10} L_{\odot}$, $\alpha \approx 1$ and the normalization is roughly $\Phi^* \approx 5 \cdot 10^{-3}\mathrm{Mpc}^{-3}$ (see Fig. 15.4.2). A galaxy with $L < 0.1L^*$ is considered a dwarf galaxy. Integrating over all luminosities we obtain the average luminosity density

$$\bar{L} = \int L \frac{dn}{dL} dL = \Gamma(2-\alpha)\Phi^* L^* \approx 10^8 \frac{L_{\odot}}{\mathrm{Mpc}^3} \tag{15.4.2}$$

Although the Schechter form seems quite universal, the values of L^*, α, Φ^* depend sensitively on galaxy type and on the waveband.

If the galaxies have an average M/L ratio around 10 then

$$\rho \approx 10^9 \frac{M_{\odot}}{\mathrm{Mpc}^3} \approx 10^{-31} \frac{\mathrm{g}}{\mathrm{cm}^3} \tag{15.4.3}$$

comparable to but smaller than the critical density. This kind of direct calculation is however very uncertain and does not lead to stringent constraints on Ω_m.

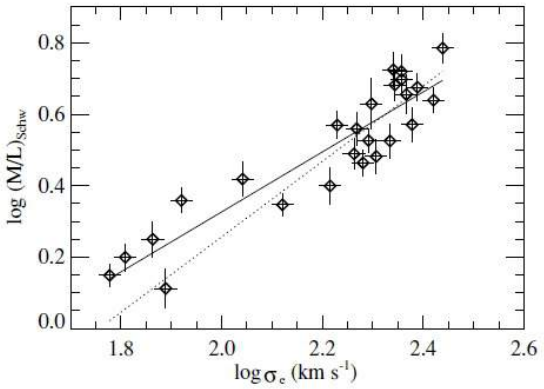


Figure 15.4.1: Relation between M/L and the velocity dispersion of elliptical and lenticular (S0) galaxies (Cappellari et al. 2006, Mon. Not. R. Astron. Soc. 366, 1126–1150 (2006)).

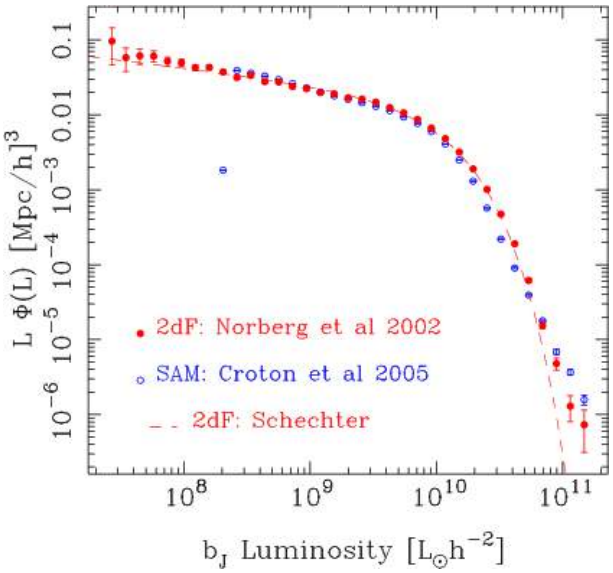


Figure 15.4.2: Galaxy luminosity function in the 2dFGRS survey, compared with a simulation and fitted by a Schechter function (from Smith 2012, 2012 MNRAS,426,531S).

Chapter 16

Cosmology with galaxy clusters

Galaxy clusters occupy a special position in cosmology, since they are the largest gravitationally bounded object in the Universe. They are therefore a bridge between linear scales, where the memory of the initial conditions is still fully traceable, and the strongly non-linear scales in which other non gravitational phenomena take place. Galaxy clusters can be studied by looking at the dynamics of member galaxies, at the X-ray emission of the hot intra cluster gas, at weak and strong lensing, and at the Sunyaev-Zel'dovich effect on the CMB photons. We will discuss some of these topics in this chapter.

16.1 Quick summary

- Galaxy clusters are the largest gravitational bounded object on the Universe
- They can be studied by internal dynamics, strong and weak lensing, X-ray emission of the hot intracluster component, Sunyaev-Zel'dovich effect
- All these probes point to a large amount of dark matter, ten times the baryonic component
- Clusters are a direct probe of cosmology also because their number density depend sensitively on perturbation growth and on volume measurements

16.2 Mass of clusters

Galaxy clusters are groups of hundreds to thousands galaxies within roughly 1 Mpc radius from their center. The closest cluster is Virgo (15 Mpc/h away); the closest rich and regular cluster is Coma (65 Mpc/h). The most regular among them appear as relatively isolated, almost spherical, groups in equilibrium (i.e. without significant subclustering neither in space nor in velocity).

There are at least three independent methods to determine the mass of clusters: (i) hydrostatic equilibrium between the intra-cluster medium (ICM) and the gravitational potential, (ii) dynamics of member galaxies, and (iii) lensing. We discuss here the first method.

Hydrostatic equilibrium for the ICM gas means that the gradient of the pressure $P_{\rm gas}$ equals the gravitational force:

$$\nabla P_{\rm gas} = -\rho_{\rm gas} \nabla \Phi_N \,, \tag{16.2.1}$$

where $\rho_{\rm gas}$ is the density and Φ_N is the gravitational potential. Assuming spherical symmetry we obtain

$$\frac{dP_{\rm gas}}{dr} = -\frac{GM\rho_{\rm gas}}{r^2} \,, \tag{16.2.2}$$

where we have used $\Phi_N = -GM/r$. Assuming the ideal gas equation of state

$$P_{\rm gas} = \frac{N}{V} k_B T = \frac{\rho_{\rm gas}}{\mu m_p} k_B T \,, \tag{16.2.3}$$

where $\mu \approx 0.6$ is the mean molecular weight for a gas with the expected primordial composition[a] and m_p is the proton mass, we obtain for the mass within a radius r:

$$M(r) = -\frac{r}{G}\frac{k_B T}{\mu m_p}\left(\frac{d\ \ln\rho_{\rm gas}}{d\ \ln r} + \frac{d\ \ln T}{d\ \ln r}\right) . \tag{16.2.4}$$

This provides a relation between the gas temperature T, the density profile $\rho_{\rm gas}$, and the total cluster mass profile $M(r)$. In turn, the gas temperature can be estimated by comparing the X-ray bremsstrahlung emission with plasma models. The gas density profile is often parametrized by the so-called β-model distribution

$$\rho_{\rm gas} = \frac{\rho_0}{[1+(r/r_c)^2]^{3\beta/2}} , \tag{16.2.5}$$

where $\beta = \mu m_p \sigma_r^2/(k_B T)$ is the ratio of the gas kinetic energy (σ_r is the line-of-sight velocity dispersion) to temperature. If, in addition, the temperature gradient d $\ln T$/d $\ln r$ is negligible (isothermal distribution) then the mass-temperature (M-T) relation reduces to

$$M(r) = \frac{3\beta k_B T(r)}{G\mu m_p}\frac{r^3}{r_c^2+r^2} \approx (1.1\times 10^{14} h^{-1} M_\odot)\beta\frac{T(r) r^3}{r_c^2+r^2} , \tag{16.2.6}$$

where r and T are in units of h^{-1} Mpc and keV, respectively. Although β is in principle measurable, it is always left as a free parameter in order to take into account at some level departures from the various assumptions (spherical model, ideal gas equation of state, isothermal distribution, etc.).

More complicated, and hopefully more realistic, models for the M-T relation have been proposed. Using such mass-temperature relations the mass of several clusters has been established, for instance, by the satellites Chandra and XMM-Newton. Averaging over many clusters it is also possible to fit a universal simple mass-temperature relation. The simple fit provided by Vikhlinin et al (2006) is

$$M = M_5\left(\frac{T}{5\,{\rm keV}}\right)^\alpha , \tag{16.2.7}$$

with $\alpha \approx$ 1.5-1.6 and $M_5 \approx 10^{14} M_\odot$. A value $\alpha = 3/2$ is indeed predicted for a virialized cluster, since in this case the velocity $V_{\rm vir}^2$ scales as $M/R \sim R^2$ i.e. as $M^{2/3}$ and the gas kinetic energy is proportional to the temperature, so that $V_{\rm vir}^2 \propto M^{2/3} \propto T$. Ultimately, a calibration of the mass-temperature relation will be provided by lensing mass estimations . Once one has a well-calibrated M-T relation, it is possible to infer the cluster masses directly by measuring the temperature of the hot gas through a comparison of their X-ray spectra to plasma models (the Bremmstrahlung spectrum goes like $\exp(-h\nu/k_B T_g)$ and contains often lines from highy ionized iron). From the mass of several clusters one can finally reconstruct the mass function and compare it to the theoretical prediction.

16.3 Baryon fraction[b]

Clusters can contribute to constrain dark energy parameters in another way, first proposed by Sasaki and Pen, expanding over previous work. As we have seen for the supernovae, what is needed for cosmology is not necessarily a standard candle but rather a *standardizable* candle, i.e. a source whose absolute luminosity depends in a known way on an independent observable. If in clusters the mass of baryons that emit light, either X-ray emitting hot intracluster gas or optical galaxies, is a fixed universal fraction of the total mass, then by estimating the total mass we can estimate the total baryon mass and the total luminosity. This works just as for the supernovae: there, we estimate the total luminosity correlating it with the light-curve width; here, we correlate it with the total mass. In both cases we do not need to know the value of the absolute luminosity but only that it is constant or varies in a controlled way.

[a] The molecular weight is the mass of a molecule in units of the proton mass, or of 1/12 of ^{12}C. A fully ionized gas can be considered composed of a mixture of "molecules" formed by either nuclei (mostly protons) or electrons. Since the electron mass is negligible, the mean molecular weight is 1/2. Adding a bit of Helium nuclei we obtain $\mu \approx$0.6.

[b] Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

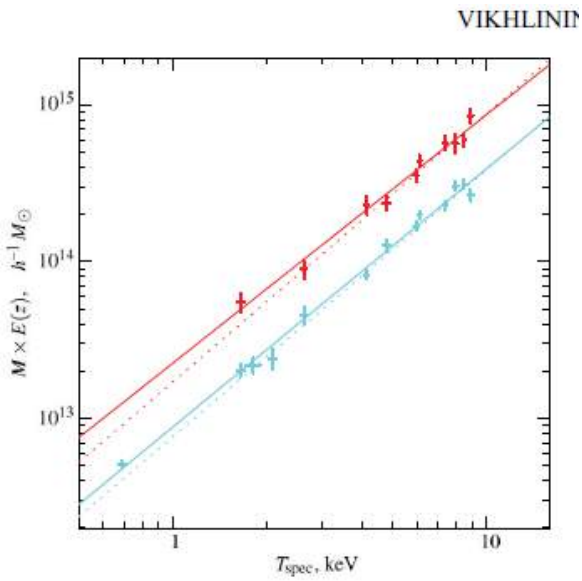


Figure 16.2.1: X-ray temperature vs. mass in clusters selected by the Chandra satellite (Vikhlinin et al. Astrophys.J.640:691-709, 2006)

In clusters most of the baryons are actually in the intra-cluster medium, so for sake of simplicity we only consider the X luminosity. The fundamental assumption is that

$$\frac{M_{\rm gas}}{M_{\rm tot}} = \frac{\Omega_b}{\Omega_m} = \text{constant}\,, \tag{16.3.1}$$

for all clusters. This is indeed likely because clusters are very large: to make up their mass, one has to pile up all the matter in a radius of roughly 10 Mpc. It is difficult to imagine such large volumes containing wildly varying proportions of baryons and dark matter. There would simply be no time for any reasonable process to segregate matter on such large scales.

So at least in standard cosmology, one expects all clusters to contain the fixed ratio of baryons to total matter set by cosmology. Now, the X-ray thermal bremsstrahlung luminosity that comes from those baryons is proportional to the volume $V \propto r^3$ of the emitting region and to the square of the electron density ρ_e, i.e. to $\rho_e^2 r^3$. Since the mass $M_{\rm gas}$ is in proportion to $\rho_e V$, it follows that $L_X \propto M_{\rm gas}^2/r^3$ or $M_{\rm gas} \propto (L_X r^3)^{1/2}$. We also notice that the X-luminosity is measured by an observed flux $\mathcal{F}_X = L_X/(4\pi d_L^2)$, so we can also write $M_{\rm gas} \propto d_L r^{3/2}$. On the other hand, from the hydrostatic equilibrium condition (16.2.4), we deduce that the total mass is $M_{\rm tot}(r) \propto r$ (see also Eq. 16.2.6 at large r), if we assume an isothermal distribution and that $d\,\ln\rho_{\rm gas}/d\,\ln r$ depends weakly on r (which is true for instance for all power-law $\rho_{\rm gas} \sim r^n$). So finally we have

$$\frac{M_{\rm gas}}{M_{\rm tot}} \propto \frac{d_L r^{3/2}}{r} \propto d_L r^{1/2}\,. \tag{16.3.2}$$

There is a final step to make. The size r of the emission region is seen under the angle $\theta = r/d_A$ (d_A is the angular diameter distance) and therefore the gas fraction within a fixed angle θ scales as

$$f_{\rm gas} = \left.\frac{M_{\rm gas}}{M_{\rm tot}}\right|_{<\theta} = A_1 d_L r^{1/2} = A_2 d_L d_A^{1/2} = A_3 d_A^{3/2}\,, \tag{16.3.3}$$

where A_1, A_2, A_3 are constants or observable quantities like $\theta$and z. Note that we have used the Etherington relation in the last step. The A_i's factors contain a lot of interesting physics but no cosmological parameters, so we are not concerned with them here. Then we see that $f_{\rm gas} d_A^{-3/2}$ is an observable quantity independent of cosmological parameters. So for instance if we have two clusters, taking the ratio of $f_{gas} d_A^{-3/2}$ gives the ratio of their angular diameter distances; so if we know the distance of one cluster we can estimate the distance to

the other. So fitting $f_{\rm gas} d_A^{-3/2}$ to the real data obtained by converting X-ray flux and temperature within the angle θ, we can constrain the cosmological parameters in d_A.

The simple $f_{\rm gas}$ prediction above relies on many things, from hydrostatic equilibrium to universal composition. Some approximations are easy to improve. For instance we can take into account the baryons contained in the galaxies rather than in the ICM. Other effects can be estimated from N-body, such as the typical departure from hydro-equilibrium or from universal composition. Some other uncertainties can be marginalized over in the likelihood. Allowing for considerable freedom in parametrizing these effects, a table of constraints on various cosmological parameters have been derived in Allen et al. (2008) from 42 clusters observed by the Chandra X-ray satellite. The constraint from the $f_{\rm gas}$ test alone gives for the equation of state of dark energy $w_{\rm DE} = -1.14 \pm 0.31$, for flat space and constant $w_{\rm DE}$ (all results here and below are at 1σ). The results in Ettori et al. on a different cluster dataset give $\Omega_m^{(0)} = 0.32^{+0.04}_{-0.05}$ and $w_{\rm DE} = -1.1^{+0.60}_{-0.45}$. In combination with SN Ia and CMB, these constraint tightens to $w_{\rm DE} = -0.98 \pm 0.07$.

16.4 Virial theorem

The mass of clusters can also be estimated by the dynamics of the member galaxies, through the virial theorem and the Jeans equations. Given a system of N particles of equal masses m with position $\mathbf{r}_k$ and momenta $\mathbf{p}_k = m\dot{\mathbf{r}}_k$, we define the quantity

$$G = \sum_k^N \mathbf{r}_k \cdot \mathbf{p}_k \tag{16.4.1}$$

summing over all particles. We assume the system is at rest, i.e. the average $\langle \dot{\mathbf{r}}_k \rangle$ is zero. The derivative of G is

$$\begin{aligned} \frac{dG}{dt} &= \sum_k (m\dot{\mathbf{r}}_k \cdot \dot{\mathbf{r}}_k + \dot{\mathbf{p}}_k \cdot \mathbf{r}_k) && (16.4.2)\\ &= \sum_k (m\dot{r}_k^2 + \mathbf{F}_k \cdot \mathbf{r}_k) && (16.4.3)\\ &= 2T + \sum_k \mathbf{F}_k \cdot \mathbf{r}_k && (16.4.4) \end{aligned}$$

where T is the total kinetic energy

$$T = \frac{1}{2} m \sum_k \dot{\mathbf{r}}_k^2 = \frac{1}{2} M \sigma_{rr}^2 \tag{16.4.5}$$

where $\sigma_{rr}^2 = \frac{1}{N}\sum_k \dot{\mathbf{r}}_k^2 - \langle \dot{\mathbf{r}}_k \rangle^2 = \frac{1}{N}\sum_k \dot{\mathbf{r}}_k^2$ is the particle velocity dispersion and $M = Nm$ the total mass, and $\mathbf{F}_k$ is the force acting on the k-th particle. Now we write this force as the sum of all the forces from the other particles

$$\mathbf{F}_k = \sum_{j=1}^N \mathbf{F}_{jk} \tag{16.4.6}$$

so that

$$\sum_k \mathbf{F}_k \cdot \mathbf{r}_k = \sum_k (\sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_{j>k} \mathbf{F}_{jk}\mathbf{r}_k) = \sum_k \sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_k \sum_{j>k} \mathbf{F}_{kj}\mathbf{r}_j = \sum_k \sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_j \sum_{k>j} \mathbf{F}_{jk}\mathbf{r}_k = \tag{16.4.7}$$

$$\sum_k (\sum_{j<k} \mathbf{F}_{jk}\mathbf{r}_k + \sum_{j>k} \mathbf{F}_{jk}\mathbf{r}_k) = \sum_k \sum_{j<k} \mathbf{F}_{jk}(\mathbf{r}_k - \mathbf{r}_j) \tag{16.4.8}$$

where we have used Newton's third law, $\mathbf{F}_{jk} = -\mathbf{F}_{kj}$ and the identity $\sum_k \sum_{j<k} = \sum_j \sum_{k>j}$. Then if we have a potential that depends only on the distance $r_{jk} = \sqrt{(r_{j,x} - r_{k,x})^2 + (r_{j,y} - r_{k,y})^2 + (r_{j,z} - r_{k,z})^2}$ (eg a

gravitational or Coulomb potential) we have

$$\frac{dr_{jk}}{d\mathbf{r}_{jk}} = \{\frac{dr_{jk}}{d(r_{j,x}-r_{k,x})}, \frac{dr_{jk}}{d(r_{j,y}-r_{k,y})}, \frac{dr_{jk}}{d(r_{j,z}-r_{k,z})}\} \quad (16.4.9)$$

$$= \frac{\mathbf{r}_k - \mathbf{r}_j}{r_{jk}} \quad (16.4.10)$$

and therefore

$$\mathbf{F}_{jk} = -\nabla_{r_{jk}} V = -\frac{dV}{dr_{jk}}\frac{dr_{jk}}{d\mathbf{r}_{jk}} = -\frac{dV}{dr_{jk}}\left(\frac{\mathbf{r}_k - \mathbf{r}_j}{r_{jk}}\right) \quad (16.4.11)$$

Then we have

$$\frac{dG}{dt} = 2T + \sum_k \sum_{j<k} \mathbf{F}_{jk}(\mathbf{r}_k - \mathbf{r}_j) \quad (16.4.12)$$

$$= 2T - \sum_k \sum_{j<k} \frac{dV}{dr_{jk}}\left(\frac{|\mathbf{r}_k - \mathbf{r}_j|^2}{r_{jk}}\right) \quad (16.4.13)$$

$$= 2T - \sum_k \sum_{j<k} \frac{dV}{dr_{jk}} r_{jk} \quad (16.4.14)$$

For a stationary object (often denoted as a *virialized* system), G does not depend on time and then

$$2T = \sum_k \sum_{j<k} \frac{dV}{dr_{jk}} r_{jk} \quad (16.4.15)$$

For the gravitational potential, $V(r_{jk}) = m^2 r_{jk}^{-1}$ and we obtain

$$2T = m^2 \sum_k \sum_{j<k} r_{jk}^{-1} = -V_{tot} \quad (16.4.16)$$

where V_{tot} is the total gravitational potential energy. A generalization to $V \sim r^n$ is immediate. So by measuring the total kinetic energy of a system, i.e. in practice its velocity dispersion, we can estimate the total potential energy, which depends essentially on the mass distribution.

For instance, for a homogeneous sphere of uniform density ρ, the potential energy of every shell of thickness dr is $dV = -G(\frac{4\pi}{3}\rho r^3)(4\pi r^2 \rho dr) r^{-1}$ so we have

$$m^2 \sum_k \sum_{j<k} r_{jk}^{-1} = -\int_0^R dV = \frac{16\pi^2 G}{3}\int_0^R \rho^2 r^4 dr \quad (16.4.17)$$

$$= \frac{16\pi^2 G\rho^2}{3}\frac{R^5}{5} \quad (16.4.18)$$

$$= \frac{3}{5}\frac{GM^2}{R} \quad (16.4.19)$$

so that in this case

$$\sigma_{rr}^2 = \frac{3}{5}\frac{GM}{R} \quad (16.4.20)$$

If we use as velocity dispersion the line-of-sight velocity dispersion (the only observable one) then this value has to be multiplied by 3 since it represents only one of the three components of velocity. Typical values for galaxy clusters are $\sigma_{rr} \approx 1000$ km/sec and $R \approx 1$Mpc so

$$M \approx 10^{45} kg \approx 10^{15} M_{\odot} \quad (16.4.21)$$

The luminosity of a cluster is around $L = 10^{13} L_{\odot}$ in the visible bands, so the mass-to-light ratio is roughly 100 times the solar one. Even accounting for a large quantity of ionized gas that emits in the X-ray band and not in the visible, the mass of clusters is at least ten times larger than the baryonic mass. This is one of the strongest evidences in favor of the existence of dark matter.

16.5 The abundance of clusters[c]

Galaxy clusters are helpful in cosmology also under another aspect. The number density of clusters is in fact a function of cosmological parameters.

A cluster forms when an overdensity grows enough to begin gravitational collapse. If we can identify at which density a collapse begins we can estimate how many clusters should form at any given epoch. Imagine a shell of matter at distance R from the center of a spherical structure of mass $M(R)$ and density ρ. Newtonian mechanics tells us that the shell will move according to

$$\ddot{R} = -\frac{GM(R)}{R^2} = -\frac{4}{3}\pi G\rho R \tag{16.5.1}$$

We assume that the mass $M(R)$ is a constant. Multiplying on both sides by $2\dot{R}$ we can integrate and obtain

$$\dot{R}^2 = \frac{GM}{R} - C \tag{16.5.2}$$

where C is an integration constant. This is the cycloid equation, whose parametric solution is

$$R = GM(1-\cos\tau)C^{-1} \tag{16.5.3}$$

$$t = GM(\tau-\sin\tau)C^{-3/2} \tag{16.5.4}$$

where $\tau \in (0, 2\pi)$. As τ moves from 0 to π and then to 2π, the radius R expands, reaches a maximum and then collapses to zero. In a matter dominated universe the background density ρ_{bg} evolves as $\sim M(R_0 a)^{-3}$, if R_0 is the initial radius of the shell and $a = a_0(t/t_0)^{2/3}$ the scale factor in MDE. At the same time, the density of the overdensity changes as $\rho \sim MR^{-3}$. Then we can define the density contrast

$$\delta = \frac{\rho - \rho_{bg}}{\rho_{bg}} = (\frac{a(t)R_0}{R})^3 - 1 \tag{16.5.5}$$

Once we have $R(t)$ from (16.5.3) we obtain also $\delta(\tau)$ as

$$\delta = \frac{9}{2}\frac{(\tau-\sin\tau)^2}{(1-\cos\tau)^3} - 1 \tag{16.5.6}$$

(we fixed $\delta(\tau \to 0) = 0$). As expected, for $\tau \to 2\pi$ one has $\delta \to \infty$, i.e. collapse to a singularity. A related quantity is the evolution of δ_L, the density contrast assuming a linear growth, i.e. $\delta_L = \delta_{L(in)}(a/a_{in}) = \delta_{L(in)}(t/t_{in})^{2/3}$. One finds

$$\delta_L = \frac{3}{5}\left[\frac{3}{4}(\tau-\sin\tau)\right]^{2/3} \tag{16.5.7}$$

where δ_L is normalized by imposing $\delta = \delta_L$ for $\tau \ll 1$. Now, when $\tau \to 2\pi$ one finds

$$\delta_L \to \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \equiv \delta_c \approx 1.686 \tag{16.5.8}$$

which is considered then the threshold for collapse if a linear standard evolution is assumed. In reality, before reaching the singularity the system will deviate from a pure radial behavior and the particles will start to orbit around the center, leading to a stable virialized structure like a galaxy or a cluster, but one can approximate the virialization epoch with the collapse epoch.

If all overdensities above $\delta_L = 1.686$ form stable structures, one may just evaluate the number density of clusters that are expected to grow above this value at any given time. One can assume that during the linear regime the fluctuations of any given size R are gaussian distributed so that the number of over- or under-dense regions goes as

$$N(\delta) \sim \exp -\frac{1}{2}\frac{\delta^2}{\sigma^2(z)} \tag{16.5.9}$$

[c]Adapted from Amendola & Tsujikawa, *Dark Energy. Theory and Observations*, CUP 2010.

where σ^2 is the variance of the distribution when filtered on the scale R obtained (this is a crucial point) within the *linear* approximation. Then the number of collapsed regions of size R (or corresponding mass M) will be proportional to the probability of collapse, i.e.

$$p(M,z) = \frac{1}{\sqrt{2\pi}\sigma(z)} \int_{\delta_c}^{\infty} d\delta \exp\left[-\frac{1}{2}\frac{\delta^2}{\sigma^2(z)}\right] = \frac{1}{2}\mathrm{erfc}\left[\frac{\delta_c}{\sqrt{2}\sigma(z)}\right] \tag{16.5.10}$$

This approximated scheme allows one to estimate the number density of collapsed structure of any size R by simply using an extrapolation of linear quantities, without solving the difficult non-linear problem. The probability of collapse for any given range of masses dM will be

$$dp = \frac{\partial p}{\partial M} dM \tag{16.5.11}$$

If we partition a volume V into regions of mass M that occupy a volume $V_M = M/\rho$, then there will be V/V_M of such regions and a fraction $N = dp(V/V_M)$ of them will collapse, i.e. a number density

$$dn = \frac{N}{V} = \frac{dp}{V_M} = \frac{\rho}{M}\frac{\partial p}{\partial M} dM \tag{16.5.12}$$

$$= \frac{\sqrt{2}}{\pi}\frac{\rho}{M^2}\frac{\delta_c}{\sigma_M}\left|\frac{d\log\sigma_M}{d\log M}\right| e^{-\frac{1}{2}\frac{\delta_c^2}{\sigma_M^2}} dM \tag{16.5.13}$$

The theoretical prediction depends then on the variance σ_M^2 at any given epoch and mass and on the collapse threshold δ_c, which however is found to be only mildly dependent on cosmology. The variance σ_M^2 can be written in linear growth approximation as $\sigma_M^2(z) = G^2(z)\sigma_M^2(0)$ where $G(z)$ is the growth function. It can be compared to observations once we have a reliable estimator of the cluster masses, for instance using the X-ray temperature scaling Eq. (16.2.7) or through the Sunyaev-Zel'dovich effect. This gives then a test of the power spectrum normalization σ_M^2. Since also Ω_m enters the growth function $G(z)$, the test constrains some combination of σ_8 and Ω_m.

16.6 Sunyaev-Zel'dovich effect

CMB photons passing through a cluster of galaxies have a 1% chance of scattering with a electron of the hot intra-cluster medium, the fully ionized, hot ($10^{7\div 8}$K) gas component that is trapped in galaxy clusters by their gravitational field. The scattering is an inverse Compton scattering, meaning that the photons gain, rather than lose, energy. This implies that CMB photons of low energy are boosted to high energy and the black-body spectrum is therefore distorted: this is the thermal Sunyaev-Zel'dovich (SZ) effect. Since we still approximate the distorted spectrum by a black-body spectrum, we just express the SZ distortion by a frequency-dependent change in CMB temperature (Fig. 16.6.1).

The relative energy injection to the photons after a single scattering is

$$\frac{k_B T_e}{m_e c^2} \tag{16.6.1}$$

where T_e is the electron temperature and m_e the electron mass. Taking into account the Thomson cross section σ_T, the total number of scatterings in a electron plasm with number density n_e along the line of sight $d\ell$ is $n_e\sigma_T d\ell$ and the total energy shift is then

$$y = \int n_e \frac{k_B T_e}{m_e c^2}\sigma_T d\ell \tag{16.6.2}$$

(called Compton y-parameter). Then the change in effective black-body temperature for the distribution of photons as a function of the dimensionless frequency $x = h\nu/k_B T_{CMB}$ is (non-relativistic limit)

$$\frac{\Delta T}{T}\Big|_{SZ} = \left(x\frac{e^x+1}{e^x-1} - 4\right) y \tag{16.6.3}$$

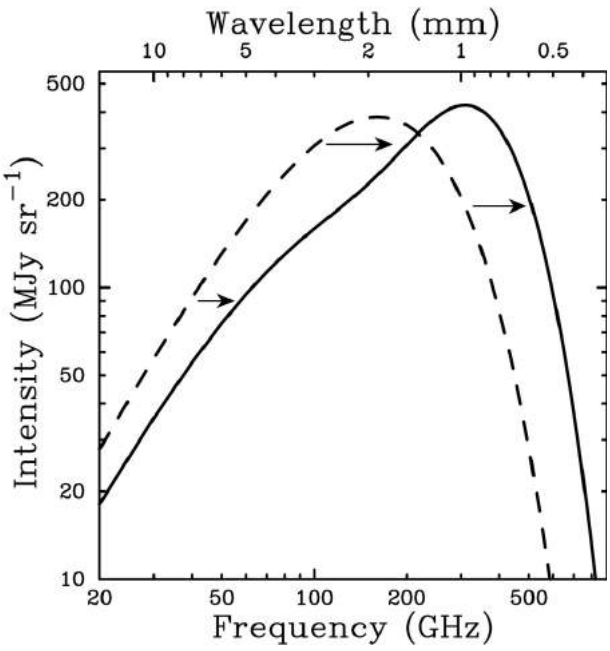


Figure 16.6.1: The distortion of the Cosmic Microwave Background (CMB) spectrum due to the Sunyaev-Zel'dovich effect (SZ) (solid line), here amplified by a factor of 1000 for readability. The CMB effective temperature decreases at frequencies below approximately 218 GHz and increases at higher frequencies. (From Carlstrom et al., 2002 ARA&A..40..643C)

This function changes sign around 220 GHz. Clusters will then appear on CMB maps as circular shadows below this frequency and as sources above it. Typical values of y are around 10^{-4}. The Planck satellite (2015) has detected more than a thousand SZ clusters in its CMB maps.

An important property of the SZ effect is that it is independent of redshift, so it can be used to map clusters at very high redshifts (of course within the experiment resolution).

The total integrated SZ (ISZ) effect in a cluster is obtained by integrating over the solid angle occupied by the cluster on the sky

$$\Delta_{ISZ} = \int \Delta T d\Omega \approx D_A^{-2} \int n_e \frac{k_B T}{m_e c^2} \sigma_T d\ell dA = \frac{k_B \sigma_T}{m_e c^2} D_A^{-2} \int n_e T dV \quad (16.6.4)$$

$$= \alpha N_e T_e D_A^{-2} = \alpha' M T_e D_A^{-2} \quad (16.6.5)$$

where N_e is the total number of electrons, T_e the density-averaged temperature, α, α' are constants independent of the cluster, M is the total mass of the cluster (assuming is universally proportional to N_e, i.e. to the gas mass) and D_A is the angular diameter distance of the cluster, which enters because by definition $d\Omega = dA/D_A^2$. So if we can estimate N_e or M we can obtain the distance to the cluster. In fact M can be independently estimated through X-ray observations (see Eq. 16.2.7) so we can use clusters also to map the cosmic expansion independently of supernovae or other distance estimators. The current results are however still not competitive.

An additional SZ effect, called kinetic SZ, is due to the Doppler shift induced on the CMB photons if the cluster is moving with velocity v_{pec} with respect to the CMB; the kinetic SZ is proportional to $v_{pec,\ell}/c$, where $v_{pec,\ell}$ is the projection along the line of sight. The kinetic effect at first order does not distort the black body spectrum and can therefore be distinguished from the thermal one.

Appendix: The Jeans equations

The virial theorem applies to averages over the entire system. If we wish to reconstruct the density profile of galaxies and clusters we need to consider the equilibrium dynamics in more detail.

A flow of particles that are neither destroyed nor created and do not collide against each other is governed by the collisionless Boltzmann equation. Stars in a smooth galactic gravitational potential, or galaxies in galaxy clusters, satisfy these conditions since they very rarely collide. Let the number dN of particles in a space volume dV and with velocities within a velocity volume $dv_x dv_y dv_z$ be

$$dN = f(t, \mathbf{x}, \mathbf{v}) dx dy dz dv_x dv_y dv_z \tag{16.6.6}$$

where f is the *distribution function* and the space-velocity volume element is called the phase-space volume. Since the particle number is conserved, the number of particles with velocity v_x entering in a space volume $dxdydz$ in direction x in the time interval dt is $\rho dxdydx = \rho v_x dtdydz$ where ρ is the number density of particles, minus the number exiting from the other side; if the particles move only along x, this has to be equal to the change in the number of particles inside the volume itself, $d\rho dV$. Then we have the continuity equation (the incoming velocity is taken negative)

$$d\rho dV = -d(\rho v_x) dt dy dz \tag{16.6.7}$$

or

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho v_x)}{\partial x} \tag{16.6.8}$$

For particle moving in any direction and also with any velocity, we have the general phase-space continuity equation

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{6} \frac{\partial(f \dot{w}_i)}{\partial w_i} = 0 \tag{16.6.9}$$

where $w = \{x, y, z, v_x, v_y, v_z\} = \{\mathbf{x}, \mathbf{v}\}$ is the phase-space vector of coordinates. Now we have $\dot{v}_i = -\frac{\partial \Phi}{\partial x_i}$ where Φ is the gravitational potential, and

$$\frac{\partial \dot{w}_i}{\partial w_i} = \{\frac{\partial v_i}{\partial x_i}, -\frac{\partial}{\partial v_i}\frac{\partial \Phi}{\partial x_i}\} = 0 \tag{16.6.10}$$

since the velocity vector is independent of the coordinates and the gravitational potential Φ is independent of the velocities. Then finally

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_{i=1}^{6} \dot{w} \frac{\partial f}{\partial w_i} = 0 \tag{16.6.11}$$

or the collisionless Boltzmann equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla \Phi \cdot \nabla_{\mathbf{v}} f = 0 \tag{16.6.12}$$

This equation can be written in any system of coordinates, eg cylindrical (useful for disk-like galaxies) or spherical (for spherical galaxies or clusters). In spherical coordinates we have

$$\frac{\partial f}{\partial t} + \dot{r}\frac{\partial f}{\partial r} + \dot{\theta}\frac{\partial f}{\partial \theta} + \dot{\phi}\frac{\partial f}{\partial \phi} + \dot{v}_r \frac{\partial f}{\partial v_r} + \dot{v}_\theta \frac{\partial f}{\partial v_\theta} + \dot{v}_\phi \frac{\partial f}{\partial v_\phi} = 0 \tag{16.6.13}$$

where we can also write $\dot{r} = v_r$, $\dot{\theta} = v_\theta / r$ and $\dot{\phi} = v_\phi / r \sin\theta$ (here we have defined the velocity components as the projection of the cartesian velocity along the directions r, θ, ϕ, respectively).

The relation of the accelerations in terms of the potential derivatives should be derived by transforming the corresponding cartesian expressions. Here we give directly the equations:

$$\dot{v}_r = \frac{v_\theta^2 + v_\phi^2}{r} - \frac{\partial \Phi}{\partial r} \tag{16.6.14}$$

$$\dot{v}_\theta = \frac{v_\phi^2 \cot\theta - v_r v_\theta}{r} - \frac{1}{r}\frac{\partial \Phi}{\partial \theta} \tag{16.6.15}$$

$$\dot{v}_\phi = \frac{-v_\phi v_r - v_\phi v_\theta \cot\theta}{r} - \frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi} \tag{16.6.16}$$

We assume from now on that the potential is purely spherical. Then we can write

$$\frac{\partial f}{\partial t}+v_r\frac{\partial f}{\partial r}+\frac{v_\theta}{r}\frac{\partial f}{\partial \theta}+\frac{v_\phi}{r\sin\theta}\frac{\partial f}{\partial \phi}+\left(\frac{v_\theta^2+v_\phi^2}{r}-\frac{\partial\Phi}{\partial r}\right)\frac{\partial f}{\partial v_r}+\left(\frac{v_\phi^2\cot\theta-v_rv_\theta}{r}\right)\frac{\partial f}{\partial v_\theta}+\left(\frac{-v_\phi v_r-v_\phi v_\theta\cot\theta}{r}\right)\frac{\partial f}{\partial v_\phi}=0 \tag{16.6.17}$$

Then we multiply by v_r and obtain

$$v_r\frac{\partial f}{\partial t}+v_r^2\frac{\partial f}{\partial r}+\frac{v_rv_\theta}{r}\frac{\partial f}{\partial \theta}+\frac{v_rv_\phi}{r\sin\theta}\frac{\partial f}{\partial \phi}+v_r\left(\frac{v_\theta^2+v_\phi^2}{r}-\frac{\partial\Phi}{\partial r}\right)\frac{\partial f}{\partial v_r}$$
$$+v_r\left(\frac{v_\phi^2\cot\theta-v_rv_\theta}{r}\right)\frac{\partial f}{\partial v_\theta}+v_r\left(\frac{-v_\phi v_r-v_\phi v_\theta\cot\theta}{r}\right)\frac{\partial f}{\partial v_\phi}=0 \tag{16.6.18}$$

Now we put $\nu(t,\mathbf{x})=\int f dv_r dv_\theta dv_\phi$, which is then a density independent of velocity, and integrate the above equation over the velocities. We have for instance terms like

$$\int\frac{\partial f}{\partial v_r}dv_rdv_\theta dv_\phi = \frac{\partial}{\partial v_r}\nu(\mathbf{x})=0 \tag{16.6.19}$$
$$\int v_r\frac{\partial f}{\partial r}dv_rdv_\theta dv_\phi = -\int\frac{\partial v_r}{\partial r}f dv_rdv_\theta dv_\phi=-\frac{\partial}{\partial r}\bar{v}_r \tag{16.6.20}$$
$$\int v_rv_\theta\frac{\partial f}{\partial \theta}dv_rdv_\theta dv_\phi = -\int\frac{\partial(v_rv_\theta)}{\partial\theta}f dv_rdv_\theta dv_\phi=-\frac{\partial}{\partial\theta}\sigma_{r\theta}+\frac{\partial}{\partial\theta}(\bar{v}_r\bar{v}_\theta) \tag{16.6.21}$$

(we applied integration by parts) where $\bar{v}_r$ is the average of v_r and

$$\sigma_{ij}^2=\int v_iv_jf dv_rdv_\theta dv_\phi-\bar{v}_i\bar{v}_j \tag{16.6.22}$$

is the velocity covariance. We now simplify our problem by assuming that: a) the system is stationary (all time derivatives vanish and $\bar{v}_r=0$; b) the kinematics has spherical symmetry: i.e. there is no rotation ($\bar{v}_\theta=\bar{v}_\phi=0$) and also $\sigma_{r\theta}^2=\sigma_{r\phi}^2=\sigma_{\theta\phi}^2=0$; c) finally, we also impose $\sigma_{\theta\theta}^2=\sigma_{\phi\phi}^2\equiv\sigma_t^2$. Then we have the Jeans equation

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma_{rr}^2)+2\frac{\sigma_{rr}^2-\sigma_t^2}{r}=-\frac{\partial\Phi}{\partial r}=-\frac{GM(r)}{r^2} \tag{16.6.23}$$

or

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma_{rr}^2)+2\sigma_{rr}^2\frac{\beta}{r}=-\frac{\partial\Phi}{\partial r}=-\frac{GM(r)}{r^2} \tag{16.6.24}$$

where

$$\beta=1-\frac{\sigma_t^2}{\sigma_{rr}^2} \tag{16.6.25}$$

is the anisotropy parameter, and it can take values in $-\infty<\beta\leq 1$. Notice that here the mass M includes all sources of gravity while the density ν might refer only to a particular population, eg stars or galaxies, and not necessarily to all the matter in the system. If velocities are isotropic, $\sigma_{rr}^2=\sigma_t^2$ we obtain the condition of spherical hydrostatic equilibrium

$$\frac{1}{\nu}\frac{\partial}{\partial r}(\nu\sigma_{rr}^2)=-\frac{GM(r)}{r^2} \tag{16.6.26}$$

The Jeans equation can be written in a form that directly gives the mass profile as a function of observables

$$M(r)=-\frac{r\sigma_{rr}^2}{G}\left[\frac{d\,\ln\nu}{d\,\ln r}+\frac{d\,\ln\sigma_{rr}^2}{d\,\ln r}+2\beta(r)\right] \tag{16.6.27}$$

Notice the similarity to the hydrostatic equation (16.2.4). However, the function $\beta(r)$ is very difficult to estimate, since we only directly measure the line-of-sight velocity.

Chapter 17

Observing the diffused gas

17.1 The 21cm line and the epoch of reionization

After the end of recombination around $z = 1000$, the cosmic hydrogen remained neutral until the UV photons emitted by the first stars began to ionize it at redshifts around 10-20 (see Fig. 17.2.1). Investigating this Epoch of Reionization (EoR) would be extremely useful to understand the early evolution of stars and galaxies. Neutral cold hydrogen is hard to observe because it has no energy levels in the optical nor in the radio, the two best windows for ground observations. However, the two possible spin states, parallel or antiparallel with respect to the proton, have a slightly different energy, the parallel state being higher by roughly $E_{01} = h\nu = 5 \cdot 10^{-6}$ eV, as first pointed out by Van de Hulst in 1942. A transition is accompanied by absorption or emission of a photon of wavelenght 21 cm. At redshift z, this becomes $21(1 + z)$ cm, so of the order of meters. Being a sharp monochromatic radio wavelength, the 21 cm line can easily be seen through dust and allows both very accurate galactic rotation curves and precise redshift measurement at very high redshifts. The presentation below is very sketchy, due to the complexity of the problem. Much more information can be found e.g. in Furlanetto et al., astro-ph/0608032.

In the following we use subscript 0 for the antiparallel singlet state (multiplicity $g_0 = 1$) and subscript 1 for the parallel triplet state ($g_1 = 3$). Neutral hydrogen before and during the EoR is excited mainly by the CMB radiation with temperature $T_\gamma = T_0(1 + z)$, where $T_0 = 2.7$K, which is hot enough to saturate the higher levels, $T_\gamma \gg T_S$, where we define a spin temperature T_S as the temperature one should have in thermodynamic equilibrium to maintain a density of population n_0, n_1 in the two states. We have then

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-\frac{E_{01}}{k_B T_S}} \approx \frac{g_1}{g_0} = 3 \tag{17.1.1}$$

because the spin temperature turns out to be close to T_γ. That is, three atoms out of four will be in the excited state. We can also write the total number density of HI atoms as $n_{HI} = n_0 + n_1 = 4n_0$.

To observe the 21 cm signal at the EoR, one compares at a given redshift z the radiation intensity of the 21 cm line with the CMB temperature. Depending on whether the spin temperature is higher or smaller that T_γ, one will see absorption or emission. In this way, it will be possible to create maps of HI emission or absorption at several z's. So far, this task is still to be achieved but several large-scale projects are being developed to realize it, the most ambitious being the Square Kilometer Array (SKA), to be set up in South Africa and Australia.

Let us now imagine a background radiation field T_γ whose brigthness or specific intensity is I_γ and a gas of HI at spin temperature T_S and brigthness I_S. (Brightness is related to the flux density S_ν as $S_\nu = \int I_\nu \cos\theta d\Omega$, while the total flux over all frequencies is $S = \int S_\nu d\nu$.) Since 21cm is much longer than the peak CMB radiation, the Rayleigh-Jeans region of the black body spectrum is a good approximation, and the brightness and the temperature are simply proportional, $T = Ic^2/2k_B\nu^2$. We can write the observed temperature at frequency ν by solving the radiative transfer equation along a radial coordinate s,

$$\frac{dI_\nu}{ds} = -\alpha(\nu) I_\nu + \alpha(\nu) I_S \tag{17.1.2}$$

which simply says that the intensity of a radiation field (here identified with the CMB radiation) along propagation decreases because of absorption and increases because of the medium emissivity I_S, assumed not to

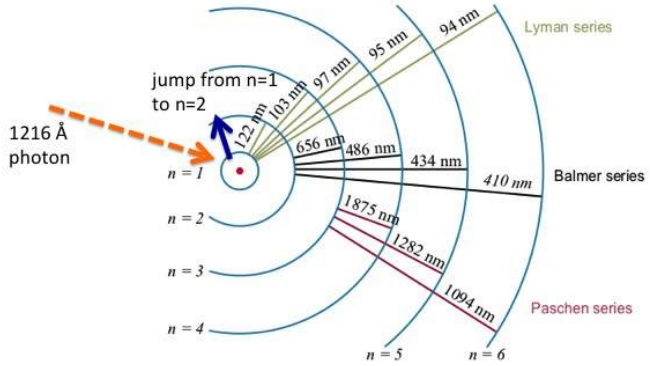


Figure 17.1.1: Absorption of a Lyman-α line (adapted from Wikicommons, authors OrangeDog and Szdori, license CC-BY-2.5).

depend on s. We define the optical depth

$$\tau = \int \alpha(\nu) ds \tag{17.1.3}$$

where $\alpha(\nu)$ is the absorption coefficient at frequency ν and ds integrates along the line of sight. (Note that all frequencies in what follows are rest frame values, not redshifted ones.) The solution for the observable brightness I_B is

$$I_B = I_S(1 - e^{-\tau}) + I_\gamma e^{-\tau} \tag{17.1.4}$$

Since in the Rayleigh-Jeans regime T is proportional to I, we can also write

$$T_B = T_S(1 - e^{-\tau}) + T_\gamma e^{-\tau} \tag{17.1.5}$$

where also T_B scales with redshift as $(1 + z)$. As intuitive, if the material is optically thick, $\tau \gg 1$, we see the spin temperature, if optically thin, $\tau \ll 1$, we see the CMB background. This relation can also be written as

$$T_{B0}(1 + z) = T_S(1 - e^{-\tau}) + T_0(1 + z)e^{-\tau} \tag{17.1.6}$$

and therefore

$$\delta T_B \equiv T_{B0} - T_0 = \frac{T_S - T_\gamma}{1 + z}(1 - e^{-\tau}) \tag{17.1.7}$$

The quantity δT_B, called brigthness, is the main observable: the difference between the observed and the CMB temperature at the desired frequency. Now we need to estimate τ and T_S.

The absorption coefficient gives the loss of energy per unit frequency of a radiation field resulting from absorption of photons $(0 \to 1)$ minus stimulated emission $(1 \to 0)$. Since τ is dimensionless, $\alpha(\nu)$ has dimensions m^{-1}. We need to introduce then the Einstein coefficients B_{ab} for the transition $a \to b$: the Einstein coefficient is the probability per unit time per unit spectral energy density, where the spectral energy is the energy density per unit frequency, so B_{ab} has units $J^{-1}m^3s^{-2}$. The rate (i.e. the probability per unit time) of absorption/emission is then given by $B_{ab}h\nu n\phi(\nu)$, where n is the number density of absorbers (atoms), $\phi(\nu)$ the line profile normalized

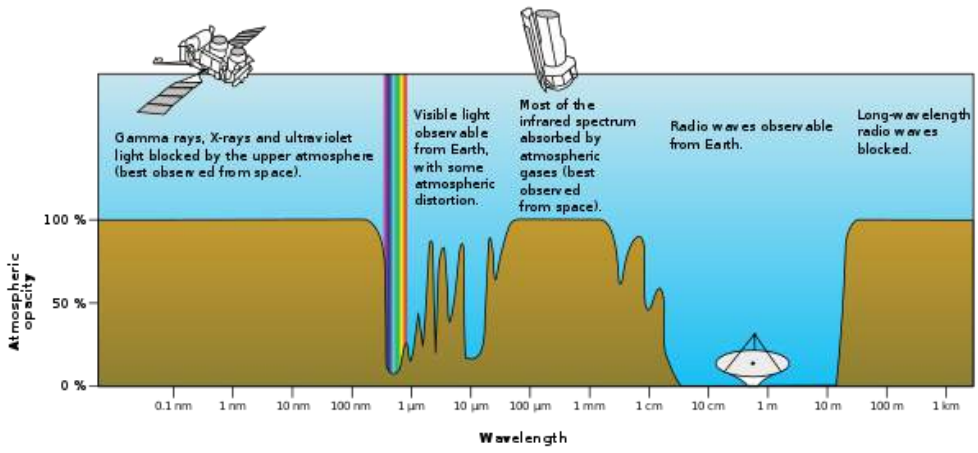


Figure 17.1.2: Windows in the atmospheric absorption. (Image: NASA)

so that $\int \phi(nu)d\nu = 1$ and $h\nu n\phi(\nu)d\nu$ is the radiation energy density in the range $d\nu$. Dividing the rate by c, one gets $\alpha_{ab} = B_{ab}h\nu n\phi(\nu)/c$, the probability of transition per unit distance. This quantity, summing all possible absorptions and substracting all possible emissions, defines the net absorption coefficient: once integrated over distance, it gives the optical depth.

Denoting with B_{01} the coefficient for absorption (from state 0 to state 1) and with B_{10} the emission one $(1 \to 0)$, the absorption coefficient is then given by the expression

$$\alpha(\nu) = \frac{h\nu}{c}(B_{01}n_0 - n_1 B_{10})\phi(\nu) \tag{17.1.8}$$

The Einstein coefficients are related by $B_{01}/B_{10} = g_1/g_0$ and therefore

$$\alpha(\nu) = \frac{h\nu}{c}n_0 B_{01}(1 - \frac{n_1}{n_0}\frac{B_{10}}{B_{01}})\phi(\nu) = \frac{h\nu}{c}n_0 B_{01}(1 - e^{-\frac{h\nu}{k_B T_S}})\phi(\nu) \approx \frac{h\nu}{c}n_0 B_{01}\frac{h\nu}{k_B T_S}\phi(\nu) \tag{17.1.9}$$

(we replaced E_{01} with $h\nu$). Therefore we find the dimensionless optical depth

$$\tau = \int \alpha(\nu)\phi(\nu)ds = \frac{h\nu}{c}B_{01}\frac{h\nu}{k_B T_S}\phi(\nu)\int n_0 ds = \frac{h\nu}{c}B_{01}\frac{h\nu}{k_B T_S}\frac{N_{HI}}{4}\phi(\nu) \tag{17.1.10}$$

where $N_{HI} = \int n_{HI}ds$ is the *column density* of HI atoms. The Einstein coefficients depend on the quantum mechanical properties of the atom. One can use now the standard relation among Einstein coefficients (see e.g. K. Lang, Astrophysical Formulae, Eq. 2-27)

$$B_{01} = \frac{3c^3 A_{10}}{8\pi h\nu^3} \tag{17.1.11}$$

where finally A_{10} is the spontaneous emission coefficient, estimated to be $A_{10} = 2.85 \times 10^{-15}\mathrm{s}^{-1}$ (i.e., one spin flip every 10 million years!).

The line profile is broadened, in general, by thermal motion, Doppler shift and cosmological recession. Here the dominating effect is the latter one, in which a frequency ν is spread over a range $\frac{\Delta\nu}{\nu} \approx \frac{\Delta V}{c}$ where ΔV is the velocity spread given by $\Delta V = sH(z)$. Therefore we can approximate $\phi(\nu) \approx 1/\Delta\nu \approx c/(sH(z)\nu)$. Then we have

$$\tau = A\frac{x_{HI}n_H}{T_S H(z)} \tag{17.1.12}$$

where $A = \frac{3c^3 hA_{10}}{32\pi\nu^2 k_B}$ and we write the column density across a distance s as $N_{HI} = x_{HI}n_H s$, $x_{HI}(z)$ being the ionization fraction and $n_H(z)$ the density of hydrogen atoms. We can use now the relation

$$n_H = (1+\delta_b)\rho_b = (1+\delta_b)\frac{3H_0^2}{8\pi G m_p}\Omega_{b,0}(1 - Y_{He})(1+z)^3 \tag{17.1.13}$$

(the factor of m_p, mass of the proton, is needed to convert the energy density ρ_b into the number density n_H) where δ_b is the fluctuation in the baryon content and $Y_{He} \approx 0.245$ is the fraction in mass in Helium (all the rest being hydrogen). We use also $H(z) = H_0\Omega_{m,0}^{1/2}(1+z)^{3/2}$, and finally we obtain from (17.1.7) and $\tau \ll 1$ the brigthness

$$\delta T_B \approx \frac{T_S - T_\gamma}{1+z}\tau = \frac{3}{8\pi G m_p} A(1 - \frac{T_\gamma}{T_S})\frac{x_{HI}(1+\delta_b)\Omega_{b,0}(1-Y_{He})H_0^2(1+z)^2}{T_S H(z)} \tag{17.1.14}$$

$$= \frac{3H_0 A(1-Y_{He})}{8\pi G m_p}(1 - \frac{T_\gamma}{T_S})x_{HI}(1+z)^{1/2}\frac{\Omega_{b,0}}{\Omega_{m,0}^{1/2}} \tag{17.1.15}$$

$$\approx 27 x_{HI}(1+\delta_b)(1 - \frac{T_\gamma}{T_S})(\frac{1+z}{10}\frac{0.15}{\Omega_{m,0}h^2})^{1/2}(\frac{\Omega_{b,0}h^2}{0.023})mK \tag{17.1.16}$$

where the last expression allows an immediate estimation of the order of magnitude, since all factors are of order unity. This expression can be further corrected by multiplying it by $H/(dv/dr + H)$, where dv/dr gives the contribution to the line width due to the peculiar velocities along the line of sight. If $T_S > T_\gamma$, the brigthness δT_B is positive (emission). Detection of δT_B can therefore map the history of reionization $x_{HI}(z)$, measure a combination of Ω_b and Ω_m, and map the spatial and time structure of the fluctuations in the baryon component δ_b.

To estimate T_S we should take into account three processes: 1) absorption and stimulated emission of CMB photons; 2) collisions with other hydrogen atoms, free electrons, and protons; and 3) Lyman-α photons (UV) scattering. We denote with C_{10}, P_{10} the de-excitation rates per atom from collisions and UV scattering, and with C_{01}, P_{01} the corresponding excitation rates. At equilibrium we have then

$$n_1(C_{10} + P_{10} + A_{10} + B_{10}I_\gamma) = n_0(C_{01} + P_{01} + A_{01} + B_{01}I_\gamma) \tag{17.1.17}$$

Now we can replace I_γ with T_γ using the Rayleigh-Jeans relation. Then we can invoke detailed equilibrium and write

$$\frac{C_{01}}{C_{10}} = \frac{g_1}{g_0}e^{-\frac{h\nu}{k_B T_K}} \approx 3(1 - \frac{h\nu}{k_B T_K}) \tag{17.1.18}$$

where T_K is the kinetic temperature, and similarly define an effective "color temperature" of the UV ionizing field T_c via

$$\frac{P_{01}}{P_{10}} \equiv 3(1 - \frac{h\nu}{k_B T_c}) \tag{17.1.19}$$

Putting all this into (17.1.17) we find

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_c T_K^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha} \tag{17.1.20}$$

where

$$x_c = \frac{C_{10}}{A_{10}}\frac{h\nu}{k_B T_\gamma} \tag{17.1.21}$$

$$x_\alpha = \frac{P_{10}}{A_{10}}\frac{h\nu}{k_B T_\gamma} \tag{17.1.22}$$

are the coupling coefficients for collisions and scattering, respectively. In the limit in which the couplings are small, $T_S = T_\gamma$ as one should expect. In most cases we can assume $T_c \approx T_\gamma$. The exact estimation of T_S can be achieved only numerically. Depending on which process dominates, one can have T_S larger or small than T_γ. During the most interesting phase of reionization, at redshifts around 10, one has positive brightness, as can be seen from Fig. (17.1.3).

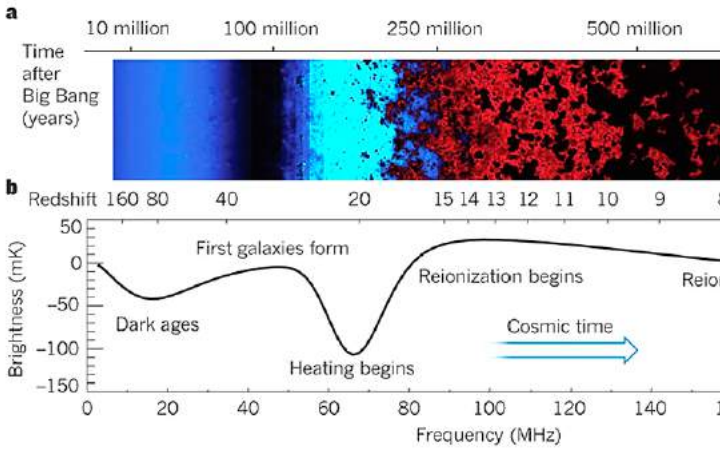


Figure 17.1.3: 21cm signal from ionized hydrogen (from Pritchard and Loeb, 2010, Nature 468, 772, reproduced with permission).

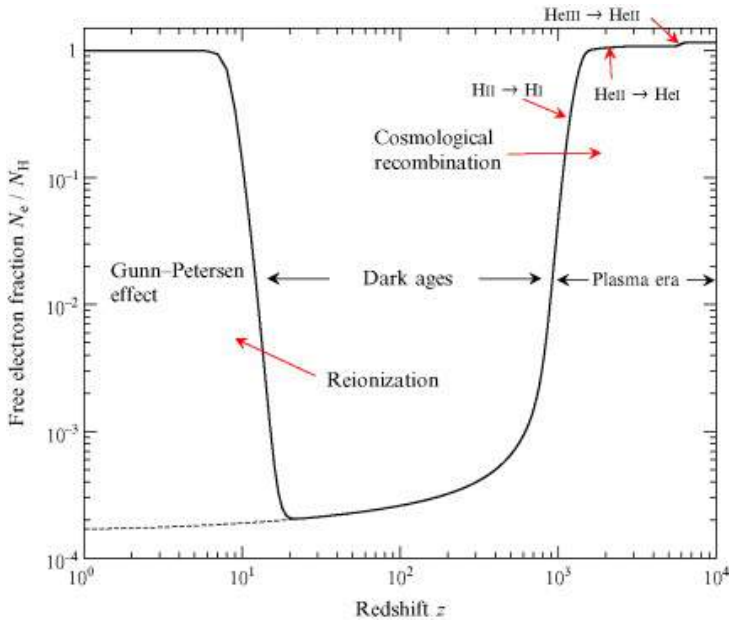


Figure 17.2.1: Reionization history (fraction of free electrons ≈ fraction of ionized hydrogen atoms). Image from Glover et al., *Atomic, Molecular, Optical Physics in the Early Universe, 2014,* adapted from Sunyaev and Chluba (RMA, 2009).

17.2 Lyman-α forest

The EoR ends around $z \approx 6$ but pockets of neutral hydrogen remain inside or nearby structures, e.g. along filaments, protected by the UV photons by their own high density. Since these clouds are associated with large scale structure, they can constitute a valuable way to reconstruct matter fluctuations at high redshifts. The clouds can be seen as Lyman-α absorbers (transition from $n = 2$ to $n = 1$ orbitals) along the line of sight of distant quasars, as in Fig. (12.7.1). Since the Lyman-α occurs at a rest frame wavelength of 1215 angstroms ($\approx$100 nm), it falls in the UV region. When redshifted by a factor of $1 + z$ equal to 3 to 6, therefore, falls in the visible part and can be detected from the ground. Distant quasars at redshifts larger than 3 show hundreds of these lines, each one generated by absorption of different clouds distributed with different redshifts along the line of sight. After $z \approx 6$, the end of the EoR, the lines merge into a continuous trough, called Gunn-Peterson trough: at earlier epochs the medium becomes more and more neutral.

Lyman alpha forest systems have a density of 10^{14} atoms per square centimeter. The so-called Lyman limit systems are 10^3 times denser, while damped Lyman alpha systems another factor of 10^3 more. Some recent surveys like the 2dF-SDSS quasar survey (2SLAQ) collected 100k quasars as of 2010, while BOSS DR9 obtains 180k quasars, 62k at redshifts >2.2. This high statistics allows the estimation of the 3D matter power spectrum from the 1D lines of sight: this is particularly useful both because the Lyman$-\alpha$ lines probe an epoch which is out of reach for today's redshift surveys and because at that epoch the fluctuations are linear down to a much smaller scale than today's structure (fluctuations at redshift z are roughly $(1 + z)$ times weaker than today). The power spectrum reconstruction however requires the use of N-body simulations to understand how well the Lyman$-\alpha$ forest traces the matter fluctuations. In Fig. (12.1.1) one can see how the Lyman$-\alpha$ reconstruction extends the reach of the power spectrum to small scales.

Chapter 18

Dark matter

We have mentioned several times the dark matter and have seen some of the evidence in favor of it. This chapter is devoted to the main observational techniques to "observe" dark matter beyond those already mentioned and to a short review of the candidates.

Quick summary

- Dark matter is very likely non-baryonic and composed of free particles.
- Several particle physics models of DM have been proposed but the so-called WIMPs (predicted in supersymmetry) are probably the most widely accepted candidate
- WIMPs of mass 10-100 GeV would have the right amount of density to explain all of the DM
- DM can be detected directly or indirectly, or even produced in accelerators, but so far no conclusive evidence has been reported
- Direct detection might occur when the galaxy halo DM particles cross the Earth and hit a nucleus in the underground detectors
- Indirect detection occurs if DM annihilates when encountering other DM particles in high-density regions (central Milky Way, dwarf galaxies, clusters etc) or decays: we can then observe the products as high energy monochromatic photons, neutrinos, charged leptons.
- Current cold DM models cannot explain easily the galaxy inner profiles and the number of dwarf satellites: this could be a problem related to baryonic physics or require a modification of the standard DM models.

18.1 Dark matter candidates

Dark matter can be composed in principle by any sort of matter that escapes direct observation in every electromagnetic waveband, from γ-rays to radio. However we know from big bang nucleosynthesis and CMB constraints that the baryons can only amount to 5% at most of the cosmic density, while we need around 30% of matter to explain spiral rotation curves, $X-$ray ICM, CMB, cosmic expansion etc. This DM must be cold enough to remain within galaxy and cluster halos: relativistic particles would easily escape the gravitational potentials. Moreover, the DM particles obviously have to have a mass sufficient to provide the required density. So the natural candidates are massive, cold, stable particles that do not interact electromagnetically. Since the existence of an abundant number of micro black holes and in general of unseen compact objects is severely limited by many constraints (for instance, limits coming from microlensing), it is likely that the DM is a gas of such particles freely floating around in galaxies and clusters.

DM particles that feel weak interaction (beside of course gravitational interaction) are called WIMPs (weakly interacting massive particles). In this case, one can estimate their present number density since we know the epoch when weak interacting particles left thermal equilibrium with the other particles (*freezing*) in the early

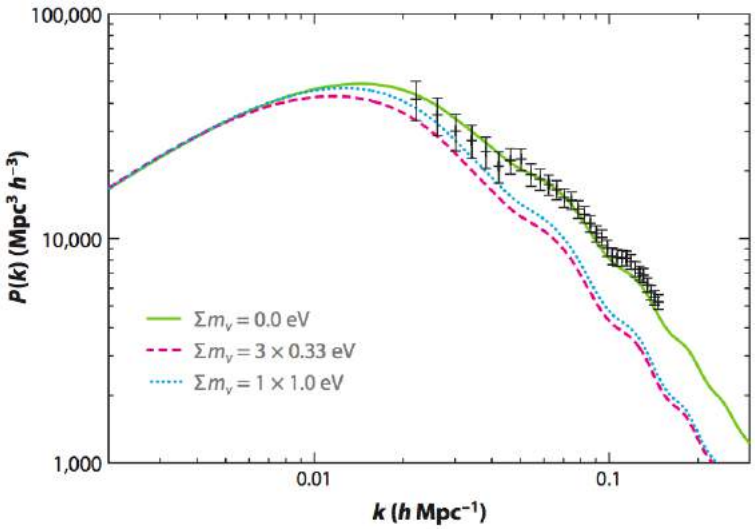


Figure 18.1.1: Matter power spectrum in models with massless and massive neutrinos (from Wong, Y., Ann.Rev. Nucl.Part.Sci. 61, 2011). The data points are from the 2dF survey (Cole et al. 2005).

universe. It turns out that particles with weak-scale mass around 100 GeV will have automatically more or less the density required to provide all the DM we need: this is called the WIMP miracle. This feature makes the WIMPs the perfect candidate for DM. However, so far no WIMP has ever been captured or produced.

Supersymmetric theories provide a host of new partner particles to the standard model, all with masses above several GeV (otherwise they would have been produced in accelerators; this does not apply to the gravitino however). The lightest supersymmetric particle must be stable because there are no symmetry-conserving particles they can decade to, so supersymmetry provide naturally at least one kind of WIMP, often identified with the neutralino o gravitino or, generically, LSP, "lightest sypersymmetric particle", whichever it be. Most experiments are therefore geared at capturing 10-100 GeV WIMPs passing through the Earth. The fact that the LHC is not finding new physics at these mass scales is generating tension with the WIMP-DM connection scenario.

Another dark matter candidate is the axion, a particle proposed in order to understand why QCD does not violate parity to a higher degree. Its mass is constrained to be very small, of the order of μeV, but it is still of the "cold" type because it has never been in thermal equilibrium with the other particles. Searches for the axion have been unsuccessful so far and large chunks of the interesting parameter space have been excluded.

Many other types of DM have been proposed, also of the "warm" type, for instance sterile neutrinos with mass around keV. One type of DM is certainly realized in nature: neutrinos. Since their mass is constrained to be below roughly 1eV, they are born relativistic (temperature of freezing in the early universe much higher than their mass) and might be becoming non-relativistic just now. Such a small mass implies that they cannot contribute to a large fraction of the DM, at most 1%. Moreover, their large velocity will not allow them to cluster along with baryonic matter so that they will in general decrease the overall matter clustering. In fact, by measuring accurately the matter power spectrum one can infer upper limits to the abundance, and therefore mass, of the neutrinos. In agreement with neutrino oscillation experiments, cosmology constrains the neutrino mass to be at or below the eV level (see Fig. 18.1.1).

18.2 Direct detection

If dark matter forms a giant halo around the bulge and disk of the Milky Way then the DM particles should cross the Earth and can be in principle detected through their weak interaction with nuclei. The density of the halo should be roughly 100 times the background density, i.e. around 0.3 proton mass per cm^3 so if the DM particles have a mass of say 100 GeV there should be one DM particle every liter of space. If the DM halo is slowly rotating, the Earth will feel a wind of DM particles at a velocity of 220 km/sec (the velocity of the Sun around the Milky Way center), with a little annual modulation along the Earth' orbit around the Sun, for a

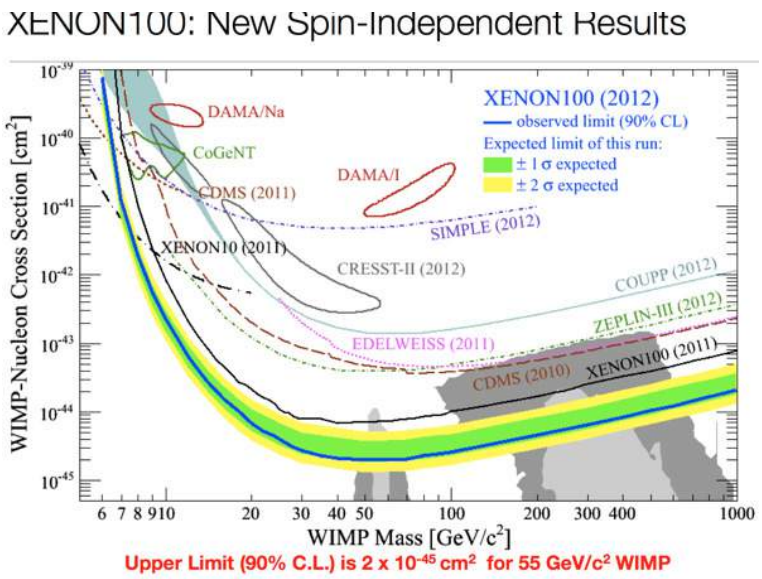


Figure 18.2.1: Typical exclusion plot cross section versus DM mass collecting the results of many experiments (XENON collaboration, 2012). The closed contours (DAMA, CoGent, CRESST) are positive detections, which appear therefore in contrast with the excluded regions from other experiments. (Credits: http://xenon.astro.columbia.edu/XENON100_Experiment/ and Phys. Rev. Lett. 109, 181301 (2012))

total rate of $10^{4\div5}$ particles per square centimeter per second.

Some of these non-relativistic particles will hit, although with a very low probability, the nuclei of the detectors currently located in various underground laboratories, shielded from the cosmic ray noise by kilometers of rock. A typical event rate could be up to a few events per day per kilogram of detector. Since we expect a good mass range to be around 10-100 GeV, the DM detectors normally employ nuclei with similar mass like Germanium, Silicon, Iodium in order to maximize the recoil. The energy of the recoil is in any case extremely small, up to 100 keV. The recoil can then be measured because, depending on the detector substance, it creates crystal dislocations, ionization, phonons, scintillations.

The main problem is separating true DM events from background, mostly due to natural radioactivity of the rocks and of the detector material itself. Another way is to search for the small annual modulation due to the Earth orbit: the experiment DAMA at Gran Sasso, Italy, claims since several years a positive detection but so far it has not received robust confirmation. Most experiments produced regions of excluded parameter space in the plane cross section versus mass (Fig. 18.2.1). Low cross sections and low masses are clearly more difficult to constrain and ample regions of parameter space are still unconstrained.

Another form of "almost direct" detection is DM production at high-energy accelerator like LHC at CERN. If the collision of two protons or other particles generates products that do not add up to the original energy, it means some particles have escaped detection. A careful series of tests on backgrounds and on already known elusive particles might in principle lead to the conclusion that the escaped particles are DM candidates.

18.3 Indirect detection

DM particles can be indirectly detected if we can observe signals that are linked to DM decay or self-annihilation. One possibility is that after many interactions some DM particles lose energy and are eventually gravitationally captured by the Sun (or even the Earth). If they accumulate for billion years in the core of the Sun their density might become large enough that their annihilation rate becomes significant. Neutrinos generated by the annihilation will be much more energetic than thermonuclear solar neutrinos and will impact the Earth atmosphere and rock, creating muons. If in a underground laboratory a highly energetic muon traveling *upward* is observed, it cannot come from cosmic rays hitting the atmosphere, so one possibility is that it comes from neutrinos generated in the Earth or Sun's core (when the Sun is below ground, i.e. at night) by DM annihilation.

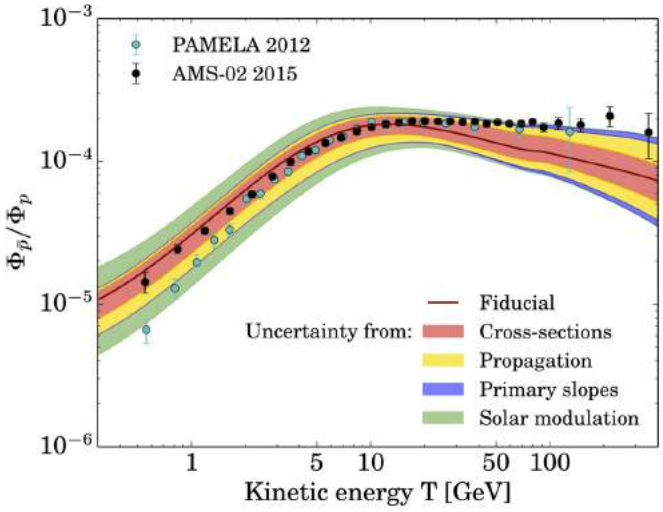


Figure 18.3.1: Excess of antiproton/proton ratio from the experiments PAMELA and AMS-02. The data are just above the uncertainty band (from Giesen et al. 2015, Journal of Cosmology and Astroparticle Physics, Issue 09, article id. 023, (2015)).

The high-energy neutrinos can also be directly captured by neutrino detectors like AMANDA and IceCube.

Another signal from annihilating or decaying DM can be obtained looking for 100 GeV-range γ-ray photons from space, perhaps from regions where we expect a high angular concentration of DM, for instance in the direction of M31 (Andromeda galaxy), the Large Magellanic Cloud, dwarf galaxies in the Local Group, the Milky Way center, galaxy clusters. Since the annihilation is a two-particle process, it is proportional to the DM density squared and this might help revealing the effect and distinguishing it from astrophysical backgrounds (for instance pulsars and supernovae remnants). The photons should have exactly the energy of the DM particles and therefore appear as a sharp line on the continuum coming from high-density regions (redshifted if coming from a moving source). Notwithstanding various temptative detections, so far the satellite Fermi-LAT has excluded WIMP masses up to 100 GeV. Cherenkov detectors, like HESS, MAGIC, VERITAS on Earth can also collect the particle showers generated from the impact of the high-energy photons with the atmosphere.

DM annihilation/decay products could also be pairs of charged fermions (protons, electrons and their antiparticles). The antiparticles will generate an excess over the standard astrophysical background particle/antiparticle ratio and could be revealed by the satellite PAMELA and the instrument AMS on board the ISS. Here again several detection claims have arisen but none has proven to be conclusive evidence of new physics (see Fig. 18.3.1).

18.4 The problems of the cold dark matter

Although the CDM is widely regarded as the most likely form of dark matter, it is not without problems. The main problem is the excess of small substructure we see (and expect to see) in N-body simulations with respect to what we observe in galaxies (Fig. 18.4.1). The high-mass end disagreement is not so critical because large halos will in general host groups or clusters instead of individual galaxies. Since CDM does not dissipate efficiently because weakly interacting, and because the theoretical power spectrum has no natural cut-off scale, the CDM lumps formed under gravitational instability live practically forever as satellites within the halos of large galaxies as ours. But around the Milky Way we observe only a few dozen dwarf galaxies, although we cannot easily estimate how many (perhaps hundreds) we are missing due to low surface brightness. Of course we do not know whether every DM lump should contain stars or instead just remain dark either because not enough baryons are collected or because they remain too hot to collapse into stars. So in fact the DM subclustering problems is probably just a manifestation of our uncertainties about baryonic physics, that is, of exactly how

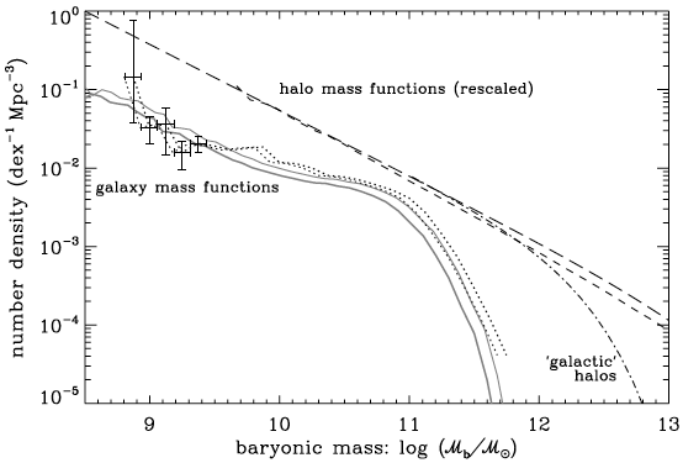


Figure 18.4.1: Halo mass function from N-body simulations (dahed lines) compared with the galaxy mass/luminosity function from real data (continuous and dotted lines). The discrepancy means that there is a non-linear relation between halos and galaxies. In particular, at the small mass end, the simulations show much more halos than galaxies (from Baldry et al., *On the galaxy stellar mass function, the mass–metallicity relation and the implied baryonic mass function*, Mon. Not. R. Astron. Soc. 388, 945–959 (2008)).

and when exactly stars finally form.

The clustering problem of DM arises at small scales and its consequence is not just an excess of satellites but also a inner halo profile that seem much steeper (cuspy) that the one that best fits galaxy rotation curves: this is called the cusp-core problem. Here again the DM evidence comes from N-body simulations, well fitted by the steep NFW profile (12.7.2), while observations refer to luminous sources, so baryonic physics might very likely play an important role. Alternatively, DM could depart from the simplest theoretical expectations and allow for a stronger self-interaction that introduces some form of dissipative friction that erases substructure.

Appendix

Probability distributions

Some well-known probability **distributions**.

Poisson distribution.

The Poisson distribution gives the probability to have N events when η are expected. For instance, if we throw at random balls in a set of boxes, so that the boxes contain on average η balls each, the Poisson distribution gives us the probability to find N balls in a given box. It is given by the PDF

$$P(N;\eta) = \frac{e^{-\eta}\eta^N}{N!} \tag{18.4.1}$$

Although the Poisson distribution is defined over integer numbers, we can approximate it as a continuous distribution for large η. The PDF is normalized to unity, and its average is of course η, while its variance is also η. For instance, if the average number of balls is 3, there is a 5% chance to get no balls in a given box. This gives a simple prescription for the variance to be expected in a distribution of particles, eg galaxies, when the process that created the distribution is supposed to be uniform: if we find N galaxies in a given volume, then we can estimate the variance as N, and therefore the error on N will be $\sqrt{N}$. This is the often employed Poissonian error.

Gaussian distribution

The Gaussian distribution is

$$P(x) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] \tag{18.4.2}$$

Its mean is x_0 and its variance is σ. The Gaussian distribution describes the distribution of a random variable when the variable is the result of many independent processes; for instance, a measure is subject to many experimental uncertainties, most of which are independent, and so the measure itself can be approximately distributed as a gaussian around its true value.

χ^2 distribution

The χ^2 distribution is

$$P(\chi^2;n) = \frac{(\chi^2)^{n/2-1}\exp(-\chi^2/2)}{2^{n/2}\Gamma(n/2)} \tag{18.4.3}$$

Its mean is n and its variance is $2n$. The χ^2 distribution is useful because a sum of n squared Gaussian variables is distributed as a χ^2 variable. Since a product of Gaussian distribution can be expressed as a function of the sum of squared Gaussian variables, the χ^2 PDF gives the resulting distribution.