

# Chapter 8

## The Standard Model of Cosmology and the Dark Universe



*This chapter introduces the observational data on the structure, composition, and evolution of the Universe, within the framework of the theory of general relativity, and describes the model currently providing the best quantitative description. In particular, we will illustrate the experimental evidence suggesting the existence of new forms of matter and energy, and describe the expansion, the chemical evolution, and the formation of structures, from the beginning of time—that, we believe, started with a phase transition from a singularity: the “Big Bang.”*

The origin and fate of the Universe is, for many researchers, the fundamental question. Many answers were provided over the ages, a few of them built over scientific observations and reasoning. During the last century important scientific theoretical and experimental breakthroughs occurred after Einstein’s proposal of the General Theory of Relativity in 1915, with precise and systematic measurements establishing the expansion of the Universe, the existence the cosmic microwave background, and the abundances of light elements in the Universe. The fate of the Universe can be predicted from its energy content—but, although the chemical composition of the Universe and the physical nature of its constituent matter have occupied scientists for centuries, we do not know yet this energy content well enough.

We are made of protons, neutrons, and electrons, combined into atoms in which most of the energy is concentrated in the nuclei (baryonic matter), and we know a few more particles (photons, neutrinos, ...) accounting for a limited fraction of the total energy of atoms. However, the motion of stars in galaxies as well as results about background radiation and the large-scale structure of the Universe (both will be discussed in the rest of this chapter) is inconsistent with the presently known laws of physics, unless we assume that a new form of matter exists. This matter is not

visible, showing little or no interaction with photons—we call it “dark matter”. It is, however, important in the composition of the Universe, because its energy is a factor of five larger than the energy of baryonic matter.

Recently, the composition of the Universe has become even more puzzling, as observations imply an accelerated expansion. Such an acceleration can be explained by a new, unknown, form of energy—we call it “dark energy”—generating a repulsive gravitational force. Something is ripping the Universe apart.

The current view on the distribution of the total budget between these forms of energy is shown in Fig. 1.8. Note that we are facing a new Copernican revolution: we are not made of the same matter that most of the Universe is made of. Moreover, the Universe displays a global behavior difficult to explain, as we shall see in Sect. 8.1.1.

Today, at the beginning of the twenty-first century, the Big Bang model with a large fraction of dark matter (DM) and dark energy is widely accepted as “the standard model of cosmology,” but no one knows what the “dark” part really is, and thus the Universe and its ultimate fate remain basically unknown.

## 8.1 Experimental Cosmology

About one century ago, we believed that the Milky Way was the only galaxy; today, we have a more refined view of the Universe, and the field of experimental cosmology probably grows at a faster rate than any other field in physics. In the last century, we obtained unexpected results about the composition of the Universe, and its global structure.

### 8.1.1 *The Universe Is Expanding*

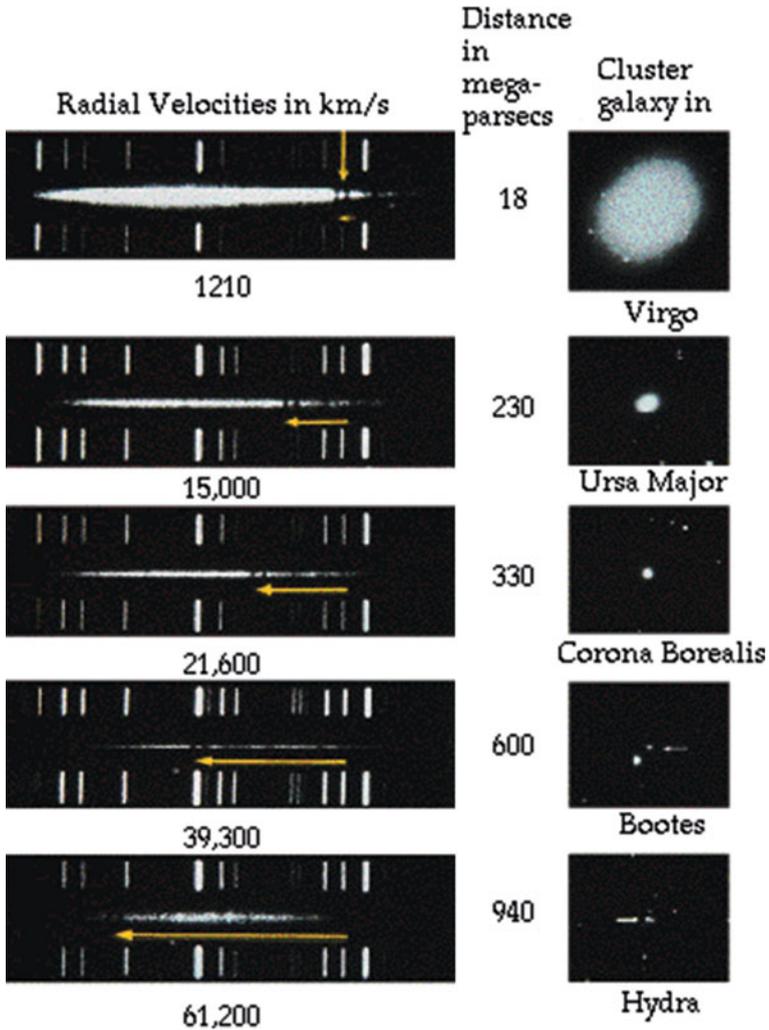
As introduced in Chap. 1, striking evidence that the Universe is expanding comes from the observation that most galaxies are receding in all directions with radial velocities  $v$  proportional to their distance  $d$  from us. This is the famous Hubble law

$$v = H_0 d, \quad (8.1)$$

where  $H_0 \simeq 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$  is the so-called Hubble constant (we shall see that it is not at all constant and can change during the history of the Universe) which is often expressed as a function of a dimensionless parameter  $h$  defined as

$$h = \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}. \quad (8.2)$$

However, velocity and distance are not directly measured. The main observables are the redshift  $z$ —i.e., the fractional wavelength shift observed in specific absorption lines (hydrogen, sodium, magnesium, ...) of the measured spectra of objects (Fig. 8.1)



**Fig. 8.1** Wavelength shifts observed in spectra of galaxies depending on their distance. From J. Silk, “The Big Bang,” Times Books 2000

$$z = \frac{\lambda_{\text{observed}} - \lambda_{\text{emitted}}}{\lambda_{\text{emitted}}} = \frac{\Delta\lambda}{\lambda_{\text{emitted}}}, \tag{8.3}$$

and the apparent luminosity of the celestial objects (stars, galaxies, supernovae, ...), for which we assume we know the intrinsic luminosity.

A redshift occurs whenever  $\Delta\lambda > 0$  which is the case for the large majority of galaxies. There are notable exceptions ( $\Delta\lambda < 0$ , a blueshift) as the one of M31, the

nearby Andromeda galaxy, explained by a large intrinsic velocity (peculiar velocity) oriented toward us.

Wavelength shifts were first observed by the US astronomer James Keeler at the end of the nineteenth century in the spectrum of the light reflected by the rings of Saturn, and later on, at the beginning of twentieth century, by the US astronomer Vesto Slipher, in the spectral lines of several galaxies. In 1925 spectral lines had been measured for around 40 galaxies.

These wavelength shifts were (and still often are) incorrectly identified as simple special relativistic Doppler shifts due to the movement of the sources. In this case  $z$  would be given by

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1, \quad (8.4)$$

which in the limit of small  $\beta$  becomes

$$z \simeq \beta; \quad (8.5)$$

in terms of  $z$  the Hubble law can then be written as:

$$z \simeq \frac{H_0}{c} d. \quad (8.6)$$

However, the limit of small  $\beta$  is not valid for high redshift objects with  $z$  as high as 11 that have been observed in the last years—the list of the most distant object comprises more than 100 objects with  $z > 7$  among galaxies (the most abundant category), black holes, and even stars. On the other hand, high redshift supernovae (typically  $z \sim 0.1$  to 1) have been extensively studied. From these studies an interpretation of the expansion based on special relativity is clearly excluded: one has to invoke general relativity.

In terms of general relativity (see Sect. 8.2) the observed redshift is not due to any movement of the cosmic objects but to the expansion of the proper space between them. This expansion has no center: an observer at any point of the Universe will see the expansion in the same way with all objects in all directions receding with radial velocities given by the same Hubble law and not limited by the speed of light (in fact for  $z \gtrsim 1.5$  radial velocities are, in a large range of cosmological models, higher than  $c$ ): it is the distance scale in the Universe that is changing.

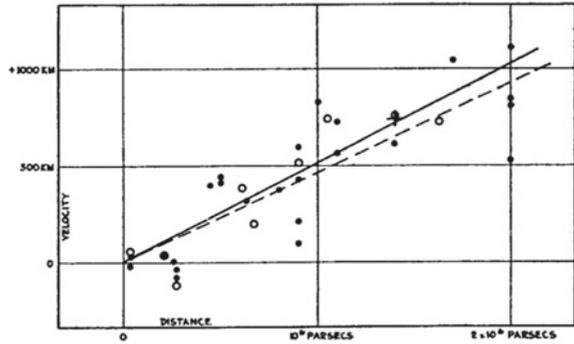
Let us now write the distance between two objects as

$$d = a(t)x, \quad (8.7)$$

where  $a(t)$  is a scale that may change with time and  $x$  by definition is the present ( $t = t_0$ ) distance between the objects ( $a(t_0) = 1$ ) that does not change with time (comoving distance). Then

$$\dot{d} = \dot{a}x; \quad v = H_0 d$$

**Fig. 8.2** The velocity–distance relation measured by Hubble (the original “Hubble plot”). From E. Hubble, Proceedings of the National Academy of Sciences 15 (1929) 168



with

$$H_0 = \left. \frac{\dot{a}(t)}{a(t)} \right|_{t=t_0} . \tag{8.8}$$

In this simple model the Hubble constant is just the expansion rate of the distance scale in the Universe.

Let us come back to the problem of the measurement of distances. The usual method to measure distances is to use reference objects (standard candles), for which the absolute luminosity  $L$  is known. Then, assuming isotropic light emission in an Euclidean Universe (see Sect. 8.2) and measuring the corresponding light flux  $f$  on Earth, the distance  $d$  can be estimated as

$$d = \sqrt{\frac{L}{4\pi f}} . \tag{8.9}$$

In his original plot shown in Fig. 8.2 Hubble used as standard candles Cepheid<sup>1</sup> stars, as well as the brightest stars in the Galaxy, and even entire galaxies (assuming the absolute luminosity of the brightest stars and of the Galaxies to be approximately constant).

The original Hubble result showed a linear correlation between  $v$  and  $d$ , but the slope (the Hubble constant) was wrong by a factor of 7 due to an overall calibration error caused mainly by a systematic underestimate of the absorption of light by dust.

A constant slope would mean that the scale distance  $a(t)$  discussed above would increase linearly with time:

$$a(t) = a(t_0) + \dot{a}(t - t_0) ,$$

---

<sup>1</sup>Cepheids are variable red supergiant stars with pulsing periods strongly correlated with their absolute luminosity. This extremely useful propriety was discovered by the US astronomer Henrietta Leavitt at the beginning of twentieth century and has been used by Hubble to demonstrate in 1924 that the Andromeda Nebula M31 was too far to be part of our own Galaxy, the Milky Way.

i.e.,

$$\frac{a(t)}{a(t_0)} = 1 + H_0(t - t_0). \quad (8.10)$$

Hubble suggested in his original article, under the influence of a model by de Sitter, that this linear behavior could be just a first-order approximation. In fact until recently (1998) most people were convinced that at some point the expansion should be slowed down under the influence of gravity which should be the dominant (attractive) force at large scale. This is why the next term added to the expansion is usually written by introducing a deceleration parameter  $q_0$  (if  $q_0 > 0$  the expansion slows down) defined as

$$q_0 = - \frac{\ddot{a}a}{\dot{a}^2} \Big|_{t=t_0} = - \frac{\ddot{a}}{H_0^2 a} \Big|_{t=t_0}, \quad (8.11)$$

and then

$$\frac{a(t)}{a(t_0)} \simeq 1 + H_0(t - t_0) - \frac{1}{2}q_0H_0^2(t - t_0)^2. \quad (8.12)$$

The relation between  $z$  and  $d$  must now be modified to include this new term.

However, in an expanding Universe the computation of the distance is much more subtle. Various distance measures are usually defined between two objects: in particular, the proper distance  $d_p$  and the luminosity distance  $d_L$ .

- $d_p$  is defined as the length measured on the spatial geodesic connecting the two objects at a fixed time (a geodesic is defined to be a curve whose tangent vectors remain parallel if they are transported along it. Geodesics are (locally) the shortest path between points in space, and describe locally the infinitesimal path of a test inertial particle). It can be shown (see Ref. [F8.2]) that

$$d_p \simeq \frac{c}{H_0} z \left( 1 - \frac{1 + q_0}{2} z \right); \quad (8.13)$$

for small  $z$  the usual linear Hubble law is recovered.

- $d_L$  is defined as the distance that is experimentally determined using a standard candle assuming a static and Euclidean Universe as noted above:

$$d_L = \sqrt{\frac{L}{4\pi f}}. \quad (8.14)$$

The relation between  $d_p$  and  $d_L$  depends on the curvature of the Universe (see Sect. 8.2.3). Even in a flat (Euclidean) Universe (see Sect. 8.2.3 for a formal definition; for the moment, we rely on an intuitive one, and think of flat space as a space in which the sum of the internal angles of a triangle is always  $\pi$ ) the flux of light emitted by an object with a redshift  $z$  and received at Earth is attenuated by a factor  $(1 + z)^2$  due to the dilation of time ( $\gamma \simeq (1 + z)$ ) and the increase of the photon's wavelength ( $a^{-1} = (1 + z)$ ). Then if the Universe was basically flat

$$d_L = d_p (1 + z) \simeq \frac{c}{H_0} z \left[ 1 + \frac{1 - q_0}{2} z \right]. \quad (8.15)$$

To experimentally determine  $q_0$  one needs to extend the range of distances in the Hubble plot by a large amount. New and brighter standard candles are needed.

### 8.1.2 Expansion Is Accelerating

Type Ia supernovae have been revealed themselves as an optimal option to extend the range of distances in the Hubble plot. Supernovae Ia occur whenever, in a binary system formed by a white dwarf (a compact Earth-size stellar endproduct of mass close to the solar mass) and another star (for instance a red giant, a luminous giant star in a late phase of stellar evolution), the white dwarf accretes matter from its companion reaching a total critical mass of about 1.4 solar masses. At this point a nuclear fusion reaction starts, leading to a gigantic explosion (with a luminosity about  $10^5$  times larger than the brightest Cepheids; see Fig. 8.3 for an artistic representation).

The results obtained by the “Supernova Cosmology Project” and by the “High- $z$  Supernova Search Team” resulted in extended Hubble plots (Fig. 8.4) that were a surprise and triggered a revolution in our understanding of the content and evolution of the Universe.<sup>2</sup> The striking point is that the fit to the experimental supernova ( $z$ ,  $d$ ) data leads to negative values of  $q_0$  meaning that, contrary to what was expected, the expansion of the Universe is nowadays accelerating.

An alternative method to the use of standard candles to determine extragalactic distances is the use of “standard rulers”. Let us suppose that we know the absolute length  $l$  of an object (the standard ruler) that is placed at some distance transversally to the observation line. Then the distance of the object can be obtained from its angular size  $\delta\theta$  by the simple formula:

$$d_A = \frac{l}{\delta\theta}, \quad (8.16)$$

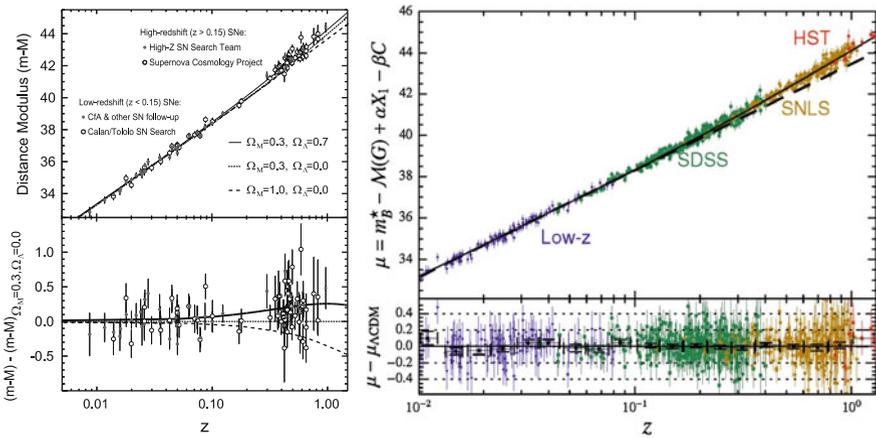
where  $d_A$  is known as the angular diameter distance. In a curved and/or expanding Universe  $d_A$  does not coincide with the proper ( $d_p$ ) and the luminosity ( $d_L$ ) distances defined above but it can be shown (see Ref. [8.2]) that:

---

<sup>2</sup>The Supernova Cosmology Project is a collaboration, led by Saul Perlmutter, dedicated to the study of distant supernovae of type Ia, that started collecting data in 1988. Another collaboration also searching for distant supernovae of type Ia was formed by Brian Schmidt and Adam Riess in 1994, the High- $z$  Supernova Search Team. These teams found over 50 distant supernovae of type Ia for which the light received was weaker than expected—which implied that the rate of expansion of the Universe was increasing. Saul Perlmutter, born in 1959 in Champaign–Urbana, IL, USA, Ph.D. from University of California, Berkeley; Brian P. Schmidt, USA and Australian citizen, born in 1967 in USA, Ph.D. from Harvard; Adam G. Riess, born in 1969 in Washington, DC, USA, Ph.D. from Harvard, all professors in the USA, were awarded the 2011 Nobel Prize in Physics “for the discovery of the accelerating expansion of the Universe through observations of distant supernovae.”



**Fig. 8.3** Artistic representation of the formation and explosion of a supernova Ia (Image from A. Hardy, David A. Hardy/[www.astroart.org](http://www.astroart.org))



**Fig. 8.4** Left: The “Hubble plot” obtained by the “High-z Supernova Search Team” and by the “Supernova Cosmology Project.” The lines represent the prediction of several models with different energy contents of the Universe (see Sect. 8.4). The best fit corresponds to an accelerating expansion scenario. From “Measuring Cosmology with Supernovae,” by Saul Perlmutter and Brian P. Schmidt; Lecture Notes in Physics 2003, Springer. Right: an updated version by the “Supernova Legacy Survey” and the “Sloan Digital Sky Survey” projects, M. Betoule et al. arXiv:1401.4064

$$d_A = \frac{d_L}{(1+z)^2} . \tag{8.17}$$

Several candidates for standard rulers have been discussed in the last years and, in particular, the observation of Baryon Acoustic Oscillations (BAO) opened a new and promising path. BAO use the Fourier transform of the distance correlation function between specific astrophysics objects (for instance luminous red galaxies, blue galaxies) to discover, as function of the redshift  $z$ , the clustering scales of the baryonic matter. These scales are related to the evolution of initial density perturbations in the early Universe (see Sect. 8.3). The correlation function  $\xi$  between pairs of

galaxies is just the excess probability that the two galaxies are at the distance  $r$  and thus a sharp peak in  $\xi(r)$  will correspond in its Fourier transform to an oscillation spectrum with a well-defined frequency.

### 8.1.2.1 Dark Energy

There is no classical explanation for the accelerated expansion of the Universe. A new form of energy is invoked, permeating the space and exerting a negative pressure. This kind of energy can be described in the general theory of relativity (see later) and associated, e.g., to a “cosmological constant” term  $\Lambda$ ; from a physical point of view, it corresponds to a “dark” energy component—and to the present knowledge has the largest energy share in the Universe.

In Sect. 8.4 the current overall best picture able to accommodate all present experimental results (the so-called  $\Lambda$ CDM model) will be discussed.

## 8.1.3 Cosmic Microwave Background

In 1965 Penzias and Wilson,<sup>3</sup> two radio astronomers working at Bell Laboratories in New Jersey, discovered by accident that the Universe is filled with a mysterious isotropic and constant microwave radiation corresponding to a blackbody temperature around 3 K.

Penzias and Wilson were just measuring a small fraction of the blackbody spectrum. Indeed they were measuring the region in the tail around wavelength  $\lambda \sim 7.5$  cm while the spectrum peaks around  $\lambda \sim 2$  mm. To fully measure the density spectrum it is necessary to go above the Earth’s atmosphere, which absorbs wavelengths lower than  $\lambda \sim 3$  cm. These measurements were eventually performed in several balloon and satellite experiments. In particular, the Cosmic Background Explorer (COBE), launched in 1989, was the first to show that in the 0.1 to 5 mm range the spectrum, after correction for the proper motion of the Earth, is well described by the Planck blackbody formula

$$\varepsilon_\gamma(\nu) d\nu = \frac{8\pi h}{c^3} \frac{\nu^3 d\nu}{e^{\frac{h\nu}{k_B T}} - 1}, \quad (8.18)$$

where  $k_B$  is the Boltzmann constant. Other measurements at longer wavelengths confirmed that the cosmic microwave background (CMB) spectrum is well described by the spectrum of a single temperature blackbody (Fig. 8.5) with a mean temperature of

---

<sup>3</sup>Arno Penzias (1933–) was born in Munich, Germany. In 1939 his family was rounded up for deportation, but they managed to escape to the USA, where he could graduate in Physics at Columbia University. Robert Wilson (1936–) grew up in Huston, Texas, and studied at Caltech. They shared the 1978 Nobel prize in Physics “for their discovery of the cosmic microwave background radiation.”

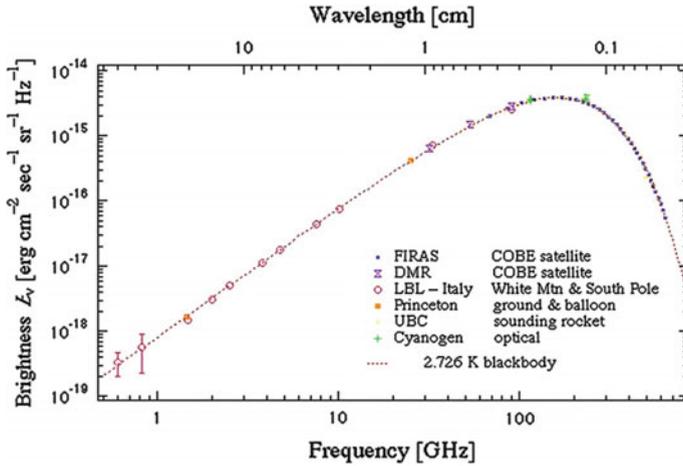


Fig. 8.5 The CMB intensity plot as measured by COBE and other experiments (from [http://aether.lbl.gov/www/projects/cobe/CMB\\_intensity.gif](http://aether.lbl.gov/www/projects/cobe/CMB_intensity.gif))

$$T = (2.726 \pm 0.001) \text{ K} .$$

The total photon energy density is then obtained by integrating the Planck formula over the entire frequency range, resulting in the Stefan–Boltzmann law

$$\epsilon_\gamma = \frac{\pi^2 (k_B T)^4}{15 (\hbar c)^3} \simeq 0.26 \text{ eV cm}^{-3} ; \tag{8.19}$$

moreover, the number density of photons is given by

$$n_\gamma \simeq \frac{2.4}{\pi^2} \left( \frac{k_B T}{\hbar c} \right)^3 \simeq 410 \text{ cm}^{-3} . \tag{8.20}$$

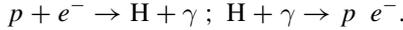
The existence of CMB had been predicted in the 1940s by George Gamow, Robert Dicke, Ralph Alpher, and Robert Herman in the framework of the Big Bang model.

### 8.1.3.1 Recombination and Decoupling

In the Big Bang model the expanding Universe cools down going through successive stages of lower energy density (temperature) and more complex structures. Radiation materializes into pairs of particles and antiparticles, which, in turn, give origin to the existing objects and structures in the Universe (nuclei, atoms, planets, stars, galaxies, ...). In this context, the CMB is the electromagnetic radiation left over when electrons and protons combine to form neutral atoms (the, so-called, *recombination*

phase). After this stage, the absence of charged matter allows photons to be basically free of interactions, and evolve independently in the expanding Universe (*photon decoupling*).

In a simple, but reasonable, approximation (neglecting heavier elements, in particular helium) recombination occurs as the result of the balance between the formation and the photodisintegration of hydrogen atoms:



If these reactions are in equilibrium at a given temperature  $T$  (high enough to allow the photodisintegration and low enough to consider  $e$ ,  $p$ ,  $\text{H}$  as nonrelativistic particles) the number density of electrons, protons, and hydrogen atoms may be approximated by the Maxwell–Boltzmann distribution (see Sect. 8.3.1)

$$n_x = g_x \left( \frac{m_x k_B T}{2\pi \hbar^2} \right)^{\frac{3}{2}} e^{-\frac{m_x c^2}{k_B T}} , \quad (8.21)$$

where  $g_x$  is a statistical factor accounting for the spin (the subscript  $x$  refers to each particle type).

The ratio  $n_{\text{H}} / (n_p n_e)$  can then be approximately modeled by the Saha equation

$$\frac{n_{\text{H}}}{n_p n_e} \simeq \left( \frac{m_e k_B T}{2\pi \hbar^2} \right)^{-\frac{3}{2}} e^{\frac{Q}{k_B T}} , \quad (8.22)$$

where

$$Q = (m_p + m_e - m_{\text{H}}) c^2 \simeq 13.6 \text{ eV} \quad (8.23)$$

is the hydrogen binding energy.

Defining  $X$  as the fractional ionization ( $X = 1$  for complete ionization, whereas,  $X = 0$  when all protons are inside neutral atoms),

$$X = \frac{n_p}{n_p + n_{\text{H}}} , \quad (8.24)$$

and assuming that there is zero total net charge, ( $n_p = n_e$ ), the Saha equation can be rewritten as

$$\frac{1 - X}{X} \simeq n_p \left( \frac{m_e k_B T}{2\pi \hbar^2} \right)^{-\frac{3}{2}} e^{\left( \frac{Q}{k_B T} \right)} . \quad (8.25)$$

On the other hand at thermal equilibrium, the energy density of photons as a function of the frequency  $\nu$  follows the usual blackbody distribution corresponding, as we have seen before, to a photon density number of:

$$n_\gamma \simeq \frac{2.4}{\pi^2} \left( \frac{k_B T}{\hbar c} \right)^3. \quad (8.26)$$

For the typical photodisintegration temperatures ( $k_B T \sim 13.6 \text{ eV}$ )

$$n_\gamma \gg n_B,$$

where  $n_B$  is the total number of baryons, which in this simple approximation is defined as

$$n_B = n_p + n_H = \frac{n_p}{X}. \quad (8.27)$$

The baryon to photon ratio is thus

$$\eta = \frac{n_B}{n_\gamma} = \frac{n_p}{X n_\gamma} \ll 1. \quad (8.28)$$

After decoupling,  $n_B$  and  $n_\gamma$  evolve independently both as  $a(t)^{-3}$ , where  $a(t)$  is the scale factor of the Universe, see Sect. 8.1.1. Thus,  $\eta$  is basically a constant, which can be measured at the present time through the measurement of the content of light elements in the Universe (see Sect. 8.1.4):

$$\eta \sim (5 - 6) \times 10^{-10}. \quad (8.29)$$

The Saha equation can then be written as a function of  $\eta$  and  $T$ , and used to determine the recombination temperature (assuming  $X \sim 0.5$ ):

$$\frac{1 - X}{X^2} = 2 \simeq 3.84 \eta \left( \frac{k_B T}{m_e c^2} \right)^{\frac{3}{2}} e^{\frac{Q}{k_B T}}. \quad (8.30)$$

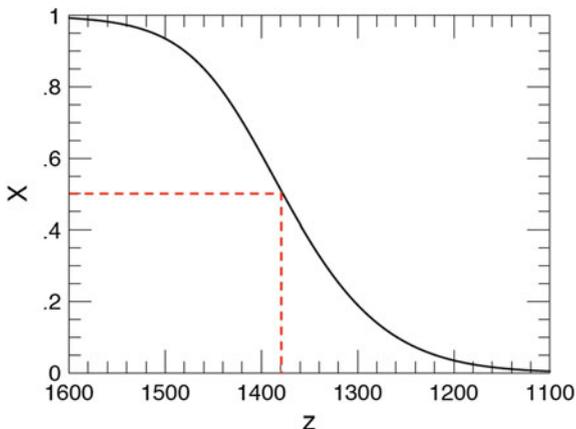
The solution of this equation gives a remarkably stable value of temperature for a wide range of  $\eta$ . For instance  $\eta \sim 5.5 \times 10^{-10}$  results into

$$k_B T_{\text{rec}} \simeq 0.323 \text{ eV} \implies T_{\text{rec}} \simeq 3740 \text{ K}. \quad (8.31)$$

This temperature is much higher than the measured CMB temperature reported above. The difference is attributed to the expansion of the Universe between the recombination epoch and the present. Indeed, as discussed in Sect. 8.1.1, the photon wavelength increases during the expansion of a flat Universe by a factor  $(1 + z)$ . The entire CMB spectrum was expanded by this factor, and then it can be estimated that recombination had occurred (see Fig. 8.6) at

$$z_{\text{rec}} \sim 1300 - 1400. \quad (8.32)$$

**Fig. 8.6**  $X$  as a function of  $z$  in the Saha equation. Time on the abscissa increases from left to right (as  $z$  decreases). Adapted from B. Ryden, lectures at ICTP Trieste, 2006



After recombination the Universe became substantially transparent to photons. The photon decoupling time is defined as the moment when the interaction rate of photons  $\Gamma_{\gamma_{\text{scat}}}$  equals the expansion rate of the Universe (which is given by the Hubble parameter)

$$\Gamma_{\gamma_{\text{scat}}} \sim H. \quad (8.33)$$

The dominant interaction process is the photon–electron Thomson scattering. Then

$$\Gamma_{\gamma_{\text{scat}}} \simeq n_e \sigma_T c,$$

where  $n_e$  and  $\sigma_T$  are, respectively, the free electron density number and the Thomson cross section.

Finally, as  $n_e$  can be related to the fractional ionization  $X(z)$  and the baryon density number ( $n_e = X(z) n_B$ ), the redshift at which the photon decoupling occurs ( $z_{\text{dec}}$ ) is given by

$$X(z_{\text{dec}}) n_B \sigma_T c \sim H. \quad (8.34)$$

However, the precise computation of  $z_{\text{dec}}$  is subtle. Both  $n_B$  and  $H$  evolve during the expansion (for instance in a matter-dominated flat Universe, as it will be discussed in Sect. 8.2,  $n_B(z) \propto n_{B,0}(1+z)^3$  and  $H(z) \propto H_0(1+z)^{3/2}$ ). Furthermore, the Saha equation is not valid after recombination since electrons and photons are no longer in thermal equilibrium. The exact value of  $z_{\text{dec}}$  depends thus on the specific model for the evolution of the Universe and the final result is of the order of

$$z_{\text{dec}} \sim 1100. \quad (8.35)$$

After decoupling the probability of a further scattering is extremely low except at the so-called *reionization* epoch. After the formation of the first stars, there was a period ( $6 < z < 20$ ) when the Universe was still small enough for neutral hydrogen

formed at recombination to be ionized by the radiation emitted by stars. Still, the scattering probability of CMB photons during this epoch is small. To account for it, the reionization optical depth parameter  $\tau$  is introduced, in terms of which the scattering probability is given by

$$P \sim 1 - e^{-\tau}.$$

The CMB photons follow then spacetime geodesics until they reach us. These geodesics are slightly distorted by the gravitational effects of the mass fluctuations close to the path giving rise to microlensing effects, which are responsible for a typical total deflection of  $\sim 2$  arcminutes.

The spacetime points to where the last scattering occurred thereby define with respect to any observer a region called the *last scattering surface* situated at a redshift,  $z_{lss}$ , very close to  $z_{dec}$

$$z_{lss} \sim z_{dec} \sim 1100. \quad (8.36)$$

Beyond  $z_{lss}$  the Universe is opaque to photons and to be able to observe it other messengers, e.g., gravitational waves, have to be studied. On the other hand, the measurement of the primordial nucleosynthesis (Sect. 8.1.4) allows us to indirectly test the Big Bang model at times well before the recombination epoch.

### 8.1.3.2 Temperature Fluctuations

The COBE satellite<sup>4</sup> measured the temperature fluctuations in sky regions centered at different points with Galactic coordinates  $(\theta, \Phi)$

$$\frac{\delta T(\theta, \Phi)}{\langle T \rangle} = \frac{T(\theta, \Phi) - \langle T \rangle}{\langle T \rangle} \quad (8.37)$$

---

<sup>4</sup>Three satellite missions have been launched so far to study the cosmic background radiation. The first was COBE in 1989, followed by Wilkinson Microwave Anisotropy Probe (WMAP) in 2001, both of which were NASA missions. The latest (with the best angular resolution and sensitivity), called Planck, has been launched by the European Space Agency (ESA) with a contribution from NASA in 2009. In terms of sensitivity and angular resolution, WMAP improved COBE by a factor of 40, and Planck gave a further improvement by a factor of 4; in addition Planck measures polarization. The instruments onboard Planck are a low-frequency (solid state) instrument from 30 GHz, and a bolometer—a device for measuring the power of incident electromagnetic radiation via the heating of a material with a temperature-dependent electrical resistance—for higher frequencies (up to 900 GHz). The total weight of the payload is 2 tons (it is thus classified as a large mission); it needs to be kept at cryostatic temperatures. John Mather, from the Goddard Space Flight Center, and George Smoot, at the University of California, Berkeley, shared the 2006 Nobel Prize in Physics “for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation.”

and found that, apart from a dipole anisotropy of the order of  $10^{-3}$ , the temperature fluctuations are of the order of  $10^{-5}$ : the observed CMB spectrum is remarkably isotropic.

The dipole distortion (a slight blueshift in one direction of the sky and a redshift in the opposite direction—Fig. 8.7) observed in the measured average temperature can be attributed to a global Doppler shift due to the peculiar motion (COBE, Earth, Solar system, Milky Way, Local Group, Virgo cluster, ...) with respect to a hypothetical CMB isotropic reference frame characterized by a temperature  $T$ . Indeed

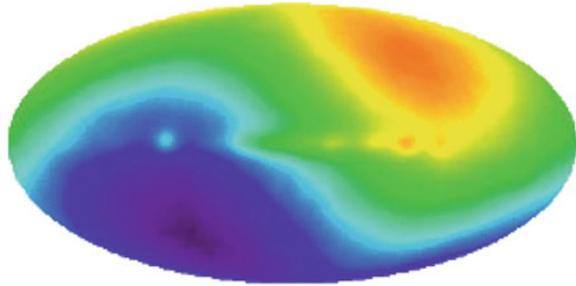
$$T^* = T \left( 1 + \frac{v}{c} \cos \theta \right) \quad (8.38)$$

with

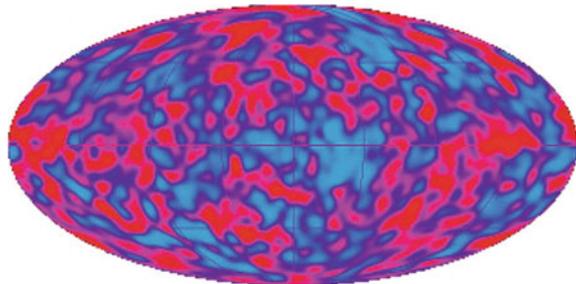
$$v = (371 \pm 1) \text{ km/s.}$$

After removing this effect, the remaining fluctuations reveal a pattern of tiny inhomogeneities at the level of the last scattering surface. The original picture from COBE (Fig. 8.8), with an angular resolution of  $7^\circ$ , was confirmed and greatly improved by the Wilkinson Microwave Anisotropy Probe WMAP, which obtained full sky maps with a  $0.2^\circ$  angular resolution. The Planck satellite delivered more recently sky maps with three times improved resolution and ten times higher sensitivity (Fig. 8.9), also covering a larger frequency range.

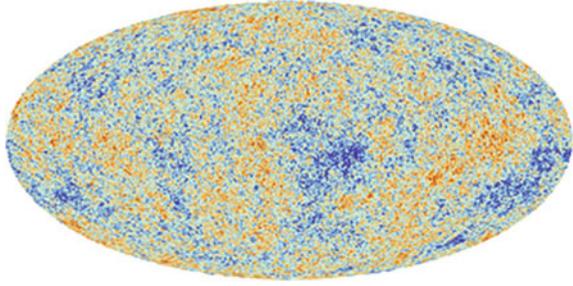
**Fig. 8.7** Sky map (in Galactic coordinates) of CMB temperatures measured by COBE after the subtraction of the emission from our Galaxy. A dipole component is clearly visible. (from <http://apod.nasa.gov/apod/ap010128.html>)



**Fig. 8.8** CMB temperature fluctuations sky map as measured by COBE after the subtraction of the dipole component and of the emission from our Galaxy. <http://lambda.gsfc.nasa.gov/product/cobe>



**Fig. 8.9** CMB temperature fluctuations sky map as measured by the Planck mission after the subtraction of the dipole component and of the emission from our Galaxy. <http://www.esa.int/spaceinimages>



Once these maps are obtained it is possible to establish two-point correlations between any two spatial directions.

Technically, the temperature fluctuations are expanded using spherical harmonics

$$\frac{\delta T}{\langle T \rangle}(\theta, \Phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} Y_{lm}^*(\theta, \Phi), \quad (8.39)$$

with

$$a_{lm} = \int_{\theta=0}^{\pi} \int_{\Phi=0}^{2\pi} \frac{\delta T}{\langle T \rangle}(\theta, \Phi) Y_{lm}^*(\theta, \Phi) d\Omega. \quad (8.40)$$

Then the correlation between two directions  $\hat{n}$  and  $\hat{n}^*$  separated by an angle  $\alpha$  is defined as

$$C(\alpha) = \left\langle \frac{\delta T}{\langle T \rangle}(\hat{n}) \frac{\delta T}{\langle T \rangle}(\hat{n}^*) \right\rangle_{\hat{n} \cdot \hat{n}^* = \cos \alpha}$$

and can be expressed as

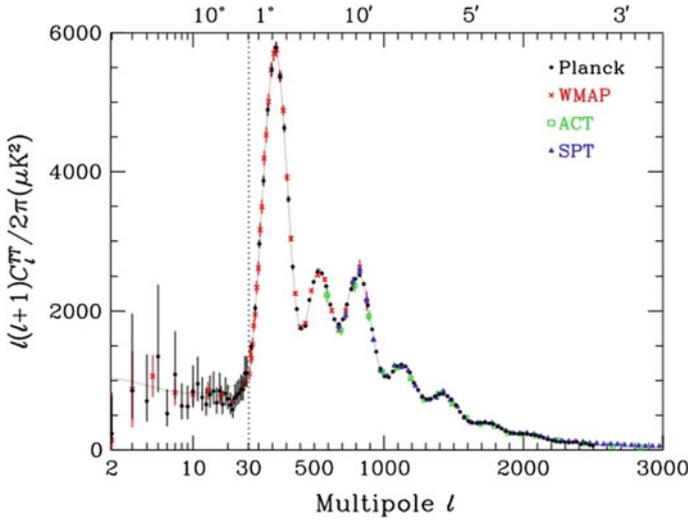
$$C(\alpha) = \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) C_l P_l(\cos \alpha),$$

where  $P_l$  are the Legendre polynomials and the  $C_l$ , the multipole moments, are given by the variance of the harmonic coefficients  $a_{lm}$ :

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^l \langle |a_{lm}|^2 \rangle. \quad (8.41)$$

Each multipole moment corresponds to a sort of angular frequency  $l$ , whose conjugate variable is an angular scale  $\alpha$  such that

$$\alpha = \frac{180^\circ}{l}. \quad (8.42)$$



**Fig. 8.10** Temperature power spectrum from the Planck, WMAP, ACT, and SPT experiments. The abscissa is logarithmic for  $l$  less than 30, linear otherwise. The curve is the best-fit Planck model. From C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)

The total temperature fluctuations (temperature power spectrum) can be then expressed as a function of the multipole moment  $l$  (Fig. 8.10, top)

$$\langle \Delta T^2 \rangle = \left( \frac{l(l+1)}{2\pi} C_l \right) \langle T \rangle^2. \tag{8.43}$$

Such a function shows a characteristic pattern with a first peak around  $l \sim 200$  followed by several smaller peaks.

The first peak at an angular scale of  $1^\circ$  defines the size of the “sound horizon” at the time of last scattering (see Sect. 8.1.4) and the other peaks (acoustic peaks) are extremely sensitive to the specific contents and evolution model of the Universe at that time. The observation of very tiny fluctuations at large scales (much greater than the horizon,  $l \ll 200$ ) leads to the hypothesis that the Universe, to be casually connected, went through a very early stage of exponential expansion, called *inflation*.

Anisotropies can also be found studying the polarization of CMB photons. Indeed at the recombination and reionization epochs the CMB may be partially polarized by Thomson scattering with electrons. It can be shown that linear polarization may be originated by quadrupole temperature anisotropies. In general the polarization pattern is decomposed in two orthogonal modes, respectively, called B-mode (curl-like) and E-mode (gradient-like). The E-mode comes from density fluctuations, while primordial gravitational waves are expected to display both polarization modes. Gravitational lensing of the CMB E-modes may also be a source of B-modes. E-modes were first measured in 2002 by the DASI telescope in Antarctica and later

on the Planck collaboration published high-resolution maps of the CMB polarization over the full sky. The detection and the interpretation of B-modes are very challenging since the signals are tiny and foreground contaminations, as the emission by Galactic dust, are not always easy to estimate. The arrival angles of CMB photons are smeared, due to the microlensing effects, by dispersions that are function of the integrated mass distribution along the photon paths. It is possible, however, to deduce these dispersions statistically from the observed temperature angular power spectra and/or from polarized E- and B-mode fields. The precise measurement of these dispersions will give valuable information for the determination of the cosmological parameters. It will also help constraining parameters, such as the sum of the neutrino masses or the dark energy content, that are relevant for the growth of structures in the Universe, and evaluating contributions to the B-mode patterns from possible primordial gravity waves.

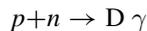
The detection of gravitational lensing was reported by several experiments such as the Atacama Cosmology Telescope, the South Pole Telescope, and the POLARBEAR experiment. The Planck collaboration has measured its effect with high significance using temperature and polarization data, establishing a map of the lensing potential.

Some of these aspects will be discussed briefly in Sect. 8.3, but a detailed discussion of the theoretical and experimental aspects of this fast-moving field is far beyond the scope of this book.

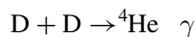
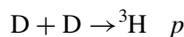
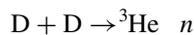
### 8.1.4 Primordial Nucleosynthesis

The measurement of the abundances of light elements in the Universe (H, D,  $^3\text{He}$ ,  $^4\text{He}$ ,  $^6\text{Li}$ ,  $^7\text{Li}$ ) is the third observational “pillar” of the Big Bang model, after the Hubble expansion and the CMB. As it was proposed, and first computed, by the Russian American physicists Ralph Alpher and George Gamow in 1948, the expanding Universe cools down, and when it reaches temperatures of the order of the nuclei-binding energies per nucleon ( $\sim 1-10$  MeV) nucleosynthesis occurs if there are enough protons and neutrons available. The main nuclear fusion reactions are

- proton–neutron fusion:

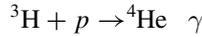
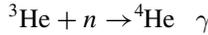
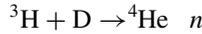


- deuterium–deuterium fusion:

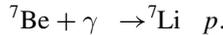
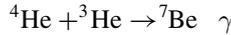
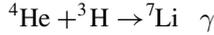
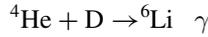


- other  $^4\text{He}$  formation reactions:





- and finally the lithium and beryllium formation reactions (there are no stable nuclei with  $A = 5$ ):



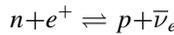
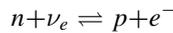
The absence of stable nuclei with  $A = 8$  basically stops the primordial Big Bang nucleosynthesis chain. Heavier nuclei are produced in stellar (up to Fe), or supernova nucleosynthesis.<sup>5</sup>

The relative abundance of neutrons and protons, in case of thermal equilibrium at a temperature  $T$ , is fixed by the ratio of the usual Maxwell–Boltzmann distributions (similarly to what was discussed for the recombination—Sect. 8.1.3):

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{\frac{3}{2}} \exp\left(-\frac{(m_n - m_p)c^2}{kT}\right). \quad (8.44)$$

If  $k_B T \gg (m_n - m_p)c^2 \rightarrow n_n/n_p \sim 1$ ; if  $k_B T \ll (m_n - m_p)c^2 \rightarrow n_n/n_p \sim 0$ .

Thermal equilibrium is established through the weak processes connecting protons and neutrons:



as long as the interaction rate of these reactions  $\Gamma_{n,p}$  is greater than the expansion rate of the Universe,

$$\Gamma_{n,p} \geq H.$$

$\Gamma$  and  $H$  diminish during the expansion, the former much faster than the latter. Indeed in a flat Universe dominated by radiation (Sect. 8.2)

$$\Gamma_{n,p} \sim G_F T^5, \quad (8.45)$$

---

<sup>5</sup>Iron ( ${}^{56}\text{Fe}$ ) is the stable element for which the binding energy per nucleon is largest (about 8.8 MeV); it is thus the natural endpoint of fusion processes of lighter elements, and of fission of heavier elements.

$$H \sim \sqrt{g^*} T^2, \quad (8.46)$$

where  $G_F$  is the Fermi weak interaction constant and  $g^*$  the number of degrees of freedom that depends on the relativistic particles content of the Universe (namely on the number of generations of light neutrinos  $n_\mu$ , which, in turn, allows to set a limit on  $n_\mu$ ).

The exact calculation of the freeze-out temperature  $T_f$  at which

$$\Gamma_{n,p} \sim H$$

is out of the scope of this book. The values obtained for  $T_f$  are a little below the MeV scale:

$$k_B T_f \sim 0.8 \text{ MeV}. \quad (8.47)$$

At this temperature

$$\frac{n_n}{n_p} \sim 0.2.$$

After the freeze-out this ratio would remain constant if neutrons were stable. However, as we know, neutrons decay via beta decay,

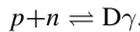
$$n \rightarrow p e^- \bar{\nu}_e.$$

Therefore, the  $n_n/n_p$  ratio will decrease slowly while all the neutrons will not be bound inside nuclei, so that

$$\frac{n_n}{n_p} \sim 0.2 e^{-t/\tau_n} \quad (8.48)$$

where  $\tau_n \simeq 885.7$  s is the neutron lifetime.

The first step of the primordial nucleosynthesis is, as we have seen, the formation of deuterium via proton–neutron fusion



Although the deuterium binding energy, 2.22 MeV, is higher than the freeze-out temperature, the fact that the baryons to photons ratio  $\eta$  is quite small ( $\eta \sim (5 - 6) \times 10^{-10}$ ) makes photodissociation of the deuterium nuclei possible at temperatures lower than the blackbody peak temperature  $T_f$  (the Planck function has a long tail). The relative number of free protons, free neutrons, and deuterium nuclei can be expressed, using a Saha-like equation (Sect. 8.1.3), as follows:

$$\frac{n_D}{n_p n_n} \simeq \frac{g_D}{g_p g_n} \left( \frac{m_D}{m_p m_n} \right)^{\frac{3}{2}} \left( \frac{k_B T}{2\pi \hbar^2} \right)^{-\frac{3}{2}} e^{\frac{Q}{k_B T}}, \quad (8.49)$$

where  $Q$  is now given by

$$Q = (m_p + m_n - m_D) c^2 \sim 2.22 \text{ MeV}.$$

Expressing  $n_p$  as a function of  $\eta$  and  $n_\gamma$  and performing an order of magnitude estimation, we obtain

$$\frac{n_D}{n_n} \propto \eta n_\gamma \left( \frac{m_p c^2 k_B T}{\pi \hbar^2} \right)^{-\frac{3}{2}} e^{\frac{Q}{k_B T}}. \quad (8.50)$$

Replacing now  $n_\gamma$  by the Planck distribution

$$\frac{n_D}{n_n} \propto \eta \left( \frac{k_B T}{m_p c^2} \right)^{\frac{3}{2}} e^{\frac{Q}{k_B T}}. \quad (8.51)$$

This is analogous to the formulation of the Saha equation used to determine the recombination temperature (Sect. 8.1.3). As we have shown its solution (for instance for  $(n_D/n_n) \sim 1$ ) gives a remarkably stable value of temperature. In fact there is a sharp transition around  $k_B T_D \sim 0.1 \text{ MeV}$ : above this value neutrons and protons are basically free; below this value all neutrons are substantially bound first inside D nuclei and finally inside  ${}^4\text{He}$  nuclei, provided that there is enough time before the fusion rate of nuclei becomes smaller than the expansion rate of the Universe. Indeed, since the  ${}^4\text{He}$  binding energy per nucleon is much higher than those of D,  ${}^3\text{H}$ , and  ${}^3\text{He}$ , and since there are no stable nuclei with  $A = 5$ , then  ${}^4\text{He}$  is the favorite final state.

The primordial abundance of  ${}^4\text{He}$ ,  $Y_p$ , is defined usually as the fraction of mass density of  ${}^4\text{He}$  nuclei,  $\rho({}^4\text{He})$ , over the total baryonic mass density,  $\rho$  (Baryons)

$$Y_p = \frac{\rho({}^4\text{He})}{\rho(\text{Baryons})}. \quad (8.52)$$

In a crude way let us assume that after nucleosynthesis all baryons are H or  ${}^4\text{He}$ , i.e., that

$$\rho(\text{H}) + \rho({}^4\text{He}) \simeq 1.$$

Thus

$$Y_p = 1 - \frac{\rho(\text{H})}{\rho(\text{Baryons})} = 1 - \frac{n_p - n_n}{n_p + n_n} = \frac{2 \frac{n_n}{n_p}}{1 + \frac{n_n}{n_p}}. \quad (8.53)$$

For  $(n_n/n_p) \sim 0.2$ ,  $Y_p = 0.33$ .

In fact due to the decay of neutrons between  $k_B T_f \sim 0.8 \text{ MeV}$  and  $k_B T_D \sim 0.1 \text{ MeV}$

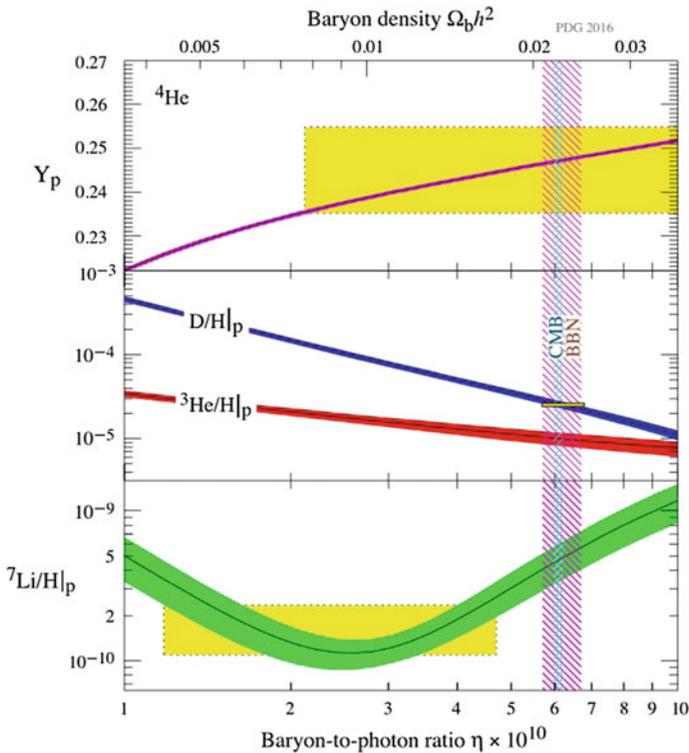
$$\frac{n_n}{n_p} \sim 0.13 - 0.15$$

and the best estimate for  $Y_p$  is in the range

$$Y_p \sim 0.23 - 0.26. \quad (8.54)$$

Around one-quarter of the primordial baryonic mass of the Universe is due to  ${}^4\text{He}$  and around three quarters is made of hydrogen. There are however small fractions of D,  ${}^3\text{He}$ , and  ${}^3\text{H}$  that did not turn into  ${}^4\text{He}$ , and there are, thus, tiny fractions of  ${}^7\text{Li}$  and  ${}^7\text{Be}$  that could have formed after the production of  ${}^4\text{He}$  and before the dilution of the nuclei due to the expansion of the Universe. Although their abundances are quantitatively quite small, the comparison of the expected and measured ratios are important because they are rather sensitive to the ratio of baryons to photons,  $\eta$ .

In Fig. 8.11 the predicted abundances of  ${}^4\text{He}$ , D,  ${}^3\text{He}$ , and  ${}^7\text{Li}$  computed in the framework of the standard model of Big Bang nucleosynthesis as a function of  $\eta$  are compared with measurements (for details see the Particle Data Book). An increase in  $\eta$  will increase slightly the deuterium formation temperature  $T_D$  (there are less  $\gamma$  per baryon available for the photodissociation of the deuterium), and therefore, there is



**Fig. 8.11** The observed and predicted abundances of  ${}^4\text{He}$ , D,  ${}^3\text{He}$ , and  ${}^7\text{Li}$ . The bands show the 95% CL range. Boxes represent the measured abundances. The narrow vertical band represents the constraints at 95% CL on  $\eta$  (expressed in units of  $10^{10}$ ) from the CMB power spectrum analysis while the wider is the Big Bang nucleosynthesis concordance range. From C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016)

more time for the development of the chain fusion processes ending at the formation of the  ${}^4\text{He}$ . Therefore, the fraction of  ${}^4\text{He}$  will increase slightly, in relative terms, and the fraction of D and  ${}^3\text{He}$  will decrease much more significantly, again in relative terms. The evolution of the fraction of  ${}^7\text{Li}$  is, on the contrary, not monotonous; it shows a minimum due to the fact that it is built up from two processes that have a different behavior (the fusion of  ${}^4\text{He}$  and  ${}^3\text{H}$  is a decreasing function of  $\eta$ ; the production via  ${}^7\text{Be}$  is an increasing function of  $\eta$ ).

Apart from the measured value for the fraction of  ${}^7\text{Li}$  all the other measurements converge to a common value of  $\eta$  that is, within the uncertainties, compatible with the value indirectly determined by the study of the acoustic peaks in the CMB power spectrum (see Sect. 8.4).

### 8.1.5 Astrophysical Evidence for Dark Matter

Evidence that the Newtonian physics applied to visible matter does not describe the dynamics of stars, galaxies, and galaxy clusters were well established in the twentieth century.

As a first approximation, one can estimate the mass of a galaxy based on its brightness: brighter galaxies contain more stars than dimmer galaxies. However, there are other ways to assess the total mass of a galaxy. In spiral galaxies, for example, stars rotate in quasi-circular orbits around the center. The rotational speed of peripheral stars depends, according to Newton's law, on the total mass of the galaxy, and one has thus an independent measurement of this mass. Do these two methods give consistent results?

In 1933 the Swiss astronomer Fritz Zwicky applied for the first time the virial theorem to the Coma cluster of galaxies<sup>6</sup>; his choice was motivated by the fact that Coma is a regular and nearly spherical well-studied cluster. We recall that the virial theorem states that, for a stationary self-gravitating system, twice its total kinetic energy  $K$  plus its potential energy  $U$  vanishes. Explicitly, denoting by  $v$  the total velocity of a galaxy in the cluster, we have  $K = M v^2/2$  and for a spherical system  $U = -\alpha G M^2/R$ , where the constant  $\alpha$  depends on the density profile and it is generally of order one. Since a generic astronomical object is not at rest with respect to the Sun (because of the expansion of the Universe, of the peculiar motion, etc.), the application of the virial theorem to Coma requires the velocity to be measured with respect to its center-of-mass. Accordingly,  $v^2$  should be replaced by  $\sigma^2$ , where  $\sigma$  is the three-dimensional velocity dispersion of the Coma galaxies. Further, since only the line-of-sight velocity dispersion  $\sigma_{\parallel}$  of the galaxies can be measured, Zwicky made the simplest possible assumption that Coma galaxies are isotropically distributed, so that  $\sigma = \sqrt{3} \sigma_{\parallel}$ . As far as the potential energy is concerned, Zwicky assumed that

---

<sup>6</sup>The word *virial* comes from the latin *vis*, i.e., strength, or force; the term was coined by German physicist and mathematician Rudolf Clausius, one of the founders of the science of thermodynamics, around 1870.

galaxies are uniformly distributed inside Coma, which yields  $\alpha = 3/5$ . Thus, the virial theorem now reads

$$\sigma_{\parallel, \text{vir}}^2 = \frac{G M_{\text{gal}}}{5 R_{\text{Coma}}}, \quad (8.55)$$

where  $M_{\text{gal}}$  is the total mass of Coma in term of galaxies (no intracluster gas was known at that time). Zwicky was able to measure the line of sight of only seven galaxies of the cluster; assuming them to be representative of the whole galaxy population of Coma he found  $\langle v_{\parallel} \rangle \simeq 7.31 \times 10^8 \text{ cm s}^{-1}$  and  $\sigma_{\parallel, \text{obs}} \simeq 6.57 \times 10^7 \text{ cm s}^{-1}$ . Further, from the measured angular diameter of Coma and its distance as derived from the Hubble law he estimated  $R_{\text{Coma}} \simeq 10^{24} \text{ cm}$ . Finally, Zwicky supposed that Coma contains about  $N = 800$  galaxies with mass  $m_{\text{gal}} \simeq 10^9 M_{\odot}$  – which at that time was considered typical for galaxies—thereby getting  $M_{\text{Coma}} \simeq 8 \times 10^{11} M_{\odot}$ . Therefore Eq. (8.55) yields  $\sigma_{\parallel} \simeq 4.62 \times 10^6 \text{ cm s}^{-1}$ . Since  $\sigma_{\parallel}^2 \propto M_{\text{gal}}$ , in order to bring  $\sigma_{\parallel, \text{vir}}$  in agreement with  $\sigma_{\parallel, \text{obs}}$ , Zwicky had to increase  $M_{\text{gal}}$  by a factor of about 200 (he wrote 400), thereby obtaining for the Coma galaxies  $M_{\text{gal}} \simeq 2 \times 10^{11} M_{\odot}$  (he wrote  $4 \times 10^{11} M_{\odot}$ ). Thus, Zwicky ended up with the conclusion that Coma galaxies have a mass about two orders of magnitudes larger than expected: his explanation was that these galaxies are totally dominated by dark matter.

Despite this early evidence, it was only in the 1970s that scientists began to explore this discrepancy in a systematic way and that the existence of dark matter started to be quantified. It was realized that the discovery of dark matter would not only have solved the problem of the lack of mass in clusters of galaxies, but would also have had much more far-reaching consequences on our prediction of the evolution and fate of the Universe.

An important observational evidence of the need for dark matter was provided by the rotation curves of spiral galaxies—the Milky Way is one of them. Spiral galaxies contain a large population of stars placed on nearly circular orbits around the Galactic center. Astronomers have conducted observations of the orbital velocities of stars in the peripheral regions of a large number of spiral galaxies and found that the orbital speeds remain constant, contrary to the expected prediction of reduction at larger radii. The mass enclosed within the orbits radius must therefore gradually increase in regions even beyond the edge of the visible galaxy.

Later, another independent confirmation of Zwicky’s findings came from gravitational lensing. Lensing is the effect that bends the light coming from distant objects, due to large massive bodies in the path to the observer. As such it constitutes another method of measuring the total content of gravitational matter. The obtained mass-to-light ratios in distant clusters match the dynamical estimates of dark matter in those clusters.

### 8.1.5.1 How Is Dark Matter Distributed in Galaxies?

The first systematic investigation of the distribution of DM contained in spiral galaxies was carried out by Vera Rubin and collaborators between 1980 and 1985 using

stars as DM tracers. Since in spiral galaxies stars move on nearly circular orbits, the gravitational acceleration equals the centripetal force. Thus, by denoting by  $\mu$  the mass of a star, we have  $\mu v^2/r = G\mu M(r)/r^2$  where  $M(r)$  is the total mass inside the radius  $r$  of the orbit of the star:

$$v(r) = \sqrt{\frac{GM(r)}{r}}. \quad (8.56)$$

Thus, from the kinematic measurements of the *rotation curve*  $v(r)$  one can infer the dynamics of the galaxy. If all galactic mass were luminous, then at large enough distance from the center most of the mass would be well inside  $r$ , thereby implying that  $M(r) \simeq \text{constant}$ ; Eq. 8.56 yields  $v(r) \propto 1/\sqrt{r}$ . This behavior is called *Keplerian* because it is identical to that of the rotation velocity of the planets orbiting the Sun. Yet, the observations of Rubin and collaborators showed that  $v(r)$  rises close enough to the center, and then reaches a maximum and stays constant as  $r$  increases, failing to exhibit the expected Keplerian fall-off. According to Eq. 8.56 the observed behavior implies that in order to have  $v(r) = \text{constant}$  it is necessary that  $M(r) \propto r$ . But since

$$M(r) = 4\pi \int_0^r dr' r'^2 \rho(r'), \quad (8.57)$$

where  $\rho(r)$  is the mass density, the conclusion is that at large enough galactocentric distance the mass density goes like  $\rho(r) \propto 1/r^2$ . In analogy with the behavior of a self-gravitating isothermal gas sphere, the behavior is called *singular isothermal*. As a consequence, spiral galaxies turn out to be surrounded in first approximation by a singular isothermal halo made of dark matter. In order to get rid of the central singularity, it is often assumed that the halo profile is pseudo-isothermal, assuming a density:

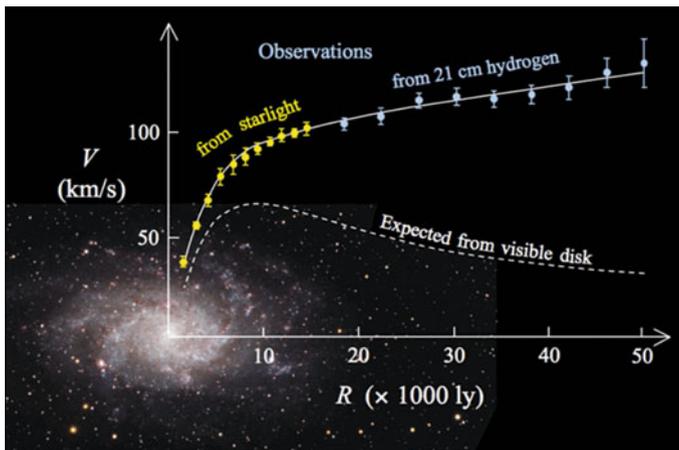
$$\text{Pseudo-Isothermal} : \rho_{\text{iso}}(r) = \frac{\rho_0}{1 + (r/r_s)^2}. \quad (8.58)$$

While strongly suggestive of the existence of dark halos around spiral galaxies, optical studies have the disadvantage that typically at the edge of the stellar disk the difference between a constant rotation curve and a Keplerian one is about 15%, too small to draw waterproof conclusion when errors are taken into account. Luckily, the disks of spirals also contain neutral atomic hydrogen (HI) clouds; like stars they move on nearly circular orbits, but the gaseous disk extends typically twice, and in some cases even more. According to relativistic quantum mechanics, the nonrelativistic ground state of hydrogen at  $E \simeq -13.6 \text{ eV}$  splits into a pair of levels, depending on the relative orientation of the spins of the proton and the electron; the energy splitting is only  $\delta E \simeq 5.9 \mu\text{eV}$  (hyperfine splitting). Both levels are populated thanks to collisional excitation and interaction with the CMB; thus, HI clouds can be detected by radio-telescopes since photons emitted during the transition to the ground state have a wavelength of about 21 cm. In 1985 Van Albada, Bahcall, Begeman and Sancisi performed this measurement for the spiral NGC 3198, whose gaseous disks is

more extended than the stellar disk by a factor of 2.7, and could construct the rotation curve out to 30 kpc. They found that the flat behavior persists, and this was regarded as a clear-cut evidence for dark matter halos around spiral galaxies. Measurements now include a large set of galaxies (Fig. 8.12), including the Milky Way (Fig. 8.13).

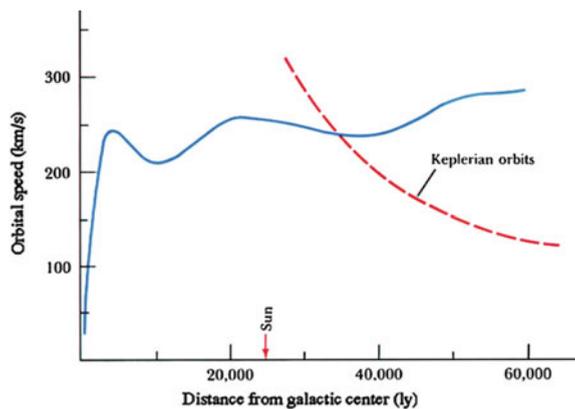
Profiles obtained in numerical simulations of dark matter including baryons are steeper in the center than those obtained from simulations with dark matter only. The Navarro, Frenk, and White (NFW) profile, often used as a benchmark, follows a  $r^{-1}$  distribution at the center. On the contrary, the Einasto profile does not follow a power law near the center of galaxies, is smoother at kpc scales, and seems to fit better more recent numerical simulations. A value of about 0.17 for the shape parameter  $\alpha$  in Eq. 8.59 is consistent with present data and simulations. Moore and collaborators have suggested profiles steeper than NFW.

The analytical expression of these profiles are



**Fig. 8.12** Rotation curve of the galaxy M33 (from Wikimedia Commons, public domain)

**Fig. 8.13** Rotation curve of the Milky Way (from <http://abyss.uoregon.edu>)



$$\begin{aligned}
 \text{NFW} : \quad \rho_{\text{NFW}}(r) &= \rho_s \frac{r_s}{r} \left(1 + \frac{r}{r_s}\right)^{-2} \\
 \text{Einasto} : \quad \rho_{\text{Einasto}}(r) &= \rho_s \exp\left\{-\frac{2}{\alpha} \left[\left(\frac{r}{r_s}\right)^\alpha - 1\right]\right\} \\
 \text{Moore} : \quad \rho_{\text{Moore}}(r) &= \rho_s \left(\frac{r_s}{r}\right)^{1.16} \left(1 + \frac{r}{r_s}\right)^{-1.84}
 \end{aligned} \tag{8.59}$$

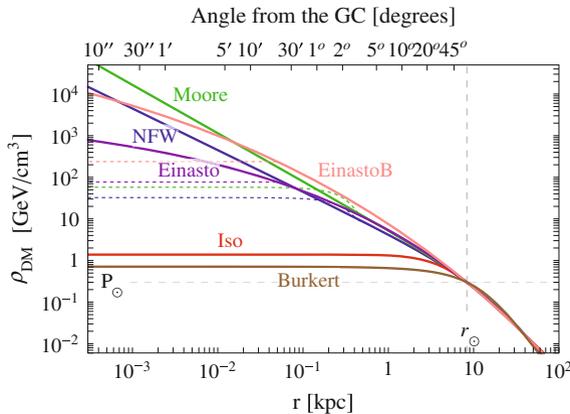
Presently, there are no good observational measurements of the shape of the Milky Way near the Galactic center; this is why one usually assumes a spherically symmetrical distribution. Figure 8.14 compares these different profiles with the constraint to fit the velocities in the halo of our Galaxy.

In the neighborhood of the solar system one has a DM density

$$\rho_{\text{DM, local}} \simeq 0.4 \text{ GeV/cm}^3,$$

i.e., five orders of magnitude larger than the total energy density of the Universe.

To distinguish between the functional forms for the halos is not easy. They vary among each other only in the central region, where the luminous matter is dominant. Needless to say, the high-density, central region is the most crucial for detection—and uncertainties there span three orders of magnitude. Also because of this, one of the preferred targets for astrophysical searches for DM are small satellite galaxies of the Milky Way, the so-called dwarf spheroidals (dSph), which typically have a number of stars  $\sim 10^3\text{--}10^8$ , to be compared with the  $\sim 10^{11}$  of our Galaxy. For these galaxies the



**Fig. 8.14** Comparison of the densities as a function of the radius for DM profiles used in the literature, with values adequate to fit the radial distribution of velocities in the halo of the Milky Way. The curve EinastoB indicates an Einasto curve with a different  $\alpha$  parameter. From M. Cirelli et al., “PPPC 4 DM ID: A Poor Particle Physicist Cookbook for Dark Matter Indirect Detection”, arXiv:1012.4515, JCAP 1103 (2011) 051

ratio between the estimate of the total mass  $M$  inferred from the velocity dispersion (velocities of single stars are measured with an accuracy of a few kilometers per second thanks to optical measurements) and the luminous mass  $L$ , inferred from the count of the number of stars, can be very large. The dwarf spheroidal satellites of the Milky Way could become tidally disrupted if they did not have enough dark matter. In addition these objects are not far from us: a possible DM signal should not be attenuated by distance dimming. Table 8.1 shows some characteristics of dSph in the Milky Way; their position is shown in Fig. 8.15.

The observations of the dynamics of galaxies and clusters of galaxies, however, are not the only astrophysical evidence of the presence of DM. Cosmological models for the formation of galaxies and clusters of galaxies indicate that these structures fail to form without DM.

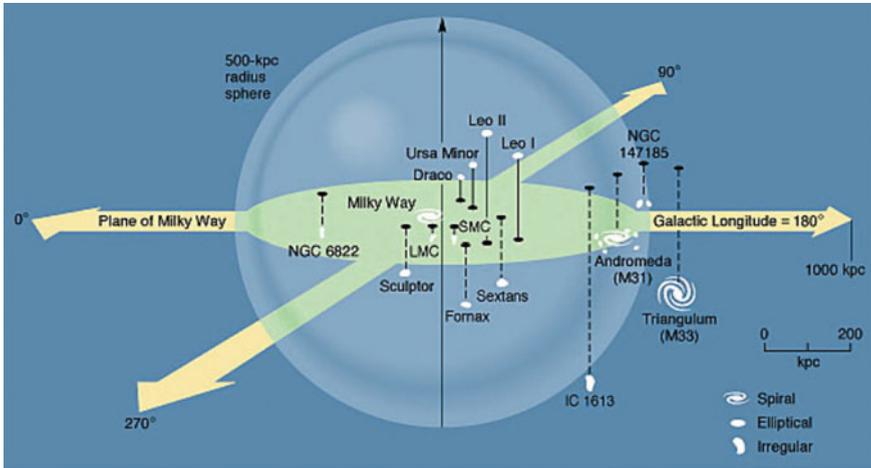
### 8.1.5.2 An Alternative Explanation: Modified Gravity

The dependence of  $v^2$  on the mass  $M(r)$  on which the evidence for DM is based relies on the virial theorem, stating that the kinetic energy is on average equal to the absolute value of the total energy for a bound state, defining zero potential energy at infinite distance. The departure from this Newtonian prediction could also be related to a departure from Newtonian gravity.

Alternative theories do not necessarily require dark matter, and replace it with a modified Newtonian gravitational dynamics. Notice that, in a historical perspective, deviations from expected gravitational dynamics already led to the discovery of previously *unknown matter* sources. Indeed, the planet Neptune was discovered following the prediction by Le Verrier in the 1840s of its position based on the detailed observation of the orbit of Uranus and Newtonian dynamics. In the late nineteenth century, the disturbances to the expected orbit of Neptune led to the discovery of

**Table 8.1** A list of dSph satellites of the Milky Way that may represent the best candidates for DM searches according to their distance from the Sun, luminosity, and inferred  $M/L$  ratio

dSph	$D_{\odot}$ (kpc)	$L$ ( $10^3 L_{\odot}$ )	$M/L$ ratio
Segue 1	23	0.3	>1000
UMa II	32	2.8	1100
Willman 1	38	0.9	700
Coma Berenices	44	2.6	450
UMi	66	290	580
Sculptor	79	2200	7
Draco	82	260	320
Sextans	86	500	90
Carina	101	430	40
Fornax	138	15500	10



**Fig. 8.15** The Local Group of galaxies around the Milky Way (from <http://abyss.uoregon.edu/~js/ast123/lectures/lec11.html>). The largest galaxies are the Milky Way, Andromeda, and M33, and have a spiral form. Most of the other galaxies are rather small and with a spheroidal form. These orbit closely the large galaxies, as is also the case of the irregular Magellanic Clouds, best visible in the Southern hemisphere, and located at a distance of about 120,000 ly, to be compared with the Milky Way radius of about 50,000 ly

Pluto. On the other hand, the precession of the perihelion of Mercury, which could not be *quantitatively* explained by Newtonian gravity, confirmed the prediction of general relativity—and thus a modified dynamics.

The simplest model of modified Newtonian dynamics is called MOND; it was proposed in 1983 by Milgrom, suggesting that for extremely small accelerations the Newton’s gravitational law may not be valid—indeed Newton’s law has been verified only at reasonably large values of the gravitational acceleration. MOND postulates that the acceleration  $a$  is not linearly dependent on the gradient of the gravitational field  $\phi_N$  at small values of the acceleration, and proposes the following modification:

$$\mu\left(\frac{a}{a_0}\right) a = |-\nabla\phi_N| . \tag{8.60}$$

The function  $\mu$  is positive, smooth, and monotonically increasing; it is approximately equal to its argument when the argument takes small values compared to unity (deep MOND limit), but approaches unity when that argument is large.  $a_0$  is a constant of the order of  $10^{-10} \text{ m s}^{-2}$ .

Let us now consider again stars orbiting a galaxy with speed  $v(r)$  at radius  $r$ . For large  $r$  values,  $a$  will be smaller than  $a_0$  and we can approximate  $\mu(x) \simeq x$ . One has then

$$\frac{v^4}{r^2} \simeq a_0 \frac{GM}{r^2} .$$

In this limit, the rotation curve flattens at a typical value  $v_f$  given by

$$v_f = (M G a_0)^{1/4}. \quad (8.61)$$

MOND explains well the shapes of rotation curves; for clusters of galaxies one finds an improvement but the problem is not completely solved.

The likelihood that MOND is the full explanation for the anomaly observed in the velocities of stars in the halo of galaxies is not strong. An explanation through MOND would require an ad hoc theory to account for cosmological evidence as well. In addition the observation in 2004 of the merging galaxy cluster 1E0657-58 (the so-called bullet cluster), has further weakened the MOND hypothesis. The bullet cluster consists of two colliding clusters of galaxies, at a distance of about 3.7 Gly. In this case (Fig. 8.16), the distance of the center of mass to the center of baryonic mass cannot be explained by changes in the gravitational law, as indicated by data with a statistical significance of  $8\sigma$ .

One could also consider the fact that galaxies may contain invisible matter of known nature, either baryons in a form which is hard to detect optically, or massive neutrinos—MOND reduces the amount of invisible matter needed to explain the observations.



**Fig. 8.16** The matter in the “bullet cluster” is shown in this composite image (from <http://apod.nasa.gov/apod/ap060824.html>, credits: NASA/CXC/CfA/ M. Markevitch et al.). In this image depicting the collision of two clusters of galaxies, the bluish areas show the distributions of dark matter in the clusters, as obtained from gravitational lensing, and the red areas correspond to the hot X-ray emitting gases. The individual galaxies observed in the optical image data have a total mass much smaller than the mass in the gas, but the sum of these masses is far less than the mass of dark matter. The clear separation of dark matter and gas clouds is a direct evidence of the existence of dark matter

### 8.1.6 Age of the Universe: A First Estimate

The age of the Universe is an old question. Has the Universe a finite age? Or is the Universe eternal and always equal to itself (steady state Universe)?

For sure the Universe must be older than the oldest object that it contains and the first question has been then: how old is the Earth? In the eleventh century, the Persian astronomer Abu Rayhan al-Biruni had already realized that Earth should have a finite age, but he just stated that the origin of Earth was too far away to possibly measure it. In the nineteenth century the first quantitative estimates finally came. From considerations, both, on the formation of the geological layers, and on the thermodynamics of the formation and cooling of Earth, it was estimated that the age of the Earth should be of the order of tens of millions of years. These estimates were in contradiction with both, some religious beliefs, and Darwin's theory of evolution. Rev. James Ussher, an Irish Archbishop, published in 1650 a detailed calculation concluding that according to the Bible "God created Heaven and Earth" some six thousand years ago, more precisely "at the beginning of the night of October 23rd in the year 710 of the Julian period", which means 4004 B.C.. On the other hand, tens or even a few hundred million years seemed to be a too short time to allow for the slow evolution advocated by Darwin. Only the discovery of radioactivity at the end of nineteenth century provided precise clocks to date rocks and meteorite debris with, and thus to allow for reliable estimates of the age of the Earth. Surveys in the Hudson Bay in Canada found rocks with ages of over four billion ( $\sim 4.3 \times 10^9$ ) years. On the other hand measurements on several meteorites, in particular on the Canyon Diablo meteorite found in Arizona, USA, established dates of the order of  $(4.5-4.6) \times 10^9$  years. Darwin had the time he needed!

The proportion of elements other than hydrogen and helium (defined as the metallicity) in a celestial object can be used as an indication of its age. After primordial nucleosynthesis (Sect. 8.1.4) the Universe was basically composed by hydrogen and helium. Thus the older (first) stars should have lower metallicity than the younger ones (for instance our Sun). The measurement of the age of low metallicity stars imposes, therefore, an important constraint on the age of the Universe. Oldest stars with a well-determined age found so far are, for instance, HE 1523-0901, a red giant at around 7500 light-years away from us, and HD 140283, denominated the Methuselah star, located around 190 light years away. The age of HE 1523-0901 was measured to be 13.2 Gyr, using mainly the decay of uranium and thorium. The age of HD 140283 was determined to be  $(14.5 \pm 0.8)$  Gyr.

The "cosmological" age of the Universe is defined as the time since the Big Bang, which at zeroth order is just given by the inverse of the Hubble constant:

$$t_0 \simeq \frac{1}{H_0} \simeq 14 \text{ Gyr}. \quad (8.62)$$

A more precise value is determined by solving the equations of evolution of the Universe, the so-called Friedmann equations (see Sect. 8.2), for a given set of the

cosmological parameters. Within the  $\Lambda$ CDM model (see Sect. 8.4) the best-fit value, taking into account the present knowledge of such parameters, is

$$t_0 = (13.80 \pm 0.04) \text{ Gyr} . \quad (8.63)$$

Within uncertainties both the cosmological age and the age of the first stars are compatible, but the first stars had to be formed quite early in the history of the Universe.

Finally, we stress that a Universe with a finite age and in expansion will escape the nineteenth-century Olbers' Paradox: "How can the night be dark?" This paradox relies on the argument that in an infinite static Universe with uniform star density (as the Universe was believed to be by most scientists until the mid of last century) the night should be as bright as the day. In fact, the light coming from a star is inversely proportional to the square of its distance, but the number of stars in a shell at a distance between  $r$  and  $(r + dr)$  is proportional to the square of the distance  $r$ . From this it seems that any shell in the Universe should contribute the same amount light. Apart from some too crude approximations (such as, not taking into account the finite life of stars), redshift and the finite size of the Universe solve the paradox.

## 8.2 General Relativity

Special relativity, introduced in Chap. 2, states that one cannot distinguish on the basis of the laws of physics between two inertial frames moving at constant speed one with respect to the other. Experience tells that it is possible to distinguish between an inertial frame and an accelerated frame. Can the picture change if we include gravity?

In classical mechanics, gravity is a force and determines the movement of a body according to Newton's second law. The gravitational force is proportional to the body's gravity charge, which is the gravitational mass  $m_g$ ; this, in turn, is proportional to the inertial mass  $m_I$ , that characterizes the body's inertia to be accelerated by a force. The net result is that the local acceleration of a body,  $g$ , due to a gravitational field created by a mass  $M$  at a distance  $r$ , is proportional to the ratio  $m_g/m_I$

$$F_g = m_g G \frac{M}{r^2} = F_g = m_I g ,$$

and

$$g = \frac{m_g}{m_I} G \frac{M}{r^2} ,$$

where  $G$  is the universal gravitational constant.

Thus if  $m_g$  were proportional to  $m_I$  the movement of a body in a gravitational field would be independent of its mass and composition. In fact the experiments of Galilei on inclined planes showed the universality of the movement of rolling

balls of different compositions and weights. Such universality was also found by measuring the period of pendulums with different weights and compositions but identical lengths, first again by Galilei, and later on with a much higher precision (better than 0.1%) by Newton. Nowadays,  $m_g/m_I$  is experimentally known to be constant for all bodies, independent of their nature, mass, and composition, up to a relative resolution of  $5 \times 10^{-14}$ . We then *choose*  $G$  in such a way that  $m_g/m_I \equiv 1$ . Space-based experiments, allowing improved sensitivities up to  $10^{-17}$  on  $m_g/m_I$ , are planned for the next years.

### 8.2.1 *Equivalence Principle*

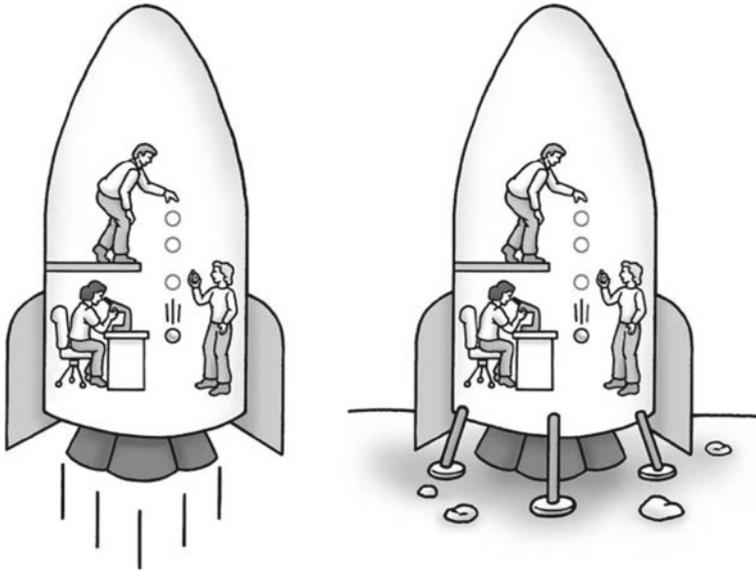
It is difficult to believe that such a precise equality is just a coincidence. This equality has been thus promoted to the level of a principle, named the *weak equivalence principle*, and it led Einstein to formulate the *strong equivalence principle* which is a fundamental postulate of General Relativity (GR). Einstein stated that it is not possible to distinguish infinitesimal movements occurring in an inertial frame due to gravity (which are proportional to the gravitational mass), from movements occurring in an accelerated frame due to “fictitious” inertial forces (which are proportional to the inertial mass).

A ball dropped in a gravitational field has, during an infinitesimal time interval, the same behavior that a free ball has in an accelerated frame if the acceleration  $a$  of the accelerated frame is opposite to the local acceleration  $g$  of gravity (Fig. 8.17). No experiment can distinguish between the two scenarios.

### 8.2.2 *Light and Time in a Gravitational Field*

In the same way, if an observer is inside a free-falling elevator, gravity is locally canceled out by the “fictitious” forces due to the acceleration of the frame. Free-falling frames are equivalent to inertial frames. A horizontal light beam in such a free falling elevator then moves in a straight line for an observer inside the elevator, but it curves down for an observer outside the elevator (Fig. 8.18). Light therefore curves in a gravitational field.

The bending of light passing near the Sun was discussed by Newton himself and computed by Cavendish and Soldner to be of about 0.9 arcsecond for a light ray passing close to the Sun’s limb; this result in its deduction assumes the Newton corpuscular theory of light. However, Einstein found, using the newborn equations of GR, a double value and then a clear test was in principle possible through the observation of the apparent position of stars during a total solar eclipse. In May 1919 Eddington and Dyson led two independent expeditions, respectively, to the equatorial islands of São Tomé and Príncipe and to Sobral, Brazil. The observations were perturbed by clouds (Príncipe) and by instrumental effects (Sobral) but nevertheless the announcement by Eddington that Einstein’s predictions were confirmed had an

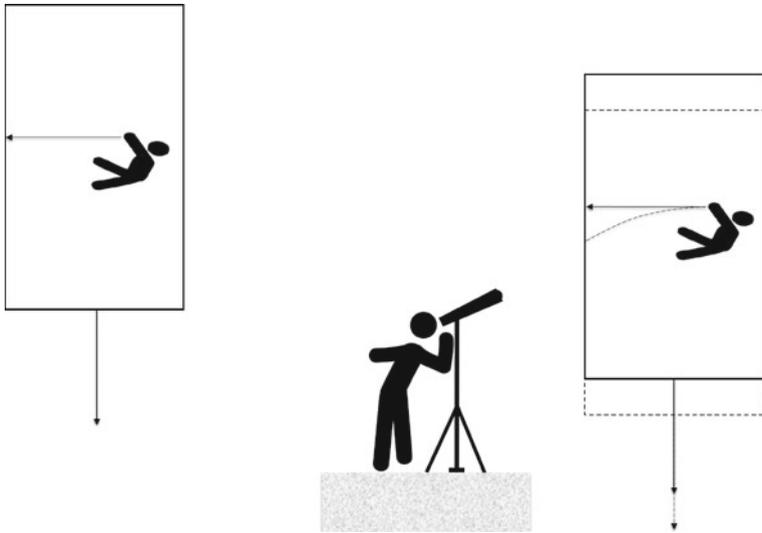


**Fig. 8.17** Scientists performing experiments in an accelerating spaceship moving with an upward acceleration  $g$  (left) obtain the same results as if they were on a planet with gravitational acceleration  $g$  (right). From A. Zimmerman Jones, D. Robbins, “String Theory For Dummies”, Wiley 2009

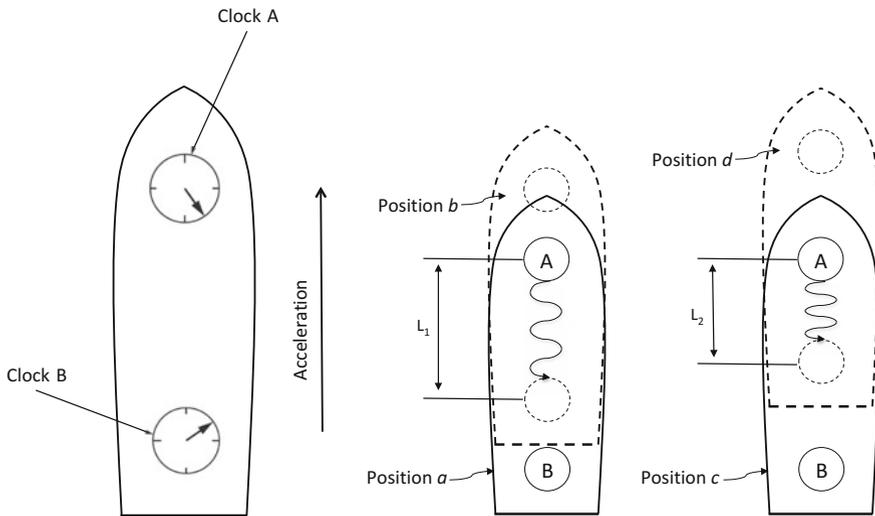
enormous impact on public opinion and made general relativity widely known. Further and more robust observations were carried on in the following years and the predictions of general relativity on light deflection were firmly confirmed.

Now we want to use the principle of equivalence for predicting the influence of the gravitational field on the measurement of time intervals. We shall follow the line of demonstration by Feynman in his famous Lectures on Physics.

Suppose we put a clock A at the “head” of a rocket uniformly accelerating, and another identical clock B at the “tail,” as in Fig. 8.19, left. Imagine that the front clock emits a flash of light each second, and that you are sitting at the tail comparing the arrival of the light flashes with the ticks of clock B. Assume that the rocket is in the position  $a$  of Fig. 8.19, right, when clock A emits a flash, and at the position  $b$  when the flash arrives at clock B. Later on the ship will be at position  $c$  when the clock A emits its next flash, and at position  $d$  when you see it arrive at clock B. The first flash travels the distance  $L_1$  and the second flash travels the shorter distance  $L_2$ , because the ship is accelerating and has a higher speed at the time of the second flash. You can see, then, that if the two flashes were emitted from clock A one second apart, they would arrive at clock B with a separation somewhat less than one second, since the second flash does not spend as much time on the way. The same will also happen for all the later flashes. So if you were sitting in the tail you would conclude that clock A was running faster than clock B. If the rocket is at rest in a gravitational



**Fig. 8.18** Trajectory of a light beam in an elevator freely falling seen by an observer inside (left) and outside (right): Icons made by Freepik from [www.flaticon.com](http://www.flaticon.com)



**Fig. 8.19** Rocket on the left: Two clocks onboard an accelerating rocket. Two rockets on the right: Why the clock at the head appears to run faster than the clock at the tail

field, the principle of equivalence guarantees that the same thing happens. We have the relation

$$(\text{Rate at the receiver}) = (\text{Rate of emission}) \left( 1 + \frac{gH}{c^2} \right)$$

where  $H$  is the height of the emitter above the receiver. This time dilation due to the gravitational field can also be seen as due to the differences in the energy losses of the photons emitted in both elevators by “climbing” out the gravitational field. In fact in a weak gravitational field the variation of the total energy of a particle of mass  $m$ , assuming the equivalence principle, is independent of  $m$ :

$$\frac{\Delta E}{E} \simeq \frac{mgH}{mc^2} = \frac{gH}{c^2}.$$

Since, for a photon, energy and frequency are related by the Planck formula  $E = h\nu$ :

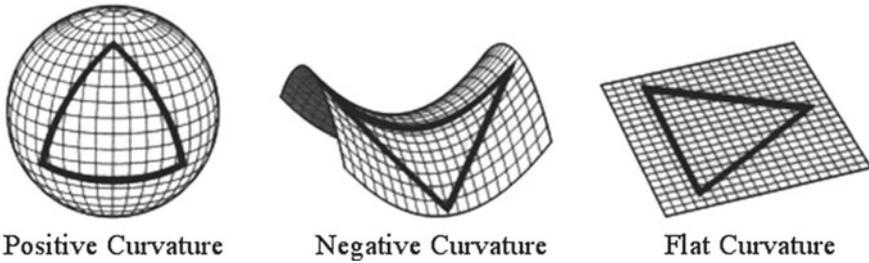
$$\frac{\Delta E}{E} = \frac{\Delta\nu}{\nu} \sim \frac{\Delta\lambda}{\lambda} \sim \frac{gH}{c^2}.$$

### 8.2.3 Flat and Curved Spaces

Gravity in GR is no longer a force (whose sources are the masses) acting in a flat spacetime Universe. Gravity is embedded in the geometry of spacetime that is determined by the energy and momentum contents of the Universe.

Classical mechanics considers that we are living in a Euclidean three-dimensional space (flat, i.e., with vanishing curvature), where through each point outside a “straight” line (a geodesic) there is one, and only one, straight line parallel to the first one; the sum of the internal angles of a triangle is  $180^\circ$ ; the circumference of a circle of radius  $R$  is  $2\pi R$ , and so on. However, it is interesting to consider what would happen if this were not the case.

To understand why a different approach could be interesting, let us consider that we are living on the surface of a sphere (the Earth is approximately a sphere). Such a surface has positive (ideally constant) curvature at all points (i.e., the spherical surface stays on just one side of the tangent planes to the surface at any given point): the small distance between two points is now the length of the arc of circle connecting the two points and whose center coincides with the center of the sphere (geodesic line in the sphere), and this is as close as we can get to a *straight line*; the sum of the angles of a triangle is greater than  $180^\circ$ ; the circumference of a circle of radius  $R$  is less than  $2\pi R$ . Alternatively, let us imagine that we were living on a saddle, which has a negative curvature (the surface can curve away from the tangent plane in two different directions): then the sum of the angles of a triangle is less than  $180^\circ$ ; the perimeter of a circumference is greater than  $2\pi R$ , and so on. The three cases are visualized in Fig. 8.20. The metric of the sphere and of the saddle are not Euclidean.



**Fig. 8.20** 2D surfaces with positive, negative, and null curvatures (from <http://thesimplephysicist.com>, © 2014 Bill Halman/tdotwebcreations)

### 8.2.3.1 2D Space

In a flat 2D surface (a plane) the square of the distance between two points is given in Cartesian coordinates by

$$ds^2 = dx^2 + dy^2 \tag{8.64}$$

or

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{8.65}$$

with

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8.66}$$

The metric  $g_{\mu\nu}$  of the 2D flat surface is constant and the geodesics are straight lines.

The metric of a 2D spherical surface is a little more complex. The square of the distance between two neighboring close points situated on the surface of a sphere with radius  $a$  embedded in our usual 3D Euclidean space (Fig. 8.21) is given in spherical coordinates by

$$ds^2 = a^2 d\theta^2 + a^2 \sin^2 \theta d\varphi^2. \tag{8.67}$$

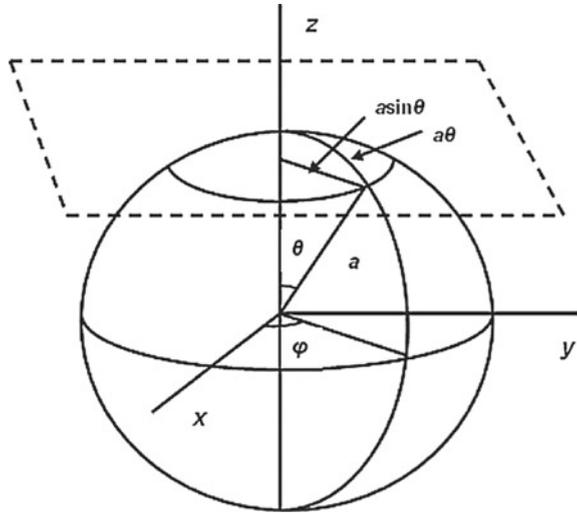
The maximum distance between two points on the sphere is bounded by  $d = \sqrt{s^2} = \pi a$ : the two points are the extrema of a half great circle.

Now the matrix representing the metric in spherical coordinates,

$$g_{\mu\nu} = \begin{pmatrix} a^2 & 0 \\ 0 & a^2 \sin^2 \theta \end{pmatrix}, \tag{8.68}$$

is no longer constant, because of the presence of the  $\sin^2 \theta$  term. It is not possible to cover the entire sphere with one unique plane without somewhat distorting the plane, although it is always possible to define locally at each point one tangent plane.

**Fig. 8.21** Distances on a sphere of radius  $a$ . From A. Tan et al. DOI:10.5772/50508



The geodesics are not straight lines; they are indeed part of great circles, as it can be deduced directly from the metrics and its derivatives.

This metric can now be written introducing a new variable  $r = \sin \theta$  as

$$ds^2 = a^2 \left( \frac{dr^2}{1 - Kr^2} + r^2 d\varphi^2 \right) \tag{8.69}$$

with

$$K = 1$$

for the case of the sphere.<sup>7</sup> Indeed, the sphere has a positive ( $K = 1$ ) curvature at any point of its surface. However, the above expressions are valid both for the case of negative ( $K = -1$ ) and null ( $K = 0$ ) curvature. In the case of a flat surface, indeed, the usual expression in polar coordinates is recovered:

$$ds^2 = a^2 (dr^2 + r^2 d\varphi^2) . \tag{8.70}$$

The distance between two points with the same  $\varphi$  and, respectively,  $r_1 = 0$  and  $r_2 = R/a$ , is given by:

$$s = \int_0^{\frac{R}{a}} a \frac{dr}{\sqrt{1 - Kr^2}} = a S_k \tag{8.71}$$

---

<sup>7</sup> $r$  has range  $[0, 1]$ ;  $K$  is the curvature, which, in general, can be  $-1, 0,$  or  $+1$ . The more general change of coordinates  $r' = a \sin \theta$  does not result in anything new, and can be recast in the form used above after setting  $r = r'/a$ . Of course with the  $r'$  coordinate, the curvature is not normalized, and can be, generically, negative, zero, or positive.

with

$$S_k = \begin{cases} \arcsin(R/a) & \text{if } K = 1 \\ R/a & \text{if } K = 0 \\ \operatorname{arcsinh}(R/a) & \text{if } K = -1 \end{cases} . \tag{8.72}$$

The area of the sphere is now given by

$$A = 4 \pi a^2 S_k^2 . \tag{8.73}$$

The relation between the proper distance and the luminosity distance (Sect. 8.1.1) is now

$$d_L = d_p \frac{a}{R} S_k (1 + z) , \tag{8.74}$$

and the metric can also be written in a more compact form using the function  $S_k$ :

$$ds^2 = a^2 (dr^2 + S_k^2 d\varphi^2) . \tag{8.75}$$

### 8.2.3.2 3D Space

For a homogeneous and isotropic 3D space the previous formula can be generalized (now  $r$  and  $\theta$  are independent variables) leading to:

$$ds^2 = a^2 \left[ \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2\theta d\varphi^2) \right] .$$

### 8.2.3.3 4D Spacetime

For a spatially homogeneous and isotropic 4D spacetime the generalization leads to the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, sometimes just called the Robertson-Walker metric ( $c = 1$ ):

$$ds^2 = dt^2 - a^2(t) \left[ \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2\theta d\varphi^2) \right] \tag{8.76}$$

where  $a(t)$  is a radial scale factor which may depend on  $t$  (allowing for the expansion/contraction of the Universe).

Introducing the solid angle,  $d\Omega^2 = d\theta^2 + \sin^2\theta d\varphi^2$ , the FLRW metric can be written as

$$ds^2 = dt^2 - a^2(t) \left( \frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right) . \tag{8.77}$$

Finally, the Robertson–Walker metric can also be written using the functions  $S_k$  introduced above as

$$ds^2 = dt^2 - a^2(t) ( dr^2 + S_k^2 d\Omega^2 ). \quad (8.78)$$

The special relativity Minkowski metric is a particular case ( $K = 0$ ,  $at = \text{constant}$ ) of the FLRW metric.

The geodesics in a 4D spacetime correspond to the extremal (maximum or minimum depending on the metric definition) world lines joining two events in spacetime and not to the 3D space paths between the two points. The geodesics are determined, as before, just from the metric and its derivatives.

### 8.2.4 Einstein's Equations

In GR the world lines of freely falling test particles are just the geodesics of the 4D spacetime of the Universe we are living in, whose geometry is locally determined by its energy and momentum contents as expressed by Einstein's equations (which, below, are in the form where we neglect a cosmological constant term, see later)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} = \frac{8\pi}{c^4}T_{\mu\nu}.$$

In the equations above  $G_{\mu\nu}$  and  $R_{\mu\nu}$  are, respectively, the Einstein and the Ricci tensors, which are built from the metric and its derivatives;  $\mathcal{R}$  is the Ricci scalar ( $\mathcal{R} = g^{\mu\nu}R_{\mu\nu}$ ) and  $T_{\mu\nu}$  is the energy–momentum tensor.

The energy and the momentum of the particles determine the geometry of the Universe which then determines the trajectories of the particles. Gravity is embedded in the geometry of spacetime. Time runs slower in the presence of gravitational fields.

Einstein's equations are tensor equations and thus independent on the reference frame (the covariance of the physics laws is automatically ensured). They involve 4D symmetric tensors and represent in fact 10 independent nonlinear partial differential equations whose solutions, the metrics of spacetime, are in general difficult to sort out. However, in particular and relevant cases, exact or approximate solutions can be found. Examples are the Minkowski metric (empty Universe); the Schwarzschild metric (spacetime metric outside a noncharged spherically symmetric nonrotating massive object—see Sect. 8.2.8); the Kerr metric (a cylindrically symmetric vacuum solution); the FLRW metric (homogeneous and isotropic Universe—see Sect. 8.2.5).

Einstein introduced at some point a new metric proportional term in his equations (a quantity  $\Lambda$  constant in space and time, the so-called “cosmological constant”):

$$G_{\mu\nu} + g_{\mu\nu}\Lambda = \frac{8\pi}{c^4}T_{\mu\nu}. \quad (8.79)$$

His motivation was to allow for static cosmological solutions, as this term can balance gravitational attraction. Although later on Einstein discarded this term (the static

Universe would be unstable), the recent discovery of the accelerated expansion of the Universe might give it again an essential role (see Sects. 8.2.5 and 8.4).

The energy–momentum tensor  $T^{\mu\nu}$  in a Universe of free noninteracting particles with four-momentum  $p_i^\mu$  moving along trajectories  $r_i t$  is defined as

$$T^{\mu 0} = \sum_i p_i^\mu t \delta^3(r - r_i t) \tag{8.80}$$

$$T^{\mu k} = \sum_i p_i^\mu t \frac{dx_i^k}{dt} \delta^3(r - r_i t) . \tag{8.81}$$

The  $T^{\mu 0}$  terms can be seen as “charges” and the  $T^{\mu k}$  terms as “currents”, which then obey a continuity equation ensuring energy–momentum conservation. In general relativity local energy–momentum conservation generalizes the corresponding results in special relativity,

$$\frac{\partial}{\partial x^0} T^{\mu 0} + \nabla_i T^{\mu i} = 0, \text{ or } \frac{\partial}{\partial x^\nu} T^{\mu\nu} = 0. \tag{8.82}$$

To get an intuitive grasp of the physical meaning of the energy–momentum tensor, let us consider the case of a special relativistic perfect fluid (no viscosity). In the rest frame of a fluid with energy density  $\rho$  and pressure  $\mathcal{P}$

$$T^{00} = c^2 \rho ; T^{0i} = 0 ; T^{ij} = \mathcal{P} \delta_{ij} . \tag{8.83}$$

Pressure has indeed the dimension of an energy density ( $\delta W = F \cdot dx = \mathcal{P} dV$ ) and accounts for the “kinetic energy” of the fluid.

To appreciate a fundamental difference from the Newtonian case, we quote that for a perfect fluid with energy density  $\rho$  and pressure  $\mathcal{P}$  the weak gravity field predicted by Newton is given by

$$\nabla^2 \phi = 4 \pi G \rho, \tag{8.84}$$

from which we see that pressure does not contribute. On the contrary, the weak field limit of Einstein’s equations is

$$\nabla^2 \phi = 4 \pi G \left( \rho + \frac{3 \mathcal{P}}{c^2} \right) . \tag{8.85}$$

Remembering that, in the case of a relativistic fluid

$$\mathcal{P} \sim \frac{1}{3} \rho c^2 \tag{8.86}$$

the weak gravitational field is then determined by

$$\nabla^2 \phi = 8 \pi G \rho, \quad (8.87)$$

which shows that the gravitational field predicted by general relativity is twice the one predicted by Newtonian gravity. Indeed, the observed light deflection by Eddington in 1919 at S. Tomé and Príncipe islands in a solar eclipse was twice the one expected according to classical Newtonian mechanics.

Once the metric is known, the free fall trajectories of test particles are obtained “just” by solving the geodesic equations

$$\frac{d^2 x^\sigma}{d\tau^2} + \Gamma_{\mu\nu}^\sigma \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0, \quad (8.88)$$

where  $\Gamma_{\mu\nu}^\sigma$  are the Christoffel symbols given by

$$\Gamma_{\mu\nu}^\sigma = \frac{g^{\rho\sigma}}{2} \left( \frac{\partial g_{\nu\rho}}{\partial x^\mu} + \frac{\partial g_{\mu\rho}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\rho} \right). \quad (8.89)$$

In the particular case of flat space in Cartesian coordinates the metric tensor is everywhere constant,  $\Gamma_{\mu\nu}^\sigma = 0$ , and then

$$\frac{d^2 x^\mu}{d\tau^2} = 0.$$

The free particles classical straight world lines are then recovered.

### 8.2.5 *The Friedmann–Lemaître–Robertson–Walker Model (Friedmann Equations)*

The present standard model of cosmology assumes the so-called cosmological principle, which in turn assumes a homogeneous and isotropic Universe at large scales. Homogeneity means that, in Einstein’s words, “all places in the Universe are alike” and isotropic just means that all directions are equivalent.

The FLRW metric discussed before (Sect. 8.2.3) embodies these symmetries leaving two independent functions,  $a(t)$  and  $K(t)$ , which represent, respectively, the evolution of the scale and of the curvature of the Universe. The Russian physicist Alexander Friedmann in 1922, and independently the Belgian Georges Lemaître in 1927, solved Einstein’s equations for such a metric leading to the famous Friedmann equations, which are still the starting point for the standard cosmological model, also known as the Friedmann–Lemaître–Robertson–Walker (FLRW) model.

The Friedmann equations can be written (with the convention  $c = 1$ ) as

$$\left( \frac{\dot{a}}{a} \right)^2 + \frac{K}{a^2} = \frac{8\pi G}{3} \rho + \frac{\Lambda}{3} \quad (8.90)$$

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3}(\rho + 3\mathcal{P}) + \frac{\Lambda}{3}. \quad (8.91)$$

These equations can be combined into a thermodynamics-like equation

$$\frac{d}{dt}(\rho a^3) = -\mathcal{P} \frac{d}{dt}(a^3), \quad (8.92)$$

where by identifying  $a^3$  with the volume  $V$  we can recognize adiabatic energy conservation

$$dE = -\mathcal{P} dV.$$

Moreover, remembering that the Hubble parameter is given by (Eq. 8.8):

$$H = \frac{\dot{a}}{a},$$

the first Friedmann equation is also often written as

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3} \rho + \frac{\Lambda}{3}, \quad (8.93)$$

which shows that the Hubble constant is not a constant but a parameter that evolves with the evolution of the Universe.

### 8.2.5.1 Classical Newtonian Mechanics

“Friedmann-like” equations can also be formally deduced in the framework of classical Newtonian mechanics, as follows:

1. From Newton law of gravitation and from the Newton second law of motion we can write

$$m\ddot{R} = -\frac{GMm}{R^2} \implies \left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3} \rho. \quad (8.94)$$

2. From energy conservation

$$\frac{1}{2}m\dot{R}^2 - \frac{GMm}{R} = \text{constant} \implies \left(\frac{\dot{a}}{a}\right)^2 - \frac{\text{constant}}{a^2} = \frac{8\pi G}{3} \rho. \quad (8.95)$$

The two Friedmann equations are “almost” recovered. The striking differences are that in classical mechanics the pressure does not contribute to the “gravitational mass”, and that the  $\Lambda$  term must be introduced by hand as a form of repulsive potential.

The curvature of spacetime is, in this “classical” version, associated to (minus) the total energy of the system, which can somehow be interpreted as a “binding energy”.

### 8.2.5.2 Single Component Universes

The two Friedmann equations determine, once the energy density  $\rho$  and the pressure  $\mathcal{P}$  are known, the evolution of the scale  $a(t)$  and of the curvature  $K(t)$  of the Universe. However,  $\rho$  and  $\mathcal{P}$  are nontrivial quantities, depending critically on the amount of the different forms of energy and matter that exist in the Universe at each evolution stage.

In the simplest case of a Universe with just nonrelativistic particles (ordinary baryonic matter or “cold”—i.e., nonrelativistic—dark matter) the pressure is negligible with respect to the energy density ( $\mathcal{P} \ll \rho_m c^2$ ) and the Friedmann equations can be approximated as

$$\frac{d}{dt} (\rho_m a^3) = 0 ; \left( \frac{\ddot{a}}{a} \right) = -\frac{4\pi G}{3} \rho_m. \quad (8.96)$$

Solving these equations one finds

$$\rho_m \propto \frac{1}{a^3} ; a(t) \propto t^{\frac{2}{3}}. \quad (8.97)$$

In general for a Universe with just one kind of component characterized by an equation of state relating  $\rho$  and  $\mathcal{P}$  of the type  $\mathcal{P} = \alpha\rho$ , the solutions are

$$\rho \propto a^{-3(\alpha+1)} ; a(t) \propto t^{\frac{2}{3(\alpha+1)}}. \quad (8.98)$$

For instance, in the case of a Universe dominated by relativistic particles (radiation or hot matter),  $\alpha = 1/3$ , and we obtain

$$\rho_\gamma \propto \frac{1}{a^4} ; a(t) \propto t^{\frac{1}{2}}. \quad (8.99)$$

This last relation can be interpreted by taking, as an example, a photon-dominated Universe, where the decrease in the number density of photons ( $n_\gamma \propto a^{-3}$ ) combines with a decrease in the mean photon energy ( $E_\gamma \propto a^{-1}$ ) corresponding to wavelength dilation.

### 8.2.5.3 Static Universe and Vacuum Energy Density

To model a static Universe ( $\dot{a} = 0, \ddot{a} = 0$ ) one should have:

$$\frac{K}{a^2} = \frac{8\pi G}{3} \rho ; \rho + 3\mathcal{P} = 0. \quad (8.100)$$

$K$  should then be positive ( $K = 1$ ) and  $\mathcal{P} = -\frac{1}{3}\rho$ . This requires a “new form of energy” with negative  $\alpha$ , which can be related to the “cosmological” constant term.

By reading this term on the right-hand side of Einstein’s equations, we can formally include it in the energy–momentum tensor, thus defining a “vacuum” tensor as

$$T_{\mu\nu}^{\Lambda} = g_{\mu\nu}\Lambda = \begin{pmatrix} \rho_{\Lambda} & 0 & 0 & 0 \\ 0 & -\rho_{\Lambda} & 0 & 0 \\ 0 & 0 & -\rho_{\Lambda} & 0 \\ 0 & 0 & 0 & -\rho_{\Lambda} \end{pmatrix} \tag{8.101}$$

with

$$\rho_{\Lambda} = \frac{\Lambda}{8\pi G}. \tag{8.102}$$

This implies an equation of state of the form ( $\alpha = -1$ ):

$$\mathcal{P}_{\Lambda} = -\rho_{\Lambda}. \tag{8.103}$$

Therefore, in a static Universe we would have

$$\rho = \rho_m + \rho_{\Lambda} \tag{8.104}$$

and

$$\rho_m = 2 \rho_{\Lambda}. \tag{8.105}$$

### 8.2.5.4 De Sitter Universe

In a Universe dominated by the cosmological constant ( $\rho \equiv \rho_{\Lambda}$ ), as first discussed by de Sitter,

$$\frac{d}{dt} (\rho_{\Lambda} a^3) = \rho_{\Lambda} \frac{d}{dt} (a^3) \tag{8.106}$$

and

$$H^2 + \frac{K}{a^2} = \frac{\Lambda}{3} \tag{8.107}$$

implying

$$\rho_{\Lambda} = \text{constant} ; a(t) \sim e^{Ht} \tag{8.108}$$

with

$$H = \sqrt{\frac{\Lambda}{3}}. \tag{8.109}$$

Thus the de Sitter Universe has an exponential expansion while its energy density remains constant.

### 8.2.6 Critical Density of the Universe; Normalized Densities

The curvature of the Universe depends, according to Friedmann equations, on the energy density of the Universe according to

$$\frac{K}{a^2} = \frac{8\pi G}{3}\rho - H^2. \quad (8.110)$$

Therefore, if

$$\rho = \rho_{\text{crit}} = \frac{3H^2}{8\pi G} \quad (8.111)$$

one obtains

$$K = 0$$

and the Universe is, in this case, spatially flat.

For the present value of  $H_0$  this corresponds to

$$\rho_{\text{crit}} \sim 5 \text{ GeV}/\text{m}^3, \quad (8.112)$$

i.e., less than 6 hydrogen atoms per cubic meter. The number of baryons per cubic meter one obtains from galaxy counts is, however, twenty times smaller—consistently with the result of the fit to CMB data.

#### 8.2.6.1 Normalized Densities $\Omega_i$ , $H$ , and $q_0$

The energy densities of each type of matter, radiation, and vacuum are often normalized to the critical density as follows:

$$\Omega_i = \frac{\rho_i}{\rho_{\text{crit}}} = \frac{8\pi G}{3H^2} \rho_i. \quad (8.113)$$

By defining also a normalized “curvature” energy density as

$$\Omega_K = -\frac{K}{H^2 a^2} = -\frac{K}{\dot{a}^2}, \quad (8.114)$$

the first Friedmann equation

$$\frac{8\pi G}{3H^2} \rho - \frac{K}{H^2 a^2} = 1 \quad (8.115)$$

takes then a very simple form

$$\Omega_m + \Omega_\gamma + \Omega_\Lambda + \Omega_K = 1. \quad (8.116)$$

On the other hand, it can be shown that taking into account the specific evolution of each type of density with the scale parameter  $a$ , the evolution equation for the Hubble parameter can be written as

$$H^2 = H_0^2 (\Omega_{0\Lambda} + \Omega_{0K} a^{-2} + \Omega_{0m} a^{-3} + \Omega_{0\gamma} a^{-4}), \quad (8.117)$$

where the subscripts 0 indicate the values at present time ( $t = t_0, a_0 = 1$ ).

Since the scale factor  $a$  is related to the redshift  $z$ , as discussed in Sect. 8.1.1, by

$$(1 + z) = a^{-1},$$

the Hubble evolution equation can be written as

$$H^2(z) = H_0^2 (\Omega_{0\Lambda} + \Omega_{0K}(1+z)^2 + \Omega_{0m}(1+z)^3 + \Omega_{0\gamma}(1+z)^4). \quad (8.118)$$

Finally, the deceleration parameter  $q_0$  can also be expressed as a function of the normalized densities  $\Omega_i$ . In fact  $q_0$  was defined as (Sect. 8.1.1)

$$q_0 = -\frac{\ddot{a}}{H_0^2 a}. \quad (8.119)$$

Now, using the second Friedmann equation

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3} (\rho + 3\mathcal{P})$$

and the equations of state

$$\mathcal{P}_i = \alpha_i \rho_i,$$

one obtains

$$\begin{aligned} q_0 &= -\frac{\ddot{a}}{H_0^2 a} = \frac{1}{2} \frac{8\pi G}{3H_0^2} \sum_i \rho_i (1 + 3\alpha_i) \\ q_0 &= \frac{1}{2} \sum_i \Omega_i (1 + 3\alpha_i) \\ q_0 &= \frac{1}{2} \Omega_{0m} + \Omega_{0\gamma} - \Omega_{0\Lambda}. \end{aligned} \quad (8.120)$$

These equations in  $H$  and  $q_0$  are of the utmost importance, since they connect directly the experimentally measured quantities  $H_0$  and  $q_0$  to the densities of the various energy species in the Universe.

### 8.2.6.2 Experimental Determination of the Normalized Densities

The total density of baryons, visible or invisible, as inferred from nucleosynthesis, is about 0.26 baryons per cubic meter, i.e.,

$$\Omega_b \sim 0.049 \pm 0.003. \quad (8.121)$$

A small fraction of this is luminous—i.e., visible energy.

The currently most accurate determination of the overall densities comes from global fits of cosmological parameters to recent observations (see later). Using measurements of the anisotropy of the CMB and of the spatial distribution of galaxies, as well as the measured acceleration, the data indicate a fraction of nonbaryonic DM over the total energy content of the Universe given by

$$\Omega_{DM, nonbaryonic} \sim 0.258 \pm 0.011. \quad (8.122)$$

According to the same fits, the total baryonic matter density is

$$\Omega_b \sim 0.048 \pm 0.002. \quad (8.123)$$

Part of the baryonic matter may contribute to DM in form of nonluminous Dark Matter, e.g., massive compact objects or cold molecular gas clouds (see later).

In summary, a remarkable agreement of independent astrophysical observations with cosmological global fits indicates that the energy content of DM in the Universe could be about 25% of the total energy of the Universe, compared to some 5% due to ordinary matter.

The dark energy density can be measured from the observed curvature in the Hubble plot and from the position of the “acoustic peaks” in the angular power spectrum of the temperature fluctuations in the CMB:

$$\Omega_\Lambda \sim 0.692 \pm 0.012. \quad (8.124)$$

Dark energy dominates thus the energy content of the Universe.

The Friedmann equation 8.90 can also be rewritten as

$$\Omega = \frac{\rho}{\rho_{crit}} = 1 + \frac{K}{H^2 a^2}, \quad (8.125)$$

where the *closure parameter*  $\Omega$  is the sum of  $\Omega_m$ ,  $\Omega_\gamma$  and  $\Omega_\Lambda$ , with  $\Omega_\gamma \simeq 5 \times 10^{-5}$  being negligible. This means that, in general,  $\Omega$  is a function of time, unless  $\Omega = 1$  and thus  $K = 0$  (flat Universe).

The present experimental data indicate a value

$$\Omega \sim 1.0002 \pm 0.0026 : \quad (8.126)$$

it would look very strange if this were a coincidence, unless  $\Omega$  is identically one. For this reason this fact is at the heart of the standard model of cosmology, the  $\Lambda$ CDM model, which *postulates*  $\Omega = 1$ .

### 8.2.7 Age of the Universe from the Friedmann Equations and Evolution Scenarios

The evolution of the Hubble parameter can be used to estimate the age of the Universe for different composition of the total energy density. Indeed

$$H = \frac{\dot{a}}{a} = \frac{1}{a} \frac{da}{dt} = - \left( \frac{dz/dt}{1+z} \right)$$

$$dt = - \frac{dz}{(1+z) H}$$

$$(t_0 - t) = \frac{1}{H_0} \int_0^z \frac{dz}{(1+z) (\Omega_\Lambda + \Omega_K(1+z)^2 + \Omega_m(1+z)^3 + \Omega_\gamma(1+z)^4)^{1/2}} . \quad (8.127)$$

The solution to this equation has to be obtained numerically in most realistic situations. However, in some simplified scenarios, an analytical solution can be found. In particular for matter ( $\Omega_m = 1$ ) and radiation ( $\Omega_\gamma = 1$ ) dominated Universes the solutions are, respectively,

$$t_0 = \frac{2}{3H_0} \quad (8.128)$$

and

$$t_0 = \frac{1}{2H_0} . \quad (8.129)$$

In a flat Universe with matter and vacuum energy parameters close to the ones presently measured ( $\Omega_m = \Omega_{DM,nonbaryonic} + \Omega_B \simeq 0.3$ ,  $\Omega_\Lambda \simeq 0.7$ ) we obtain

$$t_0 \sim \frac{0.96}{H_0} . \quad (8.130)$$

#### 8.2.7.1 Evolution Scenarios

Friedmann equations have four independent parameters which can be chosen as:

- the present value of the Hubble parameter,  $H_0$ ;
- the present value of the energy density of radiation,  $\Omega_\gamma$  (we shall omit the subscript 0);
- the present value of the energy density of matter,  $\Omega_m$ ;
- the present value of the energy density of vacuum,  $\Omega_\Lambda$ .

If we know these parameters, the geometry and the past and future evolutions of the Universe are determined provided the dynamics of the interactions, annihilations and creations of the different particle components (see Sect. 8.3.1) are neglected. The solutions to these equations, in the general multicomponent scenarios, cannot be expressed in closed, analytical form, and require numerical approaches.

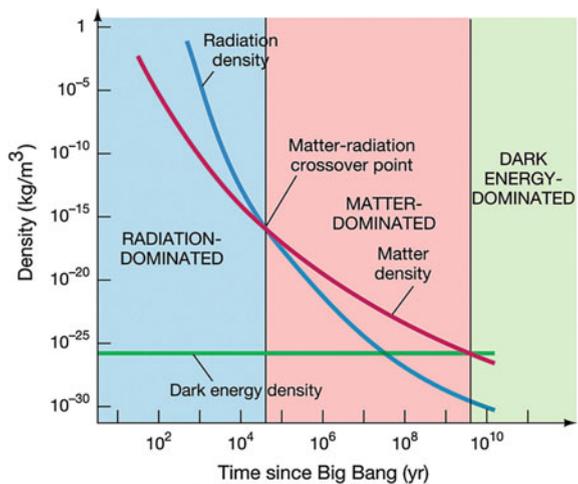
However, as we have discussed above, the evolution of the energy density of the different components scales with different powers of the scale parameter of the Universe  $a$ . Therefore, there are “eras” where a single component dominates. It is then reasonable to suppose that, initially, the Universe was radiation dominated (apart for a very short period where it is believed that inflation occurred—see Sect. 8.3.2), then that it was matter dominated and finally, at the present time, that it is the vacuum energy (mostly “dark” energy, i.e., not coming from quantum fluctuations of the vacuum of the known interactions) that is starting to dominate (Fig. 8.22).

The crossing point ( $a = a_{cross}$ ) between the matter and radiation ages can be obtained, in first approximation, by just equalizing the corresponding densities:

$$\Omega_\gamma(a_{cross}) = \Omega_m(a_{cross})$$

$$\Omega_\gamma(a_0) \left(\frac{a_{cross}}{a_0}\right)^{-4} = \Omega_m(a_0) \left(\frac{a_{cross}}{a_0}\right)^{-3}$$

**Fig. 8.22** The different ages the Universe passed through since the Big Bang (from [http://scienceblogs.com/startswithabang/files/2013/06/AT\\_7e\\_Figure\\_27\\_01.jpeg](http://scienceblogs.com/startswithabang/files/2013/06/AT_7e_Figure_27_01.jpeg), © 2011 Pearson education, Pearson Addison Wesley)



$$\left(\frac{a_{cross}}{a_0}\right)^{-1} = 1 + z_{cross} = \frac{\Omega_m(a_0)}{\Omega_\gamma(a_0)}. \quad (8.131)$$

The time after the Big Bang when this crossing point occurs can approximately be obtained from the evolution of the scale factor in a radiation dominated Universe

$$a_{cross} \sim \left(2 H_0 \sqrt{\Omega_\gamma(a_0)} t_{cross}\right)^{\frac{1}{2}}, \quad (8.132)$$

or

$$t_{cross} \sim a_{cross}^2 \left(2 H_0 \sqrt{\Omega_\gamma(a_0)}\right)^{-1}. \quad (8.133)$$

Using the current best-fit values for the parameters (see Sect. 8.4), we obtain

$$z_{cross} \sim 3\,200 \implies t_{cross} \sim 7 \times 10^4 \text{ years}. \quad (8.134)$$

After this time (i.e., during the large majority of the Universe evolution) a two-component (matter and vacuum) description should be able to give a reasonable, approximate description. In this case the geometry and the evolution of the Universe are determined only by  $\Omega_m$  and  $\Omega_\Lambda$ . Although this is a restricted parameter phase space, there are several different possible evolution scenarios as shown in Fig. 8.23:

1. If  $\Omega_m + \Omega_\Lambda = 1$  the Universe is flat but it can expand forever ( $\Omega_\Lambda > 0$ ) or eventually recollapses ( $\Omega_\Lambda < 0$ ).
2. If  $\Omega_m + \Omega_\Lambda > 1$  the Universe is closed (positive curvature).
3. If  $\Omega_m + \Omega_\Lambda < 1$  the Universe is open (negative curvature).
4. In a small phase space region with  $\Omega_m + \Omega_\Lambda > 1$  and  $\Omega_\Lambda > 0$  there is a solution for which the Universe bounces between a minimum and a maximum scale factor.

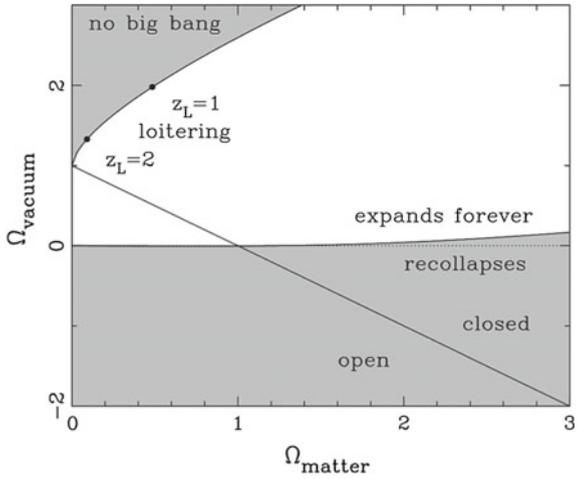
Some of these evolution scenarios are represented as functions of time in Fig. 8.24 for selected points in the parameter space discussed above. The green curve represents a flat, matter-dominated, critical density Universe (the expansion rate is slowing down forever). The blue curve shows an open, low density, matter-dominated Universe (the expansion is slowing down, but not as much). The orange curve shows a closed, high-density Universe (the expansion reverts to a “big crunch”). The red curve shows a Universe with a large fraction of “dark energy” (the expansion of the Universe accelerates).

The present experimental evidence (see Sect. 8.4) highly favors the “dark energy” scenario, leading to a cold thermal death of the Universe.

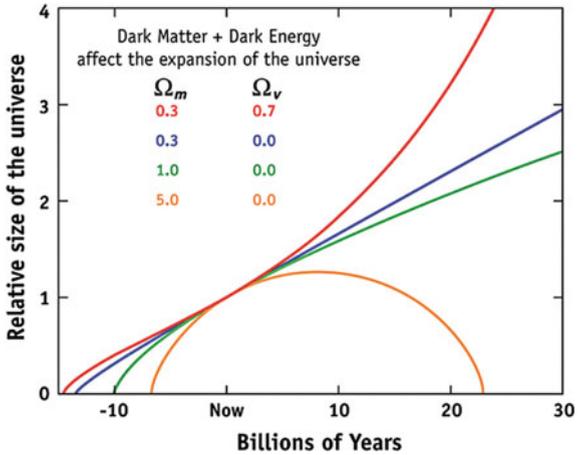
### 8.2.8 Black Holes

The first analytical solution of Einstein’s equations was found in 1915, just a month after the publication of Einstein’s original paper, by Karl Schwarzschild, a German

**Fig. 8.23** Different scenarios for the expansion of the Universe. The Hubble constant was fixed to  $H_0 = 70$  (km/s)/Mpc. From J.A. Peacock, “Cosmological Physics”, Cambridge University Press 1998



**Fig. 8.24** Evolution of the Universe in a two-component model (matter and vacuum) for different  $(\Omega_m, \Omega_\Lambda)$  values. (from [http://map.gsfc.nasa.gov/universe/uni\\_fate.html](http://map.gsfc.nasa.gov/universe/uni_fate.html))



physicist who died one year later from a disease contracted on the First World War battlefield.

Schwarzschild’s solution describes the gravitational field in the vacuum surrounding a single, spherical, nonrotating massive object. In this case the space–time metric (called the Schwarzschild metric) can be expressed as

$$ds^2 = \left(1 - \frac{r_S}{r}\right) c^2 dt^2 - \left(1 - \frac{r_S}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2 \theta d\phi^2), \quad (8.135)$$

with

$$r_S = \frac{2GM}{c^2} \simeq 2.7 \text{ km} \frac{M}{M_\odot}. \quad (8.136)$$

In the weak field limit,  $r \rightarrow \infty$ , we recover flat spacetime. According to this solution, a clock with period  $\tau^*$  placed at a point  $r$  is seen by an observer placed at  $r = \infty$  with a period  $\tau$  given by:

$$\tau = \left(1 - \frac{r_S}{r}\right)^{-1} \tau^*. \quad (8.137)$$

In the limit  $r \rightarrow r_S$  (the Schwarzschild radius) the metric shows a coordinate singularity: the time component goes to zero and the radial component goes to infinity. From the point of view of an asymptotic observer, the period  $\tau^*$  is seen now as infinitely large. No light emitted at  $r = r_S$  is able to reach the  $r > r_S$  world. This is what is usually called, following John Wheeler, a “black hole”.

The existence of objects so massive that light would not be able to escape from them, was already predicted in the end of the eighteenth century by Michell in England and independently by Laplace in France. They just realized that, if the escape velocity from a massive object would have been greater than the speed of light, then the light could not escape from the object:

$$v_{\text{esc}} = \sqrt{\frac{2GM}{r}} > c. \quad (8.138)$$

Thus an object with radius  $R$  and a mass  $M$  would be a “black hole” if:

$$M > \frac{Rc^2}{2G} : \quad (8.139)$$

the “classical” radius and the Schwarzschild radius coincide.

The singularity observed in the Schwarzschild metric is not in fact a real physics singularity; it depends on the reference frame chosen (see [F8.3] for a discussion). An observer in free-fall frame will cross the Schwarzschild surface without feeling any discontinuity; (s)he will go on receiving signals from the outside world but (s)he will not be able to escape from the unavoidable, i.e., from crunching, at last, at the center of the black hole (the real physical singularity).

Schwarzschild black holes are however just a specific case. In 1963, New Zealand mathematician Roy Kerr found an exact solution to Einstein’s equations for the case of a rotating noncharged black hole and two years later the US Ezra Newman extended it to the more general case of rotating charged black holes. In fact it can be proved that a black hole can be completely described by three parameters: mass, angular momentum, and electric charge (the so-called no-hair theorem).

Black holes are not just exotic solutions of the General Theory of Relativity. They may be formed either by gravitational collapse or particle high-energy collisions. While so far there is no evidence of their formation in human-made accelerators, there is striking indirect evidence that they are part of several binary systems and that they are present in the center of most galaxies, including our own (the Milky

Way hosts in its center a black hole of roughly 4 million solar masses, as determined from the orbit of nearby stars). Extreme high-energy phenomena in the Universe, generating the most energetic cosmic rays, may also be caused by supermassive black holes inside AGN (Active Galactic Nuclei—see Chap. 10).

### 8.2.9 Gravitational Waves

Soon after the discovery of the electromagnetic radiation the existence of gravitational waves was suggested. The analogy was appealing but it took a long way before a firm prediction by Einstein and only very recently the direct experimental detection has been possible (see Chap. 10). According to Einstein's equations the structure of spacetime is determined by the energy-momentum distributions but the solutions of such equations are far from being trivial. In particular, in the case of gravitational waves, where the components of the spacetime metric have to be time dependent (contrary for instance to the cases discussed in the previous section where the metric was assumed to be static), general exact analytic solutions are, still nowadays, impossible to obtain.

The spacetime metric out of the gravitational sources is basically flat and small perturbation of the metric components may be considered (linearized gravity). Let us then write the space metric in free space (weak field approximation) as:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (8.140)$$

where  $\eta_{\mu\nu}$  is the Minkowski metric and  $h_{\mu\nu} \ll 1$  for all  $\mu, \nu$ .

Choosing the appropriate coordinate system, the “transverse traceless” (TT) gauge (for a detailed discussion see for example [F8.6]), Einstein's equations in vacuo can, in this approximation, be written as:

$$\left( \frac{\partial^2}{\partial t^2} - \nabla^2 \right) h_{\mu\nu} = 0 \text{ or more briefly } \square h_{\mu\nu} = 0. \quad (8.141)$$

This is a wave equation whose simplest solutions are plane waves:

$$h_{\mu\nu} = A_{\mu\nu} e^{ik_a x^a}, \quad (8.142)$$

where  $A_{\mu\nu}$  and  $k_a$  are respectively the wave amplitude and the wave vector. These waves are transverse,

$$A_{\mu\nu} k^\mu = 0, \quad (8.143)$$

and they propagate along light rays, i.e.,  $k_a$  is a null-vector:

$$k_{\mu}k^{\mu} = 0. \quad (8.144)$$

Their propagation velocity is thus the light velocity  $c$  (remember that  $c = 1$  in the metric we have chosen), which is a non trivial result. These equations were derived directly from Einstein's equations.

Assuming a propagation along the  $z$ -axis with an energy  $w$ :

$$k_{\mu} = (w, 0, 0, w), \quad (8.145)$$

it can be shown that, in this gauge, only four components of  $A_{\mu\nu}$  ( $A_{xx} = -A_{yy}$ ;  $A_{xy} = A_{yx}$ ), may be nonzero. The general solution for a propagation along the  $z$  axis with fixed frequency  $w$  can thus be written as:

$$h_{\mu\nu}(z, t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{xx} & A_{xy} & 0 \\ 0 & A_{xy} & -A_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} e^{iw(z-t)}. \quad (8.146)$$

Then, whenever  $A_{xy} = 0$ , the space-time metric produced by such a wave is given by:

$$ds^2 = dt^2 - ((1 + h_+)dx^2) + (1 - h_+)dy^2 + dz^2 \quad (8.147)$$

with

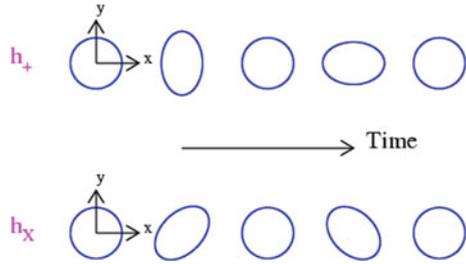
$$h_+ = A_{xx}e^{iw(z-t)}. \quad (8.148)$$

The effects of this wave in the transverse space axes  $x$  and  $y$  are opposite: while one expands, the other contracts and vice-versa. For instance, this gravitational wave would change the distance  $L$  between two masses placed on the  $x$  axis by  $dL = L h_+$ . This wave is said to be “plus” polarized (denoted by  $+$ ). On the other hand, if  $A_{xx} = 0$  a similar effect may be observed for axis rotated by  $45^\circ$  and then the wave is said to be “cross” polarized (denoted by  $\times$ ). Such effects are graphically represented in Fig. 8.25.

The amplitudes of such effects are however quite tiny if the sources are quite far ( $h_+$  is proportional to  $1/R$  where  $R$  is the distance to the source). The relative change of the distance between two tests masses at Earth, the *strain*, which is the variable measured by gravitational wave detectors (see Sect. 4.6), is of the order of  $10^{-23}$  for the Hulse–Taylor binary pulsar and of  $10^{-21}$  for the coalescence of a binary stellar-mass black hole system (see Sect. 10.4.4).

In summary, gravitational waves are “ripples in space-time” propagating in free space with the velocity of light and their effect on the relative distances between free mass particles have been detected, as it will be discussed in Chap. 10.

**Fig. 8.25** Graphical representation of the effects of polarized waves (top, + polarization; bottom, × polarization) From K. Riles, “Gravitational Waves: Sources, Detectors and Searches”, Prog. Part. Nucl. Phys. 68 (2013) 1



### 8.3 Past, Present, and Future of the Universe

#### 8.3.1 Early Universe

In its “first” moments the Universe, according to the Big Bang model, was filled with a high-density, hot (high-energy) gas of relativistic particles at thermal equilibrium. The assumption of thermal equilibrium is justified since the interaction rate per particle  $\Gamma$  ( $\Gamma = n\sigma v$ , where  $n$  is the number density,  $\sigma$  is the cross section, and  $v$  is the relative velocity) and the Hubble parameter  $H$  ( $H^2 \sim 8\pi G\rho/3$ ) evolve with the energy density  $\rho$  as

$$\Gamma \propto n \propto \rho ; H \propto \rho^{\frac{1}{2}} . \tag{8.149}$$

Thus at some point, going back in time, we should have had

$$\frac{\Gamma}{H} \gg 1 . \tag{8.150}$$

Since the early Universe was radiation dominated (Sect. 8.2),

$$\rho_\gamma \propto \frac{1}{a^4} ; a(t) \propto t^{\frac{1}{2}} . \tag{8.151}$$

The temperature is, by definition, proportional to the mean particle energy and thus, in the case of radiation, it increases proportionally to the inverse of the Universe scale:

$$T \propto a^{-1} . \tag{8.152}$$

On the other hand, at a temperature  $T$  the number density, the energy density, and the pressure of each particle type can be calculated (neglecting chemical potentials) by standard quantum statistical mechanics:

$$n_i = \frac{g_i}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi p^2}{e^{\left(\frac{e_i}{k_B T}\right)} \pm 1} dp , \tag{8.153}$$

$$\rho_i c^2 = \frac{g_i}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi p^2}{e^{\left(\frac{e_i}{k_B T}\right)} \pm 1} e_i dp, \quad (8.154)$$

$$\mathcal{P}_i = \frac{g_i}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi p^2}{e^{\left(\frac{e_i}{k_B T}\right)} \pm 1} \frac{p_i c^2}{3 E_i} dp, \quad (8.155)$$

where  $g_i$  are the internal degrees of freedom of the particles—the  $+/-$  signs are for bosons (Bose–Einstein statistics) and fermions (Fermi–Dirac statistics), respectively.

For  $k_B T \gg m_i c^2$  (relativistic limit)

$$n_i = \begin{cases} g_i \frac{\zeta(3)}{\pi^2} \left(\frac{k_B T}{\hbar c}\right)^3, & \text{for bosons} \\ \frac{3}{4} \left[ g_i \frac{\zeta(3)}{\pi^2} \left(\frac{k_B T}{\hbar c}\right)^3 \right], & \text{for fermions} \end{cases} \quad (8.156)$$

$$\rho_i c^2 = \begin{cases} g_i \frac{\pi^2}{30} k_B T \left(\frac{k_B T}{\hbar c}\right)^3, & \text{for bosons} \\ \frac{7}{8} \left[ g_i \frac{\pi^2}{30} k_B T \left(\frac{k_B T}{\hbar c}\right)^3 \right], & \text{for fermions} \end{cases} \quad (8.157)$$

$$\mathcal{P}_i = \frac{\rho_i c^2}{3},$$

where  $\zeta$  is the Riemann zeta function ( $\zeta(3) \simeq 1.20206$ ).

For a nonrelativistic particle with  $m_x c^2 \sim k_B T$  the classical Maxwell–Boltzmann distribution is recovered:

$$n_x = g_x \left( \frac{m_x k_B T}{2\pi\hbar^2} \right)^{\frac{3}{2}} e^{-\left(\frac{m_x c^2}{k_B T}\right)}.$$

The total energy density in the early Universe can be obtained summing over all possible relativistic particles and can be written as

$$\rho c^2 = g_{ef}^* \frac{\pi^2}{30} k_B T \left(\frac{k_B T}{\hbar c}\right)^3, \quad (8.158)$$

where  $g_{ef}^*$  is defined as the total “effective” number of degrees of freedom and is given by

$$g_{ef}^* = \sum_{\text{bosons}} g_i + \frac{7}{8} \sum_{\text{fermions}} g_j. \quad (8.159)$$

However, the interaction rate of some relativistic particles (like neutrinos, see below) may become at some point smaller than the expansion rate of the Universe and then they will be no more in thermal equilibrium with the other particles. It is said that they *decouple* and their temperature will evolve as  $a^{-1}$  independently of the temperature of the other particles. The individual temperatures  $T_i, T_j$  may be introduced in the

definition of the “effective” number of degrees of freedom as

$$g_{ef} = \sum_{\text{bosons}} g_i \left( \frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_j \left( \frac{T_j}{T} \right)^4 \quad (8.160)$$

( $g_{ef}$  is of course a function of the age of the Universe). At a given time all the particles with  $M_x c^2 \ll k_B T$  contribute.

The total energy density determines the evolution of the Hubble parameter

$$H^2 \sim \frac{8\pi G}{3} \frac{\pi^2}{30} g_{ef} k_B T \left( \frac{k_B T}{\hbar c} \right)^3 \quad (8.161)$$

$$H \sim \left( \frac{4\pi^3 G}{45 (\hbar c)^3} \right)^{1/2} \sqrt{g_{ef}} (k_B T)^2, \quad (8.162)$$

or, introducing the Planck mass (Sect. 2.10),

$$H \sim 1.66 \sqrt{g_{ef}} \frac{(k_B T)^2}{\hbar c^2 m_P}. \quad (8.163)$$

Remembering (Sect. 8.2) that in a radiation dominated Universe the Hubble parameter is related to time just by

$$H = \frac{1}{2t}, \quad (8.164)$$

time and temperature are related by

$$t = \left( \frac{45 (\hbar c)^3}{16 \pi^3 G} \right)^{1/2} \frac{1}{\sqrt{g_{ef}}} \frac{1}{(k_B T)^2}, \quad (8.165)$$

or using standard units

$$t = \frac{2.4}{\sqrt{g_{ef}}} \left( \frac{1 \text{ MeV}}{k_B T} \right)^2 \text{ s}, \quad (8.166)$$

which is a kind of rule of thumb formula for the early Universe.

Finally, the expansion of the Universe is assumed to be adiabatic. In fact there is, by definition, no outside system and the total entropy is much higher than the small variations due to irreversible processes. The entropy of the early Universe can then be assumed to be constant.

Remembering that the entropy  $S$  can be defined as

$$S = \frac{(\rho c^2 + \mathcal{P})}{k_B T} V, \quad (8.167)$$

the entropy density  $s$  is then given by:

$$s = \frac{\rho c^2 + \mathcal{P}}{k_B T}. \quad (8.168)$$

Summing over all possible particle types

$$s = g_{ef}^s \frac{2\pi^2}{45} \left( \frac{k_B T}{\hbar c} \right)^3, \quad (8.169)$$

where  $g_{ef}^s$  is defined similarly to  $g_{ef}$  as

$$g_{ef}^s = \sum_{\text{bosons}} g_i \left( \frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_j \left( \frac{T_j}{T} \right)^3. \quad (8.170)$$

At  $k_B T \sim 1$  TeV ( $T \sim 10^{16}$  K,  $t \sim 10^{-12}$  s) all the standard model particles should contribute. In the SM there are six different types of bosons ( $\gamma$ ,  $W^\pm$ ,  $Z$ ,  $g$ ,  $H^0$ ) and 24 types of fermions and antifermions (quarks and leptons): thus the total “effective” number of degrees of freedom is

$$g_{ef}^* = 106.75. \quad (8.171)$$

At early times the possibility of physics beyond the standard model (Grand Unified Theories (GUT) with new bosons and Higgs fields, SuperSymmetry with the association to each of the existing bosons or fermions of, respectively, a new fermion or boson, ...) may increase this number. The way up to the Planck time ( $\sim 10^{-43}$  s), where general relativity meets quantum mechanics and all the interactions may become unified, remains basically unknown. Quantum gravity theories like string theory or loop quantum gravity have been extensively explored in the last years, but, for the moment, are still more elegant mathematical constructions than real physical theories. The review of such attempts is out of the scope of the present book; only the decoupling of a possible stable heavy dark matter particle will be discussed in the following.

At later times the temperature decreases, and  $g_{ef}$  decreases as well. At  $k_B T \sim 0.2$  GeV hadronization occurs and quarks and gluons become confined into massive hadrons. At  $k_B T \sim 1$  MeV ( $t \sim 1$  s) the light elements are formed (primordial nucleosynthesis, see Sect. 8.1.4). Around the same temperature neutrinos also decouple as it will be discussed below. At  $k_B T \sim 0.8$  eV the total energy density of nonrelativistic particles is higher than the total energy density of relativistic particles and then the Universe enters a matter-dominated era (see Sect. 8.2.5). Finally at  $k_B T \sim 0.3$  eV recombination and decoupling occur (see Sect. 8.1.3). At that moment, the hot plasma of photons, baryons, and electrons, which was coherently oscillating under the combined action of gravity (attraction) and radiation pressure (repulsion), breaks apart: photons propagate away originating the CMB while the baryon oscillations stop (no more radiation pressure) leaving a density excess at a fixed radius (the sound horizon) which, convoluted with the initial density fluctuations, are the seeds for the

subsequent structure formation. This entire evolution scenario is strongly constrained by the existence of dark matter which is gravitationally coupled to baryons.

### 8.3.1.1 Neutrino Decoupling and $e^+ e^-$ Annihilations

Decoupling (also called freeze-out) of neutrinos occurs, similarly to what was discussed in Sect. 8.1.4 in the case of the primordial nucleosynthesis, whenever the neutrinos interaction rate  $\Gamma_\nu$  is of the order of the expansion rate of the Universe

$$\Gamma_\nu \sim H.$$

Neutrinos interact just via weak interactions (like  $\nu e^- \rightarrow \nu e^-$ ) and thus their cross sections have a magnitude for  $k_B T \sim \sqrt{s} \ll m_W$  of the order of

$$\sigma \sim G_F^2 s \sim G_F^2 (k_B T)^2. \quad (8.172)$$

The neutrino interaction rate  $\Gamma_\nu$  is proportional to  $T^5$  ( $\Gamma_\nu = n\sigma v$  and  $n \propto T^3$ ,  $v \sim c$ ) while  $H$ , as seen above (Eq. 8.166), is proportional to  $T^2$ . Therefore, there will be a crossing point, which, indeed, occurs for temperatures around a few MeV.

Before decoupling photons and neutrinos have the same temperature. But from this point on, neutrinos will have basically no more interactions and, thus, their temperature decreases just with  $a^{-1}$ , while photons are still in thermal equilibrium with a plasma of electrons and positrons through production ( $\gamma\gamma \rightarrow e^+ e^-$ ) and annihilation ( $e^+ e^- \rightarrow \gamma\gamma$ ) reactions. For temperatures below 1 MeV this equilibrium breaks down as the production reaction is no more possible ( $m_e c^2 \sim 0.5$  MeV). However, entropy should be conserved and therefore

$$g_{ef}^s T^3 = \text{constant}; \quad g_{ef}^{e\gamma} T_{e\gamma}^3 = g_{ef}^\gamma T_\gamma^3.$$

Before decoupling

$$g_{ef}^{e\gamma} = 2 \times 2 \times \frac{7}{8} + 2 = \frac{11}{2} \quad (8.173)$$

and after decoupling

$$g_{ef}^\gamma = 2. \quad (8.174)$$

Therefore,

$$\frac{T_\gamma}{T_{e\gamma}} \sim \left(\frac{11}{4}\right)^{1/3} \simeq 1.4. \quad (8.175)$$

The temperature of photons after the annihilation of the electrons and positrons is thus higher than the neutrino temperature at the same time (the so-called *reheating*).

The temperature of the cosmic neutrino background is therefore nowadays around 1.95 K, while the temperature of the CMB is around 2.73 K (see Sect. 8.1.3). The ratio

between the number density of cosmic background neutrinos (and antineutrinos) and photons can then be computed using Eq. 8.158 as:

$$\frac{N_\nu}{N_\gamma} = 3 \frac{3}{11},$$

where the factor 3 takes into account the existence of three relativistic neutrino families. Reminding that nowadays  $N_\gamma \simeq 410/\text{cm}^3$ , the number density of cosmological neutrinos should be:

$$N_\nu \simeq 340/\text{cm}^3.$$

The detection of such neutrinos (which are all around) remains an enormous challenge to experimental particle and astroparticle physicists.

### 8.3.2 Inflation and Large-Scale Structures

The early Universe should have been remarkably flat, isotropic, and homogeneous to be consistent with the present measurements of the total energy density of the Universe (equal or very near of the critical density) and with the extremely tiny temperature fluctuations ( $\sim 10^{-5}$ ) observed in the CMB. On the contrary, at scales  $\sim 50$  Mpc, the observed Universe is filled with rather inhomogeneous structures, like galaxies, clusters, superclusters, and voids. The solution for this apparent paradox was given introducing in the very early Universe an exponential superluminal (recessional velocities much greater than the speed of light) expansion. It is the, so-called, *inflation*.

#### 8.3.2.1 The Inflaton Field

The possibility of a kind of exponential expansion was already discussed above in the framework of a Universe dominated by the cosmological constant (Sect. 8.2). The novelty was to introduce a mechanism that could, for a while, provide a vacuum energy density and a state equation ( $\mathcal{P} = \alpha\rho$ , with  $\alpha < -1/3$ ) ensuring thus the necessary negative pressure. A scalar field filling the entire Universe (the “inflaton field”) can serve these purposes.

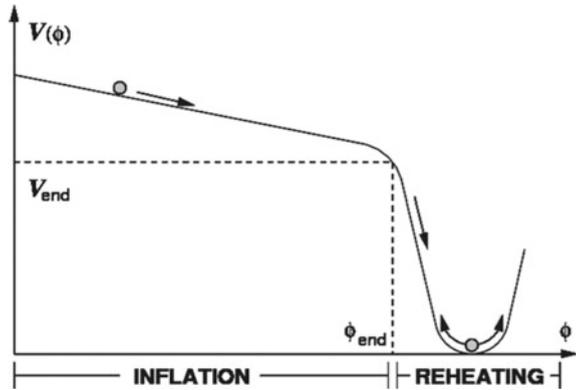
In fact the energy density and the pressure of a scalar field  $\phi(t)$  with an associated potential energy  $V(\phi)$  are given by (For a discussion see [F 8.5]):

$$\rho = \frac{1}{2} \frac{1}{\hbar c^3} \dot{\phi}^2 + V(\phi); \quad \mathcal{P} = \frac{1}{2} \frac{1}{\hbar c^3} \dot{\phi}^2 - V(\phi). \quad (8.176)$$

Thus whenever

$$\frac{1}{2\hbar c^3} \dot{\phi}^2 < V(\phi)$$

**Fig. 8.26** An example of an inflaton potential. (from K. Dimopoulos, J. Phys.: Conf. Ser. 283 012010 (doi:10.1088/1742-6596/283/1/012010))



an exponential expansion occurs. This condition is satisfied by a reasonably flat potential, like the one sketched in Fig. 8.26.

In the first phase the inflaton field rolls down slowly starting from a state of false vacuum ( $\phi = 0, V(\phi) \neq 0$ ) and inflation occurs. In this case the inflation period ends when the potential changes abruptly its shape going to a minimum (the true vacuum). The field then oscillates around this minimum dissipating its energy: this process will refill the empty Universe originated by the exponential expansion with radiation (*reheating*), which will then be the starting point of a “classical” hot big bang expansion.

During the inflation period a superluminal expansion thus occurs

$$a(t) \sim e^{Ht} \tag{8.177}$$

with (see Sect. 8.2.5)

$$H \sim \sqrt{\frac{8\pi G}{3c^2} \rho}.$$

In this period the scale factor grows as

$$\frac{a(t_f)}{a(t_i)} \sim e^N, \tag{8.178}$$

with  $N$  (the number of  $e$ -foldings, i.e., of expansions by a factor of  $e$ ), typically, of the order of  $10^2$ .

### 8.3.2.2 Flatness, Horizon, and Monopole Problems

The energy density evolves with (see Sect. 8.2.5)

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

Then, at the Planck time, the energy density will be very close to the critical density as it was predicted extrapolating back the present measured energy density values to the early moments of the Universe (the so-called flatness problem). For example, at the epoch of the primordial nucleosynthesis ( $t \sim 1$  s) the deviation from the critical density should be  $\lesssim 10^{-12} - 10^{-16}$ .

The exponential expansion will also give a solution to the puzzle that arises from the observations of the extreme uniformity of the CMB temperature measured all over the sky (the so-called horizon problem).

In the standard Big Bang model the horizon distance (the maximum distance light could have traveled since the origin of time) at last scattering ( $t_{ls} \sim 3 \cdot 10^5$  years,  $z_{ls} \sim 1100$ ) is given by

$$d_H = a(t_{ls}) \int_0^{t_{ls}} \frac{c dt}{a(t)}. \quad (8.179)$$

If there was no expansion  $d_H$  would be as expected to be just  $d_H = c t_{ls}$ .

Basically, the horizon distance is just a consequence of the finite speed of light (which also solves the Olbers' paradox as referred in Sect. 8.1.6).

In a similar way the proper distance from last scattering to the present ( $t_0 \sim 14$  Gyr) is given by

$$D = a(t_0) \int_{t_{ls}}^{t_0} \frac{c dt}{a(t)}. \quad (8.180)$$

In the Big Bang model there was (see Sect. 8.2.5) first a radiation dominated expansion followed by a matter-dominated expansion with scale parameter evolution  $a(t) \propto t^{\frac{1}{2}}$  and  $a(t) \propto t^{\frac{2}{3}}$ , respectively. The crossing point was computed to be around  $t_{cross} \sim 7 \times 10^4$  years.

Then, assuming that during most of the time the Universe is matter dominated (the correction due to the radiation dominated period is small),

$$d_H \sim 3 t_{ls},$$

$$D \sim 3 t_0.$$

The regions causally connected at the time of last scattering (when the CMB photons were emitted) as seen by an observer on Earth have an angular size of

$$\delta\theta \sim \frac{d_H}{D} (1 + z_{ls}) \frac{180^\circ}{\pi} \sim 1^\circ - 2^\circ, \quad (8.181)$$

where the  $(1 + z_{ls})$  factor accounts for the expansion between the time of last scattering and the present.

Regions separated by more than this angular distance have in the standard Big Bang model no way to be in thermal equilibrium. Inflation, by postulating a superluminal expansion at a very early time, ensures that the entire Universe that we can now observe was causally connected in those first moments before inflation.

Finally, according to the Big Bang picture, at the very early moments of the Universe all the interactions should be unified. When later temperature decreases, successive phase transitions, due to spontaneous symmetry breaking, originated the present world we live in, in which the different interactions are well individualized. The problem is that Grand Unified Theories (GUT) phase transition should give rise to a high density of magnetic monopoles. Although none of these monopoles were ever observed (the so-called “monopole problem”), if inflation had occurred just after the GUT phase transition the monopoles (or any other possible relics) would be extremely diluted and this problem would be solved.

It is then tempting to associate the inflaton field to some GUT breaking mechanism, but it was shown that potentials derived from GUTs do not work, and for this reason the inflaton potential is still, for the moment, an empirical choice.

### 8.3.2.3 Structure Formation

Nowadays, the most relevant and falsifiable aspect of inflationary models is their predictions for the origin and evolution of the structures that are observed in the present Universe.

Quantum fluctuations of the inflaton field originate primeval density perturbations at all distance scales. During the inflationary period all scales that can be observed today went out of the horizon (the number of  $e$ -foldings is set accordingly) to reenter later (starting from the small scales and progressively moving to large scales) during the classical expansion (the horizon grows faster than the Universe scale). They evolve under the combined action of gravity, pressure, and dissipation, giving rise first to the observed acoustic peaks in the CMB power spectrum and, finally, to the observed structures in the Universe.

The spatial density fluctuations are usually decomposed into Fourier modes labeled by their wave number  $k$  or by their wavelength  $\lambda = 2\pi/k$ , and

$$\frac{\delta\rho}{\rho}(\mathbf{r}) = A \int_{-\infty}^{\infty} \delta_k e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k.$$

Each distance scale corresponds then to a density fluctuation wave characterized by amplitude and dispersion. Generic inflationary models predict density perturbations that are adiabatic (the perturbations in all particle species are similar if they are originated by one single field), Gaussian (the amplitudes follow a Gaussian probability distribution), and obeying a scalar power law spectrum of the type

$$\langle |\delta_k|^2 \rangle \sim A_s k^{n_s-1}.$$

If  $n_s = 1$  (Harrison–Zel’dovich spectrum) the amplitudes in the corresponding gravitational potential are equal at all scales.

This power spectrum is distorted (in particular for high  $k$ , i.e., small scales) as each scale mode will reenter the horizon at a different moment and thus will evolve differently.

In the radiation-dominated phase baryonic matter is coupled to photons and thus the density perturbation modes that had reentered the horizon cannot grow due the existence of a strong radiation pressure which opposes gravity (the sound speed is high and therefore the Jeans scale, at which one would expect collapse, is greater than the horizon). These perturbations on very small scales can be strongly (or even completely) suppressed, while on larger scales a pattern of acoustic oscillations is built up.

At recombination baryons and photons decouple, the sound speed decreases dramatically, and the Jeans scale goes to zero (there is no more photon pressure to sustain gravitation). The baryonic density perturbations will then grow by coalescing onto already formed DM halos.

The regions with matter overdensities at recombination will originate cold spots in the CMB. In fact, it can be shown that due to the combined action of the perturbed gravitational potential  $\phi$  and the Doppler shift, the temperature fluctuations at each point in space are proportional to the gravitational potential

$$\frac{\delta T}{\langle T \rangle} \cong \frac{1}{3} \Delta \phi.$$

Then the pattern of density acoustic oscillations at recombination remains imprinted in the CMB power spectrum, with the positions and amplitudes of the observed peaks strongly correlated with the Universe model parameters. For instance the position of the first peak is, as it was discussed in Sect. 8.1.3, a measurement of the size of the sound horizon at recombination, and thus strongly constrains the curvature of the Universe, while its amplitude depends on the baryon/photon ratio.

The pattern of the density oscillations at recombination should be also somehow imprinted in the matter distribution in the Universe, what starts to be revealed by the observation of the Baryon Acoustic Oscillations (see Sect. 8.1.1).

Dark matter is, by definition, not coupled to photons and therefore it is not subject to any dramatic change at recombination time. Whenever dark matter became cold (nonrelativistic) associated density perturbations could start to grow and build gravitational potential wells that have, then, been “filled” by baryons after recombination and boosted the formation of gravitational structures. The relative proportion of hot (for instance neutrinos) and cold (for instance WIMPs) dark matter may lead to different scenarios for the formation of large-scale structures. In presence of hot dark matter, a top to bottom formation scenario (from superclusters to galaxies) is favored, while in a cold dark matter (CDM) scenario, it is just the contrary: this second case is in agreement with observational evidence of the existence of supernovae almost as old as the Universe.

## 8.4 The $\Lambda$ CDM Model

The  $\Lambda$ CDM model, also denominated as the *concordance model* or the *Standard Model of Cosmology*, is a parametrization of the Big Bang cosmological model based on general relativity with a reduced set of parameters. We can assume the evolution of the Universe under GR to be represented through the first Friedmann equation

$$\boxed{H^2 = \frac{8\pi G}{3} \rho + \frac{\Lambda}{3}} - \frac{K}{a^2} \quad (8.182)$$

$K$  being the curvature of space and  $\rho$  the density. The  $\Lambda$ CDM model postulates that we live in a flat Universe ( $K = 0$  and  $\Omega_m + \Omega_\gamma + \Omega_\Lambda = 1$ ) with  $\Omega_m = \Omega_b + \Omega_c$ ,  $\Omega_b$  being the baryonic density and  $\Omega_c$  the cold dark matter (CDM) density. The Universe is dominated by dark energy in the form of a nonzero cosmological constant  $\Lambda$  and cold dark matter, CDM. The  $\Lambda$ CDM model also assumes homogeneity, isotropy, and a power law spectrum of primordial fluctuations. It is the simplest model describing the existence and structure of the CMB, of the large-scale structure in the distribution of galaxies, of the abundances of nucleons, of the accelerating expansion of the universe.

The assumption that ( $\Omega_m + \Omega_\gamma + \Omega_\Lambda = 1$ ) is motivated by the fact that observations are consistent with this value with extreme accuracy. Indeed

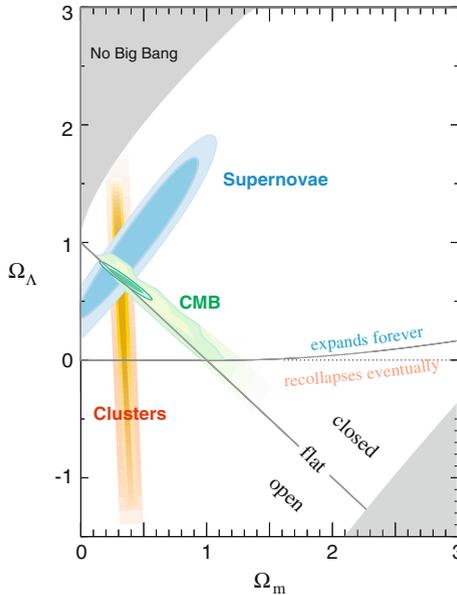
$$\Omega_m + \Omega_\gamma + \Omega_\Lambda = 1.0002 \pm 0.0026. \quad (8.183)$$

Since at present  $\Omega_\gamma \simeq 0$ , then  $\Omega_\Lambda \simeq 1 - (\Omega_b + \Omega_c)$ . The minimal  $\Lambda$ CDM model has six free parameters, which can be chosen as:

1.  $H_0$ , the Hubble parameter;
2.  $\Omega_b$ , the baryonic matter density in units of the critical density;
3.  $\Omega_c$ , the cold dark matter density in units of the critical density;
4.  $\tau$ , the optical depth to reionization (see Sect. 8.1.3.1);
5.  $A_s$  and  $n_s$ , related to the primordial fluctuation spectrum (we shall not make use of these parameters in the following).

The first evidence for a nonzero cosmological constant came from the observations by the ‘‘Supernova Cosmology Project’’ and by the ‘‘High- $z$  Supernova Search Team’’, showing that the Universe is in a state of accelerated expansion (see Sect. 8.1.1). In 2003 it was already possible to conclude that  $\Omega_m \simeq 0.3$  and  $\Omega_\Lambda \simeq 0.7$  (Fig. 8.27). The present best fit for observational data by the PDG (2018) provides for the main  $\Lambda$ CDM parameters from 1 to 4 the following values:

1.  $H_0 = (100 \times h) \text{ km s}^{-1} \text{ Mpc}^{-1}$ , with  $h = 0.678 \pm 0.009$
2.  $\Omega_b = (0.02226 \pm 0.00023)/h^2$
3.  $\Omega_c = (0.1186 \pm 0.0020)/h^2$
4.  $\tau = 0.066 \pm 0.016$ .



**Fig. 8.27** Confidence regions in the plane  $(\Omega_m, \Omega_\Lambda)$ . Credit: <http://supernova.lbl.gov>

Relaxing some of the assumptions of the standard  $\Lambda$ CDM model, extra parameters like, for example, the total mass of the neutrinos, the number of neutrino families, the dark energy equation of state, the spatial curvature, can be added.

As it is the case for particle physics, in the beginning of the twenty-first century we have a standard model also for cosmology that describes with remarkable precision the high-quality data sets we were able to gather in the last years. Although we do not yet know how to deduce the parameters of this SM from first principles in a more complete theory, we do have, nevertheless, realized that a slight change in many of these parameters would jeopardize the chance of our existence in the Universe. Are we special?

At the same time, additional questions pop up. What is dark matter made of? And how about dark energy? Why the “particle physics” vacuum expectation value originated from quantum fluctuations is 120 orders of magnitude higher than what is needed to account for dark energy?

Finally, the standard model of cosmology gives us a coherent picture of the evolution of the Universe (Figs. 8.28 and 8.29) starting from close to Planck time, where even General Relativity is no longer valid. What happened before? Was there a single beginning, or our Universe is one of many? What will happen in the future? Is our Universe condemned to a thermal death? Questions for the twenty-first century. Questions for the present students and the future researchers.

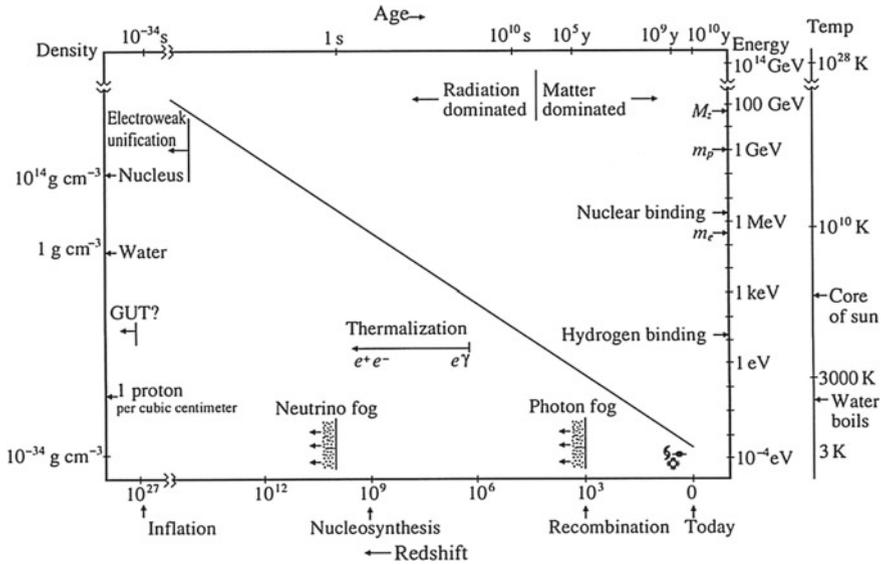


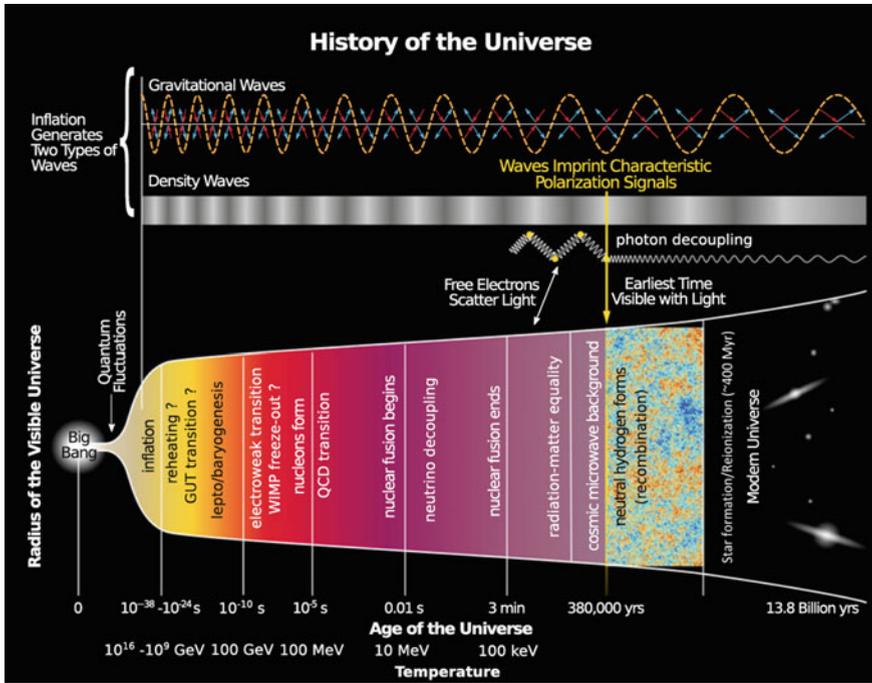
Fig. 8.28 The density, temperature, age, and redshift for the several Universe epochs. From E. Linder, “First principles of Cosmology,” Addison-Wesley 1997

### 8.4.1 Dark Matter Decoupling and the “WIMP Miracle”

The  $\Lambda$ CDM model assumes that dark matter is formed by stable massive nonrelativistic particles. These particles must have an interaction strength weaker than the electromagnetic one—otherwise they would have been found (see later); the acronym WIMP (Weakly Interactive Massive Particle) is often used to name them, since for several reasons that will be discussed below, the favorite theoretical guess compatible with experiment is that they are heavier than  $M_Z/2 \sim 45$  GeV. The lightest supersymmetric particle, possibly one of the neutralinos  $\chi$  (the lightest SuperSYmmetric particles, see the previous Chapter), is for many the most likely candidate; we shall use often the symbol  $\chi$  to indicate a generic WIMP. WIMPs must be neutral and, if there is only one kind of WIMP, we can assume that they coincide with their antiparticle (as it is the case for the neutralino).

1. We can think that in the early Universe, in the radiation dominated era, WIMPs were produced in collisions between particles of the thermal plasma. Important reactions were the production and annihilation of WIMP pairs in particle-antiparticle collisions. At temperatures corresponding to energies much higher than the WIMP mass,  $k_B T \gg m_\chi c^2$ , the colliding particle-antiparticle pairs in the plasma had enough energy to create WIMP pairs, the rate of the process being

$$\Gamma_\chi = \langle \sigma v \rangle n_\chi$$



**Fig. 8.29** Timeline of the Universe. Adapted from G. Sigl, “Astroparticle Physics: Theory and Phenomenology”, Springer 2017. Taken from Yinweichen – Own work, CC BY-SA 3.0, Wikimedia Commons

where  $n_\chi$  is the number density of WIMPs,  $\sigma$  the annihilation cross section, and  $v$  the speed. The inverse reactions converting pairs of WIMPs into SM particles were in equilibrium with the WIMP-producing processes.

- As the Universe expanded, temperature decreased, and the number of particles capable to produce a WIMP decreased exponentially as the Boltzmann factor

$$e^{-\left(\frac{m_\chi c^2}{k_B T}\right)}. \tag{8.184}$$

In addition, the expansion decreased the density  $n_\chi$ , and with it the production and annihilation rates.

- When the mean free path for WIMP-producing collisions became of the same order of the radius:

$$\lambda = \frac{1}{n_\chi \sigma} \sim \frac{v}{H}$$

or equivalently the WIMP annihilation rate became smaller than the expansion rate of the universe  $H$ :

$$\Gamma_\chi \sim n_\chi \langle \sigma v \rangle \sim H, \tag{8.185}$$

production of WIMPs ceased (decoupling). After this, the number of WIMPs in a comoving volume remained approximately constant and their number density decreased as  $a^{-3}$ . The value of the decoupling density is therefore a decreasing function of  $\langle \sigma v \rangle$ , where the velocity  $v$  is small for a large mass particle. In Fig. 8.30 the number density of a hypothetical dark matter particle as a function of time (expressed in terms of the ratio  $m_\chi c^2/k_B T$ ) for different assumed values of  $\langle \sigma v \rangle$  is shown.

A numerical solution provides

$$k_B T_{\text{dec}} \sim \frac{m_\chi c^2}{x} \tag{8.186}$$

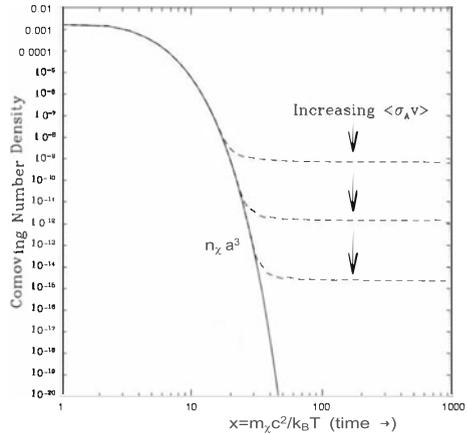
with  $x \sim 20\text{--}50$  in the range  $10 \text{ GeV} \lesssim m_\chi c^2 \lesssim 10 \text{ TeV}$ , and

$$\left( \frac{\Omega_\chi}{0.2} \right) \sim \frac{x}{20} \left( \frac{3 \text{ pb}}{\sigma} \right). \tag{8.187}$$

An important property illustrated in Fig. 8.30 is that smaller annihilation cross sections lead to larger relic densities: the weakest wins. This fact can be understood from the fact that WIMPs with stronger interactions remain in thermodynamical equilibrium for a longer time: hence they decouple when the Universe is colder, and their density is further suppressed by a smaller Boltzmann factor. This leads to the inverse relation between  $\Omega_\chi$  and  $\sigma$  in Eq. 8.187.

4. If the  $\chi$  particle interacts via weak interactions (Chap. 6) its annihilation cross section for low energies can be expressed as

**Fig. 8.30** The comoving number density of a nonrelativistic massive particle as a function of time (expressed in terms of the ratio  $\frac{m_\chi c^2}{k_B T}$ ) for different values of  $\langle \sigma v \rangle$ . Adapted from D. Hooper, “TASI 2008 Lectures on Dark Matter”, arXiv:0901.4090 [hep-ph]



$$\sigma \sim \frac{g_W^4}{m_\chi^2} \quad (8.188)$$

where  $g_W$  is the weak elementary coupling constant,  $g_W^4 \simeq 90 \text{ nb GeV}^2$ . Inserting for  $m_\chi^2$  a value of the order of 100 GeV in Eq. 8.187 one finds the right density of dark matter to saturate the energy budget of the Universe with just one particle, and no need for a new interaction.

Eq. 8.187 is often expressed using the thermally-averaged product of the cross section times velocity  $\langle\sigma v\rangle$ . For  $x \sim 20$ ,  $v \sim c/3$ , and one has

$$\langle\sigma v\rangle \sim 3 \text{ pb} \times 10^{10} \frac{\text{cm}}{\text{s}} = 3 \times 10^{-26} \frac{\text{cm}^3}{\text{s}}.$$

The value  $\langle\sigma v\rangle \sim 3 \times 10^{-26} \text{ cm}^3/\text{s}$  is a benchmark value for the velocity-averaged annihilation cross section of dark matter particles.

An appropriate relation between  $g_\chi$  and  $m_\chi$  can thus ensure a density of particles at decoupling saturating the total DM content of the Universe. In addition the expected values for a WIMP with  $m_\chi \sim m_Z \sim 100 \text{ GeV}$  and  $g_\chi \sim g_W \sim 0.6$ , corresponding to the electroweak coupling, provides the right scale for the observed dark matter density ( $\Omega_\chi \sim 0.2\text{--}0.3$ , see Sect. 8.4); this coincidence is called the *WIMP miracle*. A WIMP can indeed be the mysterious missing dark particle, but the WIMP miracle is not the only possible solution: we take it just as a benchmark. In the opinion of Andrej Sacharov, dark matter could just be gravitationally coupled—and if he was right, it will be extremely difficult to detect it experimentally. A value  $\langle\sigma v\rangle$  of the order of  $\sim 3 \times 10^{-26} \text{ cm}^3/\text{s}$  is the resulting benchmark value for the velocity-averaged annihilation cross section of dark matter particles in a range of weak interactions and of DM masses of the order<sup>8</sup> of 50 GeV–10 TeV.

## 8.5 What Is Dark Matter Made of, and How Can It Be Found?

Observations indicate a large amount of dark matter or substantial modifications of the standard theory of gravitation (see Sect. 8.1.5).

Dark matter is unlikely to consist of baryons.

- First, the  $\Lambda$ CDM model (Sect. 8.4) computes the total content of baryonic DM (i.e., nonluminous matter made by ordinary baryons) from the fit to the CMB spectrum, and the result obtained is only some 4% of the total energy of the Universe; the

---

<sup>8</sup>Strictly speaking, the Fermi model of the weak interactions entailing a cross section proportional to  $1/s$  starts failing at energies  $\gg 100 \text{ GeV}$ , since the squares of the masses of the vector bosons have to be considered. However, the “WIMP miracle” is still granted up to some 10 TeV.

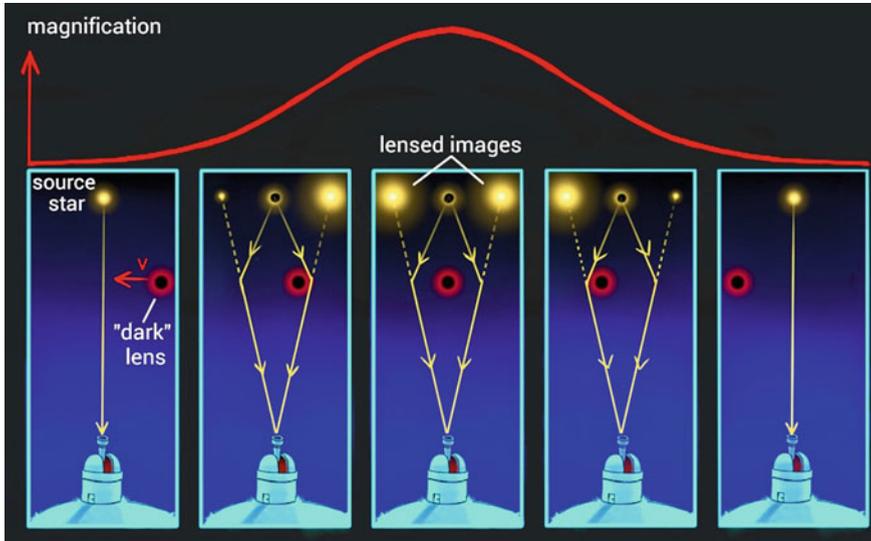


Fig. 8.31 Principle of gravitational microlensing. By Adam Rogers, blog “The Amateur Realist”

structure of the Universe, computed from astrophysical simulations, is consistent with the fractions within the  $\Lambda$ CDM model.

- Second, the abundances of light elements depend on the baryon density, and the observed abundances are again consistent with those coming from the fit to  $\Omega_b$  coming from the CMB data.

A direct search for baryonic dark matter has been however motivated by the fact that some of the hypotheses on which cosmological measurements are based might be wrong (as in the case of MOND, for example).

Baryonic DM should cluster into massive astrophysical compact objects, the so-called MACHOs,<sup>9</sup> or into molecular clouds.

The result of observations is that the amount of DM due to molecular clouds is small.

The main baryonic component should be thus concentrated in massive objects (MACHOs), including black holes. We can estimate the amount of this component using the gravitational field generated by it: a MACHO may be detected when it passes in front of a star and the star light is bent by the MACHO’s gravity. This causes more light to reach the observer and the star to look brighter, an effect known as gravitational microlensing (Fig. 8.31), very important also in the search for extrasolar planets (see Chap. 11). Several research groups have searched for MACHOs and

<sup>9</sup>MACHO is a generic name for a compact structure composed of baryonic matter, which emits little or no radiation, and are thus very difficult to detect. MACHOs may be black holes or neutron stars, as well as brown or very faint dwarf stars, or large planets.

found that only less than 20% of the total DM can be attributed to them. Therefore, MACHOs do not solve the missing mass problem.

Candidates for nonbaryonic DM must interact very “weakly” with electromagnetic radiation (otherwise they would not be dark), and they must have the right density to explain about one-quarter of the energy content of the Universe. A new particle of mass above the eV and below some  $M_Z/2$  would have been already found by LEP: DM particles must be very heavy or very light if they exist. They must also be stable on cosmological timescales (otherwise they would have decayed by now). We use the acronyms WIMP (weakly interacting massive particle) to indicate possible new “heavy” particles, and WISP (weakly interacting slim particle, or sub-eV particle) to indicate possible new light particles. Part of the rationale for WIMPs has been discussed in Sect. 8.4.1.

We shall present in this chapter the results of direct searches for dark matter, searches at accelerators, and shortly of indirect searches; a more detailed discussion of indirect signatures in the context of multimessenger astrophysics will be presented in Chap. 10.

### 8.5.1 WISPs: Neutrinos, Axions and ALPs

Among WISPs, neutrinos seem to be an obvious candidate. However, they have a free-streaming length larger than the size of a supercluster of galaxies (they thus enter in the category of the so-called “hot” dark matter). If neutrinos were the main constituent of dark matter, the first structures would have the sizes of superclusters; this is in contrast with the deep field observations from the Hubble Space Telescope (which looked in the past by sampling the Universe in depth). Observations from the Planck satellite allow to set an upper limit at 95% CL

$$\Omega_\nu \leq 0.004.$$

After having excluded known matter as a possible DM candidate, we are only left with presently unknown—although sometimes theoretically hypothesized—matter.

The axion is a hypothetical light pseudoscalar (spin-parity  $0^+$ ) particle originally postulated to explain the so-called strong  $CP$  problem. In principle,  $CP$  should not be a symmetry of the QCD Lagrangian; however,  $CP$  (and  $T$ ) appear to be conserved, as opposed to what happens for weak interactions; this fact has been verified with very good accuracy. To fix this problem, Peccei and Quinn (1977) proposed a new global symmetry, spontaneously broken at a very-high-energy scale, and giving rise to an associated boson called the axion (see Sect. 7.3.2). Being pseudoscalar (like the  $\pi^0$ ), the axion can decay into two photons, at a rate determined by the (small) coupling  $g_{A\gamma\gamma} \equiv 1/M$ —all quantities here are expressed in NU. The standard axion mass  $m_A$  is related to the coupling by the formula

$$\frac{m_A}{1 \text{ eV}} \simeq \frac{1}{M/6 \times 10^6 \text{ GeV}}. \quad (8.189)$$

The axion lifetime would then be proportional to  $1/M^5$ , which is larger than the age of the Universe for  $m_A > 10 \text{ eV}$ . An axion below this mass would thus be stable.

Since the axion couples to two photons, in a magnetic or electric field it could convert to a photon; vice versa, a photon in an external magnetic or electric field could convert into an axion (Primakoff effect); the amplitude of the process would be proportional to  $g_{A\gamma\gamma}$ .

Axion-like particles (ALPs) are a generalization of the axion: while the axion is characterized by a strict relationship between its mass  $m_A$  and  $g_{A\gamma\gamma} = 1/M$ , these two parameters are unrelated for ALPs. Depending on the actual values of their mass and coupling constant, ALPs can play an important role in cosmology, either as cold dark matter particles or as quintessential dark energy.

In order to account for dark matter, that is, to reach an energy density of the order of the critical density, axion masses should be at least  $0.1 \text{ meV}$ . Light axions and ALPs could still be DM candidates, since they are produced nonthermally via Bose-Einstein condensation, and thus they can be “cold”.

**Axion and ALP Searches.** Attempts are being made to directly detect axions mostly by:

1. Using the *light-shining-through-a-wall* (LSW) technique: a laser beam travels through a region of high magnetic field, allowing the possible conversion of photons into axions. These axions can then pass through a wall, and on the other side they can be converted back into photons in a magnetic field. An example is the OSQAR experiment at CERN.
2. Trying to spot *solar axions* using helioscopes: the CAST (CERN Axion Solar Telescope) experiment looks for the X-rays that would result from the conversion of solar axions produced in the Sun back into photons, using a 9-tons superconducting magnet.
3. Searching for axions in the local galactic dark matter halo (haloscopes). Axion conversion into photons is stimulated by strong magnetic field in a microwave cavity. When the cavity’s resonant frequency is tuned to the axion mass, the interaction between local axions and the magnetic field is enhanced. The Axion Dark Matter eXperiment (ADMX) in Seattle uses a resonant microwave cavity within a 8 T superconducting magnet.

Indirect searches are also possible.

4. The vacuum magnetic birefringence (VMB) in high magnetic fields due to photon–axion mixing can be investigated. Different polarizations often experience a different refractive index in matter—a common example is a uniaxial crystal. The vacuum is also expected to become birefringent in presence of an external magnetic field perpendicular to the propagation direction, due to the orientation of the virtual  $e^+e^-$  loops. The magnitude of this birefringence could be enhanced by the presence of an axion field, which provides further magnetic-

dependent mixing of light to a virtual field (experiment PVLAS by E. Zavattini and collaborators, 2006).

5. Study of possible anomalies in the cooling times of stars and of cataclysmic stellar events. An example is given by SNe, which produce vast quantities of weakly interacting particles, like neutrinos and possibly gravitons, axions, and other unknown particles. Although this flux of particles cannot be measured directly, the properties of the cooling depend on the ways of losing energy. The results on the cooling times and the photon fluxes (since photons are coupled to axions) constrain the characteristics of the invisible axions: emission of very weakly interacting particles would “steal” energy from the neutrino burst and shorten it. The best limits come from SN1987A. However, significant limits come also from the cooling time of stars on the horizontal branch in the color-magnitude diagram, which have reached the helium burning phase.
6. ALPs can also directly affect the propagation of photons coming from astrophysical sources, by mixing to them. This possibility has been suggested in 2007 by De Angelis, Roncadelli, and Mansutti (DARMa), and by Simet, Hooper, and Serpico. The conversion of photons into axions in the random extragalactic magnetic fields, or at the source and in the Milky Way, could give rise to a sort of cosmic light-shining-through-a-wall effect. This might enhance the yield of very-high-energy photons from distant active galactic nuclei, which would be otherwise suppressed by the interaction of these photons with the background photons in the Universe (see Chap. 10). These effects are in the sensitivity range of *Fermi*-LAT and of the Cherenkov telescopes.
7. The line emission from the two-photon decay of axions in galaxy clusters can be searched with optical and near-infrared telescopes.

With negative results, experimental searches have limited the region of mass and coupling allowed for ALPs. The limit

$$g_{A\gamma\gamma} < 6.6 \times 10^{-11} \text{ GeV}^{-1} \quad (8.190)$$

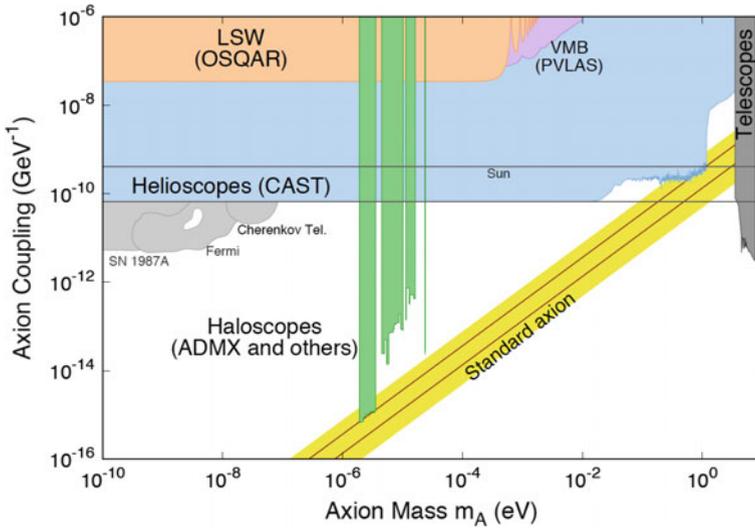
represents the strongest constraint for a wide mass range.

A hint for ALPs comes from possible anomalies in the propagation of very-high-energy photons from astrophysical sources (see Chap. 10).

A summary of exclusion limits, and of a possible observational window indicated by the cosmological propagation of VHE photons (see Chap. 10), is shown in Fig. 8.32. The topic is very hot and many new experimental results are expected in the next years.

## 8.5.2 *WIMPs*

If dark matter (DM) particles  $\chi$  are massive they must be “weakly” (i.e., with a strength corresponding to the weak interaction or even weaker) interacting (WIMPs).



**Fig. 8.32** Axion and ALP coupling to photons versus the ALP mass. The labels are explained in the text. Adapted from C. Patrignani et al. (Particle Data Group), *Chin. Phys. C*, 40, 100001 (2016) and 2017 update

A lower limit for the strength of the interaction is given by the gravitational strength. They must be neutral and, for a large range of interaction strengths, of mass larger than  $M_Z/2$ , otherwise they would have been found at the LEP  $e^+e^-$  collider.

The “WIMP miracle”, discussed in Sect. 8.4.1, guarantees that a single type of WIMP of mass  $m_\chi$  of the order of 50 GeV - few TeV, emerged from a standard thermal decoupling, can saturate the energy budget of the Universe for dark matter if the interaction characterizing WIMPs is the well-known electroweak interaction. WIMPs should be stable or should have a lifetime large enough in order to have survived from the early Universe until the present time.

If DM can be explained by just one particle  $\chi$ , coincident with its antiparticle, we expect an annihilation cross section  $\sigma_{\text{ann}}$

$$\sigma_{\text{ann}} \sim 3 \text{ pb} . \tag{8.191}$$

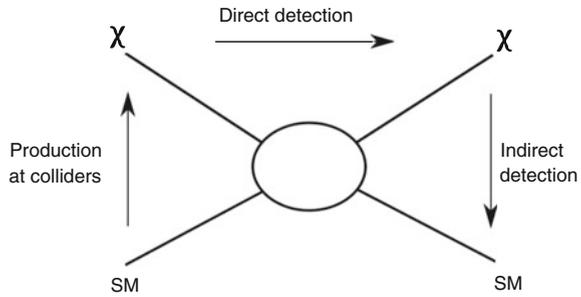
and a product of the cross section to the average velocity

$$\langle \sigma_{\text{ann}} |v_\chi| \rangle \simeq 3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1} . \tag{8.192}$$

The results in Eqs. 8.191 and 8.192 are a natural benchmark for the behavior of WIMPs, and fit well with the dynamics of electroweak interactions (Sect. 8.4.1).

Several extensions to the SM have proposed WIMP candidates, most notably supersymmetric models (SUSY) with  $R$ -parity conservation, in which the lightest supersymmetric particle, the putative neutralino  $\chi$ , is stable and thus a serious can-

**Fig. 8.33** Different processes used to investigate on WIMPs



didate (Sect. 7.6.1) with a range of annihilation cross sections including the desired ones—the spectrum of cross section can vary in some 5 orders of magnitude depending on the many free parameters of SUSY. For this reason the neutralino is usually thought to be a “natural” DM candidate. However, more general models are also allowed.

WIMPs could be detected:

1. At accelerators, where they can be produced.
2. Directly, via elastic scattering with targets on Earth. If the DM conjecture is correct, we live in a sea of WIMPs. For a WIMP mass of 50 GeV, there might be in our surroundings some  $10^5$  particles per cubic meter, moving at a speed smaller than the revolution velocity of the Earth around the Sun. From astrophysical observations the local WIMP density is about  $0.4 \text{ GeV/cm}^3$ ; the velocity distribution is Maxwellian, truncated by the Galactic escape velocity of 650 km/s. For a mass of 50 GeV, the RMS velocity is comparable to the speed of the solar system in the Galaxy,  $\sim 230 \text{ km/s}$ . Direct detection relies on observation of the scattering or other interaction of the WIMPs inside low-background Earth-based detectors.
3. Indirectly, by their decay products if they are unstable (WIMPs can be unstable, provided their lifetime is larger than the Hubble time), or by their self-annihilation products in high-density DM environments. The annihilation products of pairs of WIMPs—for example, in the halo of the Galaxy, or as a result of their accumulation in the core of the Sun or of the Earth, is likely to happen if the WIMP is a boson or a Majorana fermion as the SUSY neutralino.

These three techniques are complementary (Fig. 8.33), but results are often difficult to compare. In this chapter we shall discuss the techniques and summarize the main results on 1. and 2.; we shall explain the observables related to 3. and we shall discuss the experimental results in Chap. 10, in the context of multimessenger astrophysics.

### 8.5.2.1 Production and Detection of WIMPs at Accelerators

WIMPs can be created at colliders, but they are difficult to detect, since they are neutral and weakly interacting. However, it is possible to infer their existence. Their

signature would be missing momentum when one tries to reconstruct the dynamics of a collision into a final state involving dark matter particles and standard model particles—notice that a collision producing dark matter particles only would not be triggered. There has been a huge effort to search for the appearance of these new particles.

The production of WIMPs is severely constrained by LEP up to a mass close to  $M_Z/2$ . WIMPs with  $m_\chi < m_H/2 \sim 63$  GeV can be constrained with the branching ratio for invisible Higgs boson decays which is measured at LHC to be  $< 0.2$ . This does not appear as a strong constraint, but in many scenarios the Higgs boson coupling to WIMPs is stronger than to SM particles.

Accelerator searches are complementary to the direct searches that will be described later; however, to compare with noncollider searches, the limits need to be translated via a theory into upper limits on WIMP-nucleon scattering or on WIMP annihilation cross sections, introducing model dependence—for example, the comparison can be done in the framework of SUSY (Fig. 8.36). In particular, searches at accelerators can exclude the region below 10 GeV and a cross section per nucleon of the order of  $10^{-44}$  cm<sup>2</sup>, where direct searches are not very sensitive.

### 8.5.2.2 Direct Detection of WIMPs in Underground Detectors

Experimental detection is based on the nuclear recoil that would be caused by WIMP elastic scattering.

WIMP velocities in the Earth's surroundings are expected to be of one order of magnitude smaller than the Galactic escape velocity, i.e., they are nonrelativistic: thermalized WIMPs have typical speeds

$$\sqrt{\langle v_\chi^2 \rangle} \simeq \sqrt{\frac{2k_B T}{m_\chi}} \simeq 27 \left( \frac{100 \text{ GeV}}{m_\chi} \right)^{1/2} \text{ m/s}.$$

These are smaller than the velocity  $v_\odot$  of the solar system with respect to the center of the Galaxy, which is of the order of  $10^{-3} c$ .

If the Milky Way's dark halo is composed of WIMPs, then, given the DM density in the vicinity of the solar system and the speed of the solar system with respect to the center of the Galaxy, the  $\chi$  flux on the Earth should be about

$$\Phi_\chi \simeq v_\odot n_{\text{DM, local}} \simeq 10^5 \frac{100 \text{ GeV}}{m_\chi} \text{ cm}^{-2} \text{ s}^{-1}$$

(a local dark matter density of  $0.4 \text{ GeV/cm}^3$  has been used to compute the number density of DM particles). This flux is rather large and a potentially measurable fraction might scatter off nuclei.

The kinematics of the scattering is such that the transferred energy is in the keV range. The recoil energy  $E_K$  of a particle of mass  $M$  initially at rest after a

nonrelativistic collision with a particle of mass  $m_\chi$  traveling at a speed  $10^{-3}c$  is approximately

$$E_K \simeq 50 \text{ keV} \left[ \frac{M}{100 \text{ GeV}} \left( \frac{2}{1 + M/m_\chi} \right)^2 \right]. \quad (8.193)$$

The expected number of collisions is some  $10^{-3}$  per day in a kilogram of material for a 50 GeV particle weakly interacting.

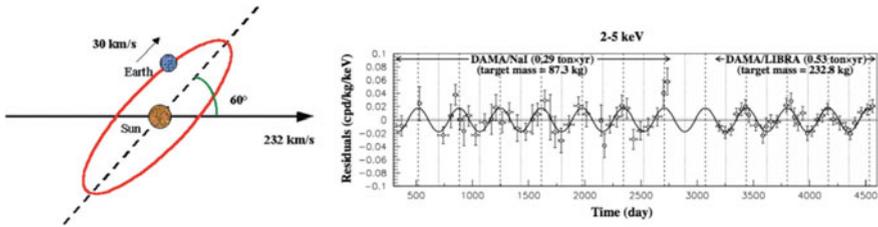
Translating a number of collisions into a cross section per nucleon is not trivial in this case. The WIMP-nucleon scattering cross section has a spin-dependent (SD) and a spin-independent (SI) part. When the scattering is coherent, the SI cross section has a quadratic dependence on the mass number  $A^2$ , which leads to strong enhancement for heavy elements. A nucleus can only recoil coherently for  $A \ll 50$ . SD scattering on the other hand depends on the total nuclear angular momentum. In the case of spin-dependent interaction the cross section is smaller by a factor of order  $A - A^2$  than for coherent scattering.

Detectors sensitive to WIMP interactions should have a low energy threshold, a low-background noise, and a large mass. The energy of a nucleus after a scattering from a WIMP is converted into a signal corresponding to (1) ionization, (2) scintillation light; (3) vibration quanta (phonons). The main experimental problem is to distinguish the genuine nuclear recoil induced by a WIMP from the huge background due to environmental radioactivity. It would be useful to do experiments which can measure the nuclear recoil energy and if possible the direction. The intrinsic rejection power of these detectors can be enhanced by the simultaneous detection of different observables (for example, heat and ionization or heat and scintillation).

The WIMP rate may be expected to exhibit some angular and time dependence. For example, there might be a daily modulation because of the shadowing effects of the Earth when turned away from the Galactic center (GC). An annual modulation in the event rate would also be expected as the Earth's orbital velocity around the Sun (about 30 km/s) adds to or subtracts from the velocity of the solar system with respect to the GC (about 230 km/s), so that the number of WIMPs intercepted per unit time varies (Fig. 8.34, left).

The detectors have then to be well isolated from the environment, possibly shielded with active and passive materials, and constructed with very low activity materials. In particular, it is essential to operate in an appropriate underground laboratory to limit the background from cosmic rays and from natural radioactivity. There are many underground laboratories in the world, mostly located in mines or in underground halls close to tunnels, and the choice of the appropriate laboratory for running a low-noise experiment is of primary importance. Just to summarize some of the main characteristics,

- The thickness of the rock (to isolate from muons and from the secondary products of their interaction).
- The geology (radioactive materials produce neutrons that should be shielded) and the presence of Radon.



**Fig. 8.34** Left: the directions of the Sun's and the Earth's motions during a year. Assuming the WIMPs to be on average at rest in the Galaxy, the average speed of the WIMPs relative to the Earth is modulated with a period of 1 year. Right: annual modulation of the total counting rate (background plus possible dark matter signal) in 7 years of data with the DAMA detector. A constant counting rate has been subtracted. From R. Bernabei et al., Riv. Nuovo Cim. 26 (2003) 1

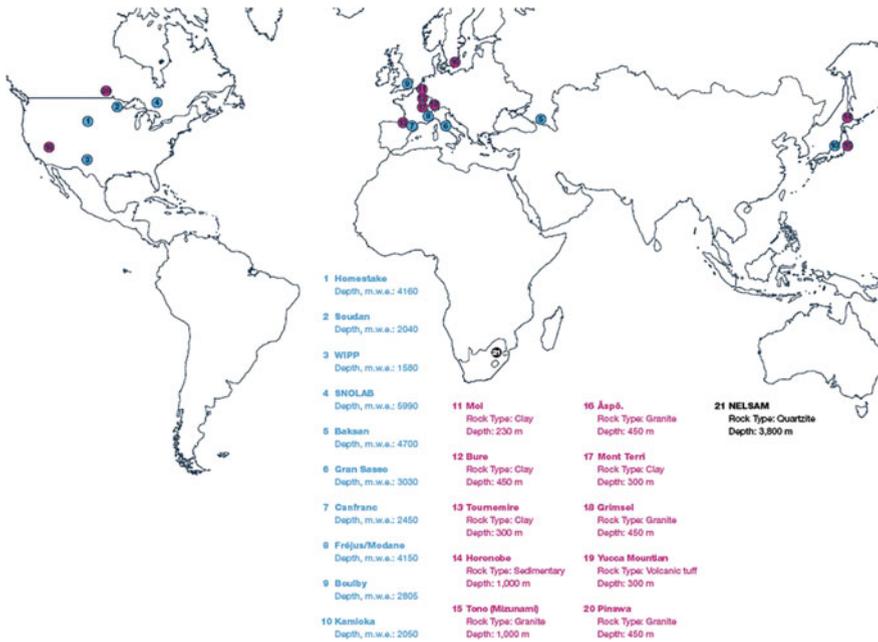
- The volume available (none of the present installations could host a megaton detector).
- The logistics.

Some of the largest underground detectors in the world are shown in Fig. 8.35.

As an example, the INFN Gran Sasso National Laboratory (LNGS), which is the largest underground European laboratory, hosts some 900 researchers from 30 different countries. LNGS is located near the town of L'Aquila, about 120 kilometers from Rome. The underground facilities are located on one side of the highway tunnel crossing the Gran Sasso mountain; there are three large experimental halls, each about 100 m long, 20 m wide, and 18 m high. An average 1400 m rock coverage gives a reduction factor of one million in the cosmic ray flux; the neutron flux is thousand times less than the one at the surface. One of the halls points to CERN, allowing long-baseline accelerator neutrino experiments.

Essentially three types of detectors operate searching directly for dark matter in underground facilities all around the world.

- Semiconductor detectors. The recoil nucleus or an energetic charged particle or radiation ionizes the traversed material and produces a small electric signal proportional to the deposited energy. Germanium crystals, which have a very small value of the gap energy (3 eV) and thus have a good resolution of 1 per thousand at 1 MeV, are commonly used as very good detectors since some years. The leading detectors are the CDMS, CoGeNT, CRESST, and EDELWEISS experiments. The bolometric technique (bolometers are ionization-sensitive detectors kept cold in a Wheatstone bridge; the effects measured are: the change in electric resistance consequent to the heating, i.e., the deposited energy, and ionization) increases the power of background rejection, and allows a direct estimate of the mass of the scattering particle.
- Scintillating crystals. Although their resolution is worse than Germanium detectors, no cooling is required. The scintillation technique is simple and well known, and large volumes can be attained because the cost per mass unit is low. However, these detectors are not performant enough to allow an event-by-event analysis. For



**Fig. 8.35** Underground laboratories for research in particle physics (1–10) listed with their depth in meters water equivalent. Laboratories for research in the million-year scale isolation of nuclear waste are also shown (11–20). The NELSAM laboratory (21) is for earthquake research. From [www.deepscience.org](http://www.deepscience.org)

this reason, some experiments are looking for a time-dependent modulation of a WIMP signal in their data. As the Earth moves around the Sun, the WIMP flux should be maximum in June (when the revolution velocity of the Earth adds to the velocity of the solar system in the Galaxy) and minimum in December, with an expected amplitude variation of a few percent. DAMA (now called in its upgrade DAMA/LIBRA) is the first experiment using this detection strategy. The apparatus is made of highly radio-pure NaI(Tl) crystals, each with a mass of about 10 kg, with two PMTs at the two opposing faces.

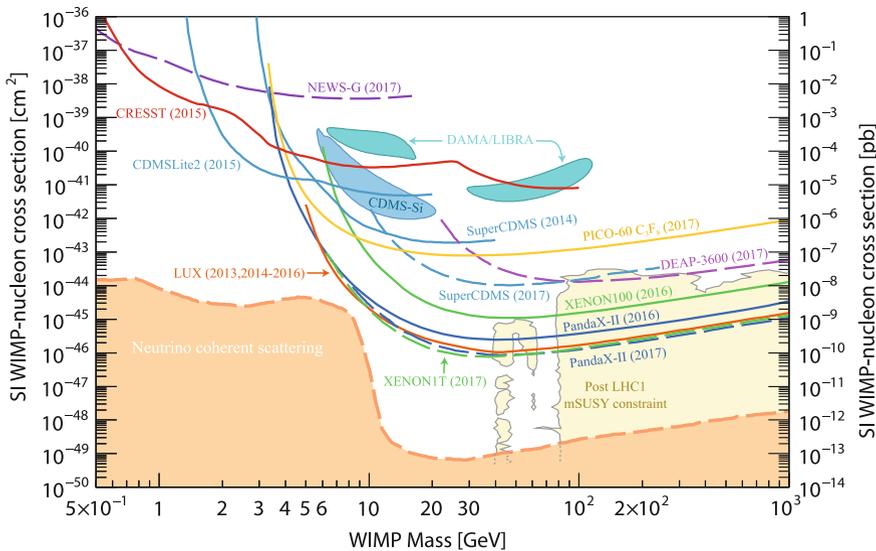
- Noble liquid detectors. Certainly the best technique, in particular in a low-background environment, it uses noble elements as detectors (this implies low background from the source itself) such as argon ( $A = 40$ ) and xenon ( $A = 131$ ). Liquid xenon (LXe) and liquid argon (LAr) are good scintillators and ionizers in response to the passage of radiation. Using pulse-shape discrimination of the signal, events induced by a WIMP can be distinguished from background electron recoil. The main technique is to the present knowledge the “double phase” technique. A vessel is partially filled with noble liquid, with the rest of the vessel containing the same element in a gaseous state. Electric fields of about 1 kV/cm and 10 kV/cm are established across the liquid and gas volumes, respectively. An

interaction in the liquid produces excitation and ionization processes. Photomultiplier tubes are present in the gas volume and in the liquid. The double phase allows reconstruction of the topology of the interaction (the gas allowing a TPC reconstruction), thus helping background removal. The leading experiments are:

- The XENON100 detector, a 165 kg liquid xenon detector located in LGNS with 62kg in the target region and the remaining xenon in an active veto together with high purity Germanium detectors. A new liquid xenon-based project, XENON1t, is planned in the LNGS, with 3.5 tons of liquid xenon.
- The LUX detector, a 370 kg xenon detector installed in the Homestake laboratory (now called SURF) in the US. LUX was decommissioned in 2016 and a new experiment, LUX-ZEPLIN (LZ), with 7 tons of active liquid xenon is in preparation.

Whatever the detector is, the energy threshold is a limiting factor on the sensitivity at low WIMP masses; but for high values of  $m_\chi$  the flux decreases as  $1/m_\chi$  and the sensitivity for fixed mass densities also drops. The best sensitivity is attained for WIMP masses close to the mass of the recoiling nucleus.

The experimental situation is not completely clear (Fig. 8.36). Possible WIMP detection signals were claimed by the experiment DAMA, based on a large scintillator



**Fig. 8.36** Compilation of experimental results on cross sections of WIMPs versus masses. The areas labeled as DAMA/LIBRA and CDMS-Si indicate regions of possible signals from those experiments. Supersymmetry implications are also shown. New experiments to hunt for dark matter are becoming so sensitive that neutrino will soon show up as background; the “neutrino floor” is shown in the plot. From C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016) and 2017 update, in which experiments are also described in detail

(NaI(Tl)) volume, and the CRESST and CoGeNT data show some stress with respect to experiments finding no signal. The data analyzed by DAMA corresponded to 7 years of exposure with a detector mass of 250 kg, to be added to 6 years of exposure done earlier with a detector mass of 100 kg. Based on the observation of a signal at  $9.3\sigma$  (Fig. 8.34, right) modulated with the expected period of 1 year and the correct phase (with a maximum near June 2, as expected from the Earth’s motion around the Sun), DAMA proposes two possible scenarios: a WIMP with  $m_\chi \simeq 50$  GeV and a cross section per nucleon  $\sigma \simeq 7 \times 10^{-6}$  pb, and a WIMP with  $m_\chi \simeq 8$  GeV and  $\sigma \simeq 10^{-3}$  pb. The DAMA signal is controversial, as it has not presently been reproduced by other experiments with comparable sensitivity but with different types of detectors (we remind that there is some model dependence in the rescaling from the probability of interaction to the cross section per nucleon).

In the next years the sensitivity of direct DM detectors will touch the “neutrino floor” for WIMP masses above 10 GeV, in particular thanks to the DARWIN detector, a 50-ton LXe detector planned to start in the mid-2020s at LNGS.

In the meantime the DarkSide collaboration at LNGS has proposed a 20-ton liquid argon dual-phase detector, with the goal to be sensitive to a cross section of  $9 \times 10^{-48} \text{cm}^2$  for a mass of  $1 \text{ TeV}/c^2$ , based on extrapolations of the demonstrated efficiency of a 50 kg pathfinder.

### 8.5.2.3 Indirect Detection of WIMPs

WIMPs are likely to annihilate in pairs; it is also possible that they are unstable, with lifetimes comparable with the Hubble time, or larger. In these cases one can detect secondary products of WIMP decays. Let us concentrate now on the case of annihilation in pairs—most of the considerations apply to decays as well.

If the WIMP mass is below the  $W$  mass, the annihilation of a pair of WIMPs should proceed mostly through  $f\bar{f}$  pairs. The state coming from the annihilation should be mostly a spin-0 state (in the case of small mutual velocity the  $s$ -wave state is favored in the annihilation; one can derive a more general demonstration using the Clebsch–Gordan coefficients). Helicity suppression entails that the decay in the heaviest accessible fermion pair is preferred, similar to what seen in Chap. 6 when studying the  $\pi^\pm$  decay (Sect. 6.3.4): the decay probability into a fermion–antifermion pair is proportional to the square of the mass of the fermion. In the mass region between 10 and 80 GeV, the decay into  $b\bar{b}$  pairs is thus preferred (this consideration does not hold if the decay is radiative, and in this case a generic  $f\bar{f}$  pair will be produced). The  $f\bar{f}$  pair will then hadronize and produce a number of secondary particles.

In the case of the annihilation in the cores of stars, the only secondary products which could be detected would be neutrinos. However, no evidence for a significant extra flux of high-energy neutrinos from the direction of the Sun or from the Earth’s core has ever been found.

One could have annihilations in the halos of galaxies or in accretion regions close to black holes or generic cusps of dark matter density. In this case one could have

generation of secondary particles, including gamma rays, or antimatter which would appear in excess to the standard rate.

We shortly present here the possible scenarios for detections, which will be discussed in larger details in Chap. 10, in the context of multimessenger astrophysics.

**Gamma Rays.** The self-annihilation of a heavy WIMP  $\chi$  can generate photons (Fig. 8.37) in three main ways.

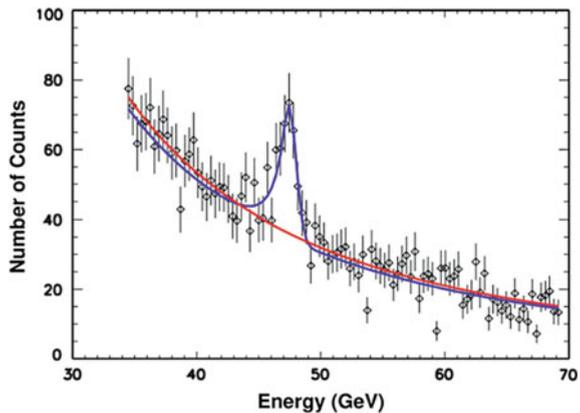
- (a) Directly, via annihilation into a photon pair ( $\chi\chi \rightarrow \gamma\gamma$ ) or into a photon— $Z$  pair ( $\chi\chi \rightarrow \gamma Z$ ) with  $E_\gamma = m_\chi$  or  $E_\gamma = (4m_\chi^2 - m_Z^2)/4m_\chi$ , respectively; these processes give a clear signature at high energies, as the energy is monochromatic, but the process is suppressed at one loop, so the flux is expected to be very faint.
- (b) Via annihilation into a quark pair which produces jets emitting in turn a large number of  $\gamma$  photons ( $q\bar{q} \rightarrow \text{jets} \rightarrow \text{many photons}$ ); this process produces a continuum of gamma rays with energies below the WIMP mass. The flux can be large but the signature might be difficult to detect, since it might be masked by astrophysical sources of photons.
- (c) Via internal bremsstrahlung; also in this case one has an excess of low energy gamma rays with respect to a background which is not so well known. Besides the internal bremsstrahlung photons, one will still have the photons coming from the processes described at the two previous items.

The  $\gamma$ -ray flux from the annihilation of a pair of WIMPs of mass  $m_\chi$  can be expressed as the product of a particle physics component times an astrophysics component:

$$\frac{dN}{dE} = \frac{1}{4\pi} \underbrace{\frac{\langle\sigma_{\text{ann}}v\rangle}{2m_\chi^2} \frac{dN_\gamma}{dE}}_{\text{Particle Physics}} \times \underbrace{\int_{\Delta\Omega-l.o.s.} dl(\Omega)\rho_\chi^2}_{\text{Astrophysics}}. \quad (8.194)$$

The particle physics factor contains  $\langle\sigma_{\text{ann}}v\rangle$ , the velocity-weighted annihilation cross section (there is indeed a possible component from cosmology in  $v$ ), and  $dN_\gamma/dE$ ,

**Fig. 8.37**  $\gamma$ -ray signature of neutralino self-annihilation or of neutralino decay. Simulation from the *Fermi*-LAT collaboration



the  $\gamma$ -ray energy spectrum for all final states convoluted with the respective branching ratios. The part of the integral over line of sight (*l.o.s.*) in the observed solid angle of the squared density of the dark matter distribution constitutes the astrophysical contribution.

It is clear that the expected flux of photons from dark matter annihilations, and thus its detectability, depend crucially on the knowledge of the annihilation cross section  $\sigma_{\text{ann}}$  (which even within SUSY has uncertainties of one to two orders of magnitude for a given WIMP mass) and of  $\rho_\chi$ , which is even more uncertain, and enters squared in the calculation. Cusps in the dark matter profile, or even the presence of local clumps, could make the detection easier by enhancing  $\rho_\chi$ —and we saw that the density in the cusps is uncertain by several orders of magnitude within current models (Sect. 8.1.5.1). In the case of WIMP decays, the density term will be linear.

The targets for dark matter searches should be not extended, with the highest density, with no associated astrophysical sources, close to us, and possibly with some indication of small luminosity/mass ratio from the stellar dynamics.

- The Galactic center is at a distance of about 8 kpc from the Earth. A black hole of about  $3.6 \times 10^6$  solar masses, Sgr A\*, lies there. Because of its proximity, this region might be the best candidate for indirect searches of dark matter. Unfortunately, there are other astrophysical  $\gamma$ -ray sources in the field of view (e.g., the supernova remnant Sgr A East), and the halo core radius makes it an extended rather than a point-like source.
- The best observational targets for dark matter detection outside the Galaxy are the Milky Way's dwarf spheroidal satellite galaxies (for example, Carina, Draco, Fornax, Sculptor, Sextans, Ursa Minor). For all of them (e.g., Draco), there is observational evidence of a mass excess with respect to what can be estimated from luminous objects, i.e., a high M/L ratio. In addition, the gamma-ray signal expected in the absence of WIMP annihilation is zero.

The results of the experimental searches will be discussed in Sect. 10.5.3.

**Neutrinos.** Neutrino–antineutrino pairs can also be used for probing WIMP annihilation or decay, along the same line discussed for gamma rays, apart from the fact that neutrino radiation is negligible. Besides the smaller astrophysical background, the advantage of neutrinos is that they can be observed even if the annihilation happens in the cores of opaque astrophysical objects (the Sun or compact objects in particular); apart from these cases the sensitivity of the gamma-ray channel is by far superior, due to the experimental difficulty of detecting neutrinos for the present and next generation of detectors.

**Matter–Antimatter and Electron Signatures.** Another indirect manifestation of the presence of WIMPs would be given by their decay (or self-annihilation) producing democratically antimatter and matter.

A possible observable could be related to electron and positron pairs. A smoking gun would be the presence of a peak in the energy of the collected electrons, indicating a two-body decay. A shoulder reaching  $m_\chi/2$  could also be a signature, but, in this last case, one could hypothesize astrophysical sources as well.

An excess of antimatter with respect to the prediction of models in which antimatter is just coming from secondary interactions of cosmic rays and astrophysical sources could be seen very clearly in the positron and antiproton spectrum. The PAMELA space mission observed a positron abundance in cosmic radiation higher than that predicted by current models (see Chap. 10). This has been confirmed by the AMS-02 mission, reaching unprecedented accuracy. AMS-02 has also found an excess of antiprotons with respect to models in which only secondary production is accounted. A smoking gun signature for the origin of positrons from the decay of a  $\chi$  or from a  $\chi\chi$  annihilation would be a steep drop-off of the ratio at a given energy. A more detailed discussion of experimental data will be presented in Chap. 10.

### 8.5.3 *Other Nonbaryonic Candidates*

Additional candidates, more or less theoretically motivated, have been proposed in the literature. We list them shortly here; they are less economic than the ones discussed before (WIMPs in particular).

**Sterile Neutrinos.** A possible DM candidate is a “sterile” neutrino, i.e., a neutrino which does not interact via weak interactions. We know that such neutrino states exist: the right-handed component of neutrinos in the standard model are sterile. Constraints from cosmology make it, however, unlikely that light sterile neutrinos can be the main component of dark matter. Sterile neutrinos with masses of the order of the keV and above could be, with some difficulty, accommodated in the present theories.

**Kaluza–Klein States.** If particles propagate in extra spacetime dimensions, they will have an infinite spectroscopy of partner states with identical quantum numbers; these states could be a DM candidate.

**Matter in Parallel Branes; Shadow or Mirror Matter.** Some theories postulate the presence of matter in parallel branes, interacting with our world only via gravity or via a super-weak interaction. In theories popular in the 1960s, a “mirror matter” was postulated to form astronomical mirror objects; the cosmology in the mirror sector could be different from our cosmology, possibly explaining the formation of dark halos. This mirror-matter cosmology has been claimed to explain a wide range of phenomena.

**Superheavy Particles (WIMPzillas).** Superheavy particles above the GZK cutoff (WIMPzillas) could have been produced in the early Universe; their presence could be detected by an excess of cosmic rays at ultrahigh energies.

### Further Reading

[F8.1] J. Silk, “The big bang”, Times Books 2000.

- [F8.2] B. Ryden, “Introduction to Cosmology”, Cambridge 2016. This book provides a clear introduction to cosmology for upper-level undergraduates.
- [F8.3] E.F. Taylor and J.A. Wheeler, “Exploring Black Holes, introduction to general relativity”, Addison-Wesley 2000. This book provides an enlightening introduction to the physics of black holes emphasizing how they are “seen” by observers in different reference frames.
- [F8.4] M.V. Berry, “Principles of Cosmology and Gravitation”, Adam Hilger 1989. This book presents the fundamentals of general relativity and cosmology with many worked examples and exercises without requiring the use of tensor calculus.
- [F8.5] V. Mukhanov, “Physical Foundations of Cosmology”, Cambridge 2005. This book provides a comprehensive introduction to inflationary cosmology at early graduate level.
- [F8.6] B. Schutz, “A first Course in General Relativity”, second edition, Cambridge University Press 2009. This is a classic and comprehensive textbook.
- [F8.7] R. Feynman, “The Feynman Lectures on Physics”, [www.feynmanlectures.caltech.edu](http://www.feynmanlectures.caltech.edu). The classic book by Feynman on Web.

## Exercises

1. *Cosmological principle and Hubble law.* Show that the Hubble law does not contradict the cosmological principle (all points in space and time are equivalent).
2. *Olbers Paradox.* Why is the night dark? Does the existence of interstellar dust (explanation studied by Olbers himself) solve the paradox?
3. *Steady state Universe.* In a steady state Universe with Hubble law, matter has to be permanently created. Compute in that scenario the creation rate of matter.
4. *Blackbody form of the Cosmic Microwave Background.* In 1965 Penzias and Wilson discovered that nowadays the Universe is filled with a cosmic microwave background which follows an almost perfect Planck blackbody formula. Show that the blackbody form of the energy density of the background photons was preserved during the expansion and the cooling that had occurred in the Universe after photon decoupling.
5. *The CMB and our body.* If CMB photons are absorbed by the human body (which is a reasonable assumption), what is the power received by a human in space because of CMB?
6. *CMB, infrared and visible photons.* Estimate the number of near-visible photons ( $\lambda$  from  $0.3 \mu\text{m}$  to  $1 \mu\text{m}$ ) in a cubic centimeter of interstellar space. Estimate the number of far-infrared photons in the region of  $\lambda$  from  $1000 \mu\text{m}$  to  $1 \mu\text{m}$ .
7. *Requirements for a cosmic neutrino background detector.* Let the typical energy of a neutrino in the Cosmic Neutrino Background be  $\sim 0.2 \text{ meV}$ . What is the approximate interaction cross section for cosmic neutrinos? How far would typically a cosmic neutrino travel in ice before interacting?
8. *Dark Matter and mini-BHs.* If BHs of mass  $10^{-8} M_{\odot}$  made up all the dark matter in the halo of our Galaxy, how far away would the nearest such BH on average? How frequently would you expect such a BH to pass within 1 AU of the Sun?

9. *Nucleosynthesis and neutron lifetime.* The value of the neutron lifetime, which is abnormally long for weak decay processes (why?), is determinant in the evolution of the Universe. Discuss what would have been the primordial fraction of He if the neutron lifetime would have been one-tenth of its real value.
10. *GPS time corrections.* Identical clocks situated in a GPS satellite and at the Earth surface have different periods due general relativity effects. Compute the time difference in one day between a clock situated in a satellite in a circular orbit around Earth with a period of 12 h and a clock situated on the Equator at the Earth surface. Consider that Earth has a spherical symmetry and use the Schwarzschild metric.
11. *Asymptotically Matter-dominated Universe.* Consider a Universe composed only by matter and radiation. Show that whatever would have been the initial proportion of matter and radiation energy densities this Universe will be asymptotically matter dominated.
12. *Cosmological distances.* Consider a light source at a redshift of  $z = 2$  in an Einstein-de Sitter Universe. (a) How far has the light from this object traveled to reach us? (b) How distant is this object today?
13. *Decoupling.* What are the characteristic temperatures (or energies) at which (a) neutrinos decouple; (b) electron-positron pairs annihilate; (c) protons and neutrons drop out of equilibrium; (d) light atomic nuclei form; (e) neutral He atoms form; (f) neutral hydrogen atoms form; (g) photons decouple from baryonic matter?
14. *Evolution of momentum.* How does the momentum of a free particle evolve with redshift (or scale factor)?
15.  *$\Lambda$ CDM and distances.* Estimate the expected apparent magnitude of a type Ia supernova (absolute magnitude  $M \simeq -19$  at a redshift  $z = 1$  in the  $\Lambda$ CDM Universe.
16. *Flatness of the Early Universe.* The present experimental data indicate a value for the normalized total energy density of the Universe compatible with one within a few per mil. Compute the maximum possible value of  $|\Omega - 1|$  at the scale of the electroweak symmetry breaking consistent with the measurements at the present time.
17. *WIMP “miracle”.* Show that a possible Weak Interacting Massive Particle (WIMP) with a mass of the order of  $m_\chi \sim 100$  GeV would have the relic density needed to be the cosmic dark matter (this is the so-called WIMP “miracle”).
18. *Recoil energy in a DM detector.* Calculate the recoil energy of a target nucleus in a DM detector.