Cosmology

Prof. Dr. Matthias Bartelmann
Institut für Theoretische Astrophysik
Universität Heidelberg

revised by
Prof. Dr. Christoph Pfrommer
Leibniz-Institut für Astrophysik Potsdam (AIP)
University of Potsdam

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Contents

Chapter 1

The Homogeneous Universe

1.1 Geometry and Dynamics

1.1.1 Assumptions

- Cosmology rests on two fundamental assumptions:

  1. When averaged over sufficiently large scales, the observable properties of the Universe are isotropic, i.e. independent of direction.

     It remains to be clarified what *sufficiently large* scales are. Nearby galaxies are very anisotropically distributed, distant galaxies approach isotropy, the microwave background is almost perfectly isotropic.

  2. Cosmological principle: Our position in the Universe is by no means preferred to any other.

     This reflects Copernican revolution of the world model, when it was realised that the Earth is not at the centre of the Universe.

  By the second assumption, the first must hold for every observer in the Universe. If the Universe is in fact isotropic around all of its points, it is also homogeneous. Thus, these two assumptions are often phrased as: *the Universe is homogeneous and isotropic*.

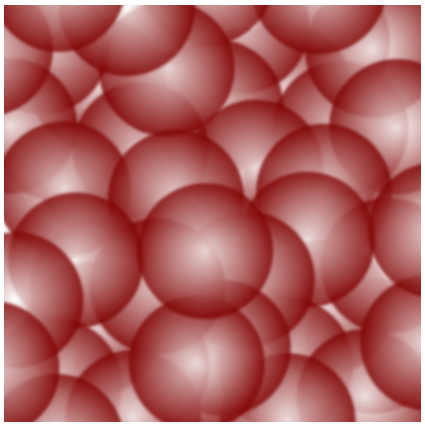
- These are bold assumptions, which have to be justified. An ideally homogeneous and isotropic universe would not allow us to exist. We need to carefully study how an idealised world model following from these two assumptions can accommodate structures.

- Of the four interactions (strong, weak, electromagnetic and gravitational), the strong and the weak interactions are limited to length scales typical for subatomic elementary-particle interactions. Electromagnetism is limited in range by the shielding of opposite charges, although magnetic fields can bridge very large scales. The remaining force relevant for cosmology is gravity.

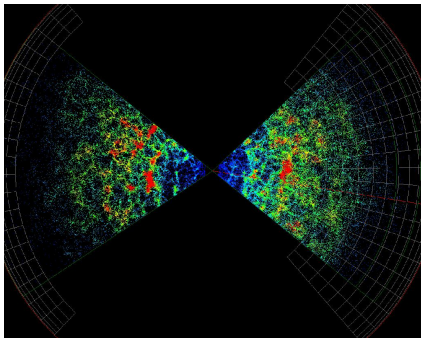
- Gravity is described by General Relativity. Newtonian gravity was constructed for isolated bodies and has fundamental difficulties in explaining space filled with homogeneous matter.

- General Relativity describes space-time as a four-dimensional manifold whose metric tensor $g_{\mu\nu}$ is a dynamical field. Its dynamics is governed by Einstein's field equations which couple the metric to the matter-energy content of space-time.

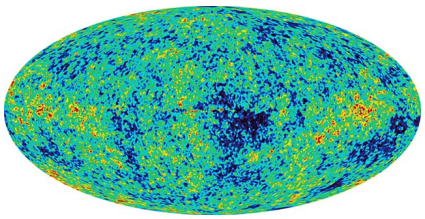
- As the structure of space-time determines the motion of matter and energy, which in turn determine the structure of space-time,



If the universe is isotropic about all points, it must be homogeneous.



The galaxy distribution is manifestly anisotropic...



... but the microwave background is phantastically isotropic.

general relativity is inevitably non-linear (in contrast to electrodynamics). Solutions of Einstein's field equations are thus typically very difficult to construct.

1.1.2 Metric

- Due to symmetry, the 4×4 tensor $g_{\mu\nu}$ has ten independent components, the time-time component g_{00}, the three space-time components g_{0i}, and the six space-space components g_{ij}. Greek indices run over space and time, i.e. they take on values $0 \leq \mu \leq 3$, while Latin indices indicate spatial components only, $1 \leq i \leq 3$.

- The two fundamental assumptions greatly simplify the metric. Phrased in a more precise language, they read:

  1. When averaged over sufficiently large scales, there exists a mean motion of matter and energy in the Universe with respect to which all observable properties are isotropic.
  2. All fundamental observers, i.e. imagined observers following this mean motion, experience the same history of the Universe, i.e. the same averaged observable properties, provided they set their clocks suitably.

- Consider the line element $\mathrm{d}s$,

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \ . \tag{1.1}$$

  Spatial coordinates attached to fundamental observers are called comoving coordinates. In such coordinates, $\mathrm{d}x^i = 0$ for fundamental observers. Requiring that their proper time equals the coordinate time $\mathrm{d}t$, we must have

$$\mathrm{d}s^2 = g_{00}\mathrm{d}t^2 = -c^2\mathrm{d}t^2 \ \Rightarrow \ g_{00} = -c^2 \ . \tag{1.2}$$

- Isotropy requires that clocks can be synchronised such that $g_{0i} = 0$. If that was impossible, the components of g_{0i} singled out a preferred direction in space, violating isotropy. Thus

$$g_{0i} = 0 \ . \tag{1.3}$$

- The line element is thus reduced to

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + g_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{1.4}$$

  Therefore, spacetime can be decomposed into spatial hypersurfaces of constant time, i.e. it permits a *foliation*. Without violating isotropy and homogeneity, the spatial hypersurfaces can then be scaled by a function $a(t)$ which may only depend on time,

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\mathrm{d}l^2 \ , \tag{1.5}$$

where $\mathrm{d}l$ is the line element of homogeneous and isotropic three-space. If the scale function depended on the spatial coordinates, the expansion could not maintain spatial homogeneity. A special case of (1.5) is Minkowski space, for which $\mathrm{d}l$ is the Euclidean line element.

- Isotropy requires three-space to have spherical symmetry. We thus introduce polar coordinates (w, θ, ϕ) where w is the radial coordinate and (θ, ϕ) are the polar angles:

$$\mathrm{d}l^2 = \mathrm{d}w^2 + f_K^2(w)\left[\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right] = \mathrm{d}w^2 + f_K^2(w)\mathrm{d}\omega^2 \ , \quad (1.6)$$

where $\mathrm{d}\omega$ is the solid-angle element. The radial function $f_K(w)$ is permitted because the relation between the radial coordinate w and the area of spheres of constant w is still arbitrary.

- The metric expressed by the line element (1.6) is manifestly isotropic. It can be shown that homogeneity requires $f_K(w)$ to be trigonometric, hyperbolic, or linear in w,

$$f_K(w) = \begin{cases} K^{-1/2}\sin(K^{1/2}w) & (K > 0) \\ w & (K = 0) \\ |K|^{-1/2}\sinh(|K|^{1/2}w) & (K < 0) \end{cases} \ , \quad (1.7)$$

where K is a constant parameterising the curvature of spatial hypersurfaces. $f_K(w)$ and $|K|^{-1/2}$ have the dimension of a length.

- An alternative form of the line element $\mathrm{d}s$ is obtained substituting the radial coordinate by r for $f_K(w)$, then

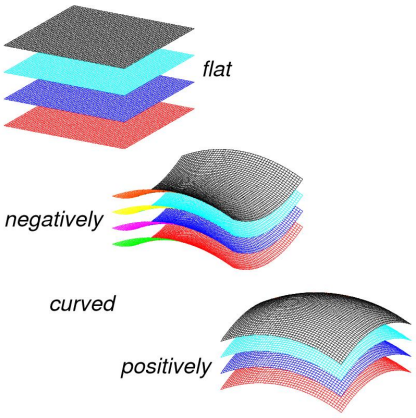
$$\mathrm{d}l^2 = \frac{\mathrm{d}r^2}{1 - Kr^2} + r^2\mathrm{d}\omega^2 \ . \quad (1.8)$$

this is often used, but has the disadvantage of becoming singular for $K > 0$ and $r = K^{-1/2}$.

- We thus arrive at the metric for the homogeneous and isotropic universe,

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left[\mathrm{d}w^2 + f_K^2(w)\mathrm{d}\omega^2\right] \quad (1.9)$$

with $f_K(w)$ given by (1.7). This is called Robertson-Walker metric.



The space-time of the universe can be foliated into flat or positively or negatively curved spatial hypersurfaces.

1.1.3 Cosmological Redshift

- Spatial hypersurfaces can expand or shrink controlled by the scale function $a(t)$. This leads to a red- or blueshift of photons propagating through space-time.

- Consider light emitted from a comoving source at time t_e reaching a comoving observer at $w = 0$ at time t_o. Since $ds = 0$ for light, the metric (1.9) requires

$$c|dt| = a(t)dw \,, \tag{1.10}$$

where the modulus on the left-hand side indicates that time can run with or against w, depending on whether w is measured towards or from the observer.

- The coordinate distance between source and observer is

$$w_{eo} = \int_{t_e}^{t_o} dw = \int_{t_e}^{t_o} \frac{c dt}{a(t)} = \text{const.} \,, \tag{1.11}$$

thus the derivative of w_{eo} with respect to the emission time t_e must vanish,

$$\frac{dw_{eo}}{dt_e} = \frac{c}{a(t_o)}\frac{dt_o}{dt_e} - \frac{c}{a(t_e)} \;\Rightarrow\; \frac{dt_o}{dt_e} = \frac{a_o}{a_e} \,. \tag{1.12}$$

- Time intervals dt_e at the source are thus changed until they arrive at the observer in proportion to changes in the scale of the universe between emission and absorption.

- Let $dt = \nu^{-1}$ be the cycle time of a light wave, then

$$\frac{\nu_e}{\nu_o} = \frac{\lambda_o}{\lambda_e} = 1 + \frac{\lambda_o - \lambda_e}{\lambda_e} = 1 + z = \frac{a(t_o)}{a(t_e)} \,. \tag{1.13}$$

Thus, light is red- or blueshifted by the same amount as the Universe expanded or shrunk between emission and observation. Conventionally, the scale factor at the time of observation is set to unity, $a(t_o) = 1$, and the scale factor at emission abbreviated by $a(t_e) =: a$. Then,

$$a = \frac{1}{1+z} \,, \quad z = \frac{1}{a} - 1 \,. \tag{1.14}$$

1.1.4 Dynamics

- The dynamics of the metric (1.9) is reduced to the dynamics of the scale factor $a(t)$. Differential equations for $a(t)$ now follow from Einstein's field equations, which read

$$G_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta} + \Lambda g_{\alpha\beta} \,. \tag{1.15}$$

Λ is the cosmological constant originally introduced by Einstein in order to allow static cosmological models.

- $G_{\alpha\beta}$ is the Einstein tensor constructed from the curvature tensor, which depends on the metric tensor and its first and second derivatives.

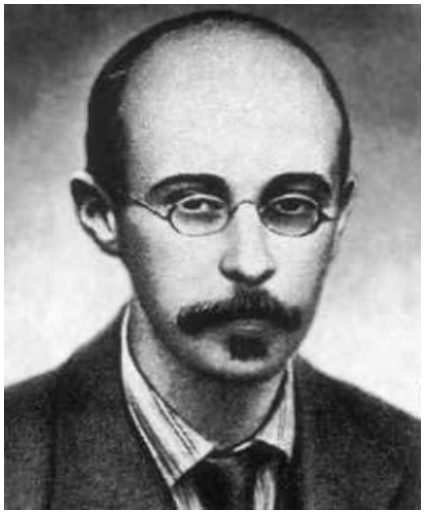
- $T_{\alpha\beta}$ is the energy-momentum tensor of the cosmic fluid, which must be of the form of the stress-energy tensor of a perfect fluid, characterised by pressure p and (energy) density ρ, which can only be functions of time because of homogeneity,

$$p = p(t) , \quad \rho = \rho(t) . \tag{1.16}$$

- When specialised to the metric (1.9), Einstein's equations (1.15) reduce to two differential equations for the scale factor $a(t)$ (see e.g. Chapter 12 in the script on General Relativity):

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3} \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} . \end{aligned} \tag{1.17}$$

  These are Friedmann's equations. A Robertson-Walker metric whose scale factor satisfies (1.17) is called Friedmann-Lemaître-Robertson-Walker metric. The scale factor is uniquely determined once its value at a fixed time t is chosen. We set $a = 1$ today.



Alexander Friedmann

- The Friedmann equations can be combined to yield the adiabatic equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\rho c^2\right) + p\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\right) = 0 \tag{1.18}$$

  which intuitively states energy conservation: the left-hand side is the change in internal energy, the right-hand side is the pressure work. This is the first law of thermodynamics in absence of heat flow (which would violate isotropy).

- Since energy conservation (1.18) follows from the Friedmann equations (1.17), any two equations from (1.17) and (1.18) can be used equivalently to all three of them. We follow common practise and use the first-order equation from (1.17), which we will call *the* Friedmann equation henceforth, and (1.18) where needed.

Georges Lemaître

1.1.5 Remark on Newtonian Dynamics

- Note that (1.17) can also be derived from Newtonian gravity, except for the Λ term. The argument runs like this: in a homogeneous and isotropic universe, a spherical region of radius R can be identified around an arbitrary point, the matter density within that sphere must be homogeneous. The matter surrounding the sphere cannot have any influence on its dynamics because it would have to pull into some direction, which would violate isotropy. Thus, the size of the sphere is arbitrary.

- Suppose now a test mass m is located on the boundary of the sphere. It's equation of motion is

$$\ddot{r} = -\frac{G}{r^2}\left(\frac{4\pi}{3}r^3\rho\right) = -\frac{4\pi G}{3}r\rho \ . \tag{1.19}$$

  this is already the second eq. (1.17) except for the pressure term.

- The pressure term adds to the density because pressure is a consequence of particle motion, i.e. the kinetic energy of particles, which is equivalent to a mass density and thus acts gravitationally. For particles with a mean squared velocity $\langle v^2\rangle$ we have

$$p = \frac{\rho}{3}\left\langle v^2\right\rangle = \frac{1}{3}E_{\rm kin} \ \Rightarrow \ \rho_{\rm p} = \frac{E_{\rm kin}}{c^2} = \frac{3p}{c^2} \ , \tag{1.20}$$

  thus the pressure adds an equivalent mass density $\rho_{\rm p}$, which we have to add to ρ; (1.19) thus reads

$$\ddot{r} = -\frac{4\pi G}{3}r\left(\rho + \frac{3p}{c^2}\right) \ . \tag{1.21}$$

- In analogy to (1.18), energy conservation requires

$$3r^2\dot{r}\rho c^2 + r^3\dot{\rho}c^2 = 3pr^2\dot{r} \ . \tag{1.22}$$

  Dividing by r and combining terms yields

$$2r\dot{r}\rho + \left(\rho + \frac{3p}{c^2}\right)r\dot{r} + r^2\dot{\rho} = 0 \ . \tag{1.23}$$

  Eliminating the term in brackets with (1.21) yields

$$2\dot{r}\ddot{r} = \frac{8\pi G}{3}(2r\dot{r}\rho + r^2\dot{\rho}) \ \Rightarrow \ \frac{{\rm d}(\dot{r}^2)}{{\rm d}t} = \frac{8\pi G}{3}\frac{{\rm d}(\rho r^2)}{{\rm d}t} \ . \tag{1.24}$$

- Integrating, we find

$$\left(\frac{\dot{r}}{r}\right)^2 = \frac{8\pi G}{3}\rho + \frac{C}{c^2} \tag{1.25}$$

  with a constant of integration C. Putting $K = -C/c^2$ yields the first eq. (1.17) without the Λ term.

- We thus find that Friedmann's equations can be derived from Newtonian dynamics if we account for the mass density equivalent to the energy density related to pressure and solve the equation of motion of a self-gravitating homogeneous sphere taking energy conservation into account. The Λ term is purely relativistic.

1.2 Parameters, Age and Distances

1.2.1 Forms of Matter

- Two forms of matter can broadly be distinguished, relativistic and non-relativistic. They are often called radiation and dust, respectively.

- For relativistic bosons and fermions, the pressure is

$$p = \frac{\rho c^2}{3} \tag{1.26}$$

  while non-relativistic matter is well approximated as pressure-free, $p = 0$, because the pressure is much smaller than the rest-mass energy ρc^2 it needs to be compared with.

- For non-relativistic matter, (1.18) reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\rho c^2\right) = 0 \;\Rightarrow\; \frac{\dot{\rho}}{\rho} = -3\frac{\dot{a}}{a} \tag{1.27}$$

  which implies

$$\rho(t) = \rho_0 a^{-3}\,, \tag{1.28}$$

  with the present density ρ_0 and using the convention that $a = 1$ today. This behaviour simply reflects the fact that the density of non-relativistic matter is decreasing because of dilution as space is expanding.

- For relativistic matter, (1.18) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\rho c^2\right) + \frac{\rho c^2}{3}\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\right) = 0 \;\Rightarrow\; \frac{\dot{\rho}}{\rho} = -4\frac{\dot{a}}{a} \tag{1.29}$$

  implying

$$\rho(t) = \rho_0 a^{-4} \tag{1.30}$$

  From (1.28) and (1.30) we see that the density of relativistic particles drops faster by one more power of a because particles are diluted and lose energy as they are redshifted.

- We have thus exploited the adiabatic equation for deriving the dependence of density on the scale factor for non-relativistic and relativistic matter. Inserting (1.28) and (1.30) into the Friedmann equation as appropriate, we thus obtain a single equation for the dynamics of the scale factor.

1.2.2 Parameters

- It is convenient to introduce parameters, most of which are dimension-less. The *Hubble parameter* is defined as the relative expansion rate,

$$H(t) := \frac{\dot{a}}{a} , \quad H_0 := H(t_0) . \tag{1.31}$$

Its value at the present time t_0 is the *Hubble constant*. It has the unit of an inverse time, but is commonly expressed in units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ because it quantifies by how much the recession velocity of cosmic objects grows as their distance increases. The Hubble constant is frequently expressed by the dimension-less parameter h,

$$H_0 = 100\,h\,\frac{\mathrm{km}}{\mathrm{s\,Mpc}} = 3.22 \times 10^{-18}\,h\,\mathrm{s^{-1}} . \tag{1.32}$$

- The inverse of the Hubble constant is the Hubble time,

$$t_\mathrm{H} := \frac{1}{H_0} = 3.10 \times 10^{17}\,h^{-1}\,\mathrm{s} = 9.82 \times 10^{9}\,h^{-1}\,\mathrm{yr} . \tag{1.33}$$

The Hubble time times the speed of light is the Hubble radius,

$$r_\mathrm{H} := \frac{c}{H_0} = 9.30 \times 10^{27}\,h^{-1}\,\mathrm{cm} = 3.01 \times 10^{3}\,h^{-1}\,\mathrm{Mpc} . \tag{1.34}$$

- The critical density is defined as

$$\rho_\mathrm{cr}(t) := \frac{3H^2(t)}{8\pi G} , \quad \rho_\mathrm{cr0} := \rho_\mathrm{cr}(t_0) = \frac{3H_0^2}{8\pi G} . \tag{1.35}$$

Writing it in the form

$$\frac{4\pi G}{3}\left(\frac{\rho_\mathrm{cr} a^3}{a}\right) = \frac{\dot{a}^2}{2} \tag{1.36}$$

illustrates that in a sphere filled with matter of critical density the gravitational potential is exactly balanced by the specific kinetic energy.

- The critical density today is

$$\rho_\mathrm{cr0} = 1.86 \times 10^{-29}\,h^2\,\mathrm{g\,cm^{-3}} , \tag{1.37}$$

corresponding to one proton mass in approximately $10^5\,\mathrm{cm}^3$ of the cosmic volume, or about a galaxy mass per Mpc^3.

- Densities expressed in units of the critical density are the dimension-less density parameters

$$\Omega(t) := \frac{\rho(t)}{\rho_\mathrm{cr}(t)} , \quad \Omega_0 := \Omega(t_0) = \frac{\rho(t_0)}{\rho_\mathrm{cr0}} . \tag{1.38}$$

- The density parameter corresponding to the cosmological constant, also often called cosmological constant, is

$$\Omega_\Lambda(t) = \frac{\Lambda c^2}{3H^2(t)} , \quad \Omega_{\Lambda 0} := \Omega_\Lambda(t_0) = \frac{\Lambda c^2}{3H_0^2} . \tag{1.39}$$

- Distinguishing the densities of radiation, ρ_R, and non-relativistic matter, ρ_M, we introduce the two density parameters

$$\Omega_{\mathrm{r}0} = \frac{\rho_{\mathrm{r}0}}{\rho_{\mathrm{cr}0}} , \quad \Omega_{\mathrm{m}0} = \frac{\rho_{\mathrm{m}0}}{\rho_{\mathrm{cr}0}} . \tag{1.40}$$

  Using (1.28) and (1.30) yields

$$\rho_\mathrm{r} = \Omega_{\mathrm{r}0}\rho_{\mathrm{cr}0}a^{-4} , \quad \rho_\mathrm{m} = \Omega_{\mathrm{m}0}\rho_{\mathrm{cr}0}a^{-3} . \tag{1.41}$$

- Replacing $\rho \to (\rho_\mathrm{r} + \rho_\mathrm{m})$ in Friedmann's equation then yields

$$H^2(a) = H_0^2\left[\Omega_{\mathrm{r}0}a^{-4} + \Omega_{\mathrm{m}0}a^{-3} + \Omega_{\Lambda 0} - \frac{Kc^2}{a^2H_0^2}\right] . \tag{1.42}$$

  Specialising to $a = 1$, we have $H^2(a = 1) = H_0^2$ on the left-hand side and solving for the K-dependent term, we find

$$-\frac{Kc^2}{H_0^2} = 1 - \Omega_{\mathrm{r}0} - \Omega_{\mathrm{m}0} - \Omega_{\Lambda 0} =: \Omega_\mathrm{K} , \tag{1.43}$$

  the curvature parameter.

- We thus arrive at the final form for Friedmann's equation

$$\begin{aligned} H^2(a) &= H_0^2\left[\Omega_{\mathrm{r}0}a^{-4} + \Omega_{\mathrm{m}0}a^{-3} + \Omega_{\Lambda 0} + \Omega_\mathrm{K}a^{-2}\right] \\ &=: H_0^2E^2(a) . \end{aligned} \tag{1.44}$$

  It is mostly used in this form for practical calculations.

- Note that all density contributions in square brackets scale with different powers of a. Their relative importance thus changes over time. Today, the radiation density is much smaller than the matter density. However, going back in time, the radiation density grows faster than the matter density, so there is a time t_eq before which radiation dominates. Expressing t_eq by the scale factor a_eq, we have from (1.41)

$$a_\mathrm{eq} = \frac{\Omega_{\mathrm{r}0}}{\Omega_{\mathrm{m}0}} . \tag{1.45}$$

  Before that, the universe is called radiation-dominated. Later, matter dominates while curvature is still negligible. Finally curvature becomes important and Ω_Λ may take over.

- The density parameters change with time. Ignoring radiation density, one has for non-relativistic matter

$$\Omega_{\rm m}(a) = \frac{8\pi G}{3H^2(a)}\rho_{\rm m0}a^{-3} = \frac{\Omega_{\rm m0}}{a + \Omega_{\rm m0}(1-a) + \Omega_{\Lambda 0}(a^3 - a)} \,, \tag{1.46}$$

and for the density parameter corresponding to the cosmological constant

$$\Omega_\Lambda(a) = \frac{\Lambda}{3H^2(a)} = \frac{\Omega_{\Lambda 0}a^3}{a + \Omega_{\rm m0}(1-a) + \Omega_{\Lambda 0}(a^3 - a)} \,. \tag{1.47}$$

- Two interesting consequences follow from eqs. (1.46) and (1.47): first, they imply $\Omega_{\rm m}(a) \to 1$ and $\Omega_\Lambda(a) \to 0$ for $a \to 0$ regardless of their present values $\Omega_{\rm m0}$ and $\Omega_{\Lambda 0}$. Second, if $\Omega_{\rm m0} + \Omega_{\Lambda 0} = 1$, this remains valid for $a < 1$.

1.2.3 Parameter Values

- The cosmological parameters, most notably H_0, $\Omega_{\rm m0}$ and $\Omega_{\Lambda 0}$, were highly insecure for most of the last century. Only recently, the situation has much improved mainly because of the microwave-background measurements and wide-field galaxy surveys like the 2-Degree-Field (2dF) survey and Sloan Digital Sky Survey (SDSS).

- Combining cosmic microwave-background measurements with baryonic acoustic oscillations and measurements of the Hubble constant, the cosmological parameters are now constrained as follows (all errors are 1-σ error margins):

Hubble constant	h	$0.704^{+0.013}_{-0.014}$
dark-matter density	$\Omega_{\rm d0}$	0.227 ± 0.014
cosmological constant	$\Omega_{\Lambda 0}$	$0.728^{+0.015}_{-0.016}$
baryon density	$\Omega_{\rm B}$	0.0456 ± 0.0016
radiation density	$\Omega_{\rm r0}$	$(4.67 \pm 0.26) \times 10^{-5}$

- Since $\Omega_{\rm K}$ is very close to zero, we will assume $\Omega_{\rm K} = 0$ in most of what follows.

- The Hubble constant is

$$H_0 = 73 \pm 2\,{\rm km\,s^{-1}\,Mpc^{-1}} = (2.35 \pm 0.06) \times 10^{-18}\,{\rm s^{-1}} \,, \tag{1.48}$$

i.e. the Hubble time is

$$\frac{1}{H_0} = (4.25 \pm 0.12) \times 10^{17}\,{\rm s} = (1.35 \pm 0.04) \times 10^{10}\,{\rm yr} \,. \tag{1.49}$$

- From (1.45), the scale factor at matter-radiation equality is

$$a_{\rm eq} = (1.95 \pm 0.19) \times 10^{-4} \,. \tag{1.50}$$

1.2.4 Age and Expansion of the Universe

- Since $H = \dot{a}/a$, the age of the Universe is determined by

$$\frac{\mathrm{d}a}{\mathrm{d}t} = H_0 a E(a) \;\Rightarrow\; H_0 t = \int_0^a \frac{\mathrm{d}a'}{a' E(a')} \,, \tag{1.51}$$

  as follows from Friedmann's equation (1.43) where we have assumed that time starts running when $a = 0$. This integral cannot generally be solved analytically, but limiting cases are interesting to study.

- Early Universe: in the early Universe, radiation dominates because its contribution scales with a^{-4} in Friedmann's equation. During that time, $E(a) = \Omega_{\mathrm{r0}}^{1/2} a^{-2}$ and

$$H_0 t = \frac{a^2}{2\sqrt{\Omega_{\mathrm{r0}}}} \;\Leftrightarrow\; a = \left[2\sqrt{\Omega_{\mathrm{r0}}} H_0 t\right]^{1/2} \,. \tag{1.52}$$

  Thus, at early times, the expansion of the Universe scales like $a \propto \sqrt{t}$ until the radiation density drops near the density of non-relativistic matter. At matter-radiation equality, the age of the universe is

$$t_{\mathrm{eq}} = 1.65 \times 10^{10}\,\mathrm{s} = 522\,\mathrm{yr} \,. \tag{1.53}$$

- Early matter-dominated era: after non-relativistic matter starts dominating, and before curvature becomes important, we may approximate $E(a) = \sqrt{\Omega_{\mathrm{m0}}} a^{-3/2}$ and obtain

$$H_0 t = \frac{2a^{3/2}}{3\sqrt{\Omega_{\mathrm{m0}}}} \;\Leftrightarrow\; a = \left[\frac{3}{2}\sqrt{\Omega_{\mathrm{m0}}} H_0 t\right]^{2/3} \,. \tag{1.54}$$

  Thus the expansion scales like $a \propto t^{2/3}$. This case is called the Einstein-de Sitter limit and is often used in studies of the early universe.

- Very late Universe: if $\Omega_\Lambda \neq 0$, it dominates at late times. Then, $E(a) = \sqrt{\Omega_\Lambda}$ and

$$H_0 t = \frac{\ln a}{\sqrt{\Omega_\Lambda}} \;\Rightarrow\; a \propto \exp\left[\sqrt{\Omega_\Lambda} H_0 t\right] \,, \tag{1.55}$$

  where we have ignored the lower integration limit because the approximation of a dominating cosmological constant is only valid after finite time. Then, the Universe expands exponentially, i.e. the cosmological constant is driving the Universe exponentially apart. This case is called the de Sitter limit and plays and important role in the theory of cosmological inflation.

- We have seen in (1.53) that the period of radiation domination is brief. For most of the cosmic time, radiation is negligible and matter, cosmological constant and curvature co-exist in comparable densities. We shall now study a few interesting simplified cases ignoring the contribution from the radiation density.

- Einstein-de Sitter universe: if $\Omega_\Lambda = 0$ and $\Omega_{\rm m0} = 1$, (1.51) holds throughout cosmic history, and

$$H_0 t = \frac{2}{3} a^{3/2} \;\Leftrightarrow\; a = \left(\frac{3}{2} H_0 t\right)^{2/3} . \tag{1.56}$$

  The age of such a Universe today is

$$t_0 = \frac{2}{3H_0} = 8.97 \times 10^9 \,\mathrm{yr} . \tag{1.57}$$

  This case is historically important.

- In a flat universe with $\Omega_{\rm m0} \neq 0$ and $\Omega_\Lambda = 1 - \Omega_{\rm m0} \neq 0$, the curvature term vanishes and

$$H_0 t = \int_0^a \frac{\sqrt{a'}\,\mathrm{d}a'}{\sqrt{\Omega_{\rm m0} + \Omega_\Lambda a'^3}} . \tag{1.58}$$

  This can be integrated substituting $x := a^{3/2}$ and yields

$$H_0 t = \frac{2}{3\sqrt{1-\Omega_{\rm m0}}} \,\mathrm{arcsinh}\left[\sqrt{\frac{1-\Omega_{\rm m0}}{\Omega_{\rm m0}}} a^{3/2}\right] . \tag{1.59}$$

  The age of a universe with $\Omega_{\rm m0} = 0.24$ is

$$t(a=1) = \frac{1.03}{H_0} = 1.40 \times 10^{10} \,\mathrm{yr} . \tag{1.60}$$

- The expansion of the spatially flat model becomes exponential when

$$\sqrt{\frac{1-\Omega_{\rm m0}}{\Omega_{\rm m0}}} a^{3/2} \gtrsim 1 \;\Rightarrow\; a \gtrsim \left(\frac{\Omega_{\rm m0}}{1-\Omega_{\rm m0}}\right)^{1/3} \approx 0.68 \tag{1.61}$$

- As (1.55) shows, a universe expanding with H_0 today may never reach $a = 0$ going back in time. The fact that the universe is expanding today does thus not imply that it originated in a Big Bang!

- However, it is straightforward to see that there must have been a Big Bang because we know from the existence of the microwave background that the radiation density is finite, from the existence of luminous material that the matter density is finite, and from the existence of objects with very high redshifts z that the scale factor of the universe must have been as small as $1/(1+z)$ or smaller in the past.

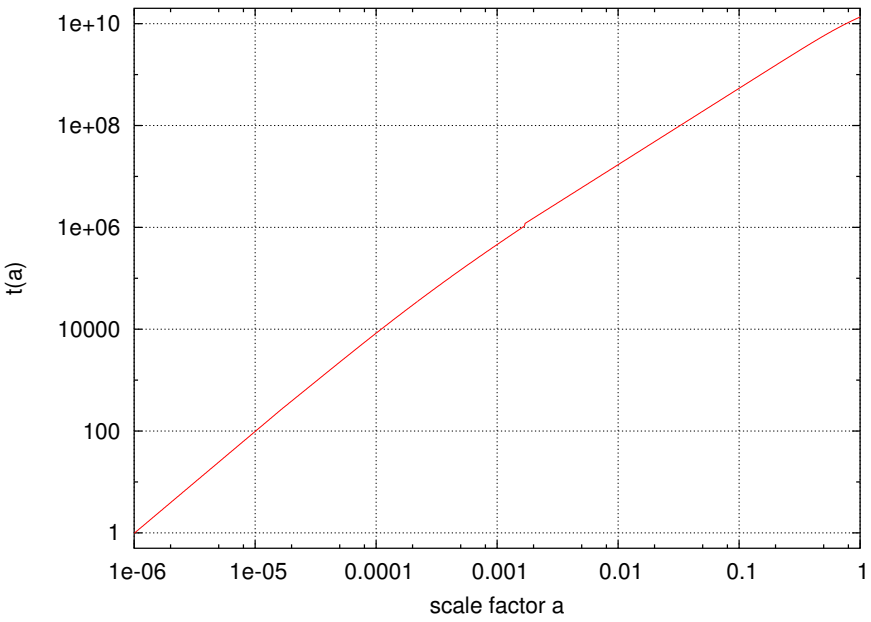


Figure 1.1: Cosmic age $t(a)$ as a function of the scale factor a

1.2.5 Distance measures

- Distance measures are no longer unique in General Relativity. In Euclidean geometry, a distance between two points is defined by a measurement connecting the points at the same instant of time. This is generally impossible for two reasons. First, what is considered simultaneous at the two points depends on their relative motion. Second, connecting the points requires time because of the finite speed of light. Distances in cosmology thus need to be defined according to idealisations or measurement prescriptions, which generally lead to different expressions.

- Distance measures relate emission events on a source's world line to an observation event on an observer's world line. The emission and observation times be t_2 and t_1, respectively, are uniquely related to the scale factors a_2 and $a_1 > a_2$ of the universe at t_2 and t_1, which can in turn be expressed by the redshifts z_2 and $z_1 < z_2$.

- The *proper distance* $D_{\rm prop}$ is the distance measured by the time required for light to travel from a source to an observer. It is thus determined by $\mathrm{d}D_{\rm prop} = -c\mathrm{d}t = -c\mathrm{d}a/\dot{a}$. The minus sign is required because $D_{\rm prop}$ should increase away from, while t and a increase towards the observer. Thus

$$D_{\rm prop}(z_1, z_2) = c \int_{a(z_2)}^{a(z_1)} \frac{\mathrm{d}a}{\dot{a}} = \frac{c}{H_0} \int_{a(z_2)}^{a(z_1)} \frac{\mathrm{d}a}{aE(a)} \,. \qquad (1.62)$$

The integrand is the same as in (1.51), thus

$$D_{\rm prop}(z_1,z_2) = \frac{2}{3\sqrt{1-\Omega_{\rm m0}}}\left[\operatorname{arcsinh}\left(\sqrt{\frac{1-\Omega_{\rm m0}}{\Omega_{\rm m0}}}a_1^{3/2}\right) - \operatorname{arcsinh}\left(\sqrt{\frac{1-\Omega_{\rm m0}}{\Omega_{\rm m0}}}a_2^{3/2}\right)\right] \tag{1.63}$$

for a spatially-flat universe.

- The *comoving distance* $D_{\rm com}$ is the distance on the spatial hypersurface at t = const. between the world lines of a source and an observer comoving with the mean cosmic flow. This is the coordinate distance between source and observer, thus $\mathrm{d}D_{\rm com} = \mathrm{d}w$. Since light rays propagate according to $\mathrm{d}s = 0$, $a\mathrm{d}w = -c\mathrm{d}t = -c\mathrm{d}a/\dot{a}$, thus

$$D_{\rm com}(z_1,z_2) = c\int_{a(z_2)}^{a(z_1)}\frac{\mathrm{d}a}{a\dot{a}} = \frac{c}{H_0}\int_{a(z_2)}^{a(z_1)}\frac{\mathrm{d}a}{a^2E(a)} =: w(z_1,z_2)\ . \tag{1.64}$$

- The *angular diameter distance* $D_{\rm ang}$ is defined in analogy to the relation in Euclidean space between the area δA and the solid angle $\delta\omega$ of an object, $\delta\omega\, D_{\rm ang}^2 = \delta A$. Since the solid angle of spheres of constant radial coordinate w is scaled by $f_K(w)$ in (1.6), we must have

$$\frac{\delta A}{4\pi a_2^2 f_K^2[w(z_1,z_2)]} = \frac{\delta\omega}{4\pi}\ . \tag{1.65}$$

In words, the area of the object must be related to the area of the full sphere like the solid angle of the object to the solid angle of the sphere. It follows

$$D_{\rm ang}(z_1,z_2) = \left(\frac{\delta A}{\delta\omega}\right)^{1/2} = a(z_2)f_K[w(z_1,z_2)]\ , \tag{1.66}$$

as the coordinate distance $w(z_1,z_2) = D_{\rm com}(z_1,z_2)$ and $D_{\rm ang}(z_1,z_2) = a(z_2)f_K[D_{\rm com}(z_1,z_2)]$.

- A fourth important distance measure is the *luminosity distance* $D_{\rm lum}$, which is defined in analogy to the Euclidean relation between the intrinsic luminosity of an object and its flux. Counting emitted and absorbed photons and taking redshift into account, one finds

$$D_{\rm lum}(z_1,z_2) = \left[\frac{a(z_1)}{a(z_2)}\right]^2 D_{\rm ang}(z_1,z_2) = \frac{a(z_1)^2}{a(z_2)}f_K[D_{\rm com}(z_1,z_2)]\ . \tag{1.67}$$

The first equality in (1.67) is the *Etherington relation* and valid in arbitrary spacetimes. It is physically intuitive because photons are

redshifted by a_1/a_2 between emission and absorption, their arrival times are stretched by a_1/a_2, and they are spatially diluted by a factor $(a_1/a_2)^2$. This yields a factor $(a_1/a_2)^4$ between luminosity and flux, and thus a factor $(a_1/a_2)^2$ in the luminosity distance.

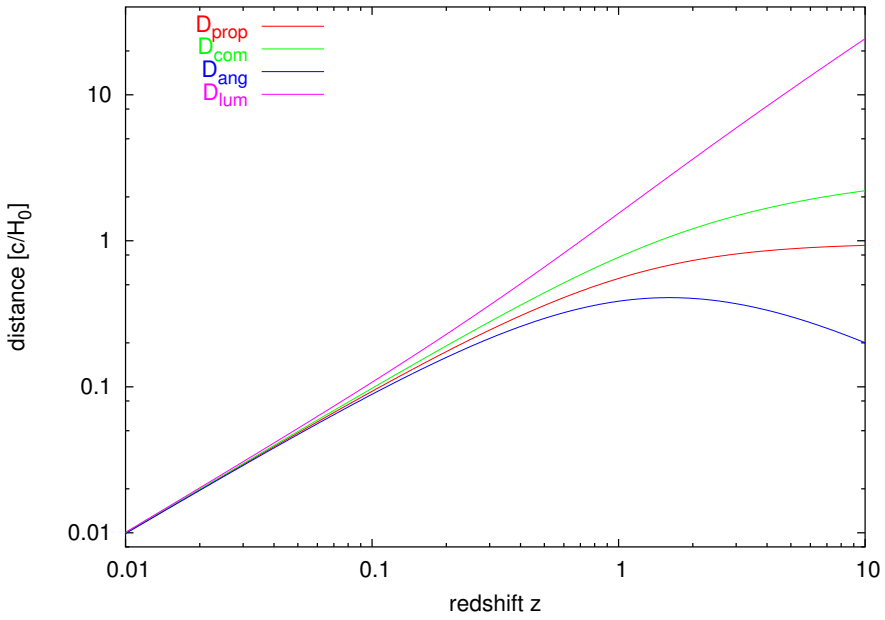


Figure 1.2: Four different distance measures in a spatially-flat universe with $\Omega_{\mathrm{m0}} = 0.3$.

- These distance measures can be vastly different at moderate and high redshifts. For $z \ll 1$, $a \approx 1 - z$, and $E(a) \approx 1$, then

$$D = \frac{cz}{H_0} + O(z^2) \tag{1.68}$$

for all distance measures introduced above. This local distance measure is also known as "Hubble's law".

- The angular-diameter distance from redshift zero to redshift z for an Einstein-de Sitter universe is

$$D_{\mathrm{ang}}(z) = \frac{2c}{H_0}\frac{1}{1+z}\left[1 - \frac{1}{(1+z)^{1/2}}\right] . \tag{1.69}$$

This shows that cosmological distances need not be monotonic. in fact, $D_{\mathrm{ang}}(z)$ has a maximum for $z = 5/4$ in the Einstein-de Sitter case (1.69) and gently decreases for increasing z. This is a consequence of space-time curvature, to be distinguished from spatial curvature!

1.2.6 Horizons

- Between times t_1 and $t_2 > t_1$, light can travel across the comoving distance

$$\Delta w(t_1, t_2) = \int_{t_1}^{t_2} \frac{c\mathrm{d}t}{a(t)} = c \int_{a(t_1)}^{a(t_2)} \frac{\mathrm{d}a}{a\dot{a}} \tag{1.70}$$

  cf. (1.64).

- As $t \to 0$, $a \to 0$. The curvature and cosmological-constant terms in the first eq. (1.17) become negligible and

$$\dot{a} = a\sqrt{\frac{8\pi G}{3}\rho} \,. \tag{1.71}$$

  Let $\rho \propto \rho_0 a^{-n}$, then

$$\Delta w(t_1, t_2) = \frac{c}{H_0}\sqrt{\Omega_0}^{-1} \int_{a(t_1)}^{a(t_2)} \frac{\mathrm{d}a}{a^{2-n/2}} \propto a^{n/2-1} \,, \tag{1.72}$$

  which diverges for $a \to 0$ if $n < 2$.

- Thus, if $n > 2$, light can only travel by a finite distance between the Big Bang and any later time, thus any particle in the Universe can only be influenced by events within a finite region. There exists a *particle horizon*.

- A simpler definition of a horizon is often used, namely the time-dependent Hubble radius

$$r_{\mathrm{H}}(t) = \frac{c}{H(t)} = \frac{c}{H_0}\frac{a^{3/2}}{\sqrt{\Omega_{\mathrm{m}0}}}\left(1 + \frac{a_{\mathrm{eq}}}{a}\right)^{-1/2} \,, \tag{1.73}$$

  where we have used the Einstein-de Sitter limit (1.54). Particularly important for structure formation is the Hubble radius at $a = a_{\mathrm{eq}}$,

$$r_{\mathrm{H,eq}} = \frac{c}{H_0}\frac{a_{\mathrm{eq}}^{3/2}}{\sqrt{2\Omega_{\mathrm{m}0}}} \,. \tag{1.74}$$

- As $t \to \infty$, suppose $a \propto t^m$, then

$$\Delta w(t_1, t_2) \propto t^{1-m} \,, \tag{1.75}$$

  which converges for $m > 1$. This happens if the expansion of the Universe is dominated by the cosmological constant at late times.

- Then, the region which can be seen by a particle remains finite. There exists an *event horizon*.

1.3 Thermal Evolution

1.3.1 Assumptions

- *The universe expands adiabatically* – isotropy requires the universe to expand adiathermally: no heat can flow because flow directions would violate isotropy. Adiathermal expansion is adiabatic if it is reversible, but irreversible processes may occur. However, the cosmic microwave and neutrino backgrounds dominate by far the entropy of the universe, thus entropy generation is completely negligible.

- *Thermal equilibrium can be maintained despite the expansion* – thermal equilibrium can only be maintained if the interaction rate of particles is higher than the expansion rate of the Universe. The expansion rate of the Universe is highest at early times, so thermal equilibrium may be difficult to maintain as $t \to 0$. Nonetheless, for $t \to 0$, particle densities grow so fast that interaction rates are indeed higher than the expansion rate. As the Universe expands, particle species drop out of equilibrium.

- *The cosmic "fluids" can be treated as ideal gases* – Characteristics of an ideal gas are: There are no long-range interactions between particles. The particles interact only by direct collisions which is a good approximation for weakly interacting particles like neutrinos, though this is even valid for charged particles because oppositely charged particles shield each other. As a consequence the internal energy of an ideal gas does not depend on the occupied volume. Cosmic "fluids" can be treated as possibly relativistic quantum gases.

- Those assumptions are the starting point of our considerations. They need to be verified as we go along.

1.3.2 Quantum Statistics

- We will need many relations later for the behaviour of ideal quantum gases which we now derive in a brief detour.

- If a thermodynamic system has fixed internal energy, particle number N, and volume, it is called a micro-canocical ensemble. Its density in phase space is constant.

- If only the mean internal energy is specified, the ensemble is canonical. The probability of finding a quantum state (symbolically labelled by α) with energy ϵ_α occupied is given by the Boltz-

mann factor

$$f_\alpha = \frac{\mathrm{e}^{-\epsilon_\alpha/kT}}{Z_\mathrm{c}} \,, \quad Z_\mathrm{c} = \sum_\alpha \mathrm{e}^{-\epsilon_\alpha/kT} \,, \tag{1.76}$$

where T is the temperature, and Z_C is the canonical partition sum over all accessible quantum states. The canonical phase-space distribution minimises the Helmholtz free energy $F(T,V,N) = -kT\ln Z_\mathrm{c}$.

- If, in addition, only the mean number of particles is specified, the ensemble is grand-canonical. All accessible quantum states (labelled by α) are then occupied by an unknown number N_α of particles such that $\sum_\alpha N_\alpha = N$. The total energy of that ensemble is $E(N_\alpha) = \sum_\alpha \epsilon_\alpha N_\alpha$. The phase-space distribution function of a grand-canonical ensemble is

$$f_\alpha = \frac{\mathrm{e}^{-[E(N_\alpha)-\mu N_\alpha]/kT}}{Z_\mathrm{gc}} \,, \quad Z_\mathrm{gc} = \sum_{N=0}^{\infty} \mathrm{e}^{\mu N/kT} \sum_{\{N_\alpha\}} \mathrm{e}^{-E(N_\alpha)/kT} \,, \tag{1.77}$$

where μ is the chemical potential and Z_gc is the grand-canonical partition sum, in which the second sum is over all sets $\{N_\alpha\}$ of occupation numbers which sum up to N. The grand-canonical phase-space distribution minimises the grand-canonical potential $\Phi(T,V,\mu) = -kT\ln Z_\mathrm{gc}$.

- We now evaluate the grand-canonical partition sum:

$$Z_\mathrm{gc} = \sum_{N=0}^{\infty} \sum_{\{N_\alpha\}} \mathrm{e}^{-\sum_\alpha(\epsilon_\alpha-\mu)N_\alpha/kT} \,. \tag{1.78}$$

Although the second sum is constrained, we have to sum over all possible particle numbers N. Thus, ultimately all possible sets of occupation numbers N_α occur, and

$$Z_\mathrm{gc} = \sum_{N_\alpha} \prod_\alpha \mathrm{e}^{-(\epsilon_\alpha-\mu)N_\alpha/kT} = \prod_\alpha Z_\alpha \,, \tag{1.79}$$

with

$$Z_\alpha := \sum_{N_\alpha} \mathrm{e}^{-(\epsilon_\alpha-\mu)N_\alpha/kT} \,. \tag{1.80}$$

- For fermions, $N_\alpha = 0, 1$ because of Pauli's exclusion principle, while for bosons, $N_\alpha = 0, 1, \ldots, \infty$. Thus

$$Z_\alpha = \begin{cases} 1 + \mathrm{e}^{-(\epsilon_\alpha-\mu)/kT} & \text{fermions} \\ \left(1 - \mathrm{e}^{-(\epsilon_\alpha-\mu)/kT}\right)^{-1} & \text{bosons} \,, \end{cases} \tag{1.81}$$

where we have used the geometrical series

$$\sum_{n=0}^{\infty} \mathrm{e}^{-nx} = \sum_{n=0}^{\infty} \left(\mathrm{e}^{-x}\right)^n = \frac{1}{1-\mathrm{e}^{-x}} \,. \tag{1.82}$$

- The mean occupation number of a quantum state α is

$$\bar{N}_\alpha = \frac{1}{Z_\alpha} \sum_\alpha N_\alpha \mathrm{e}^{-(\epsilon_\alpha - \mu)N_\alpha/kT} = \frac{kT}{Z_\alpha} \frac{\partial Z_\alpha}{\partial \mu} \,, \tag{1.83}$$

  which leads to the well-known result

$$\bar{N}_\alpha = \frac{1}{\mathrm{e}^{(\epsilon_\alpha - \mu)/kT} \pm 1} \,, \tag{1.84}$$

  where the + sign applies to fermions, the − sign to bosons.

1.3.3 Properties of Ideal Quantum Gases

- In thermal equilibrium with a heat bath of temperature T, the chemical potential of a system with N particles must vanish, $\mu = 0$: the Helmholtz free energy $F(T, V, N) = E - TS$ is minimised in equilibrium for a system at constant T and V, so from $\mathrm{d}F = -S\,\mathrm{d}T - P\mathrm{d}V + \mu \mathrm{d}N = 0$

$$\left(\frac{\partial F}{\partial N} \right)_{T,V} = \mu = 0 \,. \tag{1.85}$$

- The particle momentum $\vec{p} = \hbar \vec{k}$ is generally related to energy by

$$\epsilon(p) = \sqrt{c^2 p^2 + m^2 c^4} \,. \tag{1.86}$$

- For particles confined in a volume V, the number of states per k-space element is

$$\mathrm{d}N = g \frac{V}{(2\pi)^3} \mathrm{d}^3 k \,, \tag{1.87}$$

  where g is the statistical weight, e.g. the spin degeneracy factor. Summations over quantum states are now replaced by integrals over k space weighted according to (1.84).

- Using (1.84), the spatial particle number density in thermal equilibrium is

$$n = \frac{g}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi p^2 \mathrm{d}p}{\exp[\epsilon(p)/kT] \pm 1} \,. \tag{1.88}$$

  The mean energy density is the number of states per phase-space volume element, times the mean occupation number, times the energy per state, integrated over momentum space,

$$u = \frac{g}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi p^2\, \epsilon(p)\, \mathrm{d}p}{\exp[\epsilon(p)/kT] \pm 1} \,. \tag{1.89}$$

- Integrals like those in (1.88) and (1.89) are most easily carried out by substituting the geometrical series (1.82),

$$\begin{aligned}\int_0^\infty \frac{x^m \mathrm{d}x}{e^x - 1} &= \int_0^\infty \frac{x^m e^{-x} \mathrm{d}x}{1 - e^{-x}} = \int_0^\infty \mathrm{d}x x^m e^{-x} \sum_{n=0}^{\infty} e^{-nx} \\ &= \sum_{n=1}^{\infty} \int_0^\infty \mathrm{d}x x^m e^{-nx} = m! \zeta(m+1) \,. \end{aligned} \tag{1.90}$$

  For fermions, use

$$\frac{1}{e^x + 1} = \frac{1}{e^x - 1} - \frac{2}{e^{2x} - 1} \,. \tag{1.91}$$

- Using (1.79), (1.81) and (1.86), the grand-canonical potential can be written as

$$\Phi(T, V, \mu) = \mp kT \frac{gV}{(2\pi\hbar)^3} \int_0^\infty \mathrm{d}p 4\pi p^2 \ln\left[1 \pm e^{\mu/kT} e^{-\epsilon(p)/kT}\right] \,, \tag{1.92}$$

  with the upper sign applying to fermions, the lower to bosons. From the expressions for the Helmholtz free energy F, the grand-canonical potential Φ and the thermodynamic Euler relation,

$$\begin{aligned} F(T, V, N) &= U - TS \\ \Phi(T, V, \mu) &= F - \mu N = U - TS - \mu N \\ U &= TS - PV + \mu N \,, \end{aligned} \tag{1.93}$$

  we find the simple relation

$$\Phi = -PV \;\Rightarrow\; P = -\frac{\Phi}{V} \,, \tag{1.94}$$

  which enables us to directly compute the pressure of quantum gases. Likewise, from the total differential of the grand-canonical potential, $\mathrm{d}\Phi(T, V, \mu) = -S\,\mathrm{d}T - P\mathrm{d}V - N\mathrm{d}\mu$, we find the entropy as

$$S = -\frac{\partial \Phi}{\partial T} \,. \tag{1.95}$$

- Example: relativistic bosons have $\epsilon = cp$, and in thermal equilibrium their chemical potential vanishes, $\mu = 0$. Their grand-canonical potential is thus

$$\Phi(T, V, \mu) = kT \frac{gV}{(2\pi\hbar)^3} \int_0^\infty 4\pi p^2 \mathrm{d}p \ln\left[1 - e^{-cp/kT}\right] \,. \tag{1.96}$$

  We substitute $x := cp/kT$ and find

$$\Phi(T, V, \mu) = \frac{gV}{2\pi^2\hbar^3} \frac{(kT)^4}{c^3} \int_0^\infty x^2 \mathrm{d}x \ln\left(1 - \mathrm{e}^{-x}\right) \,. \tag{1.97}$$

The integral over the logarithm can be solved as follows:

$$\begin{aligned}\int_0^\infty x^m \ln\left(1-\mathrm{e}^{-x}\right)\mathrm{d}x &= \left.\frac{x^{m+1}}{m+1}\ln\left(1-\mathrm{e}^{-x}\right)\right|_0^\infty \\ &- \int_0^\infty \frac{x^{m+1}}{m+1}\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}\mathrm{d}x \\ &= -m!\zeta(m+2)\,, \end{aligned} \tag{1.98}$$

where (1.90) was inserted. We thus find the grand-canonical potential

$$\Phi(T,V,\mu) = -gV\frac{\pi^2}{90}\frac{(kT)^4}{(\hbar c)^3}\,, \tag{1.99}$$

from which we obtain the pressure

$$P_\mathrm{B} = -\frac{\Phi}{V} = g\,\frac{\pi^2}{90}\,\frac{(kT)^4}{(\hbar c)^3}\,, \tag{1.100}$$

and the entropy density

$$s = \frac{S}{V} = -\frac{1}{V}\frac{\partial\Phi}{\partial T} = gk\frac{2\pi^2}{45}\left(\frac{kT}{\hbar c}\right)^3\,. \tag{1.101}$$

- Summarising, these equations yield the following expressions for the number, energy, entropy densities and the pressure of relativistic boson and fermion gases as well as a non-relativistic gas in thermal equilibrium (i.e., $\mu = 0$):

	relativistic distributions: Bose-Einstein	relativistic distributions: Fermi-Dirac	non-relativistic: Maxwell-Boltzmann
n	$n_\mathrm{B} = g_\mathrm{B}\,\frac{\zeta(3)}{\pi^2}\left(\frac{kT}{\hbar c}\right)^3$	$n_\mathrm{F} = \frac{3}{4}\frac{g_\mathrm{F}}{g_\mathrm{B}}n_\mathrm{B}$	$g\left(\frac{mkT}{2\pi\hbar^2}\right)^{3/2}\exp\left(-\frac{mc^2}{kT}\right)$
u	$u_\mathrm{B} = g_\mathrm{B}\,\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3}$	$u_\mathrm{F} = \frac{7}{8}\frac{g_\mathrm{F}}{g_\mathrm{B}}u_\mathrm{B}$	$n\left(\frac{3}{2}kT + mc^2\right)$
P	$P_\mathrm{B} = \frac{u_\mathrm{B}}{3} = g_\mathrm{B}\,\frac{\pi^2}{90}\,\frac{(kT)^4}{(\hbar c)^3}$	$P_\mathrm{F} = \frac{u_F}{3} = \frac{7}{8}\frac{g_\mathrm{F}}{g_\mathrm{B}}P_\mathrm{B}$	$nkT \ll u$
s	$s_\mathrm{B} = g_\mathrm{B}k\,\frac{2\pi^2}{45}\left(\frac{kT}{\hbar c}\right)^3$	$s_\mathrm{F} = \frac{7}{8}\frac{g_\mathrm{F}}{g_\mathrm{B}}s_\mathrm{B}$	$nk\left\{\ln\left[\frac{1}{n}\left(\frac{mkT}{2\pi\hbar^2}\right)^{3/2}\right]+\frac{5}{2}\right\}$

(1.102)

- Some numbers are useful for later estimates. Note: $1\,\mathrm{eV} = 1.6 \times 10^{-12}\,\mathrm{erg}$ correspond to $T = 1.16 \times 10^4\,\mathrm{K}$.

$$\begin{aligned} n_\mathrm{B} &= 10 g_\mathrm{B} \left(\frac{T}{\mathrm{K}}\right)^3 \mathrm{cm}^{-3} = 1.6 \times 10^{13} g_\mathrm{B} \left(\frac{kT}{\mathrm{eV}}\right)^3 \mathrm{cm}^{-3} \\ u_\mathrm{B} &= 3.8 \times 10^{-15} g_\mathrm{B} \left(\frac{T}{\mathrm{K}}\right)^4 \frac{\mathrm{erg}}{\mathrm{cm}^3} = 2.35 \times 10^{-3} g_\mathrm{B} \left(\frac{kT}{\mathrm{eV}}\right)^4 \frac{\mathrm{erg}}{\mathrm{cm}^3} \\ \frac{s_\mathrm{B}}{k} &= 36 g_\mathrm{B} \left(\frac{T}{\mathrm{K}}\right)^3 \mathrm{cm}^{-3} = 5.7 \times 10^{13} g_\mathrm{B} \left(\frac{kT}{\mathrm{eV}}\right)^3 \mathrm{cm}^{-3} \,. \end{aligned} \quad (1.103)$$

1.3.4 Adiabatic Expansion of Ideal Gases

- For relativistic boson or fermion gases in thermal equilibrium, the pressure is a third of the energy density,

$$P = \frac{u}{3} = \frac{E}{3V} \,. \quad (1.104)$$

- The first law of thermodynamics in absence of heat transfer, $\mathrm{d}E + P\mathrm{d}V = 0$, then implies

$$\mathrm{d}E = -P\mathrm{d}V = 3\mathrm{d}(PV) \;\Rightarrow\; P \propto V^{-4/3} \,, \quad (1.105)$$

  i.e. the *adiabatic index* is $\gamma = 4/3$. For non-relativistic ideal gases, $\gamma = 5/3$.

- According to (1.102), pressure P scales with temperature T^4 for relativistic particles, thus

$$T \propto V^{-1/3} \propto a^{-1} \,, \quad (1.106)$$

  where a is the cosmological scale factor. The temperature of non-relativistic gases drops faster,

$$T \propto PV \propto V^{-5/3+1} \propto a^{-2} \,. \quad (1.107)$$

- The result (1.105) is very important for cosmology. It implies that the photon temperature drops inversely proportional to the scale factor, which has an important consequence for the spectrum of the microwave background, as we shall see later.

1.3.5 Particle Freeze-Out

- We have to verify the basic assumption that thermal equilibrium can be maintained against the rapid expansion of the universe at early times. For doing so, we compare the expansion rate of the universe to the interaction rate of particles.

- At early times, curvature and cosmological constant are negligible, thus Friedmann's equation implies

$$\dot{a} = a\sqrt{\frac{8\pi G}{3}\rho} \,. \tag{1.108}$$

The expansion time-scale $t_{\rm exp}$ can be approximated by

$$t_{\rm exp} \approx \frac{a}{\dot{a}} = \sqrt{\frac{3}{8\pi G\rho}} \approx (G\rho)^{-1/2} \,. \tag{1.109}$$

During the radiation-dominated era in the early universe, $\rho \propto a^{-4}$, thus

$$t_{\rm exp} \propto a^2 \,, \tag{1.110}$$

as we have already seen in (1.49) in the context of how the young universe ages. The expansion time-scale thus increases rapidly as the universe expands away from the Big Bang.

- Thermal equilibrium is maintained predominantly by two-body interactions. The number of collision partners found by a particle travelling for a time interval dt with velocity v relative to the cosmic rest frame through a particle population with number density n is

$$\mathrm{d}N = n\langle\sigma v\rangle \mathrm{d}t \,, \tag{1.111}$$

where σ is the collision cross section, which typically depends on relative velocity v and is thus averaged with v.

- The collision rate experienced by a single particle species is thus

$$\Gamma := \frac{\mathrm{d}N}{\mathrm{d}t} = n\langle\sigma v\rangle \propto n \propto T^3 \propto a^{-3} \,, \tag{1.112}$$

where we have used (1.102) and (1.106) which are both valid throughout the radiation-dominated early phase of the universe. The collision time-scale is thus

$$t_{\rm coll} = \Gamma^{-1} \propto a^3 \,. \tag{1.113}$$

- As $a \to 0$, the ratio between expansion and collision time scales is $t_{\rm exp}/t_{\rm coll} \propto a^{-1} \to \infty$, which implies that the collisions have a much shorter time scale than the expansion in the early universe. Thermal equilibrium can thus be maintained despite the expansion *in particular* at early times. As the universe keeps expanding, collisions become rare and thermal equilibrium will ultimately break down.

- In absence of collisions, the continuity equation for the number density n of a particle species is

$$\dot{n} + \vec{\nabla}\cdot(n\vec{v}) = 0 \,. \tag{1.114}$$

In the homogeneous and isotropic universe, n is spatially constant, and $\vec{v} = H\vec{r}$, where $\vec{r}$ is the physical distance of a particle from the origin. Since $\vec{\nabla} \cdot \vec{r} = 3$, we thus have

$$\dot{n} + 3Hn = 0 \,. \tag{1.115}$$

- The right-hand side of (1.115) will deviate from zero in presence of collisions and thermal particle creation. We saw in (1.112) that the collision rate is $\Gamma = n\langle\sigma v\rangle$. Likewise, the source term for thermal particle creation is $S = \langle\sigma v\rangle n_{\rm T}^2$. Thus, the continuity equation changes to read

$$\dot{n} + 3Hn = -\Gamma n + S = -\Gamma n\left(1 - \frac{n_{\rm T}^2}{n^2}\right) \,. \tag{1.116}$$

- We now introduce the comoving number density $N := a^3 n$. substituting from $\dot{N} = a^3(3Hn + \dot{n})$ in (1.116) yields

$$\dot{N} = -\Gamma N\left(1 - \frac{N_{\rm T}^2}{N^2}\right) \,. \tag{1.117}$$

Substituting further

$$\frac{\mathrm{d}}{\mathrm{d}t} = \dot{a}\frac{\mathrm{d}}{\mathrm{d}a} = aH\frac{\mathrm{d}}{\mathrm{d}a} = H\frac{\mathrm{d}}{\mathrm{d}\ln a} \,, \tag{1.118}$$

yields

$$\frac{\mathrm{d}\ln N}{\mathrm{d}\ln a} = -\frac{\Gamma}{H}\left(1 - \frac{N_{\rm T}^2}{N^2}\right) \,. \tag{1.119}$$

- Thus, if the comoving number density is thermal, $N = N_{\rm T}$, it does not change. If N deviates from $N_{\rm T}$, it needs to change for readjustment to its thermal equlibrium value $N_{\rm T}$. This is impossible if $\Gamma \ll H$ because then the rate of change becomes too small. Then, the particles freeze out of thermal equilibrium.

- For relativistic particles, $n \propto T^3 \propto a^{-3}$, thus $N = a^3 n = \text{const.}$. According to the freeze-out equation (1.119),

$$\frac{\mathrm{d}\ln N}{\mathrm{d}\ln a} = 0 \;\Rightarrow\; N = N_{\rm T} \,. \tag{1.120}$$

This implies that relativistic particle species retain their thermal-equilibrium density regardless of Γ/H, i.e. even after freeze-out.

- For non-relativistic particles, the comoving number density in thermal equilibrium is

$$N_{\rm T} \propto T^{-3/2}\mathrm{e}^{-mc^2/kT} \,. \tag{1.121}$$

For $kT \lesssim mc^2$, $N_{\rm T}$ drops exponentially, i.e. very quickly $N_{\rm T} \ll N$, then

$$\frac{{\rm d}\ln N}{{\rm d}\ln a} \approx -\frac{\Gamma}{H} \to 0 \,, \tag{1.122}$$

as the collision rate falls below the expansion rate. The actual comoving number density of particles then remains constant, while its thermal-equilibrium value drops to zero.

1.4 Recombination and Nucleosynthesis

1.4.1 The Neutrino Background

- Neutrinos are kept in thermal equilibrium by the weak interaction,

$$\nu + \bar{\nu} \leftrightarrow e^+ + e^- \,, \tag{1.123}$$

  which freezes out when the temperature drops to

$$T_\nu \approx 10^{10.5}\,\mathrm{K} \approx 2.7\,\mathrm{MeV} \,. \tag{1.124}$$

- Due to their low mass, neutrinos are ultra-relativistic when they freeze out of equilibrium, thus their comoving number density is that of an ideal, relativistic fermion gas.

- The electron-positron decay reaction,

$$e^+ + e^- \leftrightarrow 2\gamma \,, \tag{1.125}$$

  is suppressed a little later, when the temperature drops below,

$$T \approx 2m_\mathrm{e}c^2 \approx 1\,\mathrm{MeV} \approx 10^{10}\,\mathrm{K} \,, \tag{1.126}$$

  because photons are no longer energetic enough for electron-positron pair production afterwards.

- Electrons and positrons annihilate shortly after neutrino freeze-out. Their decay entropy thus heats the photon gas, but not the neutrinos. The temperature of the photon gas is therefore higher than that of the neutrino gas.

- The entropies before and after electron-positron annihilation must be equal. Let primes denote quantities before annihilation, then the entropy densities must satisfy

$$s'_{\mathrm{e}^+} + s'_{\mathrm{e}^-} + s'_\gamma = s_\gamma \,. \tag{1.127}$$

- Before annihilation, the temperatures of electrons, positrons and photons can be considered equal because thermal equilibrium was maintained, $T'_{\mathrm{e}^+} = T'_{\mathrm{e}^-} = T'_\gamma =: T'$.

- The statistical weights of electrons, positrons and photons are all $g_{\mathrm{e}^+} = g_{\mathrm{e}^-} = g_\gamma = 2$. Their entropy densities therefore differ only by the fermion factor $7/8$ from (1.102),

$$s'_{\mathrm{e}^+} = s'_{\mathrm{e}^-} = \frac{7}{8} s'_\gamma \,, \tag{1.128}$$

and since they are proportional to T^3, the temperature T after annihilation follows from (1.127) as,

$$\left(2 \cdot \frac{7}{8} + 1\right)(T')^3 = T^3$$
$$\Rightarrow \quad T = \left(\frac{11}{4}\right)^{1/3} T' \approx 1.4\, T' \,. \qquad (1.129)$$

Hence the photon temperature is approximately 40% higher today than the neutrino temperature.

1.4.2 The Entropy of the Universe Today

- Entropies of different particle species are additive. We generalize the results provided in (1.102) to multiple (non-interacting) relativistic species and calculate the entropy density of the Universe

$$s = k\frac{2\pi^2}{45}\left(\frac{kT}{\hbar c}\right)^3 \left[\sum_{i=\text{bosons}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{j=\text{fermions}} g_j \left(\frac{T_j}{T}\right)^3\right]. \quad (1.130)$$

- Using (1.102), we obtain the number density of relic CMB photons in terms of their temperature today as

$$n_{\gamma,0} = \frac{2\zeta(3)}{\pi^2}\left(\frac{kT_{\gamma,0}}{\hbar c}\right)^3, \qquad (1.131)$$

where the factor 2 accounts for the two polarization degrees of freedom of photons.

- Today, the entropy of the Universe is dominated by CMB photons and three species of neutrinos and their anti-partners,

$$s_0 = k\frac{\pi^4 n_{\gamma,0}}{45\zeta(3)}\left[2 + \frac{7}{8}6\left(\frac{T_{\nu,0}}{T_{\gamma,0}}\right)^3\right] = k\frac{\pi^4 n_{\gamma,0}}{45\zeta(3)}\frac{43}{11} \approx 7k n_{\gamma,0}, \quad (1.132)$$

where we used (1.129), (1.131), and $\zeta(3) \approx 1.202$.

1.4.3 Photons and Baryons

- Assuming for simplicity that all baryons are locked up in hydrogen, the number density of baryons today is

$$n_{\rm B} = \frac{\rho_{\rm B}}{m_{\rm p}} = \frac{\Omega_{\rm B}}{m_{\rm p}}\frac{3H_0^2}{8\pi G} = 1.1 \times 10^{-5}\, \Omega_{\rm B}\, h^2\, {\rm cm}^3 \,, \qquad (1.133)$$

where $m_{\rm p}$ is the proton mass, and $\Omega_{\rm B}$ is the baryon density parameter, defined as in (1.39).

- As we shall see later, the baryon density parameter is constrained to be
$$\Omega_{\rm B} h^2 \approx 0.025 \,, \tag{1.134}$$
i.e. baryons contribute only $\approx$ 10% − 20% of the matter in the Universe.

- The photon number density today is given by the temperature of the microwave background through (1.102),
$$n_\gamma = 407\,{\rm cm}^{-3} \,. \tag{1.135}$$

- Both $n_{\rm B}$ and n_γ scale with temperature $\propto T^3 \propto a^{-3}$, implying that their ratio is constant,
$$\eta := \frac{n_{\rm B}}{n_\gamma} = 2.7 \times 10^{-8}\, \Omega_{\rm B} h^2 \,. \tag{1.136}$$

- There is approximately a billion photons per baryon in the universe. The entropy of the photon gas dominates the entropy of the universe by a huge margin, justifying the assumption of adiabatic expansion, because any contribution to the entropy due to irreversible processes can be neglected compared to the photon entropy.

- It is unclear how η is set. It is a fundamental physical problem why there are baryons in the universe, because they should have annihilated with anti-baryons. There must have been an asymmetry between baryons and antibaryons, which is possible under the Sakharov conditions (CP violation, interactions changing baryon number, departure from thermodynamic equilibrium, e.g. during phase transitions).

- When we speak of "the temperature of the universe" from now on, we refer to the temperature of the photon gas.

- The smallness of η will turn out to be very important for nucleosynthesis and the recombination of the universe, i.e. its transition from the fully ionised to the neutral state.

1.4.4 The Recombination Process

- As the temperature drops, electrons and protons combine to form hydrogen atoms when the reaction
$$e^- + p^+ \leftrightarrow H + \gamma \,, \tag{1.137}$$
freezes out.

- For determining how recombination proceeds, we need to minimise the Helmholtz free energy $F(T, V, N)$, which is related to the canonical partition function Z_c,

$$F(T, V, N) = -kT \ln Z_\mathrm{c} \ . \tag{1.138}$$

- For the process (1.137), the canonical partition function is given by

$$Z_\mathrm{c} = \frac{Z_\mathrm{e}^{N_\mathrm{e}} Z_\mathrm{p}^{N_\mathrm{p}} Z_\mathrm{H}^{N_\mathrm{H}}}{N_\mathrm{e}! N_\mathrm{p}! N_\mathrm{H}!} \ , \tag{1.139}$$

  where $Z_{\mathrm{e,p,H}}$ and $N_{\mathrm{e,p,H}}$ are the canonical partition functions and numbers of electrons, protons, and hydrogen atoms, respectively. Here we assumed independent, indistinguishable particles and a density that is low enough to neglect degeneracy effects, i.e., $g_i = 1$ for $i \in \{\mathrm{e}, \mathrm{p}, \mathrm{H}\}$. The photons do not contribute because they provide the heat bath controlling the temperature T.

- The baryon number is $N_\mathrm{B} = N_\mathrm{p} + N_\mathrm{H}$, the electron number is $N_\mathrm{e} = N_\mathrm{p}$, thus $N_\mathrm{H} = N_\mathrm{B} - N_\mathrm{e}$. Given the total baryon number, all other numbers can be expressed by the electron number N_e.

- Since the numbers $N_{\mathrm{e,p,H}}$ will be very large, we can use Stirling's formula for the factorials, $\ln N! \approx N \ln N - N$.

- We now need to minimise the Helmholtz free energy with respect to N_e:

$$\begin{aligned} \frac{\partial F}{\partial N_\mathrm{e}} &= 0 \\ &= \frac{\partial}{\partial N_\mathrm{e}} \Big[N_\mathrm{e} \ln Z_\mathrm{e} + N_\mathrm{p} \ln Z_\mathrm{p} + N_\mathrm{H} \ln Z_\mathrm{H} \\ &- N_\mathrm{e}(\ln N_\mathrm{e} - 1) - N_\mathrm{p}(\ln N_\mathrm{p} - 1) - N_\mathrm{H}(\ln N_\mathrm{H} - 1) \Big] \\ &= \ln Z_\mathrm{e} + \ln Z_\mathrm{p} - \ln Z_\mathrm{H} - 2 \ln N_\mathrm{e} + \ln(N_\mathrm{B} - N_\mathrm{e}) \end{aligned} \tag{1.140}$$

  where we have used

$$\frac{\partial N_\mathrm{p}}{\partial N_\mathrm{e}} = 1 \ , \quad \frac{\partial N_\mathrm{H}}{\partial N_\mathrm{e}} = -1 \ . \tag{1.141}$$

- For the electron number, (1.140) implies

$$\frac{N_\mathrm{e}^2}{N_\mathrm{B} - N_\mathrm{e}} = \frac{Z_\mathrm{e} Z_\mathrm{p}}{Z_\mathrm{H}} \ . \tag{1.142}$$

- Following (1.153) (Die Referenz kommt erst später im Text), the canonical partition function for a single particle species is

$$Z = \frac{4\pi g V}{(2\pi\hbar)^3} \int_0^\infty \mathrm{d}p p^2 \mathrm{e}^{-(\epsilon-\mu)/kT} \ , \tag{1.143}$$

where $\epsilon = mc^2 + p^2/(2m)$ in the non-relativistic limit. Thus

$$Z = \frac{gV(2\pi mkT)^{3/2}}{(2\pi\hbar)^3} \mathrm{e}^{-(mc^2-\mu)/kT} \,. \tag{1.144}$$

- The total chemical potential must vanish in equilibrium [cf. (1.162)], thus $\mu_\mathrm{e} + \mu_\mathrm{p} = \mu_\mathrm{H}$, and the ionisation potential of hydrogen is $\chi = (m_\mathrm{e} + m_\mathrm{p} - m_\mathrm{H})c^2 = 13.6\,\mathrm{eV}$. inserting (1.144) into (1.142) and using these relations yields

$$\frac{x^2}{1-x} = \frac{(2\pi m_\mathrm{e} kT)^{3/2}}{(2\pi\hbar)^3 n_\mathrm{B}} \mathrm{e}^{-\chi/kT} \,, \tag{1.145}$$

where $x = N_\mathrm{e}/N_\mathrm{B}$ is the ionisation degree, and $n_\mathrm{B} = N_\mathrm{B}/V$ is the number density of baryons. This is Saha's equation.

- According to (1.136) and (1.102), the baryon density is

$$n_\mathrm{B} = \eta n_\gamma = 2\eta \frac{\zeta(3)}{\pi^2} \left(\frac{kT}{\hbar c}\right)^3 \,, \tag{1.146}$$

which yields

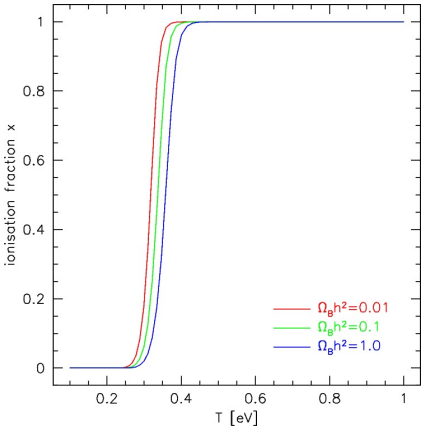
$$\frac{x^2}{1-x} = \frac{\sqrt{\pi}}{4\sqrt{2}\zeta(3)\eta} \left(\frac{m_\mathrm{e}c^2}{kT}\right)^{3/2} \mathrm{e}^{-\chi/kT} \approx \frac{0.26}{\eta} \left(\frac{m_\mathrm{e}c^2}{kT}\right)^{3/2} \mathrm{e}^{-\chi/kT} \,. \tag{1.147}$$

- For recombination to be considered finished, $x \ll 1$ and $x^2/(1-x) \approx x^2$. Since $1/\eta$ is a huge number, $kT \ll \chi$ is required for x to be small. For example, putting $x = 0.1$ yields $kT_\mathrm{rec} = 0.3\,\mathrm{eV}$, or

$$T_\mathrm{rec} \approx 3500\,\mathrm{K} \,. \tag{1.148}$$

- Since $\chi = 13.6\,\mathrm{eV}$, one would naively expect $T_\mathrm{rec} \approx 10^5\,\mathrm{K}$. The very large photon-to-baryon ratio $1/\eta$ delays recombination considerably.

- Strictly, Saha's equation is invalid for cosmological recombination because it assumes thermal equilibrium between the reaction partners, which breaks down as recombination proceeds. However, due to the rapid progress of recombination, the deviation between the ionisation degree predicted by Saha's equation and by an exact treatment remains small.



Ionisation fraction as a function of temperature for three different values of the baryon density parameter. Once it sets in, recombination completes very quickly.

Two-Photon Recombination

- Direct hydrogen recombination produces energetic photons. The final transition to the ground state is Lyman-α ($2P \to 1S$), so that the energy of the emitted photon is $h\nu \geq E_{\mathrm{Ly}\alpha} = 3\chi/4 = 10.2\,\mathrm{eV}$.

- The abundant Lyα photons keep reionising the cosmic gas because they cannot stream away as from hydrogen clouds. The energy loss due to cosmic expansion is slow.

- Recombination can only proceed by production of photons with lower energy than Lyα. This is possible through the forbidden transition $2S \rightarrow 1S$, which requires the emission of two photons.

- This process is slow, hence recombination proceeds at a somewhat lower rate than predicted by Saha's equation.

Thickness of the Recombination Shell

- Recombination is not instantaneous, but requires a finite time interval. There is thus a "recombination shell" with finite thickness.

- The optical depth along a light ray through the recombination shell is

$$\tau = \int n_{\rm e}\sigma_{\rm T}{\rm d}r = n_{\rm B}\sigma_{\rm T}\int x{\rm d}r\ , \tag{1.149}$$

  where $\sigma_{\rm T}$ is the Thomson scattering cross section,

$$\sigma_{\rm T} = \frac{8\pi}{3}\left(\frac{e^2}{m_{\rm e}c^2}\right)^2 = 6.65\times10^{-25}\,{\rm cm}^2\ , \tag{1.150}$$

  and ${\rm d}r = c{\rm d}t = c{\rm d}a/\dot{a}$ is the proper length interval.

- The probability distribution for a photon to be scattered between z and $z - {\rm d}z$ is

$$p(z){\rm d}z = {\rm e}^{-\tau}\frac{{\rm d}\tau}{{\rm d}z}{\rm d}z\ . \tag{1.151}$$

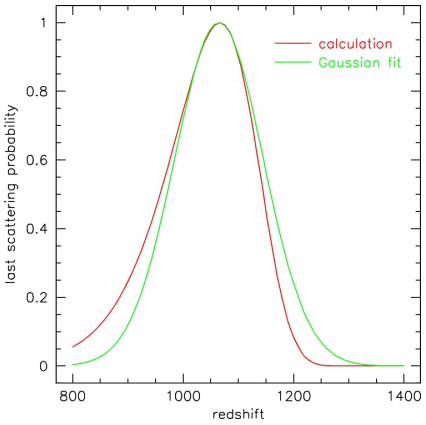
  This distribution is well described by a Gaussian with mean $\bar{z} = 1100$ and standard deviation $\sigma_z \approx 80$.

- The finite width of the last-scattering shell implies that microwave background photons seen today were released at different redshifts. Since the plasma cooled as recombination proceeded, the CMB photons were released at different temperatures. Since $T = T_0(1+z)$,

$$\delta T \approx T_0\delta z \approx T_0\sigma_z \approx 200\,{\rm K}\ . \tag{1.152}$$

  This is a sizeable temperature difference.

- Photons were redshifted after their emission. Those emitted earlier from somewhat hotter plasma were redshifted somewhat



Detailed calculation of and Gaussian fit to the last-scattering probability distribution as a function of redshift.

more, and vice versa for photons emitted later. These effects cancel exactly in Friedmann-Lemaître models because $T \propto a^{-1}$. Despite the CMB photons originate from plasma with a range of temperatures, the CMB is thus expected to have a Planck spectrum of a *single* temperature.

1.4.5 Nucleosynthesis

- As the universe expands and cools, it passes through a temperature range which allows the fusion of light nuclei. The faster the expansion, the less time there is for nucleosynthesis, thus the light-element abundances measure the expansion rate in the early universe.

- Protons and neutrons form at the QCD phase transition when $kT \approx 160\,\mathrm{MeV}$. Afterwards, they can interconvert through the weak interaction, e.g.

$$n + \nu_\mathrm{e} \leftrightarrow p + e^- \,, \tag{1.153}$$

  and remain in thermal equilibrium until weak interactions freeze out at $kT \approx 800\,\mathrm{keV}$.

- At this point, the neutron-to-proton number-density ratio was

$$\frac{n_\mathrm{n}}{n_\mathrm{p}} = \mathrm{e}^{-\Delta mc^2/kT} = \frac{1}{6} \,, \tag{1.154}$$

  where $\Delta mc^2 = 1.4\,\mathrm{MeV}$ is the mass difference between neutrons and protons.

- Fusion builds upon two-body processes because the probability for others is too low. The first element to form is deuterium, the next are helium isotopes, followed by Lithium. Examples are

$$\begin{aligned} n + p &\rightarrow {}^2\mathrm{H} + \gamma \\ {}^2\mathrm{H} + {}^2\mathrm{H} &\rightarrow {}^3\mathrm{He} + n \\ {}^3\mathrm{He} + {}^2\mathrm{H} &\rightarrow {}^4\mathrm{He} + p \\ {}^2\mathrm{H} + {}^2\mathrm{H} &\rightarrow {}^3\mathrm{H} + p \\ {}^4\mathrm{He} + {}^3\mathrm{H} &\rightarrow {}^7\mathrm{Li} + \gamma \,. \end{aligned} \tag{1.155}$$

  The absence of stable nuclei with atomic weights $A = 5$ and $A = 8$ and increasing coulomb barriers make the production of heavier elements highly inefficient.

- The equilibrium of deuterium formation $n + p \leftrightarrow {}^2\mathrm{H} + \gamma$ is controlled by Saha's equation. As for recombination, high photon density prevents ${}^2\mathrm{H}$ formation through photo dissociation until

temperature has dropped well below $kT \approx 2\,\mathrm{MeV}$ corresponding to the binding energy. ^{2}H formation is delayed until $kT \approx 80\,\mathrm{keV}$, about three minutes after the Big Bang.

- This is well before matter-radiation equality, thus the density of relativistic particles (photons, neutrinos, others?) (Others?) controls the expansion rate, and baryon-to-photon ratio η is the only relevant parameter,

$$\eta = 10^{-10}\eta_{10}\,, \quad \eta_{10} = 273\Omega_{\mathrm{B}}h^2\,. \tag{1.156}$$

- Deuterium is crucial. If too much ^{2}H is formed, neutrons are locked up, no heavier elements can form. If too little ^{2}H is formed, an important agent for further fusion is missing. The ^{2}H production rate needs to be "just right",

$$n_{\mathrm{B}}\langle\sigma v\rangle t \approx 1\,. \tag{1.157}$$

This is the Gamow criterion.

- The velocity-averaged fusion cross section $\langle\sigma v\rangle$ is known. The time t is determined by the expansion rate, i.e. the photon density or photon temperature T. The Gamow criterion can thus be used for estimating T from constraints on the baryon density n_{B}.

- Neutrons are in equilibrium with protons until $kT \approx 800\,\mathrm{keV}$ and consumed in efficient fusion after $kT \approx 80\,\mathrm{keV}$. In between, they decay with a half-life of

$$t_{\mathrm{n}} = 886.7 \pm 1.9\,\mathrm{s}\,. \tag{1.158}$$

Accordingly, the neutron-to-proton ratio drops to

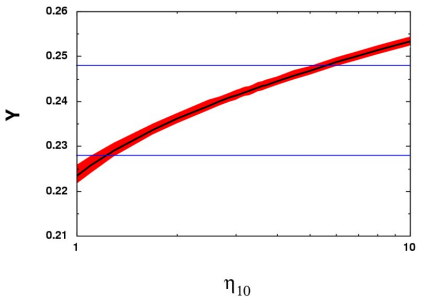
$$\frac{n_{\mathrm{n}}}{n_{\mathrm{p}}} = \frac{1}{7}\,. \tag{1.159}$$

- Once ^{2}H exists, neutrons are efficiently locked up into ^{4}He because of its high binding energy. The expected primordial ^{4}He abundance by mass (Y_{p}) is thus

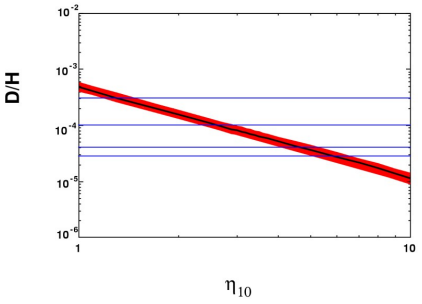
$$Y_{\mathrm{p}} = 1 - X_{\mathrm{p}} \approx 1 - \frac{n_{\mathrm{p}} - n_{\mathrm{n}}}{n_{\mathrm{p}} + n_{\mathrm{n}}} = \frac{2n_{\mathrm{n}}}{n_{\mathrm{p}} + n_{\mathrm{n}}} = \frac{2(n_{\mathrm{n}}/n_{\mathrm{p}})}{1 + n_{\mathrm{n}}/n_{\mathrm{p}}} = \frac{1}{4}\,, \tag{1.160}$$

where X_{p} denotes the primordial hydrogen abundance by mass. This number is relatively insensitive to the baryon density, and thus to η.

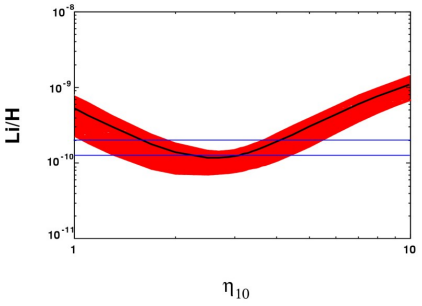
- Expected trends of light-element abundances with η are:
  - Gentle increase of Y_{p} with increasing η as nucleosynthesis starts earlier.



Helium abundance as a function of η



Deuterium abundance as a function of η



Lithium abundance as a function of η

  - ^{2}H and ^{3}He are burnt by fusion, thus their abundances decrease as η increases, which implies more available time for burning.
  - ^{7}Li is destroyed by protons at low η with an efficiency increasing with η. Its precursor ^{7}Be is produced more efficiently as η increases. Thus, a ^{7}Li valley is formed.

- Element abundances are calculated using Monte-Carlo codes. The main uncertainties are the interaction rates and the half-life of free neutrons. 2-σ prediction uncertainties are $\sim 0.4\%$ for ^{4}He, $\sim 15\%$ for ^{2}H and ^{3}He, and $\sim 42\%$ for ^{7}Li at $\eta_{10} = 5$.

- Comparison with observations is difficult because light elements get produced and consumed (e.g. in stars) during cosmic history. Objects need to be found which either retain the primordial element mix, or in which abundance changes can be constrained:

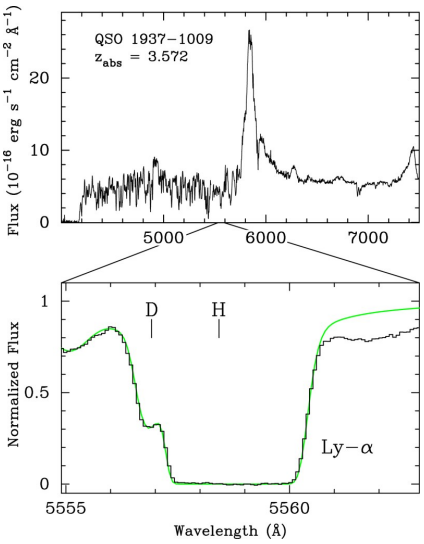
  - ^{2}H is observed in neutral hydrogen gas via resonant UV absorption from the ground state, or via the hyperfine transition of the ground state, or via ^{2}H-H molecule lines.
  - ^{3}He$^+$ is observed via the hyperfine transition of the ground state.
  - ^{4}He is probed by optical recombination line emission in ionised hydrogen (HII-regions).
  - ^{7}Li is observed in the spectra of cool, low-mass stars in the Galactic halo (very old, local stellar population).

  Heavy elements are formed by stars as early as $z \sim 6$, so observations need to concentrate on gas with lowest metal abundance. Possibly observed dependence of light-element abundances on metal abundance may allow extrapolation to zero enrichment.

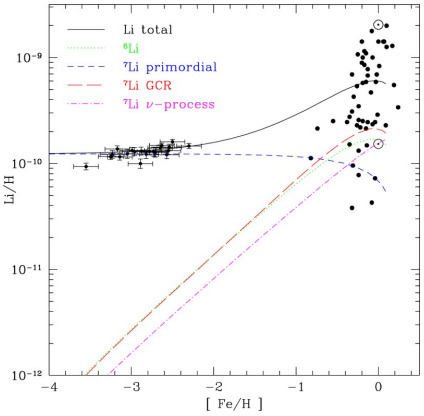
- It is assumed that evolutionary corrections for ^{2}H, ^{4}He and ^{7}Li are low or negligible, but highly uncertain for ^{3}He because of later production in pre-main sequence stars and destruction in stellar interiours.

- ^{2}H is ideal baryometer because of monotonic abundance decrease with increasing η. Destroyed by later fusion, so observed abundances are lower bound to primordial abundance. (Folgender Satz hat keinen Hauptsatz) Can be observed in high-z quasar spectra which require high resolution to allow accurate continuum subtraction, corrections for saturation and velocity shifts in hydrogen lines. Such measurements find

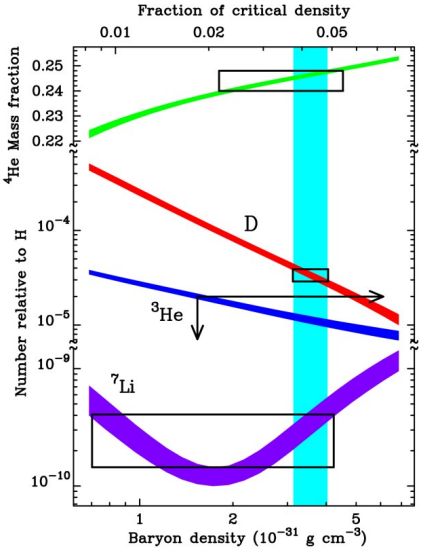
$$\frac{n_{^2\mathrm{H}}}{n_{\mathrm{H}}} = (3-4)\times 10^{-5}\ , \tag{1.161}$$



Deuterium line in a high-redshift quasar spectrum.



The Spite plateau in the Lithium abundance.



Results from Big-Bang nucleosynthesis.

at 95% confidence. Substantial depletion is unlikely because it should have increased metal abundance. Somewhat lower values are seen in the interstellar medium consistent with consumption.

- ^{4}He observations suffer from systematic uncertainties due to necessary metallicity corrections, the interpretation of stellar absorption spectra and collisional excitation of observed recombination lines. A conservative range for the ^{4}He abundance is

$$Y_\mathrm{p} = 0.238 \pm 0.01 \; . \tag{1.162}$$

- ^{7}Li is observed in low-metallicity halo stars which should have locked up very nearly primordial gas, but they may have processed it. Cool stellar atmospheres are difficult to model. Stellar rotation is important because it induces mixing. ^{7}Li may also have been produced by cosmic-ray spallation on the interstellar medium.

- ^{7}Li abundance against iron abundance shows the *Spite plateau* with very little dispersion,

$$A_\mathrm{Li,p} := 12 + \log \frac{n_{^7\mathrm{Li}}}{n_\mathrm{H}} = 2.2 \pm 0.1 \; . \tag{1.163}$$

  Necessary corrections seem to be moderate.

- Results from Big-Bang nucleosynthesis theory and observations can be summarised as follows:

  - Through (1.156), density of *visible* baryons implies $\eta_{10} \geq 1.5$.
  - ^{2}H abundance (1.161) implies $4.2 \leq \eta_{10} \leq 6.3$
  - ^{7}Li abundance predicted assuming this range of η_{10} is $2.1 \leq A_\mathrm{Li,p} \leq 2.8$, consistent with the observed value (1.163).
  - This yields $0.244 \leq Y_\mathrm{p} \leq 0.250$, overlapping with measured range (1.162).

- The baryon density implied by Big-Bang nucleosynthesis is

$$\Omega_\mathrm{B} h^2 = 0.019 \pm 0.0024 \; , \tag{1.164}$$

  at 95% significance. It is mainly based on the high-z deuterium abundance, but yields a consistent set of light-element abundances.

Chapter 2

The Inhomogeneous Universe

2.1 The Growth of Perturbations

2.1.1 Newtonian Equations

- There are pronounced structures in the universe on scales from stars to galaxy clusters and filaments. While filaments and the voids they surround can reach sizes of $\sim 50\,h^{-1}$Mpc, they are still small compared to the Hubble radius. In this chapter, we describe the basic theory for structure growth in the expanding universe.

- Strictly, this theory should be worked out in the framework of general relativity, which is a complicated exercise. With the inhomogeneities being "small", i.e. much smaller than the typical scale of the universe, we can neglect effects of curvature and the finite speed of information propagation and work within the framework of Newtonian dynamics.

- The dynamics of stars in galaxies, and of galaxies in galaxy clusters, shows that these objects need to contain much more matter than can be inferred from the light they emit. This is evidence for the existence of "dark matter" in the universe which dominates its matter content.

- We thus need to describe inhomogeneities in a cosmic fluid which contains at least radiation, dark matter, and baryonic matter and which moves according to Newtonian gravity.

- We begin with the continuity equation, which formulates mass conservation,
$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \,, \tag{2.1}$$
where $\rho(t, \vec{x})$ and $\vec{v}(t, \vec{x})$ are the density and velocity of the cosmic fluid at position $\vec{x}$ and time t. In contrast to the homogeneous universe, they now depend on position.

- The second equation is Euler's equation which formulates the conservation of momentum,
$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{\vec{\nabla} p}{\rho} - \vec{\nabla}\Phi \,. \tag{2.2}$$
The terms on the right-hand side represent the pressure-gradient and gravitational forces.

- The Newtonian gravitational potential Φ satisfies the Poisson equation
$$\nabla^2 \Phi = 4\pi G \rho \,. \tag{2.3}$$

2.1.2 Perturbation Equations

- We now decompose density and velocity into their homogeneous background values ρ_0 and $\vec{v}_0$ and small perturbations $\delta\rho$ and $\delta\vec{v}$,

$$\rho(t,\vec{x}) = \rho_0(t) + \delta\rho(t,\vec{x}) \;, \quad \vec{v}(t,\vec{x}) = \vec{v}_0(t) + \delta\vec{v}(t,\vec{x}) \;. \tag{2.4}$$

- Let $\vec{r}$ and $\vec{x}$ be physical and comoving coordinates, respectively, then $\vec{r} = a\vec{x}$ and the velocity is

$$\vec{v} = \dot{\vec{r}} = \dot{a}\vec{x} + a\dot{\vec{x}} = H\vec{r} + a\dot{\vec{x}} = \vec{v}_0 + \delta\vec{v} \;, \tag{2.5}$$

  i.e. $\vec{v}_0 = H\vec{r}$ is the Hubble velocity, and $\delta\vec{v} = a\dot{\vec{x}}$ is the peculiar velocity deviating from the Hubble flow.

- Inserting (2.4) into (2.1) and keeping only terms up to first order yields

$$\frac{\partial(\rho_0 + \delta\rho)}{\partial t} + \vec{\nabla}\cdot(\rho_0\vec{v}_0 + \delta\rho\vec{v}_0 + \rho_0\delta\vec{v}) = 0 \;. \tag{2.6}$$

  The background quantities ρ_0 and $\vec{v}_0$ need to satisfy mass conservation separately,

$$\frac{\partial\rho_0}{\partial t} + \rho_0\vec{\nabla}\cdot\vec{v}_0 = \frac{\partial\rho_0}{\partial t} + 3H\rho_0 = 0 \;, \tag{2.7}$$

  where $\vec{v}_0 = H\vec{r}$ and $\vec{\nabla}\cdot\vec{r} = 3$ were used. Thus

$$\frac{\partial\delta\rho}{\partial t} + \vec{v}_0\cdot\vec{\nabla}\delta\rho + \rho_0\vec{\nabla}\cdot\delta\vec{v} + \delta\rho\vec{\nabla}\cdot\vec{v}_0 = 0 \;. \tag{2.8}$$

- Defining the density contrast,

$$\delta := \frac{\delta\rho}{\rho_0} \;, \tag{2.9}$$

  we find

$$\frac{\partial\delta\rho}{\partial t} = \delta\dot{\rho}_0 + \dot{\delta}\rho_0 = -\delta\rho_0\vec{\nabla}\cdot\vec{v}_0 + \dot{\delta}\rho_0 \;, \tag{2.10}$$

  using the unperturbed continuity equation (2.7). The perturbed continuity equation (2.8) can now be written

$$\dot{\delta} + \vec{v}_0\cdot\vec{\nabla}\delta + \vec{\nabla}\cdot\delta\vec{v} = 0 \;. \tag{2.11}$$

- Likewise, we split the momentum conservation equation (2.2) into unperturbed and perturbed parts, where we introduce the pressure and potential perturbations δp and $\delta\Phi$, and keep only terms to linear order,

$$\frac{\partial\delta\vec{v}}{\partial t} + (\delta\vec{v}\cdot\nabla)\vec{v}_0 + (\vec{v}_0\cdot\vec{\nabla})\delta\vec{v} = -\frac{\vec{\nabla}\delta p}{\rho_0} - \vec{\nabla}\delta\Phi \;. \tag{2.12}$$

  Written in components, the term $(\delta\vec{v}\cdot\vec{\nabla})\vec{v}_0$ reads

$$\left[(\delta\vec{v}\cdot\vec{\nabla})\vec{v}_0\right]_i = \left(\delta v_j\frac{\partial}{\partial r_j}\right)Hr_i = H\delta_{ij}(\delta\vec{v})_j = H(\delta\vec{v})_i \;. \tag{2.13}$$

- Treated similarly, the Poisson equation becomes

$$\nabla^2 \delta\Phi = 4\pi G \rho_0 \delta \; . \tag{2.14}$$

- We now convert to comoving coordinates, $\vec{x} = \vec{r}/a$ and comoving peculiar velocities, $\vec{u} := \delta\vec{v}/a$, and introduce the gradient with respect to the comoving coordinates,

$$\vec{\nabla}_r = \frac{1}{a}\vec{\nabla}_x \; . \tag{2.15}$$

- Likewise, we have to transform the time derivative. The total differential of an arbitrary function $f(\vec{r}, t)$ is

$$\begin{aligned} \mathrm{d}f &= \frac{\partial f}{\partial t}\mathrm{d}t + \vec{\nabla}_r f \cdot \mathrm{d}\vec{r} = \frac{\partial f}{\partial t}\mathrm{d}t + \vec{\nabla}_r f \cdot a(H\vec{x}\mathrm{d}t + \mathrm{d}\vec{x}) \\ &= \left(\frac{\partial f}{\partial t} + H\vec{x}\cdot\vec{\nabla}_x f\right)\mathrm{d}t + \vec{\nabla}_x f \cdot \mathrm{d}\vec{x} \; , \end{aligned}$$

  hence, the partial time derivative in physical coordinates needs to be replaced according to

$$\frac{\partial}{\partial t} + H\vec{x}\cdot\vec{\nabla}_x \to \frac{\partial}{\partial t} \; . \tag{2.16}$$

  In order to keep notation simple, $\vec{\nabla}$ abbreviates $\vec{\nabla}_x$ hereafter.

- We are now left with the three perturbation equations

$$\begin{aligned} \dot{\delta} + \vec{\nabla}\cdot\vec{u} &= 0 \\ \dot{\vec{u}} + 2H\vec{u} &= -\frac{\vec{\nabla}\delta p}{a^2\rho_0} - \frac{\vec{\nabla}\delta\Phi}{a^2} \\ \nabla^2\delta\Phi &= 4\pi G\rho_0 a^2 \delta \; , \end{aligned} \tag{2.17}$$

  for the four variables δ, $\vec{u}$, δp, and $\delta\Phi$. The over-dots denote partial time derivatives. We additionally need an equation of state linking the pressure fluctuation to the density fluctuation,

$$\delta p = \delta p(\delta) = c_\mathrm{s}^2 \delta\rho = c_\mathrm{s}^2 \rho_0 \delta \; , \tag{2.18}$$

  with the sound speed c_s.

2.1.3 Density Perturbations

- Taking the divergence of the Euler equation, we find an equation for $\vec{\nabla}(\dot{\vec{u}}) = \mathrm{d}(\vec{\nabla}\cdot\vec{u})/\mathrm{d}t$, which can be inserted into the total time derivative of the continuity equation. This yields the single equation for the density contrast

$$\ddot{\delta} + 2H\dot{\delta} = \left(4\pi G\rho_0\delta + \frac{c_\mathrm{s}^2\nabla^2\delta}{a^2}\right) \; . \tag{2.19}$$

- We can decompose δ into plane waves,

$$\delta(\vec{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \hat{\delta}(\vec{k}, t) \mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}} \,, \tag{2.20}$$

  introducing the time-dependent Fourier amplitudes $\hat{\delta}(\vec{k}, t)$ and decoupling the time evolution from the spatial dependence. Inserted into (2.19), this yields

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \hat{\delta}\left(4\pi G\rho_0 - \frac{c_\mathrm{s}^2 k^2}{a^2} \,.\right) \tag{2.21}$$

- Starting from special-relativistic fluid mechanics, and ignoring pressure gradients, the perturbation equations for an ideal relativistic fluid (e.g. photons) can be derived in a very similar way. This leads to the replacements $\vec{\nabla}\cdot(\rho\vec{v}) \to \vec{\nabla}\cdot[(\rho + p/c^2)\vec{v}]$ in the continuity equation, and $\rho \to \rho + 3p/c^2$ in the Poisson equation. With the pressure $p = \rho c^2/3$ for radiation domination, this means substituting $\vec{\nabla}\cdot(\rho\vec{v}) \to 4/3\vec{\nabla}\cdot(\rho\vec{v})$ in the continuity equation and $\rho \to 2\rho$ in the Poisson equation. Ignoring pressure forces, the result is the evolution equation

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = 2 \cdot \frac{4}{3} \cdot 4\pi G\rho_0\hat{\delta} = \frac{32\pi}{3} G\rho_0\hat{\delta} \,. \tag{2.22}$$

- On a static background, $H = 0$, and (2.21) becomes the oscillator equation

$$\ddot{\hat{\delta}} + \omega_0^2\hat{\delta} = 0 \,, \quad \omega_0 := \sqrt{\frac{c_\mathrm{s}^2 k^2}{a^2} - 4\pi G\rho_0} \,. \tag{2.23}$$

  The oscillation frequency is real for sufficiently large comoving wave numbers k,

$$k \geq k_\mathrm{J} := \frac{2\sqrt{\pi G\rho_0}a}{c_\mathrm{s}} \,. \tag{2.24}$$

  k_J defines the comoving Jeans length

$$\lambda_\mathrm{J} := \frac{2\pi}{k_\mathrm{J}} = \frac{c_\mathrm{s}}{a}\sqrt{\frac{\pi}{G\rho_0}} \,. \tag{2.25}$$

  Perturbations smaller than the Jeans length oscillate. Others grow or decay. The Jeans length can be empirically derived by balancing the sound crossing time, $t_\mathrm{s} = a\lambda_\mathrm{J}/c_\mathrm{s} = 2\pi a/(k_\mathrm{J}c_\mathrm{s})$, with the gravitational free-fall time, $t_\mathrm{ff} = \sqrt{\pi/(G\rho_0)}$, which yields the desired result (2.24).

- We now study the behaviour of perturbations on scales much larger than the Jeans length, or in pressure-less fluids. If $\Omega = 1$, the perturbation equations read

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \frac{3}{2}H^2\hat{\delta} \,, \quad \ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = 4H^2\hat{\delta} \,, \tag{2.26}$$

for the matter- and radiation-dominated cases, respectively, for which we have from (1.52) and (1.49)

$$\frac{\dot{a}}{a} = H(t) = \frac{2}{3t} \,, \quad \frac{\dot{a}}{a} = H(t) = \frac{1}{2t} \,. \tag{2.27}$$

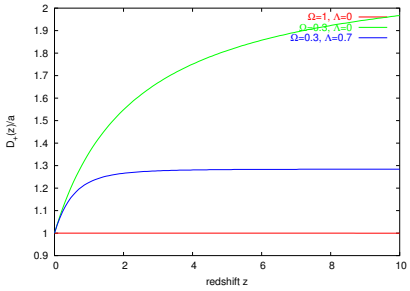
- The *ansatz* $\hat{\delta}(\vec{k}, t) \propto t^n$ yields

$$n^2 + \frac{n}{3} - \frac{2}{3} = 0 \,, \quad n^2 - 1 = 0 \tag{2.28}$$

hence $n = -1, 2/3$ in the matter-dominated and $n = \pm 1$ in the radiation-dominated cases, which translates to

$$\hat{\delta} \propto \left\{ \begin{array}{c} a \\ a^{-3/2} \end{array} \right\} \text{ matter-dominated era} \qquad \left\{ \begin{array}{c} a^2 \\ a^{-2} \end{array} \right\} \text{ radiation-dominated era} \tag{2.29}$$

Decaying modes are irrelevant for cosmic structure growth, so $\delta \propto a^2$ during the radiation-dominated era, and $\delta \propto a$ afterwards. The phases of the waves determine whether a given cosmological patch develops into an underdense region (i.e., a void) or a galaxy cluster. Constructive interference of the growing modes causes the development of overdensities, which then collapse into galaxies (in the case of small-scale modes) or clusters (for large-scale modes). Destructive interference leads to the growth of voids.



Linear growth factor D_+/a as a function of redshift for different cosmologies.

- During the matter-dominated era in models with $\Omega_{\mathrm{m},0} \neq 1$ and $\Omega_{\Lambda,0}$, the linear evolution of the density contrast follows

$$\delta(a) = \delta_0 D_+(a) \,, \tag{2.30}$$

with the linear growth factor

$$\begin{aligned} D_+(a) &= \frac{g(a)}{g(1)} \,, \\ g(a) &= \frac{5}{2}\Omega_{\mathrm{m}0} E(a) \int_0^a \frac{\mathrm{d}a'}{a'^3 E^3(a')} \,. \end{aligned} \tag{2.31}$$

In excellent approximation,

$$g(a) = \frac{5a}{2}\Omega_{\mathrm{m}} \left[\Omega_{\mathrm{m}}^{4/7} - \Omega_\Lambda + \left(1 + \frac{1}{2}\Omega_{\mathrm{m}}\right)\left(1 + \frac{1}{70}\Omega_\Lambda\right) \right]^{-1} \,. \tag{2.32}$$

The Λ term causes a faster growth of structure at earlier times and slows down structure formation at late times (in comparison to an Einstein-de-Sitter universe). Well after the dominance of the cosmological constant, structure formation eventually comes to a halt and structures that have not had time to virialize yet (such as super clusters) can be pulled apart by the gravitational repulsion of the cosmological constant such that they will never collapse in the future.

- The sound speed defines the Jeans length, below which perturbations cannot grow, but oscillate. For dark matter consisting of weakly interacting massive particles, for instance, the concept of a sound speed makes no sense because the dark matter behaves like an ensemble of collision-less particles. In that case, one can show that the comoving Jeans length (2.24) is replaced by

$$\lambda_\mathrm{J} = \frac{\left\langle v^{-2}\right\rangle^{-1/2}}{a}\sqrt{\frac{\pi}{G\rho_0}}\,, \tag{2.33}$$

where v is the velocity dispersion of the particles. Perturbations in collision-less matter smaller than the Jeans length are thus prevented from growing because their gravity is insufficient for keeping their particles bound.

- (Hypothetical) forms of dark matter with $v \to 0$ are called "cold dark matter" (CDM). They have $\lambda_\mathrm{J} \to 0$, hence structures can grow on all scales. If v is finite as it would be for neutrinos, the matter is called "hot dark matter" (HDM).

2.1.4 Velocity Perturbations

- Ignoring pressure gradients, the second equation (2.17) says

$$\dot{\vec{u}} + 2H\vec{u} = -\frac{\vec{\nabla}\delta\Phi}{a^2}\,. \tag{2.34}$$

The peculiar velocity field must thus be aligned with the gradient of the potential perturbation. We attempt solving the continuity equation using the *ansatz* $\vec{u} = u(t)\vec{\nabla}\delta\Phi$,

$$u(t)\vec{\nabla}^2\delta\Phi = u(t)\,4\pi G\rho_0 a^2\delta = -\dot{a}\frac{\mathrm{d}\delta}{\mathrm{d}a}\,. \tag{2.35}$$

- For linearly growing perturbations, we have

$$\frac{\mathrm{d}\delta}{\mathrm{d}a} = \delta_0\frac{\mathrm{d}D_+(a)}{\mathrm{d}a} = \frac{\delta}{a}\frac{\mathrm{d}\ln D_+(a)}{\mathrm{d}\ln a} =: \frac{\delta}{a}f(\Omega_\mathrm{m})\,, \tag{2.36}$$

where

$$f(\Omega_\mathrm{m}) := \frac{\mathrm{d}\ln D_+(a)}{\mathrm{d}\ln a} \approx \Omega_\mathrm{m}^{0.6}\,, \tag{2.37}$$

is an excellent approximation. Moreover, we insert

$$4\pi G\rho_0 = 4\pi G\frac{3H^2}{8\pi G}\Omega_\mathrm{m} = \frac{3H^2\Omega_\mathrm{m}}{2} \ , \tag{2.38}$$

into (2.35) and find

$$u(t) = -\frac{2f(\Omega_\mathrm{m})}{3a^2H\Omega_\mathrm{m}} \ . \tag{2.39}$$

- The peculiar velocity field satisfying the continuity equation can thus be written as

$$\delta\vec{v} = a\vec{u} = -\frac{2f(\Omega_\mathrm{m})}{3aH\Omega_\mathrm{m}}\vec{\nabla}\delta\Phi \ . \tag{2.40}$$

  This solution states that a potential well (that is caused by an upwards fluctuation of the density) gravitationally attracts matter in the surroundings and pulls on it. This causes this matter to flow in (hence the minus sign) and to accrete, which further deepens the well. Additional solutions are possible which exhibit vorticity, $\vec{u} = \vec{\nabla}\times\vec{a}$, such that $\vec{\nabla}\cdot\vec{u} = 0$. Since δ can either grow or decay, $\dot{\delta} = 0$ and $\vec{\nabla}\cdot\vec{u} = 0$ can occur only where $\delta = 0$.

2.2 Statistics and Non-linear Evolution

2.2.1 Power Spectra

- We have seen before (2.20) that it is convenient to decompose the density contrast δ into plane waves. We introduce the Fourier transform $\hat{\delta}$ of the density contrast δ as

$$\delta(\vec{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\hat{\delta}(\vec{k})\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}}\,, \quad \hat{\delta}(\vec{k}) = \int \mathrm{d}^3x\delta(\vec{x})\mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}\,. \tag{2.41}$$

- The density contrast is a random field, which must be isotropic and homogeneous in order to comply with the fundamental cosmological assumptions. This means that the statistical properties of δ, e.g. its mean or variance, do not change under rotations and translations.

- By definition, the mean of the density contrast vanishes,

$$\langle\delta\rangle = \left\langle\frac{\rho-\rho_0}{\rho_0}\right\rangle = \frac{\langle\rho\rangle}{\rho_0} - 1 = 0\,. \tag{2.42}$$

  The variance of δ *in Fourier space* defines the power spectrum $P(k)$,

$$\langle\hat{\delta}(\vec{k})\hat{\delta}^*(\vec{k}')\rangle =: (2\pi)^3P(k)\delta_\mathrm{D}(\vec{k}-\vec{k}')\,, \tag{2.43}$$

  where δ_D is Dirac's delta distribution, which ensures that modes of different wave vector $\vec{k}$ are uncorrelated in Fourier space in order to ensure homogeneity. The power spectrum cannot depend on the direction of $\vec{k}$ because of isotropy.

- The correlation function of δ in real space is defined as

$$\xi(y) := \langle\delta(\vec{x})\delta(\vec{x}+\vec{y})\rangle\,, \tag{2.44}$$

  where the average extends over all positions $\vec{x}$ and orientations of $\vec{y}$. The correlation function measures the coherence of the density contrast between all points on the sky separated by a distance $|\vec{y}|$. Again, ξ cannot depend on the direction of $\vec{y}$ because of isotropy.

- Inserting the Fourier integrals for $\delta(\vec{x})$ in (2.44), we find

$$\begin{aligned}\xi(y) &= \left\langle\int\frac{\mathrm{d}^3k}{(2\pi)^3}\int\frac{\mathrm{d}^3k'}{(2\pi)^3}\hat{\delta}(\vec{k})\hat{\delta}(\vec{k}')\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}}\mathrm{e}^{-\mathrm{i}\vec{k}'(\vec{x}+\vec{y})}\right\rangle \\ &= \int\frac{\mathrm{d}^3k}{(2\pi)^3}\int\frac{\mathrm{d}^3k'}{(2\pi)^3}\langle\hat{\delta}(\vec{k})\hat{\delta}^*(\vec{k}')\rangle\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{x}}\mathrm{e}^{+\mathrm{i}\vec{k}'(\vec{x}+\vec{y})} \\ &= \int\frac{\mathrm{d}^3k}{(2\pi)^3}P(k)\mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{y}}\,,\end{aligned} \tag{2.45}$$

which states that the correlation function is the Fourier transform of the power spectrum (and vice versa). Hence, both statistical measures carry an equivalent amount of information. Simplifying furthermore, we obtain

$$\begin{aligned}\xi(y) &= 2\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \int_0^\pi \sin\theta \mathrm{d}\theta \mathrm{e}^{-\mathrm{i}ky\cos\theta} \\ &= 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \frac{\sin ky}{ky} \,, \end{aligned} \tag{2.46}$$

where θ was the angle between vectors $\vec{k}$ and $\vec{y}$. Obviously, the variance of δ is the correlation function at $y = 0$,

$$\sigma^2 = 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \,. \tag{2.47}$$

- The variance in real space depends on the scale which we are considering. Let us introduce

$$\bar{\delta}(\vec{x}) := \int \mathrm{d}^3 y \delta(\vec{x}) W_R(|\vec{x} - \vec{y}|) \,, \tag{2.48}$$

i.e. the density contrast field averaged on the scale R with a *window function* W_R. The idea of the window function is that it approaches a finite constant well within R, and drops to zero outside R.

- The Fourier convolution theorem says $\widehat{f * g} = \hat{f}\hat{g}$, i.e. the Fourier transform of a convolution is the product of the Fourier transforms of the convolved functions. Applying this to (2.47) yields $\hat{\bar{\delta}} = \hat{\delta}\hat{W}_R$. thus, the power spectrum of the density contrast filtered on the scale R is $\bar{P}(k) = P(k)\hat{W}_R^2(k)$. Using (2.47), the variance of the filtered density-contrast field is

$$\sigma_R^2 = 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \hat{W}_R^2(k) \,. \tag{2.49}$$

The variance on a scale of $8\,h^{-1}$ Mpc, σ_8, is often used for characterising the amplitude of the power spectrum.

2.2.2 Evolution of the Power Spectrum

- We have seen in (2.29) that density perturbations grow $\propto a^2$ during the radiation-dominated era, and $\propto a$ afterwards.

- As the universe expands, the Hubble radius grows, and thus the scale of perturbations which can be in causal contact. A density perturbation mode is said to "enter the horizon" when its wave length λ equals the Hubble radius.

- Modes entering the horizon while radiation dominates feel the radiation pressure, which almost completely stops the growth of the density perturbation until matter starts dominating and radiation pressure quickly becomes negligible. Accordingly, modes which are small enough to enter the horizon before $a_{\rm eq}$ are relatively suppressed compared to larger modes which enter the horizon afterwards.

- Modes of comoving wave number k enter the horizon at $a_{\rm eq}$ if

$$\lambda = \lambda_0 = a_{\rm eq}\frac{2\pi}{k_0} = r_{\rm H,eq} = \frac{c}{H_0}\frac{a_{\rm eq}^{3/2}}{\sqrt{2\Omega_{\rm m0}}} \; . \tag{2.50}$$

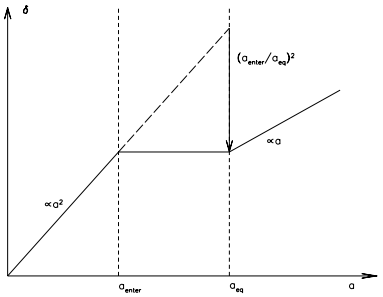
  Thus, the wave number of modes entering the horizon at $a_{\rm eq}$ is

$$k_0 = 2\pi\frac{H_0}{c}\sqrt{\frac{2\Omega_{\rm m0}}{a_{\rm eq}}} = 2\pi\frac{H_0}{c}\Omega_{\rm m0}\sqrt{\frac{2}{\Omega_{\rm r0}}} \; . \tag{2.51}$$

  Modes larger than this, i.e. with $k < k_0$, continue growing. Modes with $k > k_0$ stop growing when they enter the horizon at $a_{\rm enter}$ and continue only after $a_{\rm eq}$ when radiation ceases to dominate.

- According to (1.68), the Hubble radius scales like $\propto a^2$ during radiation domination and $\propto a^{3/2}$ later, hence $a_{\rm enter}$ is determined by

$$\begin{aligned} a_{\rm enter}\lambda &= a_{\rm enter}\frac{2\pi}{k} \propto \begin{cases} a_{\rm enter}^2 & (a_{\rm enter} < a_{\rm eq}) \\ a_{\rm enter}^{3/2} & (a_{\rm enter} > a_{\rm eq}) \end{cases} \\ \Rightarrow \qquad a_{\rm enter} &\propto \begin{cases} k^{-1} & (a_{\rm enter} < a_{\rm eq}) \\ k^{-2} & (a_{\rm enter} > a_{\rm eq}) \end{cases} \end{aligned} \tag{2.52}$$



Growth suppression during the radiation-dominated era.

- While the growth of small modes is suppressed, modes larger than λ_0 continue growing $\propto a^2$ during radiation domination, hence the relative suppression of the small modes is

$$f_{\rm sup} = \left(\frac{a_{\rm enter}}{a_{\rm eq}}\right)^2 = \left(\frac{k_0}{k}\right)^2 \; . \tag{2.53}$$

- Suppose the initial power spectrum at very early times is $P_{\rm i}(k)$. If modes enter the horizon before $a_{\rm eq}$, the spectrum is $P_{\rm enter}(k) = a_{\rm enter}^4 P_{\rm i}(k)$, and $P_{\rm enter}(k) = a_{\rm enter}^2 P_{\rm i}(k)$ if they enter afterwards. In both cases, $P_{\rm enter}(k) = k^{-4}P_{\rm i}(k)$ because of (2.52).

- The total power in density fluctuations on scales $2\pi/k$ is $k^3P(k)$. Assuming that the power entering the horizon should not depend on time, the initial power spectrum must satisfy

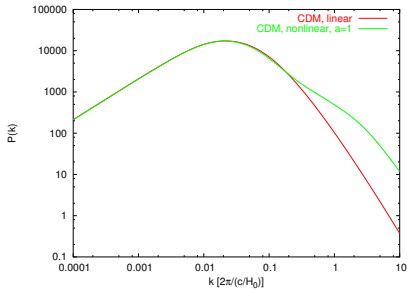
$$k^3 P_{\rm enter}(k) = k^3 \cdot k^{-4} P_{\rm i}(k) = \text{const.} \;\Rightarrow\; P_{\rm i}(k) \propto k \; . \tag{2.54}$$

  This is called the Harrison-Zel'dovich-Peebles spectrum.

- For $k < k_0$ the shape of the spectrum is unchanged because all such modes grow similarly. For $k > k_0$, suppression $\propto f_{\rm sup}^2 \propto k^{-4}$ sets in. Thus, we expect the spectrum to behave like

$$P(k) \propto \begin{cases} k & (k < k_0) \\ k^{-3} & (k \gg k_0) \end{cases} \tag{2.55}$$

This is the shape of the spectrum for cold dark matter (CDM). For hot dark matter (HDM), it is cut off above the Jeans wave number $k_{\rm J}$ corresponding to the finite velocity dispersion of the hot particles.



Linear and non-linear CDM power spectra.

2.2.3 The Zel'dovich Approximation

- Once the density contrast δ approaches unity, the linear description of its evolution will break down. A kinematical treatment for following the evolution further into the non-linear regime was invented by Zel'dovich.

- It starts by decomposing the cosmic fluid into particles and writing their (physical) trajectories as

$$\vec{r}(t) = a(t)\vec{x} + b(t)\vec{f}(\vec{x}) \,, \tag{2.56}$$

where $\vec{x}$ is the particle's position at some very early time. The first term describes the universal expansion, the second the peculiar motion. We assume that the *displacement field* $\vec{f}$ is irrotational,

$$\vec{f}(\vec{x}) = \vec{\nabla}\psi(\vec{x}) \,, \tag{2.57}$$

with some scalar potential $\psi(\vec{x})$.

- Since trajectories cannot get lost, the evolution of physical density is given by the Jacobian determinant of the mapping $\vec{x} \to \vec{r}$,

$$\rho = \rho_0 \det{}^{-1}\left[\frac{\partial r_i}{\partial x_j}\right] = \rho_0 \det{}^{-1}\left[a(t)\delta_{ij} + b(t)\frac{\partial f_i}{\partial x_j}\right] \,. \tag{2.58}$$

- Let $(\lambda_1, \lambda_2, \lambda_3)$ be the eigenvalues of the *deformation tensor* $f_{ij} := \partial f_i/\partial x_j = \partial^2\psi/\partial x_i \partial x_j$, then the density is

$$\rho = \frac{\rho_0}{(a + b\lambda_1)(a + b\lambda_2)(a + b\lambda_3)} \,, \tag{2.59}$$

where ρ_0 is the mean density at the present time. The mean density at later times is $\bar{\rho} = \rho_0 a^{-3}$, i.e. the density contrast is

$$\begin{aligned} \delta &= \frac{1}{(1 + b/a\lambda_1)(1 + b/a\lambda_2)(1 + b/a\lambda_3)} - 1 \\ &\approx -\frac{b}{a}(\lambda_1 + \lambda_2 + \lambda_3) = -\frac{b}{a}\vec{\nabla}\cdot\vec{f}. \end{aligned} \tag{2.60}$$

- The velocity perturbation has to satisfy the continuity equation $\vec{\nabla} \cdot \vec{u} = -\dot{\delta}$. Evaluating this equation in the appropriate approximation yields

$$-\dot{\delta} = \left(\frac{\dot{b}}{a} - \frac{\dot{a}b}{a^2}\right)\vec{\nabla} \cdot \vec{f} = H\left(\frac{\mathrm{d}b(a)}{\mathrm{d}a} - \frac{b}{a}\right)\vec{\nabla} \cdot \vec{f} \stackrel{!}{=} \vec{\nabla} \cdot \vec{u}, \quad (2.61)$$

thus

$$\vec{u} = H\left(\frac{\mathrm{d}b(a)}{\mathrm{d}a} - \frac{b}{a}\right)\vec{f}\,. \quad (2.62)$$

- From the growth of the linear density perturbations (2.30), we can immediately infer that

$$\frac{b}{a} = D_+(a)\,, \quad \delta_0 = -\vec{\nabla} \cdot \vec{f}\,, \quad (2.63)$$

thus

$$\frac{\mathrm{d}b}{\mathrm{d}a} = D_+ + a\frac{\mathrm{d}D_+}{\mathrm{d}a} = D_+[1 + f(\Omega_\mathrm{m})]\,, \quad (2.64)$$

and

$$\vec{u} = HD_+(a)f(\Omega_\mathrm{m})\vec{f} \quad (2.65)$$

i.e. the displacement field $\vec{f}$ is directly proportional to the velocity perturbation $\vec{u}$.

- Combining results, the particle trajectories according to the Zel'dovich approximation are

$$\vec{r} = a\left[\vec{x} + D_+(a)\vec{f}\right] = a\left[\vec{x} + \frac{\vec{u}}{Hf(\Omega_\mathrm{m})}\right]\,. \quad (2.66)$$

- An important result can be derived from the Zel'dovich approximation assuming that the density contrast, and thus the perturbation of the gravitational potential, are Gaussian random fields. The theory of multivariate Gaussians allows to derive the probability distribution $p(\lambda_1, \lambda_2, \lambda_3)$ for the eigenvalues of the deformation tensor F_{ij}. The result is

$$p(\lambda_1, \lambda_2, \lambda_3) = \frac{15^3}{8\pi\sqrt{5}\sigma^6}|(\lambda_3 - \lambda_2)(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)| \quad (2.67)$$
$$\times \quad \exp\left\{-\frac{3}{2\sigma^2}\left[2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2) - (\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)\right]\right\}\,,$$

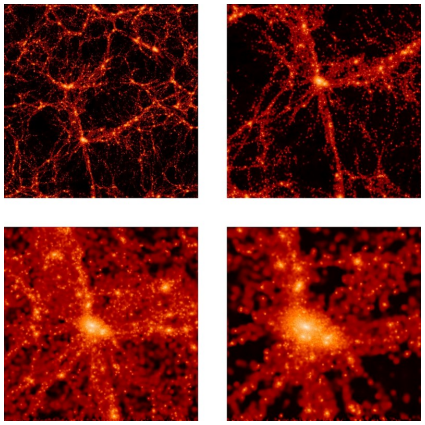
with σ^2 from (2.47). This result shows that the probability for two eigenvalues of F_{ij} to be equal is zero, implying that isotropic collapse is excluded. Forming structures will therefore be anisotropic, progressively flattening as the collapse proceeds. The resulting flattened mass distributions were called "pancakes" by Zel'dovich.

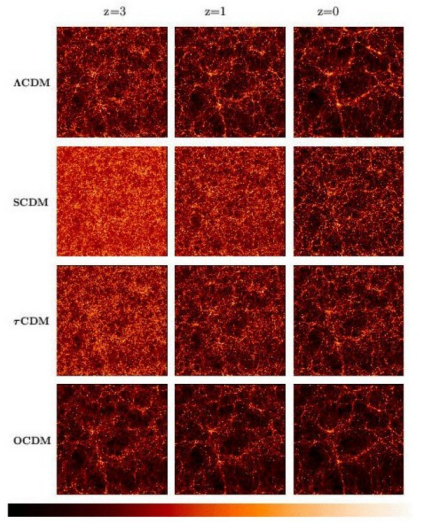
2.2.4 Nonlinear Evolution

- When the density contrast reaches unity, linear perturbation theory breaks down. The Zel'dovich approximation breaks down when trajectories cross because they just pass each other, ignoring their gravitational interaction.

- For a correct treatment, one has to resort to numerical simulations. They decompose the matter distribution into particles whose initial velocities are typically slightly perturbed according to some assumed power spectrum. The particles are then transported to redshifts high enough for linear evolution to hold on all scales considered. For later evolution, the equations of motion for all particles are solved. In the following, the most popular numerical techniques shall be presented:

  1. Ideally, particles move under the influence of the gravity from all other particles, but direct summation of all the gravitational forces of $N-1$ particles on N particles becomes prohibitively time-consuming and the scheme attains numerical complexity of $O(N^2)$ for every timestep. Several approximation schemes are therefore being employed.
  2. The particle-mesh (PM) algorithm computes the gravitational potential of the particle distribution on a grid (mesh) by solving Poisson's equation in Fourier space, making use of fast-Fourier techniques, thereby reducing the numerical complexity to $O(N \log N)$. The gravitational forces are then given by the gradients of the potential at the particle positions. This technique has a spatial resolution limited by the size of the mesh cells which makes it impractical for many modern applications.
  3. The particle-particle particle-mesh (P^3M) algorithm improves the PM technique by adding corrections for nearby particles which are determined by direct summation. Here, the numerical complexity is also $O(N \log N)$ provided the number of particle-particle operations per timestep is kept constant. However, this is not the case for high-resolution "zoom" simulations of individually forming objects in cosmological environments.
  4. Tree codes bundle distant particles into groups whose gravitational force on a particle is approximated as if they were point masses, or masses whose spatial distribution has a few low-order multipoles only, e.g. the monopole corresponding to a point mass, plus a dipole corresponding to a linear deformation, and so on. Depending on the solid angle that is subtended by the group on the sky seen by the particle, the





The VIRGO Collaboration 1996

nonlinear structure evolution, simulated in different cosmologies (Virgo collaboration)

"tree" is opened into its branches and leaves, i.e., higher-order multipoles of this group are considered. Alternatively, the monopole of that group, which is centered on its center-of-mass, is subdivided into the monopole moments of subgroups of the parent group, which subtend a smaller angle on the sky. It turns out that the numerical complexity of this technique is also $O(N \log N)$. The particle tree is updated as the evolution proceeds.

- Non-linear evolution causes density-perturbation modes to couple: while modes of different wave lengths evolve independently during linear evolution, mode coupling in the non-linear evolution moves power from large to small scales as structures collapse. The effect on the power spectrum is that the amplitude on small scales is increased at the expense of intermediate scales. Large scales continue to evolve linearly and independently.

- Even if the original density perturbation field δ is Gaussian, it must develop non-Gaussianities during non-linear evolution. This is evident because $\delta \geq -1$ by definition, but can become arbitrarily large. An originally Gaussian distribution of δ thus becomes increasingly skewed as it develops a tail towards infinite δ.

- Typical behaviour seen in numerical simulations shows the formation of "pancakes" and filaments as predicted by the theory of Gaussian random fields. Gravitational fragmentation of filaments into individual lumps causes galaxy-sized dark matter overdensities to form, which are called haloes. In the ΛCDM universe, those merge into galaxy groups which gradually stream towards the higher-density regions and larger mass concentrations at the intersections of filaments—galaxy clusters. Those form at the sites of constructive interference of long waves in the primordial fluctuations and are enhanced through gravitational collapse. Ongoing gravitational pull on the surrounding regions causes galaxy- and group-sized haloes to continuously merge into clusters which sit atop the cosmic mass hierarchy of haloes and thereby present the largest gravitationally collapsed objects to date. Giant voids form as matter accumulates in the walls of the cosmic network. Equivalently, the formation of voids can be considered to result from destructive interference of waves in the primordial fluctuations.

2.3 Spherical Collapse

2.3.1 Collapse of a Homogeneous Overdense Sphere

- The distribution of the dark matter in the universe can be considered as composed of individual so-called haloes, approximately spherical overdense clouds of dark matter which can reach highly non-linear densities in their centres.

- An approximate understanding of the parameters of such haloes and their relation to the dark-matter density contrast can be obtained by studying the dynamics of a spherical, homogeneous overdensity, leading to the so-called spherical collapse model.[1] While considering a *spherical* collapse of matter formally contradicts Zel'dovich's collapse analysis, an exact analytical solution that results from such an analysis nevertheless provides useful insights into non-linear collapse of more realistic situations. In particular the analysis (1) relates time (or redshift) at which the object collapses to its initial (linear) overdensity and (2) it maps the collapse time (redshift) to the final density of dark matter haloes that formed by collapse.

- The measured temperature anisotropies in the cosmic microwave background imply $\delta \ll 1$ at recombination. Thus, non-linear collapse happens at $a \gg a_{\rm rec}$, i.e., in the matter- or vacuum-dominated eras. We make the following assumptions in our analysis.

  - We consider a spherical perturbation that has initially a uniform overdensity.
  - The fluid is assumed to have zero pressure and is collisionless (i.e., the analysis applies to dark matter and not baryons). Later stages of baryonic collapse are different from that of dark matter since baryons additionally feel the pressure force, which causes the development of shocks in converging flows. However, since baryons only contribute 16% of the total mass, they do not appreciably change the collapse of dark matter.
  - For simplicity, we set $\Omega = \Omega_{\rm m} = 1$, i.e., a flat matter-dominated universe. This can be generalized to cases with $\Omega_{\rm m0} \neq 1$ and $\Omega_\Lambda \neq 1$.

- We consider a sphere of mass M and proper radius R and assume that the universe outside the sphere remains spherically symmetric such that it exerts no gravitational force on the matter in the

[1] An alternative derivation of the spherical collapse that employs the equations of motions in dimensionless form is provided in Appendix A.1.

sphere. Since M = const., we have

$$\frac{\mathrm{d}^2R}{\mathrm{d}t^2} = -\frac{GM}{R^2} \,, \tag{2.68}$$

which can be integrated to yield

$$\frac{1}{2}\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 - \frac{GM}{R} = \phi \,. \tag{2.69}$$

- We consider the gravitationally bound case, for which the energy per units mass is $\phi < 0$ and which leads to collapse. Adopting $R = 0$ at $t = 0$, we can integrate this equation and obtain

$$t = \int_0^R \frac{\mathrm{d}r}{\sqrt{2\,(GM/r + \phi)}} = \frac{A}{\sqrt{2|\phi|}} \int_{\theta(0)}^{\theta(R)} \frac{\sin\theta \mathrm{d}\theta}{\sqrt{2/(1-\cos\theta) - 1}} \,, \tag{2.70}$$

where we suitably changed the integration variable, using the transformation $r = A(1 - \cos\theta)$, where $A = GM/(2|\phi|)$. Employing trigonometric identities, we obtain

$$t = \frac{A}{\sqrt{2|\phi|}} \int_{\theta(0)}^{\theta(R)} (1 - \cos\theta)\mathrm{d}\theta = \frac{A}{\sqrt{2|\phi|}}(\theta - \sin\theta) \,. \tag{2.71}$$

Thus, the spherical collapse problem has the following parametric solution, which describes a cycloid,

$$R = A(1 - \cos\theta) \,, \quad A = \frac{GM}{2|\phi|} \,, \tag{2.72}$$

$$t = B(\theta - \sin\theta) \,, \quad B = \frac{GM}{(2|\phi|)^{3/2}} \,. \tag{2.73}$$

- The solution is characterised by an initial expansion phase from $R = 0$ at $\theta = 0$. It reaches a maximum radius $R_{\rm ta} = A$ at $\theta_{\rm ta} = \pi$ at which it turns around and collapses back to $R = 0$ at $\theta_{\rm c} = 2\pi$. In principle, it re-expands for $\theta > 2\pi$ but in practice, other physical effects become important and complicate things. The corresponding times are $t_{\rm ta} = \pi B$ for the maximum (turn-around) radius and $t_{\rm c} = 2\pi B = 2t_{\rm ta}$ for collapse at $R = 0$.

2.3.2 Connection to Linear Perturbation Theory

- The mean density inside the sphere is (2.72)

$$\rho = \frac{M}{4\pi/3\,R^3} = \frac{3M}{4\pi A^3}\frac{1}{(1-\cos\theta)^3} \,, \tag{2.74}$$

while the mean density of the background universe with $\Omega_{\rm m0} = 1$ is

$$\bar{\rho} = \frac{3H^2}{8\pi G} = \frac{1}{6\pi G t^2} = \frac{1}{6\pi G B^2}\frac{1}{(\theta - \sin\theta)^2}\ , \tag{2.75}$$

with $H = 2/(3t)$. The overdensity of the sphere (which is generally non-linear) can be obtained by combining these equations to yield

$$1 + \delta = \frac{\rho}{\bar{\rho}} = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}\ . \tag{2.76}$$

- To make the connection to linear perturbation theory, we consider the behaviour of the collapse at small t, which corresponds to small θ. Performing a Taylor series expansion of $\cos\theta$ and $\sin\theta$, we obtain

$$1 + \delta = 1 + \frac{3}{20}\theta^2 + O(\theta^4)\ , \tag{2.77}$$

$$t = \frac{B}{6}\theta^3 + O(\theta^5)\ . \tag{2.78}$$

Solving for θ gives (using $t_{\rm ta} = \pi B$)

$$\theta = \left(\frac{6t}{B}\right)^{1/3} + \ldots = (6\pi)^{1/3}\left(\frac{t}{t_{\rm ta}}\right)^{1/3} + \ldots, \quad \text{for } t \ll t_{\rm ta}. \tag{2.79}$$

- Thus, $\theta \ll 1$ corresponds to $t \ll t_{\rm ta}$. Substituting (2.79) into (2.77) gives

$$\delta = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_{\rm ta}}\right)^{2/3} \ll 1, \quad \text{for } t \ll t_{\rm ta}. \tag{2.80}$$

This yields the scaling of the density contrast in the spherical collapse model, $\delta \propto t^{2/3} \propto a$ (since $t \propto a^{3/2}$ in the Einstein-de Sitter model), which is exactly the behaviour of the growing mode solution of linear perturbation theory. Note that the decaying mode solution is absent due to our choice of initial conditions at $t = 0$.

- A corollary emerges from (2.80) that if the sphere has a uniform initial overdensity ($\delta_{\rm i}$) at some early time ($t_{\rm i}$), then *all* interior spheres will have the same $t_{\rm ta}$ and hence the sphere remains uniform as it collapses!

- There is an important distinction between (1) the real overdensity and (2) the overdensity *extrapolated* according to linear theory,

$$\delta_{\rm lin} = \delta_{\rm i}\left(\frac{t}{t_{\rm i}}\right)^{2/3} = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_{\rm ta}}\right)^{2/3} \quad \text{for all } t. \tag{2.81}$$

The maximum expansion radius at turnaround ($t = t_{\rm ta}$) is

$$\delta_{\rm lin}(t_{\rm ta}) = \frac{3}{20}(6\pi)^{2/3} \approx 1.062 \tag{2.82}$$

while the real (non-linear) overdensity is according to (2.76)

$$1 + \delta(t_{\rm ta}) = \frac{9\pi^2}{16} \approx 5.55. \tag{2.83}$$

- At collapse ($t = t_{\rm c} = 2t_{\rm ta}$), we have

$$\delta_{\rm c} := \delta_{\rm lin}(t_{\rm c}) = \frac{3}{20}(12\pi)^{2/3} \approx 1.686. \tag{2.84}$$

In terms of the initial overdensity $\delta_{\rm i}$, collapse happens at time

$$t_{\rm c} = t_{\rm i}\left(\frac{\delta_{\rm c}}{\delta_{\rm i}}\right)^{3/2} \propto \delta_{\rm i}^{-3/2}, \tag{2.85}$$

$$1 + z_{\rm c} = (1 + z_{\rm i})\left(\frac{\delta_{\rm i}}{\delta_{\rm c}}\right) \propto \delta_{\rm i}\ , \tag{2.86}$$

since $t \propto a^{3/2} \propto (1 + z)^{-3/2}$. Thus, perturbations that are initially more overdense collapse earlier! Generally, $\delta_{\rm c} = \delta_{\rm c}(\Omega_{\rm m}, \Omega_\Lambda)$, but the dependence on $\Omega_{\rm m}$ and Ω_Λ is weak so our result applies quite generally although it was derived for the Einstein-de Sitter model.

2.3.3 Final Density of a Collapsed Halo

- According to the spherical top hat collapse model, a uniform sphere collapses to a point of infinite density and then re-expands. In a realistic situation, the sphere contains inhomogeneities that generate tangential random velocities in the dark matter during collapse. This leads to an equilibrium configuration where the dark matter velocity dispersion balances its gravity. This relaxation process is called virialisation.

- We assume that the *final* dark matter halo is in dynamical equilibrium and obeys the virial theorem

$$2K_{\rm f} + V_{\rm f} = 0\ , \tag{2.87}$$

where K denotes the total kinetic energy in random motions, V is the total gravitational binding energy, and we neglected the surface pressure term due to further infalling material. We have

$$K_{\rm f} = \frac{M}{2}\sigma_{\rm f}^2\ , \quad \text{and} \tag{2.88}$$

$$V_{\rm f} = -G\int_0^{R_{\rm f}} \frac{4\pi r^3 \rho}{3r}{\rm d}m_{\rm shell} = -\frac{3}{5}\frac{GM^2}{R_{\rm f}}\ , \tag{2.89}$$

where σ is the three-dimensional velocity dispersion, ${\rm d}m_{\rm shell} = 4\pi r^2 \rho {\rm d}r$ and we assumed a uniform sphere of radius $R_{\rm f}$. Hence

$$E_{\rm f} = K_{\rm f} + V_{\rm f} = \frac{1}{2}V_{\rm f} = -\frac{3}{10}\frac{GM^2}{R_{\rm f}}\ . \tag{2.90}$$

- At turn-around, the sphere is at rest, i.e., $K_{\rm ta} = 0$. The total energy at turn-around is

$$E_{\rm ta} = V_{\rm ta} = -\frac{3}{5}\frac{GM^2}{R_{\rm ta}} \ . \tag{2.91}$$

  Since dark matter is collisionless, the conservation of total energy during the collapse yields $E_{\rm f} = E_{\rm ta}$ and hence, $R_{\rm f} = R_{\rm ta}/2$.

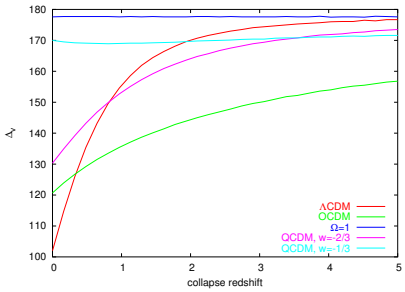
- The final density is thus $\rho_{\rm f} = 8\rho(t_{\rm ta})$. Assuming that virialisation happens at $t \approx t_{\rm c}$ and since $\bar{\rho} \propto t^{-2}$ and $t_{\rm c} = 2t_{\rm ta}$, the overdensity of the final halo is

$$1 + \delta_{\rm v} := 1 + \delta_{\rm c} = \frac{\rho_{\rm c}}{\bar{\rho}\,(t_{\rm c}/t_{\rm ta})^{-2}} = 32\,[1 + \delta(t_{\rm ta})] = 18\pi^2 = 178 \ . \tag{2.92}$$

  Hence, the final halo density is

$$\rho_{\rm f} = (1 + \delta_{\rm v})\bar{\rho}(t_{\rm c}) = 18\pi^2\bar{\rho}(t_{\rm c}) \ . \tag{2.93}$$

  $\delta_{\rm lin}$ and $\delta_{\rm v} := \Delta_{\rm v}$ are widely used in cosmology to characterise dark matter haloes. Other popular choices are $\Delta_{\rm v} = 100, 200, 500$, where each definition has its merits and shortcomings.



Virial overdensity in different cosmologies as a function of the halo collapse redshift.

- These two parameters derived from the spherical collapse model, $\delta_{\rm c}$ and $\Delta_{\rm v}$, are very widely used in cosmology for characterising dark-matter haloes and their formation. Extending these calculations into more general cosmological models is surprisingly difficult and requires numerical solutions of the underlying differential equations. Approximations to the solutions for $\Omega_{\rm m} < 1$ are

$$\delta_{\rm c} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \begin{cases} (1.0 + 0.0406\log_{10}\Omega_{\rm m}) & (\Omega_{\Lambda 0} = 0) \\ (1.0 + 0.0123\log_{10}\Omega_{\rm m}) & (\Omega_{\Lambda 0} = 1 - \Omega_{\rm m0}) \end{cases} \tag{2.94}$$

  and

$$\Delta_{\rm v} = 9\pi^2 \begin{cases} \left[1 + 0.1210(\Omega_{\rm m} - 1) + \Omega_{\rm m}^{0.6756}\right] & (\Omega_{\Lambda 0} = 0) \\ \left[1 + 0.7076(\Omega_{\rm m} - 1) + \Omega_{\rm m}^{0.4403}\right] & (\Omega_{\Lambda 0} = 1 - \Omega_{\rm m0}) \end{cases} \tag{2.95}$$

  where $\Omega_{\rm m}$ is the matter density parameter at the redshift of halo collapse.

2.3.4 The Press-Schechter Mass Function

- An important piece of information is the distribution of haloes over mass, the so-called mass function, which gives the number density of haloes at redshift z within the mass range between M and $M + {\rm d}M$.

- A characteristic length scale $R(M)$ can be assigned to a halo of mass M, which is defined as the radius of a homogeneous sphere filled with the mean cosmic matter density having mass M,

$$\frac{4\pi}{3}R^3\rho_{\rm cr}\Omega_{\rm m} = M \;\Rightarrow\; R(M) = \left(\frac{3M}{4\pi\rho_{\rm cr}\Omega_{\rm m}}\right)^{1/3} , \qquad (2.96)$$

  where $\Omega_{\rm m}$ and $\rho_{\rm cr}$ have to be evaluated at the redshift required.

- Aiming at haloes of mass M, we consider the density contrast field filtered on the scale $R(M)$. We therefore use $\bar{\delta}$ as defined in (2.48), i.e. the density contrast convolved with a window function W_R which has a characteristic scale $R = R(M)$.

- It will be convenient to scale halo masses with the so-called non-linear mass, which is the mass M_* for whose characteristic length scale $R(M_*) =: R_*$ the variance (2.49) of the density contrast becomes $\delta_{\rm c}^2$,

$$\sigma_{R_*}^2 = 4\pi \int_0^\infty \frac{k^2{\rm d}k}{(2\pi)^3} P(k)\hat{W}_{R_*}^2(k) = \delta_{\rm c}^2 \; . \qquad (2.97)$$

- For a Gaussian random field, the probability of finding at a given point $\vec{x}$ in space a filtered density contrast $\bar{\delta}(\vec{x})$ is

$$p(\bar{\delta}, a) = \frac{1}{\sqrt{2\pi\sigma_R^2(a)}} \exp\left[-\frac{\bar{\delta}^2(\vec{x})}{2\sigma_R^2(a)}\right] , \qquad (2.98)$$

  where we have explicitly noted that the variance σ will depend on time or equivalently on the scale factor a through the linear growth factor, $\sigma_R(a) = \sigma_R D_+(a)$.

- Press & Schechter suggested that the probability of finding the filtered density contrast at or above the linear density contrast for spherical collapse, $\bar{\delta} > \delta_{\rm c}$, is equal to the fraction of the cosmic volume filled with haloes of mass M,

$$F(M, a) = \int_{\delta_{\rm c}}^\infty {\rm d}\bar{\delta} p(\bar{\delta}, a) = \frac{1}{2}{\rm erfc}\left(\frac{\delta_{\rm c}}{\sqrt{2}\sigma_R(a)}\right) , \qquad (2.99)$$

  where ${\rm erfc}(x)$ is the complementary error function. Obviously, this equation implies that the fraction of cosmic volume filled with haloes of fixed mass M is a highly sensitive function of the scale factor a.

- The distribution of haloes over masses M is simply $\partial F(M)/\partial M$, so we have to relate σ_R to M, which is accomplished by the characteristic radius $R(M)$,

$$\frac{\partial}{\partial M} = \frac{{\rm d}\sigma_R(a)}{{\rm d}M}\frac{\partial}{\partial\sigma_R(a)} = \frac{{\rm d}\sigma_R}{{\rm d}M}\frac{\partial}{\partial\sigma_R} , \qquad (2.100)$$

where we have inserted the variance σ_R on the scale R at the present epoch. Using

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{erfc}(x) = -\frac{2}{\sqrt{\pi}}\mathrm{e}^{-x^2} , \tag{2.101}$$

we find

$$\left|\frac{\partial F(M)}{\partial M}\right| = \frac{1}{\sqrt{2\pi}}\frac{\delta_\mathrm{c}}{\sigma_R D_+(a)}\left|\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}\right|\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2 D_+^2(a)}\right) , \tag{2.102}$$

where the absolute values have been added to ensure positiveness of the Press-Schechter mass function.

- The normalisation of the mass function is wrong, however. It is easy to see that

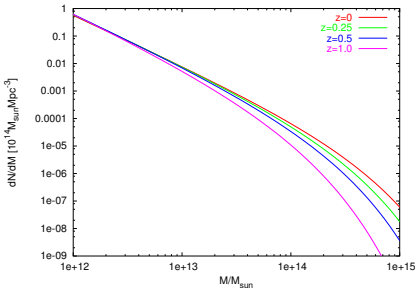
$$\int_0^\infty \left|\frac{\partial F(M)}{\partial M}\right| \mathrm{d}M = \frac{1}{2} \tag{2.103}$$

the reason for this problem is quite subtle, as we shall see later. for now, we will arbitrarily multiply the mass function by a factor of two.

- This fraction of the cosmic volume filled with haloes of masses within $[M, M + \mathrm{d}M]$ is converted to a (comoving) number density by dividing with the mean volume M/ρ_0 occupied by M

$$\begin{aligned} f(M,a)\mathrm{d}M &\equiv \frac{\partial n(M,a)}{\partial M}\mathrm{d}M \qquad (2.104)\\ &= \sqrt{\frac{2}{\pi}}\frac{\rho_0\delta_\mathrm{c}}{\sigma_R D_+(a)}\left|\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}\right|\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2 D_+^2(a)}\right)\frac{\mathrm{d}M}{M} . \end{aligned}$$

- The Press-Schechter mass function (2.104) has turned out to describe the mass distribution of dark-matter haloes in cosmological simulations remarkably well. Only recently have modifications been applied in order to improve its agreement with large, high-resolution simulations, or to take into account that halo collapse is not expected to proceed spherically, but elliptically.



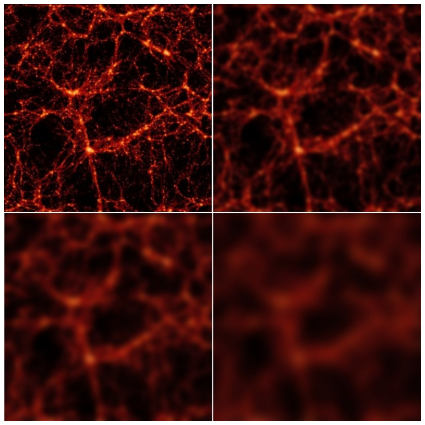
Press-Schechter mass function for the ΛCDM model at four different redshifts

2.4 Halo Formation as a Random Walk

2.4.1 Correct Normalisation of the Press-Schechter Mass Function

- The normalisation problem, however, is embarrassing and needs to be resolved. The solution was given with an elegant argument interpreting the statistics of halo formation in terms of a random walk.

- Suppose the density-contrast field δ is given. A large sphere is centred on some point $\vec{x}$ and its radius gradually shrunk. For each radius R of the sphere, the density contrast $\bar{\delta}$ averaged within R is measured and monitored as a function of R. By choosing a window function W_R in the definition (2.48) of $\bar{\delta}$ whose Fourier transform has a sharp cut-off in k space, $\bar{\delta}$ will undergo a random walk because decreasing R corresponds to adding shells in k space which are independent of the modes which are already included.

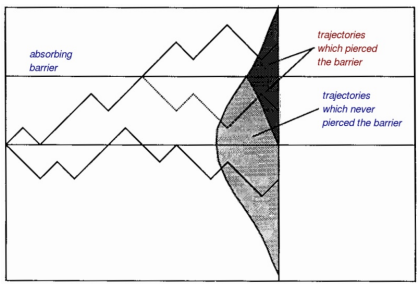


Progressive smoothing of the density field.

- $\bar{\delta}(\vec{x})$ is thus following a random trajectory. A halo is expected to be formed at $\vec{x}$ if $\bar{\delta}(\vec{x})$ reaches δ_c for some radius R. If $\bar{\delta}(\vec{x}) < \delta_c$ for some radius, it may well exceed δ_c for a smaller radius. Or, if $\bar{\delta}(\vec{x}) \geq \delta_c$ for some radius, it may well drop below δ_c for a smaller radius.

- For determining halo numbers correctly, it is thus necessary to count all points in space which are part of haloes of any mass. As R is shrunk around a point $\vec{x}$, that point must be counted as being part of a halo if there is a radius R for which $\bar{\delta}(\vec{x}) \geq \delta_c$.

- In the terminology of the random walk, we need to introduce an *absorbing barrier* at δ_c such that points $\vec{x}$ with trajectories $\bar{\delta}(\vec{x})$ vs. R which hit the barrier are removed from counting them as not being parts of haloes. To accomplish this, we follow the strategy of counting trajectories that do *not* make it into haloes such that the complement of that union represent trajectories of haloes.



Random walk with an absorbing barrier.

- A trajectory meeting the boundary has equal probability for moving above or below. For any *forbidden* trajectory continuing above the boundary, there is an *allowed* mirror trajectory continuing below it, and conversely. For any trajectory reaching a point $\bar{\delta} < \delta_c$ exclusively along *allowed* trajectories, there is a path reaching its mirror point on the line $\bar{\delta} = \delta_c$ exclusively along *forbidden* trajectories, and conversely. Thus, the probability for reaching a point $\bar{\delta} < \delta_c$ along *allowed* trajectories exclusively below the barrier is the probability for reaching it along *any* trajectory, minus the probability for reaching its mirror point

$\delta_c + (\delta_c - \bar{\delta}) = 2\delta_c - \bar{\delta}$ along *forbidden* trajectories,

$$p_s(\bar{\delta})d\bar{\delta} = \frac{1}{\sqrt{2\pi}\sigma_R}\left[\exp\left(-\frac{\bar{\delta}^2}{2\sigma_R^2}\right) - \exp\left(-\frac{(2\delta_c - \bar{\delta})^2}{2\sigma_R^2}\right)\right] , \quad (2.105)$$

where σ_R is the variance of $\bar{\delta}$ on the scale R, as before.

- (2.105) is the probability distribution for the averaged density contrast to fall within $[\bar{\delta}, \bar{\delta} + d\bar{\delta}]$ and *not* to exceed δ_c when averaged on *any* scale. The probability for $\bar{\delta}$ to *exceed* δ_c on some scale is thus

$$1 - P_s = 1 - \int_{-\infty}^{\delta_c} d\bar{\delta} p_s(\bar{\delta}) = \mathrm{erfc}\left(\frac{\delta_c}{\sqrt{2}\sigma_R}\right) , \quad (2.106)$$

without the factor 1/2 in (2.99). The rest of the derivation of the Press-Schechter mass function proceeds as before.

2.4.2 Extended Press-Schechter Theory

- Considering the random walk of the density contrast field when averaged over increasing or decreasing scales allows the statistics of haloes to be greatly extended. In order to simplify notation, we abbreviate $S := \sigma_R^2$.

- First, we note that we can either consider the barrier height δ_c to be constant while σ_R is increasing with time, or σ_R to be constant, while δ_c is decreasing with time, because only the ratio δ_c/σ_R enters the relevant expressions. Thus, the barrier can be considered moving towards zero as time progresses,

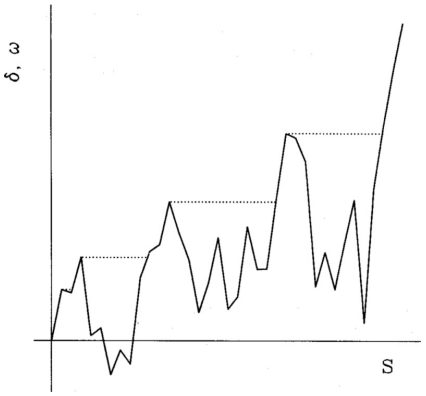
$$\omega := \frac{\delta_c}{D_+(a)} , \quad (2.107)$$

reflecting the fact that halo collapse becomes easier as structure formation proceeds. Since $\delta_c(a)$ decreases monotonically with increasing time, it can uniquely be used instead of time. The evolution of a halo can now be expressed as a random walk in S as time proceeds, or ω decreases.

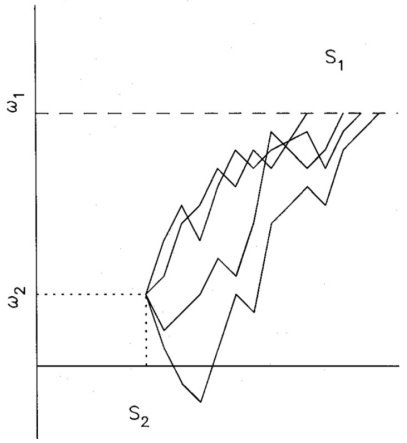
- Second, we note that

$$\begin{aligned} -\frac{\partial P_s}{\partial S}dS &= -\frac{\partial}{\partial S}\int_{-\infty}^{\delta_c} d\bar{\delta} p_s(\bar{\delta}) \\ &=: p_S(S,\omega)dS = \frac{\omega}{\sqrt{2\pi S^3}} e^{-\omega^2/2S} dS , \end{aligned} \quad (2.108)$$

is the probability for $\bar{\delta}$ to hit the barrier δ_c for the first time when the variance is increased from S to $S + dS$. It represents the fraction of mass in haloes of a mass M corresponding to the scale R.



Trajectory of a halo in the S-ω plane. Increasing S means decreasing mass, and ω decreases with time.



Trajectories of low-mass haloes at early time, forming a massive halo at a later time

- Consider now a trajectory passing through the barrier ω_2 for the first time at S_2, continuing to eventually pass through $\omega_1 > \omega_2$ at some $S_1 > S_2$. It represents a halo of mass M_1 corresponding to S_1 which, at a later time corresponding to ω_2, reaches mass $M_2 > M_1$ corresponding to S_2. The conditional probability for the halo to pass within $[S_1, S_1 + \mathrm{d}S_1]$ at ω_1, starting from S_2 at ω_2 is, according to (2.108),

$$p_{S_1}(S_1, \omega_1 | S_2, \omega_2)\mathrm{d}S_1 = \frac{\omega_1 - \omega_2}{\sqrt{2\pi}(S_1 - S_2)^{3/2}} \exp\left[-\frac{(\omega_1 - \omega_2)^2}{2(S_1 - S_2)}\right]\mathrm{d}S_1 \tag{2.109}$$

because the probability (2.108) only needs to be transformed shifting the origin of trajectories from $(S, \omega) = (0, 0)$ to $(S, \omega) = (S_2, \omega_2)$.

- From (2.109) and Bayes' theorem on conditional probabilities, we can straightforwardly derive the probability for a halo which for the first time reaches ω_1 at S_1 to reach ω_2 for the first time at S_2:

$$\begin{aligned}
& p_{S_2}(S_2, \omega_2 | S_1, \omega_1)\mathrm{d}S_2\, p_S(S_1, \omega_1)\mathrm{d}S_1 \\
= \;& p_{S_1}(S_1, \omega_1 | S_2, \omega_2)\mathrm{d}S_1\, p_S(S_2, \omega_2)\mathrm{d}S_2 \\
\Rightarrow \quad & p_{S_2}(S_2, \omega_2 | S_1, \omega_1)\mathrm{d}S_2 \\
= \;& \frac{p_{S_1}(S_1, \omega_1 | S_2, \omega_2)\mathrm{d}S_1\, p_S(S_2, \omega_2)\mathrm{d}S_2}{p_S(S_1, \omega_1)\mathrm{d}S_1} \\
= \;& \frac{1}{\sqrt{2\pi}}\left[\frac{S_1}{S_2(S_1 - S_2)}\right]^{3/2} \frac{\omega_2(\omega_1 - \omega_2)}{\omega_1} \\
\times \;& \exp\left[-\frac{(\omega_2 S_1 - \omega_1 S_2)^2}{2S_1 S_2(S_1 - S_2)}\right]\mathrm{d}S_2\,.
\end{aligned} \tag{2.110}$$

This provides the conditional probability for a halo of mass M_1 to have merged to form a halo of mass between M_2 and $M_2 + \mathrm{d}M_2$.

- The expected transition rate from S_1 to S_2 within the times t_1 and t_2 corresponding to ω_1 and ω_2 is determined by (2.110) taking the limit $\omega_2 \to \omega_1 =: \omega$,

$$\begin{aligned}
& \frac{\mathrm{d}^2 p_{S_2}}{\mathrm{d}S_2 \mathrm{d}\omega}(S_1 \to S_2 | \omega)\mathrm{d}S_2\mathrm{d}\omega \\
= \;& \frac{1}{\sqrt{2\pi}}\left[\frac{S_1}{S_2(S_1 - S_2)}\right]^{3/2} \exp\left[-\frac{\omega^2(S_1 - S_2)}{2S_1 S_2}\right]\mathrm{d}S_2\mathrm{d}\omega\,.
\end{aligned} \tag{2.111}$$

This gives the merger rate, i.e. the probability that, in the time interval corresponding to $\mathrm{d}\omega$, a halo of mass M_1 will merge with another halo of mass $M_2 - M_1$.

- We finally need to substitute the masses M_1 and M_2 for S_1 and S_2, and the time for ω. We wish to know the probability for a halo

of mass M to accrete another halo of mass ΔM within the time interval $\mathrm{d}t$ at time t. The transformation is

$$\frac{\mathrm{d}^2 p_M}{\mathrm{d}\ln\Delta M \mathrm{d}t}(M_1 \to M_2|t) = \frac{\mathrm{d}S_2}{\mathrm{d}\ln\Delta M}\left|\frac{\mathrm{d}\omega}{\mathrm{d}t}\right|\frac{\mathrm{d}^2 p_{S_2}}{\mathrm{d}S_2 \mathrm{d}\omega} \,. \tag{2.112}$$

- By the definition (2.107), the derivative of ω with respect to t is

$$\left|\frac{\mathrm{d}\omega}{\mathrm{d}t}\right| = \frac{\delta_\mathrm{c}}{D_+^2(a)} D'_+(a)\dot{a} = H\frac{\delta_\mathrm{c}}{D_+(a)}\frac{\mathrm{d}\ln D_+(a)}{\mathrm{d}\ln a} \,, \tag{2.113}$$

  where H is the Hubble parameter at scale factor a.

- Since $\Delta M = M_2 - M_1$, and S was introduced for σ_R^2, we have

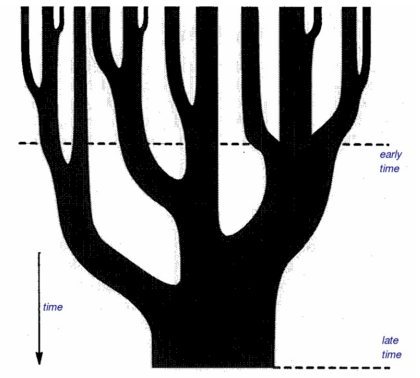
$$\frac{\mathrm{d}S_2}{\mathrm{d}\ln\Delta M} = \Delta M \frac{\mathrm{d}\sigma_R^2(M_2)}{\mathrm{d}M_2} \,. \tag{2.114}$$

- With expressions (2.113) and (2.114), the merger probability (2.112) becomes

$$\begin{aligned}\frac{\mathrm{d}^2 p_M}{\mathrm{d}\ln\Delta M \mathrm{d}t} &= \sqrt{\frac{2}{\pi}}\frac{H\delta_\mathrm{c}}{\sigma_{R2} D_+}\frac{\mathrm{d}\ln D_+}{\mathrm{d}\ln a}\Delta M \frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}(M+\Delta M) \\ &\times \left(1 - \frac{\sigma_{R2}^2}{\sigma_R^2}\right)^{-3/2} \\ &\times \exp\left[-\frac{\delta_\mathrm{c}^2}{2\sigma_{R2}^2 D_+^2}\left(1 - \frac{\sigma_{R2}^2}{\sigma_R^2}\right)\right] \,, \end{aligned} \tag{2.115}$$

  where $\sigma_{R2} := \sigma_R(M_2) = \sigma_R(M + \Delta M)$.

- In much the same way, the random-walk interpretation of halo growth allows deducing halo-survival times and other interesting quantities related to halo growth.



A "merger tree", i.e. a graphical representation of the accretion history of a halo

2.5 Halo Density Profiles

- Generally, a self-gravitating system of particles does not have an equilibrium state. The virial theorem demands that its total energy ($E = K + V$) is minus half its potential energy (V),

$$2K + V = E + K = 0 \quad \Rightarrow \quad K = -E = -\frac{V}{2}. \tag{2.116}$$

  Since $V < 0$ for self-gravitating systems, any inevitable energy loss, e.g., through the ejection of a body by means of three-body encounters, makes the potential energy become more negative. As a result, the halo becomes more tightly bound, which in turn increases its energy loss because the dynamical timescale is reduced by this contraction according to

$$t_{\rm dyn} \sim \left(\frac{R_{\rm g}^3}{GM}\right)^{1/2} \sim (G\rho)^{-1/2}\,, \tag{2.117}$$

  where $R_{\rm g} = GM/v^2$ is the gravitational radius. Thus, any halo density profile must reflect a potentially long-lived, but transient state.

- Knowing global halo properties like their mass, their distribution in mass and redshift, and their growth over time, their internal density profiles are an important characteristic. We will discuss two widely used models for the density profiles.

2.5.1 Isothermal Sphere

- A simple analytic model for the density profile is the isothermal sphere, which is a spherically-symmetric, self-gravitating system of non-interacting particles whose kinetic energy is characterised by a constant "temperature" $T = m/k\sigma^2$ where σ denotes the three-dimensional velocity dispersion.

- The equations describing the isothermal sphere are thus the Euler equation of hydrostatic equilibrium,

$$\frac{{\rm d}p}{{\rm d}r} = -\frac{GM(r)}{r^2}\rho\,, \tag{2.118}$$

  and the equation of state for the ideal gas

$$p = \frac{\rho}{m}kT\,, \tag{2.119}$$

  where m is the mean mass of the particles constituting the sphere.

- Inserting (2.119) into (2.118) yields

$$\frac{kT}{m}\frac{\mathrm{d}\ln\rho}{\mathrm{d}r} = -\frac{G}{r^2}\int_0^r 4\pi\rho(r')r'^2\mathrm{d}r' \ , \tag{2.120}$$

  where we have expressed the mass as an integral over the density. Differentiation with respect to r yields the second-order differential equation for ρ,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\ln\rho}{\mathrm{d}r}\right) = -\frac{4\pi Gm}{kT}r^2\rho \ . \tag{2.121}$$

- One solution of (2.121) is singular and can be obtained by means of a power-law ansatz in r to yield

$$\rho_1(r) = \frac{\sigma^2}{2\pi Gr^2} \qquad \sigma^2 := \frac{kT}{m} \ , \tag{2.122}$$

  where σ is the (radially constant) velocity dispersion of the particles. The mass and circular velocity of the *singular isothermal sphere (SIS)* are given by

$$M(<r) = \frac{2\sigma^2}{G}r \quad \text{and} \quad v_\mathrm{c}^2 = \frac{GM(<r)}{r} = 2\sigma^2 \ . \tag{2.123}$$

- The solution to (2.121) depends on the boundary conditions. It turns out that there is a second solution, which has a finite central density ρ_0. To find this solution, we have to identify a characteristic length scale such that we can obtain a general solution in terms of dimensionless variables. The dimensional variables in (2.121) are G, ρ, and the combination $\sigma^2 = kT/m$. Those can be combined to yield a length scale, $\sigma/\sqrt{G\rho}$, which represents the typical distance a particle travels in the central dynamical time. We define the *King radius* at which the density profile cores out,

$$r_0 := \left(\frac{9\sigma^2}{4\pi G\rho_0}\right)^{1/2} \ , \tag{2.124}$$

  and the dimensionless variables

$$\tilde{\rho} = \frac{\rho}{\rho_0} \ , \quad \text{and} \quad \tilde{r} = \frac{r}{r_0} \ . \tag{2.125}$$

- (2.121) cast into dimensionless form reads

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{r}}\left(\tilde{r}^2\frac{\mathrm{d}\ln\tilde{\rho}}{\mathrm{d}\tilde{r}}\right) = -9\tilde{r}^2\tilde{\rho} \ . \tag{2.126}$$

  The (numerical) solution is obtained by integrating this differential equation outwards from $\tilde{r} = 0$ with the central boundary conditions $\tilde{\rho}(0) = 1$ and $\mathrm{d}\tilde{\rho}/\mathrm{d}\tilde{r} = 0$ (the second condition is necessary

since $M(\tilde{r})$ vanishes at $\tilde{r} = 0$). The resulting second solution can be approximated by

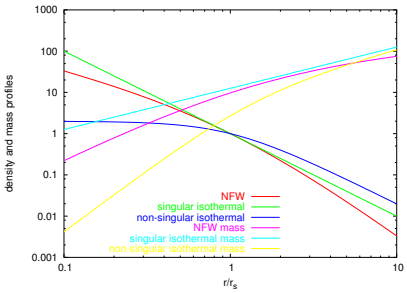
$$\tilde{\rho}_2(\tilde{r}) = \begin{cases} \left(1 + \tilde{r}^2\right)^{3/2} & \tilde{r} \lesssim 3, \\ \frac{2}{9}\tilde{r}^{-2} & \tilde{r} \gtrsim 3 \end{cases} , \tag{2.127}$$

i.e., the SIS is the asymptotic solution at large $\tilde{r}$. Note that by defining dimensionless variables, we have reduced the *family* of solutions with different densities and temperatures to a single solution for appropriately scaled variables.

- Both solutions have the advantage that they reproduce the flat rotation curves observed in spiral galaxies. The rotational velocity $v_{\rm rot}$ of a particle orbiting at radius r is determined by

$$v_{\rm rot}^2 = \frac{GM}{r} , \tag{2.128}$$

which is constant at $r \gg r_0$ for both density profiles of the isothermal sphere. However, the temperature within a stable "gas" sphere cannot be constant because particles would evaporate from it. Besides, the mass of the isothermal sphere diverges linearly as $r \to \infty$. To get a halo of finite mass, we must truncate it at some large radius by confining it with an external "pressure" that is practise is provided by accretion of mass. The isothermal profile is thus at best an approximation for the inner parts of haloes.



Singular and non-singular isothermal and NFW density and mass profiles.

2.5.2 Navarro-Frenk-White (NFW) Density Profile

- Numerical simulations of halo formation in the cold dark matter model consistently show density profiles like

$$\rho(r) = \frac{\rho_{\rm s}}{x(1+x)^2} , \quad x := \frac{r}{r_{\rm s}} , \tag{2.129}$$

which have a characteristic scale radius $r_{\rm s}$ beyond which they fall off $\propto r^{-3}$, and within which the density profile flattens considerably.

- Using the identity $x/(1+x)^2 \equiv (1+x)^{-1} - (1+x)^{-2}$, the mass of such haloes within radius r can easily be derived,

$$M(r) = 4\pi\rho_{\rm s} r_{\rm s}^3 \int_0^x \frac{x'\mathrm{d}x'}{(1+x')^2} = 4\pi\rho_{\rm s} r_{\rm s}^3 \left[\ln(1+x) - \frac{x}{1+x}\right] . \tag{2.130}$$

It rises $\propto x^2$ for small x and diverges logarithmically for $x \to \infty$. The divergence is not a fundamental problem because the halo profile must become invalid at the latest where ρ drops to the cosmic background density. In practise, the assumption of spherical symmetry starts to break down earlier, and becomes invalid at scales beyond the virial radius.

- The virial radius $r_{\rm vir}$ of a halo is often defined as the radius r_{200} enclosing a mean overdensity of 200 times the *critical* cosmic density, i.e., the contribution of Λ to the critical density is included in the estimate of the reference density. Since Ω_Λ is constant, this has the advantage that the mass of a halo that has decoupled from the cosmic expansion can only grow by accretion but remains constant otherwise (for $a > a_{\rm eq,\Lambda}$). The factor 200 is a rough approximation to the density contrast of $18\pi^2 \approx 178$ expected at virialisation in the spherical collapse model. This implies

$$M_{200}\left(\frac{4\pi}{3}r_{200}^3\right)^{-1} = 200\rho_{\rm cr}(a) = 200\frac{3H^2}{8\pi G}\ , \tag{2.131}$$

  where M_{200} is often identified with the total halo mass M. We obtain

$$r_{200} = \left(\frac{GM_{200}}{100H^2}\right)^{1/3}\ . \tag{2.132}$$

- Other frequent definitions use the radius enclosing a mean overdensity of 200 times the *mean* matter density (i.e., without the Λ contribution),

$$M_{200\rm m}\left(\frac{4\pi}{3}r_{200\rm m}^3\right)^{-1} = 200\rho_{\rm cr}(a)\Omega_{\rm m}(a)\ . \tag{2.133}$$

  This definition has the advantage that haloes of the same mass but at different redshifts show the same amount of kinetic pressure contribution or velocity anisotropy as a function of radius, i.e., this definition is close to a dynamical definition of the virial radius. However, it requires the knowledge of the ab initio unknown cosmological parameter $\Omega_{\rm m0}$ and has the property that the halo mass increases at late times because of the redshift dilution of the mean matter density as $\rho_{\rm m} = \rho_{\rm m0}a^{-3}$ even in the absence of mass accretion! Sometimes, people prefer a redshift dependent overdensity $\Delta(a)$ from the solution of a spherical top-hat perturbation *at* the time of collapse (2.95) rather than a constant overdensity threshold. While this property is easily calculable in simulations, the collapse time of a cluster is inaccessible in observations which jeopardises detailed comparisons of observations and theory.

- The ratio $c := r_{200}/r_{\rm s}$ is called *concentration* of the halo . It turns out to be a function of halo mass and redshift and to depend on cosmological parameters. Generally, c is the higher the earlier haloes form. This reflects the hierarchical growth of haloes and implies that smaller haloes form earlier when the mean background density was higher. As a result, these haloes have a higher density at small scales in comparison to larger halos when radii are scaled to R_{200}. Given the halo mass M, the (virial) radius is given by (2.132), the concentration parameter gives $r_{\rm s} = r_{200}/c$,

and the scale density ρ_s is then determined from (2.130) by the requirement that $M(r_{200}) = M_{200}$. Thus, the profile (2.129) is essentially determined by a single parameter, e.g. its mass.

- It is currently unclear how the density profile arises. Also, its slope near the core is being discussed.

Chapter 3

The Early Universe

3.1 Structures in the Cosmic Microwave Background

3.1.1 Simplified Theory of CMB Temperature Fluctuations

The Dipole

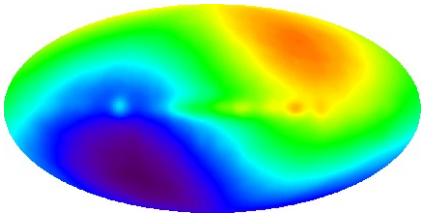
- We saw earlier that the universe is filled with a radiation background which has an ideal Planck spectrum with a temperature of 2.726 K. This cosmic microwave background is spectacularly isotropic, i.e. its temperature is almost the same everywhere on the sky.

- The Earth is not at rest with respect to the microwave background. Its motion around the Sun, combined with the Sun's motion around the centre of the Milky Way, combined with the Milky Way's motion within the Local Group, combined with the motion of the Local Group towards the Virgo cluster, causes an effective net motion with velocity v with respect to the CMB.

- As can be shown by a Lorentz transformation from the CMB rest frame to the rest frame of the Earth, this motion causes a dipolar pattern in the CMB temperature,

$$T(\theta) = T_0 \left(1 + \frac{v}{c} \cos\theta\right) + O\left(\frac{v^2}{c^2}\right) \tag{3.1}$$

  where T_0 is the mean CMB temperature and θ is the angle between the line-of-sight and the direction of motion. The CMB temperature is slightly enhanced towards the direction of motion, and decreased in its antidirection, corresponding to the Doppler shift.



CMB dipole as measured by COBE

- The COBE satellite determined the velocity of the Earth with respect to the CMB to be

$$v = (371 \pm 1)\,\mathrm{km\,s^{-1}}\ , \tag{3.2}$$

  pointing towards the Galactic coordinates

$$l = (264.3 \pm 0.2)^\circ\ , \quad b = (48.1 \pm 0.1)^\circ\ , \tag{3.3}$$

  the amplitude of the dipole is thus of order 10^{-3} K.

Expectations from Structure Growth

- Structures exist in the universe with a density contrast well above unity which, at the time when the CMB decoupled, must have had a density contrast of

$$\delta(a_{\rm CMB}) = \frac{\delta(a=1)}{D_+(a_{\rm CMB})} \gtrsim a_{\rm CMB}^{-1} \approx 10^{-3} \ . \tag{3.4}$$

  If the CMB energy density u were of equal magnitude, temperature fluctuations in the CMB should be of order 10^{-3} K, because

$$u \propto T^4 \ \Rightarrow \ \frac{\delta u}{u} = 4\frac{\delta T}{T} \tag{3.5}$$

  i.e. of the same order as the CMB dipole.

- After the detection of the CMB in 1965, temperature fluctuations were sought at this level, but not found. It was realised later that the problem can be solved if dark matter does not electromagnetically interact, because then structures can form in the dark matter much before decoupling without leaving a direct imprint on the CMB temperature fluctuations. This is the strongest argument that dark matter should not interact electromagnetically, and probably only through the weak interaction.

- Based on the assumption of weakly interacting dark matter, the expected temperature fluctuations in the CMB are expected to be of order $\delta T/T \approx 10^{-5}$, i.e. in the regime of micro-Kelvins. They were finally detected at this level by COBE in 1992.

Perturbation Equations and the Sachs-Wolfe Effect

- Studying the origin of the CMB fluctuations in detail is a complicated process. One must begin with the collisional Boltzmann equation for the photons and account for relativistic effects on the photon propagation like curvature and time delay. However, the simplified treatment shown here illustrates the main physical effects.

- The number density, energy density and pressure of the CMB photons are

$$n \propto T^3 \ , \quad u \propto T^4 \ , \quad p = \frac{u}{3} \propto T^4 \ . \tag{3.6}$$

  Introducing the relative temperature fluctuation $\Theta := \delta T/T_0$, where T_0 is the mean CMB temperature, we have

$$\frac{\delta n}{n_0} = 3\Theta \ , \quad \frac{\delta u}{u_0} = 4\Theta = \frac{\delta p}{p_0} \ . \tag{3.7}$$

- Ignoring expansion terms and setting $a = 1$, the continuity and Euler equations for the slightly perturbed photon gas read

$$\dot{n} + n_0 \vec{\nabla} \cdot \vec{v} = 0 \ , \quad \dot{\vec{v}} = -c^2 \frac{\vec{\nabla} \delta p}{u_0 + p_0} - \vec{\nabla} \delta\Phi \ , \tag{3.8}$$

  where $\vec{v}$ is the streaming velocity of the perturbations. They follow from the divergence of the relativistic energy-momentum tensor.

- Using (3.7) and $u_0 + p_0 = 4/3\, u_0 = 4p_0$, these equations can be written in terms of the temperature fluctuation

$$\dot{\Theta} + \frac{1}{3} \vec{\nabla} \cdot \vec{v} = 0 \ , \quad \dot{\vec{v}} = -c^2 \vec{\nabla} \Theta - \vec{\nabla} \delta\Phi \ . \tag{3.9}$$

- Inserting the divergence of the Euler equation into the time derivative of the continuity equation yields

$$\ddot{\Theta} - \frac{c^2}{3} \vec{\nabla}^2 \Theta - \frac{1}{3} \vec{\nabla}^2 \delta\Phi = 0 \ . \tag{3.10}$$

  Transforming to Fourier space, this becomes

$$\ddot{\hat{\Theta}} + \frac{c^2 k^2}{3} \hat{\Theta} + \frac{k^2}{3} \delta\hat{\Phi} = 0 \ . \tag{3.11}$$

- We now need to add a relativistic effect by hand which would appear in the equations if we derived them fully relativistically. Perturbing the metric by the potential $\delta\Phi$ causes the time delay

$$\frac{\delta t}{t} = -\frac{\delta\Phi}{c^2} \ , \tag{3.12}$$

  which causes the photons to be redshifted such that

$$\frac{\delta T}{T_0} = \Theta = -\frac{\delta\Phi}{c^2} \ . \tag{3.13}$$

  Fluctuations in the potential thus produce temperature fluctuations, and we have to add a source term

$$\ddot{\hat{\Theta}} = -\frac{\delta \ddot{\hat{\Phi}}}{c^2} \tag{3.14}$$

  to (3.11), which then reads

$$\ddot{\hat{\Theta}} + \frac{c^2 k^2}{3} \hat{\Theta} + \frac{k^2}{3} \delta\hat{\Phi} + \frac{\delta \ddot{\hat{\Phi}}}{c^2} = 0 \ . \tag{3.15}$$

- Combining temperature and potential fluctuations to form an effective temperature fluctuation $\hat{\Theta}+\delta\hat{\Phi}/c^2 =: \hat{\theta}$, we obtain the oscillator equation for $\hat{\theta}$,

$$\ddot{\hat{\theta}} + \frac{c^2k^2}{3}\hat{\theta} = 0 \,. \tag{3.16}$$

  Obviously, the solutions are trigonometric functions. If $\dot{\hat{\theta}} = 0$ at $t = 0$, the solution at the time of recombination is

$$\hat{\theta}(t_{\rm rec}) = \hat{\theta}(0)\cos\left[\frac{ck}{\sqrt{3}}t_{\rm rec}\right] \,. \tag{3.17}$$

  $c/\sqrt{3}\,t_{\rm rec} =: r_{\rm s}$ is called the sound horizon.

- The time delay (3.12) causes another temperature shift on the photons escaping from the last-scattering surface. Because of the Hubble expansion, the time delay causes a fluctuation in the scale factor at which the photons escape,

$$\frac{\delta T}{T_0} = \theta = -\frac{\delta a}{a} = -\frac{\dot{a}\delta t}{a} \,, \tag{3.18}$$

  because $T \propto a^{-1}$. In the matter-dominated era in the early universe, $a \propto t^{2/3}$, thus

$$\frac{\dot{a}}{a} = \frac{2}{3t} \;\Rightarrow\; \theta = \frac{2}{3}\frac{\delta\Phi}{c^2} \,, \tag{3.19}$$

  such that the temperature fluctuation $\hat{\Theta}$ becomes

$$\hat{\Theta} = \hat{\theta} - \frac{\delta\hat{\Phi}}{c^2} = -\frac{1}{3}\frac{\delta\hat{\Phi}}{c^2} \,. \tag{3.20}$$

  This is the Sachs-Wolfe effect.

Effects of Baryons

- Baryons couple to the photons through Compton scattering. Since the mean photon energy is of order 0.3 eV at the time of CMB decoupling, which is very small compared to the rest-mass energy of the electrons in the cosmic plasma, the limit of Thomson scattering is sufficient.

- In presence of baryons, Euler's equation must be corrected by multiplying the velocity and the potential gradient with the factor $(1 + R)$, where R is the ratio between the momentum densities of baryons and photons,

$$R := \frac{\rho_{\rm B}c^2 + p_{\rm B}}{u_0 + p_0} \approx \frac{3}{4}\frac{\Omega_{\rm B0}}{\Omega_{\rm r0}}a \,. \tag{3.21}$$

- Replacing $\vec{v} \to (1+R)\vec{v}$ and $\vec{\nabla}\delta\Phi \to (1+R)\vec{\nabla}\delta\Phi$ transforms (3.15) to

$$\ddot{\hat{\Theta}} + \frac{\dot{R}\dot{\hat{\Theta}}}{1+R} + \frac{c^2k^2}{3(1+R)}\hat{\Theta} = -\frac{k^2}{3}\delta\hat{\Phi} - \frac{\delta\ddot{\hat{\Phi}}}{c^2} \ . \tag{3.22}$$

  Thus the sound speed $c/\sqrt{3}$ is reduced by the baryons to $c/\sqrt{3(1+R)}$.

- Equation (3.22) describes sound waves in the temperature fluctuations which are driven by the gravitational potential fluctuation $\delta\Phi$ and its time derivatives, and damped by the expansion of the universe. On scales larger than the sound horizon,

$$\frac{2\pi}{k} < \frac{ct_{\rm rec}}{\sqrt{3(1+R)}} \ , \tag{3.23}$$

  these acoustic oscillations are suppressed.

Damping

- Further damping occurs due to imperfect coupling between the photons and the baryons. The photons exert a random walk and can thus diffuse across the length scale

$$\left\langle \lambda^2 \right\rangle = \left\langle \sum_{i=1}^{N} \lambda_i \sum_{j=1}^{N} \lambda_j \right\rangle = \left\langle \sum_{i=1}^{N} \lambda_i^2 \right\rangle + 2\left\langle \sum_{i>j} \lambda_i\lambda_j \right\rangle = N\lambda^2 \ , \tag{3.24}$$

  where λ is the mean-free path of the photons

$$\lambda = \frac{1}{n_{\rm e}\sigma_{\rm T}} \ , \tag{3.25}$$

  with the Thomson cross section $\sigma_{\rm T}$. We assumed in the last step of (3.24) that the individual steps are uncorrelated. Thus,

$$\lambda_{\rm D} := \left(\left\langle \lambda^2 \right\rangle\right)^{1/2} = \sqrt{N}\lambda \ . \tag{3.26}$$

  The number of collisions per unit time is

$$\mathrm{d}N = n_{\rm e}\sigma_{\rm T}c\mathrm{d}t \tag{3.27}$$

  thus,

$$\lambda_{\rm D}^2 = \int_0^{t_{\rm rec}} \frac{c\mathrm{d}t}{n_{\rm e}\sigma_{\rm T}} \ . \tag{3.28}$$

- Structures smaller than the diffusion length are damped, hence damping sets in for wave numbers

$$k > k_{\rm D} = \frac{2\pi}{\lambda_{\rm D}} \ . \tag{3.29}$$

Polarisation

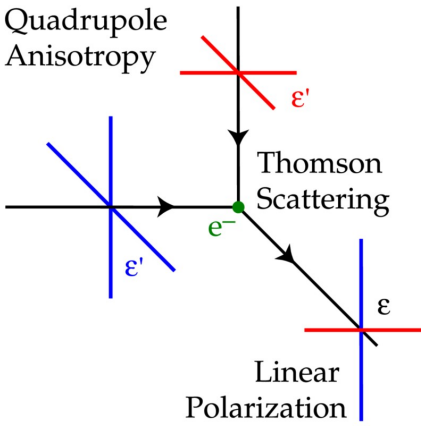
- Thomson scattering is anisotropic. Its differential cross section is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{3\sigma_\mathrm{T}}{8\pi} \left|\vec{e}' \cdot \vec{e}\right|^2 \tag{3.30}$$

  where $\vec{e}'$ and $\vec{e}$ are the unit vectors in the directions of the incoming and outgoing electric fields, respectively. Evidently, the scattered electric field with a field vector orthogonal to that of the incoming field has zero intensity.

- If the infalling radiation is isotropic, the scattered radiation is unpolarised. If, however, the infalling radiation has a quadrupolar intensity anisotropy, the scattered radiation is polarised because it has different intensities in its two orthogonal polarisation directions.

- Since the electrons within the last-scattering shell are irradiated by anisotropic light, the CMB is expected to be linearly polarised to some degree. The intensity of the polarised light should be of order 10% that of the unpolarised light, i.e. it should have an amplitude of order 10^{-6} K.



Origin of the CMB polarisation

3.1.2 CMB Power Spectra and Cosmological Parameters

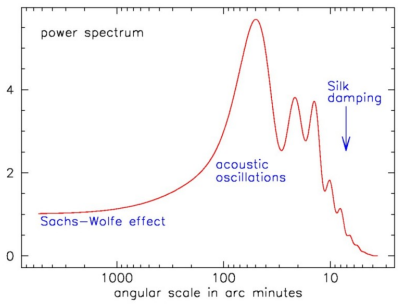
- Three effects were identified before which determine temperature fluctuations in the CMB: the Sachs-Wolfe effect on large scales, acoustic oscillations on scales smaller than the sound horizon, and damping on small scales due to photon diffusion.

- The visible temperature fluctuations on the sky are determined by the projection on the sky of photon density fluctuations in three-dimensional space. Due to that procedure, fluctuations of a single wave number k are smeared out over a range of angular scales.

- Fourier decomposition is not defined on the sphere. Instead, one has to project the temperature fluctuations onto another set of basis functions which are orthonormal on the sky. These are the spherical harmonic functions $Y_\ell^m(\vec{\theta})$. If $T(\vec{\theta})$ is the temperature at position $\vec{\theta}$ on the sky, it can be expanded into a series

$$T(\vec{\theta}) = \sum_{\ell m} a_{\ell m} Y_\ell^m(\vec{\theta}) \,, \tag{3.31}$$

  with the (generally complex) coefficients $a_{\ell m}$.



Appearance of the three most important CMB effects in the power spectrum.

- Because of the orthonormality of the spherical harmonics,

$$\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \sin\theta \mathrm{d}\theta Y_{\ell_1}^{m_1 *}(\theta,\varphi)\, Y_{\ell_2}^{m_2 *}(\theta,\varphi) = \delta_{\ell_1\ell_2}\delta_{m_1 m_2} \ , \qquad (3.32)$$

  the expansion coefficients are given by

$$a_{\ell m} = \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \sin\theta \mathrm{d}\theta T(\theta,\phi) Y_\ell^m(\theta,\phi) \ . \qquad (3.33)$$

Launch of the Boomerang experiment.

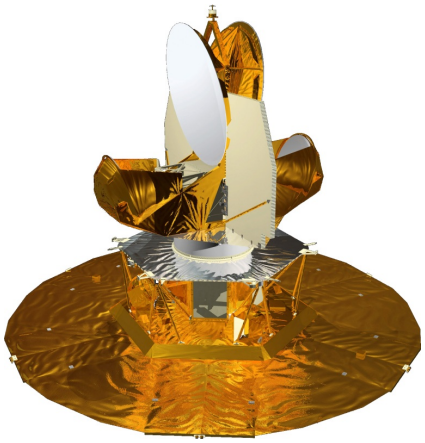
- The power spectrum of the temperature map is defined by

$$C_\ell = \left\langle |a_{\ell m}|^2 \right\rangle \qquad (3.34)$$

  which depends only on the multipole order ℓ because of statistical isotropy. Conventionally, the quantity $\ell(\ell+1)C_\ell$ is shown instead of C_ℓ because it reflects the total power contained in the multipole ℓ.

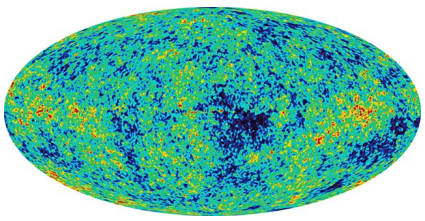
- The shape of $\ell(\ell+1)C_\ell$ is characteristic. As expected, the Sachs-Wolfe effect dominates on large scales, i.e. small ℓ, acoustic oscillations set in on scales smaller than the projection of the sound horizon on the sky, and very small scales are damped.

- The many pronounced features of the CMB power spectrum, and their tight relation to the cosmological parameters, allow cosmological parameters to be determined very accurately if the C_ℓ can be measured with high precision. This has caused substantial efforts to be put into the CMB measurements, with remarkable success.



The WMAP satellite.

- After relatively noisy measurements of the CMB on small fractions of the sky with balloon-borne experiments like Boomerang or Maxima, or ground-based experiments like Dasi, VSA or CBI, the NASA satellite "Wilkinson Microwave Anisotropy Probe" (WMAP) has obtained accurate full-sky maps of the microwave sky with an angular resolution of $\gtrsim 15'$ at frequencies between 23 and 94 GHz, and is continuing to measure. It has so far produced a CMB power spectrum which covers the first two acoustic peaks with high accuracy.



Full-sky CMB map produced by the WMAP satellite.

- Although the WMAP results alone suffer from degeneracies between different cosmological parameters, their combination with results from other cosmological experiments (in particular measurements of supernovae of type Ia, galaxy correlation functions, and structures in the distribution of neutral hydrogen) has produced the most accurate set of cosmological parameters to date:

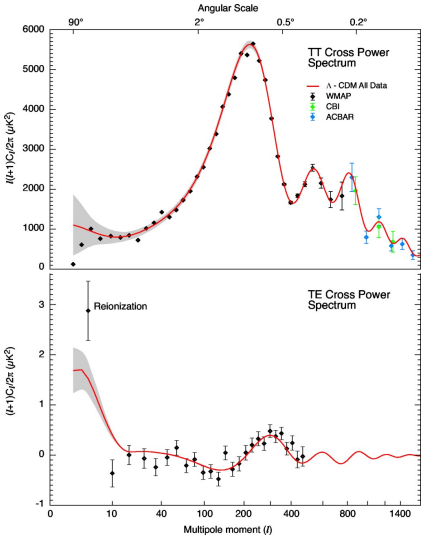
CMB temperature	T_{CMB}	$2.728 \pm 0.004\,\mathrm{K}$
total density	Ω_{tot}	$1.01^{+0.02}_{-0.01}$
matter density	Ω_{m}	$0.25^{+0.01}_{-0.03}$
baryon density	Ω_{b}	$0.045^{+0.001}_{-0.002}$
Hubble constant	h	0.73 ± 0.03
baryon-to-photon ratio	η	$6.1^{+0.3}_{-0.2} \times 10^{-10}$
fluctuation amplitude	σ_8	$0.74^{+0.05}_{-0.06}$
scalar spectral index	n_{s}	$0.951^{+0.015}_{-0.019}$
decoupling redshift	z_{dec}	1089 ± 1
age of the Universe	t_0	$13.7 \pm 0.2\,\mathrm{Gyr}$
age at decoupling	t_{dec}	$379^{+8}_{-7}\,\mathrm{kyr}$
reionisation redshift (95% c.l.)	z_{r}	$10.9^{+2.7}_{-2.3}$
reionisation optical depth	τ	0.09 ± 0.03

Most of these parameters should remain as further CMB data come in and are being analysed, but the error bars should continue to shrink. The most insecure numbers in this table are probably the redshift of reionisation and optical depth.

- The power spectrum of the polarised radiation shows similarly pronounced features as that of the temperature. Also, the structures in the polarisation map are expected to be correlated with those in the temperature map, i.e. there is a non-vanishing cross-power spectrum between temperature and polarisation.

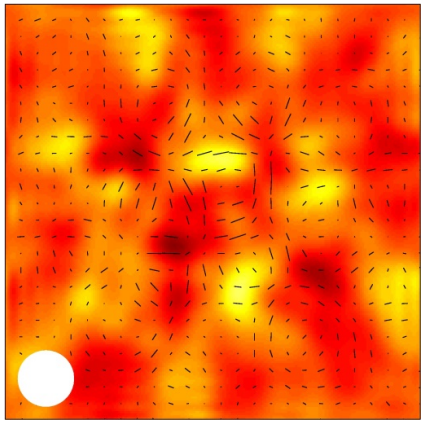
- Polarisation was first detected in the CMB by the DASI experiment located at the Amundsen-Scott station at the South Pole. Its amplitude, power spectrum and and cross-power spectrum with the temperature agree very well with expectations from theory. The WMAP satellite has measured the cross-power spectrum between temperature and polarisation only, which agrees very well with the theoretical expectations derived from the temperature power spectrum.

- The European satellite *Planck* will obtain full-sky maps of the CMB temperature and polarisation with an angular resolution of $\gtrsim 5'$ at frequencies between 30 and 857 GHz, further substantially improving upon the results from WMAP.



CMB spectrum derived from the WMAP results.

The DASI interferometer at the Amundsen-Scott station at the South Pole



Temperature and polarisation map produced by DASI.

3.1.3 Foregrounds

- Originating at redshift $z \approx 1100$, the CMB shines through the entire visible universe on its way to us. It is thus hidden behind a sequence of foreground layers.

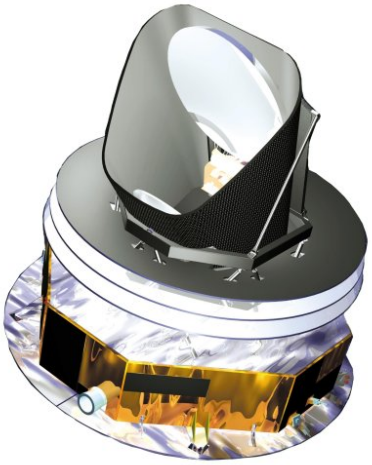
- The most important ones of those are caused by the microwave emission from our own Galaxy. Warm dust in the plane of

the Milky Way with a temperature near 20 K produces emission mainly above the CMB peak frequency. Electrons gyrating in the Galactic magnetic field emit synchrotron radiation which has a power law falling from radio frequencies into the microwave regime. Thermal free-free emission (*bremsstrahlung*) from ionised hydrogen partially falls into the microwave regime. Further sources include, e.g. the line emission from molecules like CO.

- Hot plasma in galaxy clusters inverse-Compton scatters microwave background photons to higher energies, giving rise to the so-called Sunyaev-Zel'dovich effect in the microwave regime. The characteristic spectral behaviour of that effect will enable future CMB missions to detect of order 10^4 galaxy clusters out to high redshifts.

- Other types of point source appearing in the microwave background include high-redshift galaxies, and planets, asteroids, and possibly comets in the Solar System. Also, dust in the plane of the Solar System emits the so-called Zodiacal light, which adds faint microwave emission.

- While these microwave foregrounds need to be carefully subtracted from the microwave sky to arrive at the CMB, they themselves provide important data sets for cosmology, but also for research on the Galaxy and possibly also the Solar System.



The European *Planck* satellite was launched on May 14, 2009.

3.2 Cosmological Inflation

3.2.1 Problems

Planck Scales

- Big-Bang cosmology offers a very successful, coherent picture for the evolution of the universe, but at the same time has fundamental problems.
- Evidently, the naïve picture of the Big Bang predicts the energy density to grow beyond all boundaries. Heuristically, we expect this approach to break down at the latest when quantum-mechanical effects set in. An estimate for when this may happen is given by the following argument:
- A quantum-mechanical length scale for a particle of mass m is its de Broglie wavelength,

$$\lambda_{\mathrm{dB}} = \frac{2\pi\hbar}{mc} , \tag{3.35}$$

while a gravitational length scale is given by its Schwarzschild radius,

$$r_{\mathrm{S}} = \frac{2Gm}{c^2} . \tag{3.36}$$

Quantum-mechanical effects are expected to become important in general relativity at the latest when the two become equal, which defines the Planck mass

$$m_{\mathrm{P}} = \sqrt{\frac{\hbar c}{G}} \approx 2 \times 10^{-5}\,\mathrm{g} \approx 10^{19}\,\frac{\mathrm{GeV}}{c^2} . \tag{3.37}$$

- Through (3.35), the Planck mass defines a length scale, the Planck length

$$l_{\mathrm{P}} = \frac{\hbar}{m_{\mathrm{P}} c} = \sqrt{\frac{\hbar G}{c^3}} \approx 10^{-33}\,\mathrm{cm} , \tag{3.38}$$

and a time scale, the Planck time

$$t_{\mathrm{P}} = \frac{l_{\mathrm{P}}}{c} = \sqrt{\frac{\hbar G}{c^5}} \approx 10^{-43}\,\mathrm{s} . \tag{3.39}$$

At times closer to the Big Bang than the Planck time, the purely general-relativistic treatment of cosmology is expected to break down.

The Horizon and Flatness Problems

- We have seen earlier that the particle horizon is given by

$$\Delta w(t_1, t_2) = \frac{c}{H_0\sqrt{\Omega_0}} \int_{a(t_1)}^{a(t_2)} \frac{\mathrm{d}a}{a^{2-n/2}} \,, \tag{3.40}$$

  in the early universe, i.e. before curvature and cosmological-constant terms became relevant.

- At recombination, the universe is well in the matter-dominated epoch, so we can set $n = 3$. Inserting further $a(t_1) = 0$ and $a(t_2) = a_{\mathrm{rec}}$ in (3.40) yields

$$\Delta w(0, t_{\mathrm{rec}}) = \frac{2c}{H_0}\sqrt{\frac{a_{\mathrm{rec}}}{\Omega_{\mathrm{m0}}}} \approx 175\ \sqrt{\Omega_{\mathrm{m0}}}^{-1}\, h^{-1}\,\mathrm{Mpc}\,. \tag{3.41}$$

  This is the *comoving* radius of a sphere around an given point in the recombination shell which could have causal contact with this point before recombination.

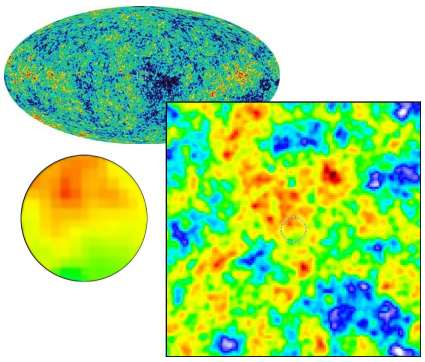
- The angular-diameter distance from us to the recombination shell for a flat universe without a cosmological constant ($\Omega_{\mathrm{m0}} = 1$ and $\Omega_\Lambda = 0$) is

$$D_{\mathrm{ang}}(0, z_{\mathrm{rec}}) \approx \frac{2c}{H_0} a_{\mathrm{rec}} \left(1 - \sqrt{a_{\mathrm{rec}}}\right) \approx \frac{2c}{H_0} a_{\mathrm{rec}} \approx 5\, h^{-1}\,\mathrm{Mpc}\,. \tag{3.42}$$

- The angular size of the particle horizon at recombination on the CMB sky is therefore

$$\theta_{\mathrm{rec}} = \frac{a_{\mathrm{rec}}\Delta w(0, a_{\mathrm{rec}})}{D_{\mathrm{ang}}(0, z_{\mathrm{rec}})} \approx \sqrt{a_{\mathrm{rec}}} \approx 1.7^\circ\,, \text{ for } \Omega_{\mathrm{m0}} = 1\,. \tag{3.43}$$

  Repeating this calculation for a critical, low-density universe with a cosmological constant ($\Omega_\Lambda = 1 - \Omega_{\mathrm{m0}}$) yields almost the same result.



Size of causally connected regions on the CMB.

- Given any point on the microwave sky, the causally connected region around it has a radius of approximately one degree, i.e. four times the radius of the full moon. How is it possible that the CMB temperature is so very closely the same all over the full sky? Points on the sky further apart than $\approx 2^\circ$ had no chance of causally interacting and "communicating" their temperature. This constitutes the *horizon problem*.

- Ignoring the cosmological-constant term, the Friedmann equation can be written

$$H^2(a) = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} = H^2(a)\left[\Omega_{\mathrm{total}}(a) - \frac{Kc^2}{a^2H^2}\right]\,. \tag{3.44}$$

Thus the deviation of $\Omega_{\rm total}$ from unity is

$$|\Omega_{\rm total} - 1| = \frac{Kc^2}{a^2H^2} \ . \tag{3.45}$$

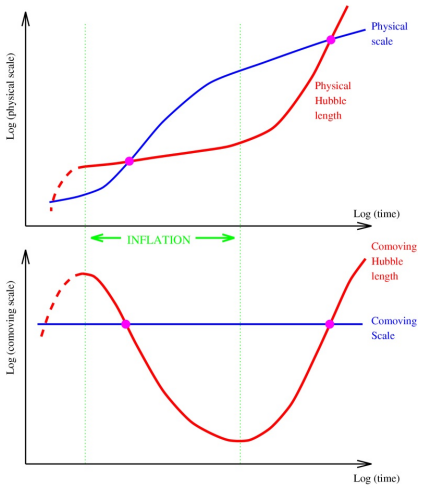
- We have already seen that $\Omega \to 1$ for $a \to 0$ during the matter-dominated era. During radiation-domination, $a^2H^2 = \dot{a}^2 \propto t^{-1}$, during the early matter-dominated era, $a^2H^2 \propto t^{-2/3}$, thus

$$|\Omega_{\rm total} - 1| \propto \begin{cases} t & \text{radiation-dominated era} \\ t^{2/3} & \text{early matter-dominated era.} \end{cases} \tag{3.46}$$

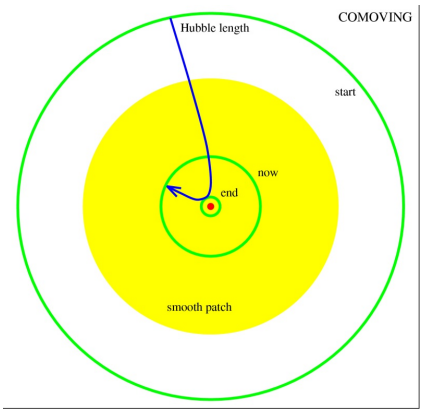
  Therefore, if there is any tiny deviation of $\Omega_{\rm total}$ from unity at early times, it moves rapidly away from unity. In order for $\Omega_{\rm total}$ to be anywhere near unity today, it must have been extremely close to unity at early times, which constitutes an uncomfortable fine-tuning problem, the *flatness problem*.

- The horizon problem is exacerbated by the observation that not only is the temperature of the CMB very nearly the same all over the sky, but also coherent structures exist in the CMB which are much larger than the horizon size at decoupling. How could these structures be formed?

- Apart from the problem of how structures can be coherent beyond the horizon scale, it remains as yet unexplained where structures originate from in the first place. Ultimately, cosmology needs to explain why there are structures rather than complete homogeneity.



Effect of a shrinking comoving Hubble radius.



Horizon and causally connected regions.

3.2.2 Inflation

The Idea of Inflation

- Returning to (3.45), we note that c/H is the Hubble radius, hence $c/(aH)$ is the *comoving* Hubble radius. At least the flatness problem could be solved if the comoving Hubble radius could *shrink* sufficiently for some time, because then the deviation of $\Omega_{\rm total}$ from unity would be driven towards zero.

- The physical picture behind a *shrinking* comoving Hubble radius is the following: the Hubble radius characterises the radius of the observable universe, thus the *comoving* Hubble radius gives the radius of the observable universe in comoving coordinates, i.e. after transforming to non-expanding coordinates. If the comoving Hubble radius could shrink during some time, the observable part of the universe could be moved within causally connected regions, thus the contents of the entire observable universe

could be brought into causal contact. After this phase ends, the observable universe can expand again, but its physical state can appear coherent everywhere thereafter.

Conditions for Inflation

- The condition for a shrinking, comoving Hubble radius is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{c}{aH}\right) < 0 \,. \tag{3.47}$$

  Since $aH = \dot{a}$, this implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{c}{\dot{a}}\right) = -\frac{c\ddot{a}}{\dot{a}^2} < 0 \;\Rightarrow\; \ddot{a} > 0 \,, \tag{3.48}$$

  i.e. it is equivalent to accelerated expansion.

- Accelerated expansion seems incompatible with gravity because the gravitational force exerted by the matter inside a representative spherical section of the universe is expected to decelerate its expansion.

- Friedmann's 2nd equation (without a cosmological constant) allows accelerated expansion if

$$\rho c^2 + 3p < 0 \,, \tag{3.49}$$

  i.e. expansion can accelerate if and only of the pressure is sufficiently negative,

$$p < -\frac{\rho c^2}{3} \,. \tag{3.50}$$

- Energy conservation requires

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\rho c^2 a^3\right) + p\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\right) = 0 \;\Rightarrow\; \dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \,. \tag{3.51}$$

  Since, by definition, the cosmological constant has $\dot{\rho} = 0$, it must correspond to a form of matter which has

$$p = -\rho c^2 \,, \tag{3.52}$$

  i.e. the cosmological constant provides a suitably exotic equation of state.

- Once the cosmological-constant term becomes appreciable in Friedmann's equation, it quickly dominates because it scales with the highest power of the scale factor a. As we have seen, it accelerates cosmic expansion, thus a grows rapidly, and the

cosmological-constant term very quickly entirely determines the dynamics. This is the case of de Sitter expansion mentioned earlier in the context of the late cosmic evolution,

$$a \propto \exp\left(\sqrt{\Omega_\Lambda} H_0 t\right) , \tag{3.53}$$

i.e. exponential expansion sets in once Λ starts dominating.

Inflation and Scalar Fields

- As an example for a simple physical system which may have negative pressure, consider a self-interacting scalar field ϕ, which has the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) , \tag{3.54}$$

where $V(\phi)$ is the interaction potential and Einstein's sum convention was assumed. We adopted the following convention for the scalar product, $\partial_\mu\phi\partial^\mu\phi = g_{\mu\nu}\partial^\mu\phi\partial^\nu\phi$ with the metric of Minkowski space given by $g_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$.

- The field ϕ has the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\mathcal{L} = \mathrm{diag}(\rho c^2, p, p, p) . \tag{3.55}$$

Here, we adopted the energy-momentum tensor of an ideal fluid that has no shear stress. Its time-time component is the energy density,

$$\rho c^2 = T_{00} = \frac{1}{2}\dot{\phi}^2 + V(\phi) + \frac{1}{2}(\vec{\nabla}\phi)^2 , \tag{3.56}$$

while the pressure is given by one third times the trace over its space-space components,

$$p = \frac{1}{3}\mathrm{trace}(T_{ii}) = \frac{1}{2}\dot{\phi}^2 - V(\phi) - \frac{1}{6}\left(\vec{\nabla}\phi\right)^2 . \tag{3.57}$$

- Due to homogeneity, the terms $\vec{\nabla}\phi$ must vanish. The requirement (3.50) then translates to

$$\frac{1}{2}\dot{\phi}^2 - V(\phi) < -\frac{1}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) , \tag{3.58}$$

which is satisfied if,

$$\dot{\phi}^2 < V(\phi) . \tag{3.59}$$

Thus the scalar field ϕ shows the desired behaviour provided its kinetic energy is sufficiently small compared to its potential energy, i.e. if it "moves" sufficiently slowly.

- Inserting the energy density of ϕ into Friedmann's equation yields

$$H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right] , \tag{3.60}$$

and the continuity equation (3.51) requires

$$\ddot{\phi} + 3H\dot{\phi} = -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} . \tag{3.61}$$

These equations determine the evolution of ϕ in the expanding cosmological background.

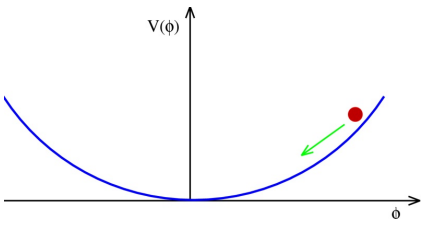
Slow-Roll Conditions

- Following the requirement (3.59), we impose the conditions

$$\dot{\phi}^2 \ll V , \quad \frac{\mathrm{d}}{\mathrm{d}t}\dot{\phi}^2 \ll \frac{\mathrm{d}V(\phi)}{\mathrm{d}t} \;\Rightarrow\; \ddot{\phi} \ll \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} , \tag{3.62}$$

for successful inflation, i.e. we want inflation to be strong and to persist sufficiently long. These conditions simplify the evolution equations to

$$H^2 \approx \frac{8\pi G}{3}V(\phi) , \quad 3H\dot{\phi} \approx -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} =: -V' . \tag{3.63}$$



Slowly rolling field in a flat potential.

- Consequently, the condition $\dot{\phi}^2 \ll V(\phi)$ can be written as

$$\dot{\phi}^2 = \left(\frac{V'}{3H}\right)^2 = \frac{(V')^2}{24\pi G V} \ll V \;\Rightarrow\; \frac{1}{24\pi G}\left(\frac{V'}{V}\right)^2 =: \epsilon \ll 1 . \tag{3.64}$$

Additionally,

$$\ddot{\phi} = -\frac{\mathrm{d}}{\mathrm{d}t}\frac{V'}{3H} = -\frac{V''\dot{\phi}}{3H} + \frac{V'\dot{H}}{3H^2} , \tag{3.65}$$

and, with

$$2H\dot{H} = \frac{8\pi G}{3}V'\dot{\phi} \;\Rightarrow\; \frac{\dot{H}}{H} = \frac{4\pi G}{3H^2}V'\dot{\phi} = \frac{\dot{\phi}}{2}\frac{V'}{V} , \tag{3.66}$$

we find

$$\ddot{\phi} = -\frac{V''\dot{\phi}}{3H} + \frac{(V')^2\dot{\phi}}{6VH} \ll V' = -3H\dot{\phi} , \tag{3.67}$$

and thus

$$\frac{V''}{3H^2} - \frac{(V')^2}{6VH^2} = \frac{1}{8\pi G}\frac{V''}{V} - \frac{3}{2}\epsilon =: \eta - \frac{3}{2}\epsilon \ll 3 . \tag{3.68}$$

- Thus, successful inflation is equivalent to the condition that the two *slow-roll parameters*

$$\epsilon := \frac{1}{24\pi G}\left(\frac{V'}{V}\right)^2 \ll 1 \ , \quad \eta := \frac{1}{8\pi G}\left(\frac{V''}{V}\right) \ll 1 \ , \tag{3.69}$$

are both much smaller than unity. Hence, we require the inflaton potential to exhibit a large gradient length L or equivalently a shallow potential slope (since $\sqrt{\epsilon} \sim V'/V \sim 1/L \ll 1$) as well as a large curvature radius R (since $\sqrt{\eta} \sim \sqrt{V''/V} \sim 1/R \ll 1$).

Amount and End of Inflation

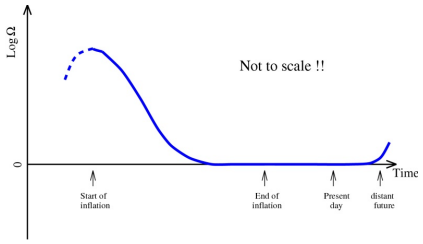
- Today's age of the universe is $t_0 \approx 4 \times 10^{17}$ s. The Planck time, which is a possible time for the onset of inflation, is $t_P \approx 10^{-43}$ s. During the radiation-dominated era,

$$|\Omega_{\text{total}} - 1| \propto t \ . \tag{3.70}$$

Thus, $\Omega_{\text{total}} \approx 1$ today can be achieved if

$$|\Omega_{\text{total}} - 1| \approx 10^{60} \tag{3.71}$$

at the onset of inflation.



Driving the universe spatially flat.

- Since $a \propto t^{1/2}$ during radiation domination, for inflation to solve the flatness problem, the comoving Hubble radius thus needs to shrink by a factor of $\approx 10^{30}$, which corresponds to an increase in the scale factor by a factor of approximately e^{60}. This would at the same time solve the horizon (or causality) problem.

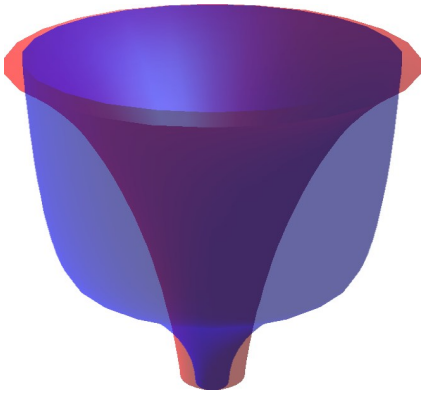
- During inflation, the energy density of the inflaton field is approximately constant since $\rho c^2 \approx V$, and the changes in V are small due to the slow-roll conditions.

- All other densities drop by huge amounts. Since $\rho \propto a^{-3}$ for non-relativistic matter and $\rho \propto a^{-4}$ for radiation, their densities decrease by factors of $\approx e^{-180}$ and $\approx e^{-240}$, respectively.

- Since there is matter and radiation in the universe today, there must be a way to convert the energy density of the inflaton field into the energy density of radiation or matter as inflation ends, i.e. when $(\epsilon, \eta) \approx 1$.

- At this time, the kinetic terms $\dot{\phi}$ and $\ddot{\phi}$ become important. The inflaton field may oscillate around the minimum of its potential energy.

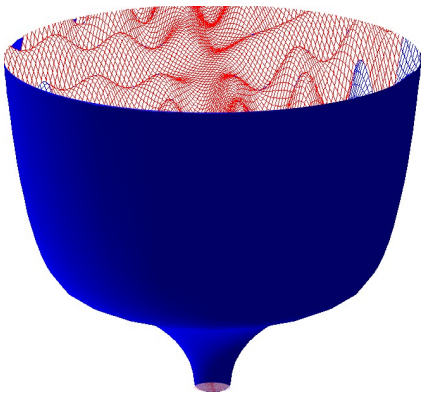
- It is assumed that the inflaton field can decay through some coupling to "ordinary" matter and thus turn its energy density back into other constituents of the cosmic fluid. In particular, this "reheating" process should produce a "hot state" in thermal equilibrium at the end of inflation in order not to spoil the successes of Big Bang nucleosynthesis. How this occurs in detail is an open question.



The universe expands beyond the horizon.

Inflation and Structure Formation

- As any other quantum field, the inflaton field must have undergone vacuum fluctuations before inflation because of the uncertainty principle.
- Once inflation sets in, the vacuum fluctuations are quickly driven outside of the horizon (or, in the language of the shrinking comoving horizon, the horizon quickly contracts below the length scale of the quantum fluctuation), where they "freeze in" because they lack causal contact.
- Assuming there exists a quantum field theoretical extension of gravitation, there are two light quantum fields that fluctuate during inflation: the "inflaton" field itself and the quantized gravitational field, which would be mediated by the hypothetical "graviton", a massless elementary particle of spin 2. These two different quantum fluctuations are inflated to macroscopic perturbations with distict properties. Since the inflaton energy density got eventually converted to radiation and matter by means of the "reheating" process, inflaton fluctuations produce fluctuations in the primordial density field (which are called "scalar" fluctuations because of the scalar nature of the inflaton fields). Quantum fluctuations in the graviton field are blown up to macroscopic gravitational waves during inflation and give rise to so-called "tensor" fluctuations (because of the tensor nature of the metric field that characterizes space-time). While an electromagnetic wave is an oscillation in the electric and magnetic fields that propagates at the speed of light, a gravitational wave is an oscillation in the gravitational field that also propagates at the speed of light. Most importantly, there exists a possibility of disentangling gravitational waves from density fluctuations, using the polarization of the CMB: while density fluctuations give only rise to primordial E mode polarization, gravitational waves produce B-mode as well as E-mode polarization on large angular scales and provide thus a unique signature to infer the existence of an inflationary epoch.



Initial quantum fluctuations are inflated to macroscopic scales.

- For a highly simplified treatment of the qualitative properties of

density fluctuations produced that way, consider a spherical overdensity. It must of course satisfy Friedmann's equation, which we write in the form (3.44),

$$H^2 = H^2\left(\Omega - \frac{Kc^2}{a^2H^2}\right) , \tag{3.72}$$

where Ω is the density parameter inside the overdensity, from which we obtain

$$\rho a^2 = \frac{3H^2a^2}{8\pi G}\Omega = \frac{3H^2a^2}{8\pi G} + \frac{3Kc^2}{8\pi G} = \frac{\rho a^2}{\Omega} + \frac{3Kc^2}{8\pi G} , \tag{3.73}$$

and thus

$$\rho a^2\left(1 - \frac{1}{\Omega}\right) = \text{const.} \tag{3.74}$$

- For a linear overdensity in the early universe, $\Omega = 1 + \delta\Omega$ with $\delta\Omega \ll 1$, thus $\delta\rho = \rho\delta\Omega \ll \rho$, and (3.74) implies

$$\rho a^2\left(1 - \frac{1}{\Omega}\right) \approx \rho a^2\delta\Omega \approx \delta\rho a^2 = \text{const.} \tag{3.75}$$

  i.e. the physical overdensity $\delta\rho$ inside the spherical perturbation must scale $\propto a^{-2}$.

- The fluctuation $\delta\Phi$ in the gravitational potential caused by the spherical overdensity is

$$\delta\Phi = \frac{G\delta M}{R} = \frac{4\pi G}{3}(aL)^3\frac{\delta\rho}{aL} = \text{const.}\, L^2 , \tag{3.76}$$

  where R is the physical radius of the sphere, and L is its comoving radius. The last equality follows because $\delta\rho \propto a^{-2}$. The potential fluctuation caused by the perturbation thus remains constant during inflation.

- The physical scale (aL) changes by $\approx$ 30 orders of magnitude during inflation, thus inflation predicts approximately identical potential fluctuations on all accessible physical scales.

- The detailed theory of the inflationary origin of structures starts with the vacuum expectation value of the inflaton field on the scale corresponding to wave number k,

$$\left\langle 0 \left| |\phi_k|^2 \right| 0 \right\rangle , \tag{3.77}$$

  and solves the equations for the field amplitudes. The result is that the root-mean-square fluctuations in the gravitational potential scale as follows,

$$\left\langle \delta\Phi^2 \right\rangle^{1/2} \propto \frac{H^2}{\dot{\phi}} , \tag{3.78}$$

  which is approximately constant because of the slow-roll conditions.

- Due to Poisson's equation, the Fourier modes of the potential and density fluctuations are related by $k^2\delta\hat{\Phi}(k) \propto -\hat{\delta}(k)$, thus the (primordial) density power spectrum predicted by inflation is

$$|\hat{\delta}(k)|^2 \propto k^4|\delta\hat{\Phi}(k)|^2 \propto k^3 P_{\rm i}(k) \;\Rightarrow\; P_{\rm i}(k) \propto k \; . \tag{3.79}$$

  This is the Harrison-Zel'dovich-Peebles spectrum which was originally required for completely different reasons. Precise calculations find

$$P_{\rm i}(k) \propto k^n \; , \tag{3.80}$$

  with $n \lesssim 1$.

- Since the density fluctuations arise from superpositions of enormous numbers of statistically independent vacuum fluctuations of the inflaton field, they are expected to be Gaussian because of the central limit theorem.

- Thus, inflation provides a physical picture for solving the horizon and flatness problems of the Big Bang theory, and at the same time provides a natural explanation for the origin of structures in the universe, which are predicted to be nearly scale-invariant and Gaussian.

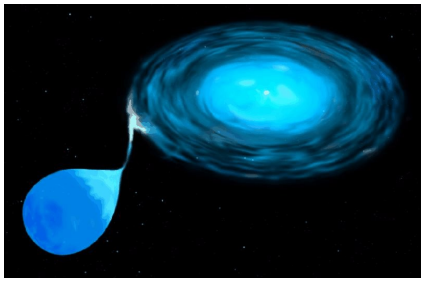
- However, it also creates new problems as it does not provide a theoretical motivation for the flat "slow-roll" shape of the inflaton potential nor does it motivate the existence of the scalar field. Besides, it requires reheating at the end of inflation to recover the hot state of the Big Bang. This could be obtained through a coupling of the inflaton field to ordinary matter, for which there is no fundamental reason in the underlying theory.

3.3 Dark Energy

3.3.1 Expansion of the Universe

- Observations force us to accept that the cosmological constant today makes up $\approx 70\%$ of the energy density of the universe.
- Measurements of the CMB power spectrum reveal that the universe is spatially flat or very close to flat, i.e. the total energy density contributed by all constituents of the cosmic fluid equals the critical density.
- We know from the CMB itself, but also from other observations, that the matter density, dark and baryonic, contributes approximately 30% to the total energy density, and the abundance of light elements requires the baryon density to be much lower. In the framework of the Friedmann model, the remaining 70% of the energy density must be contributed by the cosmological constant.
- The most important class of observations supporting this conclusion is supernovae of type Ia. The progenitor system of a type Ia supernova consists of a binary with at least one massive ($\approx 1\ M_\odot$) carbon-oxygen white dwarf:
  - The *single-degenerate scenario* assumes that the companion of the white dwarf is an evolved star. When the companion star becomes a red giant, it grows over its Roche volume and transfers mass to the white dwarf. White dwarfs are stabilised by the Fermi pressure of a degenerate electrons gas. This can only stabilise masses up to $1.4\,M_\odot$ against gravity. When the companion star feeds the white dwarf beyond this limit, a thermonuclear runaway burning is eventually triggered, which explodes the white dwarf. This scenario appears to be ruled out for explaining the majority of type Ia supernovae.
  - Alternatively, the *double-degenerate scenario* assumes the existence of a binary consisting of two carbon-oxygen white dwarfs. At the end of their evolution, they merge and cause a thermonuclear runaway burning of carbon and oxygen in the more massive progenitor. The resulting type Ia supernovae explosion generates $\approx 1\ M_\odot$ ^{56}Ni, which decays radioactively into ^{56}Ni and eventually to ^{56}Fe. This decay is responsible for the extraordinary brightness of type Ia supernovae ($\sim 10^{11}$ times more luminous in comparison to a star on the main sequence).
- The mass of the produced ^{56}Ni varies by nearly an order of magnitude and so does the intrinsic integrated bolometric luminosity of

Supernovae 1994 d



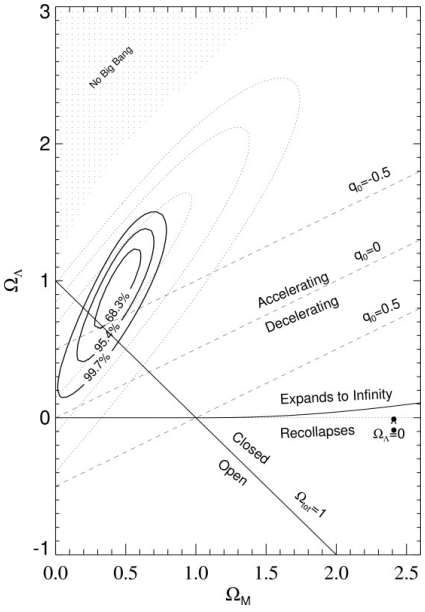
A white dwarf fed by a companion star.

type Ia supernovae. However, the more ^{56}Ni there is produced, the higher is the opacity for the optical photons, which increases the diffusion time through the explosives to the surface of the supernovae. As a result, the light curve of more luminous supernovae is broader and the peak luminosity is higher. There exists an empirical relation between the width of the light curve and the peak luminosity, which enables to calibrate them to an absolute luminosity. Thus, type Ia supernovae form a class of "standardisable candles".

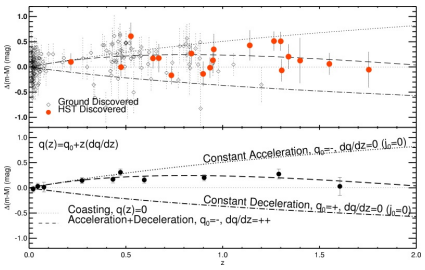
- Knowing their absolute luminosity and observing their apparent brightness, their (luminosity) distances can be inferred. Their redshift can be determined from their spectra. Thus, it is possible to reconstruct the luminosity distance as a function of redshift.

- Initially very surprisingly, the distance turns out to be significantly larger than expected in a universe without cosmological constant. Observations of type-Ia supernovae first forced cosmologists to take seriously the possibility that the universe undergoes accelerated expansion.

- Meanwhile, high-redshift supernovae have shown that the expansion of the universe turned over from deceleration to acceleration around a redshift of unity.



Cosmological parameter range compatible with SN-Ia observations.



The cosmic expansion turned from deceleration to acceleration near $z \sim 1$.

3.3.2 Modified Equation of State

- This is an unfavourable situation because we have no idea what the cosmological constant may be, and it is entirely unclear why at present the density parameters of matter and the cosmological constant should be anywhere near equality.

- A simple estimate of the energy or equivalent matter density of the cosmological constant produces an awfully wrong result. A natural density scale would be the Planck mass divided by the cubed Planck length, which gives

$$\rho = \frac{m_\mathrm{P}}{l_\mathrm{P}^3} \approx \frac{10^{-5}}{(10^{-33})^3}\,\mathrm{g\,cm^{-3}} \approx 10^{94}\,\mathrm{g\,cm^{-3}}\,, \tag{3.81}$$

  which is about 123 orders of magnitude larger than the critical density of the universe $\rho_\mathrm{cr} \approx 10^{-29}\,\mathrm{g\,cm^{-3}}$.

- A similar answer is obtained by considering the zero-point energy of a scalar field of mass m, momentum k, and frequency ω. (For this argument, we use natural units, $\hbar = c = 1$.) The zero-point energy is given by $E = \omega/2 = \sqrt{k^2 + m^2}/2$. Summing over the

zero-point energies of this field up to a cut-off scale $k_{\max} \gg m$, we obtain the vacuum energy density

$$\rho_{\rm vac} = \int_0^{k_{\max}} \frac{{\rm d}^3k}{(2\pi)^3} \frac{\sqrt{k^2+m^2}}{2} \approx \int_0^{k_{\max}} \frac{4\pi k^2 {\rm d}k}{(2\pi)^3} \frac{k}{2} = \frac{k_{\max}^4}{16\pi^2} \,. \tag{3.82}$$

In the first step, we used the fact that the integral is dominated by large-k modes since $k_{\max} \gg m$. Taking $k_{\max} = m_{\rm p}$, we obtain $\rho_{\rm vac} \approx 10^{75}$ GeV^4, which is 122 orders of magnitudes larger than the observed energy density of the cosmological constant, $\rho_\Lambda \approx 10^{-47}$ GeV^4.

- The main reasons why the cosmological constant is considered necessary are that the total matter density is much smaller than unity, while the spatial curvature of the universe is close or equal to zero, and that observations of supernovae of type Ia require the expansion of the universe to be accelerated.

- Seeking a physical explanation for the cosmological constant, it is useful to look at cosmological inflation, which also grew from the requirement of accelerated expansion. As we have seen there, this requires a form of matter whose pressure is

$$p < -\frac{1}{3}\rho c^2 \,, \tag{3.83}$$

while the cosmological constant has $p = -\rho c^2$.

- It is plausible to generalise the equation of state (3.83) as

$$p = w\rho c^2 \,, \quad w < -\frac{1}{3} \,, \tag{3.84}$$

with a parameter w which may or may not depend on time. Forms of matter with such equations of state have been termed "quintessence".

- Suppose for simplicity that w is constant. Then the continuity equation requires

$$\frac{\rm d}{{\rm d}t}\left(a^3\rho_{\rm Q}c^2\right) + w\rho_{\rm Q}c^2\frac{\rm d}{{\rm d}t}\left(a^3\right) = 0 \,, \tag{3.85}$$

which implies

$$\rho_{\rm Q} = \rho_{\rm Q0}a^{-3(1+w)} \,, \tag{3.86}$$

where $\rho_{\rm Q0}$ is the quintessence density today. Evidently, the behaviour of the cosmological constant is recovered for $w = -1$.

- Replacing Ω_Λ by $\Omega_{\rm Q}$, and ignoring the radiation density, the Friedmann equation reads

$$H^2(a) = H_0^2\left[\Omega_{\rm m0}a^{-3} + (1-\Omega_{\rm m0}-\Omega_{\rm Q0})a^{-2} + \Omega_{\rm Q0}a^{-3(1+w)}\right] \,. \tag{3.87}$$

For $w = -1/3$, the quintessence terms cancel, and the equation looks like the Friedmann equation for an open model with $\Omega_{\rm m0}$ only and $\Omega_{\rm Q0} = 0$.

- If w is not constant, the continuity equation leads to

$$\rho_{\rm Q}(a) = \rho_{\rm Q0} \exp\left[-3 \int_a^1 (1+w) {\rm d} \ln a\right] . \tag{3.88}$$

- As for cosmological inflation, a self-interacting scalar field is one candidate for a form of matter which can have negative pressure. The ratio w between pressure and density is

$$w = \frac{p}{\rho c^2} = \frac{\dot\phi^2/2 - V(\phi)}{\dot\phi^2/2 + V(\phi)} , \tag{3.89}$$

and the scalar field ϕ satisfies the evolution equation (3.61) ,

$$\ddot\phi + 3H\dot\phi + V'(\phi) = 0 . \tag{3.90}$$

3.3.3 Models of Dark Energy

- So far, the interaction potential $V(\phi)$ is completely unconstrained. One suggestion is

$$V(\phi) = \frac{\kappa}{\phi^\alpha} , \tag{3.91}$$

the so-called Ratra-Peebles potential. The constant κ has the dimension $(\text{mass})^{4+\alpha}$. It needs to be set such as to agree with the quintessence density parameter today.

- For a power-law expansion, $a \propto t^n$, the Hubble function reads $H = \dot a/a = n/t$ and the evolution equation (3.90) reduces to an algebraic equation, which admits power-law solutions for ϕ,

$$\phi = \phi_0 t^{2/(2+\alpha)} . \tag{3.92}$$

The kinetic term

$$\dot\phi = \frac{2\phi_0}{2+\alpha} t^{-\alpha/(2+\alpha)} \tag{3.93}$$

decays for $\alpha > 0$. Solving the algebraic equation imposes a consistency relation for the amplitude of the scalar field evolution (ϕ_0) and the parameters of the potential,

$$\kappa = \phi_0^{\alpha+2} \frac{2(6n - \alpha + 3\alpha n)}{\alpha(2+\alpha)^2} . \tag{3.94}$$

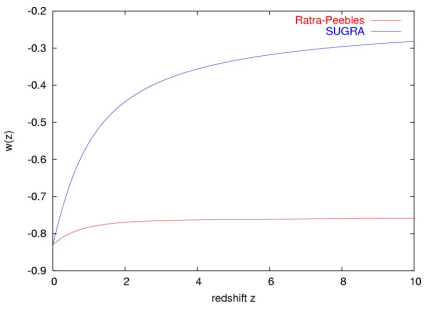
- The energy density of the quintessence field then scales as

$$\rho_{\mathrm{Q}} = \frac{1}{2}\dot{\phi}^2 + V(\phi) \propto t^{-2\alpha/(2+\alpha)} \,, \tag{3.95}$$

and its ratio to the density of matter or radiation scales as

$$\frac{\rho_{\mathrm{Q}}}{\rho} \propto t^{2-2\alpha/(2+\alpha)} = t^{4/(2+\alpha)} \,, \tag{3.96}$$

because the densities of matter and radiation both scale $\propto t^2$ while they dominate the expansion. For $\alpha \to 0$, the quintessence density ρ_{Q} approaches a constant and reproduces the behaviour of the cosmological constant. For $\alpha > -2$, the quintessence density decays more slowly than that of matter or radiation, leading ϕ to dominate the expansion of the universe at late times.



Equation-of-state parameter w as a function of redshift for two models of dark energy.

- If $\alpha > -2$, the field grows arbitrarily large in this model, thus V approaches zero. For $\alpha > 0$, the energy density $\rho_{\mathrm{Q}} \to 0$.

- A favourable aspect of the model (3.91) is that it has so-called *tracker* properties, meaning that a wide variety of initial conditions ϕ and $\dot{\phi}$ lead to the same final solution for ϕ. This may help solving the coincidence problem, which states that nearly equal values for Ω_Λ and Ω_{m} today seem to require delicate fine-tuning in the early universe.

- Another model, which is motivated by super-gravity theories, has an exponential term in addition to the power-law potential,

$$V(\phi) = \frac{K}{\phi^\alpha} \mathrm{e}^{4\pi G\phi^2} \,. \tag{3.97}$$

It shares the tracker property with the power-law model, but has a significantly different behaviour.

3.3.4 Effects on Cosmology

- The modified expansion rate in quintessence models may have pronounced cosmological consequences on age and distances, nucleosynthesis, the microwave background, structure formation and so forth.

- Since nucleosynthesis depends critically on how the expansion time scale compares to the time scales of neutron decay and the nuclear interactions, the cosmic expansion during nucleosynthesis is tightly constrained by observations of the light-element abundances. Thus, at the time of nucleosynthesis, the quintessence field must be negligible compared to the radiation density which otherwise drives the expansion.

- Changes in the expansion time scale during CMB recombination changes the width of the recombination shell and thus modifies the height of the high-order acoustic peaks. If expansion is faster, the temperature of the cosmic plasma drops more rapidly, the recombination shell becomes thinner, thus fewer small-scale fluctuations are projected onto each other looking into the recombination shell, the damping of the high-order acoustic peaks is reduced, so they can be higher.

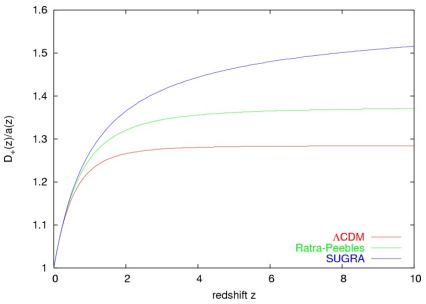
- Modified expansion behaviour changes the curvature of spacetime, and thus the angular-diameter and luminosity distances. This influences the appearance of supernovae of type Ia, the apparent size of fluctuations in the CMB, the cosmic volume of redshift shells, and the overall geometry of the universe, and thus effects like gravitational lensing.

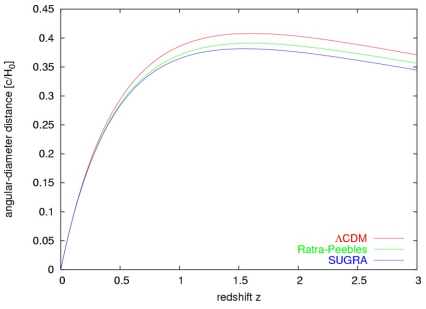
- The growth factor is modified, typically in such a way that structures form earlier in quintessence compared to cosmological-constant models. Structures are thus expected to be present at higher redshifts in quintessence models, and more pronounced at given redshifts, compared to the cosmological-constant case.

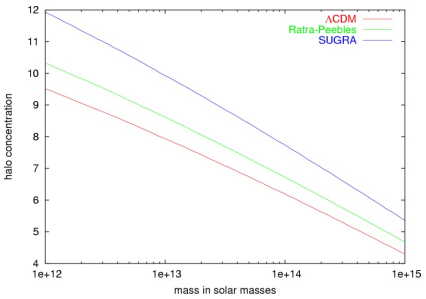
- Halo collapse against the universal expansion is modified, which implies that the spherical collapse proceeds differently. Consequently, the spherical-collapse parameters δ_c and Δ_v are modified, having pronounced effects on halo statistics (e.g. through the Press-Schechter mass function).

- The core densities of haloes appear to reflect the cosmic background density at their formation times. Since quintessence makes haloes form earlier, they tend to be denser in their cores, which may have strong effects on their appearance (e.g. through gravitational lensing, X-ray emission, and so forth).

- The modified growth factor in quintessence models changes the time evolution of fluctuations in the gravitational potential. Photons propagating from the CMB recombination shell throughout the universe thus experience changes in the gravitational potential which are stronger than in the cosmological-constant model. A larger fraction of the CMB amplitude is thus of secondary rather than primary origin, possibly changing the normalisation of the power spectrum.







Growth factor, angular-diameter distance, and halo concentrations in ΛCDM and two dark-energy models.

Chapter 4

The Late Universe

4.1 Galaxies and Gas

4.1.1 Ellipticals and Spirals

- Galaxies are objects with typical sizes of a few kpc, while their typical distances are of order Mpc, so they are clearly distinguished entities.
- Galaxies typically consist of a central, more or less amorphous, nearly spherically-symmetric part, called the bulge, and a flattened, structure, called the disk.
- Bulges contain predominantly old, metal-poor, red population-II stars which have an almost isotropic velocity dispersion.
- Disks contain more metal-rich, younger, blue population-I stars which move around the centre in nearly circular orbits.
- Galaxies are classified by the ratio between bulges and disks. Those dominated by the bulge are called ellipticals, those dominated by the disk are called spirals, and there is a continuous classification range in between, the Hubble sequence. Historically, ellipticals are also called early-type, and spirals late-type galaxies.
- Disks have near-exponential intensity profiles,

$$I(r) = I_0 \exp\left(-\frac{r}{r_0}\right) , \tag{4.1}$$

  with the scale length r_0, while bulges have the less steep de-Vaucouleurs- or $r^{1/4}$ profile,

$$I(r) = I_0 \exp\left[-\left(\frac{r}{r_0}\right)^{1/4}\right] . \tag{4.2}$$

- Other types of galaxy are less easily fit into this scheme, such as the irregular, dwarf, or blue compact galaxies.
- Spectra of ellipticals show signatures of old stellar populations. They correspond to a temperature near 5000 K, are rich in metal lines, and dominated by giant stars moving off the stellar main sequence.
- Spectra of spirals are characterised by signatures of recent star formation. They contain young, hotter, bluer stars with less absorption features. The radiation of the young stars can ionise ambient gas and thus produce narrow nebular emission lines.

Galaxy morphologies are classified by the ratio between bulges and disks.

- The metal abundances in galaxies reflect metal production by type-II supernovae, which are the end products of massive-star evolution. Typically, metal abundances increase with increasing galaxy mass (a consequence of hierarchical bottom-up growth of galaxies) and towards galaxy centres (which form stars earlier than the more dilute external regions of the galaxies).

- Ellipticals and spirals are characterised by a completely different dynamics of their collisionless stellar systems: ellipticals are supported by the three-dimensional velocity dispersion σ of the orbiting stars while spirals are rotationally supported, i.e., the stars orbit with v_c in nearly circular orbits. Galaxy luminosities and dynamical properties obey

  1. the Faber-Jackson relation (for ellipticals), and
  2. the Tully-Fisher relation (for spirals),

  $$\frac{\sigma_v}{\sigma_{v,*}} \overset{(1)}{=} \left(\frac{L}{L_*}\right)^{1/\alpha} \overset{(2)}{=} \frac{v_\mathrm{c}}{v_{\mathrm{c},*}} , \quad \alpha \simeq 3 \ldots 4 . \tag{4.3}$$

  Both velocity scales are of the same order, $v_{\mathrm{c},*} \simeq 220\,\mathrm{km\,s^{-1}} \simeq \sigma_{v,*}$. Since $v_\mathrm{c} = \sqrt{2}\sigma_v$, ellipticals with the same luminosity are more massive than spirals. While the scatter of the Faber-Jackson relation is about 1 magnitude, that of the Tully Fisher relation only amounts to about 0.4 magnitudes. By adding information about the central surface-brightness I_0, the Faber-Jackson relation for ellipticals can be replaced by a tighter relation, the *fundamental plane*, which is defined by

  $$L \propto I_0^{-0.7}\, \sigma_v^3 , \tag{4.4}$$

  with a scatter of about 0.4 magnitudes.

- Elliptical and spiral galaxy populations inhabit different regions of space. While spirals dominate in low-density regions (well outside galaxy clusters), ellipticals predominantly inhabit high-density regions like cluster cores. Apparently, disks do not survive in dense environments.

4.1.2 Spectra, Magnitudes and K-Corrections

- The intensity of electromagnetic radiation is characterised by the energy (ε) received per unit time and unit detector area (A) from unit solid angle on the sky (Ω) and per unit frequency interval. This is called the *specific intensity*,

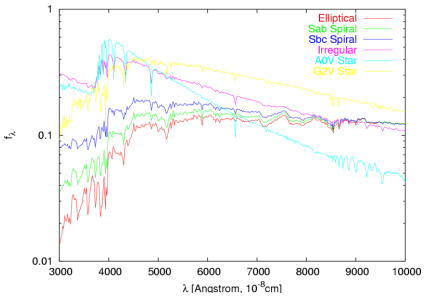
  $$I_\nu = \frac{\mathrm{d}^4\varepsilon}{\mathrm{d}t\,\mathrm{d}A\,\mathrm{d}\Omega\,\mathrm{d}\nu} . \tag{4.5}$$

When integrated over the solid angle of a source, it is called the *flux density* S_ν, which is consequently the energy received per area, time and frequency. Its conventional unit is Jansky,

$$1\,\mathrm{Jy} = 10^{-26}\,\frac{\mathrm{W}}{\mathrm{m^2\,Hz}} = 10^{-23}\,\frac{\mathrm{erg}}{\mathrm{s\,cm^2\,Hz}}\ . \tag{4.6}$$

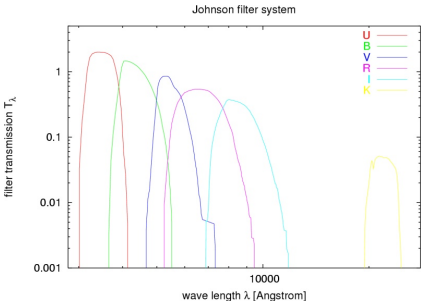
- We will loosely speak of the *flux* below, which can be specific intensity if not integrated over solid angle, flux density if integrated over solid angle, or flux if integrated over detector area. If f_ν is the flux per unit frequency, the flux f_λ per unit wavelength is

$$f_\lambda = \left|\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right| f_\nu = \frac{c}{\lambda^2} f_\nu\ . \tag{4.7}$$



Spectra of different galaxy types.

- Intensities are measured through filters with transmission functions T_ν or T_λ. Sets of transmission curves define a filter system, such as the Johnson-UBVRI system or that used by the Sloan Digital Sky Survey (SDSS).



Transmission curves of the Johnson filter system.

- The transmission curves define the effective wavelength

$$\lambda_{\mathrm{eff}} := \frac{\int \mathrm{d}\lambda \lambda T_\lambda}{\int \mathrm{d}\lambda T_\lambda}\ , \tag{4.8}$$

and the sensitivity

$$Q := \int \mathrm{d}\ln\nu T_\nu\ . \tag{4.9}$$

- At least in optical astronomy, fluxes are commonly measured in magnitudes, which provide a peculiarly defined logarithmic scale. Generally, the magnitude *difference* of two objects is

$$\Delta m = -2.5\log_{10}\left(\frac{R_1}{R_2}\right)\ ,\quad R_i = \int \mathrm{d}\lambda f_{\lambda,i} T_\lambda \tag{4.10}$$

if $R_{1,2}$ are the instrumental responses to the flux received from objects 1 and 2. The zero point is commonly defined as the instrumental response to the flux of a standard star (e.g. α Lyrae, which is an A0V star).

- For so-called *AB* magnitudes, the zero point is defined in terms of the physical flux in Jy. For instance, the *AB* magnitude system used by the SDSS is defined by

$$m = -2.5\log_{10}\frac{\int \mathrm{d}\ln\nu\, f_\nu T_\nu}{Q} - 48.6 \tag{4.11}$$

- This can directly be related to the number of electrons released in a CCD. The energy received per unit time and unit frequency interval by a telescope with collecting area A is $\mathrm{d}E = A\mathrm{d}t\mathrm{d}\nu f_\nu$. This energy comes in form of $\mathrm{d}N_\gamma = \mathrm{d}E/(h\nu)$ photons, a fraction T_ν of which can pass the filter. Thus, the number of photons arriving at the CCD, or the number of electrons released by the CCD assuming 100% efficiency of the CCD in converting photons to electrons, is
$$N_\mathrm{e} = \frac{At}{h}\int \mathrm{d}\ln\nu\, f_\nu T_\nu\,, \tag{4.12}$$
where t is the total exposure time.

- For example, an object with an AB magnitude of $m = 25$ in a given filter band with sensitivity $Q = 0.1$ has
$$\int \mathrm{d}\ln\nu\, T_\nu = 3.6\times10^{-30}\,, \tag{4.13}$$
and thus releases
$$\frac{N_\mathrm{e}}{At} = 5.5\times10^{-4} \tag{4.14}$$
electrons per second exposure time and cm^2 collecting area. Hence, a CCD attached to a telescope with 4 m mirror diameter releases ~ 70 electrons per second from such an object.

- The *absolute* magnitude M of an object is the magnitude the object would have if its distance was 10 pc from the observer. If its true (luminosity!) distance is D_L and its magnitude is m, the absolute magnitude is
$$M = m + 5\log_{10}\left(\frac{D_L}{10\,\mathrm{pc}}\right) = -2.5\log_{10}\left[\frac{R(f_\lambda)}{D_L^2}\frac{(10\,\mathrm{pc})^2}{R_{10}}\right]. \tag{4.15}$$

- For objects at cosmological distances, the K-correction must be applied which takes into account that the spectrum is redshifted with respect to the fixed filter
$$K(z) = 2.5\log_{10}\frac{\int \mathrm{d}\lambda\, f_\lambda T_\lambda}{\int \mathrm{d}\lambda\, f_{\lambda/(1+z)} T_\lambda} = 2.5\log_{10}\frac{R(f_\lambda)}{R(f_{\lambda/(1+z)})}\,. \tag{4.16}$$
This modifies the absolute magnitude according to
$$M = m+5\log_{10}\left(\frac{D_L}{10\,\mathrm{pc}}\right)+K(z) = -2.5\log_{10}\left[\frac{R(f_{\lambda/(1+z)})}{D_L^2}\frac{(10\,\mathrm{pc})^2}{R_{10}}\right]. \tag{4.17}$$

- Since $\lambda f_\lambda = \nu f_\nu$, the K-correction for power-law spectra, $f_\nu \propto \nu^{-\alpha}$, is
$$K = 2.5\log_{10}\frac{\int \mathrm{d}\ln\nu\,\nu^{-\alpha+1}\nu^2 T_\nu}{\int \mathrm{d}\ln\nu\,(1+z)^{-\alpha+1}\nu^{-\alpha+1}\nu^2 T_\nu} = 2.5(\alpha-1)\log_{10}(1+z), \tag{4.18}$$

i.e. the K-correction vanishes for spectra $\propto \nu^{-1}$. It becomes positive for bluer (steeper) spectra with $\alpha > 1$ and negative for redder (flatter) spectra.

4.1.3 Luminosity Functions

- The number density of galaxies with luminosities between L and $L+\mathrm{d}L$ is described by the luminosity function. Its measurement is quite involved because it requires a detailed understanding of the survey characteristics, in particular the survey selection function.

- Measured galaxy luminosity functions are typically well fit by the Schechter function,

$$\mathrm{d}\phi(L) = \phi_* \left(\frac{L}{L_*}\right)^{\alpha} \exp\left(-\frac{L}{L_*}\right) \frac{\mathrm{d}L}{L_*} \ . \tag{4.19}$$

- At the faint-end slope $\alpha = -1.0 \pm 0.15$, quite independent of the galaxy type. The cut-off luminosity L_* is brighter for ellipticals than for spirals. Its mean value is $M_* = -19.50 \pm 0.13$ in the photographic B_J filter band, rising from $M_* = -19.59$ for ellipticals to $M_* = -19.39$ for spirals to $M_* = -18.94$ for irregulars.

- Ellipticals contribute $\sim 35\%$ to ϕ_*, spirals $\sim 57\%$, and irregulars $\sim 8\%$. The overall normalisation is $\phi_* \approx (0.0140 \pm 0.0017)h^3$, but its exact value is uncertain because it still depends on galaxy selection, and is locally sensitive to galaxy clustering.

- A cosmologically important number to derive from the luminosity function is the luminosity density

$$\rho_L = \int_0^{\infty} L\mathrm{d}\phi(L) = \Gamma(\alpha+2)\phi_* L_* \ , \tag{4.20}$$

where

$$\Gamma(x) = \int_0^{\infty} \mathrm{e}^{-t} t^{x-1} \mathrm{d}t \tag{4.21}$$

is the gamma function.

- The galaxy luminosity function in galaxy clusters is very similar to that outside clusters at intermediate luminosities, but deviations exist at the bright and the faint ends. At the bright end, luminous cD galaxies exist in the centres of many clusters which are not simply the brightest objects drawn from a Schechter function. At the faint end, the luminosity function steepens considerably due to a dwarf population which has $\alpha \sim -1.8$. Such a dwarf galaxy population may also exist outside clusters.

- There is no compelling evidence for brighter galaxies to be more strongly clustered (luminosity segregation). However, the Butcher-Oemler effect says that the fraction of blue galaxies in clusters increases with increasing redshift. This is probably a consequence of both enhanced star formation in cluster galaxies at moderate and high redshifts, and later depletion of star-forming galaxies due to mergers.

- While the luminosity function in the (near-infrared) K band does not evolve with redshift out to $z \sim 0.6$, it exhibits strong evolution in the B band. There is a significant population of faint blue galaxies at moderate and high redshifts which seems to be actively star-forming.

- Metals (i.e. all elements heavier than helium) are produced in stars, mostly in stars more massive and less long-lived than the Sun. Since metals are produced by nuclear fusion with a mass-to-energy conversion efficiency near 1%, the luminosity density of galaxies can be related to the metal abundance. The evolution of the luminosity density with redshift then allows the metal production to be deduced as a function of redshift. In turn, this yields the star-formation rate as a function of redshift. Apparently, most stars were formed between redshifts 1 and 2.

- Approximately 10% of the energy produced during that time should be radiated in the narrow Lyman-α line, so that a population of Lyman-α emitting galaxies should be seen. This has not been observed in *targeted* observations, probably because a large amount of star formation happens in environments with neutral hydrogen (which absorbs Lyman-α radiation) or in dust-shrouded environments which scatter the radiation into the infrared. The cosmic infrared background is consistent with this picture. Indeed, blind surveys have now delivered about 200 Lyman-α emitters which are presumably associated with young starburst galaxies.

4.1.4 Correlation Functions and Biasing

- The density-fluctuation field has the power spectrum $P(k)$ defined in (2.43). Its correlation function given by (2.46), thus the power spectrum is related to the correlation function by

$$\begin{aligned} P(k) &= \int \mathrm{d}^3x\xi(x)\mathrm{e}^{\mathrm{i}\vec{k}\vec{x}} = 2\pi\int_0^\infty x^2\mathrm{d}x\,\xi(x)\int_0^\pi \sin\theta\mathrm{d}\theta\mathrm{e}^{\mathrm{i}kx\cos\theta} \\ &= 4\pi\int_0^\infty x^2\mathrm{d}x\,\xi(x)\frac{\sin kx}{kx}\,. \end{aligned} \tag{4.22}$$

- Observationally, the correlation function of the galaxies describes the excess probability above random for finding a galaxy at distance x from another. Let $\mathrm{d}V_1$ and $\mathrm{d}V_2$ be two infinitesimally small volume elements separated by r, and n the number density of galaxies. Then, the probability $\mathrm{d}P$ for finding one galaxy in $\mathrm{d}V_1$ and another in $\mathrm{d}V_2$ is $\mathrm{d}P = n^2\mathrm{d}V_1\mathrm{d}V_2$. If the galaxies are randomly distributed. If the galaxies are correlated, this probability becomes
$$\mathrm{d}P = n^2[1 + \xi(r)]\mathrm{d}V_1\mathrm{d}V_2 \ . \tag{4.23}$$

- This gives the principle for measuring $\xi(r)$: in a volume-limited survey of galaxies, count pairs of galaxies separated by a distance between r and $r + \mathrm{d}r$, and compare it to the pair counts expected if the galaxies were randomly distributed. For instance, let $\langle DD\rangle$ and $\langle RR\rangle$ be the pair counts in the data (D) and the randomised (R) galaxy surveys, then
$$\xi = \frac{\langle DD\rangle}{\langle RR\rangle} - 1 \ , \tag{4.24}$$
is one estimate for ξ.

- A simple assumption holds that the number density of galaxies is related to the density contrast by
$$\frac{\delta n}{n} =: \delta^{\mathrm{gal}} = b\delta = \delta + (b-1)\delta \ , \tag{4.25}$$
where b is the bias factor, which can be inferred from velocity measurements, as we will see in the following.

- However, in order to measure the real-space galaxy correlation function $\xi_{\mathrm{gal}}(\vec{r}) = \langle\delta_{\mathrm{gal}}(\vec{x})\delta_{\mathrm{gal}}(\vec{x}+\vec{r})\rangle$, we need three-dimensional positions of all galaxies. Using a galaxy's redshift, its observable angular position can be turned into a transverse distance. The redshift itself can be turned into a line-of-sight distance, provided the total line-of-sight velocity is solely interpreted as a Hubble velocity. If the galaxies' peculiar velocities were random, averaging over a large number of galaxies would reduce their contribution to the correlation function and result in an unbiased measurement of ξ_{gal}. However, there are two systematic effects that bias the galaxies' peculiar velocities in a unique way and preclude such an approach.

    1. On large scales, galaxies are predominantly attracted by clusters and super-clusters such that there is an infalling velocity bias. This causes galaxies that are on the near side of a cluster to recess (implying a redshift) and galaxies on the far side of a cluster to approach (implying a blueshift). Overall, this "Kaiser effect" implies a squeezing of the apparent radial scale while the transverse direction remains unaffected.

2. On small scales, the galaxies' motions within a gravitationally bound structure such as a cluster are virialized and reflect its potential depth. This "finger-of-god-effect" leads to an apparent elongation of scales along the line-of-sight while the transverse scale remains unaffected.

- To quantify this and demonstrate how this can be used to infer the linear bias parameter, we will now derive the Kaiser effect. Density perturbations δ give rise to peculiar motion and displacements
$$\delta\vec{x} = \frac{\vec{r}}{a} - \vec{x}\,, \tag{4.26}$$
from which δ can be inferred according to
$$\delta = -\vec{\nabla}\cdot\delta\vec{x}\,, \tag{4.27}$$
which follows in the framework of the Zel'dovich approximation, cf. Eqs. (2.56, 2.60 and 2.63)

- Peculiar velocities $\vec{u}$ cause displacements
$$\delta\vec{x} = \frac{\vec{u}}{Hf(\Omega)} \tag{4.28}$$
of the comoving coordinates (cf. 2.66).

- The peculiar motion adds to the Hubble velocity. The apparent comoving distance to a galaxy is inferred from its observed line-of-sight velocity
$$v = \vec{v}\cdot\vec{e}_x = a(H\vec{x} + \vec{u})\cdot\vec{e}_x\,, \tag{4.29}$$
where $\vec{e}_x$ is the line-of-sight direction.

- Interpreting the total velocity as Hubble velocity implies that the apparent comoving distance vector to a galaxy is
$$\vec{x}_{\rm app} = \frac{\vec{v}}{aH} = \vec{x}_{\rm real} + \frac{\vec{u}\cdot\vec{e}_x}{H}\vec{e}_x\,. \tag{4.30}$$

- An apparent displacement $\delta\vec{x}_{\rm app}$ is thus related to the real displacement $\delta\vec{x}_{\rm real}$ by
$$\delta\vec{x}_{\rm app} = \delta\vec{x}_{\rm real} + \frac{\vec{u}\cdot\vec{e}_x}{H}\vec{e}_x = \delta\vec{x}_{\rm real} + f(\Omega)(\delta\vec{x}_{\rm real}\cdot\vec{e}_x)\vec{e}_x\,. \tag{4.31}$$

- Because $\delta\vec{v} \propto -\vec{\nabla}\delta\Phi$, a density perturbation with wave vector $\vec{k}$ causes a displacement parallel to $\vec{k}$. Let μ be the cosine of the angle between the line-of-sight and $\vec{k}$, then $\delta\vec{x}_{\rm real}\cdot\vec{e}_x = \delta x_{\rm real}\mu$, $\vec{k}\cdot\vec{e}_x = k\mu$, and
$$\hat{\delta} = -\mathrm{i}\vec{k}\cdot\delta\hat{\vec{x}} = -\mathrm{i}k\delta\hat{x}\,. \tag{4.32}$$
From this, we obtain
$$\delta_{\rm app} = \delta_{\rm real}\left[1 + f(\Omega)\mu^2\right]\,. \tag{4.33}$$

- The apparent density contrast in the galaxy counts is thus related to the real density contrast by the term caused by the velocity perturbations plus the biasing term,

$$\delta^{\rm gal}_{\rm app} = \delta_{\rm real}\left[1 + f(\Omega)\mu^2\right] + (b-1)\delta_{\rm real} = \delta^{\rm gal}_{\rm real}\left[1 + \frac{f(\Omega)\mu^2}{b}\right] . \tag{4.34}$$

- The peculiar anisotropy caused by the factor μ^2 can be used to measure

$$\beta := \frac{f(\Omega)}{b} . \tag{4.35}$$

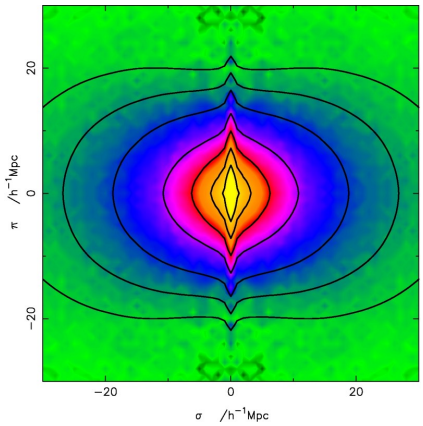
The ratio between the redshift- and real-space power spectra is

$$\frac{P_{\rm app}}{P_{\rm real}} = \left(1 + \beta\mu^2\right)^2 , \tag{4.36}$$

which can be written as

$$\frac{P_{\rm app}}{P_{\rm real}} = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right) + \left(\frac{4\beta}{3} + \frac{4\beta^2}{7}\right)P_2(\mu) + \frac{8\beta^2}{35}P_4(\mu) , \tag{4.37}$$

where $P_{2,4}(\mu)$ are the Legendre polynomials. The redshift-space power spectrum thus exhibits a characteristic quadrupolar pattern, and the ratio between quadrupole and monopole can be used to infer β.



Two-dimensional galaxy correlation function measured from the 2dF Galaxy Redshift Survey.

- On small scales, virialised motion within bound structures (e.g. galaxy clusters) leads to an apparent extension along the line-of-sight ("finger-of-god effect"). This can approximately be described by damping in Fourier space according to

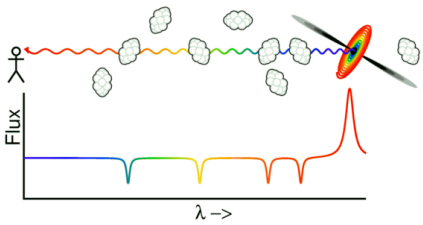
$$\hat{\delta} \to \hat{\delta}(1 + k^2\mu^2\sigma^2)^{-1/2} , \tag{4.38}$$

where σ is the velocity dispersion of the galaxies within the bound structure. The overall effect is then

$$\frac{P_{\rm app}}{P_{\rm real}} = \frac{\left(1 + \beta\mu^2\right)^2}{1 + k^2\mu^2\sigma^2} . \tag{4.39}$$

4.1.5 Intervening Gas

- The light from distant sources such as quasars (i.e., accreting supermassive black holes) passes through diffuse gas. The neutral hydrogen within these clouds absorbs rest-frame Lyman-α photons. Quasar light that is emitted blueward of Lyman-α (with an energy of $E > h\nu_{{\rm Ly}\alpha}$) is redshifted during propagation. Once it is traversing neutral hydrogen (HI) at redshift z, an absorption feature at $E = (1+z)h\nu_{{\rm Ly}\alpha}$ is produced in the observer frame. The



ubiquity of HI clouds causes the emergence of a forest of Lyman-α absorption features. Characterising the (one- and two-point) statistics of the resulting absorption lines offers an important way to study the large-scale structure and fundamental cosmology.

- The shape of absorption lines is given by the Lorentz profile

$$\frac{\mathrm{d}p}{\mathrm{d}\omega} = \frac{\Gamma/2}{(\omega-\omega_0)^2+(\Gamma/2)^2}\,, \tag{4.40}$$

  which can be considered as the probability distribution for a photon of frequency ω to be absorbed by an atom with a transition frequency ω_0. Γ is the line width.

- The Lorentz profile arises in the theory of the damped classical harmonic oscillator, where Γ is the damping rate. Quantum-mechanically, Γ^{-1} is the lifetime of the excited state resulting from the absorption.

- The natural line width defined by the decay probability of the excited state is often increased by atomic collisions, which shorten the lifetime and thus broaden the absorption line.

- If the gas moves thermally with respect to the line of sight, the resulting absorption-line profile is a convolution of the Lorentz profile with a Gaussian

$$\frac{\mathrm{d}p}{\mathrm{d}\omega} = \frac{\Gamma}{(2\pi)^{3/2}\sigma}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-v^2/2\sigma^2}\mathrm{d}v}{(\omega-\omega_0-\omega_0\,v/c)^2+(\Gamma/2)^2}\,, \tag{4.41}$$

  which is called the Voigt profile. It has a Gaussian core and Lorentzian wings.

- The absorption cross section of the Lyman-α transition of a (neutral) hydrogen atom in thermal equilibrium is

$$\sigma(\omega) = 6.9\times10^{-2}\,\frac{\mathrm{d}p}{\mathrm{d}\omega}\,\mathrm{cm}^2\,, \tag{4.42}$$

  which gives rise to the optical depth

$$\tau_{\mathrm{HI}}(\omega) = \sigma(\omega)\int n\,\mathrm{d}l := \sigma(\omega)\,N_\mathrm{c}\,, \tag{4.43}$$

  which is the cross section times the column density N_c, i.e. the hydrogen number density n integrated over the line-of-sight.

- The central optical depth of a Lyman-α line which is Doppler broadened with a velocity dispersion σ_v, the central optical depth is

$$\tau_{\mathrm{HI},0} = \left(\frac{\sigma_v}{\mathrm{km\,s^{-1}}}\right)^{-1}\left(\frac{N_\mathrm{c}}{1.86\times10^{12}\,\mathrm{cm^{-2}}}\right)\,. \tag{4.44}$$

  Typical velocity dispersions are of order a few tens of $\mathrm{km\,s^{-1}}$, thus measurable central optical depths of ~ 0.1 are reached with column densities of $N_\mathrm{c}\sim10^{12}\,\mathrm{cm^{-2}}$.

- The observed probability distribution of column densities is very wide and approximately follows a power-law

$$P(> N_\mathrm{c}) \propto N_\mathrm{c}^{-0.75} \,, \tag{4.45}$$

  up to $N_\mathrm{c} \approx 10^{21}\,\mathrm{cm}^{-2}$.

- When $N_\mathrm{c} \approx 10^{18}\,\mathrm{cm}^{-2}$, the optical depth becomes unity in the Lorentzian wings rather than the Gaussian core of the lines. Such saturated lines are called "damped" and the absorbers "damped" Lyman-α absorbers.

- If absorbers have the typical absorption cross section $\sigma(z)$ and a physical number density of $n_\mathrm{HI}(z)$, their expected number per unit redshift is

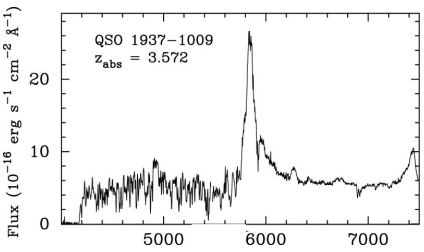
$$\mathrm{d}N = \sigma(z)\, n_\mathrm{HI}(z) \left| \frac{\mathrm{d}D_\mathrm{prop}}{\mathrm{d}z} \right| \mathrm{d}z \,, \tag{4.46}$$

  with the proper-distance D_prop given in (1.58). The redshift distribution of absorbers is the power law

$$\frac{\mathrm{d}N}{\mathrm{d}z} \propto (1+z)^{2.3\pm0.4} \,. \tag{4.47}$$

- Quasars typically have strong redshifted Lyman-α emission lines, which are absorbed by intervening neutral hydrogen gas. The total optical depth for that absorption is

$$\tau_\mathrm{HI} = \int_0^{z_\mathrm{Q}} \sigma[(1+z)\omega_0]\, n_\mathrm{HI}(z) \left| \frac{\mathrm{d}D_\mathrm{prop}}{\mathrm{d}z} \right| \mathrm{d}z \,. \tag{4.48}$$



The Lyman-α forest blueward of the Lyman-α emission line.

- If there was continuously distributed neutral hydrogen along the line-of-sight to any distant quasar, all flux blueward of the Lyman-α emission line should be absorbed, which is not observed. This indicates that the intergalactic hydrogen must be ionised.

- This *Gunn-Peterson effect* implies remarkably tight bounds on the density parameter in neutral hydrogen. For instance, the absence of complete absorption in the spectra of quasars near redshift $z_\mathrm{Q} \approx 5$ implies

$$\Omega_\mathrm{HI} \lesssim 1.5 \times 10^{-8}\, h^{-1} \,. \tag{4.49}$$

- Complete absorption has recently been detected in quasars just above redshift $z_\mathrm{Q} = 6$, which may indicate that the universe was reionised around that redshift. However, even small admixtures of neutral hydrogen are sufficient to cause complete absorption, thus reionisation may have started considerably earlier. In fact, the mean Thompson optical depth towards the CMB

$$\tau_\mathrm{e} = \sigma_\mathrm{T} \int_0^{z_\mathrm{CMB}} n_\mathrm{e}(z) \left| \frac{\mathrm{d}D_\mathrm{prop}}{\mathrm{d}z} \right| \mathrm{d}z \approx 0.09 \tag{4.50}$$

by *Planck* implies a mean reionisation redshift of $z_{\rm reion} \approx 9$ (assuming an instantaneous reionisation history).

- Hydrogen absorption lines trace the gas distribution, which should follow the density distribution of the dark matter. Lyman-α absorbers are thus an important tracer for large-scale structures and constrain the density-fluctuation power spectrum on small scales.

4.2 Gravitational Lensing

4.2.1 Assumptions, Index of Refraction

- Due to space-time curvature, masses and other concentrations of energy deflect light towards themselves, in a way similar to convex glass lenses. This gives rise to an effect called "gravitational lensing".

- Basic assumptions in conventional lensing theory are that the Newtonian gravitational potential Φ of the lensing mass is small in the sense $\Phi \ll c^2$, and that the extent of the lenses L along the line-of-sight is small compared to the Hubble length, $L \ll c/H_0$.

- Under these conditions, the Minkowski metric of flat space-time is modified. Instead of

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}\vec{x}^2 \,, \tag{4.51}$$

  the line element becomes

$$\mathrm{d}s^2 = \left(1 + \frac{2\Phi}{c^2}\right)c^2\mathrm{d}t^2 - \left(1 - \frac{2\Phi}{c^2}\right)\mathrm{d}\vec{x}^2 \,, \tag{4.52}$$

  i.e. the coefficients of $c^2\mathrm{d}t^2$ and $\mathrm{d}\vec{x}^2$ are perturbed away from unity. According to the general assumptions above, these perturbations are small.

- Since light propagates according to $\mathrm{d}s^2 = 0$, the metric (4.52) implies

$$\left(1 + \frac{\Phi}{c^2}\right)c\mathrm{d}t = \left(1 - \frac{\Phi}{c^2}\right)|\mathrm{d}\vec{x}| \,, \tag{4.53}$$

  where we have used that $(1 + 2x)^{1/2} \approx (1 + x)$ for $x \ll 1$.

- The speed of light is thus changed in presence of the perturbing potential to

$$c' = \frac{|\mathrm{d}\vec{x}|}{\mathrm{d}t} = c\left(1 + \frac{2\Phi}{c^2}\right) =: \frac{c}{n} \,, \tag{4.54}$$

  where

$$n := \left(1 - \frac{2\Phi}{c^2}\right) \geq 1 \tag{4.55}$$

  is the effective index of refraction of a weak gravitational field. Since $\Phi \leq 0$, $n \geq 1$, thus $c' \leq c$.

- Consequently, there arises a time delay compared to light propagation in vacuum. We have

$$\mathrm{d}(\Delta t) = \frac{\mathrm{d}x}{c'} - \frac{\mathrm{d}x}{c} = \frac{\mathrm{d}x}{c}(n - 1) = -\frac{2\Phi}{c^3}\mathrm{d}x \,, \tag{4.56}$$

and obtain the *Shapiro delay* in a gravitational field

$$\Delta t = -\frac{2}{c^3}\int \Phi \mathrm{d}x \,, \tag{4.57}$$

where the integral is evaluated along the line-of-sight.

4.2.2 Deflection Angle and Lens Equation

- In complete analogy to geometrical optics, we can now use Fermat's principle to calculate the deflection of light caused by the refractive index. Fermat's principle requires the light-travel time between fixed points 1 and 2 to be extremal, thus

$$\delta \int_1^2 n(\vec{x})\mathrm{d}x = 0 \,. \tag{4.58}$$

Introducing a parameter λ running along the light path, this reads

$$\int_1^2 n[\vec{x}(\lambda)]|\dot{\vec{x}}|\mathrm{d}\lambda \,, \tag{4.59}$$

with $\dot{\vec{x}} := \mathrm{d}\vec{x}/\mathrm{d}\lambda$.

- Using $|\dot{\vec{x}}| = (\dot{\vec{x}}^2)^{1/2}$, Euler's equation reads

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{\vec{x}}} - \frac{\partial L}{\partial \vec{x}} = 0 \,, \tag{4.60}$$

with $L = n(\vec{x})(\dot{\vec{x}}^2)^{1/2}$.

- The derivative $\dot{\vec{x}}$ is proportional to the tangent vector to the light ray. The curve parameter λ can be normalised such that $\dot{\vec{x}} = \vec{e}$, the unit tangent vector. We then find from Euler's equation

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}n(\vec{x})\vec{e} - \vec{\nabla}n = n\dot{\vec{e}} + (\vec{\nabla}n\cdot\vec{e})\vec{e} - \vec{\nabla}n = 0 \,. \tag{4.61}$$

Since $n - 1 \ll 1$, $\vec{\nabla}n/n = \vec{\nabla}\ln n \approx \vec{\nabla}n$, and we obtain for the change of the tangent vector along the light ray

$$\dot{\vec{e}} = \vec{\nabla}n - (\vec{\nabla}n\cdot\vec{e})\vec{e} = \vec{\nabla}_\perp n = -\frac{2}{c^2}\vec{\nabla}_\perp\Phi \,, \tag{4.62}$$

i.e. $\dot{\vec{e}}$ is determined by the component of the gradient of n perpendicular to the line-of-sight.

- The total change of the direction of $\vec{e}$ is the deflection angle

$$\hat{\vec{\alpha}} = \frac{2}{c^2}\int \vec{\nabla}_\perp\Phi \mathrm{d}l \,, \tag{4.63}$$

where the integral is carried out along an unperturbed, straight line instead of the true, curved, line-of-sight in the spirit of the Born approximation for small-angle scattering.

- According to the second assumption, the thin lenses can be projected along the line-of-sight. Their surface-mass density is

$$\Sigma(\vec{b}) = \int \rho(\vec{b}, z)\mathrm{d}z \ , \tag{4.64}$$

and their deflection angle is the superposition of the deflection angles of all infinitesimal mass elements,

$$\hat{\vec{\alpha}}(\vec{b}) = \frac{4G}{c^2} \int \mathrm{d}^2 b' \frac{\Sigma(\vec{b}')(\vec{b} - \vec{b}')}{|\vec{b} - \vec{b}'|^2} \ . \tag{4.65}$$

- If $D_{\mathrm{d,s,ds}}$ are the angular-diameter distances from the observer to the lens and the source, and from the lens to the source, respectively, the relation

$$D_{\mathrm{s}}\vec{\beta} = D_{\mathrm{s}}\vec{\theta} - D_{\mathrm{ds}}\hat{\vec{\alpha}} \tag{4.66}$$

obviously holds, where $\vec{\beta}$ and $\vec{\theta}$ are the angular positions of source and image on the sky relative to the optical axis. This is the lens equation.

- Introducing the reduced deflection angle

$$\vec{\alpha} := \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\vec{\alpha}} \tag{4.67}$$

the lens equation becomes

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}) \ . \tag{4.68}$$

- The surface-mass density Σ, scaled with the critical surface mass density

$$\Sigma_{\mathrm{cr}} := \left[\frac{4\pi G}{c^2}\frac{D_{\mathrm{d}}D_{\mathrm{ds}}}{D_{\mathrm{s}}}\right]^{-1} \tag{4.69}$$

is the *convergence* $\kappa := \Sigma/\Sigma_{\mathrm{cr}}$.

- The lensing potential is a weighted projection of the Newtonian potential

$$\psi(\vec{\theta}) := \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}}\frac{2}{c^2}\int \Phi(D_{\mathrm{d}}\vec{\theta})\,\mathrm{d}z \ . \tag{4.70}$$

Its gradient is the (reduced) deflection angle

$$\vec{\nabla}_\theta\psi(\vec{\theta}) = D_{\mathrm{d}}\vec{\nabla}_\perp\psi = \frac{2}{c^2}\frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\int \vec{\nabla}_\perp\Phi(D_{\mathrm{d}}\vec{\theta}, z)\mathrm{d}z = \vec{\alpha}(\vec{\theta}) \ . \tag{4.71}$$

Its Laplacian is the convergence

$$\Delta_\theta\psi(\vec{\theta}) = \frac{2}{c^2}\frac{D_{\mathrm{d}}D_{\mathrm{ds}}}{D_{\mathrm{s}}}\int \Delta\Phi(D_{\mathrm{d}}\vec{\theta}, z)\mathrm{d}z = 2\kappa \ , \tag{4.72}$$

where Poisson's equation and the definition of the critical surface-mass density have been used in the last steps.

4.2.3 Local Lens Mapping and Mass Reconstruction

- The local imaging properties of a lens are described by the Jacobian of the lens mapping

$$\mathcal{A} = \frac{\partial\vec{\beta}}{\partial\vec{\theta}} = \left[\delta_{ij} - \frac{\partial\alpha_i}{\partial\theta_j}\right] = \left[\delta_{ij} - \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}\right] := \left[\delta_{ij} - \psi_{ij}\right] , \quad (4.73)$$

  which is obviously symmetric. The local lens mapping is thus determined by the curvature of the lensing potential ψ.

- Images are locally magnified by a factor

$$\mu := \det\left(\frac{\partial\vec{\theta}}{\partial\vec{\beta}}\right) = \det\left(\mathcal{A}^{-1}\right) = \frac{1}{\det\mathcal{A}} . \quad (4.74)$$

- The trace of the Jacobian is

$$\mathrm{tr}\mathcal{A} = 2 - \Delta\psi = 2(1-\kappa) , \quad (4.75)$$

  subtracting it from A leaves the trace-free shear matrix

$$\Gamma_{ij} := \mathcal{A}_{ij} - \frac{\delta_{ij}}{2}\mathrm{tr}\mathcal{A} = \kappa\delta_{ij} - \psi_{ij} , \quad (4.76)$$

  which is symmetric and has the components $\gamma_1 = (\psi_{11} - \psi_{22})/2$ and $\gamma_2 = \psi_{12}$

$$\Gamma = -\begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} . \quad (4.77)$$

  Thus, the Jacobian can be decomposed into an isotropic part, responsible for isotropic image stretching, and an anisotropic, trace-free part, responsible for image distortion.

- Convergence and shear are different linear combinations of second derivatives of ψ, thus κ can be reconstructed from measurable image distortions. In Fourier space

$$\hat{\kappa} = -\frac{1}{2}\left(k_1^2 + k_2^2\right)\hat{\psi} , \quad \hat{\gamma}_1 = -\frac{1}{2}\left(k_1^2 - k_2^2\right)\hat{\psi} , \quad \hat{\gamma}_2 = -k_1k_2\hat{\psi} . \quad (4.78)$$

  Thus

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \frac{1}{k^2}\begin{pmatrix} k_1^2 - k_2^2 \\ 2k_1k_2 \end{pmatrix}\hat{\kappa} . \quad (4.79)$$

- This can easily be inverted noting that

$$\left[\frac{1}{k^2}\begin{pmatrix} k_1^2 - k_2^2 \\ 2k_1k_2 \end{pmatrix}\right]^2 = 1 , \quad (4.80)$$

  so that

$$\hat{\kappa} = \frac{1}{k^2}\begin{pmatrix} k_1^2 - k_2^2 \\ 2k_1k_2 \end{pmatrix}\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \frac{1}{k^2}\left[(k_1^2 - k_2^2)\hat{\gamma}_1 + 2k_1k_2\hat{\gamma}_2\right] . \quad (4.81)$$

This expression for $\hat{\kappa}$ is easily transformed back into configuration space

$$\kappa = \frac{1}{\pi}\int \mathrm{d}^2\theta' \,\Re\left[\mathcal{D}(\vec{\theta}-\vec{\theta}')\gamma(\vec{\theta}')\right] , \tag{4.82}$$

with $\gamma := \gamma_1 + \mathrm{i}\gamma_2$ and the kernel

$$\mathcal{D}(\vec{\theta}) = \frac{\theta_2^2 - \theta_1^2 + 2\mathrm{i}\theta_1\theta_2}{\theta^4} . \tag{4.83}$$

4.2.4 Deflection by Large-Scale Structures

- Light propagation in General Relativity, specialised to the Friedmann-Lemaître-Robertson-Walker metric, yields the result that the comoving separation of two light rays $\vec{x}$ evolves with the radial coordinate w as

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}w^2} + Kw = 0 , \tag{4.84}$$

with K given in (1.44). This is an oscillator equation with the solutions $f_K(w)$ given in (1.7).

- Near localised inhomogeneities, space-time can be approximated as Minkowskian, perturbed by the lensing potential Φ, which gives rise to the light deflection

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}w^2} = -\frac{2}{c^2}\vec{\nabla}_\perp\Phi , \tag{4.85}$$

as shown in (4.62), where the curve parameter λ has been replaced by w.

- The combined light deflection by the space-time curved on large scales, and the superposed small-scale perturbations, is thus

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}w^2} + K\vec{x} = -\frac{2}{c^2}\vec{\nabla}_\perp\Phi . \tag{4.86}$$

This is the equation for an externally driven harmonic oscillator. The solution can be found using the Green's function of the harmonic oscillator to be

$$\vec{x}(\vec{\theta}, w) = f_K(w)\vec{\theta} - \frac{2}{c^2}\int_0^w \mathrm{d}w'\, f_K(w-w')\vec{\nabla}_\perp\Phi[f_K(w')\vec{\theta}] . \tag{4.87}$$

- The deflection angle is the deviation of the true separation of the light rays from the separation expected in homogeneous space-time, divided by the distance to the sources

$$\begin{aligned}\vec{\alpha}(\vec{\theta}, w) &= \frac{f_K(w)\vec{\theta} - \vec{x}(\vec{\theta}, w)}{f_K(w)} \\ &= \frac{2}{c^2}\int_0^w \mathrm{d}w'\,\frac{f_K(w-w')}{f_K(w)}\vec{\nabla}_\perp\Phi[f_K(w')\vec{\theta}] .\end{aligned} \tag{4.88}$$

- As for the thin-lens case, where $2\kappa = \Delta\psi = \vec{\nabla}^2\psi = \vec{\nabla}\cdot\alpha$, the effective convergence is defined as

$$\kappa_{\rm eff} = \frac{1}{2}\vec{\nabla}\cdot\vec{\alpha} = \int_0^w {\rm d}w' \, \frac{f_K(w-w')f_K(w')}{f_K(w)} \left(\frac{\partial^2\Phi}{\partial x_i \partial x_i}\right)[f_K(w')\vec{\theta}] \, . \tag{4.89}$$

Inserting Poisson's equation (2.17)

$$\Delta\Phi = \frac{3H_0^2}{2a}\Omega_{\rm m0}\delta \, , \tag{4.90}$$

yields

$$\kappa_{\rm eff} = \frac{1}{c^2}\int_0^w {\rm d}w' \, W(w,w')\delta[f_K(w')\vec{\theta}] \, , \tag{4.91}$$

with

$$W(w,w') := \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \frac{\Omega_{\rm m0}}{a}\frac{f_K(w-w')f_K(w')}{f_K(w)} \, . \tag{4.92}$$

4.2.5 Limber's Equation and Weak-Lensing Power Spectra

- Given a homogeneous and isotropic random field $f(\vec{x}, w)$ with power spectrum $P_f(k)$, and a weighted projection

$$g(\vec{x}) := \int {\rm d}w \, q(w) \, f(\vec{x}, w) \, , \tag{4.93}$$

what is the power spectrum $P_g(l)$ of g, where l is a two-dimensional wave number?

- Suppose $q(z)$ is varying on much larger scales than f, Limber's equation holds

$$P_g(l) = \int {\rm d}w \, \frac{q^2(w)}{f_K^2(w)} P_f\left[\frac{l}{f_K(w)}\right] \, . \tag{4.94}$$

- Eq. (4.91) for the effective convergence is of the type (4.93), with q represented by W and f represented by δ. The condition for Limber's equation is well satisfied because the density contrast δ is varying on much smaller scales than W. Thus

$$P_\kappa(l) = \int_0^w {\rm d}w' \, \frac{W^2(w,w')}{f_K^2(w')} P_\delta\left[\frac{l}{f_K(w')}\right] \, . \tag{4.95}$$

- As in the thin-lens case, magnification and shear are defined via the Jacobian matrix of the lens mapping

$$\mathcal{A}_{ij} = \delta_{ij} - \frac{\partial\alpha_i}{\partial\theta_j} \, . \tag{4.96}$$

To first order in the $\partial\alpha_i/\partial\theta_j$, the magnification is

$$\mu = \left(1 - \frac{\partial\alpha_1}{\partial\theta_1} - \frac{\partial\alpha_2}{\partial\theta_2}\right)^{-1} = 1 + \vec{\nabla}\cdot\vec{\alpha} = 1 + 2\kappa_{\rm eff} \ . \tag{4.97}$$

- The statistics of μ and the shear γ are identical to the statistics of $\kappa_{\rm eff}$ except for constant factors. This is obvious for the statistics of the magnification fluctuation

$$\delta\mu = 2\kappa_{\rm eff} \ \Rightarrow \ P_{\delta\mu}(l) = 4P_\kappa(l) \ . \tag{4.98}$$

Considering the shear components in Fourier space, we have

$$\begin{aligned} \left\langle\hat{\gamma}_1^2\right\rangle &= \frac{(l_1^2 - l_2^2)^2}{4}\langle\hat{\psi}^2\rangle \ , \quad \left\langle\hat{\gamma}_2^2\right\rangle = (l_1 l_2)^2\langle\hat{\psi}^2\rangle \ , \\ \left\langle\hat{\kappa}_{\rm eff}^2\right\rangle &= \frac{(l_1^2 + l_2^2)^2}{4}\langle\hat{\psi}^2\rangle \ , \end{aligned} \tag{4.99}$$

and thus

$$\left\langle|\hat{\gamma}|^2\right\rangle = \frac{1}{4}(l_1^4 + 2l_1^2 l_2^2 + l_2^4)\langle\hat{\psi}^2\rangle = \frac{(l_1^2 + l_2^2)^2}{4}\langle\hat{\psi}^2\rangle = \left\langle\hat{\kappa}_{\rm eff}^2\right\rangle \ . \tag{4.100}$$

Thus the power spectra of the cosmic shear and the effective convergence are identical

$$P_\gamma(l) = P_\kappa(l) \ . \tag{4.101}$$

- Following (2.46), the correlation function of the effective convergence is

$$\xi_\kappa(\phi) = \left\langle\kappa_{\rm eff}(\vec{\theta})\kappa_{\rm eff}(\vec{\theta} + \vec{\phi})\right\rangle = \int \frac{{\rm d}^2 l}{(2\pi)^2} P_\kappa(l){\rm e}^{-{\rm i}\vec{l}\cdot\vec{\phi}} \ . \tag{4.102}$$

Note that the wave vector $\vec{l}$ is now two-dimensional, thus the integral over the angle enclosed by the vectors $\vec{l}$ and $\vec{\phi}$ yields

$$\xi_\kappa(\phi) = \int_0^\infty \frac{l{\rm d}l}{2\pi} P_\kappa(l){\rm J}_0(l\phi) \ , \tag{4.103}$$

where ${\rm J}_0(x)$ is the zeroth-order Bessel function of the first kind. This is identical to the shear correlation function ξ_γ.

- On angular scales of arc minutes, the typical expected shear- and convergence correlation functions are of order 10^{-4}, thus typical shear values on such scales are of order a few per cent.

- Albeit weak, the shear can be measured quantifying the distortions of the images of distant galaxies. The shear correlation function can then be compared to the theoretical expectation (4.103) in order to constrain cosmological parameters and the dark-matter power spectrum. This has been achieved with spectacularly solid results, leading to an independent confirmation of the standard, low-density, spatially flat cosmological model with cosmological constant.

- The cosmic-shear measurements are expected to contribute substantially to answering the question about the equation of state of the dark energy.

4.3 Galaxy Clusters

4.3.1 Galaxies in Clusters

- Galaxy clusters are a cosmologically important class of objects. They trace the most pronounced density peaks of large-scale structure. They are the largest gravitationally bound objects in the universe, assemble the latest in cosmic history, and thus reflect structure growth. They are closed objects in that their interior does not mix with outside. They are an overdense environment which impacts on the evolution of their member galaxies.

- Galaxy clusters were originally defined as regions in the sky with enhanced galaxy number density. An example are Abell's criteria: (1) at least 50 galaxies in the magnitude range $[m_3, m_3 + 2]$, where m_3 is the magnitude of the third-brightest cluster galaxy. (2) the galaxies are enclosed by the Abell radius $R_\mathrm{A} = 1.5\,h^{-1}\mathrm{Mpc}$. And (3) their redshift falls within $[0.01, 0.2]$. Abell's famous cluster catalogue is built on these criteria. Many other definitions and catalogues exist.

- Abell's catalogue contains 4076 clusters, of which 2683 have richness class $R \geq 1$. This corresponds to a local number density of rich clusters of $n \sim 10^{-5}\,h^3\,\mathrm{Mpc}^{-3}$. The mean separation between clusters is thus $\sim n^{-1/3} \sim 50\,h^{-1}\,\mathrm{Mpc}$.

- Elliptical galaxies are enriched compared to spiral galaxies in clusters. The galaxy population at intermediate luminosities is well-described by a Schechter luminosity function, but there are deviations both at the bright and the faint ends. CD galaxies are a special, bright class of objects in cluster centres. At the faint end, the luminosity function steepens considerably.

- The number density of galaxies in clusters is approximately described by a cored distribution

$$n(r) = n_0 \left(1 + \frac{r^2}{r_\mathrm{c}^2}\right)^{-3/2} , \tag{4.104}$$

  with the core radius $r_\mathrm{c} \sim 120\,h^{-1}\,\mathrm{kpc}$ and the central number density $n_0 \sim 2 \times 10^4\,h^3\,\mathrm{Mpc}^{-3}$.

- Galaxies move within the gravitational potential well of the cluster. They have a velocity distribution centred on the bulk velocity of the cluster with a velocity dispersion

$$\sigma_v^2 = \left\langle v_\parallel^2 \right\rangle - \left\langle v_\parallel \right\rangle^2 , \tag{4.105}$$

  where $v_\parallel$ is the velocity component parallel to the line-of-sight. Typical cluster velocity dispersions are of order $\sim 1000\,\mathrm{km\,s}^{-1}$.

- Moving with this velocity, galaxies take approximately a few Gyr to cross galaxy clusters, i.e. an amount of time comparable to the Hubble time. It is thus unclear whether galaxy clusters can be considered as relaxed objects in equilibrium (and the definition of equilibrium in self-gravitating systems is equally unclear).

- For a galaxy of mass m at radius R enclosing the cluster mass M, the virial theorem demands

$$2K = -V \;\Rightarrow\; 2\frac{m}{2}(3\sigma_v^2) = \frac{GMm}{R}\,, \tag{4.106}$$

where the factor 3 comes in because σ_v is the dispersion along one spatial direction only and we neglected the surface pressure term due to accretion. This yields the mass estimate

$$M \sim \frac{3R\sigma_v^2}{G} = 10^{15}\,h^{-1}\,\mathrm{M}_\odot \left(\frac{R}{1.5\,h^{-1}\,\mathrm{Mpc}}\right)\left(\frac{\sigma_v}{1000\,\mathrm{km\,s^{-1}}}\right)^2\,. \tag{4.107}$$

Although the application of the virial theorem is questionable, this mass is approximately 10 times the mass visible in galaxies. This was the first hint at a substantial amount of dark matter in the universe.

- For self-gravitating gas spheres in hydrostatic equilibrium, the hydrostatic equation reads

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{GM(r)}{r^2}\rho\,, \tag{4.108}$$

where p and ρ are the gas pressure and density, respectively. For an ideal gas, $p = \rho kT/m$, where m is the mean particle mass. Thus,

$$\frac{kT}{m}\frac{\mathrm{d}\rho}{\mathrm{d}r} + \frac{\rho k}{m}\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{GM}{r^2}\rho\,. \tag{4.109}$$

- Considering the motion of galaxies within the dark-matter dominated cluster as the motion of a gas with temperature

$$\frac{3}{2}kT = \frac{m}{2}(3\sigma_v^2) \;\Rightarrow\; T = \frac{m\sigma_v^2}{k}\,, \tag{4.110}$$

in an external potential well created by the mass M, eq. (4.109) becomes

$$M = -\frac{r\sigma_v^2}{G}\left(\frac{\mathrm{d}\ln\rho}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln\sigma_v^2}{\mathrm{d}\ln r}\right)\,, \tag{4.111}$$

where ρ is now the (number) density of galaxies.

4.3.2 X-Ray Emission

- Soon after X-ray detectors were first used in astronomy, it was detected that galaxy clusters are the brightest X-ray sources in the sky. When X-ray spectra could be taken, it was discovered that the X-ray radiation has an exponential cut-off characteristic of thermal radiation. When the sources could be spatially resolved, clusters turned out to be diffuse sources.

- The X-ray radiation thus reveals that clusters are filled with thermal gas which is hot enough for emitting X-rays. In an ionised, hot gas (a plasma), electrons scatter off the Coulomb field of ions and radiate because of their acceleration. This thermal *bremsstrahlung* (free-free emission) is the only relevant cooling process for a hot plasma of temperature $kT > 2$ keV. Below this temperature, recombination-line emission of heavy elements such as iron starts to dominate the cooling (and the associated X-ray emission; assuming typical heavy element abundances, which are are about one third of those found in the Sun).

- Heuristically, the X-ray emissivity $j_\nu(\vec{x})$ (i.e. the amount of energy emitted in photons of frequency ν per unit frequency interval $d\nu$, per unit time and unit plasma volume) must scale with the squared particle number density because it is a two-body process. It further scales with the time available for the scattering process, which is proportional to the inverse relative velocity of electron and ion, or the inverse square root of the temperature, $t \sim \Delta x/\Delta v \sim \Delta x m/\sqrt{kT}$, since $\Delta v \simeq c_\mathrm{s}$; and with the Boltzmann factor for the distribution of energy at a given temperature. Accordingly, we expect

$$j_\nu(\vec{x}) = \frac{\mathrm{d}^3 E}{\mathrm{d}\nu\,\mathrm{d}t\,\mathrm{d}V} = C\,\frac{\rho^2}{\sqrt{kT}}\mathrm{e}^{-h\nu/kT}\ , \tag{4.112}$$

  where C is a constant. This is confirmed by the theory of radiation processes. The volume emissivity is the integral of j_ν over the frequency,

$$j(\vec{x}) = \frac{\mathrm{d}^2 E}{\mathrm{d}t\,\mathrm{d}V} = \int_0^\infty \mathrm{d}\nu\, j_\nu(\vec{x}) = \tilde{C}\,\rho^2\,\sqrt{kT} \tag{4.113}$$

$$\simeq 2.5\times 10^{-23}\left(\frac{n_\mathrm{H}}{1\ \mathrm{cm}^{-3}}\right)^2\left(\frac{T}{10^8\ \mathrm{K}}\right)^{1/2}\frac{\mathrm{erg}}{\mathrm{cm}^3\ \mathrm{s}}, \tag{4.114}$$

  for a metallicity of $Z = 0.3 Z_\odot$.

- If the gas has density ρ and temperature T, eq. (4.109) requires

$$M(r) = -\frac{rkT}{Gm}\left(\frac{\mathrm{d}\ln\rho}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln T}{\mathrm{d}\ln r}\right)\ . \tag{4.115}$$

- Combining this with the mass estimate (4.111), we have

$$\sigma_v^2 \left(\frac{\mathrm{d}\ln\rho_{\rm gal}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln\sigma_v^2}{\mathrm{d}\ln r} \right) = \frac{kT}{m} \left(\frac{\mathrm{d}\ln\rho_{\rm gas}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln T}{\mathrm{d}\ln r} \right) . \quad (4.116)$$

Introducing the ratio of specific energies

$$\beta := \frac{m\sigma_v^2}{kT} , \quad (4.117)$$

yields

$$\mathrm{d}\ln\rho_{\rm gas} = \beta(\mathrm{d}\ln\rho_{\rm gal} + \mathrm{d}\ln\sigma_v^2) - \mathrm{d}\ln T . \quad (4.118)$$

Using the definition of β, $\mathrm{d}\ln\sigma_v^2 = \mathrm{d}\ln T + \mathrm{d}\ln\beta$, and (4.118) becomes

$$\mathrm{d}\ln\rho_{\rm gas} = \beta\mathrm{d}\ln\rho_{\rm gal} + (\beta-1)\mathrm{d}\ln T + \mathrm{d}\beta , \quad (4.119)$$

and thus

$$\rho_{\rm gas} \propto \rho_{\rm gal}^{\beta}\, T^{\beta-1} . \quad (4.120)$$

- Assuming isothermal gas, its distribution should thus follow the galaxy distribution to the power of the β parameter. Adopting the galaxy distribution (4.104) suggests the β profile

$$\rho_{\rm gas} = \rho_0 \left(1 + \frac{r^2}{r_0^2} \right)^{-3\beta/2} . \quad (4.121)$$

Since the X-ray emissivity is $\propto \rho^2$, this implies

$$j_\nu(r) = j_0 \left(1 + \frac{r^2}{r_0^2} \right)^{-3\beta} , \quad (4.122)$$

and, after projection, the X-ray flux per unit solid angle is given by

$$S_{\rm X} = 2\int_0^\infty \mathrm{d}z\, j_\nu[r(z)] = 2\int_{r_\perp}^\infty \frac{\mathrm{d}r\, j_\nu(r)}{\sqrt{r^2 - r_\perp^2}} \quad (4.123)$$

$$= \frac{2 j_0 r_0 \sqrt{\pi}\,\Gamma(3\beta - 1/2)}{\Gamma(3\beta)} \left(1 + \frac{r_\perp^2}{r_0^2} \right)^{-3\beta+1/2} \quad (4.124)$$

$$=: S_{\rm X0} \left(1 + \frac{\theta^2}{\theta_0^2} \right)^{-3\beta+1/2} , \quad (4.125)$$

where $z = \sqrt{r^2 - r_\perp^2}$. This routinely provides excellent fits to the X-ray surface brightness of observed clusters with $r_0 \sim 200\,h^{-1}$ kpc and $\beta \sim 2/3$.

- Such "β fits" yield the derivative $\mathrm{d}\ln\rho_{\rm gas}/\mathrm{d}\ln r$ and thus the isothermal mass estimate

$$M(r) = \frac{3\beta rkT}{Gm}\frac{r^2/r_0^2}{1+r^2/r_0^2} \ . \tag{4.126}$$

  such mass estimates can be highly misleading because of the many assumptions they rely on. (4.126) implies $M(r) \propto r$ for $r \gg r_0$.

- Assuming an NFW dark-matter density profile (2.129) and gas in hydrostatic equilibrium with it, yields density and X-ray surface-brightness profiles which can excellently be fit with β-profiles out to radii of order R_{500}, but the resulting mass profiles are wrong. At larger radii, the steeper NFW profile causes a steepening of the gas profile. Besides, the kinetic pressure contribution, the clumping factors ($C = \langle\rho^2\rangle/\langle\rho\rangle^2$), and anisotropies due to the filamentary nature of mass accretion increase steeply with radius and cause the simplified spherical approach presented her to fail at scales approaching the virial radius.

- Explaining the total X-ray luminosities of clusters requires central particle number densities of

$$\frac{\rho_0}{m} \sim 10^{-2}\,\mathrm{cm}^{-3} \ . \tag{4.127}$$

  Total gas masses are of order $\sim (10-20)\%$ of the total cluster masses, which corresponds to the cosmic baryon fraction

$$\frac{\Omega_{\rm b0}}{\Omega_{\rm m0}} = \frac{0.047}{0.3} = 16\% \ . \tag{4.128}$$

- Comparing the thermal energy content to the total (frequency-integrated) X-ray emissivity defines the cooling time

$$t_{\rm cool} = \frac{3nkT}{2j} \approx 2\ \mathrm{Gyr}\left(\frac{kT}{\mathrm{keV}}\right)^{1/2}\left(\frac{n_{\rm e}}{10^{-2}\ \mathrm{cm}^{-3}}\right)^{-1}, \tag{4.129}$$

  which drops below the Hubble time in the centres of massive clusters where gas should thus efficiently cool. If gas in pressure equilibrium cools, it becomes denser and cools even faster: this is a runaway process that should lead to a large amount of cold gas and star formation. However, the observed traces of cool gas (Hα emission) and rates of star formation are much lower than the expectations from unimpeded cooling, and recent X-ray spectra do not reveal any spectral signatures (e.g., metal lines) of cool gas ($kT \lesssim 0.5$ keV). This constitutes the famous "cooling flow problem". Therefore, there must be a way of re-heating the cooling gas in cluster cores, which could be provided by Active Galactic Nuclei (AGN) in clusters.

4.3.3 Gravitational Lensing by Galaxy Clusters

- The cores of galaxy clusters are dense enough to produce strong gravitational lensing, giving rise to strongly distorted images of background galaxies, so-called *arcs*. Assuming axial symmetry of the projected mass distribution, arcs should trace a circle with the *Einstein radius* θ_E of the cluster, which is given by the requirement that the mean cluster convergence within the Einstein radius is unity

$$\langle\kappa\rangle = \frac{M(\theta_\mathrm{E})}{\pi(D_\mathrm{d}\theta_\mathrm{E})^2}\,\frac{1}{\Sigma_\mathrm{cr}} \stackrel{!}{=} 1\ , \tag{4.130}$$

  where Σ_cr is the critical surface-mass density defined in (4.69) and D_d is the angular-diameter distance to the cluster.

- If cluster and source redshifts are known, and a cosmological model is adopted, this can be inverted to yield the cluster mass enclosed by the Einstein radius

$$M(\theta_\mathrm{E}) = \pi D_\mathrm{d}^2 \Sigma_\mathrm{cr}\,\theta_\mathrm{E}^2\ . \tag{4.131}$$

- Mass estimates obtained this way are of the same order of magnitude as those found with other techniques, but there are systematic discrepancies. In many clusters, the strong-lensing mass estimate obtained from (4.131) is substantially higher than, e.g. the X-ray mass estimate.

- The reason for such systematic deviations is that clusters are typically highly asymmetric and substructured, which gives rise to strong gravitational tidal fields. This allows strong gravitational lensing effects at a substantially lower cluster mass than that required if the clusters were symmetric.

- Away from their cores, clusters weakly deform the images of background galaxies and thus imprint their approximately tangential shear pattern on them. This distortion is observable as in cosmological weak lensing. Using (4.82), the observed shear pattern can be transformed into a mass map. Such weak-lensing mass measurements typically agree well with X-ray mass determinations.

4.3.4 Sunyaev-Zel'dovich Effects

- The CMB radiation shines through the hot plasma in galaxy clusters and must Compton-scatter off the electrons. Since they are extremely more energetic than the photons, they typically loose energy and scatter the photons to higher energy.

- The photon number is conserved, but the photon energy is increased. The resulting spectrum must thus deviate from the shape of the Planck curve which the photons have before scattering. There must be a lack of photons at low and an increase of photons at high energies compared to the Planck curve. This is the thermal Sunyaev-Zel'dovich (tSZ) effect.

- The relative intensity change at frequency ν is

$$\frac{\delta I}{I} = y \frac{2(kT)^3}{h^2} \frac{x^4 \mathrm{e}^x}{(\mathrm{e}^x - 1)^2} \left[x \coth\left(\frac{x}{2}\right) - 4 \right] , \tag{4.132}$$

  where $x := h\nu/kT$ is the dimensionless energy. Note that T is the CMB temperature as seen by the cluster, and not the electron temperature in the cluster!

- y is the Compton parameter

$$y := \int \frac{kT_\mathrm{e}}{m_\mathrm{e}c^2} \sigma_\mathrm{T} \, n_\mathrm{e} \mathrm{d}l , \tag{4.133}$$

  i.e. the typical relative energy change of a photon in Compton scattering, times the scattering probability. T_e is the electron temperature of the cluster, and σ_T is the Thomson cross section.

- The relative intensity change $\delta I/I$ is negative for frequencies below, and positive above, $x = 3.83$ or $\nu = 217\,\mathrm{GHz}$. Although the zero-crossing frequency depends on the CMB temperature which is higher at high-redshift clusters, it is later redshifted such that the *observed* zero-crossing of the tSZ effect is redshift-independent. This is a most remarkable feature of the tSZ effect.

- Clusters moving with respect to the CMB rest frame additionally Compton-scatter the CMB radiation like mirrors and thus give rise to a frequency shift called the kinetic Sunyaev-Zel'dovich (kSZ) effect. It may be possible to use this effect for measuring the bulk velocities of clusters.

4.3.5 Clusters as Cosmological Tracers

- We have seen in (2.106) that the fraction of cosmic volume filled with haloes of mass M is

$$F(M, a) = \mathrm{erfc}\left(\frac{\delta_\mathrm{c}}{\sqrt{2}\sigma_R(a)} \right) , \tag{4.134}$$

  where $\sigma_R(a)$ is the variance of dark-matter fluctuations filtered on the scale R corresponding to the cluster mass M.

- The observed fraction of the cosmic matter contained in clusters is

$$F'(M,a) = \frac{n_\mathrm{c}(a) M_\mathrm{c}(a)}{\bar{\rho}(a)} \,, \tag{4.135}$$

  where $\bar{\rho}$ is the mean cosmic density, and n_c and M_c are the number densities and masses of observed galaxy clusters. Inserting typical numbers yields

$$F'(M, a = 1) \approx 1\% \, \Omega_\mathrm{m0}^{-1} \,, \tag{4.136}$$

  for typical cluster masses of $\sim 5 \times 10^{14} \, h^{-1} \, M_\odot$.

- Equating this with the expected cluster fraction (4.134) yields an estimate for σ_R, which can be converted to the conventional normalisation parameter σ_8. Typically, values near $0.6 - 0.7$ are found, which are somewhat lower than those found from weak gravitational lensing.

- Comparing the Press-Schechter mass function to the observed mass distribution of clusters at increasing redshifts constrains structure growth as a function of cosmic time, and thus also cosmological parameters, mainly Ω_m0. The lack of strong evolution implies low density in good agreement with $\Omega_\mathrm{m0} = 0.3$.

4.3.6 Scaling Relations

- The total potential energy of a cluster is proportional to the squared mass, divided by the radius

$$V \propto -\frac{GM^2}{R} \,, \tag{4.137}$$

  and the radius scales with the mass like $R \propto M^{1/3}$ (cf. 2.132). Thus, the total potential energy is expected to scale with the mass as

$$V \propto -M^{5/3} \,. \tag{4.138}$$

- The total kinetic energy K is proportional to the temperature T times the number of particles N, i.e. to the product TM. The virial theorem requires $2K = -V$, or

$$TM \propto M^{5/3} \;\Rightarrow\; T \propto M^{2/3} \,. \tag{4.139}$$

  Two orders of magnitude in cluster mass thus correspond to a factor of ~ 20 in cluster temperature.

- The bolometric (i.e. frequency-integrated) X-ray luminosity of a cluster scales like the electron density, times the mass, times the square root of the temperature. Thus

$$L_\mathrm{X} \propto M \, \frac{M}{R^3} \, T^{1/2} \propto M \, M^{1/3} \propto M^{4/3} \propto T^2 \,, \tag{4.140}$$

because $M \propto R^3$.

- These simple scaling relations derived from gravitational physics predict a luminosity-temperature relation $L_X \propto T^{1/2}$ and a mass-temperature relation $M \propto T^3$. While the observed mass-temperature relation is close to that expectation, the luminosity-temperature relation is observed to be flatter than expected.

Appendix A

Additional Material

A.1 Spherical Collapse

A.1.1 Alternative Derivation of Spherical Collapse

- This section provides an alternative derivation of the spherical collapse model for which we write the equations of motion in dimensionless form. Such an approach has the advantange that it can be generalized to include a cosmological constant term in a straight forward manner.

- Suppose this spherical overdensity is embedded into the otherwise homogeneous, expanding background universe. As it is overdense, it will reach a maximum radius and subsequently contract and collapse. We define parameters

$$x := \frac{a}{a_{\rm ta}} , \quad y := \frac{R}{R_{\rm ta}} , \tag{A.1}$$

  i.e. x is the scale factor a in units of the scale factor $a_{\rm ta}$ when the halo reaches its turn-around radius, and y is the radius of the halo R in units of $R_{\rm ta}$.

- We restrict ourselves to the case of an Einstein-de Sitter model, for which

$$H = \frac{\dot{a}}{a} = H_0 a^{-3/2} \tag{A.2}$$

  for simplifying the notation, we introduce the scaled time $\tau := H_{\rm ta} t$, where $H_{\rm ta} = H_0 a_{\rm ta}^{-3/2}$ is the Hubble parameter at the turn-around time. Using these units, Friedmann's equation is transformed to

$$x' := \frac{{\rm d}x}{{\rm d}\tau} = \frac{1}{H_{\rm ta}} \frac{\dot{a}}{a_{\rm ta}} = \frac{H}{H_{\rm ta}} x = x^{-1/2} . \tag{A.3}$$

- The Newtonian equation of motion for the radius (i.e. for a test particle of arbitrary mass at the radius of the halo) is

$$\ddot{R} = -\frac{GM}{R^2} = -\frac{4\pi}{3} \rho_{\rm ta} R_{\rm ta}^3 \frac{G}{R^2} . \tag{A.4}$$

  Introducing τ instead of t, and expressing the density at turn-around by the critical density and the overdensity ζ of the halo with respect to the background at turn-around,

$$\rho_{\rm ta} = \frac{3H_{\rm ta}^2}{8\pi G} \zeta , \tag{A.5}$$

  we find

$$y'' = -\frac{\zeta}{2y^2} . \tag{A.6}$$

- The obvious boundary conditions for solving (A.6) are

$$y'|_{x=1} = 0 \,, \quad y|_{x=0} = 0 \,, \tag{A.7}$$

meaning that the halo starts with zero radius at $a = 0$ and reaches a maximum at $a = a_{\rm ta}$.

- Equations (A.3) and (A.6) imply

$$\tau = \frac{2}{3} x^{3/2} \,, \quad y' = \pm \sqrt{\zeta} \sqrt{\frac{1}{y} - 1} \,, \tag{A.8}$$

where the first boundary condition (A.7) was used. The plus sign applies before, the minus sign after turn-around. Integrating before turn-around, and using the second boundary condition (A.7), we find

$$\tau = \frac{1}{\sqrt{\zeta}} \left[\frac{1}{2} \arcsin(2y - 1) - \sqrt{y - y^2} + \frac{\pi}{4} \right] . \tag{A.9}$$

- Turn-around means $x = 1 = y$ and $\tau = 2/3$, which requires

$$\zeta = \left(\frac{3\pi}{4} \right)^2 . \tag{A.10}$$

From symmetry, collapse happens at twice the time required for turn-around, i.e. at $\tau = 4/3$, at which time $x = x_{\rm c} = 4^{1/3}$.

A.1.2 Collapse Parameters

- At early times, we can expand (A.9) to low order in y and find

$$\tau \approx \frac{8}{9\pi} y^{3/2} \left[1 + \frac{3y}{10} \right] . \tag{A.11}$$

The overdensity inside the halo relative to the background is

$$\Delta = \left(\frac{x}{y} \right)^3 \zeta \,, \tag{A.12}$$

because the background density scales like x^{-3} while the density within the halo scales like y^{-3}. Inserting τ from (A.8) into (A.11) and raising to the power $2/3$ yields

$$\Delta = 1 + \frac{3y}{5} \,, \tag{A.13}$$

to lowest order in y. The linear density *contrast* inside the halo when it has the radius y is therefore

$$\delta = \Delta - 1 = \frac{3y}{5} \,. \tag{A.14}$$

- Linearly extrapolating this to $x = 1$ gives the linear density contrast expected inside the halo at turn-around,

$$\delta_{\rm ta} = \frac{a_{\rm ta}}{a}\delta = \frac{\delta}{x} = \frac{3y}{5x} \,. \tag{A.15}$$

  Now,

$$\frac{1}{x} = \left(\frac{3\tau}{2}\right)^{-2/3} \approx \left(\frac{3\pi}{4}\right)^{2/3} \frac{1}{y} \tag{A.16}$$

  where we have used (A.11) to lowest order in y. Inserting this result into (A.16) yields

$$\delta_{\rm ta} = \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \approx 1.06 \,. \tag{A.17}$$

- When the halo collapses at $x_{\rm c} = 4^{1/3} = 2^{2/3}$, the *linear* density contrast inside the halo would be

$$\delta_{\rm c} = 2^{2/3}\delta_{\rm ta} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \approx 1.69 \,, \tag{A.18}$$

  this means that a halo can be considered collapsed when its density contrast expected from linear theory has reached the value of $\delta_{\rm c}$. This value depends very little on the cosmological parameters, so it can be quite generally used although it was derived for the Einstein-de Sitter model.

- When the halo reaches virial equilibrium, the potential energy of the halo is twice that at turn-around, so virialisation is expected when the radius drops to $y = 1/2$ after turn-around. Assuming virialisation happens at collapse time $x_{\rm c}$, its overdensity is

$$\Delta_{\rm v} = \left(\frac{2^{2/3}}{1/2}\right)^3 \zeta = 32\zeta = 18\pi^2 \approx 178 \tag{A.19}$$

  according to (A.12) and (A.10). A halo in virial equilibrium is thus expected to have a mean density ≈ 178 times higher than the background.

- These two parameters derived from the spherical collapse model, $\delta_{\rm c}$ and $\Delta_{\rm v}$, are very widely used in cosmology for characterising dark-matter halos and their formation.

Index