Proceedings of the 2014 Asia–Europe–Pacific School of High-Energy Physics, Puri, India, 4–17 November 2014, edited by
M. Mulders and R. Godbole, CERN Yellow Reports: School Proceedings, Vol. 2/2017, CERN-2017-005-SP (CERN, Geneva, 2017)

Cosmology

V.A. Rubakov
Institute for Nuclear Research of the Russian Academy of Sciences,
Moscow, Russia
and
Department of Particle Physics and Cosmology, Physics Faculty, Moscow State University,
Moscow, Russia

Abstract
Cosmology and particle physics are deeply interrelated. Among the common problems are dark energy, dark matter and baryon asymmetry of the Universe. We discuss these problems in general terms, and concentrate on several particular hypotheses. On the dark matter side, we consider weakly interacting massive particles and axions/axion-like particles as cold dark matter, sterile neutrinos and gravitinos as warm dark matter. On the baryon asymmetry side, we discuss electroweak baryogenesis as a still-viable mechanism. We briefly describe diverse experimental and observational approaches towards checking these hypotheses. We then turn to the earliest cosmology. We give arguments showing that the hot stage was preceded by another epoch at which density perturbations and possibly primordial gravity waves were generated. The best guess here is inflation, which is consistent with everything we know of density perturbations, but there are alternative scenarios. Future measurements of the properties of density perturbations and possible discovery of primordial gravity waves have strong potential in this regard.

Keywords
Lectures; cosmology; cosmological model; baryon asymmetry; dark matter; dark energy; nucleosynthesis.

1 Introduction

Cosmology is one of the major sources of inspiration—and confusion—for particle physicists. It gives direct evidence for the necessity to extend the Standard Model of particle physics, possibly at an energy scale that can be probed by collider experiments. Indeed, there is no doubt that most part of the mass in the present Universe is in the form of mysterious dark matter particles which are not present in the Standard Model. Also, the very existence of conventional matter in our Universe (i.e., matter–antimatter asymmetry) calls for processes with baryon number violation and substantial charge parity (CP)-violation, which have not been observed in experiments. These processes had to be rapid in the early Universe and, furthermore, the asymmetry between matter and antimatter had to be generated in a fairly turbulent cosmological epoch. Again, the conditions necessary for the generation of this asymmetry are not present in the Standard Model. Solving the problems of dark matter and matter–antimatter asymmetry are the two immediate challenges for particle physics.

Going very much back into the cosmological history, we encounter another challenging issue. It is very well known that matter in the Universe was very hot and dense early on. It is less known that the properties of the matter distribution in the past and present Universe, reflected in the properties of the cosmic microwave background (CMB), galaxy distribution etc, unambiguously tell us that the hot epoch was not the earliest. It was preceded by another, completely different epoch responsible for the generation of inhomogeneities which in the end have become galaxies and their clusters, stars and ourselves. Obviously, the very fact that we are confident about the existence of such an epoch

https://doi.org/10.23730/CYRSP-2017-002.239

is a fundamental result of theoretical and observational cosmology. The most plausible hypothesis on that epoch is cosmological inflation, though the observational support of this scenario is presently not overwhelming, and alternative possibilities have not been ruled out. For the time being it appears unlikely that we will be able to probe the physics behind that epoch in terrestrial experiments, but there is no doubt that this physics belongs to the broad domain of 'particles and fields'.

After this brief introduction, the scope of these lectures must be clear. To set the stage, we briefly consider the basic notions of cosmology. We then discuss several dark matter particle candidates and mechanisms for dark matter generation. Needless to say, these candidates do not exhaust the long list of the candidates proposed; our choice is based on a personal view of what candidates are more plausible. Our next topic is the matter–antimatter asymmetry of the Universe, and we present electroweak baryogenesis as a mechanism particularly interesting from the viewpoint of the LHC experiments. The last part of these lectures deals with cosmological perturbations, inflation (and its alternatives) and the potential of future observational data.

These lectures are meant to be self-contained, but we necessarily omit numerous details, while trying to make clear the basic ideas and results. More complete accounts of cosmology and its particle-physics aspects may be found in various books [1–6]. Dark matter candidates we consider in these lectures are reviewed in Refs. [7–10]. Electroweak baryogenesis is presented in detail in reviews [11–13]; for reference, a plausible alternative scenario, leptogenesis, is discussed in reviews [14, 15]. Aspects of inflation and its alternatives are reviewed in Refs. [16–20].

2 Expanding universe

2.1 Friedmann–Lemaître–Robertson–Walker metric

Our Universe (more precisely, its visible part) is *homogeneous and isotropic*. Clearly, this does not apply to relatively small spatial scales: there are galaxies, clusters of galaxies and giant voids. But boxes of sizes exceeding about 200 Mpc all look the same. Here the Mpc is the distance unit conventionally used in cosmology,

$$1 \text{ Mpc} \approx 3 \times 10^6 \text{ light years} \approx 3 \times 10^{24} \text{ cm} .$$

There are three types of homogeneous and isotropic three-dimensional spaces, labelled by an integer parameter $\varkappa$. These are three-sphere (closed model, $\varkappa = +1$), flat (Euclidean) space (flat model, $\varkappa = 0$) and three-hyperboloid (open model, $\varkappa = -1$). We will see that the parameter $\varkappa$ enters the dynamical equations governing the space–time fabric of the Universe.

Another basic property of our Universe is that it *expands*. This is encoded in the space–time metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\mathrm{d}\mathbf{x}^2 , \tag{1}$$

where $\mathrm{d}\mathbf{x}^2$ is the distance on a unit three-sphere, Euclidean space or hyperboloid. The metric (1) is called the Friedmann–Lemaître–Robertson–Walker (FLRW) metric, and $a(t)$ is the scale factor. In these lectures we use natural units, setting the speed of light and Planck and Boltzmann constants equal to 1,

$$c = \hbar = k_{\mathrm{B}} = 1 .$$

In these units, Newton's gravity constant is $G = M_{\mathrm{Pl}}^{-2}$, where $M_{\mathrm{Pl}} = 1.2 \times 10^{19}$ GeV is the Planck mass.

The meaning of Eq. (1) is as follows. One can check that a free mass put at a certain $\mathbf{x}$ at zero velocity will stay at the same $\mathbf{x}$ forever. In other words, the coordinates $\mathbf{x}$ are comoving. The scale factor $a(t)$ increases in time, so the distance between free masses of fixed spatial coordinates $\mathbf{x}$ grows, $\mathrm{d}l^2 = a^2(t)\mathrm{d}\mathbf{x}^2$. The space stretches out; the galaxies run away from each other.

This expansion manifests itself as a red shift. Red shift is often interpreted as the Doppler effect for a source running away from us with velocity v: if the wavelength at emission is λ_{e}, then the wavelength

we measure is $\lambda_0 = (1+z)\lambda_{\rm e}$, where $z = v/c$ (here we temporarily restore the speed of light). This interpretation is useless and rather misleading in cosmology (with respect to which reference frame does the source move?). The correct interpretation is that as the Universe expands, space stretches out and the photon wavelength increases proportionally to the scale factor a. So, the relation between the wavelengths is

$$\lambda_0 = (1+z)\lambda_{\rm e} \ , \quad \text{where } z = \frac{a(t_0)}{a(t_{\rm e})} - 1 \ ,$$

where $t_{\rm e}$ is the emission time. For $z \ll 1$, this relation reduces to the Hubble law,

$$z = H_0 r \ , \tag{2}$$

where r is the physical distance to the source and $H_0 \equiv H(t_0)$ is the present value of the Hubble parameter

$$H(t) = \frac{\dot{a}(t)}{a(t)} \ .$$

In the formulas above, we label the present values of time-dependent quantities by subscript 0; we will always do so in these lectures.

Question. Derive the Hubble law (2) for $z \ll 1$.

The red shift of an object is directly measurable. The wavelength $\lambda_{\rm e}$ is fixed by physics of the source, say, it is the wavelength of a photon emitted by an excited hydrogen atom. So, one identifies a series of emission or absorption lines, thus determining $\lambda_{\rm e}$, and measures their actual wavelengths λ_0. These spectroscopic measurements give accurate values of z even for distant sources. On the other hand, the red shift is related to the time of emission and hence to the distance to the source. Absolute distances to astrophysical sources have a lot more systematic uncertainty, and so do the direct measurements of the Hubble parameter H_0. According to the Planck Collaboration [21], the combination of observational data gives

$$H_0 = (67.8 \pm 0.9)\ \frac{\rm km}{\rm s\ Mpc} \approx (14.4 \times 10^9\ {\rm yr})^{-1} \ , \tag{3}$$

where the unit used in the first expression reflects the interpretation of red shift in terms of the Doppler shift. The fact that the systematic uncertainties in the determination of H_0 are pretty large is illustrated in Fig. 1.

Traditionally, the present value of the Hubble parameter is written as

$$H_0 = h \times 100\ \frac{\rm km}{\rm s\ Mpc} \ . \tag{4}$$

Thus, $h \approx 0.7$. We will use this value in further estimates.

2.2 Hot Universe: recombination, Big Bang nucleosynthesis and neutrinos

Our Universe is filled with CMB. The CMB as observed today consists of photons with an excellent black-body spectrum of temperature

$$T_0 = 2.7255 \pm 0.0006\ {\rm K} \ . \tag{5}$$

The spectrum has been precisely measured by various instruments, see Fig. 2, and does not show any deviation from the Planck spectrum (see Ref. [23] for a detailed review).

Once the present photon temperature is known, the number density and energy density of CMB photons are known from the Planck distribution formulas,

$$n_{\gamma,0} = 410\ \ {\rm cm}^{-3} \ , \qquad \rho_{\gamma,0} = \frac{\pi^2}{15} T_0^4 = 2.7 \times 10^{-10}\ \frac{\rm GeV}{\rm cm^3} \tag{6}$$

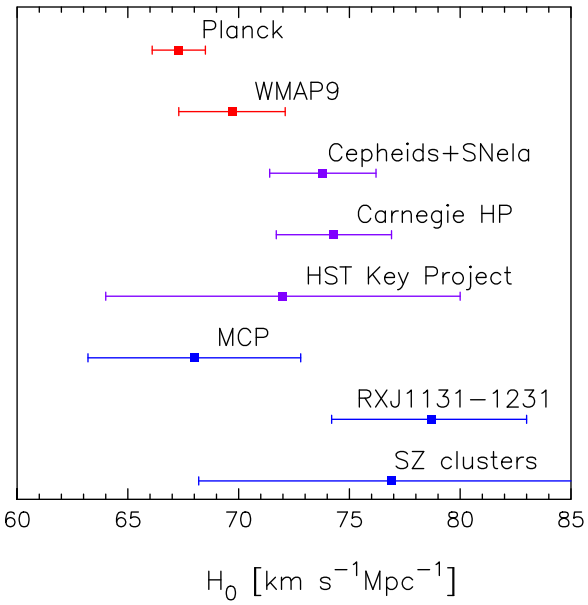

Fig. 1: Recent determinations of the Hubble parameter H_0 [22]

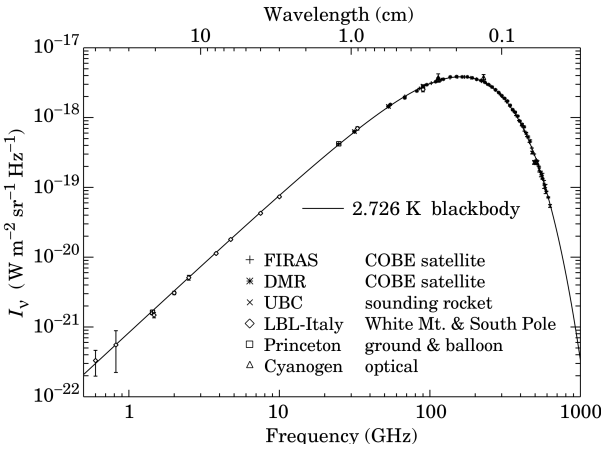

Fig. 2: Measured CMB energy spectrum as compiled in Ref. [24]

(the second expression is the Stefan–Boltzmann formula).

The CMB is a remnant of an earlier cosmological epoch. The Universe was hot at early times and, as it expands, the matter in it cools down. Since the wavelength of a photon evolves in time as $a(t)$, its energy and hence temperature scale as

$$\omega(t) \propto a^{-1}(t) \,, \qquad T(t) = \frac{a_0}{a(t)} T_0 = (1+z) T_0 \,.$$

When the Universe was hot, the usual matter (electrons and protons with a rather small admixture of light nuclei, mainly ^{4}He) was in the plasma phase. At that time photons strongly interacted with electrons due to the Thomson scattering and protons interacted with electrons via the Coulomb force, so all these particles were in thermal equilibrium. As the Universe cooled down, electrons 'recombined' with protons into neutral hydrogen atoms (helium recombined earlier), and the Universe became transparent to photons: at that time, the density of hydrogen atoms was quite small, 250 cm^{-3}. The photon last

scattering occurred at temperature and red shift

$$T_{\rm rec} \approx 3000 \text{ K} , \qquad z_{\rm rec} \approx 1090 ,$$

when the age of the Universe was about $t \approx 380$ thousand years (for comparison, its present age is about 13.8 billion years). Needless to say, CMB photons got red shifted since the last scattering, so their present temperature is $T_0 = T_{\rm rec}/(1 + z_{\rm rec})$.

The photon last scattering epoch is an important cornerstone in the cosmological history. Since after that CMB photons travel freely through the Universe, they give us a photographic picture of the Universe at that epoch. Importantly, the duration of the last scattering epoch was considerably shorter than the Hubble time $H^{-1}(t_{\rm rec})$; to a reasonable approximation, recombination occurred instantaneously. Thus, the photographic picture is only slightly washed out due to the finite thickness of the last scattering surface.

At even earlier times, the temperature of the Universe was even higher. We have direct evidence that at some point the temperature in the Universe was in the MeV range. A traditional source of evidence is the Big Bang nucleosynthesis (BBN). The story begins at a temperature of about 1 MeV, when the age of the Universe was about 1 s. Before that time neutrons were rapidly created and destroyed in weak processes like

$$\mathrm{e}^- + \mathrm{p} \longleftrightarrow \mathrm{n} + \nu_{\rm e} , \tag{7}$$

while at $T_{\rm n} \approx 1$ MeV these processes switched off, and the comoving number density of neutrons froze out. The neutron-to-proton ratio at that time was given by the Boltzmann factor,

$$\frac{n_{\rm e}}{n_{\rm p}} = \mathrm{e}^{-\frac{m_{\rm n}-m_{\rm p}}{T_{\rm n}}} .$$

Interestingly, $m_{\rm n} - m_{\rm p} \sim T_{\rm n}$, so the neutron–proton ratio at neutron freeze-out and later was neither equal to 1, nor very small. Were it equal to 1, protons would combine with neutrons into ^{4}He at a somewhat later time, and there would remain no hydrogen in the Universe. On the other hand, for very small $n_{\rm n}/n_{\rm p}$, too few light nuclei would be formed, and we would not have any observable remnants of the BBN epoch. In either case, the Universe would be quite different from what it actually is. It is worth noting that the approximate relation $m_{\rm n} - m_{\rm p} \sim T_{\rm n}$ is a coincidence: $m_{\rm n} - m_{\rm p}$ is determined by light quark masses and electromagnetic coupling, while $T_{\rm n}$ is determined by the strength of weak interactions (which govern the rates of the processes (7)) and gravity (which governs the expansion of the Universe). This is one of numerous coincidences we encounter in cosmology.

At temperatures somewhat below $T_{\rm n}$, the neutrons combined with protons into light elements in thermonuclear reactions like

$$\begin{aligned} \mathrm{p} + \mathrm{n} &\to {}^2\mathrm{H} + \gamma , \\ {}^2\mathrm{H} + \mathrm{p} &\to {}^3\mathrm{He} + \gamma , \\ {}^3\mathrm{He} + {}^2\mathrm{H} &\to {}^4\mathrm{He} + \mathrm{p} , \end{aligned} \tag{8}$$

etc, up to ^{7}Li. The abundances of light elements have been measured; see Fig. 3. On the other hand, the only parameter relevant for calculating these abundances (assuming negligible neutrino–antineutrino asymmetry) is the baryon-to-photon ratio

$$\eta_{\rm B} \equiv \eta = \frac{n_{\rm B}}{n_\gamma} , \tag{9}$$

characterizing the number density of baryons. Comparison of the BBN theory with the observational determination of the composition of the cosmic medium enables one to determine $\eta_{\rm B}$ and check the overall consistency of the BBN picture. It is even more reassuring that a completely independent measurement

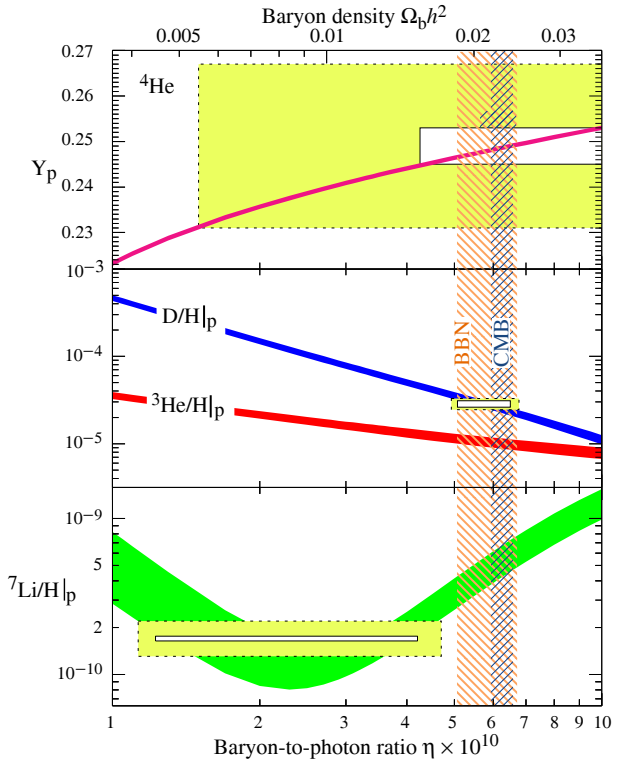

Fig. 3: Abundances of light elements, measured (boxes; larger boxes include systematic uncertainties) and calculated as functions of baryon-to-photon ratio η [25]. The determination of $\eta \equiv \eta_{\rm B}$ from BBN (vertical range marked BBN) is in excellent agreement with the determination from the analysis of CMB temperature fluctuations (vertical range marked CMB).

of $\eta_{\rm B}$ that makes use of the CMB temperature fluctuations is in excellent agreement with BBN. Thus, BBN gives us confidence that we understand the Universe at $T \sim 1$ MeV, $t \sim 1$ s. In particular, we are convinced that the cosmological expansion was governed by general relativity.

Another class of processes of interest at temperatures in the MeV range is neutrino production, annihilation and scattering,

$$\nu_\alpha + \bar{\nu}_\alpha \longleftrightarrow {\rm e}^+ + {\rm e}^-$$

and crossing processes. Here the subscript α labels neutrino flavours. These processes switch off at $T \sim$ 2–3 MeV, depending on neutrino flavour. Since then neutrinos do not interact with the cosmic medium other than gravitationally, but they do affect the properties of CMB and distribution of galaxies through their gravitational interactions. These effects are not negligible, since the energy density of relativistic neutrinos is almost the same as that of photons and, at temperature $T_{\rm rec} \simeq 3000$ K, the energy density of these relativistic species is only three times smaller than the energy density of non-relativistic particles (dark matter and baryons). Thus, observational data can be used to establish, albeit somewhat indirectly, the existence of relic neutrinos and set limits on neutrino masses. An example is shown in Fig. 4, where the number of neutrino flavours $N_{\rm eff}$ and the sum of neutrino masses are taken as free parameters. We see that cosmology *requires* relic neutrinos of at least three flavours and sets the limit on neutrino mass $m_\nu \lesssim 0.1$ eV (neutrino oscillation data tell that neutrinos with masses above 0.1 MeV are degenerate in mass). The latest Planck analysis gives [21]

$$\sum_i m_{\nu_i} < 0.23 \text{ eV} , \qquad N_{\rm eff} = 3.15 \pm 0.23 .$$

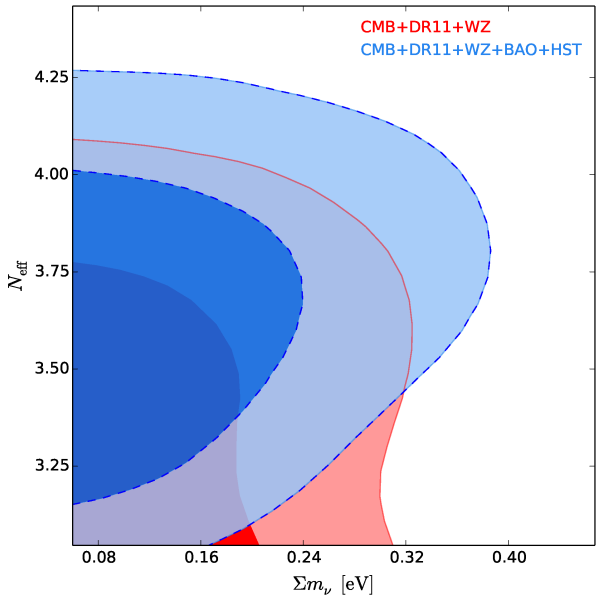

Fig. 4: Effective number of neutrino species and sum of neutrino masses allowed by cosmological observations [26].

2.3 Dynamics of expansion

The basic equation governing the expansion rate of the Universe is the Friedmann equation,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3M_{\rm Pl}^2}\rho - \frac{\varkappa}{a^2}\,, \tag{10}$$

where the dot denotes derivative with respect to time t, ρ is the *total* energy density in the Universe and $\varkappa = 0, \pm 1$ is the parameter, introduced in Section 2.1, that discriminates the Euclidean 3-space ($\varkappa = 0$) and curved 3-spaces. The Friedmann equation is nothing but the (00)-component of the Einstein equations of general relativity, $R_{00} - \frac{1}{2}g_{00}R = 8\pi T_{00}$, specified to the FLRW metric. Observationally, the spatial curvature of the Universe is very small: the last, curvature term in the right-hand side of Eq. (10) is small compared to the energy density term [21],

$$\frac{1/a^2}{8\pi\rho/(3M_{\rm Pl}^2)} < 0.005\,,$$

while the theoretical expectation is that the spatial curvature is completely negligible. Establishing that the three-dimensional space is (nearly) Euclidean is one of the profound results of CMB observations.

In what follows we set $\varkappa = 0$ and write the Friedmann equation as

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3M_{\rm Pl}^2}\rho\,. \tag{11}$$

The standard parameter used in cosmology is the critical density,

$$\rho_{\rm c} = \frac{3}{8\pi}M_{\rm Pl}^2 H_0^2 \approx 5\times 10^{-6}\,\frac{\rm GeV}{\rm cm^3}\,. \tag{12}$$

According to Eq. (11), it is equal to the sum of all forms of energy density in the present Universe. There are at least three of such forms: relativistic matter, or *radiation*, non-relativistic *matter, M* and

dark energy, Λ. For every form λ with the *present* energy density $\rho_{\lambda,0}$, one defines the parameter

$$\Omega_\lambda = \frac{\rho_{\lambda,0}}{\rho_{\rm c}} \ .$$

One finds from Eq. (11) that

$$\sum_\lambda \Omega_\lambda = 1 \ .$$

The Ω are important cosmological parameters characterizing the energy balance in the present Universe. Their numerical values are

$$\Omega_{\rm rad} = 8.7 \times 10^{-5} \ , \tag{13a}$$

$$\Omega_{\rm M} = 0.31 \ , \tag{13b}$$

$$\Omega_\Lambda = 0.69 \ . \tag{13c}$$

The value of $\Omega_{\rm rad}$ needs qualification. At early times, when the temperature exceeds the masses of all neutrino species, neutrinos are relativistic. The value of $\Omega_{\rm rad}$ in Eq. (13a) is calculated for the unrealistic case in which *all neutrinos are relativistic today*, so the radiation component even at present consists of CMB photons and three neutrino species. This prescription is convenient for studying the energy (and entropy) content in the early Universe, since it enables one to scale the energy density (and entropy) back in time in a simple way, see below. For future reference, let us give the value of the present entropy density in the Universe, pretending that neutrinos are relativistic,

$$s_0 \approx 3000 \ {\rm cm}^{-3} \ . \tag{14}$$

Question. Calculate the numerical value of Ω_γ and the entropy density of CMB photons.

Non-relativistic matter consists of baryons and dark matter. The contributions of each of these fractions are [21]

$$\Omega_{\rm B} = 0.048 \ ,$$
$$\Omega_{\rm DM} = 0.26 \ .$$

Different components of the energy density evolve differently in time. The energy of a given photon or massless neutrino scales as a^{-1}, and the number density of these species scales as a^{-3}. Therefore, the energy density of radiation scales as $\rho_{\rm rad} \propto a^{-4}$ and

$$\rho_{\rm rad}(t) = \left(\frac{a(t)}{a_0}\right)^4 \rho_{\rm rad,0} = (1+z)^4 \, \Omega_{\rm rad} \rho_{\rm c} \ . \tag{15}$$

The energy of non-relativistic matter is dominated by the mass of its particles, so the energy density scales as the number density, i.e.,

$$\rho_{\rm M}(t) = \left(\frac{a(t)}{a_0}\right)^3 \rho_{\rm M,0} = (1+z)^3 \, \Omega_{\rm M} \rho_{\rm c} \ . \tag{16}$$

Finally, the energy density of dark energy does not change in time, or changes very slowly. We assume for definiteness that ρ_Λ stays constant in time,

$$\rho_\Lambda = \Omega_\Lambda \rho_{\rm c} = {\rm const} \ . \tag{17}$$

In fact, whether or not ρ_Λ depends on time (even slightly) is a very important question. If dark energy is a cosmological constant (or, equivalently, vacuum energy), then it does not depend on time at all. Even

a slight dependence of ρ_Λ on time would mean that we are dealing with something different from the cosmological constant, like, e.g., a new scalar field with a very flat scalar potential. The existing limits on the time evolution of dark energy correspond, roughly speaking, to the variation of ρ_Λ by not more than 20% in the last 8 billion years (from the time corresponding to $z \approx 1$); usually these limits are expressed in terms of the equation-of-state parameter relating energy density and effective pressure $p_\Lambda = w_\Lambda \rho_\Lambda$:

$$w_\Lambda \approx 1.0 \pm 0.1 \; . \tag{18}$$

The relevance of the effective pressure is seen from the covariant conservation equation for the energy–momentum tensor, $\nabla_\mu T^{\mu\nu} = 0$, whose $\nu = 0$ component reads

$$\dot\rho = -3\frac{\dot a}{a}(\rho + p) \; .$$

It shows that the energy density of a component with equation of state $p = w\rho$, $w = \text{const}$ scales as $\rho \propto a^{-3(1+w)}$. As pointed out above, radiation ($w_{\rm rad} = 1/3$) and matter ($w = 0$) scale as $\rho_{\rm rad} \propto a^{-4}$ and $\rho_{\rm M} \propto a^{-3}$, respectively, while the cosmological constant case corresponds to $w_\Lambda = -1$.

Question. Show that for a gas of relativistic particles, $p = \rho/3$.

According to Eqs. (15), (16) and (17), different forms of energy dominate at different cosmological epochs. The present Universe is at the end of the transition from matter domination to Λ domination: the dark energy will 'soon' completely dominate over non-relativistic matter because of the rapid decrease of the energy density of the latter. Conversely, the matter energy density increases as we go backwards in time, and until relatively recently ($z \lesssim 0.3$) it dominated over dark energy density. At even more distant past, the radiation energy density was the highest, as it increases most rapidly backwards in time. The red shift at radiation–matter equality, when the energy densities of radiation and matter were equal, is

$$1 + z_{\rm eq} = \frac{a_0}{a(t_{\rm eq})} = \frac{\Omega_{\rm M}}{\Omega_{\rm rad}} \approx 3500$$

and, using the Friedmann equation, one finds the age of the Universe at equality

$$t_{\rm eq} \approx 50\,000 \text{ years} \; .$$

Note that recombination occurred at matter domination, but rather soon after equality. So, we have the following sequence of the regimes of evolution:

$$\ldots \Longrightarrow \text{Radiation domination} \Longrightarrow \text{Matter domination} \Longrightarrow \Lambda \text{ domination} \; .$$

The dots here denote some cosmological epoch preceding the hot stage of the evolution; as we mentioned in Section 1, we are confident that such an epoch existed, but do not quite know what it was.

2.4 Radiation domination

The epoch of particular interest for our purposes is radiation domination. By inserting $\rho_{\rm rad} \propto a^{-4}$ into the Friedmann equation (11), we obtain

$$\frac{\dot a}{a} = \frac{\text{const}}{a^2} \; .$$

This gives the evolution law

$$a(t) = \text{const} \cdot \sqrt{t} \; . \tag{19}$$

The constant here does not have physical significance, as one can rescale the coordinates $\mathbf{x}$ at some fixed moment of time, thus changing the normalization of a.

There are several points to note regarding the result (19). First, the expansion *decelerates*:

$$\ddot{a} < 0 \ .$$

This property holds also for the matter-dominated epoch, but it does not hold for the domination of the dark energy.

Question. Find the evolution laws, analogous to Eq. (19), for matter- and Λ-dominated Universes. Show that the expansion decelerates, $\ddot{a} < 0$, at matter domination and accelerates, $\ddot{a} > 0$, at Λ domination.

Second, time $t = 0$ is the Big Bang singularity (assuming erroneously that the Universe starts being radiation dominated). The expansion rate

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

diverges as $t \to 0$, and so do the energy density $\rho(t) \propto H^2(t)$ and temperature $T \propto \rho^{1/4}$. Of course, the classical general relativity and usual notions of statistical mechanics (e.g., temperature itself) are not applicable very near the singularity, but our result suggests that in the picture we discuss (hot epoch right after the Big Bang), the Universe starts its classical evolution in a very hot and dense state, and its expansion rate is very high in the beginning. It is customary to consider for illustrational purposes that the relevant quantities in the beginning of the classical expansion take the Planck values, $\rho \sim M_{\rm Pl}^4$, $H \sim M_{\rm Pl}$ etc.

Third, at a given moment of time the size of a causally connected region is finite. Consider signals emitted right after the Big Bang and travelling with the speed of light. These signals travel along the light cone with $\mathrm{d}s = 0$ and hence $a(t)\mathrm{d}x = \mathrm{d}t$. So, the coordinate distance that a signal travels from the Big Bang to time t is

$$x = \int_0^t \frac{\mathrm{d}t}{a(t)} \equiv \eta \ . \tag{20}$$

In the radiation-dominated Universe,

$$\eta = \text{const} \cdot \sqrt{t} \ .$$

The physical distance from the emission point to the position of the signal is

$$l_{\rm H}(t) = a(t)x = a(t) \int_0^t \frac{\mathrm{d}t}{a(t)} = 2t \ .$$

As expected, this physical distance is finite, and it gives the size of a causally connected region at time t. It is called the horizon size (more precisely, the size of the particle horizon). A related property is that an observer at time t can see only the part of the Universe whose current physical size is $l_{\rm H}(t)$. Both at radiation and matter domination one has, modulo a numerical constant of order 1,

$$l_{\rm H}(t) \sim H^{-1}(t) \ . \tag{21}$$

To give an idea of numbers, the horizon size at the present epoch is

$$l_{\rm H}(t_0) \approx 15 \text{ Gpc} \simeq 4.5 \times 10^{28} \text{ cm} \ .$$

Question. Find the proportionality constant in Eq. (21) for a matter-dominated Universe. Is there a particle horizon in a Universe without matter but with positive cosmological constant?

It is convenient to express the Hubble parameter at radiation domination in terms of temperature. The Stefan–Boltzmann law gives for the energy density of a gas of relativistic particles in thermal equilibrium at zero chemical potentials (chemical potentials in the Universe are indeed small)

$$\rho_{\rm rad} = \frac{\pi^2}{30} g_* T^4 \ , \tag{22}$$

with g_* being the effective number of degrees of freedom,

$$g_* = \sum_{\rm bosons} g_i + \frac{7}{8} \sum_{\rm fermions} g_i \ ,$$

where g_i is the number of spin states and the factor $7/8$ is due to Fermi statistics. Hence, the Friedmann equation (11) gives

$$H = \frac{T^2}{M_{\rm Pl}^*} \ , \qquad M_{\rm Pl}^* = \frac{M_{\rm Pl}}{1.66\sqrt{g_*}} \ . \tag{23}$$

One more point has to do with entropy: the cosmological expansion is slow, so that the entropy is conserved (modulo exotic scenarios with large entropy generation). The entropy density in thermal equilibrium is given by

$$s = \frac{2\pi^2}{45} g_* T^3 \ .$$

The conservation of entropy means that the entropy density scales *exactly* as a^{-3},

$$s a^3 = \text{const} \ , \tag{24}$$

while temperature scales *approximately* as a^{-1}. The temperature would scale as a^{-1} if the number of relativistic degrees of freedom would be independent of time. This is not the case, however. Indeed, the value of g_* depends on temperature: at $T \sim 10$ MeV relativistic species are photons, neutrinos, electrons and positrons, while at $T \sim 1$ GeV four flavours of quarks, gluons, muons and τ-leptons are relativistic too. The number of degrees of freedom in the Standard Model at $T \gtrsim 100$ GeV is

$$g_*(100 \text{ GeV}) \approx 100 \ .$$

If there are conserved quantum numbers, such as the baryon number after baryogenesis, their density also scales as a^{-3}. Hence, the time-independent characteristic of, say, the baryon abundance is the baryon-to-entropy ratio

$$\Delta_{\rm B} = \frac{n_{\rm B}}{s} \ .$$

The commonly used baryon-to-photon ratio $\eta_{\rm B}$, Eq. (9), is related to $\Delta_{\rm B}$ by a numerical factor, but this factor depends on time through g_* and stays constant only after $\mathrm{e}^+\mathrm{e}^-$ annihilation, i.e., at $T \lesssim 0.5$ MeV. Numerically,

$$\Delta_{\rm B} = 0.14 \eta_{\rm B,0} = 0.86 \times 10^{-10} \ . \tag{25}$$

3 Dark energy

Before turning to our main topics, let us briefly discuss dark energy. We know very little about this 'substance': our knowledge is summarized in Eqs. (13c) and (18). We also know that dark energy does not clump, unlike dark matter and baryons. It gives rise to the accelerated expansion of the Universe. Indeed, the solution to the Friedmann equation (11) with constant $\rho = \rho_\Lambda$ is

$$a(t) = \mathrm{e}^{H_\Lambda t} \ ,$$

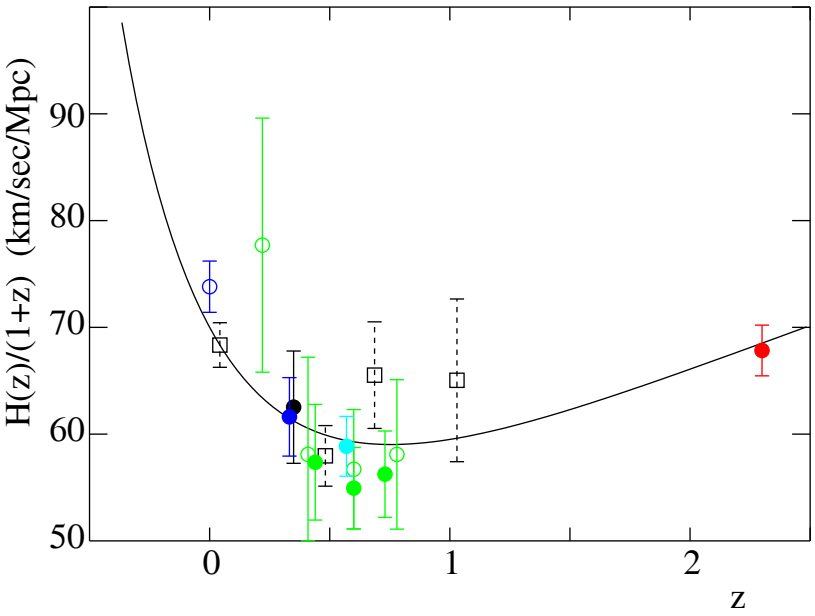

Fig. 5: Observational data on the time derivative of the scale factor as function of red shift z [27]. The change of the behaviour from decreasing to increasing with decreasing z means the change from decelerated to accelerated expansion. The theoretical curve corresponds to a spatially flat Universe with $h = 0.7$ and $\Omega_\Lambda = 0.73$.

where $H_\Lambda = (8\pi\rho_\Lambda/3M_{\rm Pl}^2)^{1/2} = \text{const}$. This gives $\ddot{a} > 0$, unlike at radiation or matter domination. The observational discovery of the accelerated expansion of the Universe was the discovery of dark energy. Recall that early on (substantial z), the Universe was matter dominated, so its expansion was decelerating. The transition from decelerating to accelerating expansion is confirmed by combined observational data, see Fig. 5, which shows the dependence on red shift of the quantity $H(z)/(1+z) = \dot{a}(t)/a_0$.

Question. Find the red shift z at which decelerated expansion turned into an accelerated one.

As a remark, the effective pressure of dark energy or any other component is defined as the (possibly time-dependent) parameter determining the spatial components of the energy–momentum tensor in a locally Lorentz frame ($a = 1$ in the FLRW context),

$$T_{\mu\nu} = \text{diag}\,(\rho, p, p, p)\ .$$

In the case of the cosmological constant, the dark energy density does not depend on time at all:

$$T_{\mu\nu} = \rho_\Lambda \eta_{\mu\nu}\ ,$$

where $\eta_{\mu\nu}$ is the Minkowski tensor. Hence, $w_\Lambda = -1$. One can view this as the characteristic of vacuum, whose energy–momentum tensor must be Lorentz-covariant. As we pointed out above, any deviation from $w = -1$ would mean that we are dealing with something other than vacuum energy density.

The problem with dark energy is that its present value is extremely small by particle-physics standards,

$$\rho_{\rm DE} \approx 4\ \text{GeV m}^{-3} = (2 \times 10^{-3}\ \text{eV})^4\ .$$

In fact, there are two hard problems. One is that particle-physics scales are much larger than the scale relevant to the dark energy density, so the dark energy density is zero to an excellent approximation. Another is that it is non-zero nevertheless, and one has to understand its energy scale. To quantify the first problem, we recall the known scales of particle physics and gravity,

$$\text{Strong interactions}: \qquad \Lambda_{\rm QCD} \sim 1\,\text{GeV}\ ,$$

$$\begin{aligned} \text{Electroweak}: &\qquad M_{\rm W} \sim 100\,\text{GeV}\,, \\ \text{Gravitational}: &\qquad M_{\rm Pl} \sim 10^{19}\,\text{GeV}\,. \end{aligned}$$

Off hand, physics at scale M should contribute to the vacuum energy density as $\rho_\Lambda \sim M^4$, and there is absolutely no reason for vacuum to be as light as it is. The discrepancy here is huge, as one sees from the above numbers.

To elaborate on this point, let us note that the action of gravity plus, say, the Standard Model has the general form

$$S = S_{\rm EH} + S_{\rm SM} - \rho_{\Lambda,0} \int \sqrt{-g}\,\mathrm{d}^4x\,,$$

where $S_{\rm EH} = -(16\pi G)^{-1}\int R\,\sqrt{-g}\,\mathrm{d}^4x$ is the Einstein–Hilbert action of general relativity, $S_{\rm SM}$ is the action of the Standard Model and $\rho_{\Lambda,0}$ is the bare cosmological constant. In order that the vacuum energy density be almost zero, one needs fantastic cancellations between the contributions of the Standard Model fields into the vacuum energy density, on the one hand, and $\rho_{\Lambda,0}$ on the other. For example, we know that quantum chromodynamics (QCD) has a complicated vacuum structure, and one would expect that the energy density of QCD should be of the order of $(1\ \text{GeV})^4$. At least for QCD, one needs a cancellation of the order of 10^{-44}. If one goes further and considers other interactions, the numbers get even worse.

What are the hints from this 'first' cosmological constant problem? There are several options, though not many. One is that the Universe could have a very long prehistory: extremely long. This option has to do with relaxation mechanisms. Suppose that the original vacuum energy density is indeed large, say, comparable to the particle-physics scales. Then there must be a mechanism which can relax this value down to an acceptably small number. It is easy to convince oneself that this relaxation could not happen in the history of the Universe we know of. Instead, the Universe should have a very long prehistory during which this relaxation process might occur. At that prehistoric time, the vacuum in the Universe must have been exactly the same as our vacuum, so the Universe in its prehistory must have been exactly like ours, or almost exactly like ours. Only in that case could a relaxation mechanism work. There are concrete scenarios of this sort [28, 29]. However, at the moment it seems that these scenarios are hardly testable, since this is prehistory.

Another possible hint is towards anthropic selection. The argument that goes back to Weinberg and Linde [30, 31] is that if the cosmological constant were larger, say, by a factor of 100, we simply would not exist: the stars would not have formed because of the fast expansion of the Universe. So, the vacuum energy density may be selected anthropically. The picture is that the Universe may be much, much larger than what we can see, and different large regions of the Universe may have different properties. In particular, vacuum energy density may be different in different regions. Now, we are somewhere in the place where one can live. All the rest is empty of observers, because there the parameters such as vacuum energy density are not suitable for their existence. This is disappointing for a theorist, as this point of view allows for arbitrary tuning of fundamental parameters. It is hard to disprove this option, on the other hand. We do exist, and this is an experimental fact. The anthropic viewpoint may, though hopefully will not, get more support from the LHC, if no or insufficient new physics is found there. Indeed, another candidate for an environmental quantity is the electroweak scale, which is fine tuned in the Standard Model in the same sense as the cosmological constant is fine tuned in gravity (in the Standard Model context, this fine tuning goes under the name of the gauge hierarchy problem).

Turning to the 'second' cosmological constant problem, we note that the scale 10^{-3} eV may be associated with some new light field(s), rather than with vacuum. This implies that ρ_Λ depends on time, i.e., $w_\Lambda \neq -1$ and w_Λ may well depend on time itself. Current data are compatible with time-independent w_Λ equal to -1, but their precision is not particularly high. We conclude that future cosmological observations may shed new light on the field content of the fundamental theory.

4 Dark matter

Unlike dark energy, dark matter experiences the same gravitational force as baryonic matter. It consists presumably of new stable massive particles. These make clumps of mass which constitute most of the mass of galaxies and clusters of galaxies. There are various ways of measuring the contribution of non-baryonic dark matter into the total energy density of the Universe (see Refs. [7–10] for details).

1. The composition of the Universe affects the angular anisotropy and polarization of CMB. Quite accurate CMB measurements available today enable one to measure the total mass density of dark matter.
2. There is direct evidence that dark matter exists in the largest gravitationally bound objects—clusters of galaxies. There are various methods to determine the gravitating mass of a cluster, and even the mass distribution in a cluster, which give consistent results. As an example, the total gravitational field of a cluster, produced by both dark matter and baryons, acts as a gravitational lens for extended light sources behind the cluster. The images of these sources enable one to reconstruct the mass distribution in the cluster. This is shown in Fig. 6. These determinations show that baryons (independently measured through their X-ray emission) make less than $1/3$ of total mass in clusters. The rest is dark matter.

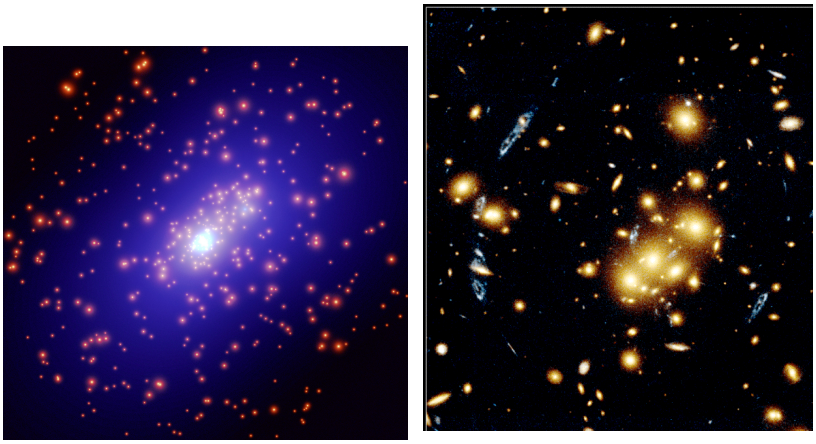

Fig. 6: Cluster of galaxies CL$0024+1654$ [32], acting as gravitational lens. Right-hand panel: cluster in visible light. Round yellow spots are galaxies in the cluster. Elongated blue images are those of one and the same galaxy beyond the cluster. Left-hand panel: reconstructed distribution of gravitating mass in the cluster; brighter regions have larger mass density.

A particularly convincing case is the Bullet Cluster, Fig. 7. Shown are two galaxy clusters that passed through each other. The dark matter and galaxies do not experience friction and thus do not lose their velocities. On the contrary, baryons in hot, X-ray-emitting gas do experience friction and hence get slowed down and lag behind dark matter and galaxies. In this way the baryons (which are mainly in hot gas) and dark matter are separated in space.

3. Dark matter exists also in galaxies. Its distribution is measured by the observations of rotation velocities of distant stars and gas clouds around a galaxy, Fig. 8. Because of the existence of dark matter away from the luminous regions, i.e., in halos, the rotation velocities do not decrease with the distance from the galactic centres; rotation curves are typically flat up to distances exceeding the size of the bright part by a factor of 10 or so. The fact that dark matter halos are so large is explained by the defining property of dark matter particles: they do not lose their energies by emitting photons and, in general, interact with conventional matter very weakly.

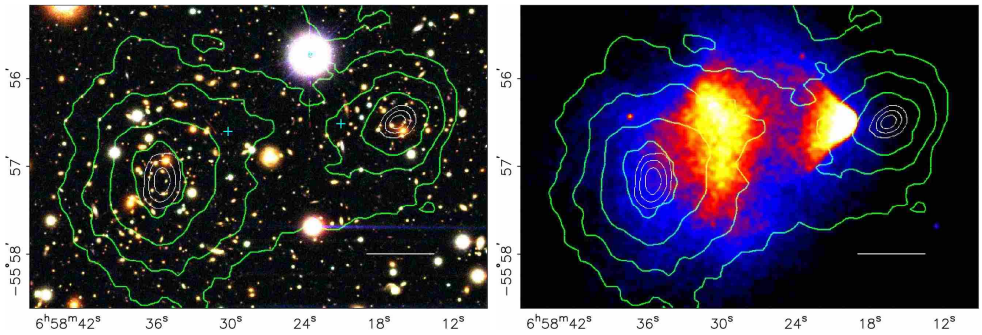

Fig. 7: Observation [33] of the Bullet Cluster 1E0657-558 at $z = 0.296$. Closed lines show the gravitational potential produced mainly by dark matter and measured through gravitational lensing. Bright regions show X-ray emission of hot baryon gas, which makes most of the baryonic matter in the clusters. The length of the white interval is 200 kpc in the comoving frame.

Dark matter is characterized by the mass-to-entropy ratio,

$$\left(\frac{\rho_{\rm DM}}{s}\right)_0 = \frac{\Omega_{\rm DM}\rho_{\rm c}}{s_0} \approx \frac{0.26 \times 5 \times 10^{-6}\ {\rm GeV\ cm^{-3}}}{3000\ {\rm cm^{-3}}} = 4 \times 10^{-10}\ {\rm GeV}\ . \tag{26}$$

This ratio is constant in time since the freeze out of dark matter density: both number density of dark matter particles $n_{\rm DM}$ (and hence their mass density $\rho_{\rm DM} = m_{\rm DM} n_{\rm DM}$) and entropy density get diluted exactly as a^{-3}.

Dark matter is crucial for our existence, for the following reason. Density perturbations in baryon–electron–photon plasma before recombination do not grow because of high pressure, which is mostly due to photons; instead, perturbations are sound waves propagating in plasma with time-independent amplitudes. Hence, in a Universe without dark matter, density perturbations in the baryonic component would start to grow only after baryons decouple from photons, i.e., after recombination. The mechanism

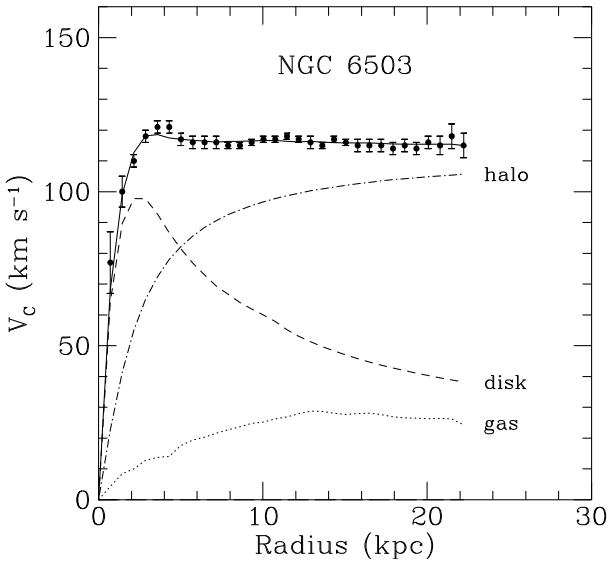

Fig. 8: Rotation velocities of hydrogen gas clouds around the galaxy NGC 6503 [34]. Lines show the contributions of the three main components that produce the gravitational potential. The main contribution at large distances is due to dark matter, labelled 'halo'.

of the growth is pretty simple: an overdense region gravitationally attracts surrounding matter; this matter falls into the overdense region, and the density contrast increases. In the expanding matter-dominated Universe this gravitational instability results in the density contrast growing like $(\delta\rho/\rho)(t) \propto a(t)$. Hence, in a Universe without dark matter, the growth factor for baryon density perturbations would be at most

$$\frac{a(t_0)}{a(t_{\rm rec})} = 1 + z_{\rm rec} = \frac{T_{\rm rec}}{T_0} \approx 10^3 \ . \tag{27}$$

Because of the presence of dark energy, the growth factor is even somewhat smaller. The initial amplitude of density perturbations is very well known from the CMB anisotropy measurements, $(\delta\rho/\rho)_{\rm i} = 5 \times 10^{-5}$. Hence, a Universe without dark matter would still be pretty homogeneous: the density contrast would be in the range of a few per cent. No structure would have been formed, no galaxies, no life. No structure would be formed in future either, as the accelerated expansion due to dark energy will soon terminate the growth of perturbations.

Since dark matter particles decoupled from plasma much earlier than baryons, perturbations in dark matter started to grow much earlier. The corresponding growth factor is larger than (27), so that the dark matter density contrast at galactic and subgalactic scales becomes of order one, perturbations enter the non-linear regime and form dense dark matter clumps at $z = 5$–10. Baryons fall into potential wells formed by dark matter, so dark matter and baryon perturbations develop together soon after recombination. Galaxies get formed in the regions where dark matter was overdense originally. For this picture to hold, dark matter particles must be non-relativistic early enough, as relativistic particles fly through gravitational wells instead of being trapped there. This means, in particular, that neutrinos cannot constitute a considerable part of dark matter.

4.1 Cold and warm dark matter

Currently, the most popular dark matter scenario is cold dark matter, CDM. It consists of particles which get out of *kinetic* equilibrium when they are non-relativistic. For dark matter particles Y which are initially in thermal equilibrium with cosmic plasma, this means that their scattering off other particles switches off at $T = T_{\rm d} \ll m_{\rm Y}$. Since then the dark matter particles move freely, their momenta decrease due to red shift, and they remain non-relativistic until now. Note that the *decoupling* temperature $T_{\rm d}$ may be much lower than the *freeze-out* temperature $T_{\rm f}$ at which the dark matter particles get out of *chemical* equilibrium, i.e., their number in the comoving volume freezes out (because, e.g., their creation and annihilation processes switch off). This is the case for many models with weakly interacting massive particles (WIMPs), a class of dark matter particles we discuss in some detail below. Note also that dark matter particles may never be in thermal equilibrium; this is the case, e.g., for axions.

An alternative to CDM is *warm dark matter*, WDM, whose particles decouple, being relativistic. Let us assume for definiteness that they are in kinetic equilibrium with cosmic plasma at temperature $T_{\rm f}$ when their number density freezes out (thermal relic). After kinetic equilibrium breaks down at temperature $T_{\rm d} \leq T_{\rm f}$, their spatial momenta decrease as a^{-1}, i.e., the momenta are of order T all the time after decoupling. Warm dark matter particles become non-relativistic at $T \sim m$, where m is their mass. Only after that do the WDM perturbations start to grow: as we mentioned above, relativistic particles escape from gravitational potentials, so the gravitational wells get smeared out instead of getting deeper. Before becoming non-relativistic, WDM particles travel the distance of the order of the horizon size; the WDM perturbations therefore are suppressed at those scales. The horizon size at the time $t_{\rm nr}$ when $T \sim m$ is of order

$$l_{\rm H}(t_{\rm nr}) \simeq H^{-1}(T \sim m) = \frac{M^*_{\rm Pl}}{T^2} \sim \frac{M^*_{\rm Pl}}{m^2} \ .$$

Due to the expansion of the Universe, the corresponding length at present is

$$l_0 = l_{\rm H}(t_{\rm nr}) \frac{a_0}{a(t_{\rm nr})} \sim l_{\rm H}(t_{\rm nr}) \frac{T}{T_0} \sim \frac{M_{\rm Pl}}{m T_0} \ , \tag{28}$$

where we neglected (rather weak) dependence on g_*. Hence, in the WDM scenario, structures of sizes smaller than l_0 are less abundant as compared to CDM. Let us point out that l_0 refers to the size of the perturbation in the linear regime; in other words, this is the size of the region from which matter collapses into a compact object.

There is a hint towards the plausibility of warm, rather than cold, dark matter. It is the dwarf-galaxy problem. According to numerical simulations, the CDM scenario tends to overproduce small objects—dwarf galaxies: it predicts hundreds of satellite dwarf galaxies in the vicinity of a large galaxy like the Milky Way, whereas only dozens of satellites have been observed so far. This argument is still controversial, but, if correct, it does suggest that the dark matter perturbations are suppressed at dwarf-galaxy scales. This is naturally the case in the WDM scenario. The present size of a dwarf galaxy is a few kpc, and the density is about 10^6 of the average density in the Universe. Hence, the size l_0 for these objects is of order 100 kpc $\simeq 3 \times 10^{23}$ cm. Requiring that perturbations of this size, but not much larger, are suppressed, we obtain from (28) the estimate for the mass of a dark matter particle

$$\text{WDM} \ : \quad m_{\rm DM} = 3\text{–}10 \ \text{keV} \ . \tag{29}$$

On the other hand, this effect is absent, i.e., dark matter is cold, for

$$\text{CDM} \ : \quad m_{\rm DM} \gg 10 \ \text{keV} \ . \tag{30}$$

Let us recall that these estimates apply to particles that are initially in kinetic equilibrium with cosmic plasma. They do *not* apply in the opposite case; an example is axion dark matter, which is cold despite being of very small axion mass.

4.2 WIMP miracle

There is a simple mechanism of the dark matter generation in the early Universe. It applies to *cold* dark matter. Because of its simplicity and robustness, it is considered by many as a very likely one, and the corresponding dark matter candidates—WIMPs—as the best candidates. Let us describe this mechanism in some detail.

Let us assume that there exists a heavy stable neutral particle Y, and that Y particles can only be destroyed or created via their pair annihilation or creation, with annihilation products being the particles of the Standard Model. The general scenario for the cosmological behaviour of Y particles is as follows. At high temperatures, $T \gg m_{\rm Y}$, the Y particles are in thermal equilibrium with the rest of the cosmic plasma; there are lots of Y particles in the plasma, which are continuously created and annihilate. As the temperature drops below $m_{\rm Y}$, the equilibrium number density decreases. At some 'freeze-out' temperature $T_{\rm f}$, the number density becomes so small that Y particles can no longer meet each other during the Hubble time, and their annihilation terminates. After that the number density of surviving Y particles decreases like a^{-3}, and these relic particles contribute to the mass density in the present Universe.

Let us estimate the properties of Y particles such that they really serve as dark matter. Elementary considerations of mean free path of a particle in gas give for the lifetime of a non-relativistic Y particle in cosmic plasma, $\tau_{\rm ann}$,

$$\langle \sigma_{\rm ann} \cdot v \rangle \cdot \tau_{\rm ann} \cdot n_{\rm Y} \sim 1 \ ,$$

where v is the relative velocity of Y particles, $\sigma_{\rm ann}$ is the annihilation cross-section at velocity v, averaging is over the velocity distribution of Y particles and $n_{\rm Y}$ is the number density. In thermal equilibrium at $T \ll m_{\rm Y}$, the latter is given by the Boltzmann law at zero chemical potential,

$$n_{\rm Y}^{\rm (eq)} = g_{\rm Y} \cdot \left(\frac{m_{\rm Y} T}{2\pi} \right)^{3/2} \mathrm{e}^{-\frac{m_{\rm Y}}{T}} \ , \tag{31}$$

where $g_{\rm Y}$ is the number of spin states of a Y particle. Let us introduce the notation

$$\langle \sigma_{\rm ann} \cdot v \rangle = \sigma_0$$

(in kinetic equilibrium, the left-hand side is the thermal average). If the annihilation occurs in an s-wave, then σ_0 is a constant independent of temperature; for a p-wave it is somewhat suppressed at $T \ll m_{\rm Y}$, namely $\sigma_0 \propto v^2 \propto T/m_{\rm Y}$. A quick way to come to correct estimate is to compare the lifetime with the Hubble time, or the annihilation rate $\Gamma_{\rm ann} \equiv \tau_{\rm ann}^{-1}$ with the expansion rate H. At $T \sim m_{\rm Y}$, the equilibrium density is of order $n_{\rm Y} \sim T^3$, and $\Gamma_{\rm ann} \gg H$ for not too small σ_0. This means that annihilation (and, by reciprocity, creation) of Y pairs is indeed rapid, and Y particles are indeed in complete thermal equilibrium with the plasma. At very low temperature, on the other hand, the equilibrium number density $n_{\rm Y}^{\rm (eq)}$ is exponentially small, and the equilibrium rate is small too, $\Gamma_{\rm ann}^{\rm (eq)} \ll H$. At low temperatures we cannot, of course, make use of the equilibrium formulas: Y particles no longer annihilate (and, by reciprocity, are no longer created), there is no thermal equilibrium with respect to creation–annihilation processes and the number density $n_{\rm Y}$ gets diluted only because of the cosmological expansion.

The freeze-out temperature $T_{\rm f}$ is determined by the relation[1]

$$\tau_{\rm ann}^{-1} \equiv \Gamma_{\rm ann} \simeq H \ , \tag{32}$$

where we use the equilibrium formulas. Making use of the relation (23) between the Hubble parameter and the temperature at radiation domination, we obtain

$$\sigma_0(T_{\rm f}) \cdot n_{\rm Y}(T_{\rm f}) \sim \frac{T_{\rm f}^2}{M_{\rm Pl}^*} \tag{33}$$

or

$$\sigma_0(T_{\rm f}) \cdot g_{\rm Y} \cdot \left(\frac{m_{\rm Y} T_{\rm f}}{2\pi}\right)^{3/2} {\rm e}^{-\frac{m_{\rm Y}}{T_{\rm f}}} \sim \frac{T_{\rm f}^2}{M_{\rm Pl}^*} \ . \tag{34}$$

The latter equation gives the freeze-out temperature, which, up to log–log corrections, is

$$T_{\rm f} \approx \frac{m_{\rm Y}}{\ln(M_{\rm Pl}^* m_{\rm Y} \sigma_0)} \tag{35}$$

(the possible dependence of σ_0 on temperature is irrelevant in the right-hand side: we are doing the calculation in the leading-log approximation anyway). Note that this temperature is somewhat lower than $m_{\rm Y}$ if the relevant microscopic mass scale is much below $M_{\rm Pl}$. This means that Y particles freeze out when they are indeed non-relativistic and get out of kinetic equilibrium at even lower temperature, hence the term 'cold dark matter'. The fact that the annihilation and creation of Y particles terminate at a relatively low temperature has to do with the rather slow expansion of the Universe, which should be compensated for by the smallness of the number density $n_{\rm Y}$.

At the freeze-out temperature, we make use of Eq. (33) and obtain

$$n_{\rm Y}(T_{\rm f}) = \frac{T_{\rm f}^2}{M_{\rm Pl}^* \sigma_0(T_{\rm f})} \ . \tag{36}$$

Note that this density is inversely proportional to the annihilation cross-section (modulo a logarithm). The reason is that for higher annihilation cross-sections, the creation–annihilation processes are longer in equilibrium, and fewer Y particles survive.

[1] In fact, we somewhat oversimplify the analysis here. The chemical equilibrium breaks down slightly earlier than what we find from Eq. (32): the corresponding temperature is obtained by equating the equilibrium creation–annihilation rate $\Gamma_{\rm ann}$ to the rate of evolution of the equilibrium number density (31), rather than to the Hubble parameter H. For $T \ll m_{\rm Y}$, this gives the equation for the temperature

$$\Gamma_{\rm ann} \simeq \frac{\dot{n}_{\rm Y}}{n_{\rm Y}} \simeq -\frac{m_{\rm Y}}{T}\frac{\dot{T}}{T} = \frac{m_{\rm Y}}{T} H(T) \ .$$

This temperature differs by the log–log correction from $T_{\rm f}$ determined from Eq. (34) and, at this temperature, one has $n_{\rm Y} \gg T^2/(M_{\rm Pl}^* \sigma_0)$, cf. Eq. (36). However, below this temperature, the annihilation of Y particles continues, and it terminates at temperature $T_{\rm f}$ determined by Eq. (32), which gives Eqs. (33) and (36). All this gives rise to log–log corrections, which we do not calculate anyway. So, our estimate for the present dark matter mass density remains valid.

Up to a numerical factor of order 1, the number-to-entropy ratio at freeze-out is

$$\frac{n_{\rm Y}}{s} \simeq \frac{1}{g_*(T_{\rm f}) M_{\rm Pl}^* T_{\rm f} \sigma_0(T_{\rm f})} \,. \tag{37}$$

This ratio stays constant until the present time, so the present number density of Y particles is $n_{\rm Y,0} = s_0 \cdot (n_{\rm Y}/s)_{\text{freeze-out}}$, and the mass-to-entropy ratio is

$$\frac{\rho_{\rm Y,0}}{s_0} = \frac{m_{\rm Y} n_{\rm Y,0}}{s_0} \simeq \frac{\ln(M_{\rm Pl}^* m_{\rm Y} \sigma_0)}{g_*(T_{\rm f}) M_{\rm Pl}^* \sigma_0(T_{\rm f})} \simeq \frac{\ln(M_{\rm Pl}^* m_{\rm Y} \sigma_0)}{\sqrt{g_*(T_{\rm f})} M_{\rm Pl} \sigma_0(T_{\rm f})} \,,$$

where we made use of (35). This formula is remarkable. The mass density depends mostly on one parameter, the annihilation cross-section σ_0. The dependence on the mass of a Y particle is through the logarithm and through $g_*(T_{\rm f})$; it is very mild. The value of the logarithm here is between 30 and 40, depending on parameters (this means, in particular, that freeze-out occurs when the temperature drops 30 to 40 times below the mass of a Y particle). Inserting $g_*(T_{\rm f}) \sim 100$, as well as the numerical factor omitted in Eq. (37), and comparing with (26), we obtain the estimate

$$\sigma_0(T_{\rm f}) \equiv \langle \sigma v \rangle (T_{\rm f}) = (1\text{–}2) \times 10^{-36} \text{ cm}^2 \,. \tag{38}$$

This is a weak-scale cross-section, which tells us that the relevant energy scale is TeV. We note in passing that the estimate (38) is quite precise and robust.

If the annihilation occurs in an s-wave, the annihilation cross-section may be parametrized as $\sigma_0 = \alpha^2/M^2$, where α is some coupling constant and M is a mass scale (which may be higher than $m_{\rm Y}$). This parametrization is suggested by the picture of Y-pair annihilation via the exchange by another particle of mass M. With $\alpha \sim 10^{-2}$, the estimate for the mass scale is roughly $M \sim 1$ TeV. Thus, with very mild assumptions, we find that the non-baryonic dark matter may naturally originate from the TeV-scale physics. In fact, what we have found can be understood as an approximate equality between the cosmological parameter, the mass-to-entropy ratio of dark matter and the particle-physics parameters,

$$\text{mass-to-entropy} \simeq \frac{1}{M_{\rm Pl}} \left(\frac{\text{TeV}}{\alpha_{\rm W}} \right)^2 \,.$$

Both are of order 10^{-10} GeV, and it is very tempting to think that this 'WIMP miracle' is not a mere coincidence. If it is not, the dark matter particles should be found at the LHC.

The most prominent candidate for WIMPs is neutralinos of the supersymmetric extensions of the Standard Model. The situation with neutralinos is somewhat tense, however. The point is that the pair annihilation of neutralinos often occurs in the p-wave, rather than the s-wave. This gives the suppression factor in $\sigma_0 \equiv \langle \sigma_{\rm ann} v \rangle$ proportional to $v^2 \sim T_{\rm f}/m_{\rm Y} \sim 1/30$. Hence, neutralinos tend to be overproduced in most of the parameter space of the Minimal Supersymmetric Standard Model (MSSM) and other models. Yet neutralinos remain a good candidate, especially at high $\tan\beta$.

A direct search for dark matter WIMPs is underway in underground laboratories. The idea is that WIMPs orbiting around the centre of our Galaxy with velocity of order 10^{-3} sometimes hit a nucleus in a detector and deposit a small energy in it. These searches have become sensitive to neutralinos, as shown in Fig. 9. Indirect searches for dark matter WIMPs include the search for neutrinos coming from the centres of the Earth and Sun (WIMPs may concentrate and annihilate there), see, e.g., Ref. [36] and positrons and antiprotons in cosmic rays (produced in WIMP annihilations in our Galaxy), see, e.g., Ref. [37]. Collider searches are sensitive to WIMPs too, see Fig. 10. We conclude that the hunt for WIMPs has entered the promising stage.

Question. Estimate the energy deposited in the XENON detector due to elastic scattering of a dark matter WIMP, for WIMP masses 10 GeV, 100 GeV and 1 TeV. Estimate the number of events per kilogram per

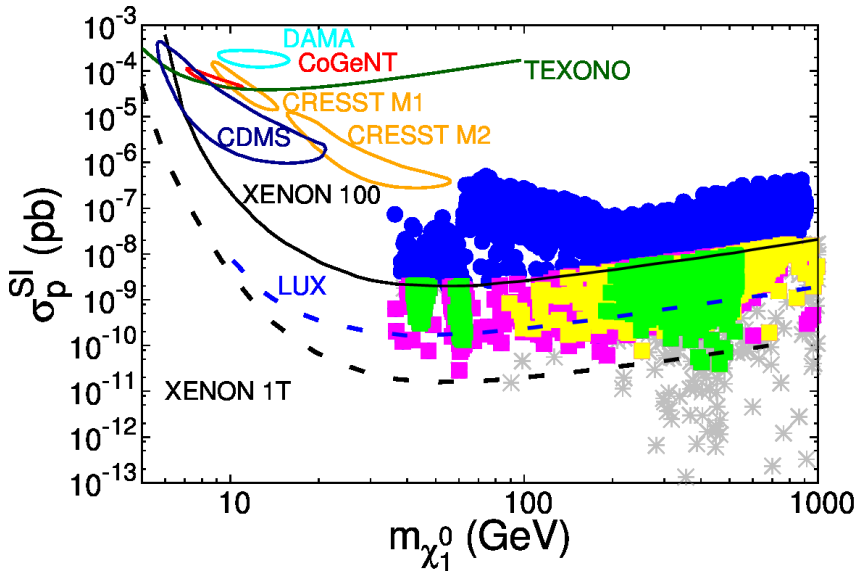

Fig. 9: MSSM predictions for spin-independent elastic neutralino–nucleon cross-section versus neutralino mass and experimentally excluded regions [35]. Shaded regions correspond to MSSM parameters consistent with collider limits and yielding $\Omega_{\rm DM} \approx 0.25$. Regions above the open solid lines are ruled out by direct searches, closed solid curves correspond to regions favoured by experiments indicated. Dashed lines are sensitivities of future direct search experiments LUX and XENON 1T.

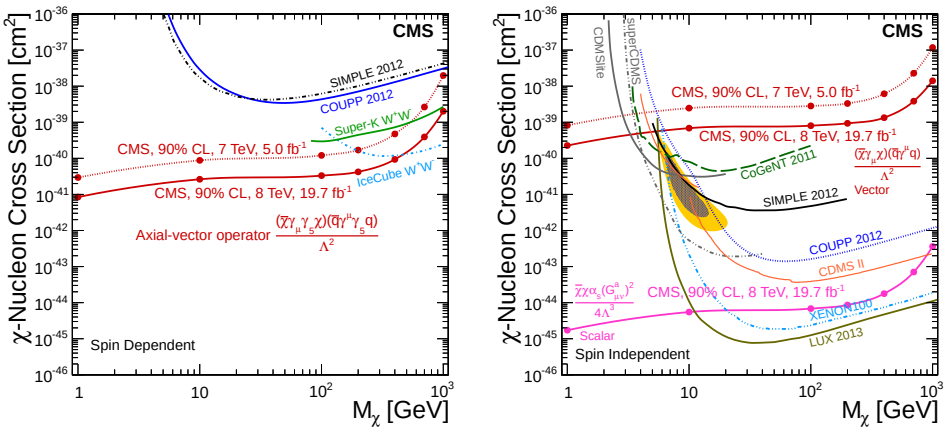

Fig. 10: Excluded regions in the parameter space $(M_{\rm X}, \sigma_{\rm pX})$ [38] for spin-dependent (left) and spin-independent (right) WIMP interactions with nucleons. Regions above the curves are ruled out at 90 % confidence level. CMS denotes searches for WIMPs at the LHC (assuming contact interaction $\mathrm{YYf_1f_2}$, where $\mathrm{f_{1,2}}$ are Standard Model fermions); IceCube and Super-K are searches for neutrinos from WIMP annihilation in the Sun; others are direct searches. The shaded region in the middle of the right-hand panel is favoured by a possible signal at the CDMS experiment.

year for the same masses and elastic cross-sections 10^{-5} pb, 10^{-9} pb and 10^{-8} pb, respectively (see Fig. 9), assuming that the WIMP mass density around the Earth is similar to the average baryon mass density, $\rho_{\rm DM} \sim 0.3$ GeV cm^{-3}, and that $v_{\rm DM} \sim 10^{-3}$.

4.3 Light long-lived particles

Many extensions of the Standard Model contain light scalar or pseudoscalar particles. In some models these new particles are so weakly interacting that their lifetime exceeds the present age of the Universe. Hence, they may serve as dark matter candidates. The best motivated of them is the axion, but there is an entire zoo of axion-like particles.

Let us consider general properties of models with light scalars or pseudoscalars. These particles should interact with the usual matter very weakly, so they must be neutral with respect to the Standard Model gauge interactions. This implies that interactions of scalars S and pseudoscalars P with gauge fields are of the form

$$\mathcal{L}_{SFF} = \frac{C_{SFF}}{4\Lambda} \cdot S F_{\mu\nu} F^{\mu\nu} \ , \quad \mathcal{L}_{PFF} = \frac{C_{PFF}}{8\Lambda} \cdot P F_{\mu\nu} F_{\lambda\rho} \epsilon^{\mu\nu\lambda\rho} \ , \tag{39}$$

where $F_{\mu\nu}$ is the field strength of the $\mathrm{SU}(3)_\mathrm{c}$, $\mathrm{SU}(2)_\mathrm{W}$ or $\mathrm{U}(1)_\mathrm{Y}$ gauge group. The parameter Λ has dimension of mass and can be interpreted as the scale of new physics related to an S and/or a P particle. This parameter has to be large; then the interactions of S and P with gauge bosons are indeed weak at low energies. Because of that, the Lagrangians (39) contain gauge-invariant operators of the lowest possible dimension. Dimensionless constants C_{SFF} and C_{PFF} are typically numbers of order 1. The terms (39) describe interactions of (pseudo)scalars with pairs of photons, gluons as well as with $\mathrm{Z}\gamma$, ZZ and $\mathrm{W}^+\mathrm{W}^-$ pairs.

Interactions with fermions can also be written on symmetry grounds. Since S and P are singlets under $\mathrm{SU}(3)_\mathrm{c} \times \mathrm{SU}(2)_\mathrm{W} \times \mathrm{U}(1)_\mathrm{Y}$, no combinations like $S\bar{f}f$ or $P\bar{f}\gamma^5 f$ are gauge invariant, so they cannot appear in the Lagrangian (hereafter f denotes the Standard Model fermions). Gauge-invariant operators of the lowest dimension have the form $H\bar{f}f$, where H is the Englert–Brout–Higgs field. Hence, the interactions with fermions are

$$\mathcal{L}SHff = \frac{Y_{SHff}}{\Lambda} \cdot SH\bar{f}f \ , \quad \mathcal{L}PHff = \frac{Y_{PHff}}{\Lambda} \cdot PH\bar{f}\gamma^5 f \ .$$

It often happens that the couplings Y_{SHff} and Y_{SPff} are of the order of the Standard Model Yukawa couplings, so upon electroweak symmetry breaking the low-energy Lagrangians have the following structure:

$$\mathcal{L}Sff = \frac{C_{Sff} m_f}{\Lambda} \cdot S\bar{f}f \ , \quad \mathcal{L}Pff = \frac{C_{Pff} m_f}{\Lambda} \cdot P\bar{f}\gamma^5 f \ , \tag{40}$$

where we assume that the dimensionless couplings C_{Sff} and C_{Pff} are also of order 1.

Making use of Eqs. (39) and (40), we estimate the partial widths of decays of P and S into the Standard Model particles:

$$\Gamma_{P(S)\to AA} \sim \frac{m_{P(S)}^3}{64\pi\Lambda^2} \ , \quad \Gamma_{P(S)\to ff} \sim \frac{m_f^2 m_{P(S)}}{8\pi\Lambda^2} \ , \tag{41}$$

where A denotes vector bosons. By requiring that the lifetime of the new particles exceeds the present age of the Universe, $\tau_{S(P)} = \Gamma_{S(P)}^{-1} > H_0^{-1}$, we find a bound on the mass of the dark matter candidates,

$$m_{P(S)} < \left(16\pi\Lambda^2 H_0\right)^{1/3} \ . \tag{42}$$

Assuming that the new physics scale is below the Planck scale, $\Lambda < M_\mathrm{Pl}$, we obtain an (almost) model-independent bound,

$$m_{P(S)} < 100 \text{ MeV} \ . \tag{43}$$

Hence, the kinematically allowed decays are $P(S) \to \gamma\gamma$, $P(S) \to \nu\bar{\nu}$ and $P(S) \to \mathrm{e}^+\mathrm{e}^-$. It follows from Eq. (41) that the two-photon decay mode dominates, unless the mass of the new particle is close to that of the electron.

Let us now consider generation of relic (pseudo)scalars in the early Universe. There are several generation mechanisms; one of them is fairly generic for the class of models we discuss. This is generation in decays of condensates (we will consider another mechanism later, in the model with axions). The picture is as follows. Let some scalar field ϕ be in a condensate in the early Universe. The condensate can be viewed as a collection of ϕ particles at rest. Equivalently, the condensate is the homogeneous scalar field that oscillates at relatively late times, when $m_\phi > H$. Let both particles, ϕ and S, interact with matter so weakly that they never get into thermal equilibrium, and let the interaction between ϕ and S have the form $\mu\phi S^2/2$, where μ is the coupling constant. Then the width of the decay $\phi \to SS$ is estimated as

$$\Gamma_{\phi\to SS} \sim \frac{\mu^2}{16\pi m_\phi} \,. \tag{44}$$

If the widths of other decay channels do not exceed the value (44), the decay of the ϕ condensate occurs at a temperature T_ϕ determined by

$$\Gamma_{\phi\to SS} \sim H(T_\phi) = \frac{T_\phi^2}{M_{\rm Pl}^*} \,.$$

Let the energy density of the ϕ condensate at that time be equal to ρ_ϕ, so that the number density of decaying ϕ particles is $n_\phi \sim \rho_\phi/m_\phi$. Immediately after the epoch of ϕ-particle decays, the number density of S particles is of order $\epsilon\rho_\phi/m_\phi$, where ϵ is the fraction of the condensate that decayed into S particles. After S particles become non-relativistic, their mass density is of order

$$\rho_S \sim \epsilon\rho_\phi \cdot \frac{m_S T^3}{m_\phi T_\phi^3} \,,$$

where we omitted the dependence on g_* for simplicity. In this way we estimate the mass fraction of S particles today,

$$\Omega_S = \frac{\rho_S}{\rho_{\rm c}} \sim \frac{m_S T_0^3}{\rho_{\rm c}} \cdot \frac{\epsilon\rho_\phi}{m_\phi T_\phi^3} \sim 0.2 \cdot \left(\frac{m_{\rm S}}{1\ {\rm eV}}\right) \cdot \frac{\epsilon\rho_\phi}{m_\phi T_\phi^3} \,. \tag{45}$$

With an appropriate choice of parameters, the correct value $\Omega_S \simeq 0.2$ can indeed be obtained. We note that the last factor on the right-hand side of Eq. (45) must be small.

4.4 Axions

Let us now turn to a concrete class of models with Peccei–Quinn symmetry and axions. This symmetry provides a solution to the *strong CP-problem,* and the existence of axions is an inevitable consequence of the construction.

The strong CP-problem [39–41] emerges in the following way. One can extend the Standard Model Lagrangian by adding the following term:

$$\Delta L = \frac{\alpha_{\rm s}}{8\pi} \cdot \theta_0 \cdot G^{\rm a}_{\mu\nu}\tilde{G}^{\mu\nu\ {\rm a}} \,, \tag{46}$$

where $\alpha_{\rm s}$ is the SU(3)$_{\rm c}$ gauge coupling, $G^{\rm a}_{\mu\nu}$ is the gluon field strength, $\tilde{G}^{\mu\nu\ {\rm a}} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}G^{\rm a}_{\lambda\rho}$ is the dual tensor and θ_0 is an arbitrary dimensionless parameter (the factor $\alpha_{\rm s}/(8\pi)$ is introduced for later convenience). The interaction term (46) is invariant under gauge symmetries of the Standard Model, but it violates P and CP. The term (46) is a total derivative, so it does not contribute to the classical field equations, and its contribution to the action is reduced to the surface integral. For any perturbative gauge field configurations (small perturbations about $G^{\rm a}_\mu = 0$), this contribution is equal to zero. However, this is not the case for configurations of instanton type. This means that CP is violated in QCD at the non-perturbative level.

Furthermore, quantum effects due to quarks give rise to the anomalous term in the Lagrangian, which has the same form as Eq. (46) with proportionality coefficient determined by the phase of the quark mass matrix $\hat{M}_{\rm q}$. The latter enters the Lagrangian as

$$\mathcal{L}{\rm m} = \bar{q}_{\rm L}\hat{M}_{\rm q}q_{\rm R} + {\rm h.c.}$$

By chiral rotation of quark fields, one makes quark masses real (i.e., physical), but that rotation induces a new term in the Lagrangian,

$$\Delta\mathcal{L}{\rm m} = \frac{\alpha_{\rm s}}{8\pi}\cdot{\rm Arg}\left({\rm Det}\hat{M}_{\rm q}\right)\cdot G^{\rm a}_{\mu\nu}\tilde{G}^{\mu\nu\ {\rm a}}\ . \tag{47}$$

There is no reason to think that ${\rm Arg}\left({\rm Det}\hat{M}_{\rm q}\right) = 0$. Neither there is a reason to think that the 'tree-level' term (46) and the anomalous contribution (47) cancel each other. Indeed, the former term is there even in the absence of quarks, while the latter comes from the Yukawa sector, as the quark masses are due to their Yukawa interactions with the Englert–Brout–Higgs field.

Thus, the Standard Model Lagrangian should contain the term

$$\Delta\mathcal{L}\theta = \frac{\alpha_{\rm s}}{8\pi}\left(\theta_0 + {\rm Arg}\left({\rm Det}\hat{M}_{\rm q}\right)\right)G^{\rm a}_{\mu\nu}\tilde{G}^{\mu\nu\ {\rm a}} \equiv \frac{\alpha_{\rm s}}{8\pi}\cdot\theta\cdot G^{\rm a}_{\mu\nu}\tilde{G}^{\mu\nu\ {\rm a}}\ . \tag{48}$$

This term violates CP, and off hand the parameter θ is of order 1.

The term (48) has non-trivial phenomenological consequences. One is that it generates the electric dipole moment (EDM) of the neutron, $d_{\rm n}$, which is estimated as [42]

$$d_{\rm n} \sim \theta\times 10^{-16}\,e\ {\rm cm}\ . \tag{49}$$

The neutron EDM has not been found experimentally, and the searches place a strong bound

$$d_{\rm n} \lesssim 3\times 10^{-26}\,e\ {\rm cm}\ . \tag{50}$$

This leads to the bound on the parameter θ,

$$|\theta| < 0.3\times 10^{-9}\ .$$

The problem to explain such a small value of θ is precisely the strong CP-problem.

A solution to this problem does not exist within the Standard Model. The solution is offered by models with axions. These models make use of the following observation. If at the classical level the quark Lagrangian is invariant under axial symmetry ${\rm U}(1)_A$ such that

$$q_{\rm L}\to {\rm e}^{{\rm i}\beta}q_{\rm L}\ ,\qquad q_{\rm R}\to {\rm e}^{-{\rm i}\beta}q_{\rm R}\ , \tag{51}$$

then the θ term would be rotated away by applying this transformation. This global symmetry is called the Peccei–Quinn (PQ) symmetry [43], ${\rm U}(1)_{\rm PQ}$. There is no PQ symmetry in the Standard Model, but one can extend the Standard Model in such a way that the classical Lagrangian is invariant under the PQ symmetry. Quark masses are not invariant under the PQ transformations (51), so PQ symmetry is *spontaneously broken*. At the classical level, this leads to the existence of a massless Nambu–Goldstone field $a(x)$, an axion. As for any Nambu–Goldstone field, its properties are determined by its transformation law under the PQ symmetry:

$$a(x)\to a(x) + \beta\cdot f_{\rm PQ}\ , \tag{52}$$

where β is the same parameter as in Eq. (51) and $f_{\rm PQ}$ is a constant of dimension of mass, the energy scale of ${\rm U}(1)_{\rm PQ}$ symmetry breaking. The mass terms in the low-energy quark Lagrangian must be

symmetric under the transformations (51) and (52), so the quark and axion fields enter the Lagrangian in the combination

$$\mathcal{L}\mathrm{m} = \bar{q}_\mathrm{R} m_\mathrm{q} \mathrm{e}^{-2\mathrm{i}\frac{a}{f_\mathrm{PQ}}} q_\mathrm{L} + \mathrm{h.c.} \tag{53}$$

Making use of Eq. (47), we find that at the quantum level the low-energy Lagrangian contains the term

$$\mathcal{L}\mathrm{a} = C_\mathrm{g} \frac{\alpha_\mathrm{s}}{8\pi} \cdot \frac{a}{f_\mathrm{PQ}} G^\mathrm{a}_{\mu\nu} \tilde{G}^{\mu\nu\ \mathrm{a}} \ , \tag{54}$$

where the constant C_g is of order 1; it is determined by PQ charges of quarks. Clearly, PQ symmetry (51) and (52) is *explicitly* broken by quantum effects of QCD, and an axion is a *pseudo*-Nambu–Goldstone boson.

Hence, the θ parameter multiplying the operator $G^\mathrm{a}_{\mu\nu}\tilde{G}^{\mu\nu\ \mathrm{a}}$ obtains a shift depending on the space–time point and proportional to the axion field,

$$\theta \to \bar{\theta}(x) = \theta + C_\mathrm{g} \frac{a(x)}{f_\mathrm{PQ}} \ . \tag{55}$$

Strong interactions would conserve CP provided the axion vacuum expectation value is such that $\langle\bar{\theta}\rangle = 0$. The QCD effects indeed do the job. They generate a non-vanishing quark condensate $\langle\bar{q}q\rangle \sim \Lambda^3_\mathrm{QCD}$ at the QCD energy scale $\Lambda_\mathrm{QCD} \sim 200$ MeV. This condensate breaks chiral symmetry and in turn generates the axion effective potential

$$V_\mathrm{a} \sim -\frac{1}{2}\bar{\theta}^2 \frac{m_\mathrm{u} m_\mathrm{d}}{m_\mathrm{u} + m_\mathrm{d}} \langle\bar{q}q\rangle + \mathcal{O}(\bar{\theta}^4) \simeq \frac{1}{8}\bar{\theta}^2 \cdot m_\pi^2 f_\pi^2 + \mathcal{O}(\bar{\theta}^4) \ , \tag{56}$$

where $m_\pi = 135$ MeV and $f_\pi = 93$ MeV are pion mass and decay constant. In fact, the axion potential must be periodic in θ with period 2π, so the expression (56) is valid for small θ only. The potential has the minimum at $\langle\bar{\theta}\rangle = 0$, so the strong CP-problem finds an elegant solution. It follows from Eqs. (55) and (56) that the axion has a mass

$$m_\mathrm{a} \approx C_\mathrm{g} \frac{m_\pi f_\pi}{2 f_\mathrm{PQ}} \ , \tag{57}$$

i.e., it is indeed a *pseudo*-Nambu–Goldstone boson.

There are various ways to implement the PQ mechanism. One is to introduce two Englert–Brout–Higgs doublets and choose the Yukawa interaction as

$$Y^\mathrm{d} \bar{Q}_\mathrm{L} H_1 D_\mathrm{R} + Y^\mathrm{u} \bar{Q}_\mathrm{L} \mathrm{i}\tau^2 H_2^* U_\mathrm{R} \ . \tag{58}$$

The two scalar fields transform under the $\mathrm{U}(1)_\mathrm{PQ}$ transformation (51) as follows:

$$H_1 \to \mathrm{e}^{2\mathrm{i}\beta} H_1 \ , \quad H_2 \to \mathrm{e}^{-2\mathrm{i}\beta} H_2 \ .$$

This ensures $\mathrm{U}(1)_\mathrm{PQ}$ invariance of the Lagrangian (58) and hence the absence of the θ term. Both scalars acquire vacuum expectation values v_1 and v_2. If no other new fields are added, we arrive at the Weinberg–Wilczek model [44, 45]. In that case, the axion field θ is the relative phase of H_1 and H_2, and the PQ scale equals the electroweak scale:

$$f_\mathrm{PQ} = 2\sqrt{v_1^2 + v_2^2} = 2v_\mathrm{SM} = 2 \times 246 \text{ GeV} \ .$$

The axion is quite heavy, $m_\mathrm{a} \sim 15$ keV, and its interaction with quarks, gluons and photons is too strong. Because of that, the Weinberg–Wilczek axion is experimentally ruled out.

This problem is solved in the Dine–Fischler–Srednicki–Zhitnitsky (DFSZ) model [46, 47] by adding a complex scalar field S which is a singlet under the Standard Model gauge group. Its interactions involve PQ invariants

$$S^\dagger S \ , \quad H_1^\dagger H_2 \cdot S^2 \ .$$

The field S transforms under $\mathrm{U}(1)_{\mathrm{PQ}}$ as $S \to \mathrm{e}^{2\mathrm{i}\beta} S$. The axion field is now a linear combination of the phases of fields H_1, H_2 and S and

$$f_{\mathrm{PQ}} = 2\sqrt{v_1^2 + v_2^2 + v_{\mathrm{s}}^2}\,, \tag{59}$$

where v_{s} is the vacuum expectation value of the field S. The latter can be large, so it is clear from Eq. (59) that the mass of the axion is small and, most importantly, its couplings to the Standard Model fields are weak: these couplings are inversely proportional to $f_{\mathrm{PQ}} \sim v_{\mathrm{s}}$. The DFSZ axion interacts with both quarks and leptons.

Another approach is called the Kim–Shifman–Vainshtein–Zakharov (KSVZ) mechanism [48, 49]. It does not require more than one Englert–Brout–Higgs field of the Standard Model. The mechanism makes use of additional quark fields Ψ_{R} and Ψ_{L}, which are triplets under $\mathrm{SU}(3)_{\mathrm{c}}$ and singlets under $\mathrm{SU}(2)_{\mathrm{W}} \times \mathrm{U}(1)_{\mathrm{Y}}$. Only these quarks transform non-trivially under $\mathrm{U}(1)_{\mathrm{PQ}}$, while the usual quarks have zero PQ charge. One also introduces a complex scalar field S, which is a singlet under the Standard Model gauge group. One writes the PQ-invariant Yukawa interaction of the new fields,

$$L = y_\Psi S \bar{\Psi}_{\mathrm{R}} \Psi_{\mathrm{L}} + \text{h.c.}\,,$$

so that S again transforms under $\mathrm{U}(1)_{\mathrm{PQ}}$ as $S \to \mathrm{e}^{2\mathrm{i}\beta} S$. PQ symmetry is spontaneously broken by the vacuum expectation value $\langle S \rangle = v_{\mathrm{s}}/\sqrt{2}$. The axion here is the phase of the field S; therefore,

$$f_{\mathrm{PQ}} = 2v_{\mathrm{s}}\,. \tag{60}$$

The KSVZ model does not contain an explicit interaction of an axion with the usual quarks and leptons.

To summarize, an axion is a light particle whose interactions with the Standard Model fields are very weak. The latter property relates to the fact that it is a pseudo-Nambu–Goldstone boson of a global symmetry spontaneously broken at the high-energy scale $f_{\mathrm{PQ}} \gg M_{\mathrm{W}}$. As for any Nambu–Goldstone field, the interactions of an axion with quarks and leptons are described by the generalized Goldberger–Treiman formula

$$\mathcal{L}\mathrm{af} = \frac{1}{f_{\mathrm{PQ}}} \cdot \partial_\mu a \cdot J^\mu_{\mathrm{PQ}}\,. \tag{61}$$

Here

$$J^\mu_{\mathrm{PQ}} = \sum_{\mathrm{f}} e_{\mathrm{f}}^{(\mathrm{PQ})} \cdot \bar{f}\gamma^\mu\gamma^5 f\,. \tag{62}$$

The contributions of fermions to the current J^μ_{PQ} are proportional to their PQ charges $e_{\mathrm{f}}^{(\mathrm{PQ})}$; these charges are model-dependent. In accord with Eq. (53), the action (61) can be integrated by parts and we obtain instead

$$\begin{aligned}\mathcal{L}\mathrm{af} &= -\frac{1}{f_{\mathrm{PQ}}} \cdot a \cdot \partial_\mu J^\mu_{\mathrm{PQ}} \\ &= -\frac{a}{f_{\mathrm{PQ}}} \cdot \sum_f 2e_f^{(\mathrm{PQ})} m_f \cdot \bar{f}\gamma^5 f\,.\end{aligned} \tag{63}$$

Besides the interaction (61), there are also interactions of axions with gluons, see Eq. (54), and photons,

$$\mathcal{L}\mathrm{ag} = C_{\mathrm{g}} \frac{\alpha_{\mathrm{s}}}{8\pi} \cdot \frac{a}{f_{\mathrm{PQ}}} \cdot G^{\mathrm{a}}_{\mu\nu} \tilde{G}^{\mu\nu\,\mathrm{a}}\,, \qquad \mathcal{L}\mathrm{a}\gamma = C_\gamma \frac{\alpha}{8\pi} \cdot \frac{a}{f_{\mathrm{PQ}}} \cdot F_{\mu\nu} \tilde{F}^{\mu\nu}\,, \tag{64}$$

where the dimensionless constants C_{g} and C_γ are also model-dependent and, generally speaking, are of order 1. The interaction terms (63) and (64) indeed have the form (39) and (40), i.e., models with axions

belong to the class of models with light, weakly interacting pseudoscalars. The axion mass, however, is not a free parameter: we find from Eq. (57) that

$$m_{\rm a} \approx m_\pi \cdot \frac{f_\pi}{2f_{\rm PQ}} \approx 0.6 \text{ eV} \cdot \left(\frac{10^7 \text{ GeV}}{f_{\rm PQ}}\right) . \tag{65}$$

The main decay channel of the light axion is decay into two photons. The lifetime $\tau_{\rm a}$ is found from Eq. (41) by setting $\Lambda = 2\pi f_{\rm PQ}/\alpha$ and using Eq. (65),

$$\tau_{\rm a} = \frac{1}{\Gamma_{{\rm a}\to\gamma\gamma}} = \frac{64\pi^3 m_\pi^2 f_\pi^2}{\alpha^2 m_{\rm a}^5} \simeq 4 \times 10^{24} \text{ s} \cdot \left(\frac{\text{eV}}{m_{\rm a}}\right)^5 .$$

By requiring that this lifetime exceeds the age of the Universe, $\tau_{\rm a} > t_0 \approx 14$ billion years, we find the bound on the mass of the axion as a dark matter candidate,

$$m_{\rm a} < 25 \text{ eV} . \tag{66}$$

There are astrophysical bounds on the strength of axion interactions $f_{\rm PQ}^{-1}$ and hence on the axion mass. Axions in theories with $f_{\rm PQ} \lesssim 10^9$ GeV, which are heavier than 10^{-2} eV, would be intensely produced in stars and supernovae explosions. This would lead to contradictions with observations. So, we are left with very light axions, $m_{\rm a} \lesssim 10^{-2}$ eV.

As far as dark matter is concerned, thermal production of axions is irrelevant. There are at least two mechanisms of axion production in the early Universe that can provide not only right axion abundance but also small initial velocities of axions. The latter property makes an axion a *cold* dark matter candidate, despite its very small mass. One mechanism has to do with decays of global strings [50]—topological defects that exist in theories with spontaneously broken global U(1) symmetry ($\rm U(1)_{PQ}$ in our case; for a discussion of this mechanism, see, e.g., Ref. [51]). Another mechanism employs an axion condensate [52–54], an homogeneous axion field that oscillates in time after the QCD epoch. This is called the axion misalignment mechanism. Let us consider the second mechanism in some detail.

As we have seen in Eq. (56), the axion potential is proportional to the quark condensate $\langle \bar{q}q \rangle$. This condensate breaks chiral symmetry. The chiral symmetry is in fact restored at high temperatures Hence, one expects that the axion potential is negligibly small at $T \gg \Lambda_{\rm QCD}$. This is indeed the case: the effective potential for the field $\bar\theta = \theta + a/f_{\rm PQ}$ vanishes at high temperatures, and this field can take any value,

$$\bar\theta_{\rm i} \in [0\,, 2\pi) \ ,$$

where we recall that the field $\bar\theta$ is a phase. There is no reason to think that the initial value $\bar\theta_{\rm i}$ is zero. As the temperature decreases, the axion mass $m(T)$ starts to get generated, so that

$$m_{\rm a}(T) \simeq 0 \qquad \text{at } T \gg \Lambda_{\rm QCD} ,$$
$$m_{\rm a}(T) \simeq m_{\rm a} \qquad \text{at } T \ll \Lambda_{\rm QCD} .$$

Hereafter $m_{\rm a}$ denotes the zero-temperature axion mass. As the mass increases, at some point the field $\bar\theta$, remaining homogeneous, starts to roll down from $\bar\theta_{\rm i}$ towards its value $\bar\theta = 0$ at the minimum of the potential. The axion field practically does not evolve when $m_{\rm a}(T) \ll H(T)$ and at the time when $m_{\rm a}(T) \sim H(T)$ it starts to oscillate. Let us estimate the present energy density of the axion field in this picture, without using the concrete form of the function $m(T)$.

The oscillations start at the time $t_{\rm osc}$ when

$$m_{\rm a}(t_{\rm osc}) \sim H(t_{\rm osc}). \tag{67}$$

At this time, the energy density of the axion field is estimated as

$$\rho_{\rm a}(t_{\rm osc}) \sim m_{\rm a}^2(t_{\rm osc}) f_{\rm PQ}^2 \bar\theta_{\rm i}^2 .$$

The oscillating axion field is the same thing as a collection of axions at rest. Their number density at the beginning of oscillations is estimated as

$$n_{\rm a}(t_{\rm osc}) \sim \frac{\rho_{\rm a}(t_{\rm osc})}{m_{\rm a}(t_{\rm osc})} \sim m_{\rm a}(t_{\rm osc}) f_{\rm PQ}^2 \bar{\theta}_{\rm i}^2 \sim H(t_{\rm osc}) f_{\rm PQ}^2 \bar{\theta}_{\rm i}^2 \ .$$

This number density, as any number density of non-relativistic particles, then decreases as a^{-3}.

The axion-to-entropy ratio at time $t_{\rm osc}$ is

$$\frac{n_{\rm a}}{s} \sim \frac{H(t_{\rm osc}) f_{\rm PQ}^2}{\frac{2\pi^2}{45} g_* T_{\rm osc}^3} \cdot \bar{\theta}_{\rm i}^2 \simeq \frac{f_{\rm PQ}^2}{\sqrt{g_*} T_{\rm osc} M_{\rm Pl}} \cdot \bar{\theta}_{\rm i}^2 \ ,$$

where we use the usual relation $H = 1.66\sqrt{g_*}T^2/M_{\rm Pl}$. The axion-to-entropy ratio remains constant after the beginning of oscillations, so the present mass density of axions is

$$\rho_{\rm a,0} = \frac{n_{\rm a}}{s} m_{\rm a} s_0 \simeq \frac{m_{\rm a} f_{\rm PQ}^2}{\sqrt{g_*} T_{\rm osc} M_{\rm Pl}} s_0 \cdot \bar{\theta}_{\rm i}^2 \ . \tag{68}$$

In fact, it is a decreasing function of $m_{\rm a}$. Indeed, $f_{\rm PQ}$ is inversely proportional to $m_{\rm a}$, see Eq. (57); at the same time, the axion obtains its mass near the epoch of QCD transition, i.e., at $T \sim \Lambda_{\rm QCD}$, so $T_{\rm osc}$ depends on $m_{\rm a}$ rather weakly.

To obtain a simple estimate, let us set $T_{\rm osc} \sim \Lambda_{\rm QCD} \simeq 200$ MeV and make use of Eq. (57) with $C_{\rm g} \sim 1$. We find

$$\Omega_{\rm a} \equiv \frac{\rho_{\rm a,0}}{\rho_{\rm c}} \simeq \left(\frac{10^{-6}\ {\rm eV}}{m_{\rm a}} \right) \bar{\theta}_{\rm i}^2 \ . \tag{69}$$

The natural assumption about the initial phase is $\bar{\theta}_{\rm i} \sim \pi/2$. Hence, an axion of mass 10^{-5}–10^{-6} eV is a good dark matter candidate. Note that an axion of lower mass $m_{\rm a} < 10^{-6}$ eV may also serve as a dark matter particle, if for some reason the initial phase $\bar{\theta}_{\rm i}$ is much smaller than $\pi/2$. This is *cold* dark matter: the oscillating field corresponds to axions at rest.

A more precise estimate is obtained by taking into account the fact that that the axion mass smoothly depends on temperature:

$$\Omega_{\rm a} \simeq 0.2 \cdot \bar{\theta}_{\rm i}^2 \cdot \left(\frac{4 \times 10^{-6}\ {\rm eV}}{m_{\rm a}} \right)^{1.2} \ .$$

We see that our crude estimate (69) is fairly accurate. Interestingly, the string mechanism of the axion production leads to the same parametric dependence of $\Omega_{\rm a}$ on the axion mass.

Search for dark matter axions with mass $m_{\rm a} \sim 10^{-5}$–10^{-6} eV is difficult, but not impossible. One way is to search for axion–photon conversion in a resonator cavity filled with a strong magnetic field. Indeed, in the background magnetic field the axion–photon interaction (second term in Eq. (64)) leads to the conversion a $\to \gamma$, and the axions of mass 10^{-5}–10^{-6} eV are converted to photons of frequency $m/(2\pi) = 2$–0.2 GHz (radio waves). Bounds on the dark matter axions are shown in Fig. 11.

4.5 Warm dark matter: sterile neutrinos and light gravitinos

As we discussed in Section 4.1, there are arguments, albeit not particularly strong, that favour warm, rather than cold, dark matter. If WDM particles are thermal relics, i.e., if they were in kinetic equilibrium at some epoch in the early Universe, then their mass should be in the range 3–10 keV. Reasonably well motivated particles of this mass are sterile neutrinos and gravitinos.

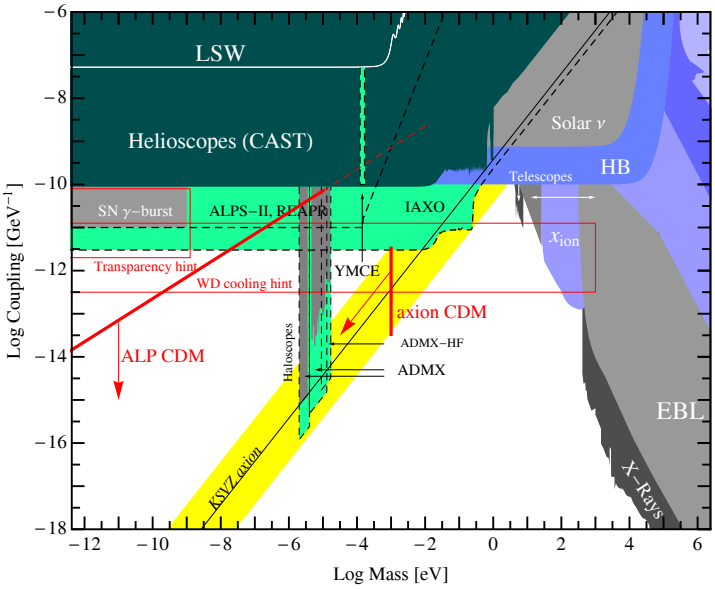

Fig. 11: Bounds on dark matter axions: axion–photon coupling versus axion mass [55]. Inclined straight line labelled 'KSVZ axion' is the prediction of the KSVZ model, shaded region along this line is the range of predictions of other axion models. Region below the line labelled ALP CDM is the range of predictions of other reasonably motivated models with axion-like particles as dark matter candidates. Dashed lines show the sensitivities of future experiments.

4.5.1 Sterile neutrinos

Sterile neutrinos are most probably required for giving masses to ordinary, 'active' neutrinos. The masses of sterile neutrinos cannot be predicted theoretically. Although sterile neutrinos of WDM mass $m_{\nu_{\rm s}} = 3$–10 keV are not particularly plausible from the particle-physics prospective, they are not pathological either. In the simplest models the creation of sterile neutrino states $|\nu_{\rm s}\rangle$ in the early Universe occurs due to their mixing with active neutrinos $|\nu_\alpha\rangle$, $\alpha = {\rm e}, \mu, \tau$. In the approximation of mixing between two states only, we have

$$|\nu_\alpha\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle \ , \qquad |\nu_{\rm s}\rangle = -\sin\theta|\nu_1\rangle + \cos\theta|\nu_2\rangle \ ,$$

where $|\nu_\alpha\rangle$ and $|\nu_{\rm s}\rangle$ are active and sterile neutrino states, $|\nu_1\rangle$ and $|\nu_2\rangle$ are mass eigenstates of masses m_1 and m_2, where we order $m_1 < m_2$, and θ is the vacuum mixing angle between sterile and active neutrinos. This mixing should be weak, $\theta \ll 1$, otherwise sterile neutrinos would decay too rapidly, see below. The heavy state is mostly sterile neutrinos $|\nu_2\rangle \approx |\nu_{\rm s}\rangle$, and $m_2 \equiv m_{\rm s}$ is the sterile neutrino mass.

The calculation of sterile neutrino abundance is fairly complicated, and we do not reproduce it here. If there is no sizeable lepton asymmetry in the Universe, the sterile neutrino production is most efficient at temperature around

$$T_* \sim \left(\frac{m_{\rm s}}{5G_{\rm F}}\right)^{1/3} \simeq 200\ {\rm MeV} \cdot \left(\frac{m_{\rm s}}{1\ {\rm keV}}\right)^{1/3} .$$

The resulting number density of sterile neutrinos is estimated as

$$\frac{n_{\nu_{\rm s}}}{n_{\nu_\alpha}} \sim T_*^3 M_{\rm Pl}^* G_{\rm F}^2 \cdot \sin^2 2\theta \sim 10^{-2} \cdot \left(\frac{m_{\rm s}}{1\ {\rm keV}}\right) \cdot \left(\frac{\sin 2\theta}{10^{-4}}\right)^2 . \qquad (70)$$

The number density of relic active neutrinos today is about 110 cm^{-3}, so we find from Eq. (70) the estimate for the present contribution of sterile neutrinos into energy density,

$$\Omega_{\nu_s} \simeq 0.2 \cdot \left(\frac{\sin 2\theta}{10^{-4}}\right)^2 \cdot \left(\frac{m_\nu}{1\ \text{keV}}\right)^2 \ . \tag{71}$$

Thus, a sterile neutrino of mass $m_\nu \gtrsim 1$ keV and small mixing angle $\theta_\alpha \lesssim 10^{-4}$ would serve as a dark matter candidate. However, this range of masses and mixing angles is ruled out. The point is that due to its mixing with an active neutrino, a sterile neutrino can decay into an active neutrino and a photon,

$$\nu_s \to \nu_\alpha + \gamma \ .$$

The sterile neutrino decay width is proportional to $\sin^2 2\theta$. If sterile neutrinos are dark matter particles, their decays would produce a narrow line in X-ray flux from the cosmos (orbiting velocity of dark matter particles in our Galaxy is small, $v \sim 10^{-3}$, hence the photons produced in their two-body decays are nearly monochromatic). Such a line has not been observed, and there exist quite strong limits. These limits, translated into limits on $\sin^2 2\theta$ as a function of sterile neutrino mass, are shown in Fig. 12; they rule out the range of masses giving the right mass density of dark matter, Eq. (71). Recall that the mass of a sterile neutrino should exceed 3 keV (in fact, a more precise limit is $m_s > 5.7$ keV [56]).

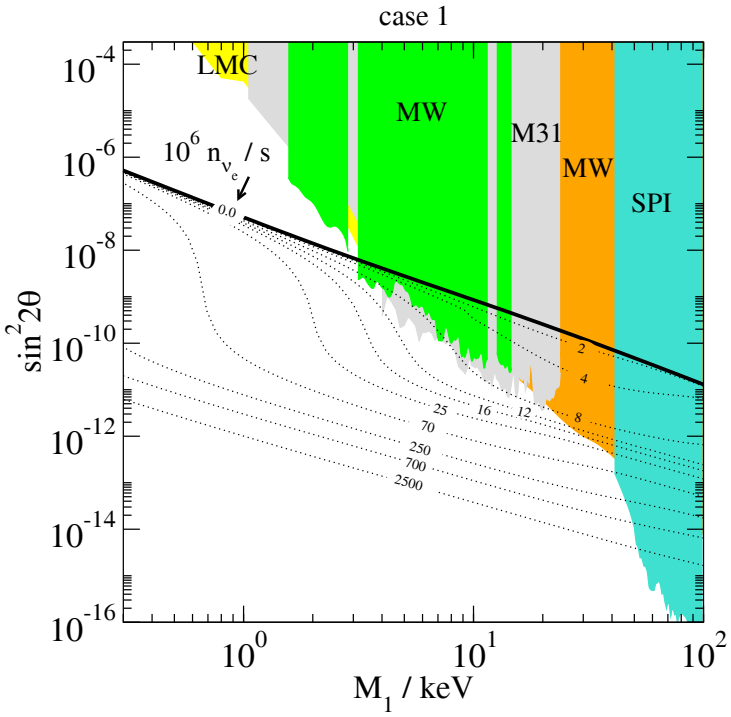

Fig. 12: Limits on sterile neutrino parameters (mass M_1, mixing angle θ) obtained from X-ray telescopes. Solid line corresponds to sterile neutrino dark matter produced in non-resonant oscillations, Eq. (71). Dashed lines show the case of resonant oscillations at non-zero lepton asymmetry; numbers in unit of 10^{-6} show the values of lepton asymmetry (lepton-to-photon ratio η_L) [57].

A (rather baroque) way out [58] is to assume that there is a fairly large lepton asymmetry in the Universe. Then the oscillations of active neutrinos into sterile neutrinos may be enhanced due to the Mikheyev-Smirnov-Wolfenstein (MSW) effect, as at some temperature they occur in the Mikheev–Smirnov resonance regime. In that case the right abundance of sterile neutrinos is obtained at smaller θ, and may be consistent with X-ray bounds. This is also shown in Fig. 12.

4.5.2 Light gravitino

A gravitino—a superpartner of a graviton—is necessarily present in supersymmetric (SUSY) theories. It acquires mass as a result of SUSY breaking (super-Higgs mechanism). The gravitino mass is of order

$$m_{3/2} \simeq \frac{F}{M_{\rm Pl}} \ ,$$

where $\sqrt{F}$ is the supersymmetry breaking scale. Hence, gravitino masses are in the right WDM ballpark for rather low supersymmetry breaking scales, $\sqrt{F} \sim 10^6$–10^7 GeV. This is the case, e.g., in the gauge-mediation scenario. With so low mass, a gravitino is the lightest supersymmetric particle (LSP), so it is stable in many supersymmetric extensions of the Standard Model. From this viewpoint gravitinos can indeed serve as dark matter particles. For what follows, important parameters are the widths of decays of other superpartners into gravitinos and the Standard Model particles. These are of order

$$\Gamma_{\tilde{\rm S}} \simeq \frac{M_{\tilde{\rm S}}^5}{F^2} \simeq \frac{M_{\tilde{\rm S}}^5}{m_{3/2}^2 M_{\rm Pl}^2} \ , \tag{72}$$

where $M_{\tilde{\rm S}}$ is the mass of the superpartner.

One mechanism of the gravitino production in the early Universe is decays of other superpartners. A gravitino interacts with everything else so weakly that once produced, it moves freely, without interacting with cosmic plasma. At production, gravitinos are relativistic and hence they are indeed *warm* dark matter candidates. Let us assume that production in decays is the dominant mechanism and consider under what circumstances the present mass density of gravitinos coincides with that of dark matter.

The rate of gravitino production in decays of superpartners of the type $\tilde{\rm S}$ in the early Universe is

$$\frac{{\rm d}(n_{3/2}/s)}{{\rm d}t} = \frac{n_{\tilde{\rm S}}}{s} \Gamma_{\tilde{\rm S}} \ ,$$

where $n_{3/2}$ and $n_{\tilde{\rm S}}$ are number densities of gravitinos and superpartners, respectively, and s is the entropy density. For superpartners in thermal equilibrium, one has $n_{\tilde{\rm S}}/s = {\rm const} \sim g_*^{-1}$ for $T \gtrsim M_{\tilde{\rm S}}$, and $n_{\tilde{\rm S}}/s \propto \exp(-M_{\tilde{\rm S}}/T)$ at $T \ll M_{\tilde{\rm S}}$. Hence, the production is most efficient at $T \sim M_{\tilde{\rm S}}$, when the number density of superpartners is still large, while the Universe expands most slowly. The density of gravitinos produced in decays of the $\tilde{\rm S}$ is thus given by

$$\frac{n_{3/2}}{s} \simeq \frac{\Gamma_{\tilde{\rm S}}}{g_*} H^{-1}(T \sim M_{\tilde{\rm S}}) \simeq \frac{1}{g_*} \cdot \frac{M_{\tilde{\rm S}}^5}{m_{3/2}^2 M_{\rm Pl}^2} \cdot \frac{M_{\rm Pl}^*}{M_{\tilde{\rm S}}^2} \ .$$

This gives the mass-to-entropy ratio today,

$$\frac{m_{3/2} n_{3/2}}{s} \simeq \sum_{\tilde{\rm S}} \frac{M_{\tilde{\rm S}}^3}{g_*^{3/2} M_{\rm Pl} m_{3/2}} \ , \tag{73}$$

where the sum runs over all superpartner species *which have ever been relativistic in thermal equilibrium*. The correct value (26) is obtained for gravitino masses in the range (29) at

$$M_{\tilde{\rm S}} = 100\text{–}300 \ {\rm GeV} \ . \tag{74}$$

Thus, the scenario with a gravitino as a warm dark matter particle requires light superpartners [59], which are to be discovered at the LHC.

A few comments are in order. First, decays of superpartners is not the only mechanism of gravitino production: gravitinos may also be produced in scattering of superpartners [60]. To avoid overproduction of gravitinos in the latter processes, one has to assume that the maximum temperature in the Universe

(e.g., reached after the post-inflationary reheating stage) is quite low, $T_{\rm max} \sim$ 1–10 TeV. This is not a particularly plausible assumption, but it is consistent with everything else in cosmology and can indeed be realized in some models of inflation. Second, existing constraints on masses of strongly interacting superpartners (gluinos and squarks) suggest that their masses exceed (74). Hence, these particles should not contribute to the sum in (73), otherwise WDM gravitinos would be overproduced. This is possible if masses of squarks and gluinos are larger than $T_{\rm max}$, so that they were never abundant in the early Universe. Third, a gravitino produced in decays of superpartners is *not* a thermal relic, as it was never in thermal equilibrium with the rest of the cosmic plasma. Nevertheless, since gravitinos are produced at $T \sim M_{\tilde{\rm S}}$ and at that time have energy $E \sim M_{\tilde{\rm S}} \sim T$, our estimate (28) does apply.

Question. Let $\tilde{\rm S}$ be the next-to-lightest superpartner which decays into a gravitino of mass $m_{3/2} = 5$ keV and a Standard Model particle. Let $\tilde{\rm S}$ be produced at the LHC at subrelativistic velocity. How far is the decay vertex of $\tilde{\rm S}$ displaced from the proton collision point? Give numerical estimates for $M_{\tilde{\rm S}} =$ 100 GeV and $M_{\tilde{\rm S}} = 1$ TeV.

4.6 Discussion

If dark matter particles are indeed WIMPs, and the relevant energy scale is of order 1 TeV, then the hot Big Bang theory will be probed experimentally up to a temperature of (a few) $\cdot$ (10–100) GeV and down to an age of 10^{-9}–10^{-11} s in the relatively near future (compare to 1 MeV and 1 s accessible today through BBN). With microscopic physics to be known from collider experiments, the WIMP density will be reliably calculated and checked against the data from observational cosmology. Thus, the WIMP scenario offers a window to a very early stage of the evolution of the Universe.

Search for dark matter axions and signals from light sterile neutrinos makes use of completely different methods. Yet there is a good chance for discovery if either of these particles make dark matter.

If dark matter particles are gravitinos, the prospect of probing quantitatively such an early stage of the cosmological evolution is not so bright: it would be very hard, if at all possible, to get an experimental handle on the gravitino mass; furthermore, the present gravitino mass density depends on an unknown reheat temperature $T_{\rm max}$. On the other hand, if this scenario is realized in nature, then the whole picture of the early Universe will be quite different from our best guess on the early cosmology. Indeed, the gravitino scenario requires a low reheat temperature, which in turn calls for a rather exotic mechanism of inflation.

The mechanisms discussed here are by no means the only ones capable of producing dark matter, and the particles we discussed are by no means the only candidates for dark matter particles. Other dark matter candidates include axinos, Q-balls, very heavy relics produced towards the end of inflation (wimpzillas) etc. Hence, even though there are grounds to hope that the dark matter problem will be solved soon, there is no guarantee at all.

5 Baryon asymmetry of the Universe

As we discussed in Section 2.4, the baryon asymmetry of the Universe is characterized by the baryon-to-entropy ratio, which at high temperatures is defined as follows:

$$\Delta_{\rm B} = \frac{n_{\rm B} - n_{\bar{\rm B}}}{s} \,,$$

where $n_{\bar{\rm B}}$ is the number density of antibaryons and s is the entropy density. If the baryon number is conserved and the Universe expands adiabatically (which is the case at least after the electroweak epoch, $T \lesssim 100$ GeV), $\Delta_{\rm B}$ is time-independent and equal to its present value $\Delta_{\rm B} \approx 0.8 \times 10^{-10}$, see Eq. (25). At early times, at temperatures well above 100 MeV, cosmic plasma contained many quark–antiquark

pairs, whose number density was of the order of the entropy density,

$$n_{\rm q} + n_{\bar{\rm q}} \sim s \ ,$$

while the baryon number density was related to densities of quarks and antiquarks as follows (baryon number of a quark equals $1/3$):

$$n_{\rm B} = \frac{1}{3}(n_{\rm q} - n_{\bar{\rm q}}) \ .$$

Hence, in terms of quantities characterizing the very early epoch, the baryon asymmetry may be expressed as

$$\Delta_{\rm B} \sim \frac{n_{\rm q} - n_{\bar{\rm q}}}{n_{\rm q} + n_{\bar{\rm q}}} \ .$$

We see that there was one extra quark per about 10 billion quark–antiquark pairs. It is this tiny excess that is responsible for the entire baryonic matter in the present Universe: as the Universe expanded and cooled down, antiquarks annihilated with quarks, and only the excessive quarks remained and formed baryons.

There is no logical contradiction to suppose that the tiny excess of quarks over antiquarks was built in as an initial condition. This is not at all satisfactory for a physicist, however. Furthermore, the inflationary scenario does not provide such an initial condition for the hot Big Bang epoch; rather, inflationary theory predicts that the Universe was baryon-symmetric just after inflation. Hence, one would like to explain the baryon asymmetry dynamically [61, 62], i.e., find the mechanism of its generation in the early Universe.

5.1 Sakharov conditions

The baryon asymmetry may be generated from an initially baryon-symmetric state only if three necessary conditions, dubbed Sakharov conditions, are satisfied. These are:

1. baryon number non-conservation;
2. C- and CP-violation;
3. deviation from thermal equilibrium.

All three conditions are easily understood. (1) If baryon number were conserved, and initial net baryon number in the Universe was zero, the Universe today would still be symmetric. (2) If C or CP were conserved, then the rate of reactions with particles would be the same as the rate of reactions with antiparticles, and no asymmetry would be generated. (3) Thermal equilibrium means that the system is stationary (no time dependence at all). Hence, if the initial baryon number is zero, it is zero forever, unless there are deviations from thermal equilibrium. Furthermore, if there are processes that violate baryon number, and the system approaches thermal equilibrium, then the baryon number tends to be washed out rather than generated.

At the epoch of the baryon-asymmetry generation, all three Sakharov conditions have to be met simultaneously. There is a qualification, however. These conditions would be literally correct if there were no other relevant quantum numbers that characterize the cosmic medium. In reality, however, lepton numbers also play a role. As we will see shortly, baryon and lepton numbers are rapidly violated by anomalous electroweak processes at temperatures above, roughly, 100 GeV. What is conserved in the Standard Model is the combination $B - L$, where L is the total lepton number. So, there are two options. One is to generate the baryon asymmetry at or below the electroweak epoch, $T \lesssim 100$ GeV, and make sure that the electroweak processes do not wash out the baryon asymmetry after its generation. This leads to the idea of electroweak baryogenesis (another possibility is Affleck–Dine baryogenesis [63]). Another is to generate $B - L$ asymmetry before the electroweak epoch, i.e., at $T \gg 100$ GeV: if the Universe

is $B - L$ asymmetric above 100 GeV, the electroweak physics reprocesses $B - L$ partially into baryon number and partially into lepton number, so that in thermal equilibrium with conserved $B - L$ one has

$$B = C \cdot (B - L) \,, \qquad L = (C - 1) \cdot (B - L) \,,$$

where C is a constant of order 1 ($C = 28/79$ in the Standard Model at $T \gtrsim 100$ GeV). In the second scenario, the first Sakharov condition applies to $B - L$ rather than baryon number itself.

Let us point out two most common mechanisms of baryon number non-conservation. One emerges in grand unified theories and is due to the exchange of supermassive particles. It is similar, say, to the mechanism of charm non-conservation in weak interactions, which occurs via the exchange of heavy W bosons. The scale of these new, baryon number violating interactions is the grand unification scale, presumably of order $M_{\rm GUT} \simeq 10^{16}$ GeV. It is rather unlikely, however, that the baryon asymmetry was generated due to this mechanism: the relevant temperature would be of order $M_{\rm GUT}$, while such a high reheat temperature after inflation is difficult to obtain.

Another mechanism is non-perturbative [39] and is related to the triangle anomaly in the baryonic current (a keyword here is 'sphaleron' [64,65]). It exists already in the Standard Model and, possibly with mild modifications, operates in all its extensions. The two main features of this mechanism, as applied to the early Universe, is that it is effective over a wide range of temperatures, 100 GeV $< T < 10^{11}$ GeV, and, as we pointed out above, that it conserves $B - L$.

5.2 Electroweak baryon number non-conservation

Let us pause here to discuss the physics behind electroweak baryon and lepton number non-conservation in a little more detail, though still at a qualitative level. A detailed analysis can be found in the book [66] and in references therein.

Let us consider the baryonic current,

$$B^\mu = \frac{1}{3} \cdot \sum_i \bar{q}_i \gamma^\mu q_i \,,$$

where the sum runs over quark flavours. Naively, it is conserved, but at the quantum level its divergence is non-zero because of the triangle anomaly (a similar effect goes under the name of the axial anomaly in the context of quantum electrodynamics (QED) and QCD),

$$\partial_\mu B^\mu = \frac{1}{3} \cdot 3_{\rm colours} \cdot 3_{\rm generations} \cdot \frac{g^2}{32\pi^2} \epsilon^{\mu\nu\lambda\rho} F^{\rm a}_{\mu\nu} F^{\rm a}_{\lambda\rho} \,,$$

where $F^{\rm a}_{\mu\nu}$ and g are the field strength of the SU(2)$_{\rm W}$ gauge field and the SU(2)$_{\rm W}$ gauge coupling, respectively. Likewise, each leptonic current ($\alpha = {\rm e}, \mu, \tau$) is anomalous in the Standard Model (we disregard here neutrino masses and mixings, which violate lepton numbers too),

$$\partial_\mu \mathcal{L}^\mu_\alpha = \frac{g^2}{32\pi^2} \cdot \epsilon^{\mu\nu\lambda\rho} F^{\rm a}_{\mu\nu} F^{\rm a}_{\lambda\rho} \,. \tag{75}$$

A non-trivial fact is that there exist large field fluctuations, $F^{\rm a}_{\mu\nu}(\mathbf{x}, t) \propto g^{-1}$, such that

$$Q \equiv \int {\rm d}^3x {\rm d}t \, \frac{g^2}{32\pi^2} \cdot \epsilon^{\mu\nu\lambda\rho} F^{\rm a}_{\mu\nu} F^{\rm a}_{\lambda\rho} \neq 0 \,. \tag{76}$$

Furthermore, for any physically relevant fluctuation the value of Q is an integer ('physically relevant' means that the gauge field strength vanishes at infinity in space–time). In four space–time dimensions such fluctuations exist only in *non-Abelian* gauge theories.

Suppose now that a fluctuation with non-vanishing Q has occurred. Then the baryon numbers at the end and beginning of the process are different,

$$B_{\rm fin} - B_{\rm in} = \int {\rm d}^3x{\rm d}t\, \partial_\mu B^\mu = 3Q \ . \tag{77}$$

Likewise,

$$\mathcal{L}\alpha,\ {\rm fin} - \mathcal{L}_{\alpha,\ {\rm in}} = Q \ . \tag{78}$$

This explains the selection rule mentioned above: B is violated, $B - L$ is not.

At zero temperature, the field fluctuations that induce baryon and lepton number violation are vacuum fluctuations, called instantons [67]. Since these are *large* field fluctuations, their probability is exponentially suppressed. The suppression factor in the Standard Model is

$$\mathrm{e}^{-\frac{16\pi^2}{g^2}} \sim 10^{-165} \ .$$

Therefore, the rate of baryon number violating processes at zero temperature is suppressed by this factor, making these processes totally negligible. On the other hand, at high temperatures there are large *thermal* fluctuations ('sphalerons') whose rate is not necessarily small. And, indeed, B-violation in the early Universe is rapid as compared to the cosmological expansion at sufficiently high temperatures, provided that (see Ref. [11] for details)

$$\langle\phi\rangle_T < T \ , \tag{79}$$

where $\langle\phi\rangle_T$ is the Englert–Brout–Higgs expectation value at temperature T.

One may wonder how baryon number is not conserved in the absence of explicit baryon number violating terms in the Lagrangian of the Standard Model. To understand what is going on, let us consider a massless *left-handed* fermion field in the background of the SU(2) gauge field $\mathbf{A}(\mathbf{x}, t)$, which depends on space–time coordinates in a non-trivial way. As a technicality, we set the temporal component of the gauge field equal to zero, $A_0 = 0$, by the choice of gauge. One way to understand the behaviour of the fermion field in the gauge field background is to study the system of eigenvalues of the Dirac Hamiltonian $\{\omega(t)\}$. The Hamiltonian is defined in the standard way

$$H_{\rm Dirac}(t) = {\rm i}\alpha^i \left(\partial_i - {\rm i}gA_i(\mathbf{x}, t)\right) \frac{1-\gamma_5}{2} \ ,$$

where $\alpha^i = \gamma^0\gamma^i$, so that the Dirac equation has the Schrödinger form,

$${\rm i}\frac{\partial\psi}{\partial t} = H_{\rm Dirac}\psi \ .$$

So, let us discuss the eigenvalues $\omega_n(t)$ of the operator $H_{\rm Dirac}(t)$, treating t as a parameter. These eigenvalues are found from

$$H_{\rm Dirac}(t)\psi_n = \omega_n(t)\psi_n \ .$$

At $\mathbf{A} = 0$, the system of levels is shown schematically in Fig. 13. Importantly, there are both positive- and negative-energy levels. According to Dirac, the lowest-energy state (Dirac vacuum) has all negative-energy levels occupied, and all positive-energy levels empty. Occupied positive-energy levels (three of them in Fig. 13) correspond to real fermions, while empty negative-energy levels describe antifermions (one in Fig. 13). Fermion–antifermion annihilation in this picture is a jump of a fermion from a positive-energy level to an unoccupied negative-energy level. As a side remark, this original Dirac picture is, in fact, equivalent to the more conventional (by now) procedure of the quantization of the fermion field, which does not make use of the notion of negative-energy levels. The discussion that follows can be translated into the conventional language; however, the original Dirac picture turns out to be a lot more

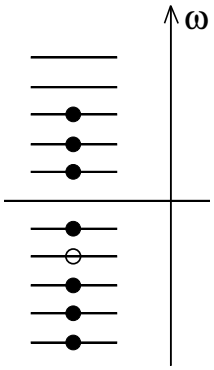

Fig. 13: Fermion energy levels at zero background gauge field

transparent in our context. This is a nice example of the complementarity of various approaches in quantum field theory.

Let us proceed with the discussion of the fermion energy levels in gauge field backgrounds. In weak background fields, the energy levels depend on time ('move'), but nothing dramatic happens. For adiabatically varying background fields, the fermions merely sit on their levels, while fast-changing fields generically give rise to jumps from, say, negative- to positive-energy levels, that is, creation of fermion–antifermion pairs. Needless to say, fermion number $(N_{\rm f} - N_{\bar{\rm f}})$ is conserved.

The situation is entirely different for the background fields with non-zero Q. The levels of left-handed fermions move as shown in the left-hand panel of Fig. 14. Some levels necessarily cross zero, and

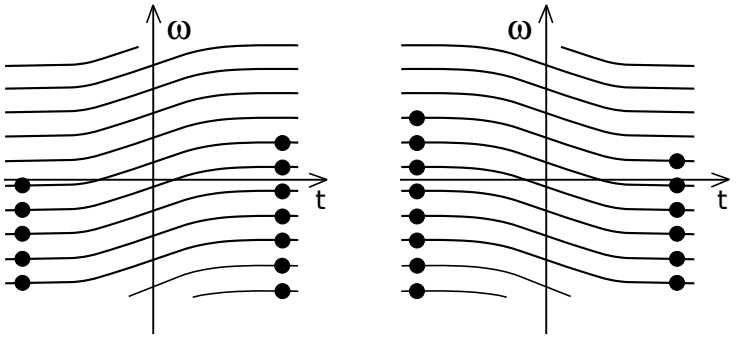

Fig. 14: Motion of fermion levels in background gauge fields with non-vanishing Q (shown is the case $Q = 2$). Left-hand panel: left-handed fermions. Right-hand panel: right-handed fermions.

the net number of levels crossing zero from below equals Q. This means that the number of left-handed fermions is not conserved: for an adiabatically varying gauge field $\mathbf{A}(\mathbf{x}, t)$, the motion of levels shown in the left-hand panel of Fig. 14 corresponds to the case in which the initial state of the fermionic system is vacuum (no real fermions or antifermions) whereas the final state contains Q real fermions (two in the particular case shown). If the evolution of the gauge field is not adiabatic, the result for the fermion number non-conservation is the same: there may be jumps from negative-energy levels to positive-energy levels or vice versa. These correspond to creation or annihilation of fermion–antifermion pairs, but the net change of the fermion number (number of fermions minus number of antifermions) remains equal to Q. Importantly, the initial and final field configurations of the gauge field may be trivial, $\mathbf{A} = 0$ (up to gauge transformation), so that fermion number non-conservation may occur due to a fluctuation that begins and ends in the gauge field vacuum. These are precisely instanton-like vacuum fluctuations. At finite temperatures, processes of this type occur due to thermal fluctuations, i.e., sphalerons.

If the same gauge field interacts also with right-handed fermions, the motion of the levels of the latter is opposite to that of left-handed fermions. This is shown in the right-hand panel of Fig. 14. The change in the number of right-handed fermions is equal to $-Q$. So, if the gauge interaction is vector-like, the total fermion number $(N_{\rm left} + N_{\rm right})$ is conserved, while chirality $(N_{\rm left} - N_{\rm right})$ is violated even for massless fermions. This explains why there is no baryon number violation in QCD. The above discussion implies, instead, that there is non-perturbative violation of chirality in QCD in the limit of massless quarks. The latter phenomenon has non-trivial consequences, which are indeed confirmed by phenomenology. In this sense anomalous non-conservation of fermion quantum numbers is an experimentally established fact.

In electroweak theory, right-handed fermions do not interact with the $\mathrm{SU}(2)_{\rm W}$ gauge field, while left-handed fermions do. Therefore, fermion number is not conserved (the anomalous relations (75) and (76) suggest that this result is valid also in the presence of the Standard Model Yukawa couplings of quarks and leptons; this is indeed the case). Since fermions of each $\mathrm{SU}(2)_{\rm W}$ doublet interact with the $\mathrm{SU}(2)_{\rm W}$ gauge bosons in one and the same way, they are equally created in a process involving a gauge field fluctuation with non-zero Q. This again leads to the relations (77) and (78), i.e., to the selection rules $\Delta B = \Delta L$ and $\Delta L_{\rm e} = \Delta L_\mu = \Delta L_\tau$.

5.3 Electroweak baryogenesis

It is tempting to make use of the electroweak mechanism of baryon number non-conservation for explaining the baryon asymmetry of the Universe. This scenario is known as electroweak baryogenesis. It meets two problems, however. One is that CP-violation in the Standard Model is too weak: the CKM mechanism alone is insufficient to generate a realistic value of the baryon asymmetry. Hence, one needs extra sources of CP-violation. Another problem has to do with departure from thermal equilibrium that is necessary for the generation of the baryon asymmetry. At temperatures well above 100 GeV, electroweak symmetry is restored, the expectation value of the Englert–Brout–Higgs field ϕ is zero, the relation (79) is valid and the baryon number non-conservation is rapid as compared to the cosmological expansion. At temperatures of order 100 GeV, the relation (79) may be violated, but the Universe expands very slowly: the cosmological time-scale at these temperatures is

$$H^{-1} = \frac{M^*_{\rm Pl}}{T^2} \sim 10^{-10}\ {\rm s}\ , \tag{80}$$

which is very large by the electroweak physics standards. The only way in which a strong departure from thermal equilibrium at these temperatures may occur appears to be the first-order phase transition.

The property that at temperatures well above 100 GeV the expectation value of the Englert–Brout–Higgs field is zero, while it is non-zero in vacuo, suggests that there may be a phase transition from the phase with $\langle\phi\rangle = 0$ to the phase with $\langle\phi\rangle \neq 0$. The situation is pretty subtle here, as ϕ is not gauge invariant and hence cannot serve as an order parameter, so the notion of phases with $\langle\phi\rangle = 0$ and $\langle\phi\rangle \neq 0$ is vague. In fact, neither electroweak theory nor most of its extensions have a gauge-invariant order parameter, so there is no real distinction between these 'phases'. This situation is similar to that in a liquid–vapour system, which does not have an order parameter and may or may not experience a vapour–liquid phase transition as temperature decreases, depending on other, external parameters characterizing this system, e.g., pressure. In the Standard Model the role of such an 'external' parameter is played by the Englert–Brout–Higgs self-coupling λ or, in other words, the Higgs boson mass.

Continuing to use somewhat sloppy terminology, we recall that in thermal equilibrium any system is at the global minimum of its *free energy*. To figure out the expectation value of ϕ at a given temperature, one introduces the temperature-dependent effective potential $V_{\rm eff}(\phi; T)$, which is equal to the free energy density in the system where the average field is pinpointed to a prescribed value ϕ, but otherwise there is thermal equilibrium. Then the global minimum of $V_{\rm eff}$ at a given temperature is at the equilibrium value of ϕ, while local minima correspond to metastable states.

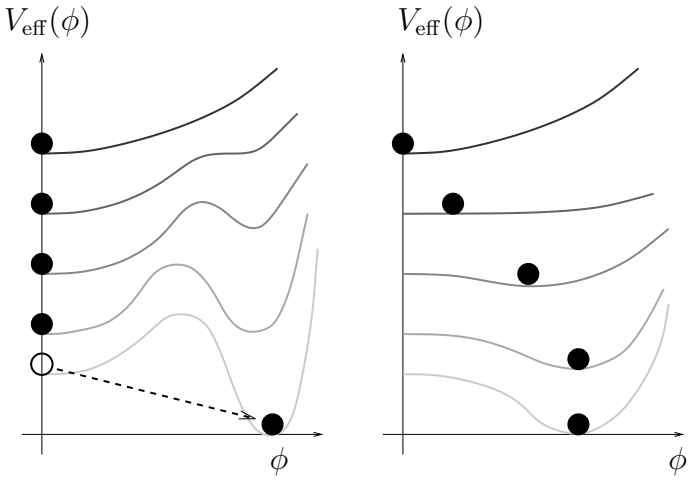

Fig. 15: Effective potential as function of ϕ at different temperatures. Left: first-order phase transition. Right: second-order phase transition. Upper curves correspond to higher temperatures. Black blobs show the expectation value of ϕ in thermal equilibrium. The arrow in the left-hand panel illustrates the transition from the metastable, supercooled state to the ground state.

The interesting case for us is the first-order phase transition. In this case, the system evolves as follows. At high temperatures, there exists one minimum of $V_{\rm eff}$ at $\phi = 0$, and the expectation value of the Englert–Brout–Higgs field is zero. As the temperature decreases, another minimum appears at finite ϕ, and then becomes lower than the minimum at $\phi = 0$; see left-hand panel of Fig. 15. However, the minima with $\phi = 0$ and $\phi \neq 0$ are separated by a barrier of $V_{\rm eff}$, the probability of the transition from the phase $\phi = 0$ to the phase $\phi \neq 0$ is very small for some time and the system gets overcooled. The transition occurs when the temperature becomes sufficiently low and the transition probability sufficiently high. This is to be contrasted to the case, e.g., of the second-order phase transition, right-hand panel of Fig. 15. In the latter case, the field slowly evolves, as the temperature decreases, from zero to non-zero vacuum value, and the system remains very close to thermal equilibrium at all times.

During the first-order phase transition, the field cannot jump from $\phi = 0$ to $\phi \neq 0$ homogeneously throughout the whole space: intermediate homogeneous configurations have free energies proportional to the volume of the system (recall that $V_{\rm eff}$ is free energy *density*), i.e., infinite. Instead, the transition occurs just like the first-order vapour–liquid transition, through boiling. Thermal fluctuations spontaneously create bubbles of the new phase inside the old phase. These bubbles then grow, their walls eventually collide and the new phase finally occupies the entire space. The Universe boils. In the cosmological context, this process happens when the bubble nucleation rate per Hubble time per Hubble volume is roughly of order 1, i.e., when a few bubbles are created in Hubble volume in Hubble time. The velocity of the bubble wall in the relativistic cosmic plasma is roughly of the order of the speed of light (in fact, it is somewhat smaller, from 0.1 to 0.01), simply because there are no relevant dimensionless parameters characterizing the system. Hence, the bubbles grow large before their walls collide: their size at collision is roughly of the order of the Hubble size (in fact, one or two orders of magnitude smaller). While the bubble is microscopic at nucleation—its size is determined by the electroweak scale and is roughly of order $(100\ \text{GeV})^{-1} \sim 10^{-16}$ cm—its size at collision of walls is macroscopic, $R \sim 10^{-2}$–10^{-3} cm, as follows from (80). Clearly, boiling is a highly non-equilibrium process, and one may hope that the baryon asymmetry may be generated at that time. And, indeed, there exist mechanisms of the generation of the baryon asymmetry, which have to do with interactions of quarks and leptons with moving bubble walls. The value of the resulting baryon asymmetry may well be of order 10^{-10}, as required by observations, provided that there is enough CP-violation in the theory.

A necessary condition for the electroweak generation of the baryon asymmetry is that the inequality (79) must be violated *just after* the phase transition. Indeed, in the opposite case the electroweak

baryon number violating processes are fast after the transition, and the baryon asymmetry, generated during the transition, is washed out afterwards. Hence, the phase transition must be of strong enough first order. This is *not* the case in the Standard Model. To see why this is so, and to get an idea of in which extensions of the Standard Model the phase transition may be of strong enough first order, let us consider the effective potential in some detail. At zero temperature, the Englert–Brout–Higgs potential has the standard form,

$$V(\phi) = -\frac{m^2}{2}|\phi|^2 + \frac{\lambda}{4}|\phi|^4 \ .$$

Here

$$|\phi| \equiv \left(\phi^\dagger \phi\right)^{1/2} \tag{81}$$

is the length of the Englert–Brout–Higgs doublet ϕ, $m^2 = \lambda v^2$ and $v = 246$ GeV is the Englert–Brout–Higgs expectation value in vacuo. The Higgs boson mass is related to the latter as follows:

$$m_{\rm H} = \sqrt{2\lambda} v \ . \tag{82}$$

Now, to the leading order of perturbation theory, the finite-temperature effects modify the effective potential into

$$V_{\rm eff}(\phi, T) = \frac{\alpha(T)}{2}|\phi|^2 - \frac{\beta}{3}T|\phi|^3 + \frac{\lambda}{4}|\phi|^4 \ . \tag{83}$$

Here $\alpha(T) = -m^2 + \hat{g}^2 T^2$, where $\hat{g}^2$ is a positive linear combination of squares of coupling constants of all fields to the Englert–Brout–Higgs field (in the Standard Model, a linear combination of g^2, g'^2 and y_i^2, where g and g' are $\mathrm{SU}(2)_{\rm W}$ and $\mathrm{U}(1)_{\rm Y}$ gauge couplings and y_i are Yukawa couplings). The phase transition occurs roughly when $\alpha(T) = 0$. An important parameter β is a positive linear combination of cubes of coupling constants of all *bosonic* fields to the Englert–Brout–Higgs field. In the Standard Model, β is a linear combination of g^3 and g'^3, i.e., a linear combination of $M_{\rm W}^3/v^3$ and $M_{\rm Z}^3/v^3$,

$$\beta = \frac{1}{2\pi}\frac{2M_{\rm W}^3 + M_{\rm Z}^3}{v^3} \ . \tag{84}$$

The cubic term in (83) is rather peculiar: in view of (81), it is not analytic in the original Englert–Brout–Higgs field ϕ. Yet this term is crucial for the first-order phase transition: for $\beta = 0$ the phase transition would be of the second order.

Question. Show that the phase transition is second order for $\beta = 0$.

The origin of the non-analytic cubic term can be traced back to the enhancement of the Bose–Einstein thermal distribution at low momenta, $p, m \ll T$,

$$f_{\rm Bose}(p) = \frac{1}{{\rm e}^{\frac{\sqrt{p^2+m_a^2}}{T}} - 1} \simeq \frac{T}{\sqrt{p^2 + m_a^2}} \ ,$$

where $m \simeq g_a|\phi|$ is the mass of the boson a that is generated due to the non-vanishing Englert–Brout–Higgs field, and g_a is the coupling constant of the field a to the Englert–Brout–Higgs field. Clearly, at $p \ll g|\phi|$ the distribution function is non-analytic in ϕ,

$$f_{\rm Bose}(p) \simeq \frac{T}{g_a|\phi|} \ .$$

It is this non-analyticity that gives rise to the non-analytic cubic term in the effective potential. Importantly, the Fermi–Dirac distribution,

$$f_{\rm Fermi}(p) = \frac{1}{{\rm e}^{\frac{\sqrt{p^2+m_a^2}}{T}} + 1} \ ,$$

is analytic in m_a^2, and hence $\phi^\dagger\phi$, so fermions do not contribute to the cubic term.

With the cubic term in the effective potential, the phase transition is indeed of the first order: at high temperatures the coefficient α is positive and large, and there is one minimum of the effective potential at $\phi = 0$, while for α small but still positive there are two minima. The phase transition occurs at $\alpha \approx 0$; at that moment

$$V_{\rm eff}(\phi, T) \approx -\frac{\beta T}{3}|\phi|^3 + \frac{\lambda}{4}|\phi|^4 \; .$$

We find from this expression that immediately after the phase transition the minimum of $V_{\rm eff}$ is at

$$\phi \simeq \frac{\beta T}{\lambda} \; .$$

Hence, the necessary condition for successful electroweak baryogenesis, $\phi > T$, translates into

$$\beta > \lambda \; . \tag{85}$$

According to (82), λ is proportional to $m_{\rm H}^2$, whereas in the Standard Model β is proportional to $(2M_{\rm W}^3 + M_{\rm Z}^3)$. Therefore, the relation (85) holds for small Higgs boson masses only; in the Standard Model one makes use of (82) and (84) and finds that this happens for $m_{\rm H} < 50$ GeV, while in reality $m_{\rm H} = 125$ GeV. In fact, in the Standard Model with $m_{\rm H} = 125$ GeV, there is no phase transition at all; the electroweak transition is a smooth crossover instead. The latter fact is not visible from the expression (83), but that expression is the lowest-order perturbative result, while the perturbation theory is not applicable for describing the transition in the Standard Model with large $m_{\rm H}$.

This discussion indicates a possible way to make the electroweak phase transition strong. What one needs is the existence of new bosonic fields that have large enough couplings to the Englert–Brout–Higgs field(s), and hence provide large contributions to β. To have an effect on the dynamics of the transition, the new bosons must be present in the cosmic plasma at the transition temperature, $T \sim 100$ GeV, so their masses should not be too high, $M \lesssim 300$ GeV. In supersymmetric extensions of the Standard Model, the natural candidate for a long time has been stop (superpartner of top-quark) whose Yukawa coupling to the Englert–Brout–Higgs field is the same as that of top, that is, large. The light stop scenario for electroweak baryogenesis would indeed work, as has been shown by the detailed analysis in Refs. [68–70].

There are other possibilities to make the electroweak transition strongly first order. Generically, they require an extension of the scalar sector of the Standard Model and predict new fairly light scalars which interact with the Standard Model Englert–Brout–Higgs field and may or may not participate in gauge interactions.

Yet another issue is CP-violation, which has to be strong enough for successful electroweak baryogenesis. As the asymmetry is generated in the interactions of quarks and leptons with the bubble walls, CP-violation must occur at the walls. Recall now that the walls are made of the scalar field(s). This points towards the necessity of CP-violation in the scalar sector, which may only be the case in a theory containing scalar fields other than the Standard Model Englert–Brout–Higgs field.

To summarize, electroweak baryogenesis requires a considerable extension of the Standard Model, with masses of new particles in the range 100–300 GeV. Hence, this mechanism will most likely be ruled out or confirmed by the LHC. We emphasize, however, that electroweak baryogenesis is not the only option at all: an elegant and well-motivated competitor is leptogenesis [14, 15, 71]; there are many other mechanisms proposed in the literature.

6 Before the hot epoch

6.1 Cosmological perturbations: preliminaries

With BBN theory and observations, and due to evidence, albeit indirect, for relic neutrinos, we are confident of the theory of the early Universe at temperatures up to $T \simeq 1$ MeV, which correspond to an

age of $t \simeq 1$ s. With the LHC, we hope to be able to learn the Universe up to temperature $T \sim 100$ GeV and age $t \sim 10^{-10}$ s. Are we going to have a handle on an even earlier epoch?

The key issue in this regard is cosmological perturbations. These are inhomogeneities in the energy density and associated gravitational potentials, in the first place. This type of inhomogeneities is called scalar perturbations, as they are described by 3-scalars. There may exist perturbations of another type, called tensors; these are primordial gravity waves. We will mostly concentrate on scalar perturbations, since they are observed; tensor perturbations are important too, and we comment on them later on. While perturbations of the present size of order 10 Mpc and smaller have large amplitudes today and are non-linear, amplitudes of all known perturbations were small in the past, and the perturbations can be described within the linearized theory. Indeed, CMB temperature anisotropy tells us that the perturbations at the recombination epoch were roughly at the level

$$\delta \equiv \frac{\delta\rho}{\rho} = 10^{-4}\text{–}10^{-5} \; . \tag{86}$$

Thus, the linearized theory works very well before recombination and somewhat later. We will be rather sloppy when talking about scalar perturbations. In general relativity, there is arbitrariness in the choice of reference frame, which can be viewed as a sort of gauge freedom. In a homogeneous and isotropic Universe, there is a preferred reference frame, in which quantities like energy density or distribution function of CMB photons are manifestly homogeneous and isotropic. It is in this frame that the metric has FLRW form (1). Once there are perturbations, no preferred reference frame exists any longer. As an example, one can choose a reference frame such that the three-dimensional hypersurfaces of constant time are hypersurfaces of constant total energy density ρ. In this frame one has $\delta\rho = 0$, so Eq. (86) does not make sense. Yet the Universe is inhomogeneous in this reference frame, since there are inhomogeneous metric perturbations $\delta g_{\mu\nu}(\mathbf{x}, t)$. We will skip these technicalities and denote the scalar perturbation by δ without specifying its gauge-invariant meaning.

Equations for perturbations are obtained by writing for every variable (including metric) an expression like $\rho(\mathbf{x}, t) = \bar{\rho}(t) + \delta\rho(\mathbf{x}, t)$ etc, where $\bar{\rho}(t)$ is the homogeneous and isotropic background, which we discussed in Section 2.3. One inserts the perturbed variables into the Einstein equations and covariant conservation equations $\nabla_\mu T^{\mu\nu} = 0$ and linearizes this set of equations. In many cases one also has to use the linearized Boltzmann equations that govern the distribution functions of particles out of thermal equilibrium; these are necessary for evaluating the linearized perturbations of the energy–momentum tensor. In any case, since the background FLRW metric (1) does not explicitly depend on $\mathbf{x}$, the linearized equations for perturbations do not contain $\mathbf{x}$ explicitly. Therefore, one makes use of the spatial Fourier decomposition

$$\delta(\mathbf{x}, t) = \int \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{x}} \delta(\mathbf{k}, t) \, \mathrm{d}^3 k \; .$$

The advantage is that modes with different momenta $\mathbf{k}$ evolve independently in the linearized theory, i.e., each mode can be treated separately. Recall that $\mathrm{d}\mathbf{x}$ is *not* the physical distance between neighbouring points; the physical distance is $a(t)\mathrm{d}\mathbf{x}$. Thus, $\mathbf{k}$ is *not* the physical momentum (wavenumber); the physical momentum is $\mathbf{k}/a(t)$. While for a given mode the comoving (or coordinate) momentum $\mathbf{k}$ remains constant in time, the physical momentum gets red shifted as the Universe expands, see also Section 2.1. In what follows we set the present value of the scale factor equal to 1 (in a spatially flat Universe this can always be done by rescaling the coordinates $\mathbf{x}$):

$$a_0 \equiv a(t_0) = 1 \; ;$$

then $\mathbf{k}$ is the *present* physical momentum and $2\pi/k$ is the present physical wavelength, which is also called the comoving wavelength.

Properties of scalar perturbations are measured in various ways. Perturbations of fairly large spatial scales (fairly low $\mathbf{k}$) give rise to CMB temperature anisotropy and polarization, so we have very

detailed knowledge of them. Somewhat shorter wavelengths are studied by analysing distributions of galaxies and quasars at present and in relatively near past. There are several other methods, some of which can probe even shorter wavelengths. There is good overall consistency of the results obtained by different methods, so we have a pretty good understanding of many aspects of the scalar perturbations.

The cosmic medium in our Universe has several components that interact only gravitationally: baryons, photons, neutrinos and dark matter. Hence, there may be and, in fact, there are perturbations in each of these components. As we pointed out in Section 4, electromagnetic interactions between baryons, electrons and photons were strong before recombination, so to a reasonable approximation these species made a single fluid, and it is appropriate to talk about perturbations in this fluid. After recombination, baryons and photons evolved independently.

By studying the scalar perturbations, we have learned a number of very important things. To appreciate what they are, it is instructive to consider first the baryon–electron–photon fluid before recombination.

6.2 Perturbations in the expanding Universe: subhorizon and superhorizon regimes.

Perturbations in the baryon–photon fluid before recombination are nothing but sound waves. It is instructive to compare the wavelength of a perturbation with the horizon size. To this end, recall (see Section 2.4) that the horizon size $l_{\rm H}(t)$ is the size of the largest region which is causally connected by the time t, and that

$$l_{\rm H}(t) \sim H^{-1}(t) \sim t$$

at radiation domination and later, see Eq. (21). The latter relation, however, holds *under the assumption that the hot epoch was the first one in cosmology*, i.e., that the radiation domination started right after the Big Bang. This assumption is in the heart of what can be called hot Big Bang theory. We will find that this assumption in fact is *not valid* for our Universe; we are going to see this ad absurdum, so let us stick to the hot Big Bang theory for the time being.

Unlike the horizon size, the physical wavelength of a perturbation grows more slowly. As an example, at radiation domination

$$\lambda(t) = \frac{2\pi a(t)}{k} \propto \sqrt{t}\,,$$

while at matter domination $\lambda(t) \propto t^{2/3}$. For obvious reasons, the modes with $\lambda(t) \ll H^{-1}(t)$ and $\lambda(t) \gg H^{-1}(t)$ are called subhorizon and superhorizon at time t, respectively. We are able to study the modes which are subhorizon *today*; longer modes are homogeneous throughout the visible Universe and are not observed as perturbations. However, *the wavelengths which are subhorizon today were superhorizon at some earlier epoch*. In other words, the physical momentum $k/a(t)$ was smaller than $H(t)$ at early times; at time $t_\times$ such that

$$q(t_\times) \equiv \frac{k}{a(t_\times)} = H(t_\times)\,,$$

the mode entered the horizon, and after that evolved in the subhorizon regime $k/a(t) \gg H(t)$, see Fig. 16. It is straightforward to see that for all cosmologically interesting wavelengths, horizon crossing occurs much later than 1 s after the Big Bang, i.e., at the time we are confident about. So, there is no guesswork at this point.

Question. Estimate the temperature at which a perturbation of comoving size 10 kpc entered the horizon.

Another way to look at the superhorizon–subhorizon behaviour of perturbations is to introduce a

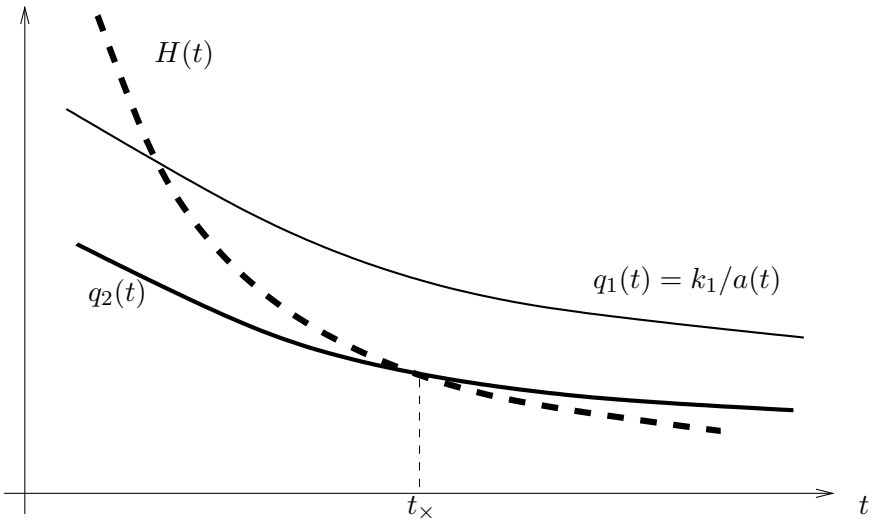

Fig. 16: Physical momenta $q(t) = k/a(t)$ (solid lines, $k_2 < k_1$) and Hubble parameter (dashed line) at radiation- and matter-dominated epochs. Here $t_\times$ is the horizon entry time.

new time coordinate (cf. Eq. (20)),

$$\eta = \int_0^t \frac{\mathrm{d}t'}{a(t')} \ . \tag{87}$$

Note that this integral converges at the lower limit in the hot Big Bang theory. In terms of this time coordinate, the FLRW metric (1) reads

$$\mathrm{d}s^2 = a^2(\eta)(\mathrm{d}\eta^2 - \mathrm{d}\mathbf{x}^2) \ .$$

In coordinates $(\eta, \mathbf{x})$, the light cones $\mathrm{d}s = 0$ are the same as in Minkowski space, and η is the coordinate size of the horizon, see Fig. 17.

Every mode of perturbation has a time-independent coordinate wavelength $2\pi/k$, and at small η it is in superhorizon regime, $2\pi/k \gg \eta$, and after horizon crossing at time $\eta_\times = \eta(t_\times)$ it becomes subhorizon.

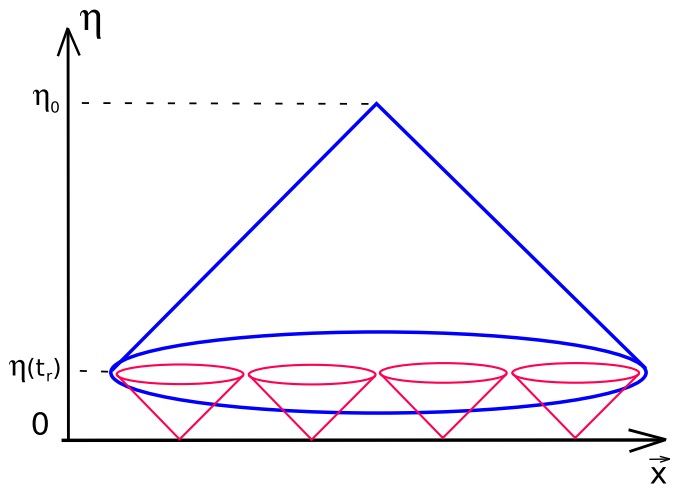

Fig. 17: Causal structure of space–time in the hot Big Bang theory. Here $\mathrm{t_r}$ is the conformal time at recombination.

6.3 Hot epoch was not the first

One immediately observes that this picture falsifies the hot Big Bang theory. Indeed, we see the horizon at recombination $l_{\rm H}(t_{\rm rec})$ at an angle $\Delta\theta \approx 2^\circ$, as schematically shown in Fig. 17. By causality, at recombination there should be no perturbations of larger wavelengths, as any perturbation can be generated within the causal light cone only. In other words, CMB temperature must be isotropic when averaged over angular scales exceeding 2°; there should be no cold or warm spots of angular size larger than 2°. Now, CMB provides us with the photographic picture shown in Fig. 18. It is seen by the naked eye that

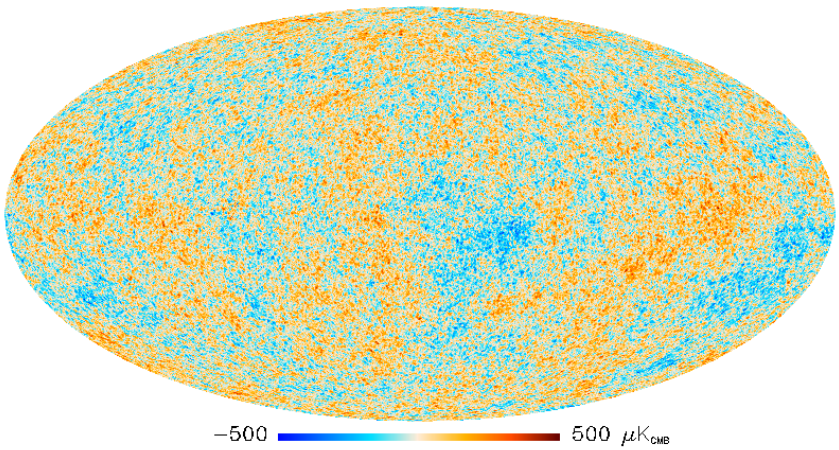

Fig. 18: CMB sky as seen by Planck

there are cold and warm regions whose angular size much exceeds 2°; in fact, there are perturbations of all angular sizes up to those comparable to the entire sky. We come to an important conclusion: *the scalar perturbations were built in at the very beginning of the hot epoch. The hot epoch was not the first, it was preceded by some other epoch, and the cosmological perturbations were generated then.*

Question. Assuming (erroneously) that there is no dark energy, and that recombination occurred deep in the matter-dominated epoch, estimate the angular scale of the horizon at recombination.

Another manifestation of the fact that the scalar perturbations were there already at the beginning of the hot epoch is the existence of peaks in the angular spectrum of CMB temperature. In general, perturbations in the baryon–photon medium before recombination are acoustic waves,

$$\delta(\mathbf{k},t) = \delta(\mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{x}} \cos\left[\int_0^t v_{\rm s} \frac{k}{a(t')} \mathrm{d}t' + \psi_{\mathbf{k}}\right] , \tag{88}$$

where $v_{\rm s}$ is sound speed, $\delta(\mathbf{k})$ is time-independent amplitude and $\psi_{\mathbf{k}}$ is time-independent phase. This expression is valid, however, in the subhorizon regime only, i.e., at late times. The two solutions in superhorizon regime at radiation domination are

$$\delta(t) = \text{const} , \tag{89a}$$

$$\delta(t) = \frac{\text{const}}{t^{3/2}} . \tag{89b}$$

Were the perturbations generated in a causal way at radiation domination, they would be always subhorizon. In that case the solutions (89) would be irrelevant, and there would be no reason for a particular

choice of phase $\psi_{\mathbf{k}}$ in Eq. (88). One would rather expect that $\psi_{\mathbf{k}}$ is a random function of $\mathbf{k}$. This is indeed the case for specific mechanisms of the generation of density perturbations at the hot epoch [72].

On the other hand, if the perturbations existed at the very beginning of the hot epoch, they were superhorizon at sufficiently early times, and were described by the solutions (89). The consistency of the whole cosmology requires that the amplitude of perturbations was small at the beginning of the hot stage. The solution (89b) rapidly decays away, and towards the horizon entry the perturbation is in constant mode (89a). So, the initial condition for the further evolution is unique modulo amplitude $\delta(\mathbf{k})$, and hence the phase $\psi(\mathbf{k})$ is uniquely determined. For modes that enter the horizon at radiation domination this phase is equal to zero and, after entering the horizon, the modes oscillate as follows:

$$\delta(\mathbf{k},t) = \delta(\mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{x}} \cos\left[\int_0^t v_{\mathrm{s}} \frac{k}{a(t')} \mathrm{d}t'\right] .$$

At recombination, the perturbation is

$$\delta(\mathbf{k},t_{\mathrm{r}}) = \delta(\mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{x}} \cos(kr_{\mathrm{s}}) , \tag{90}$$

where

$$r_{\mathrm{s}} = \int_0^{t_{\mathrm{rec}}} v_{\mathrm{s}} \frac{\mathrm{d}t'}{a(t')}$$

is the comoving size of the sound horizon at recombination, while its physical size equals $a(t_{\mathrm{rec}})r_{\mathrm{s}}$. So, we see that the density perturbation of the baryon–photon plasma at recombination *oscillates as a function of wavenumber* k. The period of this oscillation is determined by r_{s}, which is a straightforwardly calculable quantity.

So, if the perturbations existed already at the beginning of the hot stage, they show the oscillatory behaviour in momentum at the recombination epoch. This translates into an oscillatory pattern of the CMB temperature angular spectrum. Omitting details, the fluctuation of the CMB temperature is partially due to the density perturbation in the baryon–photon medium at recombination. Namely, the temperature fluctuation of photons coming from the direction $\mathbf{n}$ in the sky is, roughly speaking,

$$\delta T(\mathbf{n}) \propto \delta_\gamma(\mathbf{x}_{\mathbf{n}}, \eta_{\mathrm{rec}}) + \delta T_{\mathrm{smooth}}(\mathbf{n}) ,$$

where $T_{\mathrm{smooth}}(\mathbf{n})$ corresponds to the non-oscillatory part of the CMB angular spectrum and

$$\mathbf{x}_{\mathbf{n}} = -\mathbf{n}(\eta_0 - \eta_{\mathrm{rec}}) .$$

Here $(\eta_0 - \eta_{\mathrm{rec}})$ is the coordinate distance to the sphere of photon last scattering, and $\mathbf{x}_{\mathbf{n}}$ is the coordinate of the place where the photons coming from the direction $\mathbf{n}$ scatter the last time. The quantity $T_{\mathrm{smooth}}(\mathbf{n})$ originates from the gravitational potential generated by the dark matter perturbation; dark matter has zero pressure at all times, so there are no sound waves in this component, and there are no oscillations at recombination as a function of momentum.

One expands the temperature variation on the celestial sphere in spherical harmonics:

$$\delta T(\mathbf{n}) = \sum_{lm} a_{lm} Y_{lm}(\theta,\phi).$$

The multipole number l characterizes the temperature fluctuations at the angular scale $\Delta\theta = \pi/l$. The sound waves of momentum k are seen roughly at an angle $\Delta\theta = \Delta x/(\eta_0 - \eta_{\mathrm{rec}})$, where $\Delta x = \pi/k$ is the coordinate half-wavelength. Hence, there is the correspondence

$$l \longleftrightarrow k(\eta_0 - \eta_{\mathrm{rec}}) .$$

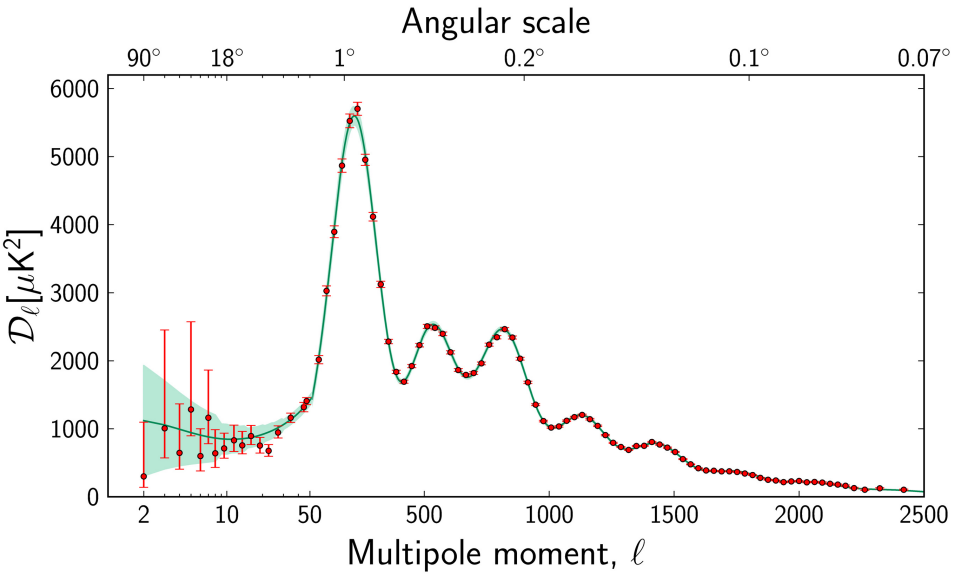

Fig. 19: The angular spectrum of the CMB temperature anisotropy [73]. The quantity on the vertical axis is D_l defined in (92). Note the unconventional scale on the horizontal axis, aimed at showing both small-l region (large angular scales) and large-l region.

Oscillations in momenta in Eq. (90) thus translate into oscillations in l, and these are indeed observed, see Fig. 19.

To understand what is shown in Fig. 19, we note that all observations today support the hypothesis that a_{lm} *are independent Gaussian random variables*. For a hypothetical ensemble of Universes like ours, the average values of products of the coefficients a_{lm} would obey

$$\langle a_{lm} a^*_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}. \tag{91}$$

This gives the expression for the temperature fluctuation:

$$\langle [\delta T(\mathbf{n})]^2 \rangle = \sum_l \frac{2l+1}{4\pi} C_l \approx \int \frac{\mathrm{d}l}{l} \mathcal{D}_l \,,$$

where

$$\mathcal{D}_l = \frac{l(l+1)}{2\pi} C_l \,. \tag{92}$$

Of course, one cannot measure the ensemble average (91). The definition of C_l used in experiments is

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2 \,,$$

where a_{lm} are measured quantities. Since we have only one Universe, this is generically different from the ensemble average (91): for given l, there are only $2l+1$ measurements, and the intrinsic statistical uncertainty—cosmic variance—is of order $(2l+1)^{-1/2}$. It is this uncertainty, rather than experimental error, that is shown in Fig. 19.

We conclude that the facts that the CMB angular spectrum has oscillatory behaviour and that there are sizeable temperature fluctuations at $l < 50$ (angular scale greater than the angular size of 2° of the horizon at recombination) unambiguously tell us that the density perturbations were indeed superhorizon at the hot cosmological stage. The hot epoch has to be preceded by some other epoch—the epoch of the generation of perturbations.

6.4 Primordial scalar perturbations

There are several things which we already know about the primordial density perturbations. By 'primordial' we mean the perturbations deep in the superhorizon regime at the radiation-domination epoch. As we already know, perturbations are time-independent in this regime, see Eq. (89a). They set the initial conditions for further evolution, and this evolution is well understood, at least in the linear regime. Hence, using observational data, one is able to measure the properties of primordial perturbations. Of course, since the properties we know of are established by observations, they are valid within certain error bars. Conversely, deviations from the results listed below, if observed, would be extremely interesting.

First, density perturbations are *adiabatic*. This means that there are perturbations in the energy density, but *not in composition*. More precisely, the baryon-to-entropy ratio and dark matter-to-entropy ratio are constant in space,

$$\delta\left(\frac{n_{\rm B}}{s}\right) = \text{const}\,, \qquad \delta\left(\frac{n_{\rm DM}}{s}\right) = \text{const}\,. \tag{93}$$

This is consistent with the generation of the baryon asymmetry and dark matter at the hot cosmological epoch: in that case, all particles were at thermal equilibrium early at the hot epoch, the temperature completely characterized the whole cosmic medium at that time and as long as physics behind the baryon asymmetry and dark matter generation is the same everywhere in the Universe, the baryon and dark matter abundances (relative to the entropy density) are necessarily the same everywhere. In principle, there may exist *entropy* (another term is isocurvature) perturbations, such that at the early hot epoch energy density (dominated by relativistic matter) was homogeneous, while the composition was not. This would give initial conditions for the evolution of density perturbations, which would be entirely different from those characteristic of the adiabatic perturbations. As a result, the angular spectrum of the CMB temperature anisotropy would be entirely different. No admixture of the entropy perturbations has been detected so far, but it is worth emphasizing that even a small admixture will show that many popular mechanisms for generating dark matter and/or baryon asymmetry have nothing to do with reality. One will have to think, instead, that the baryon asymmetry and/or dark matter were generated before the beginning of the hot stage. A notable example is the axion misalignment mechanism discussed in Section 4.4: in a latent sense, the axion dark matter exists from the very beginning in that case, and perturbations in the axion field $\delta\theta_0(\mathbf{x})$ (which may be generated together with the adiabatic perturbations) would show up as entropy perturbations in dark matter.

Second, the primordial density perturbations are *Gaussian random fields*. Gaussianity means that the three-point and all odd correlation functions vanish, while the four-point function and all higher order even correlation functions are expressed through the two-point function via Wick's theorem:

$$\begin{aligned} \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle &= 0, \\ \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4)\rangle &= \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\rangle\cdot\langle\delta(\mathbf{k}_3)\delta(\mathbf{k}_4)\rangle \\ &\quad + \text{permutations of momenta}\,. \end{aligned}$$

We note that this property is characteristic of *vacuum fluctuations of non-interacting (linear) quantum fields.* Hence, it is quite likely that the density perturbations originate from the enhanced vacuum fluctuations of non-interacting or weakly interacting quantum field(s). The free quantum field has the general form

$$\phi(\mathbf{x},t) = \int \mathrm{d}^3k \mathrm{e}^{-\mathrm{i}\mathbf{k}\mathbf{x}}\left(f^{(+)}_{\mathbf{k}}(t)a^\dagger_{\mathbf{k}} + \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{x}}f^{(-)}_{\mathbf{k}}(t)a_{\mathbf{k}}\right)\,,$$

where $a^\dagger_{\mathbf{k}}$ and $a_{\mathbf{k}}$ are creation and annihilation operators. For the field in Minkowski space–time, one has $f^{(\pm)}_{\mathbf{k}}(t) = \mathrm{e}^{\pm\mathrm{i}\omega_k t}$, while enhancement, e.g., due to the evolution in time-dependent background, means that $f^{(\pm)}_{\mathbf{k}}$ are large. But, in any case, Wick's theorem is valid, provided that the state of the system is vacuum, $a_{\mathbf{k}}|0\rangle = 0$.

We note in passing that *non-Gaussianity* is an important topic of current research. It would show up as a deviation from Wick's theorem. As an example, the three-point function (bispectrum) may be non-vanishing,

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle = \delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\, G(k_i^2;\ \mathbf{k}_1\mathbf{k}_2;\ \mathbf{k}_1\mathbf{k}_3) \neq 0\ .$$

The functional dependence of $G(k_i^2;\ \mathbf{k}_1\mathbf{k}_2;\ \mathbf{k}_1\mathbf{k}_3)$ on its arguments is different in different models of generation of primordial perturbations, so this shape is a potential discriminator. In some models the bispectrum vanishes, e.g., due to symmetries. In that case the trispectrum (connected four-point function) may be measurable instead. It is worth emphasizing that non-Gaussianity has not been detected yet.

Another important property is that the primordial power spectrum of density perturbations *is nearly, but not exactly, flat*. For a homogeneous and anisotropic Gaussian random field, the power spectrum completely determines its only characteristic, the two-point function. A convenient definition is

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = \frac{1}{4\pi k^3}\mathcal{P}(k)\delta(\mathbf{k}+\mathbf{k}')\ . \tag{94}$$

The power spectrum $\mathcal{P}(k)$ defined in this way determines the fluctuation in a logarithmic interval of momenta,

$$\langle\delta^2(\mathbf{x})\rangle = \int_0^\infty \frac{\mathrm{d}k}{k}\,\mathcal{P}(k)\ .$$

By definition, the flat spectrum is such that $\mathcal{P}$ is independent of k. In this case all spatial scales are alike; no scale is enhanced with respect to another. It is worth noting that the flat spectrum was conjectured by Harrison [74], Zeldovich [75] and Peebles and Yu [76] at the beginning of the 1970s, long before realistic mechanisms of the generation of density perturbations have been proposed.

In view of the approximate flatness, a natural parametrization is

$$\mathcal{P}(k) = A_\mathrm{s}\left(\frac{k}{k_*}\right)^{n_\mathrm{s}-1}\ , \tag{95}$$

where A_s is the amplitude, $n_\mathrm{s}-1$ is the tilt and k_* is a fiducial momentum, chosen at one's convenience. The flat spectrum in this parametrization has $n_\mathrm{s}=1$. This is inconsistent with the cosmological data, which give [21]

$$n_\mathrm{s} = 0.968 \pm 0.06\ .$$

This quantifies what we mean by a nearly, but not exactly flat, power spectrum.

6.5 Inflation or not?

The pre-hot epoch must be long in terms of the time variable η introduced in Eq. (87). What we would like to have is that a large part of the Universe (e.g., the entire visible part) be causally connected towards the end of that epoch, see Fig. 20. A long duration in η does not necessarily mean a long duration in physical time t; in fact, the physical duration of the pre-hot epoch may be tiny.

An excellent hypothesis on the pre-hot stage is inflation, the epoch of nearly exponential expansion [77–82],

$$a(t) = \mathrm{e}^{\int H\mathrm{d}t}\ , \qquad H \approx \text{const}\ .$$

Inflation makes the whole visible Universe, and likely a much greater region of space, causally connected at very early times. The horizon size at inflation is at least

$$l_\mathrm{H}(t) = a(t)\int_{t_\mathrm{i}}^{t}\frac{\mathrm{d}t'}{a(t')} = H^{-1}\mathrm{e}^{H(t-t_\mathrm{i})}\ ,$$

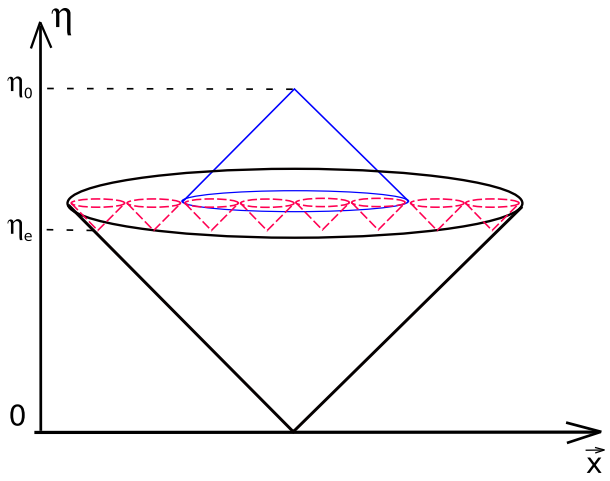

Fig. 20: Causal structure of space–time in the real Universe

where $t_{\rm i}$ is the time inflation begins, and we set $H = \text{const}$ for illustrational purposes. This size is huge for $t - t_{\rm i} \gg H^{-1}$, as desired.

Question. Assuming that at inflation $H \ll M_{\rm Pl}$, show that if the duration of inflation Δt is larger than $100H^{-1}$, the whole visible Universe is causally connected by the end of inflation. What is $100H^{-1}$ in seconds for $H = 10^{15}$ GeV? Using the time variable η, show that the causal structure of space–time in inflationary theory with $\Delta t > 100H^{-1}$ is the one shown in Fig. 20.

From the viewpoint of perturbations, the physical momentum $q(t) = k/a(t)$ decreases (gets red shifted) at inflation, while the Hubble parameter stays almost constant. So, every mode is first subhorizon ($q(t) \gg H(t)$) and later superhorizon ($q(t) \ll H(t)$). This situation is opposite to what happens at radiation and matter domination, see Fig. 21; this is precisely the prerequisite for generating the density perturbations. In fact, inflation does generate primordial density perturbations [83–87] whose properties are consistent with everything we know about them. Indeed, at the inflationary epoch, fluctuations of all

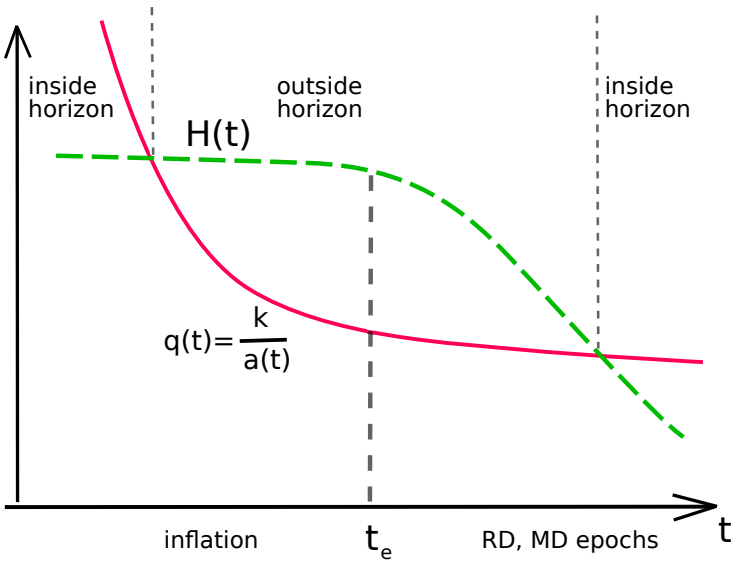

Fig. 21: Physical momentum and Hubble parameter at inflation and later: $t_{\rm e}$ is the time of the inflation end

light fields get enhanced greatly due to the fast expansion of the Universe. This is true, in particular, for the field that dominates the energy density at inflation, called an inflaton. Enhanced vacuum fluctuations of the inflaton are nothing but perturbations in the energy density at the inflationary epoch, which are reprocessed into perturbations in the hot medium after the end of inflation. The inflaton field is very weakly coupled, so the non-Gaussianity in the primordial scalar perturbations is very small [88]. In fact, it is so small that its detection is problematic even in the distant future.

The approximate flatness of the primordial power spectrum in inflationary theory is explained by the symmetry of the de Sitter space–time, which is the space–time of constant Hubble rate,

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{e}^{2Ht}\mathrm{d}\mathbf{x}^2 \,, \qquad H = \mathrm{const} \,.$$

This metric is invariant under spatial dilatations supplemented by time translations,

$$\mathbf{x} \to \lambda\mathbf{x} \,, \quad t \to t - \frac{1}{2H}\log\lambda \,.$$

Therefore, all spatial scales are alike, which is also a defining property of the flat power spectrum. At inflation, H is almost constant in time and the de Sitter symmetry is an approximate symmetry. For this reason, inflation automatically generates a nearly flat power spectrum.

The distinguishing property of inflation is *the generation of tensor modes (primordial gravity waves)* of sizeable amplitude and nearly flat power spectrum. The gravity waves are thus a smoking gun for inflation. The reason for their generation at inflation is that the exponential expansion of the Universe enhances vacuum fluctuations of all fields, including the gravitational field itself. Particularly interesting are gravity waves whose present wavelengths are huge, 100 Mpc and larger. Many inflationary models predict their amplitudes to be very large, of order 10^{-6} or so. Shorter gravity waves are generated too, but their amplitudes decay after horizon entry at radiation domination, and today they have much smaller amplitudes making them inaccessible to gravity wave detectors like LIGO or VIRGO, pulsar timing arrays etc. A conventional characteristic of the amplitude of primordial gravity waves is the tensor-to-scalar ratio

$$r = \frac{\mathcal{P}_\mathrm{T}}{\mathcal{P}} \,,$$

where $\mathcal{P}$ is the scalar power spectrum defined in Eq. (94) and $\mathcal{P}_\mathrm{T}$ is the tensor power spectrum defined in a similar way, but for transverse traceless metric perturbations h_{ij}. The result of the search for effects of the tensor modes on CMB temperature anisotropy is shown in Fig. 22. This search has already ruled out some of the popular inflationary models.

All the above referred to the simplest, single-field inflationary models. In models with more than one relevant field, the situation may be different. In particular, sizeable non-Gaussianity may be generated, while the amplitude of tensor perturbations may be very low. So, it would be rather difficult to rule out the inflationary scenario as a whole.

Inflation is not the only hypothesis proposed so far. One option is the bouncing Universe scenario, which assumes that the cosmological evolution begins from contraction, then the contracting stage terminates at some moment of time (bounce) and is followed by expansion. A version is the cycling Universe scenario with many cycles of contraction–bounce–expansion. See reviews by Lehners and Brandenberger in Ref. [16–20]. Another scenario is that the Universe starts out from a nearly flat and static state with nearly vanishing energy density. Then the energy density increases and, according to the Friedmann equation, the expansion speeds up. This goes under the name of the Genesis scenario [90]. Theoretical realizations of these scenarios are more difficult than inflation, but they are not impossible, as became clear recently.

The generation of the density perturbations is less automatic in scenarios alternative to inflation. Similarly to inflationary theory, the flatness of the scalar power spectrum is likely to be due to some symmetry. One candidate symmetry is conformal invariance [91–94]. The point is that the conformal

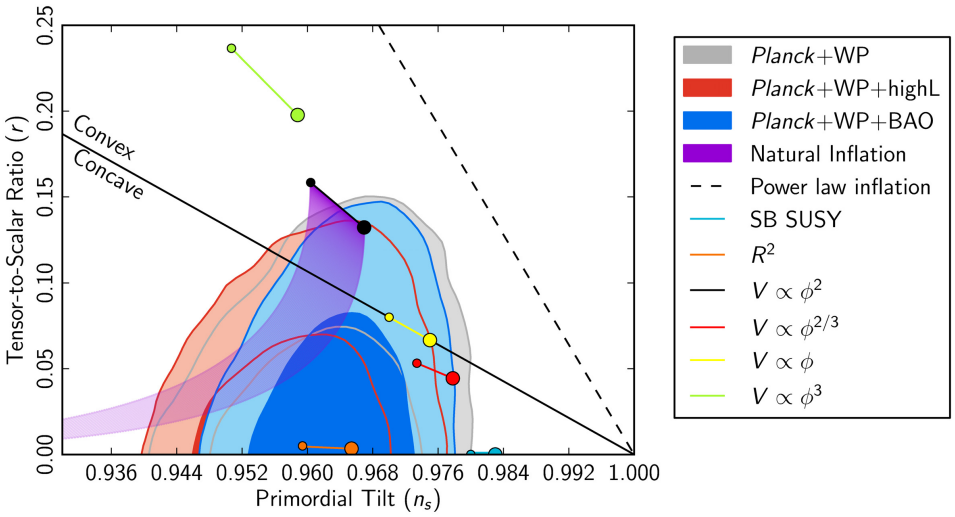

Fig. 22: Allowed regions (at 68% and 95% confidence levels) in the plane (n_s, r), where n_s is the scalar spectral index and r is the tensor-to-scalar ratio [89]. The right lower corner (the point $(1.0, 0.0)$) is the Harrison–Zeldovich point (flat scalar spectrum, no tensor modes). Intervals show predictions of popular inflationary models.

group includes dilatations, $x^\mu \to \lambda x^\mu$. This property indicates that the theory possesses no scale and has a good chance for producing the flat spectrum. A model building along this direction has begun rather recently [92–94].

6.6 Hunt continues

Until now, only very basic facts about the primordial cosmological perturbations have been observationally established. Even though very suggestive, these facts by themselves are not sufficient to unambiguously figure out what was the Universe at the pre-hot epoch of its evolution. New properties of cosmological perturbations will hopefully be discovered in the future and shed more light on this pre-hot epoch. Let us discuss some of the potential observables.

6.6.1 Tensor perturbations = relic gravity waves

As we discussed, primordial tensor perturbations are predicted by many inflationary models. On the other hand, there seems to be no way of generating a nearly flat tensor power spectrum in alternatives to inflation. In fact, most, if not all, alternative scenarios predict unobservably small tensor amplitudes. This is why we said that tensor perturbations are a smoking gun for inflation. Until recently, the most sensitive probe of the tensor perturbations has been the CMB temperature anisotropy [95–98]. However, the most promising tool is the CMB polarization. The point is that a certain class of polarization patterns (called B-mode) is generated by tensor perturbations, while scalar perturbations are unable to create it [99, 100]. Hence, dedicated experiments aiming at measuring the CMB polarization may well discover the tensor perturbations, i.e., relic gravity waves. Needless to say, this would be a profound discovery. To avoid confusion, let us note that the CMB polarization has been already observed, but it belongs to another class of patterns (so-called E-mode) and is consistent with the existence of the scalar perturbations only. The original claim of the BICEP-2 experiment [101] to detect the B-mode generated by primordial tensor perturbations was turned down [102]: the B-mode is there, but it is due to dust in our Galaxy.

6.6.2 Non-Gaussianity.

As we pointed out already, non-Gaussianity of density perturbations is very small in the simplest inflationary models. Hence, its discovery will signal that either inflation and inflationary generation of density perturbations occurred in a rather complicated way, or an alternative scenario was realized. Once the non-Gaussianity is discovered, and its shape is revealed even with modest accuracy, many concrete models will be ruled out, while at most a few will get strong support.

6.6.3 Statistical anisotropy.

In principle, the power spectrum of density perturbations may depend on the direction of momentum, e.g.,

$$\mathcal{P}(\mathbf{k}) = \mathcal{P}_0(k)\left(1 + w_{ij}(k)\frac{k_i k_j}{k^2} + \cdots\right),$$

where w_{ij} is a fundamental tensor in our part of the Universe (odd powers of k_i would contradict commutativity of the Gaussian random field $\delta(\mathbf{k})$, see Eq. (94)). Such a dependence would definitely imply that the Universe was anisotropic at the pre-hot stage, when the primordial perturbations were generated. This statistical anisotropy is rather hard to obtain in inflationary models, though it is possible in inflation with strong vector fields [103–105]. On the other hand, statistical anisotropy is natural in some other scenarios, including conformal models [106, 107]. The statistical anisotropy would show up in correlators [108, 109]

$$\langle a_{lm} a_{l'm'} \rangle \qquad \text{with } l' \neq l \text{ and/or } m' \neq m.$$

At the moment, the constraints [110, 111] on statistical anisotropy obtained by analysing the CMB data are getting into the region which is interesting from the viewpoint of some (though not many) models of the pre-hot epoch.

6.6.4 Admixture of entropy perturbations.

As we explained above, even a small admixture of entropy perturbations would force us to abandon the most popular scenarios of the generation of baryon asymmetry and/or dark matter, which assumed that it happened at the hot epoch. Once the dark matter entropy mode is discovered, the WIMP dark matter would no longer be well motivated, while other, very weakly interacting dark matter candidates, like axions or superheavy relics, would be preferred. This would redirect the experimental search for dark matter.

7 Conclusion

It is by now commonplace that the two fields studying together the most fundamental properties of matter and the Universe—particle physics and cosmology—are tightly interrelated. The present situations in these fields have much in common too. On the particle-physics side, the Standard Model has been completed by the expected discovery of the Higgs boson. On the other hand, relatively recently a fairly unexpected discovery of neutrino oscillations was made, which revolutionized our view on particles and their interactions. There are grounds to hope for even more profound discoveries, notably by the LHC experiments. While in the past there were definite predictions of the Standard Model, which eventually were confirmed, there are numerous hypotheses concerning new physics, none of which is undoubtedly plausible. On the cosmology side, the Standard Model of cosmology, ΛCDM, has been shaped, again not without an unexpected and revolutionary discovery, in this case of the accelerated expansion of the Universe. We hope for further profound discoveries in cosmology too. It may well be that we will soon learn which is the dark matter particle; again, there is an entire zoo of candidates, several of which are serious competitors. The discoveries of new properties of cosmological perturbations will hopefully reveal the nature of the pre-hot epoch. There is a clear best guess, inflation, but it is not excluded that future observational data will point towards something else.

Neither in particle physics nor in cosmology are new discoveries guaranteed, however. Nature may hide its secrets. Whether or not it does is the biggest open issue in fundamental physics.

Acknowledgement

This work is supported by the Russian Science Foundation grant 14-22-00161.

References

[1] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).

[2] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Cambridge, 2005). http://dx.doi.org/10.1017/CBO9780511790553

[3] S. Weinberg, *Cosmology* (Oxford University Press, Oxford, 2008).

[4] A.R. Liddle and D.H. Lyth, *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure* (Cambridge University Press, Cambridge, 2009). http://dx.doi.org/10.1017/CBO9780511819209

[5] D.S. Gorbunov and V.A. Rubakov, *Introduction to the Theory of the Early Universe: Hot Big Bang Theory* (World Scientific, Hackensack, NJ, 2011). http://dx.doi.org/10.1142/7874

[6] D.S. Gorbunov and V.A. Rubakov, *Introduction to the Theory of the Early Universe: Cosmological Perturbations and Inflationary Theory* (World Scientific, Hackensack, NJ, 2011). http://dx.doi.org/10.1142/7874

[7] K.A. Olive, arXiv:astro-ph/0301505.

[8] G. Bertone, D. Hooper and J. Silk, *Phys. Rep.* **405**(5–6) (2005) 279. http://dx.doi.org/10.1016/j.physrep.2004.08.031

[9] A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, *Annu. Rev. Nucl. Part. Sci.* **59** (2009) 191. http://dx.doi.org/10.1146/annurev.nucl.010909.083654

[10] M. Kawasaki and K. Nakayama, *Annu. Rev. Nucl. Part. Sci.* **63** (2013) 69. http://dx.doi.org/10.1146/annurev-nucl-102212-170536

[11] V.A. Rubakov and M.E. Shaposhnikov, *Usp. Fiz. Nauk* **166** (1996) 493 [Engl. Trans. *Phys. Usp.* **39** (1996) 461]. http://dx.doi.org/10.3367/UFNr.0166.199605d.0493

[12] M. Trodden, *Rev. Mod. Phys.* **71**(5) (1999) 1463. http://dx.doi.org/10.1103/RevModPhys.71.1463

[13] T. Konstandin, *Usp. Fiz. Nauk* **183** (2013) 785 [Engl. trans. *Phys. Usp.* **56** (2013) 747]. http://dx.doi.org/10.3367/UFNr.0183.201308a.0785

[14] W. Buchmuller, R.D. Peccei and T. Yanagida, *Annu. Rev. Nucl. Part. Sci.* **55** (2005) 311. http://dx.doi.org/10.1146/annurev.nucl.55.090704.151558

[15] S. Davidson, E. Nardi and Y. Nir, *Phys. Rep.* **466**(4–5) (2008) 105. http://dx.doi.org/10.1016/j.physrep.2008.06.002

[16] D.H. Lyth and A. Riotto, *Phys. Rep.* **314**(1–2) (1999) 16. http://dx.doi.org/10.1016/S0370-1573(98)00128-8

[17] B.A. Bassett, S. Tsujikawa and D. Wands, *Rev. Mod. Phys.* **78**(2) (2006) 537. http://dx.doi.org/10.1103/RevModPhys.78.537

[18] J.L. Lehners, *Phys. Rep.* **465**(6) (2008) 223-263. http://dx.doi.org/10.1016/j.physrep.2008.06.001

[19] A. Mazumdar and J. Rocher, *Phys. Rep.* **497**(4–5) (2011) 85. http://dx.doi.org/10.1016/j.physrep.2010.08.001

[20] R.H. Brandenberger, *Lect. Notes Phys.* **863** (2013) 333. http://dx.doi.org/10.1007/978-3-642-33036-0_12

[21] P.A.R. Ade *et al.* (Planck Collaboration), arXiv:1502.01589 [astro-ph.CO].

[22] P.A.R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **571** (2014) A16. http://dx.doi.org/10.1051/0004-6361/201321591

[23] E. Gawiser and J. Silk, *Phys. Rep.* **333** (2000) 245. http://dx.doi.org/10.1016/S0370-1573(00)00025-9

[24] K. Hagiwara *et al.* (Particle Data Group Collaboration), *Phys. Rev.* D **66**(1) (2002) 010001. http://dx.doi.org/10.1103/PhysRevD.66.010001

[25] K.A. Olive *et al.* (Particle Data Group Collaboration), *Chin. Phys.* C **38**(9) (2014) 090001. http://dx.doi.org/10.1088/1674-1137/38/9/090001

[26] E. Giusarma *et al.*, *Phys. Rev.* D **90**(4) (2014) 043507. http://dx.doi.org/10.1103/PhysRevD.90.043507

[27] N.G. Busca *et al.*, *Astron. Astrophys.* **552** (2013) A96. http://dx.doi.org/10.1051/0004-6361/201220724

[28] V.A. Rubakov, *Phys. Rev.* D **61**(6) (2000) 061501. http://dx.doi.org/10.1103/PhysRevD.61.061501

[29] P.J. Steinhardt and N. Turok, *Science* **312**(5777) (2006) 1180. http://dx.doi.org/10.1126/science.1126231

[30] S. Weinberg, *Phys. Rev. Lett.* **59**(22) (1987) 2607. http://dx.doi.org/10.1103/PhysRevLett.59.2607

[31] A.D. Linde, Inflation and quantum cosmology, in *Three Hundred Years of Gravitation*, Eds S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1987), p. 604.

[32] J.P. Kneib *et al.*, *Astrophys. J.* **598**(2) (2003) 804. http://dx.doi.org/10.1086/378633

[33] D. Clowe *et al.*, *Astrophys. J.* **648**(2) (2006) L109. http://dx.doi.org/10.1086/508162

[34] K.G. Begeman, A.H. Broeils and R.H. Sanders, *Mon. Not. R. Astron. Soc.* **249**(3) (1991) 523. http://dx.doi.org/10.1093/mnras/249.3.523

[35] T. Han, Z. Liu and A. Natarajan, *J. High Energy Phys.* **1311** (2013) 008. http://dx.doi.org/10.1007/JHEP11(2013)008

[36] A.D. Avrorin *et al.* (Baikal Collaboration), *Astropart. Phys.* **62** (2014) 12. http://dx.doi.org/10.1016/j.astropartphys.2014.07.006 http://dx.doi.org/10.1103/PhysRevLett.111.171101

[37] L. Bergstrom *et al.*, *Phys. Rev. Lett.* **111**(17) (2013) 171101. http://dx.doi.org/10.1103/physrevlett.111.171101

[38] V. Khachatryan *et al.* (CMS Collaboration), arXiv:1408.3583 [hep-ex]. http://dx.doi.org/10.1140/epjc/s10052-015-3451-4

[39] G. 't Hooft, *Phys. Rev. Lett.* **37**(1) (1976) 8. http://dx.doi.org/10.1103/PhysRevLett.37.8

[40] C.G. Callan, R.F. Dashen and D.J. Gross, *Phys. Lett.* B **63**(3) (1976) 334. http://dx.doi.org/10.1016/0370-2693(76)90277-X

[41] R. Jackiw and C. Rebbi, *Phys. Rev. Lett.* **37**(3) (1976) 172. http://dx.doi.org/10.1103/PhysRevLett.37.172

[42] J.E. Kim and G. Carosi, *Rev. Mod. Phys.* **82**(1) (2010) 557. http://dx.doi.org/10.1103/RevModPhys.82.557

[43] R.D. Peccei and H.R. Quinn, *Phys. Rev. Lett.* **38**(25) (1977) 1440. http://dx.doi.org/10.1103/PhysRevLett.38.1440

[44] S. Weinberg, *Phys. Rev. Lett.* **40**(4) (1978) 223. http://dx.doi.org/10.1103/PhysRevLett.40.223

[45] F. Wilczek, *Phys. Rev. Lett.* **40**(5) (1978) 279. http://dx.doi.org/10.1103/PhysRevLett.40.279

[46] M. Dine, W. Fischler and M. Srednicki, *Phys. Lett. B* **104**(3) (1981) 199. http://dx.doi.org/10.1016/0370-2693(81)90590-6

[47] A.R. Zhitnitsky, *Yad. Fiz.* **31**(2) (1980) 497 [Engl. trans. *Sov. J. Nucl. Phys.* **31** (1980) 260].

[48] J.E. Kim, *Phys. Rev. Lett.* **43**(2) (1979) 103. http://dx.doi.org/10.1103/PhysRevLett.43.103

[49] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Nucl. Phys.* B **166**(3) (1980) 493. http://dx.doi.org/10.1016/0550-3213(80)90209-6

[50] A. Vilenkin and A.E. Everett, *Phys. Rev. Lett.* **48**(26) (1982) 1867. http://dx.doi.org/10.1103/PhysRevLett.48.1867

[51] R.A. Battye and E.P.S. Shellard, arXiv:astro-ph/9909231.

[52] J. Preskill, M.B. Wise and F. Wilczek, *Phys. Lett.* B **120**(1–3) (1983) 127. http://dx.doi.org/10.1016/0370-2693(83)90637-8

[53] L.F. Abbott and P. Sikivie, *Phys. Lett.* B **120**(1–3) (1983) 133. http://dx.doi.org/10.1016/0370-2693(83)90638-X

[54] M. Dine and W. Fischler, *Phys. Lett.* B **120**(1–3) (1983) 137. http://dx.doi.org/10.1016/0370-2693(83)90639-1

[55] A. Ringwald, *J. Phys. Conf. Ser.* **485** (2014) 012013. http://dx.doi.org/10.1088/1742-6596/485/1/012013

[56] D. Gorbunov, A. Khmelnitsky and V. Rubakov, *J. Cosmol. Astropart. Phys.* **0810** (2008) 041. http://dx.doi.org/10.1088/1475-7516/2008/10/041

[57] M. Laine and M. Shaposhnikov, *J. Cosmol. Astropart. Phys.* **0806** (2008) 031. http://dx.doi.org/10.1088/1475-7516/2008/06/031

[58] X.-D. Shi and G.M. Fuller, *Phys. Rev. Lett.* **82**(14) (1999) 2832. http://dx.doi.org/10.1103/PhysRevLett.82.2832

[59] D. Gorbunov, A. Khmelnitsky and V. Rubakov, *J. High Energy Phys.* **0812** (2008) 055. http://dx.doi.org/10.1088/1126-6708/2008/12/055

[60] A. de Gouvea, T. Moroi and H. Murayama, *Phys. Rev.* D **56**(2) (1997) 1281. http://dx.doi.org/10.1103/PhysRevD.56.1281

[61] A.D. Sakharov, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32 [Engl. trans. *JETP Lett.* **5** (1967) 24]. http://dx.doi.org/10.1070/PU1991v034n05ABEH002497

[62] V.A. Kuzmin, *Pisma Zh. Eksp. Teor. Fiz.* **12** (1970) 335.

[63] I. Affleck and M. Dine, *Nucl. Phys.* B **249**(2) (1985) 361. http://dx.doi.org/10.1016/0550-3213(85)90021-5

[64] F.R. Klinkhamer and N.S. Manton, *Phys. Rev.* D **30**(10) (1984) 2212. http://dx.doi.org/10.1103/PhysRevD.30.2212

[65] V.A. Kuzmin, V.A. Rubakov and M.E. Shaposhnikov, *Phys. Lett.* B **155**(1–2) (1985) 36. http://dx.doi.org/10.1016/0370-2693(85)91028-7

[66] V.A. Rubakov, *Classical Theory of Gauge Fields* (Princeton University Press, Princeton, NJ, 2002).

[67] A.A. Belavin *et al.*, *Phys. Lett.* B **59**(1) (1975) 85. http://dx.doi.org/10.1016/0370-2693(75)90163-X

[68] M.S. Carena, M. Quiros and C.E.M. Wagner, *Phys. Lett.* B **380**(1–2) (1996) 81. http://dx.doi.org/10.1016/0370-2693(96)00475-3

[69] M.S. Carena, *et al.*, *Nucl. Phys.* B **503**(1–2) (1997) 387. http://dx.doi.org/10.1016/S0550-3213(97)00412-4

[70] M.S. Carena *et al.*, *Nucl. Phys.* B **650**(1–2) (2003) 24. http://dx.doi.org/10.1016/S0550-3213(02)01065-9

[71] M. Fukugita and T. Yanagida, *Phys. Lett.* B **174**(1) (1986) 45. http://dx.doi.org/10.1016/0370-2693(86)91126-3

[72] J. Urrestilla *et al.*, *J. Cosmol. Astropart. Phys.* **0807** (2008) 010. http://dx.doi.org/10.1088/1475-7516/2008/07/010

[73] P.A.R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **571** (2014) A1. http://dx.doi.org/10.1051/0004-6361/201321529

[74] E.R. Harrison, *Phys. Rev.* D **1**(10) (1970) 2726. http://dx.doi.org/10.1103/PhysRevD.1.2726

[75] Y.B. Zeldovich, *Mon. Not. R. Astron. Soc.* **160** (1972) 1P. http://dx.doi.org/10.1093/mnras/160.1.1P

[76] P.J.E. Peebles and J.T. Yu, *Astrophys. J.* **162** (1970) 815. http://dx.doi.org/10.1086/150713

[77] A.A. Starobinsky, *Pisma Zh. Eksp. Teor. Fiz.* **30** (1979) 719 [Engl. trans. *JETP Lett.* **30** (1979) 682].

[78] A.A. Starobinsky, *Phys. Lett.* B **91**(1) (1980) 99. http://dx.doi.org/10.1016/0370-2693(80)90670-X

[79] A.H. Guth, *Phys. Rev.* D **23**(2) (1981) 347. http://dx.doi.org/10.1103/PhysRevD.23.347

[80] A.D. Linde, *Phys. Lett.* B **108**(6) (1982) 389. http://dx.doi.org/10.1016/0370-2693(82)91219-9

[81] A. Albrecht and P.J. Steinhardt, *Phys. Rev. Lett.* **48**(17) (1982) 1220. http://dx.doi.org/10.1103/PhysRevLett.48.1220

[82] A.D. Linde, *Phys. Lett.* B **129**(3-4) (1983) 177. http://dx.doi.org/10.1016/0370-2693(83)90837-7

[83] V.F. Mukhanov and G.V. Chibisov, *Pisma Zh. Eksp. Teor. Fiz.* **33** (1981) 549 [Engl. trans. *JETP Lett.* **33** (1981) 532].

[84] S.W. Hawking, *Phys. Lett.* B **115**(4) (1982) 295. http://dx.doi.org/10.1016/0370-2693(82)90373-2

[85] A.A. Starobinsky, *Phys. Lett.* B **117**(3–4) (1982) 175. http://dx.doi.org/10.1016/0370-2693(82)90541-X

[86] A.H. Guth and S.Y. Pi, *Phys. Rev. Lett.* **49**(15) (1982) 1110. http://dx.doi.org/10.1103/PhysRevLett.49.1110

[87] J.M. Bardeen, P.J. Steinhardt and M.S. Turner, *Phys. Rev.* D **28**(4) (1983) 679. http://dx.doi.org/10.1016/0370-2693(82)91219-9

[88] J.M. Maldacena, *J. High Energy Phys.* **0305** (2003) 013. http://dx.doi.org/10.1088/1126-6708/2003/05/013

[89] P.A.R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **571** (2014) A22. http://dx.doi.org/10.1051/0004-6361/201321569

[90] P. Creminelli, A. Nicolis and E. Trincherini, *J. Cosmol. Astropart. Phys.* **1011** (2010) 021. http://dx.doi.org/10.1088/1475-7516/2010/11/021

[91] I. Antoniadis, P.O. Mazur and E. Mottola, *Phys. Rev. Lett.* **79**(1) (1997) 14. http://dx.doi.org/10.1103/PhysRevLett.79.14

[92] V.A. Rubakov, *J. Cosmol. Astropart. Phys.* **0909** (2009) 030. http://dx.doi.org/10.1088/1475-7516/2009/09/030

[93] P. Creminelli, A. Nicolis and E. Trincherini, *J. Cosmol. Astropart. Phys.* **1011** (2010) 021. http://dx.doi.org/10.1088/1475-7516/2010/11/021

[94] K. Hinterbichler and J. Khoury, *J. Cosmol. Astropart. Phys.* **1204** (2012) 023. http://dx.doi.org/10.1088/1475-7516/2012/04/023

[95] V.A. Rubakov, M.V. Sazhin and A.V. Veryaskin, *Phys. Lett.* B **115**(3) (1982) 189. http://dx.doi.org/10.1016/0370-2693(82)90641-4

[96] R. Fabbri and M.D. Pollock, *Phys. Lett.* B **125**(6) (1983) 445. http://dx.doi.org/10.1016/0370-2693(83)91322-9

[97] L.F. Abbott and M.B. Wise, *Nucl. Phys.* B **244**(2) (1984) 541. http://dx.doi.org/10.1016/0550-3213(84)90329-8

[98] A.A. Starobinsky, *Sov. Astron. Lett.* **11** (1985) 133.

[99] M. Kamionkowski, A. Kosowsky and A. Stebbins, *Phys. Rev. Lett.* **78**(11) (1997) 2058. http://dx.doi.org/10.1103/PhysRevLett.78.2058

[100] U. Seljak and M. Zaldarriaga, *Phys. Rev. Lett.* **78**(11) (1997) 2054. http://dx.doi.org/10.1103/PhysRevLett.78.2054

[101] P.A.R. Ade *et al.* (BICEP2 Collaboration), *Phys. Rev. Lett.* **112**(24) (2014) 241101. http://dx.doi.org/10.1103/PhysRevLett.112.241101

[102] P.A.R. Ade *et al.* (BICEP2 and Planck Collaborations), *Phys. Rev. Lett.* **114**(10) (2015) 101301. http://dx.doi.org/10.1103/PhysRevLett.114.101301

[103] M.A. Watanabe, S. Kanno and J. Soda, *Phys. Rev. Lett.* **102**(19) (2009) 191302. http://dx.doi.org/10.1103/PhysRevLett.102.191302

[104] T.R. Dulaney and M.I. Gresham, *Phys. Rev.* D **81**(10) (2010) 103532. http://dx.doi.org/10.1103/PhysRevD.81.103532

[105] A.E. Gumrukcuoglu, B. Himmetoglu and M. Peloso, *Phys. Rev.* D **81**(6) (2010) 063528. http://dx.doi.org/10.1103/PhysRevD.81.063528

[106] M. Libanov and V. Rubakov, *J. Cosmol. Astropart. Phys.* **1011** (2010) 045. http://dx.doi.org/10.1088/1475-7516/2010/11/045

[107] M. Libanov, S. Ramazanov and V. Rubakov, *J. Cosmol. Astropart. Phys.* **1106** (2011) 010. http://dx.doi.org/10.1088/1475-7516/2011/06/010

[108] L. Ackerman, S.M. Carroll and M.B. Wise, *Phys. Rev.* D **75** (2007) 083502 [Erratum *Phys. Rev.* D **80** (2009) 069901]. http://dx.doi.org/10.1103/PhysRevD.75.083502

[109] A.R. Pullen and M. Kamionkowski, *Phys. Rev.* D **76**(10) (2007) 103529. http://dx.doi.org/10.1103/PhysRevD.76.103529

[110] J. Kim and E. Komatsu, *Phys. Rev.* D **88**(10) (2013) 101301. http://dx.doi.org/10.1103/PhysRevD.88.101301

[111] G.I. Rubtsov and S.R. Ramazanov, *Phys. Rev.* D **91**(4) (2015) 043514. http://dx.doi.org/10.1103/PhysRevD.91.043514