Cosmology and dark matter

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

Abstract

Cosmology and astroparticle physics give the strongest possible evidence for the incompleteness of the Standard Model of particle physics. Leaving aside the mysterious dark energy, which may or may not be just the cosmological constant, two properties of the Universe cannot be explained by the Standard Model: dark matter and matter-antimatter asymmetry. Dark matter particles may well be discovered in foreseeable future; this issue is under intense experimental investigation. Theoretical hypotheses on the nature of the dark matter particles are numerous, so we concentrate on several well motivated candidates, such as weakly interacting massive particles, axions and sterile neutrinos, and also give examples of less motivated and more elusive candidates such as fuzzy dark matter. This gives an idea of the spectrum of conceivable dark matter candidates, while certainly not exhausting it. We then consider the matter-antimatter asymmetry and discuss whether it may result from physics at 100 GeV–TeV scale. Finally, we turn to the earliest epoch of the cosmological evolution. Although the latter topic does not appear immediately related to contemporary particle physics, it is of great interest due to its fundamental nature. We emphasize that the cosmological data, notably, on cosmic microwave background anisotropies, unequivocally show that the well understood hot stage was not the earliest one. The best guess for the earlier stage is inflation, which is consistent with everything known to date; however, there are alternative scenarios. We discuss the ways to study the earliest epoch, with emphasis on future cosmological observations.

Keywords

Cosmology; Dark matter; Dark energy; Axions; Baryon asymmetry; Lectures.

1 Introduction

It is a commonplace by now that cosmology and astroparticle physics, on the one side, and particle physics, on the other, are deeply interrelated. Indeed, the gross properties of the Universe—the existence of dark matter and the very presence of conventional, baryonic matter—call for the extension of the Standard Model of particle physics. A fascinating possibility is that the physics behind these phenomena is within reach of current or future terrestrial experiments. The experimental programs in these directions are currently intensely pursued.

Another aspect of cosmology, which currently does not appear directly related to terrestrial particle physics experiments, is the earliest epoch of the evolution of the Universe. On the one hand, there is no doubt that the usual hot epoch was preceded by another, much less conventional stage. This knowledge comes from the study of inhomogeneities in the Universe through the measurements of cosmic microwave background (CMB) anisotropies, as well as matter distribution (galaxies, clusters of galaxies, voids) in the present and recent Universe. On the the other hand, we know only rather general properties of the cosmological perturbations, which, we are convinced, were generated before the hot epoch. For this reason, we cannot be sure about the earliest epoch; the best guess is inflation, but alternatives to inflation have not yet been ruled out. It is conceivable that future cosmological observations will be

able to disentangle between different hypotheses; it is amazing that the study of the Universe at large will possibly reveal the properties of the very early epoch characterized by enormous energy density and evolution rate.

Cosmology and astroparticle physics is a large area of research, so we will be unable to cover it to any level of completeness. On the dark matter side, the number of proposals for dark matter objects invented by theorists in more than 30 years is enormous, so we do not attempt even to list them. Instead, we concentrate on a few hypotheses which may or may not have to do with reality. Namely, we study reasonably well motivated candidates—weakly interacting massive particles (WIMPs), axions, sterile neutrinos—and also discuss more exotic possibilities. On the baryon asymmetry side, we focus on scenarios for its generation which employ physics accessible by terrestrial experiments. A particular mechanism of this sort is the electroweak baryogenesis. The last part of these lectures deals with the earliest cosmology—inflation and its alternatives.

To end up this introduction, we point out that most of the topics we discuss are studied, in one or another way, in books [1]. There are of course numerous reviews, some of which will be referred to in appropriate places.

2 Homogeneous and isotropic Universe

2.1 Friedmann–Lemaître–Robertson–Walker metric

When talking about the Universe, we will always mean its visible part. The visible part is, almost for sure, a small, and maybe even tiny patch of a huge space; for the time being (at least) we cannot tell what is outside the part we observe. At large scales the (visible part of the) Universe is *homogeneous and isotropic*: all regions of the Universe are the same, and no direction is preferred. Homogeneous and isotropic three-dimensional spaces can be of three types. These are three-sphere, flat (Euclidean) space and three-hyperboloid.

A basic property of our Universe is that it expands: the space stretches out. This is encoded in the space-time metric (Friedmann–Lemaître–Robertson–Walker, FLRW)

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

where $d\mathbf{x}^2$ is the distance on unit three-sphere or Euclidean space or hyperboloid, $a(t)$ is the scale factor. Observationally, the three-dimensional space is Euclidean (flat) to good approximation (see, however, Ref. [2] where it is claimed that Planck lensing data prefers a closed Universe), so we will treat $d\mathbf{x}^2 = \delta_{ij}dx^i dx^j$, $i, j = 1, 2, 3$, as line interval in three-dimensional Euclidean space.

The coordinates $\mathbf{x}$ are comoving. This means that they label positions of free, static particles in space (one has to check that world lines of free static particles obey $\mathbf{x}$ = const; this is indeed the case). As an example, distant galaxies stay at fixed $\mathbf{x}$ (modulo peculiar motions, if any). In our expanding Universe, the scale factor $a(t)$ increases in time, so the distance between free masses of fixed spatial coordinates $\mathbf{x}$ grows, $dl^2 = a^2(t)d\mathbf{x}^2$. The galaxies run away from each other.

Since the space stretches out, so does the wavelength of a photon; the photon experiences redshift. If the wavelength at emission (say, by distant star) is λ_e, then the wavelength we measure is

$$\lambda_0 = (1 + z)\lambda_e \,, \quad \text{where} \quad z = \frac{a(t_0)}{a(t_e)} - 1 \,.$$

Here t_e is the time at emission, and z is redshift. Hereafter we denote by subscript 0 the quantities measured at the present time. We sometimes set $a_0 \equiv a(t_0) = 1$ and put ourselves at the origin of coordinate frame, then $|\mathbf{x}|$ is the *present* distance to a point with coordinates $\mathbf{x}$. We also call this the *comoving distance*.

Clearly, the further from us is the source, the longer it takes for light, seen by us today, to travel, i.e., the larger $t_0 - t_e$. High redshift sources are far away from us both in space and in time. For not so

distant sources, we have $t_0 - t_{\mathrm{e}} = r$, where r is the physical distance to the source[1]. For $z \ll 1$ we thus have the Hubble law,

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

$H_0 \equiv H(t_0)$ is the Hubble constant, i.e., the present value of the Hubble parameter

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

The value of the Hubble constant is a subject of some controversy. While the redshift of an object can be measured with high precision (λ_e is the wavelength of a photon emitted by an excited atom; one identifies a series of emission lines, thus determining λ_e, and measures their actual wavelengths λ_0, both with spectroscopic precision; absorption lines are used as well), absolute distances to astrophysical sources have considerable systematic uncertainty. The precise value of H_0 will not be important for our semi-quantitative discussions; we quote here the value found by the Planck collaboration [3],

$$H_0 = (67.7 \pm 0.4) \, \frac{\mathrm{km/s}}{\mathrm{Mpc}} \approx (14.4 \cdot 10^9 \, \mathrm{yrs})^{-1} \ . \tag{3}$$

Here Mpc is the length unit used in cosmology and astrophysics,

$$1 \, \mathrm{Mpc} \approx 3 \cdot 10^6 \, \mathrm{light \ years} \approx 3 \cdot 10^{24} \, \mathrm{cm} \ .$$

The funny unit used in the first expression in Eq. (3) has to do with (somewhat misleading) interpretation of redshift as Doppler effect: galaxies run away from us at velocity $v = z$. To account for uncertainties in H_0 one writes for the present value of the Hubble parameter

$$H_0 = h \cdot 100 \, \frac{\mathrm{km/s}}{\mathrm{Mpc}} \ . \tag{4}$$

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

Concerning length scales characteristic of various objects, we quote the following:

- sizes of visible parts of dwarf galaxies are of order 1 kpc and even smaller;
- sizes of visible parts of galaxies like ours are of order 10 kpc;
- dark halos of galaxies extend to distances of order 100 kpc and larger;
- clusters of galaxies have sizes of order $1 - 3$ Mpc;
- the homogeneity scale[2] today is of order 200 Mpc;
- the size of the visible Universe is 14 Gpc.

2.2 Cosmic microwave background

One of the fundamental discoveries of 1960's was cosmic microwave background (CMB). These are photons with black-body spectrum of temperature

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

Measurements of this spectrum are quite precise and show no deviation from the Planck spectrum (although some deviations are predicted, see Ref. [4] for review). The energy density of CMB photons is given by the Stefan–Boltzmann formula

$$\rho_{\gamma,0} = \frac{\pi^2}{15} T_0^4 = 2.7 \cdot 10^{-10} \, \frac{\mathrm{GeV}}{\mathrm{cm}^3} \ . \tag{6}$$

---

[1] Hereafter we use the natural units, with the speed of light, Planck and Boltzmann constants equal to 1, $c = \hbar = k_{\mathrm{B}} = 1$. Then Newton's gravity constant is $G = M_{\mathrm{Pl}}^{-2}$, where $M_{\mathrm{Pl}} = 1.2 \cdot 10^{19}$ GeV is the Planck mass.

[2] Regions of this size and larger look all the same, while smaller regions differ from each other; they contain different numbers of galaxies.

while the number density of CMB photons is $n_{\gamma,0} = 410 \; \text{cm}^{-3}$.

The discovery of CMB has shown that the Universe was hot at early times, and cooled down due to expansion. As we pointed out, the wavelength of a photon increases in time as $a(t)$, so the energies and hence temperature of photons scale as

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

Importantly, the energy density of CMB photons scales as

$$\rho_\gamma \propto T^4 \propto a^{-4} \,.$$

This is in contrast with the scaling of energy density (mass density) of non-relativistic particles (baryons, dark matter)

$$\rho_\text{M} \propto a^{-3} \,,$$

which is obtained by simply noting that the mass in comoving volume remains constant.

2.3 Friedmann equation

The expansion of the spatially flat Universe is governed by the Friedmann equation,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3 M_\text{Pl}^2} \rho \,, \tag{7}$$

where ρ is the *total* energy density in the Universe. This is nothing but the (00)-component of the Einstein equations of General Relativity, $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu}$, specified to spatially flat FLRW metric and homogeneous and isotropic matter.

One conventionally defines the parameter (critical density),

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

It is equal to the sum of all forms of energy density in the *present* Universe.

2.4 Present composition of the Universe

The *present* composition of the Universe is characterized by the parameters

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

where λ labels various forms of energy: relativistic matter ($\lambda = rad$), non-relativistic matter ($\lambda = M$), dark energy ($\lambda = \Lambda$). Clearly, Eq. (7) gives

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

Let us quote the numerical values:

$$\Omega_\text{rad} = 8.6 \cdot 10^{-5} \,, \tag{9a}$$
$$\Omega_\text{M} = 0.31 \,, \tag{9b}$$
$$\Omega_\Lambda = 0.69 \,. \tag{9c}$$

The value in Eq. (9a) 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 early Universe, since the energy density of relativistic neutrinos scales in the same way as that of photons,

$$\rho_\nu \propto T^4 \propto a^{-4} ,$$

and at temperatures above neutrino masses (but below 1 MeV) we have

$$\rho_\nu == \Omega_\nu \rho_c \left(\frac{a_0}{a}\right)^4 .$$

Non-relativistic matter consists of baryons and dark matter. Their contributions are [3]

$$\Omega_B = 0.049 , \tag{10a}$$
$$\Omega_{DM} = 0.26 . \tag{10b}$$

As we pointed out above, energy densities of various species evolve as follows:

– radiation (photons and neutrinos at temperatures above neutrino mass):

$$\rho_{rad}(t) = \left(\frac{a_0}{a(t)}\right)^4 \rho_{rad,0} = (1+z)^4 \, \Omega_{rad}\rho_c . \tag{11}$$

– Non-relativistic matter:

$$\rho_M(t) = \left(\frac{a_0}{a(t)}\right)^3 \rho_{M,0} = (1+z)^3 \, \Omega_M\rho_c . \tag{12}$$

– The dark energy density does not change in time, or changes very slowly. In what follows we take it constant in time,

$$\rho_\Lambda = \Omega_\Lambda \rho_c = \text{const} . \tag{13}$$

This assumption is not at all innocent. It means that dark energy is assumed to be a cosmological constant. However, even slight dependence of ρ_Λ on time would mean that we are dealing with something different from the cosmological constant. In that case the dark energy density would be associated with some field; there are various theoretical proposals concerning the properties of such a field. Present data is consistent with time-independent ρ_Λ, but the precision of this statement is not yet very high. It is extremely important to study the time-(in)dependence of ρ_Λ with high precision; several experiments are aimed at that.

2.5   Cosmological epochs
The Friedmann equation (7) is now written as

$$H^2(t) = \frac{8\pi}{3M_{Pl}^2}[\rho_\Lambda + \rho_M(t) + \rho_{rad}(t)]$$
$$= H_0^2 \left[\Omega_\Lambda + \Omega_M \left(\frac{a_0}{a(t)}\right)^3 + \Omega_{rad} \left(\frac{a_0}{a(t)}\right)^4\right]$$

This shows that the dominant term in the right hand side at early times (small $a(t)$) was ρ_{rad}, i.e., the expansion was dominated by ultrarelativistic particles (radiation). This is radiation domination epoch.

133

Then the term ρ_M took over, and matter dominated epoch began. The redshift at radiation–matter equality, when the energy densities of radiation and matter were equal, is

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

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

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

The present Universe is at the end of the transition from matter domination to Λ-domination: the dark energy density ρ_Λ will completely dominate over non-relativistic matter in future.

So, we have the following sequence of the regimes of evolution:

$$\cdots \Longrightarrow \text{Radiation domination} \Longrightarrow \text{Matter domination} \Longrightarrow \Lambda\text{–domination} \,. \tag{14}$$

Dots here denote some cosmological epoch preceding the hot stage. We discuss this point later on.

2.6 Radiation domination

2.6.1 Expansion law

The evolution of the scale factor at radiation domination is obtained by using $\rho_\text{rad} \propto a^{-4}$ in the Friedmann equation (7):

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

This gives

$$a(t) = \text{const} \cdot \sqrt{t} \,. \tag{15}$$

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

There are several properties that immediately follow from the result Eq. (15). First, the expansion *decelerates*:

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

Second, time $t = 0$ is the Big Bang singularity (assuming, for the sake of argument, that the Universe starts right from radiation domination epoch). The expansion rate

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

diverges as $t \to 0$, and so does the energy density $\rho(t) \propto H^2(t)$ and temperature $T \propto \rho^{1/4}$. This is "classical" singularity (singularity in classical General Relativity) which, one expects, is resolved in one or another way in complete quantum gravity theory. One usually assumes (although this is not necessarily correct) that the classical expansion begins just after the Planck epoch, when $\rho \sim M_\text{Pl}^4$, $H \sim M_\text{Pl}$, etc.

2.6.2 Particle horizon

The third observation has to do with the causal structure of space-time in the Hot Big Bang Theory (theory that assumes that the evolution starts from the singularity directly into radiation domination—no dots in Eq. (14)). Consider signals emitted right after the Big Bang singularity and travelling at the speed of light. The light cone obeys $ds = 0$, and hence $a(t)dx = dt$. So, the coordinate distance that a signal travels from the Big Bang to time t is

$$x = \int_0^t \frac{dt}{a(t)} \equiv \eta \,. \tag{16}$$

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

$$l_{\text{H}}(t) = a(t)x = a(t) \int_0^t \frac{dt}{a(t)} \ . \tag{17}$$

This physical distance is finite; it is the size of a causally connected region at time t. It is called the horizon size (more precisely, the size of particle horizon). In other words, an observer at time t can have information only on the part of the Universe whose physical size at that time is $l_{\text{H}}(t)$. At radiation domination, one has

$$l_{\text{H}}(t) = 2t \ .$$

Note that this horizon size is of the order of the Hubble size,

$$l_{\text{H}}(t) \sim H^{-1}(t) \ . \tag{18}$$

The notion of horizon is straightforwardly extended to the matter dominated epoch and to the present time: relation Eq. (17) is of general nature, while the scale factor $a(t)$ has to be calculated anew. To give an idea of numbers, the horizon size at the present epoch is

$$l_{\text{H}}(t_0) \approx 14 \ \text{Gpc} \simeq 4 \cdot 10^{28} \ \text{cm} \ .$$

2.6.3 Energy density

At radiation domination, cosmic plasma is almost always in thermal equilibrium, and interactions between particles are almost always weak. So, the plasma properties are determined by thermodynamics of a gas of free relativistic particles. At different times, the number of relativistic species that contribute to the energy density is different. As an example, at temperatures above 1 MeV, but below 100 MeV, relativistic are photons, three types of neutrinos, electrons and positrons; while at temperatures of about 200 GeV all Standard Model particles are relativistic. In most cases, one can neglect chemical potentials, i.e., consider cosmic plasma symmetric under interchange of particles with antiparticles (chemical potential of photons is zero, since photons can be created in processes like $e^- e^- \rightarrow e^- e^- \gamma$; since particle and its antiparticle can annihilate into photons, e.g., $e^+ e^- \rightarrow \gamma\gamma$, chemical potentials of particles and antiparticles are equal in modulus and opposite in sign, e.g., $\mu_{e^+} = -\mu_{e^-}$; in symmetric plasma $\mu_{e^+} = -\mu_{e^-} = 0$). Then the Stefan–Boltzmann law gives for the energy density

$$\rho_{\text{rad}} = \frac{\pi^2}{30} g_* T^4 \ , \tag{19}$$

where g_* is the effective number of degrees of freedom,

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

g_i is the number of spin states of a particle i, the factor 7/8 is due to Fermi-statistics. The parameter g_* depends on temperature, and hence on time: as the temperature decreases below the mass of a particle, this particle drops out from the sum here. The formula (19) enables one to write the Friedmann equation (7) as

$$H = \frac{T^2}{M_{\text{Pl}}^*} \ , \qquad M_{\text{Pl}}^* = \frac{M_{\text{Pl}}}{1.66\sqrt{g_*}} \ . \tag{20}$$

We use this simple result in what follows.

2.6.4 Entropy

The cosmological expansion is slow, which implies conservation of entropy (modulo quite exotic scenarios with large entropy generation). The entropy density of a free relativistic gas 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},

$$sa^3 = \text{const} \ , \tag{21}$$

while temperature scales *approximately* as a^{-1} (this is because g_* depends on time). We note for future reference that the effective number of degrees of freedom in the Standard Model at $T \gtrsim 100$ GeV is

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

The present entropy density in the Universe, still with the prescription that neutrinos are relativistic, is

$$s_0 \approx 3000 \text{ cm}^{-3} \ . \tag{22}$$

The precise meaning of this number is that at high temperatures (when there is thermal equilibrium), the entropy density is $s(t) = (a_0/a(t))^3 s_0$.

Notion of entropy is convenient, in particular, for characterizing asymmetries which can exist if there are conserved quantum numbers, such as the baryon number after baryogenesis. The density of a conserved number also scales as a^{-3}, so the time independent characteristic of, say, the baryon abundance is the baryon-to-entropy ratio

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

At late times, one can use another parameter, baryon-to-photon ratio

$$\eta_{\rm B} = \frac{n_{\rm B}}{n_\gamma} \ , \tag{23}$$

where n_γ is photon number density. It is related to Δ_B by a numerical factor, but this factor depends on time through g_* and stays constant only after $e^+ e^-$-annihilation, i.e., at $T \lesssim 0.5$ MeV. Numerically,

$$\Delta_{\rm B} = 0.14 \, \eta_{\rm B,0} = 0.86 \cdot 10^{-10} \ . \tag{24}$$

In what follows we discuss the ways to obtain this number from observations.

2.7 Matter domination

At matter domination, we have $\rho \propto a^{-3}$, and the Friedmann equation (7) gives

$$a(t) = \text{const} \cdot t^{2/3}$$

Qualitatively, matter domination is similar to radiation domination: expansion is decelerated, the size of the particle horizon is of order of the Hubble size, $l_{\rm H}(t) \sim H^{-1}(t) \sim t$. An important difference between radiation and matter dominated epochs is that inhomogeneities in energy density ("scalar perturbations") grow rapidly at matter domination and slowly at radiation domination. Thus, matter domination is the epoch of structure formation in the Universe.

2.8 Dark energy domination

The expansion of the Universe is accelerated today. Within General Relativity this is attributed to dark energy. We know very little about this "substance": we know its energy density, Eq. (9c), and also know that this energy density changes in time very slowly, if at all. The latter fact is quantified in the following way. Let us denote by p the effective pressure, i.e., spatial component of the energy-momentum tensor in locally-Lorentz frame $T_{\mu\nu} = \mathrm{diag}(\rho, p, p, p)$. Then covariant conservation of the energy-momentum in an expanding Universe gives for any fraction that does not interact with other fractions

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

(note that relativistic and non-relativistic matter have $p = \rho/3$ and $p = 0$, respectively, so this equation gives for them $\rho \propto a^{-4}$ and $\rho \propto a^{-3}$, as it should). A simple parametrization of time-dependent dark energy is $p_\Lambda = w_\Lambda \rho_\Lambda$ with time-independent w_Λ. The combination of cosmological data gives [3]

$$w_\Lambda \approx -1.03 \pm 0.03 . \tag{25}$$

Thus, with reasonable precision one has $p_\Lambda = -\rho_\Lambda$, which corresponds to a time-independent dark energy density.

The solution to the Friedmann equation (7) with constant $\rho = \rho_\Lambda$ is

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

where $H_\Lambda = (8\pi\rho_\Lambda/3M_{\mathrm{Pl}}^2)^{1/2} = \mathrm{const}$. This gives an accelerated expansion, $\ddot{a} > 0$, unlike at radiation or matter domination. The transition from decelerated (matter dominated) to accelerated expansion (dark energy dominated) has been confirmed quite some time ago by combined observational data, see Fig. 1, which shows the dependence on redshift of the quantity $H(z)/(1+z) = \dot{a}(t)/a_0$.
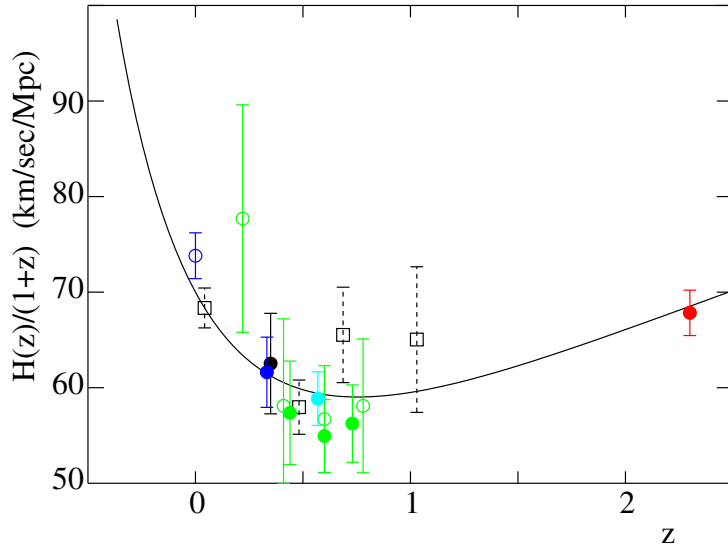

Fig. 1: Observational data on the time derivative of the scale factor as function of redshift z [5]. The change of the behavior from decreasing to increasing, as z decreases, means the change from decelerated to accelerated expansion. The theoretical curve corresponds to spatially flat Universe with $h = 0.7$ and $\Omega_\Lambda = 0.73$.

In the case of the cosmological constant, the energy-momentum tensor is proportional to the metric, and in a locally-Lorentz frame it reads

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

137

where $\eta_{\mu\nu}$ is the Minkowski tensor. Hence $w_\Lambda = -1$. One can view this as the characteristic of the 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,

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

In fact, there are two hard problems. One is that 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. We are not going to discuss these points anymore, and only emphasize that we are not aware of a compelling mechanism that solves any of the two cosmological constant problems (with possible exception of anthropic argument by Weinberg and Linde [6,7]).

3   Cornerstones of thermal history

3.1   Recombination = photon last scattering

Going back in time, we reach so high temperatures that the usual matter (electrons and protons with rather small admixture of light nuclei, mainly ^{4}He) is in the plasma phase. In plasma, photons interact with electrons due to the Thomson scattering and protons have Coulomb interaction with electrons. These interactions are strong enough to keep photons, electrons and protons in thermal equilibrium. When the temperature drops to

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

almost all electrons "recombine" with protons into neutral hydrogen atoms (helium recombined earlier). The number density of atoms at that time is quite small, $250\ \mathrm{cm}^{-3}$, so from that time on, the Universe is transparent to photons[3]. Thus, T_{rec} is the *photon last scattering temperature*. At that time the age of the Universe is $t_{\mathrm{rec}} \approx 380$ thousand years (for comparison, its present age is about 13.8 billion years).

CMB photons give us (literally!) the photographic picture of the Universe at the photon last scattering epoch. The last scattering epoch lasted considerably shorter than the then Hubble time $H^{-1}(t_{\mathrm{rec}}) \sim t_{\mathrm{rec}}$; to a meaningful (although rather crude) approximation, recombination occurred instantaneously. This is important, since in the opposite case of long recombination, the photographic picture would be strongly washed out.

This photographic picture is shown in Fig. 2. Here brighter (darker) regions correspond to higher (lower) temperatures. The relative temperature fluctuation is of the order of $\delta T/T = 10^{-4} - 10^{-5}$, so the 380 thousand year old Universe was much more homogeneous than today.

One performs Fourier decomposition of the temperature fluctuations, i.e., decomposition in spherical harmonics:

$$\frac{\delta T}{T}(\theta, \varphi) = \sum_{l,m} a_{lm} Y_{lm}(\theta, \varphi)\,.$$

Here a_{lm} are independent Gaussian random variables (no non-Gaussianities have been found so far) with $\langle a_{lm} a_{l'm'}^* \rangle \propto \delta_{ll'} \delta_{mm'}$ and $\langle a_{lm}^* a_{lm} \rangle = C_l$. The multipoles C_l, or, equivalently,

$$D_l = \frac{l(l+1)}{2\pi} C_l$$

are the main quantities of interest. The larger l, the smaller the angular scales, hence the shorter the wavelengths of density perturbations producing the temperature anisotropy.

---

[3]Modulo effects of re-ionization that occurred much later and affected a small fraction of CMB photons.
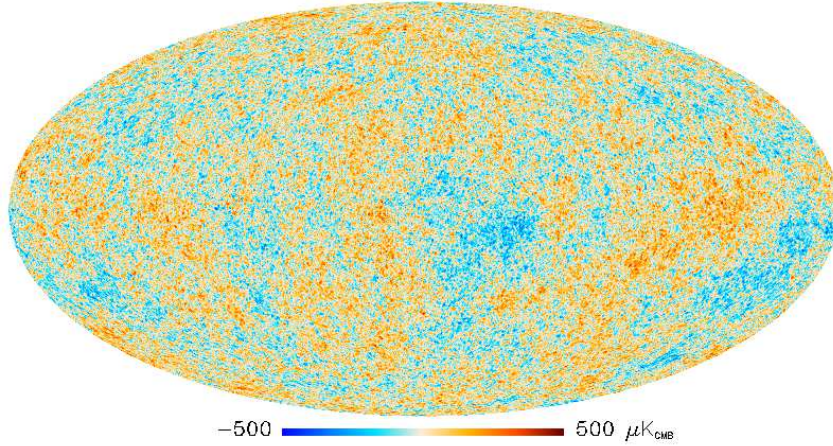
Fig. 2: CMB sky as seen by Planck.

It is worth noting that averaging here is understood in terms of an ensemble of Universes, while we have just one Universe. So, there is inevitable uncertainty in C_l, called cosmic variance. For given l, one has $(2l + 1)$ quantities a_{lm}, $m = 0, \pm 1, \ldots, \pm l$, so the uncertainty is $\Delta C_l / C_l \simeq 1/\sqrt{2l}$.

CMB temperature multipoles are shown in Fig. 3 (error bars there are due to cosmic variance, not the measurement errors). Also measured are CMB polarization multipoles and temperature-polarization cross-correlation multipoles. There is a lot of physics behind these quantities, which has to do with:

– primordial perturbations: the perturbations that are built in already at the beginning of the hot cosmological epoch, see Section 11;

– development of sound waves in the cosmic plasma from the early hot stage to recombination; gravitational potentials due to dark matter at recombination (which are sensitive to the composition of cosmic medium);

– propagation of photons after recombination (which is sensitive to expansion history of the Universe and structure formation).

Clearly, CMB measurements are a major source of the cosmological information. We come back to CMB in due course.

3.2 Big Bang nucleosynthesis

As we go back further in time, we arrive at a temperature in the Universe in the MeV range. The epoch characterized by temperatures 1 MeV–30 keV is the epoch of Big Bang nucleosynthesis. That epoch starts at a temperature of 1 MeV, when the age of the Universe is 1 s. At temperatures above 1 MeV, there are rapid weak processes like

$$\text{e}^- + \text{p} \longleftrightarrow \text{n} + \nu_e \,. \tag{26}$$

These processes keep neutrons and protons in chemical equilibrium; the ratio of their number densities is determined by the Boltzmann factor, $n_\text{n}/n_\text{p} = \exp\left(-\frac{m_\text{n} - m_\text{p}}{T}\right)$. At $T_\text{n} \approx 1$ MeV neutron-proton transitions in Eq. (26) switch off, and neutron-proton ratio is frozen out at the value

$$\frac{n_\text{e}}{n_\text{p}} = \text{e}^{-\frac{m_\text{n} - m_\text{p}}{T_\text{n}}} \,.$$
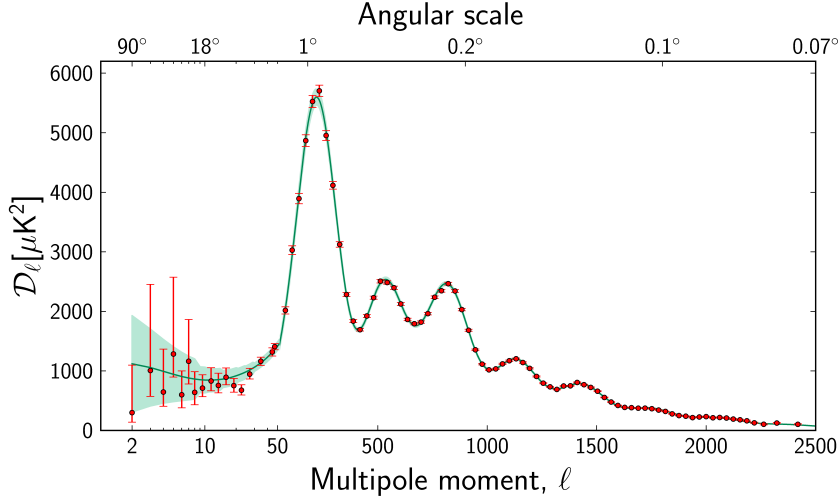
139

Fig. 3: Multipoles D_l as measured by Planck.

Interestingly, $m_n - m_p \sim T_n$, so the neutron-proton ratio at neutron freeze-out and later was neither equal to 1, nor very small. If it were equal to 1, protons would in the end combine with neutrons into ^{4}He, and there would remain no hydrogen in the Universe. On the other hand, for very small n_n/n_p, too few light nuclei would be formed, and we would not have any observable remnants of the Big Bang nucleosynthesis (BBN) epoch. In either case the Universe would be quite different from what it actually is. It is worth noting that the approximate relation $m_n - m_p \sim T_n$ is a coincidence: $m_n - m_p$ is determined by light quark masses and electromagnetic coupling, while T_n is determined by the strength of weak interactions (the rates of the processes in Eq. (26)) and gravity (the expansion of the Universe). This is one of numerous coincidences we encounter in cosmology.

At temperatures 100–30 keV, neutrons combined with protons into light elements in thermonuclear reactions

$$
\begin{aligned}
p + n &\rightarrow {}^2\text{H} + \gamma \,, \\
{}^2\text{H} + p &\rightarrow {}^3\text{He} + \gamma \,, \\
{}^3\text{He} + {}^2\text{H} &\rightarrow {}^4\text{He} + p \,,
\end{aligned}
\tag{27}
$$

etc., up to ^{7}Li. The abundances of light elements have been measured, see Fig. 4. The only parameter relevant for calculating these abundances (assuming negligible neutrino-antineutrino asymmetry) is the baryon-to-photon ratio $\eta_B \equiv \eta$, see Eq. (23), which determines the number density of baryons. Comparison of the Big Bang nucleosynthesis theory with the observational determination of the composition of cosmic medium enables one to determine η_B and check the overall consistency of the BBN picture. It is even more reassuring that a completely independent measurement of η_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.

3.3 Neutrino decoupling

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

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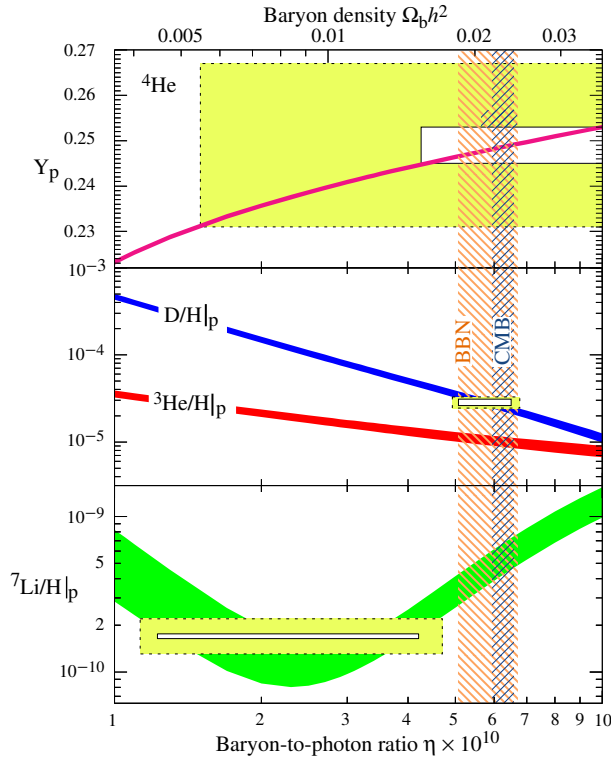\nu_\alpha + \bar{\nu}_\alpha \longleftrightarrow e^+ + e^-
$$

Fig. 4: Abundances of light elements, measured (boxes; larger boxes include systematic uncertainties) and calculated as functions of baryon-to-photon ratio η [8]. The determination of $\eta \equiv \eta_B$ from BBN (vertical range marked BBN) is in excellent agreement with the determination from the analysis of CMB temperature fluctuations (vertical range marked CMB).

and crossing processes. Here the subscript α labels neutrino flavors. These processes switch off at $T \sim 2$–3 MeV, depending on the neutrino flavor. Since then neutrinos do not interact with the cosmic medium other than gravitationally, but they do affect the properties of the CMB and the distribution of galaxies through their gravitational interactions. Thus, observational data can be used to establish, albeit somewhat indirectly, the existence of relic neutrinos and set limits on neutrino masses. We quote here the limit reported by the Planck collaboration [3]

$$\sum m_\nu < 0.12 \text{ eV} \,,$$

where the sum runs over the three neutrino species. Other analyses give somewhat weaker limits. Also, the data can be used to determine the effective number of neutrino species that counts the number of relativistic degrees of freedom [3]:

$$N_{\nu,\,\text{eff}} = 2.99 \pm 0.17 \,,$$

which is consistent with the Standard Model value $N_\nu = 3$. We see that cosmology *requires* relic neutrinos.

4 Dark matter: evidence

Unlike dark energy, dark matter experiences the same gravitational force as the baryonic matter. Dark matter is discussed in numerous reviews, see, e.g., Refs. [9–12]. 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. Dark matter is characterized by the mass-to-entropy ratio,

$$\left(\frac{\rho_{\mathrm{DM}}}{s}\right)_0 = \frac{\Omega_{\mathrm{DM}}\rho_c}{s_0} \approx \frac{0.26 \cdot 5 \cdot 10^{-6}\ \mathrm{GeV} \cdot \mathrm{cm}^{-3}}{3000\ \mathrm{cm}^{-3}} = 4 \cdot 10^{-10}\ \mathrm{GeV}\ . \qquad (28)$$

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

There are various ways of measuring the contribution of non-baryonic dark matter to the total energy density of various objects and the Universe as a whole.

4.1 Dark matter in galaxies

Dark matter exists in galaxies. Its distribution is measured by the observation of rotation velocities of distant stars and gas clouds around a galaxy, Fig. 5. If the mass was concentrated in a luminous central part of a galaxy, the velocities of objects away from the central part would decrease with the distance r to the center as $v \propto r^{-1/2}$—this immediately follows from Newton's second law

$$\frac{v^2}{r} = G\frac{M(r)}{r^2}\ .$$

In reality, 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.

4.2 Dark matter in clusters of galaxies

Dark matter makes most of the mass of the largest gravitationally bound objects—clusters of galaxies. There are various methods to determine the gravitating mass of a cluster, and mass distribution in a
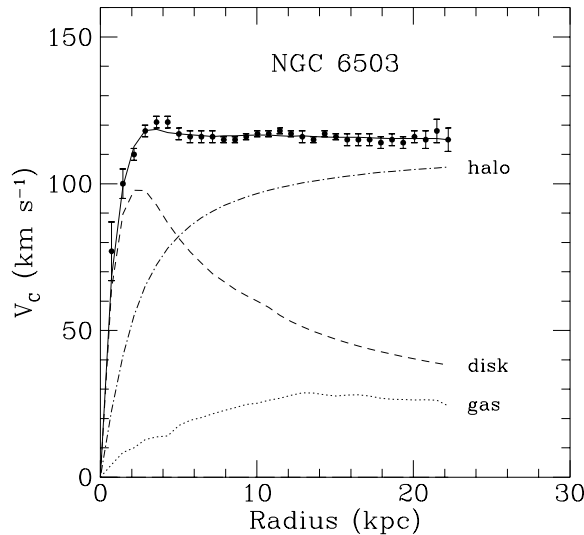

Fig. 5: Rotation velocities of hydrogen gas clouds around a galaxy NGC 6503 [13]. Lines show the contributions of the three main components that produce the gravitational potential. The main contribution at large distances is due to dark matter, labeled "halo".

cluster, which give consistent results. These include measurements of rotational velocities of galaxies in a cluster (original Zwicky argument that goes back to 1930's), measurements of temperature of hot gas (which actually makes most of the baryonic matter in clusters), observations of gravitational lensing of extended light sources (galaxies) behind the cluster, see Fig. 6. All these determinations show that baryons (independently measured through their X-ray emission) make less than 1/4 of total mass in clusters. The rest is dark matter.

Fig. 6: Cluster of galaxies CL0024 + 1654 [14], acting as gravitational lens. Right panel: cluster in visible light. Round yellow spots are galaxies in the cluster. Elongated blue strips are images of one and the same galaxy behind the cluster. Left panel: reconstructed distribution of gravitating mass in the cluster; brighter regions have larger mass density.

Concerning galaxies and clusters of galaxies, we note that there are attempts to attribute the properties of rotation curves and other phenomena, which are usually considered as evidence for dark matter, to a modification of gravity, and in this way get rid of dark matter altogether. There are several strong arguments that rule out this idea. One argument has to do with 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 the dark matter and galaxies. In this way the baryons (which are mainly in hot gas) and dark matter are separated in space. Since the baryonic mass and gravitational potentials are not concentric, one cannot attribute gravitational potentials solely to baryons, even assuming the modification of Newton's gravity law. As a remark, the fact that dark matter moves after cluster a collision considerably faster than baryonic gas means that elastic scattering between dark matter particles is weak. Quantitatively, the limit on the dark matter elastic scattering cross section is

$$\sigma_{\mathrm{DM-DM}}^{(\mathrm{el})} < 1 \cdot 10^{-24} \; \mathrm{cm}^2 \; . \tag{29}$$

This limit is not particularly strong, but it does rule out part of the parameter space of strongly interacting massive particle (SIMP) dark matter models, see Section 5.2.

4.3 Dark matter imprint in CMB

The composition of the Universe strongly affects the CMB angular anisotropy and polarization. Before recombination, the energy density perturbation is a sum of the perturbation in the baryon-electron-photon component and the dark matter component,

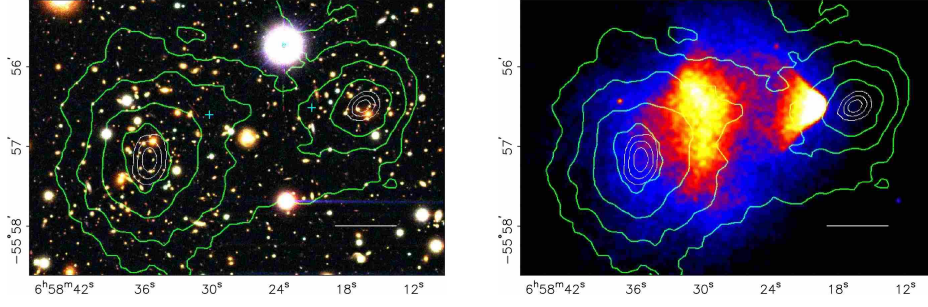$$\delta\rho = \delta\rho_{\mathrm{B}} + \delta\rho_{\mathrm{DM}}$$

143

Fig. 7: Observation [15] of the Bulet 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 in the right panel 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.

(we simplify things here, as there is also perturbation in the gravitational potentials induced by the density perturbation). The tightly coupled baryon-electron-photon plasma has high pressure (due to the photon component with $p_\gamma = \rho_\gamma/3$), so the density perturbations in this fraction undergo acoustic oscillations: every Fourier mode oscillates in time as

$$\delta\rho_{\mathrm{B}}(\mathbf{k}, t) = A(\mathbf{k})\cos\left(\int_0^t v_s \frac{k}{a(t)} dt\right) , \qquad (30)$$

where $\mathbf{k}$ is the comoving momentum (and $\mathbf{k}/a(t)$ is the physical momentum which gets redshifted), $v_s \approx 1/\sqrt{3}$ is the sound speed, and $A(\mathbf{k})$ is the amplitude that varies slowly with k (in statistical sense: $\delta\rho(\mathbf{k})$ is the Gaussian random field). We comment in Section 11 on the fact that the phase of cosine in Eq. (30) is well defined. On the contrary, dark matter is pressureless, so its perturbation is almost independent of time,

$$\delta\rho_{\mathrm{DM}} \approx \delta\rho_{\mathrm{DM}}(\mathbf{k}) ,$$

where $\delta\rho_{\mathrm{DM}}(\mathbf{k})$ slowly varies with k. At recombination time t_r, the energy density perturbation is a sum

$$\delta\rho(\mathbf{k}, t_r) = A(\mathbf{k})\cos\left(\int_0^{t_r} v_s \frac{k}{a(t)} dt\right) + \delta\rho_{\mathrm{DM}}(\mathbf{k}) . \qquad (31)$$

The first term here oscillates *as a function of k*, while the second term is a smooth, non-oscillating function of k.

Now, the behavior of $\delta\rho(t_r)$ as function of the spatial momentum k translates into the behavior of the CMB temperature fluctuation δT as function of the multipole number l. δT at a given point in space at the recombination epoch is proportional to $\delta\rho$ (here we again simplify things, this time quite considerably). We see CMB coming from a *photon last scattering sphere*; a smaller angular scale in this photographic picture corresponds to a smaller spatial scale at the recombination epoch, hence a larger multipole l corresponds to a higher three-momentum k. Thus, the oscillatory formula (31) translates into the oscillatory behavior in Fig. 3. Both, the oscillatory part of the temperature angular spectrum (which is due to the first, baryonic term in Eq. (31)) and the smooth part (due to the second, dark matter term in Eq. (31)), are clearly visible in Fig. 3. The detailed analysis of this angular spectrum enables one to determine the baryon content and the dark matter content in the Universe, Ω_{B} and Ω_{DM} quoted in Eq. (10).

4.4 Dark matter and structure formation

Dark matter is crucial for our existence, for the following reason: As we discussed above, the density perturbations in the baryon-electron-photon plasma before recombination do not grow because of high

pressure; instead, they oscillate with a 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 of the growth is qualitatively 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[4]

$$\frac{a(t_0)}{a(t_{\rm rec})} = 1 + z_{\rm rec} = \frac{T_{\rm rec}}{T_0} \approx 10^3 \ . \tag{32}$$

The initial amplitude of density perturbations is very well known from the CMB anisotropy measurements, $(\delta\rho/\rho)_i = 5 \cdot 10^{-5}$. Hence, a Universe without dark matter would still be nearly 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 the plasma much earlier than baryons, the perturbations in dark matter started to grow much earlier. The corresponding growth factor is larger than Eq. (32), so that the dark matter density contrast at galactic and sub-galactic scales becomes of order one, the perturbations enter a non-linear regime, collapse 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 work 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.

4.5 Digression – Standard ruler: BAO

Before recombination, the sound speed in the baryon-electron-photon component is about $v_s \approx 1/\sqrt{3}$. After recombination, baryons (atoms) decouple from photons, the sound speed in the baryon component is practically zero, and baryons no longer move in space. This leads to a feature in the spatial distribution of matter (galaxies) which is known as Baryon Acoustic Oscillations (BAO). It is worth noting that similar phenomenon was described by A.D. Sakharov [16] back in 1965, but in the context of a cold cosmological model (Sakharov's paper was written before the discovery of CMB).

The physics behind BAO is illustrated in Fig. 8. Suppose there is an overdense region in the very early Universe (in the beginning of the hot epoch). Importantly, the initial conditions for the baryon-electron-photon component and dark matter are the same: overdensity exists in both of them in the same place in space (this is the property of adiabatic scalar perturbations; CMB measurements ensure that primordial perturbations are indeed adiabatic). This initial condition is shown in the left panel of Fig. 8. Before recombination, dark matter perturbation stays in the same place, while the perturbation in baryon-electron-photon component moves away with the sound speed. If the initial perturbation is spherically symmetric, then the sound wave is spherical, as shown in the right panel. At recombination, the baryon perturbation is frozen in, and the whole picture expands merely due to the cosmological expansion. The comoving distance between the dark matter overdensity and the baryon overdensity shell is the comoving sound horizon at recombination

$$r_s = \int_0^{t_r} v_s \frac{k}{a(t)} dt$$

(this is precisely the argument of cosine in Eq. (31)); its present value is $r_s \simeq 150$ Mpc (we set $a_0 = 1$ here), and the value at redshift z is $r_s/(1 + z)$.

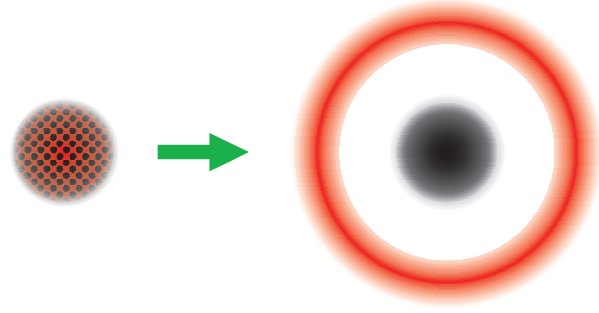[4]Because of the presence of dark energy, the growth factor is even somewhat smaller.

Fig. 8: Schematic picture of Baryon Acoustic Oscillations. Dark regions show dark matter overdensity, less dark (red) regions are the ones with baryon overdensity. Left: initial condition. Right: at recombination and later.

Due to BAO, there is correlation between the matter densities (dark matter plus baryons) separated by the comoving distance r_s. It shows up as a feature in the galaxy-galaxy correlation function $\xi(s)$, where s is the comoving distance. This bump in the correlation function was detected in Ref. [17], see Fig. 9. Clearly, BAO serves as a standard ruler at various redshifts, which can be used to study the evolution of the Universe in a not so distant past.

Currently, BAO is a very powerful tool of observational cosmology. It is used in particular to study time (in)dependence of dark energy.

The bump in the spatial correlation function translates into oscillations in momentum space, hence the name.
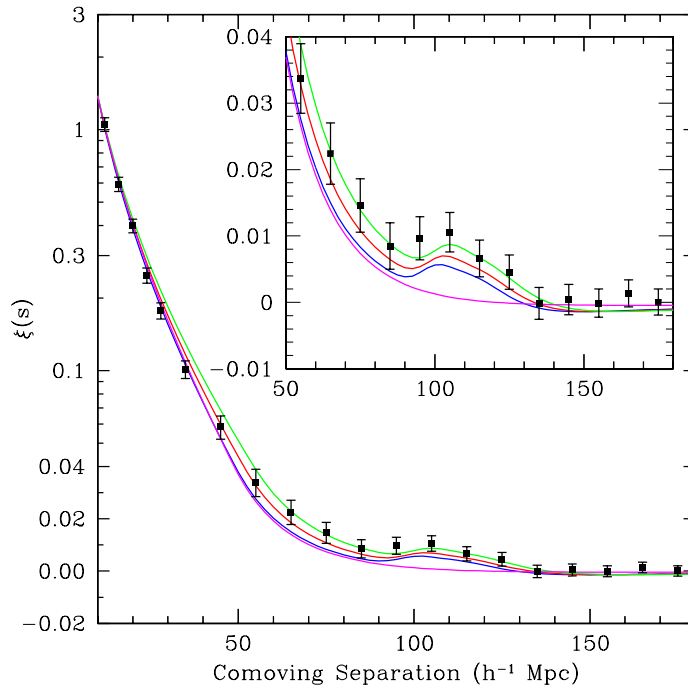
Fig. 9: The first detection of BAO: the correlation function $\xi(s)$ determined by the analysis of the Sloan Digital Sky Survey (SDSS) data on the distribution of distant galaxies. Solid lines show the predictions of various cosmological models. Green, red and blue lines correspond to $\Omega_M h^2 = 0.12, 0.13, 0.14$, respectively, with $\Omega_B h^2 = 0.024$, $n_s = 0.98$ in all cases. The magenta line corresponds to an unrealistic Universe without baryons. The parameter h is defined in Eq. (4); numerically, $h_0 \approx 0.7$.

146

5 Astrophysics: more hints on dark matter properties

Important information on dark matter properties is obtained by theoretical analysis of structure formation and its comparison with observational data. Indeed, as we discussed above, dark matter plays the key role in structure formation, so the properties of galaxies and their distribution in space potentially tell us a lot about dark matter.

Currently, theoretical studies are made mostly via numerical simulations, many of which ignore effects due to baryons (dark-matter-only). Thus, these simulations give the dark matter distribution. To compare it with observed structures, one often assumes that baryons trace dark matter, with the qualification that baryons are capable of losing their kinetic energy and forming more compact structures inside dark matter halos. In other words, a simulated dark matter collapsed clump of mass, characteristic of a galaxy, is associated with a visible galaxy, while heavier dark matter clumps are interpreted as clusters of galaxies, etc.

Currently, the most popular dark matter scenario is cold dark matter, CDM. It consists of particles whose velocities are negligible at all stages of structure formation, and whose non-gravitational interactions with themselves and with baryons are negligible too (from the viewpoint of structure formation). The CDM numerical simulations (plus the above assumption concerning baryons) are in very good agreement with observations *at relatively large spatial scales*. This is an important result that implies interesting limits on dark matter properties, which we discuss below.

However, there are astrophysical phenomena at shorter scales that may or may not hint towards something different from weakly interacting CDM. The situation is inconclusive yet, but it is worth keeping these phenomena in mind, which we now discuss in turn.

5.1 Missing satellite problem: astrophysics vs warm dark matter

It has long been known that CDM-only simulations produce a lot of small mass halos, $M \lesssim 10^9 M_\odot$ where $M_\odot$ is the Solar mass. Galaxies like the Milky Way have masses $(10^{11} - 10^{12}) M_\odot$, so we are talking about dwarf galaxies. As an example, the left panel of Fig. 10 shows the simulated dark matter distribution in a ball of radius 250 kpc around a galaxy similar to the Milky Way. Assuming that baryons trace dark matter, one observes that there must be hundreds of satellite galaxies there. The actual Milky Way satellites are shown in the right panel of Fig. 10; clearly the number of satellites is a lot smaller. This is the missing satellite problem.
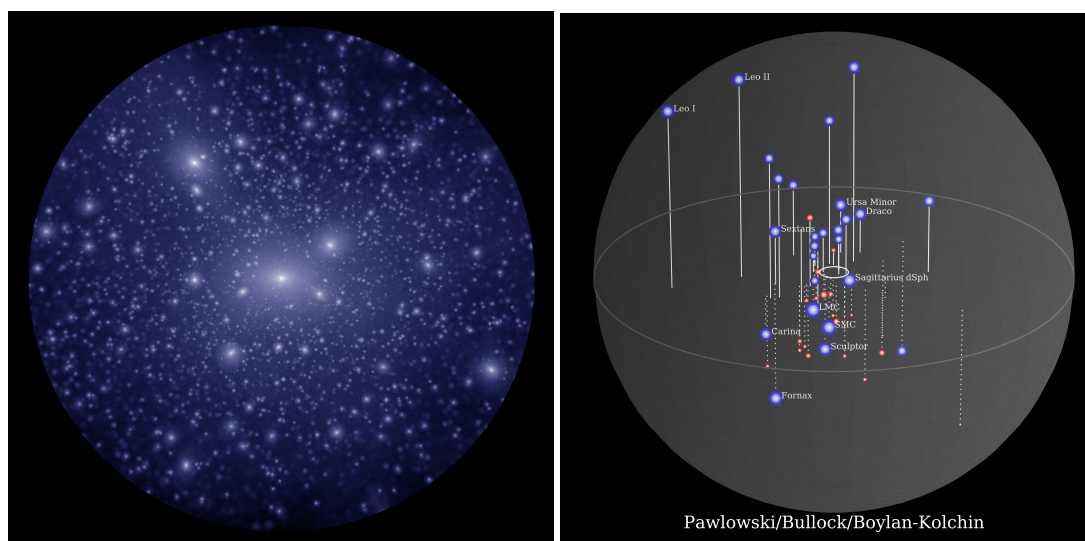

Fig. 10: Left: CDM-only simulation of 250 kpc vicinity of a galaxy like the Milky Way; right: actual distribution of satellite galaxies in 250 kpc vicinity of the Milky Way [12].

It is conceivable that this problem has astrophysical solution within the CDM model. One point is that the number of observed faint satellite galaxies around the Milky Way is not that small any longer: while a few years ago this number was about 20, it is currently about 60, and this is not a complete sample because of a limited detection efficiency—the expectation [18] for a complete sample is 150–300 with masses exceeding $10^8 M_\odot$. Another property is that dark matter halos of mass $M < 10^9 M_\odot$ appear fairly inefficient in forming a luminous component[5]—this has been suggested by simulations that include numerous effects due to baryons [19, 20]. Thus, if the CDM model is correct, and the missing satellite problem has an astrophysical solution, there must be a large number of ultra-faint dwarf galaxies with masses $(10^8 - 10^9) M_\odot$ and even larger number of non-luminous dark matter halos with $M \lesssim 10^8 M_\odot$ in the vicinity of the Milky Way. It will be possible to check this prediction in near future, notably, with the Large Synoptic Survey Telescope, LSST [22].

An alternative, particle physics solution to the missing satellite problem is *warm dark matter*, WDM. A reasonably well motivated WDM candidate is a sterile neutrino, which we discuss in Section 8. Another popular candidate is a light gravitino. In WDM case, dark matter particles decouple from the kinetic equilibrium with the baryon-photon component when they are relativistic. Let us assume for definiteness that they are in *kinetic* equilibrium with the cosmic plasma at a temperature T_f when their number density freezes out (there is no *chemical* equilibrium at $T = T_f$, otherwise the dark matter would be overabundant). After the kinetic equilibrium breaks down at a temperature $T_d \leq T_f$, the spatial momenta decrease as a^{-1}, i.e., the momenta are of the order T all the time after decoupling. When dark matter particles are relativistic, the density perturbations do not grow: relativistic particles escape from the gravitational potentials, so they do not experience the gravitational instability; in fact, the density perturbations, and hence the gravitational wells, get smeared out instead of getting deeper. WDM particles become non-relativistic at $T \sim m$, where m is their mass. Only after that the WDM perturbations start to grow. 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_{nr} when $T \sim m$ is of order

$$l_H(t_{nr}) \simeq H^{-1}(T \sim m) = \frac{M_{Pl}^*}{T^2} \sim \frac{M_{Pl}^*}{m^2} \ .$$

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

$$l_0 = l_H(t_{nr}) \frac{a_0}{a(t_{nr})} \sim l_H(t_{nr}) \frac{T}{T_0} \sim \frac{M_{Pl}}{m T_0} \ , \tag{33}$$

where we neglected the (rather weak) dependence on g_*. Hence, in WDM scenario, the structures of comoving 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.

To solve the missing satellite problem, one requires that the mass of dark matter which was originally distributed over the volume of comoving size l_0, and collapsed later on, is of order of the mass of the satellite galaxy,

$$\frac{4\pi}{3} l_0^3 \, \Omega_{DM} \rho_c \sim M_{dwarf} \ .$$

With $M_{dwarf} \sim 10^8 M_\odot$ we find $l_0 \sim 100$ kpc, and Eq. (33) gives the estimate for the mass of a dark matter particle

$$\text{WDM} \ : \quad m_{DM} = 3 \, - \, 10 \, \text{keV} \ . \tag{34}$$

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

$$\text{CDM} \ : \quad m_{DM} \gtrsim 10 \, \text{keV} \ . \tag{35}$$

---

[5]Another effect, important for satellite galaxies close to the center of the Milky Way, is the tidal force due to the gravitational potential produced by the disk of the host galaxy [21].

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

Reversing the argument, one obtains a limit on the mass of the WDM particle which decouples in the kinetic equilibrium [18],

$$m \gtrsim 4 \, \text{keV} \ . \tag{36}$$

5.1.1 Digression: phase space bound

In fact there are other ways to obtain limits on m. One has to do with the phase space density: the maximum value of the coarse grained phase space density

$$f(p, x)_{\text{coarse grained}} = \left(\frac{dN}{d^3p \, d^3x} \right)_{\text{coarse grained}}$$

does not decrease in the course of the evolution (here N is the number of particles). Indeed, the Liouville theorem tells that the microscopic phase space density is time-independent. What happens in the course of evolution is that particles penetrate initially unoccupied regions of phase space, see Fig. 11. While the maximum value of the microscopic phase space density remains constant in time, the maximum value of the coarse grained phase space density (average over phase space volume shown by dashed line in Fig. 11) decreases.
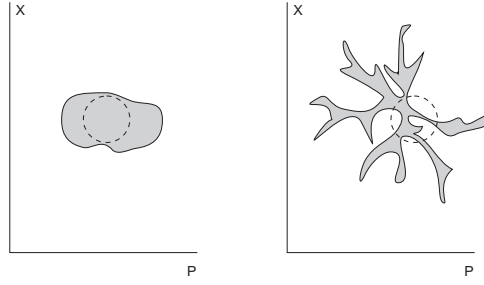
Fig. 11: Sketch of the behavior of an ensemble of particles in phase space. As the ensemble evolves, an initial compact distribution (left panel) becomes less compact.

The initial phase space density of particles in kinetic equilibrium is

$$f_i = \frac{A}{(2\pi)^3} \frac{1}{e^{p/T} + 1} \ ,$$

where we consider fermions for definiteness. The parameter A is determined by requiring that the number density n takes the prescribed value, so that

$$n_0 = \frac{\Omega_{\text{DM}} \rho_c}{m} \ .$$

We find

$$n = \int f_i d^3p = A \cdot \frac{3\zeta(3)}{4\pi^2} T^3 \ ,$$

where $\zeta(3) \approx 1.2$. So, the maximum of the initial phase space density is

$$f_{i, \max} = \frac{n}{12\pi\zeta(3)T^3} = \frac{\Omega_{\text{DM}} \rho_c}{12\pi\zeta(3)mT_{0\,\text{eff}}^3} \ ,$$

where $T_{0\,\text{eff}}$ depends on the decoupling temperature and is somewhat lower than the present photon temperature.

149

On the other hand, one can measure a quantity

$$Q = \frac{\rho_{\mathrm{DM,\,gal}}}{\langle v_{\mathrm{gal}}^2/3 \rangle^{3/2}}$$

where $\rho_{\mathrm{DM,\,gal}}$ is the mass density (say, in a central part of a dwarf galaxy), $\langle v_{\mathrm{gal}}^2 \rangle$ is the average velocity squared, and hence $\langle v_{\mathrm{gal}}^2/3 \rangle$ is the average velocity squared along the line of sight (of stars, and hence dark matter particles, in a virialized galaxy). Since $v_{\mathrm{gal}} = p_{\mathrm{gal}}/m$ and $\rho_{\mathrm{DM,\,gal}} = m n_{\mathrm{gal}}$, one obtains an estimate for the phase space density of dark matter particles in a dwarf galaxy,

$$f \simeq \frac{n_{\mathrm{gal}}}{\langle p_{\mathrm{gal}}^2 \rangle^{3/2}} = \frac{Q}{3^{3/2} m^4} \; .$$

One requires that

$$f < f_{i\,\mathrm{max}}$$

and obtains the bound on the mass of the dark matter particle

$$m \gtrsim 3 \cdot \left(\frac{Q}{\Omega_{\mathrm{DM}} \rho_{\mathrm{c}}} \right)^{1/3} T_{0\,\mathrm{eff}} \; .$$

The values of Q measured in compact dwarf galaxies are in the range

$$Q \sim (5 \cdot 10^{-3} - 2 \cdot 10^{-2}) \cdot \frac{M_\odot/\mathrm{pc}^3}{(\mathrm{km/s})^3}$$

while for the relic that decouples at $T = (1 - 100)$ MeV one has $T_{0\,\mathrm{eff}} = 2.0$ K. This gives [23, 24]

$$m \gtrsim 6 \; \mathrm{keV} \; .$$

Accidentally, this bound is similar to Eq. (36). We note that the bounds coming from the phase space density considerations are called bounds of Tremain–Gunn type.

We also note that similar (in fact, slightly stronger but less robust) bounds are obtained by the study of Lyman-α forest, see, e.g., Ref [25].

5.2   Other hints, SIMP and fuzzy DM

There are two other issues that may or may not be problematic for CDM. One is the "core-cusp problem": CDM-only simulations show singular mass density profiles (cusps) in the centers of galaxies, $\rho_{\mathrm{DM}}(r) \propto r^{-1}$, while observations imply enhanced but smooth profiles (cores). Another is the "too-big-to-fail" problem, which currently means that the densities in large satellite galaxies ($M \sim 10^{10} \; M_\odot$), predicted by CDM-only simulations are systematically higher than the observed mass densities [12].

The astrophysical solutions to these problems again have to do with baryons (supernovae feedback, etc.), and also with interactions of satellite galaxies with large host galaxy, the Milky Way, see, e.g., Refs. [12, 26] for discussion. On the particle physics side, WDM may again help out. Two other particle physics solutions are strongly interacting massive particles (SIMP) as dark matter, and fuzzy dark matter.

The idea of SIMP [27] is that dark matter is cold, but elastic scattering of dark matter particles smoothes out the cuspy mass distribution in galactic centers. Elastic scattering can also lead to a decrease of the dark matter density and thus alleviate the too-big-to-fail problem. To give an idea of the elastic scattering cross section, we take mass density of dark matter of order $\rho_{\mathrm{DM}} \sim 1 \; \mathrm{GeV/cm}^3$ and require that the mean free path of a dark matter particle is of order $l \sim 1$ kpc (typical values, by order of magnitude, both for centers of large galaxies and for dwarf galaxies),

$$1 \sim l \sigma^{(\mathrm{el})} n_{\mathrm{DM}} = l \sigma^{(\mathrm{el})} \frac{\rho_{\mathrm{DM}}}{m} \; ,$$

and obtain

$$\frac{\sigma^{(\mathrm{el})}}{m} \sim \frac{1}{l\rho_{\mathrm{DM}}} \sim 10^{-24}\frac{\mathrm{cm}^2}{\mathrm{GeV}} \ .$$

This is a very large cross section by particle physics standards, and, in view of Eq. (29), the dark matter particle must be fairly light, $m \lesssim 1\,\mathrm{GeV}$. The large elastic cross section may be due to t-channel exchange of a light mediator with $m_{med} \sim 10-100\,\mathrm{MeV}$. This mediator must decay into e^+e^-, $\gamma\gamma$, etc., otherwise it would be dark matter itself. All these features make the SIMP scenario interesting from the viewpoint of collider (search in Z-decays) and "beyond collider" experiments, such as SHiP.

Yet another proposal is fuzzy dark matter consisting of very light bosons,

$$m \sim (10^{-21} - 10^{-22})\,\mathrm{eV} \ .$$

The mechanism of their production must ensure that all of them are born with zero momenta, i.e., these particles form a scalar condensate. An oversimplified picture is that the de Broglie wavelength of these particles at velocities typical for galactic centers and dwarf galaxies, $v \sim 10\,\mathrm{km/s}$, is about 1 kpc:

$$\frac{2\pi}{mv} \sim 1\,\mathrm{kpc} \ .$$

Detailed discussion of advantages of fuzzy dark matter is given, e.g., in Ref. [28]. A way to constrain this scenario is again to study the Lyman-α forest; current constraints [29] are at the level of $2 \cdot 10^{-21}\,\mathrm{eV}$. Interestingly, effects of fuzzy dark matter may in the future be detected by the pulsar timing method [30].

From a particle physics viewpoint, fuzzy dark matter particles may emerge as pseudo-Nambu–Goldstone bosons, similar to axions. We discuss axions later, and here we borrow the main ideas. The axion-like Lagrangian for the pseudo-Nambu–Goldstone scalar field θ reads

$$L = \frac{F^2}{2}(\partial\theta)^2 - \mu^4(1 - \cos^2\theta) \approx \frac{F^2}{2}(\partial\theta)^2 - \frac{\mu^4}{2}\theta^2 \ ,$$

where F is the expectation value of a field that spontaneously breaks approximate $U(1)$ symmetry, and μ is the parameter of the explicit violation of this symmetry. Then the mass of the axion-like particle is

$$m = \frac{\mu^2}{F} \ .$$

The mechanism that creates the scalar condensate is misalignment. The initial value of θ is an arbitrary number between $-\pi$ and π, so that $\theta_i \sim 1$. The field starts to oscillate when the expansion rate becomes small enough, $H \sim m$. The calculation of the present mass density is a simplified version of the axion calculation that we give in Section 7; one finds that $\Omega_{\mathrm{DM}} \sim 0.25$ is obtained for $m = 10^{-22}\,\mathrm{eV}$ if

$$F \sim 10^{17}\,\mathrm{GeV} \ .$$

This is in the ballpark of GUT/string scales, which is intriguing.

5.3 Summary of DM astrophysics

Let us summarize the astrophysics of dark matter.

- Cold dark matter describes remarkably well the distribution and properties of structures in the Universe at relatively large scales, from galaxies like the Milky Way or somewhat smaller ($M \gtrsim 10^{11} M_\odot$), to larger structures like clusters of galaxies, filaments, etc.; also, CDM is remarkably consistent with CMB data which probe even larger scales.
- Currently, data and simulations at shorter scales are inconclusive: they may or may not show that there are "anomalies", the features that contradict the CDM model.

– It will become clear fairly soon whether these "anomalies" are real or not. The progress will come from refined simulations with all effects of baryons included, and from new instruments, notably LSST.

– If the "anomalies" are real, we will have to give up CDM, and, responding to the data, will narrow down the set of dark matter models (WDM, or SIMP, or fuzzy dark matter, or something else). This will have a profound effect on the strategy of search for dark matter particles.

– If the "anomalies" are not there, astrophysics will have to deliver the confirmation of CDM model by the discoveries of relatively light ultra-faint dwarf galaxies ($M = (10^8 - 10^9)M_\odot$) and dark objects of even smaller mass.

All this makes astrophysics a powerful tool of studying dark matter and directing particle physics in its search for dark matter particles.

6 Thermal WIMP

6.1 WIMP abundance: annihilation cross section

Thermal WIMP (weakly interacting massive particle) is a scenario featuring a simple mechanism of the dark matter generation in the early Universe. The WIMP is a *cold* dark matter candidate. Because of its simplicity and robustness, it has been considered by many as the most likely one.

Let us not go into all the details of the (fairly straightforward) calculation of the thermal WIMP abundance. These details are given in several textbooks, and also presented in proceedings of similar Schools, see, e.g., Ref. [31]. Instead, we give the main assumptions behind this mechanism and describe the main steps of the calculation.

One assumes that there exists a heavy stable neutral particle χ, and that χ particles can only be destroyed or created in the cosmic plasma via their pair-annihilation or creation, with the annihilation products being the particles of the Standard Model[6]. We note that there is a version of WIMP model in which the χ particle is not truly neutral, i.e., it does not coincide with its own antiparticle. In that case one assumes that the production and destruction occurs only via $\chi - \bar\chi$ annihilation, and there is no asymmetry between χ and $\bar\chi$ in the cosmic plasma, $n_\chi = n_{\bar\chi}$. The calculation in the $\chi - \bar\chi$ model is identical to the case of a truly neutral particle, so we consider the latter case only.

One also assumes that the χ particles are not strongly coupled, but the $\chi - \chi$ annihilation cross section is sufficiently large, so the χ particles are in complete thermal equilibrium at high temperatures. The latter assumption is justified in the end of the calculation. The thermal equilibrium means, in particular, that the abundance of χ particles is given by the standard Bose–Einstein or Fermi–Dirac distribution formula.

The cosmological behaviour of χ particles is as follows: At high temperatures, $T \gg m_\chi$, the number density of χ particles is high, $n_\chi(T) \sim T^3$. As the temperature drops below m_χ, the equilibrium number density decreases,

$$n_\chi^{(\mathrm{eq})} \propto \mathrm{e}^{-\frac{m_\chi}{T}} \ , \tag{37}$$

At some "freeze-out" temperature T_f the number density becomes so small, that χ particles can no longer meet each other during the Hubble time, and their annihilation terminates[7]. After that the number density of survived χ particles decreases as a^{-3}, and these relic particles form the CDM. The freeze-out

---

[6]The latter assumption can be relaxed: decay products of χ particles may be new particles which sufficiently strongly interact with the Standard Model particles and in the end disappear from the cosmic plasma. Also, destruction and creation of χ particles may occur via co-annihilation with their nearly degenerate partners and inverse pair creation processes; this occurs in a class of supersymmetric models where χ is the lightest supersymmetric particle and its partner is the next-to-lightest supersymmetric particle.

[7]This is a slightly oversimplified picture, which, however, gives a correct estimate, modulo a factor of order 1 in the argument of the logarithm.

temperature T_{f} is obtained by equating the mean free time of χ particle with respect to the annihilation,

$$\tau_{\mathrm{ann}}(T_{\mathrm{f}}) = (\sigma_0(T_{\mathrm{f}})n_\chi(T_{\mathrm{f}}))^{-1}$$

to the Hubble time (see Eq. (20))

$$H^{-1}(T_{\mathrm{f}}) = \frac{M_{\mathrm{Pl}}^*}{T_{\mathrm{f}}^2} \ .$$

Here we introduced the weighted annihilation cross section

$$\sigma_0(T) = \langle \sigma_{\mathrm{ann}} v \rangle_T \ ,$$

where v is the relative velocity of χ particles (in the non-relativistic regime relevant here we have $v \ll 1$), and we average over the thermal ensemble.

Thus, freeze-out occurs when

$$\sigma_0(T_{\mathrm{f}})n_\chi(T_{\mathrm{f}}) = \frac{T_{\mathrm{f}}^2}{M_{\mathrm{Pl}}^*} \ .$$

Because of exponential decay of $n_\chi^{(eq)}$ with temperature, Eq. (37), the freeze-out temperature is smaller than the mass by a logarithmic factor only,

$$T_{\mathrm{f}} \approx \frac{m_\chi}{\ln(M_{\mathrm{Pl}}^* m_\chi \sigma_0)} \ . \tag{38}$$

Note that due to the large logarithm, χ particles are indeed non-relativistic at freeze-out: their velocity squared is of order of

$$v^2(T_{\mathrm{f}}) \simeq 0.1 \ .$$

At freeze-out, the number density is

$$n_\chi(T_{\mathrm{f}}) = \frac{T_{\mathrm{f}}^2}{M_{\mathrm{Pl}}^* \sigma_0(T_{\mathrm{f}})} \ , \tag{39}$$

Note that this density is inversely proportional to the annihilation cross section (modulo logarithm). The reason is that for a higher annihilation cross section, the creation-annihilation processes are longer in equilibrium, and less χ particles survive. Up to a numerical factor of order 1, the number-to-entropy ratio at freeze-out is

$$\frac{n_\chi}{s} \simeq \frac{1}{g_*(T_{\mathrm{f}}) M_{\mathrm{Pl}}^* T_{\mathrm{f}} \sigma_0(T_{\mathrm{f}})} \ . \tag{40}$$

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

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

where we made use of (38). This formula is remarkable. The mass density depends mostly on one parameter, the annihilation cross section σ_0. The dependence on the mass of χ particle is through the logarithm and through $g_*(T_{\mathrm{f}})$; it is very mild. Plugging in $g_*(T_{\mathrm{f}}) \sim 100$, as well as a numerical factor omitted in Eq. (40), and comparing with (28) we obtain the estimate

$$\sigma_0(T_{\mathrm{f}}) \equiv \langle \sigma v \rangle(T_{\mathrm{f}}) = 1 \cdot 10^{-36} \ \mathrm{cm}^2 = 1 \ \mathrm{pb} \ . \tag{41}$$

This is a weak scale cross section, which tells us that the relevant energy scale is 100 GeV–TeV. We note in passing that the estimate (41) is quite precise and robust.

The annihilation cross section can be parameterized as $\sigma_0 = \alpha^2/M^2$ where α is some coupling constant, and M is a mass scale responsible for the annihilation processes[8] (which may be higher than m_χ). This parametrization is suggested by the picture of χ 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 mild assumptions, we find that the WIMP 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,

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

Both are of order 10^{-10} GeV, and it is very tempting to think that this "WIMP miracle" is not a mere coincidence. For long time the above argument has been—and still is—a strong motivation for WIMP searches.

6.2 WIMP candidates: "minimal" and supersymmetry; direct searches

6.2.1 "Minimal" WIMP

Even though the name—*weakly* interacting massive particle—suggests that this particle participates in the Standard Model weak interactions, in most theoretical models this is not so. An exception is the "minimal" WIMP [32]. This is a member of the electroweak multiplet with zero electric charge and zero coupling to the Z boson (couplings to the photon and Z would yield to too strong interactions with the Standard Model particles which are forbidden by direct searches). This is possible for vector-like 5-plet (weak isospin 2) with zero weak hypercharge. Another, albeit fine-tuned option is vector-like triplet (weak isospin 1) with zero weak hypercharge. Particles in vector-like representations may have "hard" masses (not given by Englert–Brout–Higgs mechanism). The right annihilation cross section in Eq. (41) is obtained for masses of these particles

$$\text{5-plet}: \quad m_5 = 9.6 \text{ TeV} , \qquad \text{3-plet}: \quad m_3 = 3 \text{ TeV} .$$

These particles are on the verge of being ruled out by direct searches.

6.2.2 Neutralino

A well motivated WIMP candidate is the neutralino of supersymmetric extensions of the Standard Model. The situation with the neutralino is rather tense, however. One point is that the pair-annihilation of neutralinos often occurs in p-wave, rather than s-wave. This gives the suppression factor in σ_0, proportional to $v^2 \sim 0.1$. Hence, neutralinos tend to be overproduced for large part of the parameter space of MSSM and other supersymmetry (SUSY) models.

Another point is the null results of the direct searches for WIMPs in underground laboratories. The idea of direct search is that WIMPs orbiting around the center of our Galaxy with velocity of order 10^{-3} sometimes hit a nucleus in a detector and deposit small energy in it. The relevant parameters for these searches are WIMP-nucleon elastic scattering cross section and WIMP mass. One distinguishes spin-independent and spin-dependent scattering. In the former case, the WIMP-*nucleus* cross section is proportional to A^2, where A is the number of nucleons in the nucleus (this is an effect of coherent scattering), while in the latter case the cross section is proportional to $J(J+1)$ where J is the spin of the nucleus.

To illustrate the progress in direct searches, we show in Fig. 12 the situation for neutralinos and the direct searches as of 1999, Ref. [33], while Fig. 13 shows the best current limits on the spin-independent cross section [34].

---

[8]For s-wave annihilation, σ_0 is independent of the particle velocity, and hence temperature; if the annihilation is in p-wave, there is an additional suppression by $v^2(T_{\text{f}}) \sim 0.1$.

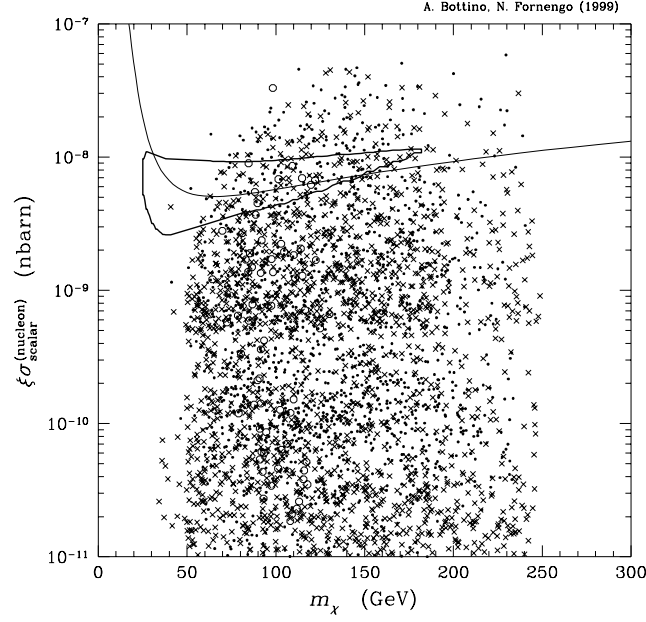A. Bottino, N. Fornengo (1999)

Fig. 12: The situation with neutralinos and the direct searches in 1999 [33]. Shown are theoretical predictions (crosses and dots) and direct detection limits (open solid line; closed solid line is DAMA hint). Vertical axis: spin-independent cross section of elastic WIMP-nucleus scattering per nucleon; parameter ξ takes value 1 for a spin-1/2 neutralino; note that 10^{-10} nbarn $= 10^{-43}$ cm^2.
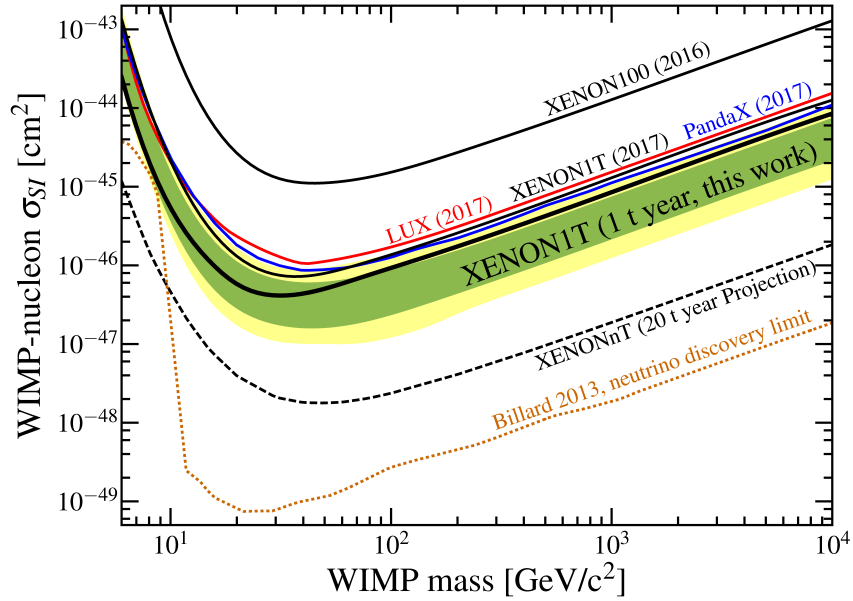

Fig. 13: Current results of direct searches for WIMPs: best limits on spin-independent WIMP-nucleus cross section per nucleon come from the XENON-1T experiment [34]. Note that the cross section 10^{-10} nbarn $= 10^{-43}$ cm^2 in the lower part of Fig. 12 is in the upper part of this figure.

Figure 14 shows both current limits (solid lines) and projected sensitivities of future dark matter detection experiments, again for spin-independent interactions [10]. We see that, on the one hand, the

155

progress in experimental search is truly remarkable, and, on the other, the null results of the searches are becoming alarming. The null results of direct (and also indirect, see below) searches are particularly worrying in view of the null results of SUSY searches at the Large Hadron Collider (LHC).
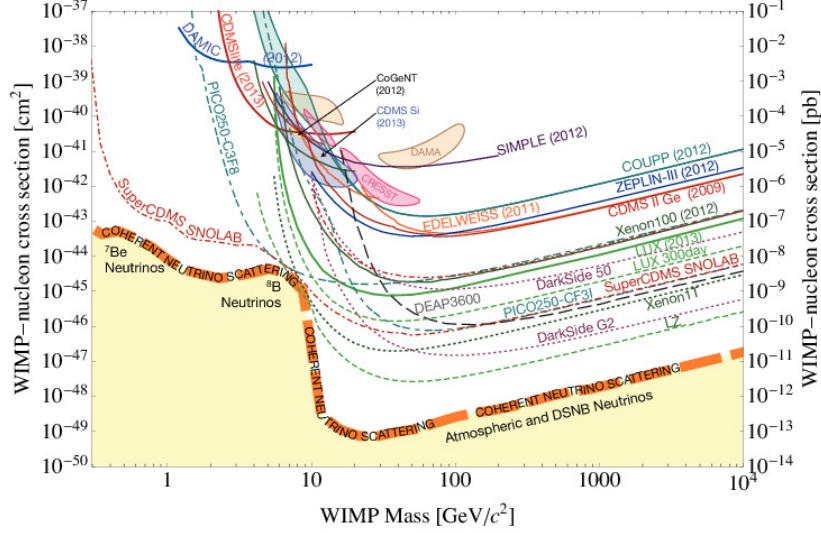
Fig. 14: Current limits and projected sensitivities of direct searches for WIMPs (spin-independent WIMP-nucleus cross section per nucleon). Yellow band in the lower part is the "neutrino floor", at which interactions of cosmic neutrinos become an important background.

6.3 Ad hoc WIMP candidates; indirect searches and the LHC

In view of the strong direct detection limits and null results of the SUSY searches at the LHC, it makes sense to consider less motivated, ad hoc WIMP candidates. The simplest assumption is that the WIMP is not nearly degenerate with any other new particle, so that the calculation of its abundance outlined above applies, and that there is one particle that mediates its pair-annihilation. This mediator can be either a Standard Model particle or a new one; we give examples of both cases. The models of this sort are often called simplified. We emphasize that the two examples of simplified models which we are going to discuss do not exhaust all possible WIMPs and mediators. Some of the models that we leave aside are actually consistent with both cosmology (they give the right value of Ω_{DM}) and experimental limits. The study of numerous simplified models is given, e.g., in Ref. [11].

With this reservation, it is fair to say that many simplified models are either already ruled out or will be ruled out soon. As one illustration, we consider the "Higgs portal", a set of models where the only field which interacts directly with WIMPs is the Englert–Brout–Higgs field. The lowest dimension Higgs-WIMP interaction terms in the cases of a spin-0 WIMP χ and spin-1/2 WIMP ψ are

$$\lambda_\chi^H \chi^* \chi H^\dagger H \ , \qquad \frac{\lambda_\psi^H}{\Lambda} \bar{\psi} \psi H^\dagger H \ ,$$

where H is the EBH field. Here λ_χ^H, λ_ψ^H are dimensionless parameters, while Λ has the dimension of mass. In both cases χ (ψ) is a Standard Model singlet with zero weak hypercharge; it has a "hard" mass $m_{\chi(\psi)}$. Since the vacuum expectation value of the EBH field H is non-zero, the above terms induce trilinear WIMP-WIMP-Higgs interactions responsible for s-channel WIMP annihilation via the Higgs exchange. It is this annihilation that is relevant in the early Universe. The trouble is that almost the entire parameter space of the Higgs portal is ruled out by direct searches. This is illustrated in Fig. 15,
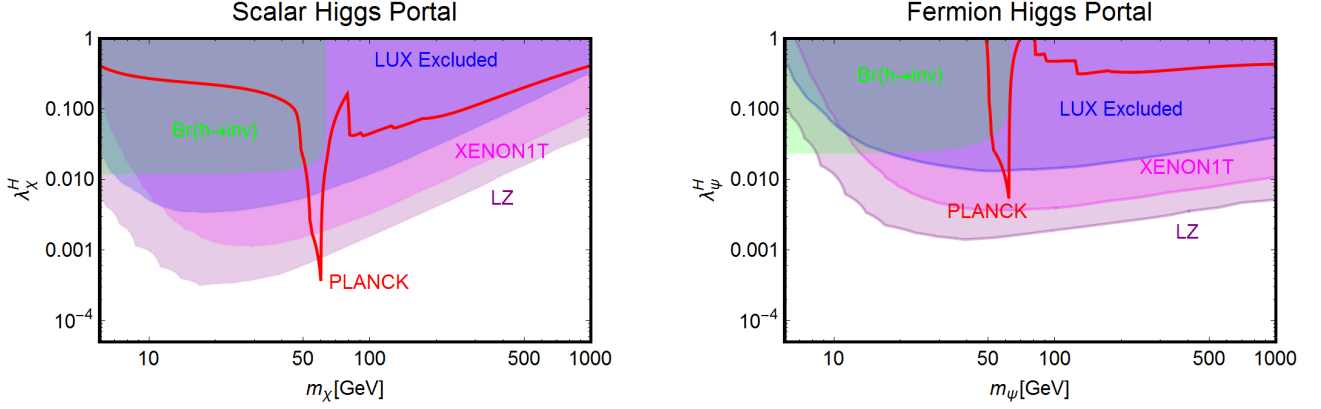
Fig. 15: Predictions from dark matter abundance (red solid lines labeled "PLANCK") and direct detection limits (shadows) in the Higgs portal models [11]. Left panel: spin-0 WIMP; right panel: spin-1/2 WIMP.

Ref. [11]. Another illustration is the Z'-portal. One assumes that both WIMP (say, spin-1/2 particle ψ) and Standard Model fermions interact with a new vector boson Z':

$$g_\psi \bar{\psi}(V_\psi - A_\psi \gamma^5)\psi Z' + \sum_f g_f \bar{f}(V_f - A_f \gamma^5)f\, Z' \,, \tag{42}$$

where sum runs over all Standard Model fermions (an important role is played by quarks). The coupling constants g_ψ, g_f are often chosen to be of order 0.5, as suggested by GUTs. Almost all parameter space of Z'-portal models with $V_\psi \neq 0$ is also ruled out by direct searches [11], as shown in Fig. 16.
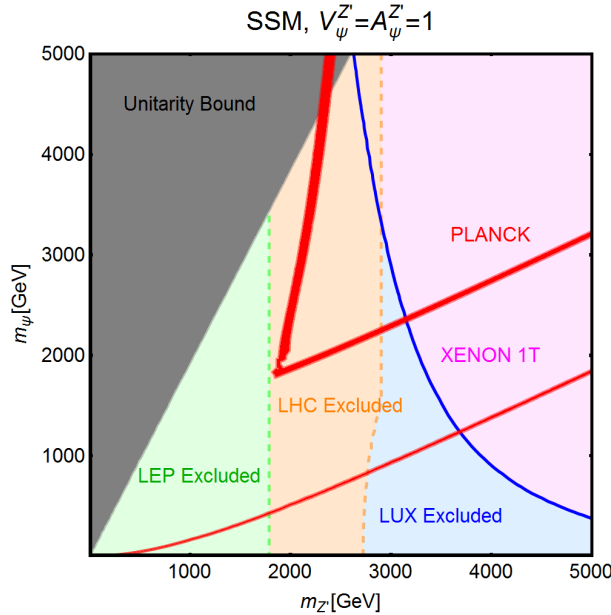
Fig. 16: Same as in Fig. 15 but for the Z'-portal in Eq. (42) with $g_\psi = g_f = 0.65$ and $V_\psi = A_\psi = V_f = A_f = 1$.

The situation is better in models with axial-vector interactions of a new vector boson (we still call it Z') with both the Standard Model particles and WIMPs,

$$V_\psi = V_f = 0 \,.$$

157

In that case, the interaction of WIMPs with a nucleons is spin-dependent, the elastic WIMP-nucleus cross section is not enhanced by A^2, so the direct detection limits are not as strong as in the case of a spin-independent interaction. An important player here is the LHC, whose limits are the most stringent [11], see Fig. 17. We see from Fig. 17 that models with $M_{Z'} \gtrsim 2.8$ TeV are capable of producing the correct abundance of dark matter and at the same time are not ruled out experimentally.
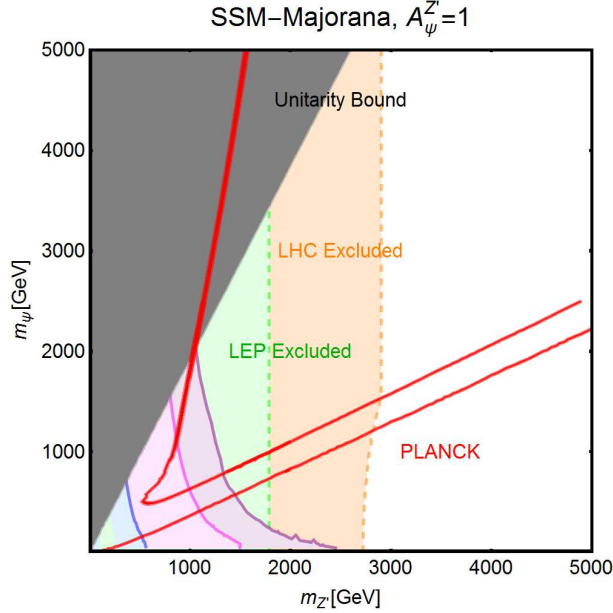
Fig. 17: Same as in Fig. 16 but for an axial-vector Z'-portal in Eq. (42) with and $V_\psi = V_f = 0$, $A_\psi = A_f = 1$.

Another way of comparing current sensitivities of direct and LHC searches is given in Figs. 18 and 19. The plots (compiled by the ATLAS collaboration) refer to the model in Eq. (42) with a vector boson Z' and coupling constants with quarks g_q, leptons g_l and WIMPs $g_\psi \equiv g_\chi$ whose values are written in the figures. Figure 18 shows the limits in the vector case, $A_\psi = A_f = 0$, $V_\psi = V_f = 1$, while Fig. 19 refers to the axial-vector case $A_\psi = A_f = 1$, $V_\psi = V_f = 0$. Clearly, the direct searches are more sensitive than the LHC in the vector case (spin-independent WIMP-nucleon elastic cross section), while the LHC wins in the axial-vector case (spin-dependent elastic cross section). Overall, the LHC has become an important source of limits on WIMPs.

Besides direct and LHC searches for cosmic and collider-produced WIMPs, respectively, important ways to address WIMPs are indirect searches. One approach is to search for high energy γ-rays which are produced in annihilations of WIMPs in various cosmic sources, from dwarf galaxies, to the Galactic center, to clusters of galaxies, and also a diffuse γ-ray flux coming from the entire Universe. This approach is particularly relevant if the WIMP annihilation proceeds in s-wave: in that case the non-relativistic annihilation rate is determined by Eq. (41), which is velocity-independent (modulo a possible Sommerfeld enhancement, see Ref. [35] for a detailed discussion). On the contrary, for p-wave annihilation the rate σv is proportional to v^2, and since the velocities in the sources are small ($v^2 \lesssim 10^{-6}$ as compared to $v^2 \simeq 0.1$ relevant to Eq. (41)), the annihilation cross section is strongly suppressed in the present Universe. Thus, meaningful limits are obtained by γ-ray observatories for WIMPs annihilating in s-wave. The current situation and future prospects are illustrated in Fig. 20, Ref. [10]. The assumption that enters this compilation is that the major WIMP annihilation channel is $b\bar{b}$. Clearly, already existing instruments, and to even larger extent future experiments are sensitive to a wide class of WIMP models.

Indirect searches for dark matter WIMPs include the search for neutrinos coming from the centers of the Earth and Sun (WIMPs may concentrate and annihilate there), see, e.g., Ref. [36], positrons and
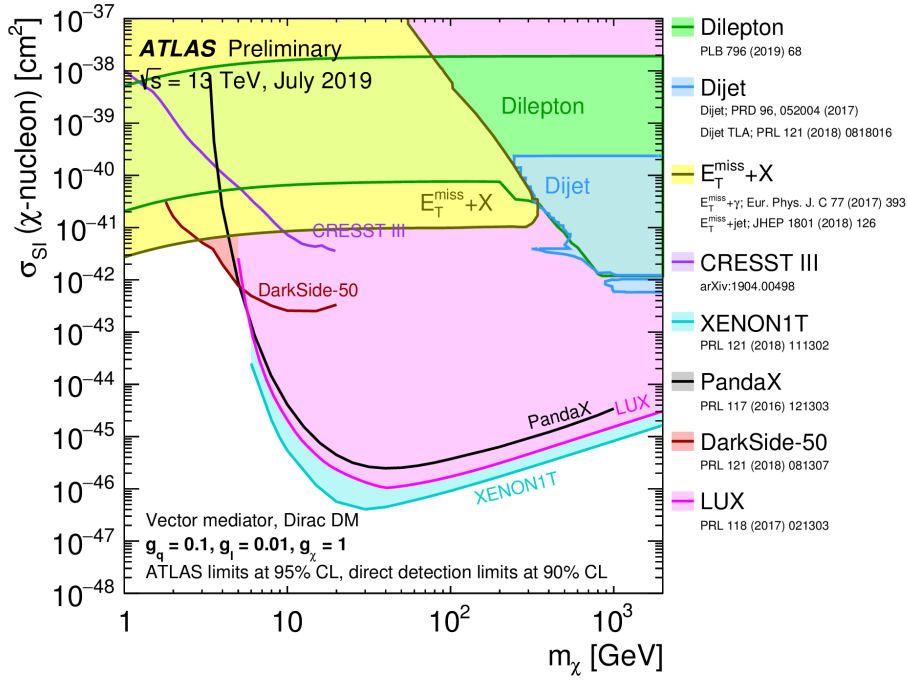
Fig. 18: LHC and direct detection limits in the case of spin-independent WIMP-nucleon elastic cross section.
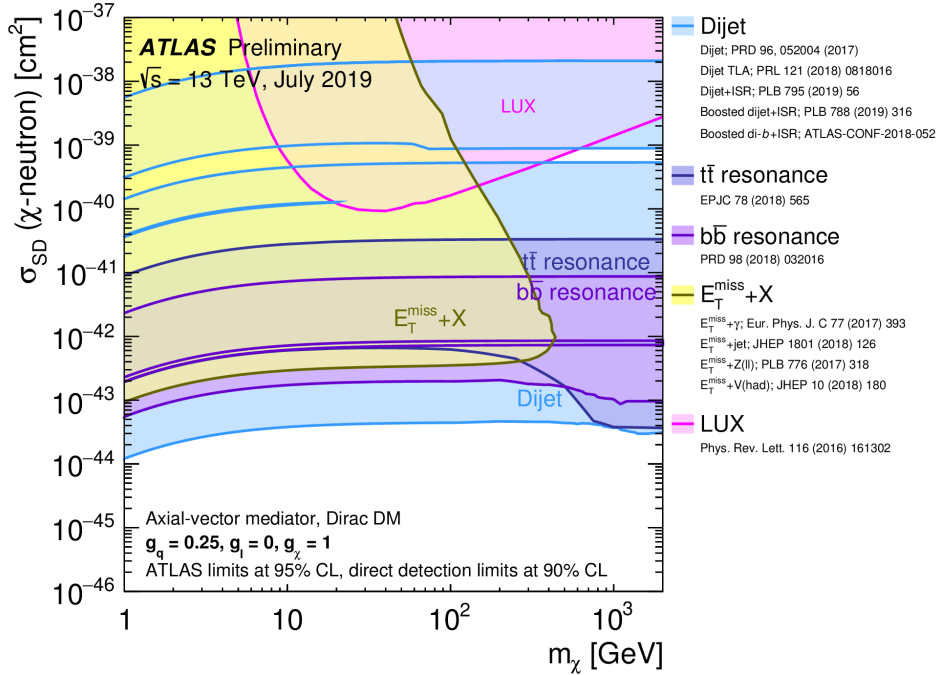

Fig. 19: The same as in Fig. 18 but for spin-dependent WIMP-nucleon elastic cross section.

antiprotons in cosmic rays (produced in WIMP annihilations in our Galaxy), see, e.g., Ref. [37]. These searches have produced interesting, albeit model-dependent limits on WIMP properties.
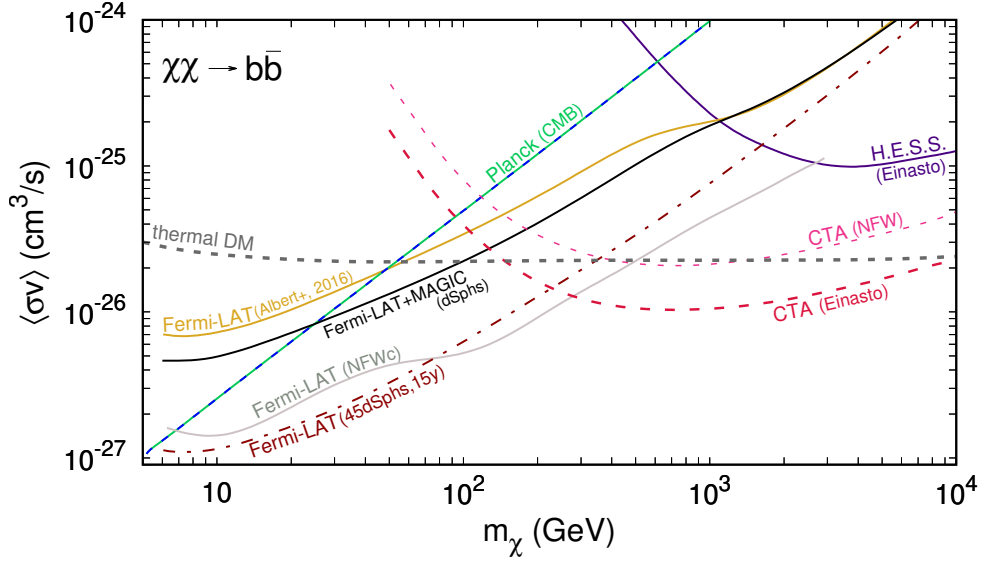
Fig. 20: Limits on WIMP annihilation cross section obtained by γ-ray telescopes (solid lines) and projected sensitivities of future γ-ray observatories (dashed lines). "NFW" and "Einasto" refer to different dark matter profiles in galaxies. Dashed line "thermal DM" is the prediction from cosmology in Eq. (41) under assumption of s-wave annihilation. Note the different units for $\langle \sigma v \rangle$ used in this figure and in Eq. (41).

6.4 WIMP summary

– While the WIMP hypothesis was very attractive for long time, and the SUSY neutralino was considered the best candidate, today the WIMP option is highly squeezed. On the one hand, the parameter space of most of the concrete models is strongly constrained by direct, LHC and indirect searches. On the other hand, SUSY searches at the LHC have moved colored superpartner masses into the TeV region, thus making SUSY less attractive from the viewpoint of solving the gauge hierarchy problem.

– This does not mean too much, however: we would like to discover *one theory* and *one point in its parameter space*.

– Hunt for WIMPs continues in numerous directions. Their potential is far from being exhausted. Concerning direct searches, we will soon face the neutrino floor problem—the situation where the cosmic neutrino background will show up. It is time to look into ways to go beneath the neutrino floor.

– With a null results of WIMP searches, it makes a lot of sense to strengthen also searches for other dark matter candidates.

7 Axions

An Axion is a consequence of the Peccei–Quinn solution to the strong CP problem. It is a pseudo-Nambu–Goldstone boson of an approximate Peccei–Quinn symmetry.

7.1 Strong CP problem

To understand the strong CP problem, we begin with considering QCD in the chiral limit $m_u = m_d = m_s = 0$. The Lagrangian is

$$L_{QCD, m=0} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \sum_i \bar{q}_i i\gamma^\mu D_\mu q_i$$

160

$$= -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} + \sum_i \left(\bar{q}_{L,i} i\gamma^\mu D_\mu q_{L,i} + \bar{q}_{R,i} i\gamma^\mu D_\mu q_{R,i} \right) \ ,$$

where $i = u, d, s$. As it stands, it is invariant under independent transformations of left and right quark fields $q_{L,i}$ and $q_{R,i}$, each with arbitrary unitary matrices. Naively, this means that the theory possesses a large symmetry

$$SU(3)_L \times U(1)_L \times SU(3)_R \times U(1)_R = SU(3)_L \times SU(3)_R \times U(1)_B \times U(1)_A \tag{43}$$

where vector $U(1)_B$ is baryon number symmetry, $q_i \to e^{i\alpha} q_i$, while axial $U(1)_A$ act as $q_i \to e^{i\beta\gamma^5} q_i$.

The symmetry in Eq. (43) is spontaneously broken: there exist quark condensates in the QCD vacuum:

$$\langle \bar{u}_L u_R \rangle = \langle \bar{d}_L d_R \rangle = \langle \bar{s}_L s_R \rangle = \frac{1}{2}\langle \bar{q}q \rangle \sim \Lambda_{\text{QCD}}^3 \tag{44}$$

The unbroken symmetry $SU(3)_V$ rotates left and right quarks together (this is the well known flavor $SU(3)$); $U(1)_B$ also remains unbroken.

Spontaneous breaking of a global symmetry always leads to the presence of a Nambu–Goldstone bosons. Naively, one expects that there are 9 Nambu–Goldstone bosons: 8 of them come from symmetry breaking $SU(3)_L \times SU(3)_R \to SU(3)_V$, and one from $U(1)_B \times U(1)_A \to U(1)_B$ (since the original symmetry is explicitly broken by quark masses, these should be pseudo-Nambu–Goldstone bosons with non-zero mass). However, there are only 8 light pseudoscalar particles whose properties are well described by Nambu–Goldstone theory: these are $\pi^\pm$, π^0, $K^\pm$, K^0, $\bar{K}^0$, η. Indeed, their masses squared are proportional to quark masses, e.g., $m_\pi^2 = (m_u + m_d)\langle \bar{q}q \rangle / f_\pi^2$. Importantly, yet another pseudoscalar η' is heavy and does not behave like pseudo-Nambu–Goldstone boson.

The reason for this mismatch (absence of the 9th pseudo-Nambu–Goldstone boson) is that $U(1)_A$ is not, in fact, a symmetry of QCD even in the chiral limit. The corresponding axial current suffers, at the quantum level, an Adler–Bell–Jackiw (triangle, or axial) anomaly,

$$\partial_\mu J_A^\mu \neq 0 \ .$$

This means that the axial charge is not conserved, and thus the $U(1)_A$ is explicitly broken. We discuss this phenomenon in little more detail in Section 10.2 in the context of electroweak baryon number non-conservation.

The strong CP problem [38–40] emerges in the following way. One considers quark mass terms in the Standard Model Lagrangian, which are obtained from the Yukawa interaction terms with a non-zero Higgs expectation value. A common believe is that one can perform unitary rotations of quark fields to make quark mass terms real (and in this way generate the Cabibbo–Kobayashi–Maskawa (CKM) matrix in quark interactions with W-bosons). This is not quite true, precisely because one cannot freely use the $U(1)_A$-rotation. In fact, by performing a $SU(3)_L \times SU(3)_R \times U(1)_B$-rotation, one casts the mass term of light quarks into the form

$$L_m = e^{i\theta} \cdot m_{ij}^{\text{CKM}} \bar{q}_{L,i} q_{R,j} + h.c. \ ,$$

where $m_{ij}^{\text{CKM}} = \text{diag}(m_u, m_d, m_s)$ is a real diagonal matrix, and θ is some phase. Naively, this phase can be rotated away by an axial rotation of all three light quark fields, $q_i \to e^{-i\theta\gamma_5/2} q_i$, but, as we discussed, this is not an innocent field redefinition. What happens instead is that this transformation generates an extra term in the QCD Lagrangian

$$\Delta L = \frac{\alpha_s}{8\pi} \cdot \theta \cdot G_{\mu\nu}^a \tilde{G}^{\mu\nu \, a} \ , \tag{45}$$

where α_s is the $SU(3)_c$ gauge coupling, $G_{\mu\nu}^a$ is the gluon field strength, $\tilde{G}^{\mu\nu \, a} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho} G_{\lambda\rho}^a$ is the dual tensor. The term (45) is invariant under gauge symmetries of the Standard Model, but it violates P and

CP symmetry. A similar term, but with another parameter θ_0 instead of θ, can already exist in the initial QCD Lagrangian. The combined parameter

$$\bar{\theta} = \theta + \theta_0$$

is a "coupling constant" that cannot be removed by field redefinition, and QCD with a non-zero $\bar{\theta}$ violates CP symmetry.

Let us show explicitly that the parameter $\bar{\theta}$ is physical, i.e., some physical quantities depend on $\bar{\theta}$. To this end, we perform a chiral rotation of light quark fields $q_i \to e^{+i\bar{\theta}\gamma_5/2}q_i$ to get rid of the term (45) and generate the phase in the quark mass terms

$$L_m = \sum_i e^{i\bar{\theta}} m_i \bar{q}_{L,i} q_{R,i} + h.c.$$

Let us consider for simplicity two light quark flavors u and d with equal masses $m_\mathrm{u} = m_\mathrm{d} \equiv m_\mathrm{q} \sim 4$ MeV and calculate the vacuum energy density in such a theory. We use perturbation theory in quark masses, and work to the leading order. Then the $\bar{\theta}$-dependent part of the vacuum energy density is $V(\bar{\theta}) = -\langle L_m \rangle$. We recall that $\langle \bar{q}q \rangle$ is non-zero in the chiral limit, see Eq. (44), and observe that it is real, provided that the term (45) is absent (no spontaneous CP violation in the chiral limit). Importantly, $\langle \bar{q}q \rangle$ does not have an arbitrary phase, since the arbitrariness of this phase would mean that $U(1)_A$ is a (spontaneously broken) symmetry, which is not the case, as we discussed above. Thus, we obtain

$$V(\bar{\theta}) = -\langle L_m \rangle = -2m_\mathrm{q}\langle \bar{q}q \rangle \cos\bar{\theta} = -\frac{m_\pi^2 f_\pi^2}{4} \cos\bar{\theta} \ . \tag{46}$$

This shows explicitly that $\bar{\theta}$ is a physically relevant parameter. We note in passing that the expression for $V(\bar{\theta})$ is, in fact, more complicated, especially for $m_u \neq m_d$ and also for three quark flavors, but the main property—minimum at $\bar{\theta} = 0$—is intact.

Thus, $\bar{\theta}$ is a new coupling constant that can take any value in the interval $(-\pi, \pi)$. There is no reason to think that $\bar{\theta} = 0$. The term (45) has a dramatic phenomenological consequence: it generates a electric dipole moment (EDM) of the neutron d_n, which is estimated as [41]

$$d_n \sim \bar{\theta} \cdot 10^{-16} \cdot e \cdot \mathrm{cm} \ . \tag{47}$$

The neutron EDM is strongly constrained experimentally,

$$d_n \lesssim 3 \cdot 10^{-26} \cdot e \cdot \mathrm{cm} \ . \tag{48}$$

This leads to the bound on the parameter $\bar{\theta}$,

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

The problem to explain such a small value of $\bar{\theta}$ 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 an axion. The idea of these models is to promote the $\bar{\theta}$-parameter to a field, which is precisely the axion field. This can be done in various ways. Two well-known ones are Dine–Fischler–Srednicki–Zhitnitsky [45,46] (DFSZ) and Kim–Shifman–Vainshtein–Zakharov [47,48] (KSVZ) mechanisms[9]. In either case, one introduces a complex scalar field Φ and makes sure that without QCD effects, the theory is invariant under global Peccei–Quinn $U(1)_{PQ}$ symmetry. Under this symmetry, the field Φ transforms as $\Phi \to e^{i\alpha}\Phi$. One also arranges that the QCD effects make this symmetry anomalous,

---

[9]Earlier and even simpler is Weinberg–Wilczek model [43,44], but it is ruled out experimentally.

very much like $U(1)_A$, so that under the $U(1)_{PQ}$-transformation, the Lagrangian obtains an additional contribution

$$\Delta L = C \frac{\alpha_s}{8\pi} \cdot \alpha \cdot G^a_{\mu\nu} \tilde{G}^{\mu\nu\,a} \,, \tag{49}$$

where C is a model-dependent constant of order 1. A simple example is the KSVZ model: one adds a new quark ψ which interacts with Φ as follows:

$$L_{int} = h\Phi \bar{\psi}_L \psi_R + h.c. \tag{50}$$

where h is Yukawa coupling. Then the Peccei–Quinn transformation is

$$\Phi \to e^{i\alpha}\Phi \,, \qquad \psi \to e^{i\alpha\gamma^5/2}\psi \,,$$

while "our" quark fields are $U(1)_{PQ}$-singlets. In the same way as above, this transformation induces the term (49), as required.

Now, one arranges the scalar potential for Φ in such a way that the Peccei–Quinn symmetry is spontaneously broken at very high energy. If not for QCD effects, the phase of Φ would be a massless Nambu–Goldstone boson, the axion. At low energies one writes $\Phi = f_{PQ} \cdot e^{i\theta(x)}$, where f_{PQ} is the Peccei–Quinn vacuum expectation value. In the absence of QCD, the field θ is rotated away from the non-derivative part of the action by the Peccei–Quinn rotation, while it reappears in the form of Eq. (49) when QCD is switched on. We see that the parameter $\bar{\theta}$ is indeed promoted to a field, and this parameter disappears upon shifting $\theta(x) \to \theta(x) - \bar{\theta}$; we are free to set $\bar{\theta} = 0$. Now, there is a potential for the field θ; it is given precisely by Eq. (46) with $\bar{\theta}$ replaced by θ. Hence, the low energy axion Lagrangian reads

$$L_a = \frac{f^2_{PQ}}{2} \partial_\mu \theta \partial^\mu \theta - V(\theta) \,.$$

As usual, the first term here comes from the kinetic term for the field Φ. We recall that the minimum of $V(\theta)$ is at $\theta = 0$; at this value CP symmetry is not violated, the strong CP problem is solved! We now make a field redefinition, $\theta = a/f_{PQ}$ and find from Eq. (46) that the quadratic axion Lagrangian is

$$L_a = \frac{1}{2} \partial_\mu a \partial^\mu a - \frac{m^2_a}{2} a^2 \,,$$

where

$$m_a = \frac{m_\pi f_\pi}{2 f_{PQ}} \,. \tag{51}$$

The axion is a *pseudo*-Nambu-Goldstone boson.

To summarize, for large Peccei–Quinn scale $f_{PQ} \gg M_W$, the axion is a light particle whose interactions with the Standard Model fields are very weak. Like for any Nambu–Goldstone field, the tree-level interactions of the axion with quarks and leptons are described by the generalized Goldberger–Treiman formula

$$L_{af} = \frac{1}{f_{PQ}} \cdot \partial_\mu a \cdot J^\mu_{PQ} \,. \tag{52}$$

Here

$$J^\mu_{PQ} = \sum_f e^{(PQ)}_f \cdot \bar{f} \gamma^\mu \gamma^5 f \,. \tag{53}$$

The contributions of fermions to the current J^μ_{PQ} are proportional to their PQ charges $e^{(PQ)}_f$; these charges are model-dependent. There is necessarily an interaction of axions with gluons, see Eq. (49),

$$L_{ag} = C_g \frac{\alpha_s}{8\pi} \cdot \frac{a}{f_{PQ}} \cdot G^a_{\mu\nu} \tilde{G}^{\mu\nu\,a} \tag{54}$$

Finally, there is an axion-photon coupling

$$L_{a\gamma} = g_{a\gamma\gamma} \cdot aF_{\mu\nu}\tilde{F}^{\mu\nu} \ , \qquad g_{a\gamma\gamma} = C_\gamma \frac{\alpha}{8\pi f_{PQ}} \ , \tag{55}$$

The dimensionless constants C_g and C_γ are model-dependent and, generally speaking, not very much different from 1. The main free parameter is f_{PQ}, while the axion mass is related to it via Eq. (51); numerically,

$$m_a = 6 \ \mu\text{eV} \ \cdot \left(\frac{10^{12} \ \text{GeV}}{f_{PQ}} \right) \ . \tag{56}$$

There are astrophysical bounds on the strength of axion interactions f_{PQ}^{-1} and hence on the axion mass. Axions in theories with $f_{PQ} \lesssim 10^9$ GeV, which are heavier than about 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, $m_a \lesssim 10^{-2}$ eV. These very light and very weakly interacting axions are interesting dark matter candidates[10].

7.2   Axions in cosmology

Axions can serve as dark matter if they do not decay within the lifetime of the Universe. The main decay channel of a light axion is the decay into two photons. The axion width is calculated as

$$\Gamma_{a\to\gamma\gamma} = \frac{m_a^3}{4\pi} \left(C_\gamma \frac{\theta}{8\pi f_{PQ}} \right)^2 \ ,$$

where the quantity in parenthesis is the axion-photon coupling, see Eq. (55). We recall the relation in Eq. (51) and obtain an axion lifetime

$$\tau_a = \frac{1}{\Gamma_{a\to\gamma\gamma}} = \frac{64\pi^3 m_\pi^2 f_\pi^2}{C_\gamma^2 \alpha^2 m_a^5} \sim 10^{24} \ \text{s} \ \cdot \left(\frac{\text{eV}}{m_a} \right)^5 \ .$$

By requiring that this lifetime exceeds the age of the Universe, $\tau_a > t_0 \approx 14$ billion years, we find a very weak bound on the mass of an axion as a dark matter candidate, $m_a < 25$ eV.

Thermal production of axions in the early Universe is not very relevant, since even if they were in thermal equilibrium at high temperatures, their thermally produced number density at present is substantially smaller than that of photons and neutrinos, and with their tiny mass they do not contribute much into the energy density[11]. This is a welcome property, since thermally produced axions, if they composed substantial part of dark matter, would be *hot* dark matter, which is ruled out.

There are at least two mechanisms of axion production in the early Universe that can provide not only the right axion abundance, but also small initial velocities of axions. The latter property makes the axion a *cold* dark matter candidate, despite its very small mass.

One mechanism [50–52] is called the misalignment scenario. It assumes that the Peccei–Quinn symmetry is spontaneously broken before the beginning of the hot epoch, $\langle\Phi\rangle \neq 0$. This is indeed the case in the inflationary framework, if f_{PQ} is higher than both the inflationary Hubble parameter (towards the inflation end) and the reheat temperature of the Universe. In this case the axion field (the phase of the field Φ) is homogeneous over the entire visible Universe, and initially it can take any value θ_0 between $-\pi$ and π. As we have seen in Eq. (46), the axion potential is proportional to the quark condensate $\langle\bar{q}q\rangle$. This condensate vanishes at high temperatures, $T \gg \Lambda_{\text{QCD}}$, and the axion potential is negligibly small. As the temperature decreases, the axion potential builds up. This is shown in Fig. 21. Accordingly, the

---

[10]We note in passing that axions may be heavy instead [49]. This case is irrelevant for dark matter.

[11]If axions were in thermal equilibrium, they contribute to the effective number of "neutrino" species N_{eff}. This contribution, however, is smaller than the current precision [3] of the determination of N_{eff}, which is equal to ± 0.17.
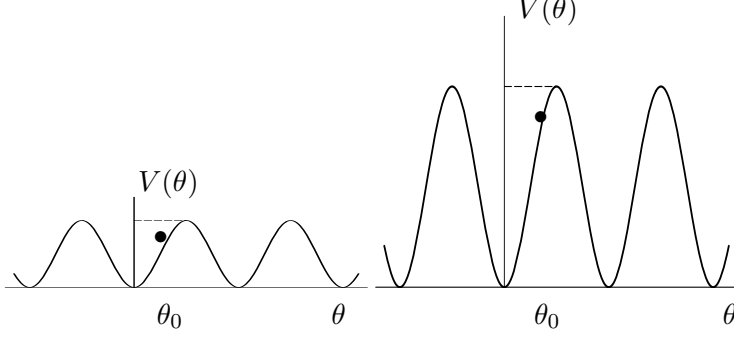
Fig. 21: The axion potential at higher temperature (left) and lower temperature (right). The bullet shows the initial value of the axion field. The field starts to roll down the potential at the time when $m(T) \sim H(T)$.

axion mass increases from zero to m_a; hereafter m_a denotes the zero-temperature axion mass. The axion field practically does not evolve when $m_a(T) \ll H(T)$ and at the time when $m_a(T) \sim H(T)$ it starts to roll down from the initial value θ_0 to the minimum $\theta = 0$ and then it oscillates. During all these stages of evolution, the axion field is homogeneous in space. The homogeneous oscillating field can be interpreted as a collection of scalar quanta with zero spatial momenta, the axion condensate. This is indeed cold dark matter.

Let us estimate the present energy density of the axion field in this picture. The oscillations start at the time t_{osc} when $m_a(t_{osc}) \sim H(t_{osc})$. At this time, the energy density of the axion field is estimated as

$$\rho_a(t_{osc}) \sim m_a^2(t_{osc})a_0^2 = m_a^2(t_{osc})f_{PQ}^2\theta_0^2 .$$

The number density of axions at rest at the beginning of oscillations is estimated to be

$$n_a(t_{osc}) \sim \frac{\rho_a(t_{osc})}{m_a(t_{osc})} \sim m_a(t_{osc})f_{PQ}^2\theta_0^2 \sim H(t_{osc})f_{PQ}^2\theta_0^2 .$$

This number density, as any number density of non-relativistic particles, then decreases as a^{-3}. The axion-to-entropy ratio at time t_{osc} is

$$\frac{n_a}{s} \sim \frac{H(t_{osc})f_{PQ}^2}{\frac{2\pi^2}{45}g_*T_{osc}^3} \cdot \theta_0^2 \simeq \frac{f_{PQ}^2}{\sqrt{g_*}T_{osc}M_{Pl}} \cdot \theta_0^2 ,$$

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

$$\rho_{a,0} = \frac{n_a}{s}m_a s_0 \simeq \frac{m_a f_{PQ}^2}{\sqrt{g_*}T_{osc}M_{Pl}}s_0 \cdot \theta_0^2 . \tag{57}$$

To obtain a simple estimate, let us set $T_{osc} \sim \Lambda_{\mathrm{QCD}} \simeq 200$ MeV and make use of Eq. (56). We find

$$\Omega_a \equiv \frac{\rho_{a,0}}{\rho_c} \simeq \left(\frac{10^{-6} \text{ eV}}{m_a}\right)\theta_0^2 . \tag{58}$$

The natural assumption about the initial phase is $\theta_0 \sim \pi/2$. Hence, an axion of mass $m_a = $ (a few) $\cdot$ 10^{-6} eV is a good dark matter candidate. Note that an axion of lower mass $m_a < 10^{-6}$ eV may also serve as a dark matter particle, if for some reason the initial phase θ_0 is much smaller than $\pi/2$.

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

$$\Omega_a \simeq 0.2 \cdot \theta_0^2 \cdot \left(\frac{4 \cdot 10^{-6} \text{ eV}}{m_a}\right)^{1.2}$$

We see that our crude estimate in Eq. (58) is fairly accurate.

We note that in the misalignment scenario, and in the inflationary framework, the initial phase θ_0 is not quite homogeneous in space. At the inflationary stage, vacuum fluctuations of all massless or light scalar fields get enhanced. As a result, scalar fields become inhomogeneous on scales exceeding the inflationary Hubble scale H_{infl}^{-1}. The amplitudes of these inhomogeneities (for canonically normalized fields) are equal to $H_{infl}/(2\pi)$. Phase perturbations give rise to perturbations of the axion dark matter energy density, which are uncorrelated with perturbations of conventional matter. These uncorrelated dark matter perturbations are called isocurvature (or entropy) modes. Cosmological observations show that their contribution cannot exceed a few per cent of the dominant adiabatic mode. This leads to a constraint [53] on the inflationary Hubble parameter H_{infl} or, equivalently, on the energy scale of the inflation (energy density of the inflaton field)

$$V_{infl}^{1/4} \lesssim 10^{13} \text{ GeV} .$$

This makes the misalignment mechanism somewhat contrived. Reversing the argument, detection of the dark matter entropy mode would be an interesting hint towards the nature of dark matter.

Another mechanism of axion production in the early Universe works under the assumption which is opposite to the main assumption of the misalignment scenario. Namely, one assumes that the Peccei–Quinn symmetry is restored at the beginning of the hot epoch, and gets spontaneously broken at a temperature of order $T \sim f_{PQ}$ at the hot stage. Then the phase of the field Φ is uncorrelated at distances exceeding the size of the horizon at that time. In principle, one should be able to predict the value of f_{PQ} and hence m_a in this scenario, since there is no uncertainty in the initial conditions. However, the dynamics in this case is quite complicated. Indeed, the uncorrelated phase gives rise to the production of global cosmic strings [54]—topological defects that exist in theories with a spontaneously broken global $U(1)$ symmetry ($U(1)_{PQ}$ in our case; for a discussion see, e.g., Ref. [55]). At the QCD transition epoch, defects of another type, axion domain walls, are created. Then all these defects get destructed, giving rise to the production of axions. The analysis of this dynamics has been made by various authors, see, e.g., Refs. [56, 57], but it is fair to say that there is no compelling prediction for m_a yet. A reasonable estimate of the axion mass is (Ref. [56] claims $m_a = 2.6 \cdot 10^{-5}$ eV)

$$m_a = (\text{a few}) \cdot 10^{-5} \text{ eV} .$$

To end up with cosmological aspects of axion dark matter, we note that it has interesting phenomenology in the present Universe. Axions tend to form mini-clusters [58] which can be disrupted and form streams of dark matter [59]. Axions also form Bose-stars [60]. All this exotica is relevant to both astrophysics and the axion search.

7.3 Axion search

The search for dark matter axions with mass $m_a \sim 10^{-5} - 10^{-6}$ eV is difficult, but not impossible. One way is to search for an axion-photon conversion in a resonant cavity filled with a strong magnetic field. Indeed, in the background magnetic field, the axion-photon interaction in Eq. (55) leads to the conversion $a \rightarrow \gamma$, see Fig. 22. Axions of mass $10^{-5} - 10^{-6}$ eV are converted to photons of frequency $\nu = m/(2\pi) = 2 - 0.2$ GHz (radio waves; $m = 10^{-6}$ eV $\longleftrightarrow \nu = 240$ MHz). To collect a reasonable number of conversion photons, one needs cavities of a high quality factor Q, which have a small bandwidths. This means that one goes in small steps in m_a, and the whole search takes a long time. This is illustrated in Fig. 23.

The hunt for dark matter axions has been intensified recently. A new set of resonant cavity experiments, CAPP, is under preparation, see Fig. 24. A new approach to search for heavier dark matter axions with $m_a \gtrsim 4 \cdot 10^{-5}$ eV has been suggested by the MADMAX interest group [64]. Other axion search experiments are reviewed in Refs. [63, 65].
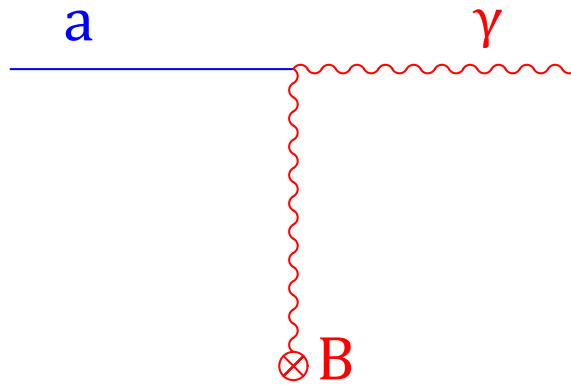
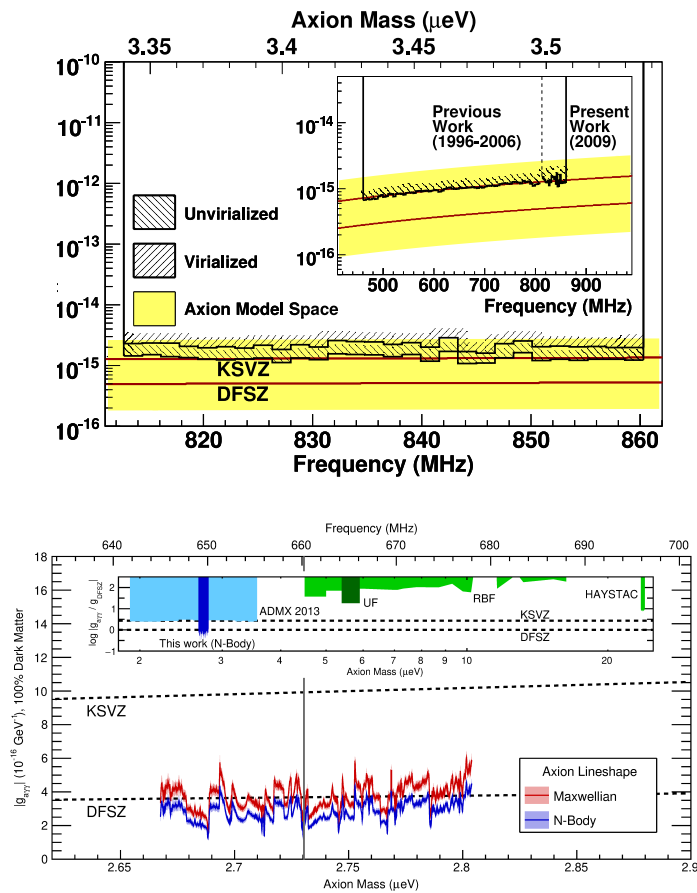**Fig. 22:** Axion-photon conversion in magnetic field.

Fig. 23: Limits on the axion-photon coupling for various axion masses. Lines labeled KSVZ and DSVZ refer to predictions of the two axion models under the assumption that axions make the whole of dark matter. Shown are limits published by ADMX collaboration in 2010 [61] (upper panel) and in 2018 [62]. Note the limited ranges of masses spanned during the long period of time. Note also that the recent limits (lower panel) reach almost entire range of axion-photon couplings predicted by various axion models.
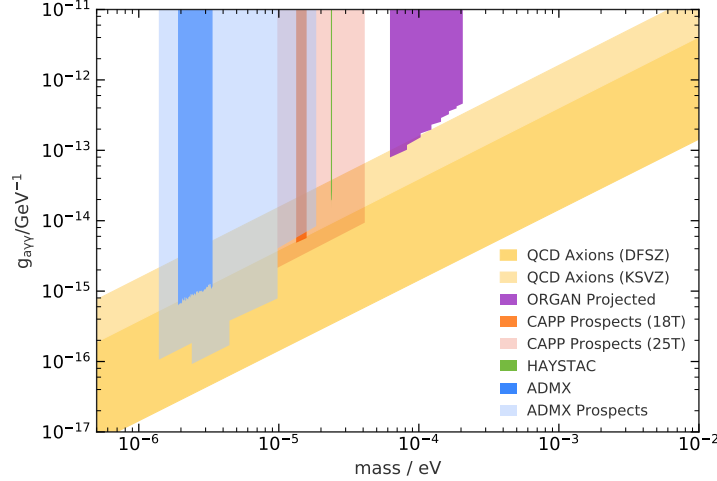
Fig. 24: Future prospects of dark matter axion searches with resonant cavities [63].

7.4 Axion-like particles (ALPs)

There may exist light, weakly interacting scalar or pseudoscalar particles other than axions. They are called axion-like particles, ALPs, and they may emerge as pseudo-Nambu–Goldstone bosons of some new approximate global symmetry. We have discussed one example, fuzzy dark matter, in Section 5.2. Unlike the axion case, where the axion-photon coupling is related, albeit in somewhat model-dependent way, to its mass via Eqs. (55) and (56), the ALP mass and coupling to photons are both arbitrary parameters. Also, ALPs may interact with the Standard Model fermions, and that coupling is again a free parameter. ALPs may or may not be dark matter candidates; searches for them is of interest independently of the dark matter problem.

If the ALP is a dark matter candidate, instruments described in previous subsection—"haloscopes"—are capable for searching for dark matter ALPs, and it makes sense to extend the search to an as wide mass range as possible. In this regard, it is worth mentioning that the CASPEr experiment [66] is going to be sensitive to very light ALPs, $m \lesssim 10^{-9}$ eV, and very small ALP-fermion couplings.

Bounds and prospects for search for light ALPs are summarized Fig. 26.

ALPs may be produced in the Sun, and their flux may be detectable by "helioscopes", instruments searching for the axion-photon conversion in the magnetic field of a magnet looking at the Sun. One such instrument, CAST, has been operating for a long time, whereas other experiments, IAXO and TASTE,
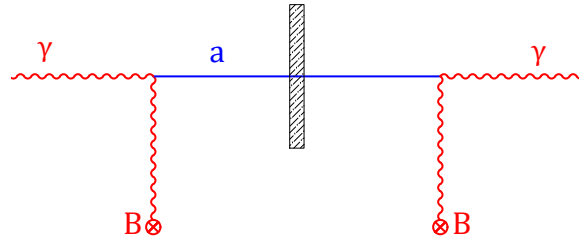
Fig. 25: "Light shining through a wall": laser light shining from the left is converted into axions in the magnetic field of a magnet placed before the wall, axions pass through the wall and are converted into photons by a magnet behind the wall; the latter photons are detected by a highly sensitive photon detector.

168
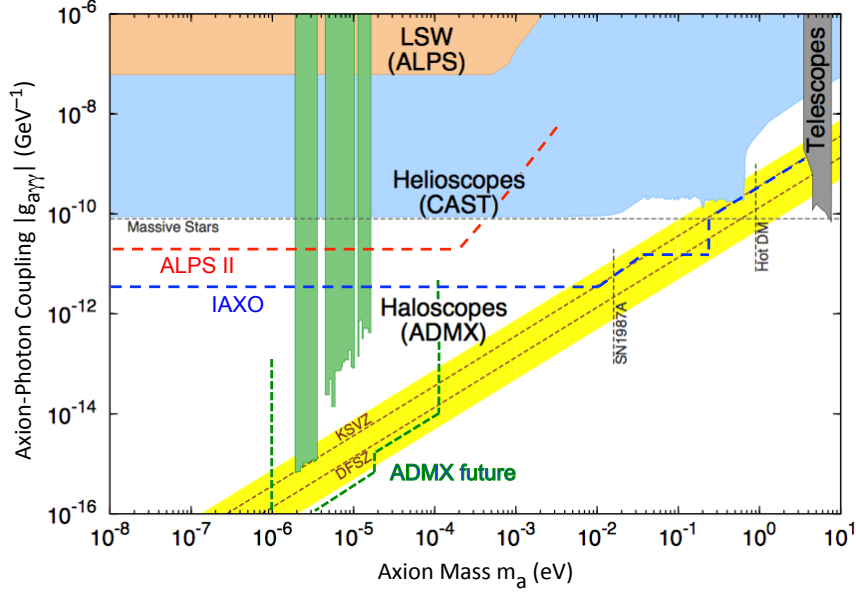
Fig. 26: Bounds on ALPs: ALP-photon coupling vs ALP mass [65]. The inclined straight strip with lines labeled "KSVZ" and "DFSZ" is the range of predictions of axion models. Shaded regions are limits from existing experiments, dashed lines show sensitivities of future searches.

are planned. Another way to search for ALPs makes use of the idea of "light shining through a wall", see Fig. 25; this idea is implemented in the ALPS-I, ALPS-II experiments. For a review of these approaches see, e.g., Ref. [65]. Finally, ALPs can be searched in beam-dump experiments and in decays of K- and B-mesons. Interesting limits are obtained by the CHARM and BaBar experiments, and a promising planned experiment is SHiP at CERN [67].

8 Warm dark matter: sterile neutrinos

As we discussed in Section 5.1, there are arguments, albeit not yet conclusive, which favor warm, rather than cold, dark matter. If WDM particles were in kinetic equilibrium at some epoch in the early Universe, then their mass should be in the range of $3 - 10$ keV. Reasonably well motivated particles of this mass are sterile neutrinos.

Sterile neutrinos—massive leptons N which do not participate in the Standard Model gauge interactions—are most probably required for giving masses to ordinary, "active" neutrinos. The masses of sterile neutrinos cannot be predicted theoretically. Although sterile neutrinos of a mass of $m_N = 3 - 10$ keV are not particularly plausible from the particle physics prospective, they are not pathological either. In the simplest case the creation of sterile neutrino states $|N\rangle$ in the early Universe occurs due to their mixing with active neutrinos $|\nu_\alpha\rangle$, $\alpha = e, \mu, \tau$. In the approximation of mixing between two states only, we have

$$|\nu_\alpha\rangle = \cos\theta |\nu_1\rangle + \sin\theta |\nu_2\rangle , \qquad |N\rangle = -\sin\theta |\nu_1\rangle + \cos\theta |\nu_2\rangle ,$$

where $|\nu_\alpha\rangle$ and $|N\rangle$ 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 neutrino. This mixing should be weak, $\theta \ll 1$, otherwise sterile neutrinos would decay too rapidly, see

below. The heavy state is mostly the sterile neutrino $|\nu_2\rangle \approx |N\rangle$, and $m_2 \equiv m_N$ is the sterile neutrino mass.

The calculation of the sterile neutrino abundance is fairly complicated, and we do not reproduce it here. If there is no sizeable lepton asymmetry in the Universe, the estimate is

$$\Omega_N \simeq 0.2 \cdot \left(\frac{\sin\theta}{10^{-4}}\right)^2 \cdot \left(\frac{m_N}{1\,\mathrm{keV}}\right)^2 . \tag{59}$$

The energy spectrum of sterile neutrinos is nearly thermal. Thus, sterile neutrino of mass $m_\nu \gtrsim 1\,\mathrm{keV}$ 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, the sterile neutrino can decay into an active neutrino and a photon, see Fig. 27,

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

The sterile neutrino decay width is proportional to $\sin^2\theta$. If sterile neutrinos are dark matter particles,
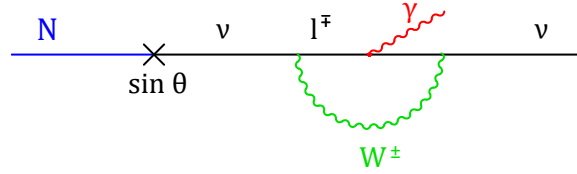
Fig. 27: Sterile neutrino decay $N \to \nu_\alpha + \gamma$.

their decays would produce a narrow line in X-ray flux from the cosmos (the orbiting velocity of dark matter particles in galaxies is small, $v \lesssim 10^{-3}$, hence the photons produced in their two-body decays are nearly monochromatic). Leaving aside a hint towards a 3.5 keV line advocated in Refs. [68,69] (see the discussion of its status in Ref. [70]), one makes use of strong limits on such a line and translates them into limits on $\sin^2\theta$. These limits as function of the sterile neutrino mass are shown in Fig. 28; they rule out the range of masses that are giving the right mass density of dark matter, Eq. (59).

A (rather baroque) way out [71] is to assume that there is a fairly large lepton asymmetry in the Universe. Then the oscillations of an active neutrino into a sterile neutrino may be enhanced due to the 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 the X-ray bounds. This is also shown in Fig. 28.

Direct laboratory searches for a sterile neutrino are currently sensitive to substantially larger sterile-active mixing angles. This is shown in Fig. 29 and also in Fig. 28, projected KATRIN limit, dashed line.

9 Dark matter summary

In the first place, 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 dark matter candidates. Other dark matter candidates include gravitinos, Q-balls, very heavy relics produced towards the end of inflation (wimpzillas), primordial black holes, etc. Hence, even though there are grounds to hope that the dark matter problem will be solved soon, there is no guarantee at all. Indeed, some of the candidates, like a gravitino or a sterile wimpzilla, interact with Standard Model particles so weakly that their direct discovery is hopeless. Concerning the candidates we have presented, we make a few comments:
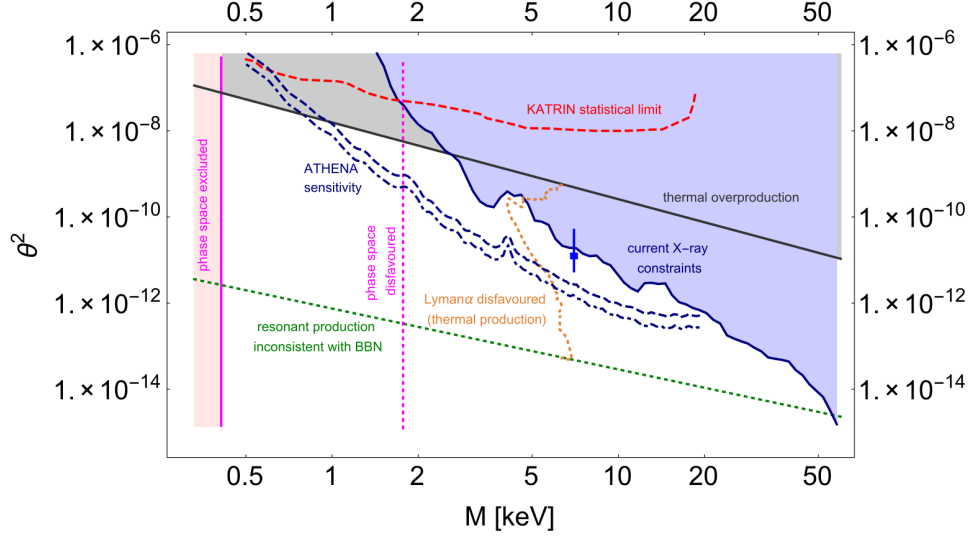
Fig. 28: Limits on sterile neutrino parameters (mass M, mixing angle squared θ^2) obtained from X-ray telescopes [70]. The straight solid line refers to sterile neutrino dark matter produced in non-resonant oscillations, Eq. (59). The region between this line and the dotted line corresponds to the resonant mechanism that works in a Universe with a fairly large lepton asymmetry. Vertical lines show very conservative limits coming from phase space and Lyman-α considerations, see Section 5.1. Regions left of these lines are disfavored. In fact, for a non-resonant mechanism, the phase space constraint is $M \gtrsim 6$ keV. The bullet with vertical interval shows the point corresponding to a putative 3.5 keV line.
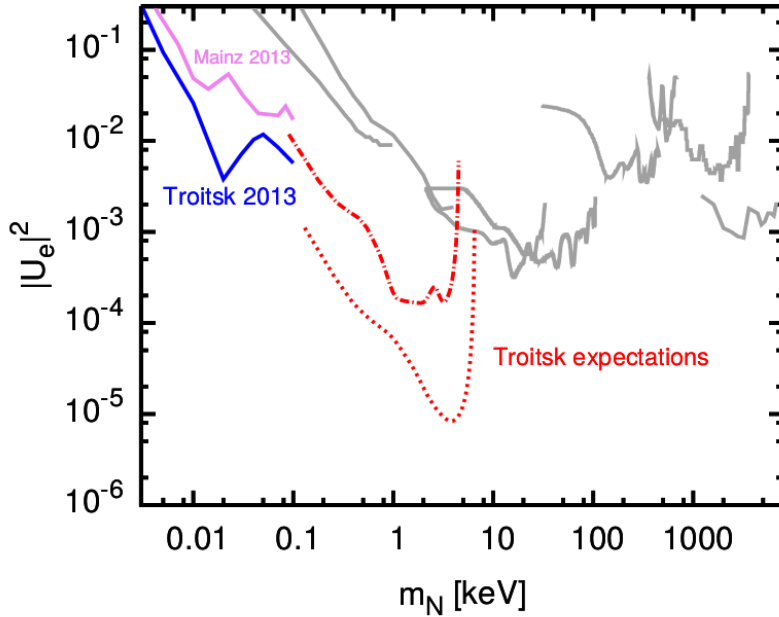
Fig. 29: Existing laboratory limits on a sterile neutrino mixing with an electron neutrino, $|U_e|^2 = \theta^2_{N\nu_e}$, and projected sensitivity of the Troitsk nu-mass experiment [72].

– With the exception of axions/ALPs, the plausible candidates are strongly constrained already. However, as we pointed out, this does not mean much, since the actual values of model parameters may still be in the unexplored region of the parameter space.

– The null results obtained so far suggest that it makes sense to look for less motivated candidates, and employ diverse search strategies. This happens already: we note in this regard existing and proposed experiments like NA64, SHiP, Troitsk nu-mass, Katrin, etc.

– Astrophysics and cosmology may well provide hints towards the nature of dark matter (CDM vs WDM vs SIMP vs fuzzy DM, etc.)

– WIMPs are attacked from different directions. 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 $10^{-9} - 10^{-11}$ s (compare to 1 MeV and 1 s accessible today through Big Bang nucleosynthesis). With microscopic physics to be known from collider experiments, the WIMP abundance 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.

– Searches for dark matter axions, ALPs or a signal from a light sterile neutrino make use of completely different methods. Yet there is a good chance for discovery, if either of these particles make dark matter.

All this shows that the situation with dark matter is controversial but extremely interesting.

10   Baryon asymmetry of the Universe

As we discussed in Section 2.6, the baryon asymmetry of the Universe is characterized by the baryon-to-entropy ratio, which at high temperatures is defined as follows,

$$\Delta_{\mathrm{B}} = \frac{n_{\mathrm{B}} - n_{\bar{\mathrm{B}}}}{s} = \frac{1}{3} \frac{n_{\mathrm{q}} - n_{\bar{\mathrm{q}}}}{s} \, ,$$

where n_{q} and $n_{\bar{\mathrm{q}}}$ are the number densities of quarks and antiquarks, respectively (baryon number of a quark equals $1/3$), 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), Δ_B is time-independent and equal to its present value $\Delta_B \approx 0.86 \cdot 10^{-10}$, see Eq. (24). At early times, at temperatures well above 100 MeV, the cosmic plasma contained many quark-antiquark pairs, whose number density was of the order of the entropy density, $n_{\mathrm{q}} + n_{\bar{\mathrm{q}}} \sim s$. Hence, in terms of quantities characterizing the very early epoch, the baryon asymmetry may be expressed as

$$\Delta_B \sim \frac{n_{\mathrm{q}} - n_{\bar{\mathrm{q}}}}{n_{\mathrm{q}} + n_{\bar{\mathrm{q}}}} \, .$$

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 would be very contrived, however. Furthermore, the inflationary scenario predicts that the Universe was baryon-symmetric at inflation (no quarks, no antiquarks). Hence, the baryon asymmetry must be explained dynamically [73, 74], by some mechanism of its generation in the early Universe.

10.1   Sakharov conditions

There are three necessary conditions for the generation of a baryon asymmetry from an initially baryon-symmetric state. These are the Sakharov conditions:

(i) baryon number non-conservation;

(ii) C- and CP violation;

(iii) deviation from a thermal equilibrium.

All three conditions are easily understood. (i) If baryon number were conserved, and the initial net baryon number in the Universe vanishes, the Universe today would still be baryon-symmetric. (ii) 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. (iii) 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 the 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 (with a qualification, see below).

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[12]. 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 [75]). 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 the thermal equilibrium with conserved $(B - L)$ one has

$$B = C \cdot (B - L) , \qquad L = (C - 1) \cdot (B - L) , \tag{60}$$

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.

There are two most commonly discussed mechanisms of baryon number non-conservation. One emerges in Grand Unified Theories and is due to the exchange of super-massive particles. The scale of these new, baryon number violating interactions is the Grand Unification scale, presumably of the order of $M_{GUT} \simeq 10^{16}$ GeV. It is not very likely, however, that the baryon asymmetry was generated due to this mechanism: the relevant temperature would have to be of order M_{GUT}, and so a high reheat temperature after inflation is difficult to obtain.

Another mechanism is non-perturbative [38] and is related to the triangle anomaly in the baryonic current (a keyword here is "sphaleron" [76, 77]). 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 $(B - L)$. A detailed analysis can be found in the book [78] and in references therein, as well as in the lecture notes of a similar School [31]. Here we only sketch its main ingredients.

10.2 Electroweak baryon number non-conservation

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 all quark flavors. Naively, it is conserved, but at the quantum level its divergence is non-zero because of the triangle anomaly (we discussed similar effect in the QCD context in

---

[12]Masses of neutrinos, if Majorana, violate lepton number. This effect, however, is by itself negligible.

Section 7.1; there, the axial current J_A^μ is not conserved even in the chiral limit),

$$\partial_\mu B^\mu = \frac{1}{3} \cdot 3_{colors} \cdot 3_{generations} \cdot \frac{g^2}{16\pi^2} F_{\mu\nu}^a \tilde{F}^{a\,\mu\nu} \,,$$

where $F_{\mu\nu}^a$ and g are the field strength of the $SU(2)_W$ gauge field and the $SU(2)_W$ gauge coupling, respectively, and $\tilde{F}^{a\,\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}^a$ is the dual tensor, cf. Eq. (45). Likewise, each leptonic current ($\alpha = e, \mu, \tau$) is anomalous in the Standard Model (we disregard here neutrino masses and mixings, which violate lepton numbers too),

$$\partial_\mu L_\alpha^\mu = \frac{g^2}{16\pi^2} F_{\mu\nu}^a \tilde{F}^{a\,\mu\nu}. \tag{61}$$

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

$$Q \equiv \int d^3x\, dt\, \frac{g^2}{16\pi^2} \cdot F_{\mu\nu}^a \tilde{F}^{a\,\mu\nu} \neq 0 \,. \tag{62}$$

Furthermore, for any physically relevant fluctuation, the value of Q is 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 in the end and beginning of the process are different,

$$B_{fin} - B_{in} = \int d^3x\, dt\, \partial_\mu B^\mu = 3Q \,. \tag{63}$$

Likewise

$$L_{\alpha,\,fin} - L_{\alpha,\,in} = Q \,. \tag{64}$$

This explains the selection rule mentioned above: B is violated, $(B - L) \equiv (B - \sum_\alpha L_\alpha)$ is not.

At zero temperature, the field fluctuations that induce baryon and lepton number violation are vacuum fluctuations are called instantons [79]. Since these are *large* field fluctuations, their probability is exponentially suppressed. The suppression factor in the Standard Model is[13]

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

Therefore, the rate of baryon number violating processes at zero temperature is 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. [80] for details)

$$\langle\phi\rangle_T < T \,, \tag{65}$$

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

10.3 Electroweak baryogenesis: what can make it work

Rapid electroweak baryon number non-conservation at high temperatures appears to open up an intriguing possibility that the baryon asymmetry was generated just by these electroweak processes. This should occur at electroweak temperatures, $T_{EW} \sim 100$ GeV, since whatever baryon asymmetry is generated by electroweak processes at higher temperatures, it would be washed out by the same processes as the Universe cools down to T_{EW}. There are two obstacles, however:

---

[13]Similar fluctuations of the gluon field in QCD are not suppressed, since QCD is strongly coupled at low energies. This explains why the axial current J_A^μ is not conserved, even approximately.

- CP violation (2nd Sakharov condition) is too weak in the Standard Model: the CKM mechanism alone is insufficient to generate an realistic value of the baryon asymmetry.
- Departure from thermal equilibrium (3d Sakharov condition) is problematic as well. At temperatures of order $T_{EW} \sim 100$ GeV, the Universe expands very slowly: the cosmological time scale at these temperatures,

$$H^{-1}(T_{EW}) = \frac{M_{\text{Pl}}^*}{T_{EW}^2} \sim 10^{-10} \text{ s} ,\tag{66}$$

is very large by the electroweak physics standards.

Let us discuss what can make the electroweak mechanism work. We begin with the second obstacle. It appears that the only way to have strong departure from a thermal equilibrium at $T_{EW} \sim 100$ GeV is a first order phase transition. Indeed, at temperatures well above 100 GeV electroweak symmetry is restored, and the expectation value of the Englert–Brout–Higgs field ϕ is zero, while it is non-zero in vacuum.
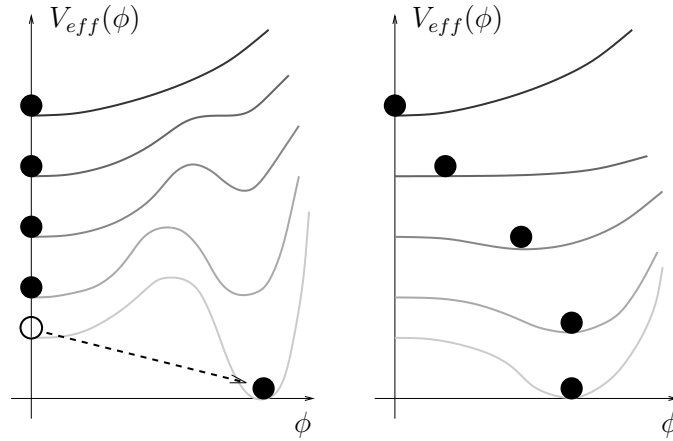
Fig. 30: Effective potential as function of ϕ at different temperatures. Left: first order phase transition. Right: second order phase transition. Upper curves correspond to higher temperatures. Black blobs show the expectation value of ϕ in thermal equilibrium. The arrow in the left panel illustrates the transition from the metastable, supercooled state to the ground state.

This suggests that there may be a phase transition from the phase with $\langle\phi\rangle = 0$ to the phase with $\langle\phi\rangle \neq 0$. In fact, the situation is 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. This is similar to a liquid-vapor system, which does not have an order parameter and, depending on the pressure, may or may not undergo a vapor-liquid phase transition as the temperature decreases.

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 $V_{\text{eff}}(\phi; T)$, which is equal to the free energy density in the system under the constraint that the average field is equal to a prescribed value ϕ, but otherwise there is thermal equilibrium. Then the global minimum of V_{eff} at a given temperature is at the equilibrium value of ϕ, while local minima correspond to metastable states.

The interesting case for us is a first order phase transition. In this case, the system evolves as follows. At high temperatures, there exists one minimum of V_{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 panel of Fig. 30. However, the minima
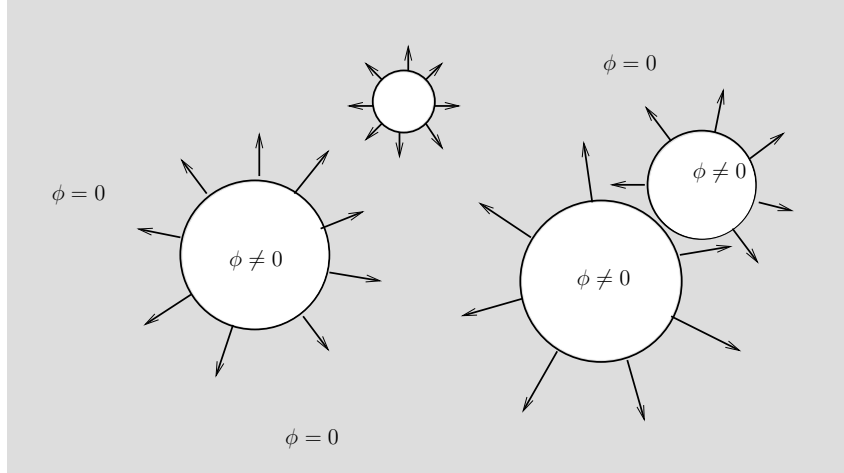
Fig. 31: First order phase transition: a boiling Universe.

with $\phi = 0$ and $\phi \neq 0$ are separated by a barrier of V_{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 panel of Fig. 30. 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.

The dynamics of a first order phase transition is highly inequilibrium. 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, Fig. 31. 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 a 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, a factor from 0.1 to 0.01). Hence, the bubbles grow large before their walls collide: their size at collision is roughly of order of the Hubble size (in fact, one or two orders of magnitude smaller). In other words, the bubbles are born microscopic, their initial sizes are determined by the electroweak scale and are roughly of order

$$R_{init} \sim (100 \, \text{GeV})^{-1} \sim 10^{-16} \, \text{cm} .$$

Their final sizes at the time the bubble walls collide are of order

$$R_{fin} \sim 10^{-2} - 10^{-3} \, \text{cm} ,$$

as follows from (66). One may hope that the baryon asymmetry may be generated during this inequilibrium process.

Does this really happen in the Standard Model? Unfortunately, no: with the Higgs boson mass $m_H = 125$ GeV, there is no phase transition in the Standard Model at all; there is smooth crossover instead [81].

Nevertheless, the first order phase transition may be characteristic of some extensions of the Standard Model. Generally speaking, one needs the existence of new bosonic fields that have large enough couplings to the Englert–Brout–Higgs field(s). To have an effect on the dynamics of the transition, the new bosons must be present in the cosmic plasma at the transition temperature, $T_{EW} \sim 100$ GeV, so their masses should not be very much higher than T_{EW}.

Let us turn to the first obstacle, CP violation. In the course of the first order phase transition, the baryon asymmetry is generated in the interactions of quarks and leptons with the bubble walls. There-

176

fore, CP violation must occur at the walls. Now, the walls are made of the scalar field(s), and 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.

In concrete models with successful electroweak baryogenesis, CP violation responsible for the baryon asymmetry often leads to sizeable electric dipole moments (EDMs) of neutron and electron. The limits on EDMs are so strong that many such models are actually ruled out. An example is the non-minimal split supersymmetric Standard Model, for which only a few years ago successful electroweak baryogenesis [82] was demonstrated. The predictions of this model for electron EDM are shown in Fig. 32. In 2016, when Ref. [82] was written, part of the parameter space was still allowed, but the recent ACME limit [83]

$$d_e < 1.1 \cdot 10^{-29} e \cdot \text{cm}$$

rules out the entire parameter space with efficient electroweak baryogenesis.
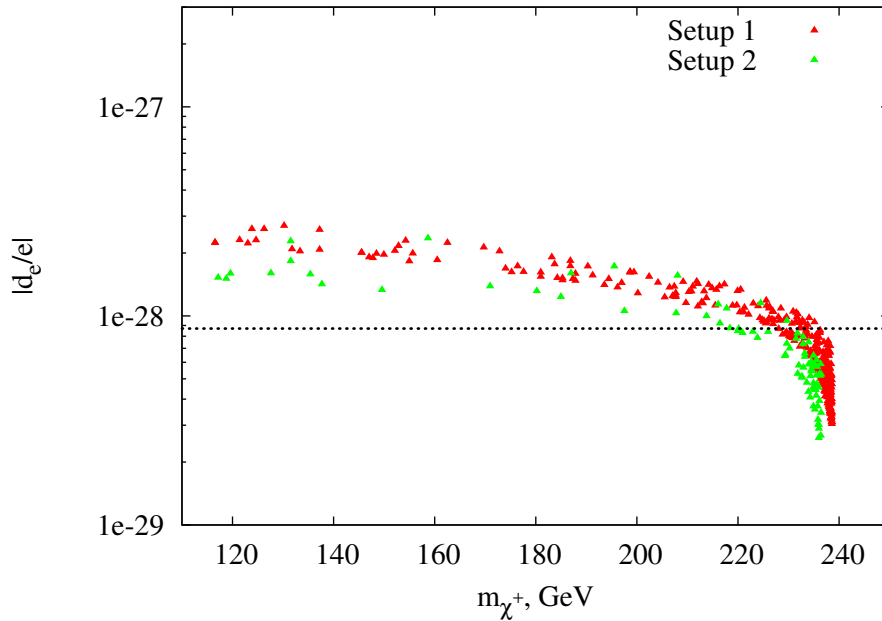
Fig. 32: Electron EDM predicted by the Non-Minimal Split Supersymmetric Standard Model with parameters suitable for electroweak baryogenesis. The current limit $d_e < 1.1 \cdot 10^{-29} e \cdot \text{cm}$ rules out all these models.

To summarize, electroweak baryogenesis requires a considerable extension of the Standard Model, often with masses of new particles in the TeV range or lower. Hence, this mechanism will most likely be ruled out or confirmed by the LHC or its successors. Moreover, limits on electron and neutron EDMs make the design of such an extension very difficult. Still, the issue is not decided yet, and the effort to construct the models with successful electroweak baryogenesis continues [84].

10.4 Baryogenesis in sterile neutrino oscillations

Let us mention another baryogenesis mechanism interesting from the viewpoint of terrestrial experiments, namely, leptogenesis in oscillations of sterile neutrinos [86, 87]. The general idea of leptogenesis [85] is that one or another mechanism generates a *lepton* asymmetry in the Universe before the electroweak transition, and electroweak sphalerons automatically reprocess part of the lepton asymmetry into a baryon asymmetry, see Eq. (60). The particular version of leptogenesis that we briefly discuss here assumes that there are at least two heavy Majorana neutrinos in the mass range $1 - 10$ GeV, and that

there is strong enough CP violation in the sterile neutrino sector. Then asymmetries in the sterile neutrino sector may be generated and transmitted to active neutrino sector via Yukawa interactions responsible for see-saw masses of the active neutrinos. In the case when there are effectively two sterile neutrino species participating in leptogenesis, the correct value of the baryon asymmetry is obtained when the two sterile neutrinos are nearly degenerate,

$$\frac{|M_1^2 - M_2^2|}{M_{1,2}^2} \lesssim 10^{-6} \; ,$$

which makes the model rather contrived. However, with three sterile neutrino species, the degeneracy is no longer required [88]. The sterile neutrinos of masses in the GeV range and parameters suitable for leptogenesis in their oscillations are typically accessible through rare decays of B-mesons, Z-bosons, as well as in future beam dump experiments such as SHiP.

An important point concerning this and virtually all other leptogenesis mechanisms is that CP violation in the sector of active neutrinos, which will hopefully be discovered in oscillation experiments, does not have direct relevance to the leptogenesis: the value of lepton and hence baryon asymmetry is determined by the CP-violating parameters *in the sterile neutrino sector*.

10.5   Baryogenesis summary

We briefly considered here two mechanisms of baryogenesis which may be directly tested, at least in principle, in particle physics experiments. These are certainly not the only mechanisms proposed, and, arguably, not the most plausible mechanisms. One particularly strong competitor is thermal leptogenesis [85], for reviews see, e.g., Ref. [89]. Its idea is that the lepton asymmetry is generated in decays of heavy Majorana sterile neutrinos. The masses of these new particles are well above the experimentally accessible energies. On the one hand, this is in line with the see-saw idea; on the other, direct proof of this mechanism does not appear possible. Interestingly, thermal leptogenesis works only with light active neutrinos: the neutrino masses inferred from cosmology and oscillation experiments are just in the right ballpark.

There are numerous alternative mechanisms of baryogenesis. To name a few, we have already mentioned the Affleck–Dine baryogenesis [75]; early discussions concentrated mostly on GUT baryogenesis [90]; there is even a possibility to generate the baryon asymmetry at the inflationary epoch [91]. Unfortunately, most of these proposals will be very difficult, if at all possible, to test. So, there is no guarantee at all that we will understand in foreseeable future the origin of matter in the Universe.

11   Before the hot epoch

With the Big Bang nucleosynthesis 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 age of $t \simeq 1$ s. With the LHC, we are learning the Universe up to temperatures $T \sim 100$ GeV and down to an age of $t \sim 10^{-10}$ s. Are we going to have a handle on even earlier epoch?

Let us summarize the current status of this issue.

– On the one hand, we are confident that the hot cosmological epoch was not the first one; it was preceded by some other, entirely different stage.
– On the other hand, we do not know for sure what was that earlier epoch; an excellent guess is inflation, but alternative scenarios are not ruled out.
– It is conceivable (although not guaranteed) that future cosmological observations will enable us to understand the nature of the pre-hot epoch.

All this makes the situation very interesting. It is fascinating that by studying the Universe at large we may be able to learn about the earliest cosmological epoch which happened at an extremely high energy density and expansion rate of our Universe.

11.1 Cosmological perturbations

The key players in this Section are cosmological perturbations. These are inhomogeneities in the energy density and associated gravitational potentials, in the first place. It is these inhomogeneities that, among other things, serve as seeds for structures—galaxies, clusters of galaxies, etc. This type of inhomogeneities is called scalar perturbations, as they are described by 3-scalars. There may exist perturbations of another type, called tensor; these are primordial gravity waves. Tensor modes have not been observed (yet), so we mostly concentrate on scalar perturbations. While perturbations of a present size of the order of 10 Mpc and smaller have large amplitudes today and are non-linear, amplitudes of all known perturbations were small in the past, and a linearized theory is applicable. Indeed, CMB temperature anisotropy tells us that the perturbations at the recombination epoch were roughly at the level of

$$\delta \equiv \frac{\delta\rho}{\rho} = 10^{-4} - 10^{-5} \; . \tag{67}$$

We are sloppy here in characterizing the scalar perturbations by the density contrast $\delta\rho/\rho$; we are going to skip technicalities and use this notation in what follows.

Linearized perturbations are most easily studied in momentum space, since the background FLRW metric in Eq. (1) does not explicitly depend on $\mathbf{x}$. The spatial Fourier transformation reads

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

Each Fourier mode $\delta(\mathbf{k}, t)$ obeys its own linearized equation and hence can be treated separately. Note that the physical distance between neighboring points is $a(t)d\mathbf{x}$. 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 redshifted as the Universe expands, see also Section 2.1. In what follows we set the present value of the scale factor equal to 1, $a_0 \equiv a(t_0) = 1$; then $\mathbf{k}$ is the *present* physical momentum and $2\pi/k$ is the present physical wavelength, which is also called comoving wavelength.

Properties of scalar perturbations are measured in various ways. Perturbations of fairly large spatial scales (fairly low $\mathbf{k}$) give rise to a 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 the 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 reasonably 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, 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.

11.2 Subhorizon and superhorizon regimes

It is instructive to compare the wavelength of a perturbation with the horizon size. To this end, recall (see Section 2.6) that the horizon size $l_{\mathrm{H}}(t)$ is the size of the largest region which is causally connected by the time t, and that

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

at the radiation domination epoch and later, see Eq. (18). 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 at the heart of what can be called the 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.

The physical wavelength of a perturbation grows slower than the horizon size. As an example, at the radiation domination epoch

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

while at the matter domination epoch $\lambda(t) \propto t^{2/3}$. For an obvious reason, the modes with $\lambda(t) \ll H^{-1}(t)$ and $\lambda(t) \gg H^{-1}(t)$ are called subhorizon and superhorizon at the time t, respectively. We are interested in the modes which are subhorizon *today*; longer modes are homogeneous throughout the visible Universe and are not observed. 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 the 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 $k/a(t) \gg H(t)$. It is straightforward to see that for all cosmologically interesting wavelengths, the horizon crossing occurs at temperatures below 1 MeV, i.e., at the time we are confident about (repeating the calculation of Section 5.1 we find that the present wavelength of order 100 kpc entered the horizon at $T \sim 4$ keV). So, there is no guesswork at this point.

Another way to look at the superhorizon–subhorizon behaviour of perturbations is to introduce a new time coordinate (cf. Eq. (16)),

$$\eta = \int_0^t \frac{dt'}{a(t')} \;. \tag{68}$$

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

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

In coordinates $(\eta, \mathbf{x})$, the light cones $ds = 0$ are the same as in Minkowski space, and η is the coordinate size of the horizon, see Fig. 33. Every mode of perturbation has the time-independent coordinate wavelength $2\pi/k$, and at small η it is in the superhorizon regime, $2\pi/k \gg \eta$.

11.3 Hot epoch was not the first

This picture falsifies the hot Big Bang theory. Indeed, within this theory, we see the horizon at recombination $l_{\mathrm{H}}(t_{\mathrm{rec}})$ at an angle $\Delta\theta \approx 2°$, as schematically shown in Fig. 33. 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, the CMB temperature must be isotropic when averaged over angular scales exceeding $2°$; there should be no cold or warm regions of an angular size larger than $2°$.

We now take a look at the CMB photographic picture shown in Fig. 2. It is seen by naked eye that 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, i.e., the cosmological perturbations were generated before the hot epoch.

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 the CMB temperature, as seen in

Fig. 3. In general, perturbations in the baryon-photon medium before recombination are acoustic waves (cf. Section 4.3),

$$\delta_B(\mathbf{k}, t) = A(\mathbf{k})e^{i\mathbf{k}\mathbf{x}} \cos\left[\int_0^t v_s \frac{k}{a(t')}dt' + \psi_{\mathbf{k}}\right] , \tag{69}$$

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

$$\delta_B(t) = \text{const} , \tag{70a}$$

$$\delta(t)_B = \frac{\text{const}}{t^{3/2}} . \tag{70b}$$

If the perturbations existed at the very beginning of the hot epoch, they were in the superhorizon regime at sufficiently early times, and were described by the solutions in Eq. (70). The consistency of the whole cosmology requires that the amplitude of the perturbations was small at the beginning of the hot stage. The solution in Eq. (70b) rapidly decays away, and towards the horizon entry the perturbation is in constant mode in Eq. (70a). So, the initial condition for the further evolution is unique modulo the amplitude $A(\mathbf{k})$, and hence the phase $\psi(\mathbf{k})$ is uniquely determined: we have $\psi(\mathbf{k}) = 0$ for modes entering horizon at the radiation domination epoch. As discussed in Section 4.3, this leads to oscillatory behavior of baryon-photon perturbations at the recombination epoch *as function of* k, and translates into oscillations of the CMB temperature multipole C_l as function of multipole number l.

Were the perturbations generated in a causal way at the radiation domination epoch, they would be always in the subhorizon regime. In that case the solutions in Eq. (70) would be irrelevant, and there would be no reason for a particular choice of phase $\psi_{\mathbf{k}}$ in Eq. (69). One would rather expect that $\psi_{\mathbf{k}}$ is a random function of $\mathbf{k}$, so $\delta_B(\mathbf{k}, t_r)$ would not oscillate as function of $\mathbf{k}$, and oscillations of C_l would not exist. This is indeed the case for specific mechanisms of the generation of density perturbations at hot epoch [92].

We conclude that the facts that the CMB angular spectrum has oscillatory behavior and that there are sizeable temperature fluctuations at $l < 50$ (angular scale greater than the angular size $2°$ of the
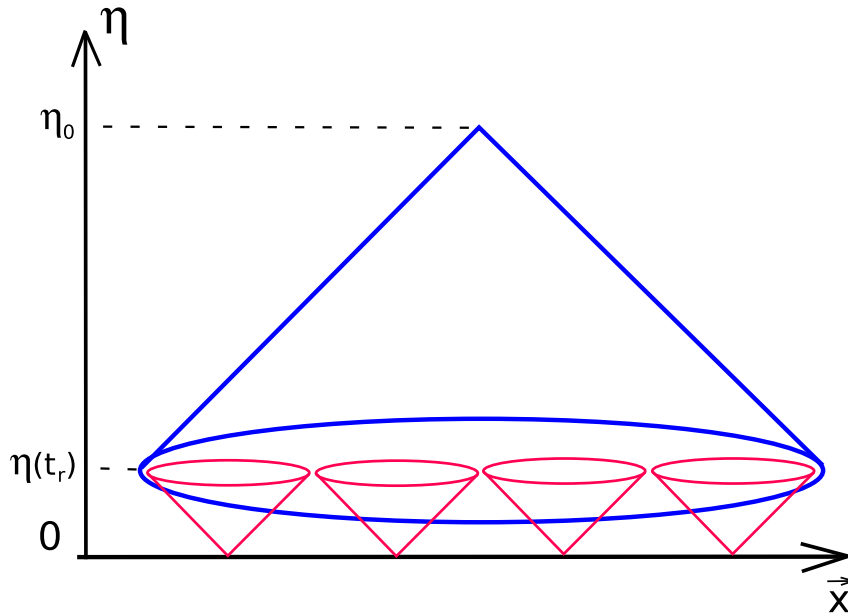
Fig. 33: Causal structure of space-time in the hot Big Bang theory. η_r and η_0 are conformal times at recombination and today, respectively.

horizon at the recombination epoch) unambiguously tell us that the density perturbations were indeed in the superhorizon regime at the hot cosmological stage. The hot epoch was preceded by some other epoch—the epoch of the generation of perturbations.

11.4 Inflation or not?

The pre-hot epoch must be long in terms of the time variable η introduced in Eq. (68). What we would like to have is that the large part of the Universe be causally connected towards the end of that epoch, see Fig. 34. Long duration in η does not necessarily mean long duration in physical time t; in fact, the
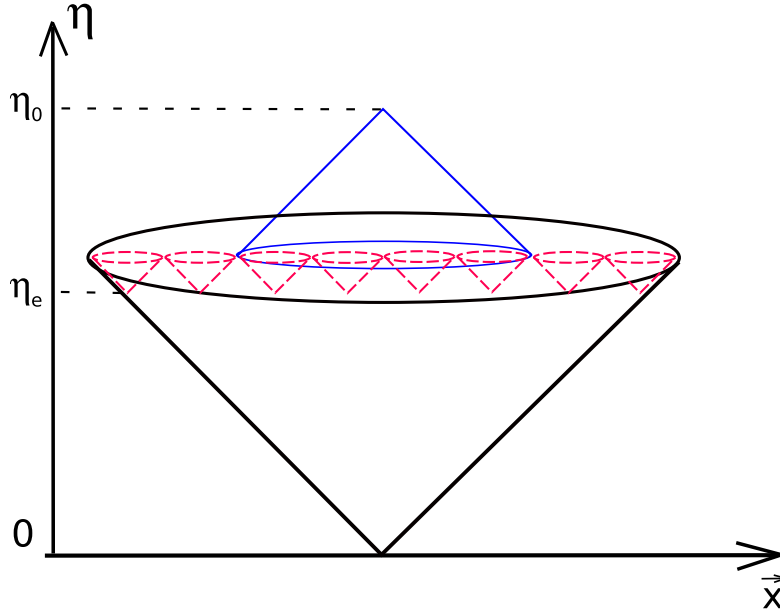
Fig. 34: Causal structure of space-time in the real Universe

pre-hot epoch may be very short in physical time.

An excellent hypothesis on the pre-hot stage is inflation, the epoch of nearly exponential expansion [93],

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

If this epoch lasts many Hubble times, the whole visible Universe, and likely much a greater region of space, is causally connected already at very early times.

From the viewpoint of perturbations, the physical momentum $q(t) = k/a(t)$ decreases (gets redshifted) at the time of inflation, while the Hubble parameter stays almost constant. So, every mode is first in the subhorizon regime ($q(t) \gg H(t)$), and later in the superhorizon regime ($q(t) \ll H(t)$). This situation is opposite to what happens at the radiation and matter domination epoch; this is precisely the pre-requisite for generating the density perturbations. Indeed, inflation does generate primordial density perturbations [94], whose properties are consistent with everything we know about them.

Inflation is not the only hypothesis proposed so far. One alternative 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 Ref. [95] for reviews. Another scenario is that the Universe starts out from 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 [96]. Theoretical realizations of these scenarios are surprisingly difficult, but not impossible, as became clear recently.

12 Towards understanding the earliest epoch

Since cosmological perturbations originate from the earliest epoch that occurred before the hot stage, properties of these perturbations will hopefully give us a clue on that epoch. Presently, we know only very basic things about the cosmological perturbations. Let us discuss this point, and at the same time consider promising directions where further study may lead to breakthrough.

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.

12.1 Adiabaticity of scalar perturbations

Primordial scalar 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 the dark matter to entropy ratio are constant in space,

$$\delta \left(\frac{n_B}{s} \right) = \text{const} , \qquad \delta \left(\frac{n_{\text{DM}}}{s} \right) = \text{const} . \tag{71}$$

This is consistent with the generation of the baryon asymmetry and of dark matter at the hot cosmological epoch: in that case, all particles were in thermal equilibrium early at the hot epoch, 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* (or isocurvature) perturbations that violate (one of) the relations in Eq. (71). No admixture of the entropy perturbations have been detected so far, but it is worth emphasizing that even 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 7.

12.2 Gaussianity

The primordial scalar perturbations are a **Gaussian random field**. Gaussianity means that the three-point and all odd correlation functions vanish, while the four-point and 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*. A free quantum field has the general form

$$\phi(\mathbf{x}, t) = \int d^3 k \, e^{-i\mathbf{k}\mathbf{x}} \left(f_{\mathbf{k}}^{(+)}(t) a_{\mathbf{k}}^\dagger + e^{i\mathbf{k}\mathbf{x}} f_{\mathbf{k}}^{(-)}(t) a_{\mathbf{k}} \right) ,$$

where $a_{\mathbf{k}}^\dagger$ and $a_{\mathbf{k}}$ are creation and annihilation operators. For the field in Minkowski space-time one has $f_{\mathbf{k}}^{(\pm)}(t) = e^{\pm i\omega_k t}$, while enhancement, e.g. due to the evolution in the time-dependent background, means that $f_{\mathbf{k}}^{(\pm)}$ are large. But in any case, Wick's theorem is valid, provided that the state of the system is the vacuum, $a_{\mathbf{k}}|0\rangle = 0$. Hence, it is quite likely that the density perturbations originate from the enhanced vacuum fluctuations of non-interacting or weakly interacting quantum field(s).

Search for *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 4-point function) may be measurable instead. For the time being, non-Gaussianity has not been detected.

Inflation does the job of producing Gaussian primordial perturbations very well. At the inflationary epoch, fluctuations of all light fields get enhanced greatly due to the fast expansion of the Universe. This is true, in particular, for the inflaton, the field that dominates the energy density at the time of inflation. Enhanced vacuum fluctuations by inflaton are reprocessed into adiabatic 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 [97]. In fact, it is so small that its detection is problematic even in the distant future. It is worth noting that this refers to the simplest, single field inflationary models. In models with more than one relevant field the situation may be different, and sizeable non-Gaussianity may be generated.

The generation of the density perturbations is less automatic in scenarios alternative to inflation. Most models proposed so far can be adjusted in such a way that non-Gaussianity is not particularly strong, but potentially observable. In many cases the bispectrum $G(k_i^2;\ \mathbf{k}_1\mathbf{k}_2;\ \mathbf{k}_1\mathbf{k}_3)$ and/or trispectrum are different from inflationary theories.

12.3 Nearly flat power spectrum

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

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \frac{1}{4\pi k^3}\mathcal{P}(k)\delta(\mathbf{k} + \mathbf{k}') \ . \tag{72}$$

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

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

By definition, the flat, scale-invariant spectrum is such that $\mathcal{P}$ is independent of k. The flat spectrum was conjectured by Harrison [98], Zeldovich [99] and Peebles and Yu [100] in the beginning of 1970's, long before realistic mechanisms of the generation of density perturbations have been proposed.

In view of the approximate flatness, a natural parametrization is

$$\mathcal{P}(k) = A_s \left(\frac{k}{k_*} \right)^{n_s - 1} \ , \tag{73}$$

where A_s is the amplitude, $(n_s - 1)$ is the tilt and k_* is a fiducial momentum, chosen at one's convenience. The flat spectrum in this parametrization has $n_s = 1$. The cosmological data give [3]

$$n_s = 0.965 \pm 0.004 \ . \tag{74}$$

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

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

$$ds^2 = dt^2 - \mathrm{e}^{2Ht}d\mathbf{x}^2 \ , \qquad H = \text{const} \ .$$

184

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, as required for the flat power spectrum. At inflation, H and the inflaton field are almost constant in time, and the de Sitter symmetry is an approximate symmetry. For this reason inflation automatically generates nearly a flat power spectrum. However, neither H nor inflaton are exactly time-independent. This naturally leads to the slight tilt in the spectrum. Overall, this picture is qualitatively consistent with the result in Eq. (74), though the quantitative prediction depends on the concrete inflationary model.

The situation is not so straightforward in alternatives to inflation: the approximate flatness of the scalar power spectrum is not at all automatic. So, one has to work hard to obtain this property. Similarly to an inflationary theory, the flatness of the scalar power spectrum may be due to some symmetry. One candidate symmetry is conformal invariance [101, 102]. The point is that the conformal group includes dilatations, $x^\mu \to \lambda x^\mu$. This property indicates that the theory possesses no scale, and has good chance for producing the flat spectrum. This idea is indeed realized at least at the toy model level.

12.4 Statistical isotropy

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

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

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 $\delta(\mathbf{k})$). Such a dependence would 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]. On the other hand, statistical anisotropy is natural in some other scenarios, including conformal models [104]. The statistical anisotropy would show up in correlations [105]

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

At the moment, the constraints [106] 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.

12.5 Tensor modes

The distinguishing property of inflation is *the generation of tensor modes (primordial gravity waves)* of sizeable amplitude and a nearly flat power spectrum. The gravity waves are thus smoking guns for inflation (although there is some debate on this point). Indeed, 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 modes. The reason for their generation at the time of 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, and periods are of the order of a billion years 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 the radiation domination epoch, and today they have much smaller amplitudes making them inaccessible to gravity wave detectors like LIGO/VIRGO, eLISA, etc. A conventional characteristic of the amplitude of primordial gravity waves is the tensor-to-scalar ratio

$$r = \frac{\mathcal{P}_T}{\mathcal{P}} \ ,$$

where $\mathcal{P}$ is the scalar power spectrum defined in Eq. (72) and $\mathcal{P}_T$ is the tensor power spectrum defined in a similar way, but for transverse traceless metric perturbations h_{ij}.

Until recently, the most sensitive probe of the tensor perturbations has been the CMB temperature anisotropy [107]. Nowadays, the best 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 [108]. 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 result of the search for effects of the tensor modes on the CMB temperature anisotropy is shown in Fig. 35. This search has already ruled out some of the popular inflationary models.
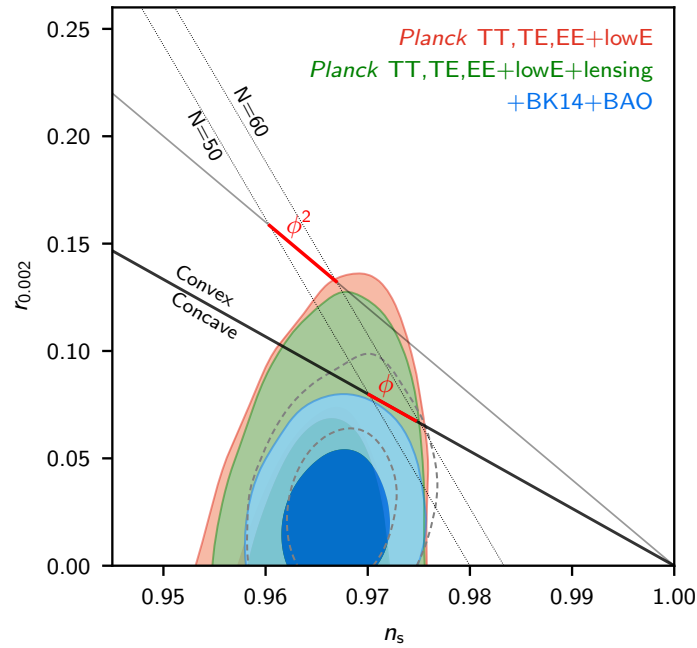
Fig. 35: Allowed regions (at 68% and 95% CL) in the plane (n_s, r), where n_s is the scalar spectral index and r is the tensor-to-scalar ratio [3], obtained by the Planck collaboration alone and by combining the Planck data with the BAO data and CMB polarization data from the BICEP2/KEK experiments. The right corner (the point $(1.0, 0.0)$) is the Harrison–Zeldovich point (flat scalar spectrum, no tensor modes). Intervals show predictions of inflationary models with quadratic and linear inflaton potentials.

13 Conclusion

The present situations in particle physics, on one side, and cosmology, on the other, have much in common. The Standard Model of particle physics and the Standard Model of cosmology, ΛCDM, have been shaped. Both fields enjoyed fairly unexpected discoveries: neutrino oscillations and accelerated expansion of the Universe.

There is strong evidence that the two Standard Models are both incomplete. Therefore, in both fields one hopes for new, revolutionary discoveries. In the context of these lectures, we hope to learn what is the dark matter particle; we may learn the origin of the matter-antimatter asymmetry in the Universe;

the discoveries of new properties of cosmological perturbations will hopefully reveal the nature of the pre-hot epoch.

However, there is no guarantee of new discoveries in particle physics or cosmology. Nature may hide its secrets. Whether or not we will be able to reveal these secrets is the biggest open question in fundamental physics.

References

[1] S. Dodelson, *Modern cosmology,* Academic Press, Amsterdam, 2003, doi:10.1016/B978-0-12-219141-1.X5019-0; V. Mukhanov, *Physical foundations of cosmology,* Cambridge University Press, Cambridge, 2005, doi:10.1017/CBO9780511790553; S. Weinberg, *Cosmology*, Oxford University Press, Oxford, 2008, WorldCat; A.R. Liddle and D.H. Lyth, *The primordial density perturbation: Cosmology, inflation and the origin of structure,* Cambridge University Press, 2009, doi:10.1017/CBO9780511819209; D.S. Gorbunov and V.A. Rubakov, *Introduction to the theory of the early universe: Hot big bang theory,* 2nd ed., World Scientific, Singapore, 2018, doi:10.1142/10447; D.S. Gorbunov and V.A. Rubakov, *Introduction to the theory of the early universe: Cosmological perturbations and inflationary theory,* World Scientific, Singapore, 2011, doi:10.1142/7873.

[2] E. Di Valentino, A. Melchiorri and J. Silk, "Planck evidence for a closed universe and a possible crisis for cosmology", *Nature Astron.* **4** (2019) 196–203, doi:10.1038/s41550-019-0906-9, arXiv:1911.02087 [astro-ph.CO].

[3] N. Aghanim *et al.* [Planck Collaboration], "Planck 2018 results. VI. Cosmological parameters", *Astron. Astrophys.* **641** (2020), A6, doi:10.1051/0004-6361/201833910, Erratum: *Astron. Astrophys.* **652** (2021), C4, doi:10.1051/0004-6361/201833910e, arXiv:1807.06209 [astro-ph.CO].

[4] E. Gawiser and J. Silk "The cosmic microwave background radiation", *Phys. Rept.* **333** (2000) 245, doi:10.1016/S0370-1573(00)00025-9, arXiv:astro-ph/0002044.

[5] N.G. Busca *et al.*, "Baryon acoustic oscillations in the Lyα forest of BOSS quasars", *Astron. Astrophys.* **552** (2013) A96, doi:10.1051/0004-6361/201220724, arXiv:1211.2616 [astro-ph.CO].

[6] S. Weinberg, "Anthropic bound on the cosmological constant", *Phys. Rev. Lett.* **59** (1987) 2607, doi:10.1103/PhysRevLett.59.2607.

[7] A.D. Linde, "Inflation and quantum cosmology", in: *Three hundred years of gravitation*, Eds. S.W. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1987) pp. 604–630.

[8] C. Amsler *et al.* [Particle Data Group], *Phys. Lett.* **B667** (2008) 1, doi:10.1016/j.physletb.2008.07.018 and 2009 partial update for the 2010 edition https://pdg.lbl.gov/2009/.

[9] K.A. Olive, "Dark matter", arXiv:astro-ph/0301505; G. Bertone, D. Hooper and J. Silk, "Particle dark matter: Evidence, candidates and constraints", *Phys. Rept.* **405** (2005) 279, doi:10.1016/j.physrep.2004.08.031, arXiv:hep-ph/0404175; A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, "The role of sterile neutrinos in cosmology and astrophysics", *Ann. Rev. Nucl. Part. Sci.* **59** (2009) 191, doi:10.1146/annurev.nucl.010909.083654, arXiv:0901.0011 [hep-ph]; M. Kawasaki and K. Nakayama, "Axions: Theory and cosmological role", *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 69, doi:10.1146/annurev-nucl-102212-170536, arXiv:1301.1123 [hep-ph];

H. Baer *et al.*, "Dark matter production in the early universe: beyond the thermal WIMP paradigm", *Phys. Rept.* **555** (2015) 1, doi:10.1016/j.physrep.2014.10.002, arXiv:1407.0017 [hep-ph].

[10] L. Roszkowski, E.M. Sessolo and S. Trojanowski, "WIMP dark matter candidates and searches—current status and future prospects", *Rept. Prog. Phys.* **81** (2018) 066201, doi:10.1088/1361-6633/aab913, arXiv:1707.06277 [hep-ph].

[11] G. Arcadi *et al.*, "The waning of the WIMP? A review of models, searches, and constraints", *Eur. Phys. J.* **C78** (2018) 203, doi:10.1140/epjc/s10052-018-5662-y, arXiv:1703.07364 [hep-ph].

[12] J.S. Bullock and M. Boylan-Kolchin, "Small-scale challenges to the ΛCDM paradigm", *Ann. Rev. Astron. Astrophys.* **55** (2017) 343, doi:10.1146/annurev-astro-091916-055313, arXiv:1707.04256 [astro-ph.CO].

[13] K.G. Begeman, A.H. Broeils and R.H. Sanders "Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics", *Mon. Not. Roy. Astron. Soc.* **249** (1991) 523, doi:10.1093/mnras/249.3.523.

[14] J.P. Kneib *et al.*, "A wide field Hubble Space Telescope study of the cluster Cl0024+1654 at Z=0.4. II. The cluster mass distribution, *Astrophys. J.* **598** (2003) 804, doi:10.1086/378633, astro-ph/0307299.

[15] D. Clowe *et al.*, "A direct empirical proof of the existence of dark matter", *Astrophys. J.* **648** (2006) L109, doi:10.1086/508162, arXiv:astro-ph/0608407.

[16] A.D. Sakharov, "The initial stage of an expanding universe and the appearance of a nonuniform distribution of matter", *Zh. Eksp. Teor. Fiz.* **49** 345, English transl. in *Sov. Phys. JETP* **22** (1966) 241, http://jetp.ras.ru/cgi-bin/e/index/e/22/1/p241?a=list.

[17] D.J. Eisenstein *et al.* [SDSS Collaboration], "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies", *Astrophys. J.* **633** (2005) 560, doi:10.1086/466512, arXiv:astro-ph/0501171.

[18] S.Y. Kim, A.H.G. Peter and J.R. Hargis, "Missing satellites problem: Completeness corrections to the number of satellite galaxies in the Milky Way are consistent with cold dark matter predictions", *Phys. Rev. Lett.* **121** (2018) 211302, doi:10.1103/PhysRevLett.121.211302, arXiv:1711.06267 [astro-ph.CO].

[19] S. Shen *et al.* "The baryon cycle of dwarf galaxies: Dark, bursty, gas-rich polluters", *Astrophys. J.* **792** (2014) 99, doi:10.1088/0004-637X/792/2/99, arXiv:1308.4131 [astro-ph.CO].

[20] Y. Revaz and P. Jablonka, "Pushing back the limits: detailed properties of dwarf galaxies in a ΛCDM universe", *Astron. Astrophys.* **616** (2018) A96, doi:10.1051/0004-6361/201832669, arXiv:1801.06222 [astro-ph.GA].

[21] S. Garrison-Kimmel *et al.*, "Not so lumpy after all: modelling the depletion of dark matter subhaloes by Milky Way-like galaxies", *Mon. Not. Roy. Astron. Soc.* **471** (2017) 1709, doi:10.1093/mnras/stx1710, arXiv:1701.03792 [astro-ph.GA].

[22] A. Drlica-Wagner *et al.* [LSST Dark Matter Group], "Probing the fundamental nature of dark matter with the Large Synoptic Survey Telescope", arXiv:1902.01055 [astro-ph.CO]; K. Bechtol *et al.*, "Dark matter science in the era of LSST", arXiv:1903.04425 [astro-ph.CO].

[23] A. Boyarsky, O. Ruchayskiy and D. Iakubovskyi, "A lower bound on the mass of dark matter particles", *JCAP* **03** (2009) 005, doi:10.1088/1475-7516/2009/03/005, arXiv:0808.3902 [hep-ph].

[24] D. Gorbunov, A. Khmelnitsky and V. Rubakov, "Constraining sterile neutrino dark matter by phase-space density observations", *JCAP* **10** (2008) 041, doi:10.1088/1475-7516/2008/10/041, arXiv:0808.3910 [hep-ph].

[25] V. Iršič *et al.*, "New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data", *Phys. Rev.* **D96** (2017) 023522, doi:10.1103/PhysRevD.96.023522, arXiv:1702.01764 [astro-ph.CO].

[26] M.R. Lovell *et al.*, "Addressing the too big to fail problem with baryon physics and sterile neutrino dark matter", *Mon. Not. Roy. Astron. Soc.* **468** (2017) 2836, doi:10.1093/mnras/stx621, arXiv:1611.00005 [astro-ph.GA].

[27] D.N. Spergel and P.J. Steinhardt, "Observational evidence for selfinteracting cold dark matter", *Phys. Rev. Lett.* **84** (2000) 3760, doi:10.1103/PhysRevLett.84.3760, arXiv:astro-ph/9909386.

[28] L. Hui *et al.*, "Ultralight scalars as cosmological dark matter", *Phys. Rev.* **D95** (2017) 043541, doi:10.1103/PhysRevD.95.043541, arXiv:1610.08297 [astro-ph.CO].

[29] V. Iršič *et al.*, "First constraints on fuzzy dark matter from Lyman-α forest data and hydrodynamical simulations", *Phys. Rev. Lett.* **119** (2017) 031302, doi:10.1103/PhysRevLett.119.031302, arXiv:1703.04683 [astro-ph.CO];
M. Nori *et al.*, "Lyman α forest and non-linear structure characterization in fuzzy dark matter cosmologies", *Mon. Not. Roy. Astron. Soc.* **482** (2019) 3227, doi:10.1093/mnras/sty2888, arXiv:1809.09619 [astro-ph.CO].

[30] A. Khmelnitsky and V. Rubakov, "Pulsar timing signal from ultralight scalar dark matter", *JCAP* **02** (2014) 019, doi:10.1088/1475-7516/2014/02/019, arXiv:1309.5888 [astro-ph.CO];
N.K. Porayko *et al.*, "Parkes pulsar timing array constraints on ultralight scalar-field dark matter", *Phys. Rev.* **D98** (2018) 102002, doi:10.1103/PhysRevD.98.102002, arXiv:1810.03227 [astro-ph.CO].

[31] V.A. Rubakov, "Cosmology", *CERN Yellow Rep. School Proc.* **2** (2017) 239, doi:10.23730/CYRSP-2017-002.239, arXiv:1804.11230 [gr-qc].

[32] M. Cirelli, N. Fornengo and A. Strumia, "Minimal dark matter", *Nucl. Phys.* **B753** (2006) 178, doi:10.1016/j.nuclphysb.2006.07.012, arXiv:hep-ph/0512090;
M. Cirelli and A. Strumia, "Minimal dark matter: Model and results", *New J. Phys.* **11** (2009) 105005, doi:10.1088/1367-2630/11/10/105005, arXiv:0903.3381 [hep-ph].

[33] A. Bottino and N. Fornengo, "Dark matter and its particle candidates", pres. at the 5th ICTP School on Nonaccelerator Particle Astrophysics, Trieste, Italy, 29 Jun.–10 Jul. 1998, hep-ph/9904469.

[34] E. Aprile *et al.* [XENON Collaboration], "Dark matter search results from a one ton-year exposure of XENON1T", *Phys. Rev. Lett.* **121** (2018) 111302, doi:10.1103/PhysRevLett.121.111302, arXiv:1805.12562 [astro-ph.CO].

[35] M. Lisanti, "Lectures on dark matter physics", Proc. Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015), Boulder, CO, USA, 1–26 Jun. 2015, Eds. J. Polchinski, P. Vieira and O. DeWolfe (World Scientific, Singapore, 2017), pp. 399-446, doi:10.1142/9789813149441_0007, arXiv:1603.03797 [hep-ph].

[36] A.D. Avrorin *et al.* [Baikal Collaboration], "Search for neutrino emission from relic dark matter in the Sun with the Baikal NT200 detector", *Astropart. Phys.* **62** (2014) 12, doi:10.1016/j.astropartphys.2014.07.006, arXiv:1405.3551 [astro-ph.HE].

[37] L. Bergström *et al.*, "New limits on dark matter annihilation from AMS cosmic ray positron data", *Phys. Rev. Lett.* **111** (2013) 171101, doi:10.1103/PhysRevLett.111.171101, arXiv:1306.3983 [astro-ph.HE].

[38] G. 't Hooft, "Symmetry breaking through Bell-Jackiw anomalies", *Phys. Rev. Lett.* **37** (1976) 8, doi:10.1103/PhysRevLett.37.8.

[39] C.G. Callan, R.F. Dashen and D.J. Gross, "The structure of the gauge theory vacuum", *Phys. Lett.* **B63** (1976) 334, doi:10.1016/0370-2693(76)90277-X.

[40] R. Jackiw and C. Rebbi, "Vacuum periodicity in a Yang-Mills quantum theory", *Phys. Rev. Lett.* **37** (1976) 172, doi:10.1103/PhysRevLett.37.172.

[41] J.E. Kim and G. Carosi, "Axions and the strong CP problem", *Rev. Mod. Phys.* **82** (2010) 557, doi:10.1103/RevModPhys.82.557, arXiv:0807.3125 [hep-ph], Erratum: *Rev. Mod. Phys.* **91** (2019) 049902, doi:10.1103/RevModPhys.91.049902.

[42] R.D. Peccei and H.R. Quinn, "CP conservation in the presence of instantons", *Phys. Rev. Lett.* **38** (1977) 1440, doi:10.1103/PhysRevLett.38.1440.

[43] S. Weinberg, "A new light boson?", *Phys. Rev. Lett.* **40** (1978) 223, doi:10.1103/PhysRevLett.40.223.

[44] F. Wilczek, "Problem of strong P and T invariance in the presence of instantons", *Phys. Rev. Lett.* **40** (1978) 279, doi:10.1103/PhysRevLett.40.279.

[45] M. Dine, W. Fischler and M. Srednicki, "A simple solution to the strong CP problem with a harmless axion", *Phys. Lett.* **B104** (1981) 199, doi:10.1016/0370-2693(81)90590-6.

[46] A.R. Zhitnitsky, "On possible suppression of the axion hadron interactions", *Yad. Fiz.* **31** (1980) 497, English transl. publ. in *Sov. J. Nucl. Phys.* **31** (1980) 260.

[47] J.E. Kim, "Weak interaction singlet and strong CP invariance", *Phys. Rev. Lett.* **43** (1979) 103, doi:10.1103/PhysRevLett.43.103.

[48] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, "Can confinement ensure natural CP invariance of strong interactions?", *Nucl. Phys.* **B166** (1980) 493, doi:10.1016/0550-3213(80)90209-6.

[49] V.A. Rubakov, *JETP Lett.* **65** (1997) 621, doi:10.1134/1.567390, arXiv:hep-ph/9703409;
P. Agrawal and K. Howe, "Factoring the strong CP problem", *JHEP* **12** (2018) 029, doi:10.1007/JHEP12(2018)029, arXiv:1710.04213 [hep-ph];
M.K. Gaillard *et al.*, "Color unified dynamical axion", *Eur. Phys. J.* **C78** (2018) 972, doi:10.1140/epjc/s10052-018-6396-6, arXiv:1805.06465 [hep-ph].

[50] J. Preskill, M.B. Wise and F. Wilczek, "Cosmology of the invisible axion", *Phys. Lett.* **B120** (1983) 127, doi:10.1016/0370-2693(83)90637-8.

[51] L.F. Abbott and P. Sikivie, "A cosmological bound on the invisible axion", *Phys. Lett.* **B120** (1983) 133, doi:10.1016/0370-2693(83)90638-X.

[52] M. Dine and W. Fischler, "The not-so-harmless axion", *Phys. Lett.* **B120** (1983) 137, doi:10.1016/0370-2693(83)90639-1.

[53] L. Visinelli and P. Gondolo, "Dark matter axions revisited", *Phys. Rev.* **D80** (2009) 035024, doi:10.1103/PhysRevD.80.035024, arXiv:0903.4377 [astro-ph.CO].

[54] A. Vilenkin and A.E. Everett, "Cosmic strings and domain walls in models with goldstone and pseudogoldstone bosons", *Phys. Rev. Lett.* **48** (1982) 1867, doi:10.1103/PhysRevLett.48.1867.

[55] R.A. Battye and E.P.S. Shellard, "Axion string cosmology and its controversies", Proc. 2nd Int. Conf. Physics beyond the Standard Model, Ed. H.V. Klapdor-Kleingrothaus (Springer, Berlin, 1999), pp. 565–572, arXiv:astro-ph/9909231.

[56] V.B. Klaer and G.D. Moore, "The dark-matter axion mass", *JCAP* **11** (2017) 049, doi:10.1088/1475-7516/2017/11/049, arXiv:1708.07521 [hep-ph].

[57] M. Kawasaki, K. Saikawa and T. Sekiguchi, "Axion dark matter from topological defects", *Phys. Rev.* **D91** (2015) 065014, doi:10.1103/PhysRevD.91.065014, arXiv:1412.0789 [hep-ph];
M. Kawasaki *et al.*, "Long-term dynamics of cosmological axion strings", PTEP **2018** (2018) 091E01, doi:10.1093/ptep/pty098, arXiv:1806.05566 [hep-ph].

[58] E.W. Kolb and I.I. Tkachev, "Axion miniclusters and Bose stars", *Phys. Rev. Lett.* **71** (1993) 3051, doi:10.1103/PhysRevLett.71.3051 [hep-ph/9303313].

[59] P. Tinyakov, I. Tkachev and K. Zioutas, "Tidal streams from axion miniclusters and direct axion searches", *JCAP* **01** (2016) 035, doi:10.1088/1475-7516/2016/01/035, arXiv:1512.02884 [astro-ph.CO].

[60] D.G. Levkov, A.G. Panin and I.I. Tkachev, "Gravitational Bose-Einstein condensation in the kinetic regime", *Phys. Rev. Lett.* **121** (2018) 151301, doi:10.1103/PhysRevLett.121.151301, arXiv:1804.05857 [astro-ph.CO].

[61] S.J. Asztalos *et al.* [ADMX Collaboration], "A SQUID-based microwave cavity search for dark-matter axions", *Phys. Rev. Lett.* **104** (2010) 041301, doi:10.1103/PhysRevLett.104.041301, arXiv:0910.5914 [astro-ph.CO].

[62] N. Du *et al.* [ADMX Collaboration], "A search for invisible axion dark matter with the Axion Dark Matter Experiment", *Phys. Rev. Lett.* **120** (2018) 151301, doi:10.1103/PhysRevLett.120.151301, arXiv:1804.05750 [hep-ex].

[63] R. Battesti *et al.*, "High magnetic fields for fundamental physics", *Phys. Rept.* **765-766** (2018) 1, doi:10.1016/j.physrep.2018.07.005, arXiv:1803.07547 [physics.ins-det].

[64] B. Majorovits *et al.* [MADMAX interest Group], "MADMAX: A new road to axion dark matter detection", *J. Phys. Conf. Ser.* 1342 (2020) 012098, doi:10.1088/1742-6596/1342/1/012098, arXiv:1712.01062 [physics.ins-det].

[65] P.W. Graham *et al.*, "Experimental searches for the axion and axion-like particles", *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 485, doi:10.1146/annurev-nucl-102014-022120, arXiv:1602.00039 [hep-ex].

[66] D. Budker *et al.*, "Proposal for a Cosmic Axion Spin Precession Experiment (CASPEr)", *Phys. Rev.* **X4** (2014) 021030, doi:10.1103/PhysRevX.4.021030, arXiv:1306.6089 [hep-ph].

[67] S. Alekhin *et al.*, "A facility to search for hidden particles at the CERN SPS: the SHiP physics case", *Rept. Prog. Phys.* **79** (2016) 124201, doi:10.1088/0034-4885/79/12/124201, arXiv:1504.04855 [hep-ph].

[68] E. Bulbul *et al.*, "Detection of an unidentified emission line in the stacked X-ray spectrum of galaxy clusters", *Astrophys. J.* **789** (2014) 13, doi:10.1088/0004-637X/789/1/13, arXiv:1402.2301 [astro-ph.CO].

[69] A. Boyarsky *et al.*, "Unidentified line in X-Ray spectra of the Andromeda galaxy and Perseus galaxy cluster", *Phys. Rev. Lett.* **113** (2014) 251301, doi:10.1103/PhysRevLett.113.251301, arXiv:1402.4119 [astro-ph.CO].

[70] A. Boyarsky *et al.*, "Sterile neutrino dark matter", *Prog. Part. Nucl. Phys.* **104** (2019) 1, doi:10.1016/j.ppnp.2018.07.004, arXiv:1807.07938 [hep-ph].

[71] X.-D. Shi and G.M. Fuller, "A new dark matter candidate: Nonthermal sterile neutrinos", *Phys. Rev. Lett.* **82** (1999) 2832 , doi:10.1103/PhysRevLett.82.2832, arXiv:astro-ph/9810076.

[72] D.N. Abdurashitov *et al.*, "The current status of "Troitsk nu-mass" experiment in search for sterile neutrino", *JINST* **10** (2015) T10005, doi:10.1088/1748-0221/10/10/T10005, arXiv:1504.00544 [physics.ins-det].

[73] A.D. Sakharov, "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32, English transl. reprinted in *Sov.Phys.Usp.*, **34** (1991) 392, doi:10.1070/PU1991v034n05ABEH002497.

[74] V.A. Kuzmin, "CP-noninvairiance and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **12** (1970) 335, English transl. in *JETP Letters* **12** (1970) 228, http://jetpletters.ru/ps/1730/article_26297.shtml.

[75] I. Affleck and M. Dine, "A new mechanism for baryogenesis", *Nucl. Phys.* **B249** (1985) 361, doi:10.1016/0550-3213(85)90021-5.

[76] F.R. Klinkhamer and N.S. Manton, "A saddle point solution in the Weinberg-Salam theory", *Phys. Rev.* **D30** (1984) 2212, doi:10.1103/PhysRevD.30.2212.

[77] V.A. Kuzmin, V.A. Rubakov and M.E. Shaposhnikov, "On the anomalous electroweak baryon number nonconservation in the early universe", *Phys. Lett.* **B155** (1985) 36, doi:10.1016/0370-2693(85)91028-7.

[78] V.A. Rubakov, *Classical theory of gauge fields* (Princeton Univ. Press, Princeton, 2002).

[79] A.A. Belavin *et al.*, "Pseudoparticle solutions of the Yang-Mills equations", *Phys. Lett.* **B59** (1975) 85, doi:10.1016/0370-2693(75)90163-X.

[80] V.A. Rubakov and M.E. Shaposhnikov, "Electroweak baryon number nonconservation in the early universe and in high-energy collisions", *Usp. Fiz. Nauk* **166** (1996) 493, doi:10.3367/UFNr.0166.199605d.0493, English vers. publ. in *Phys. Usp.* **39** (1996) 461, doi:10.1070/PU1996v039n05ABEH000145, arXiv:hep-ph/9603208.

[81] K. Kajantie *et al.*, 'The electroweak phase transition: A nonperturbative analysis", *Nucl. Phys.* **B466** (1996) 189, doi:10.1016/0550-3213(96)00052-1, arXiv:hep-lat/9510020.

[82] S.V. Demidov, D.S. Gorbunov and D.V. Kirpichnikov, "Split NMSSM with electroweak baryogenesis", *JHEP* **11** (2016) 148, doi:10.1007/JHEP11(2016)148, arXiv:1608.01985 [hep-ph], Erratum: *JHEP* **08** (2017) 080, doi:10.1007/JHEP08(2017)080.

[83] V. Andreev *et al.* [ACME Collaboration], "Improved limit on the electric dipole moment of the electron", *Nature* **562** (2018) 355, doi:10.1038/s41586-018-0599-8.

[84] T. Konstandin, "Quantum transport and electroweak baryogenesis", *Usp. Fiz. Nauk* **183** (2013) 785, doi:10.3367/UFNr.0183.201308a.0785, English vers. publ. in *Phys. Usp.* **56** (2013) 747, doi:10.3367/UFNe.0183.201308a.0785, arXiv:1302.6713 [hep-ph];
M. Chala, G. Nardini and I. Sobolev, "Unified explanation for dark matter and electroweak baryogenesis with direct detection and gravitational wave signatures", *Phys. Rev.* **D94** (2016) 055006, doi:10.1103/PhysRevD.94.055006, arXiv:1605.08663 [hep-ph];
J.M. Cline, K. Kainulainen and D. Tucker-Smith, "Electroweak baryogenesis from a dark sector", *Phys. Rev.* **D95** (2017) 115006, doi:10.1103/PhysRevD.95.115006, arXiv:1702.08909 [hep-ph];
S. Bruggisser *et al.*, "Electroweak phase transition and baryogenesis in composite Higgs models", *JHEP* **12** (2018) 099, doi:10.1007/JHEP12(2018)099, arXiv:1804.07314 [hep-ph];
I. Baldes and G. Servant, "High scale electroweak phase transition: baryogenesis & symmetry non-restoration", *JHEP* **10** (2018) 053, doi:10.1007/JHEP10(2018)053, arXiv:1807.08770 [hep-ph];
M. Carena, M. Quirós and Y. Zhang, "Electroweak baryogenesis from dark-sector CP Violation", *Phys. Rev. Lett.* **122** (2019) 201802, doi:10.1103/PhysRevLett.122.201802, arXiv:1811.09719 [hep-ph];
A. Glioti, R. Rattazzi and L. Vecchi, "Electroweak baryogenesis above the electroweak scale", *JHEP* **04** (2019) 027, doi:10.1007/JHEP04(2019)027, arXiv:1811.11740 [hep-ph].

[85] M. Fukugita and T. Yanagida, "Baryogenesis without grand unification", *Phys. Lett.* **B174** (1986) 45, doi:10.1016/0370-2693(86)91126-3.

[86] E.K. Akhmedov, V.A. Rubakov and A.Y. Smirnov, "Baryogenesis via neutrino oscillations", *Phys. Rev. Lett.* **81** (1998) 1359, doi:10.1103/PhysRevLett.81.1359, arXiv:hep-ph/9803255.

[87] T. Asaka and M. Shaposhnikov, "The νMSM, dark matter and baryon asymmetry of the universe", *Phys. Lett.* **B620** (2005) 17, doi:10.1016/j.physletb.2005.06.020, arXiv:hep-ph/0505013.

[88] M. Drewes and B. Garbrecht, "Leptogenesis from a GeV seesaw without mass degeneracy", *JHEP* **03** (2013) 096, doi:10.1007/JHEP03(2013)096, arXiv:1206.5537 [hep-ph].

[89] W. Buchmuller, R.D. Peccei and T. Yanagida, "Leptogenesis as the origin of matter", *Ann. Rev. Nucl. Part. Sci.* **55** (2005) 311, 10.1146/annurev.nucl.55.090704.15558, arXiv:hep-ph/0502169;
S. Davidson, E. Nardi and Y. Nir, "Leptogenesis", *Phys. Rept.* **466** (2008) 105, doi:10.1016/j.physrep.2008.06.002, arXiv:0802.2962 [hep-ph];
C.S. Fong, E. Nardi and A. Riotto, "Leptogenesis in the universe", *Adv. High Energy Phys.* **2012** (2012) 158303, doi:10.1155/2012/158303, arXiv:1301.3062 [hep-ph].

[90] E.W. Kolb and M.S. Turner, "Grand Unified Theories and the origin of the baryon asymmetry", *Ann. Rev. Nucl. Part. Sci.* **33** (1983) 645, doi:10.1146/annurev.ns.33.120183.003241;
A. Riotto and M. Trodden, "Recent progress in baryogenesis", *Ann. Rev. Nucl. Part. Sci.* **49** (1999) 35, doi:10.1146/annurev.nucl.49.1.35, arXiv:hep-ph/9901362.

[91] E. Babichev, D. Gorbunov and S. Ramazanov, "Dark matter and baryon asymmetry from the very dawn of the Universe", *Phys. Rev.* **D97** (2018) 123543, doi:10.1103/PhysRevD.97.123543, arXiv:1805.05904 [astro-ph.CO].

[92] J. Urrestilla *et al.*, "Cosmic microwave anisotropies from BPS semilocal strings", *JCAP* **07** (2008) 010 , doi:10.1088/1475-7516/2008/07/010, arXiv:0711.1842 [astro-ph].

[93] A.A. Starobinsky, "Spectrum of relict gravitational radiation and the early state of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **30** (1979) 719, English transl. publ. in *JETP Lett.* **30** (1979) 682, http://jetpletters.ru/ps/1370/article_20738.shtml;
A.A. Starobinsky, "A new type of isotropic cosmological models without singularity", *Phys. Lett.* **B91** (1980) 99, doi:10.1016/0370-2693(80)90670-X;
A.H. Guth, "The inflationary universe: A possible solution to the horizon and flatness problems", *Phys. Rev.* **D23** (1981) 347, doi:10.1103/PhysRevD.23.347;
A.D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems", *Phys. Lett.* **B108** (1982) 389, doi:10.1016/0370-2693(82)91219-9;
A. Albrecht and P.J. Steinhardt, "Cosmology for Grand Unified Theories with radiatively induced symmetry breaking", *Phys. Rev. Lett.* **48** (1982) 1220, doi:10.1103/PhysRevLett.48.1220;
A.D. Linde, "Chaotic inflation", *Phys. Lett.* **B129** (1983) 177, doi:10.1016/0370-2693(83)90837-7.

[94] V.F. Mukhanov and G.V. Chibisov, "Quantum fluctuation and nonsingular universe", *Pisma Zh. Eksp. Teor. Fiz.* **33** (1981) 549, English vers. publ. in *JETP Lett.* **33** (1981) 532, http://jetpletters.ru/ps/1510/article_23079.shtml;
S.W. Hawking, "The development of irregularities in a single bubble inflationary universe", *Phys. Lett.* **B115** (1982) 295, doi:10.1016/0370-2693(82)90373-2;
A.A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations", *Phys. Lett.* **B117** (1982) 175, doi:10.1016/0370-2693(82)90541-X;
A.H. Guth and S.Y. Pi, "Fluctuations in the new inflationary universe", *Phys. Rev. Lett.* **49** (1982) 1110, doi:10.1103/PhysRevLett.49.1110;
J.M. Bardeen, P.J. Steinhardt and M.S. Turner, "Spontaneous creation of almost scale-free density perturbations in an inflationary universe", *Phys. Rev.* **D28** (1983) 679, doi:10.1103/PhysRevD.28.679.

[95] J.L. Lehners, "Ekpyrotic and cyclic cosmology", *Phys. Rept.* **465** (2008) 223, doi:10.1016/j.physrep.2008.06.001, arXiv:0806.1245 [astro-ph];
R.H. Brandenberger, "Unconventional cosmology", *Lect. Notes Phys.* **863** (2013) 333, doi:10.1007/978-3-642-33036-0_12, arXiv:1203.6698 [astro-ph.CO].

[96] P. Creminelli, A. Nicolis and E. Trincherini, "Galilean genesis: An alternative to inflation", *JCAP* **11** (2010) 021, doi:10.1088/1475-7516/2010/11/021, arXiv:1007.0027 [hep-th].

[97] J.M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models", *JHEP* **05** (2003) 013, doi:10.1088/1126-6708/2003/05/013, arXiv:astro-ph/0210603.

[98] E.R. Harrison, "Fluctuations at the threshold of classical cosmology", *Phys. Rev.* **D1** (1970) 2726, doi:10.1103/PhysRevD.1.2726.

[99] Y.B. Zeldovich, "A hypothesis, unifying the structure and the entropy of the universe", *Mon. Not. Roy. Astron. Soc.* **160** (1972) 1P, doi:10.1093/mnras/160.1.1P.

[100] P.J.E. Peebles and J.T. Yu, 'Primeval adiabatic perturbation in an expanding universe", *Astrophys. J.* **162** (1970) 815, doi:10.1086/150713.

[101] I. Antoniadis, P.O. Mazur and E. Mottola, "Conformal invariance and cosmic background radiation", *Phys. Rev. Lett.* **79** (1997) 14, doi:10.1103/PhysRevLett.79.14, arXiv:astro-ph/9611208.

[102] V.A. Rubakov, "Harrison–Zeldovich spectrum from conformal invariance", *JCAP* **09** (2009) 030, doi:10.1088/1475-7516/2009/09/030, arXiv:0906.3693 [hep-th];
K. Hinterbichler and J. Khoury, "The pseudo-conformal universe: Scale invariance from spontaneous breaking of conformal symmetry", *JCAP* **04** (2012) 023, doi:10.1088/1475-7516/2012/04/023, arXiv:1106.1428 [hep-th].

[103] M.A. Watanabe, S. Kanno and J. Soda, "Inflationary universe with anisotropic hair", *Phys. Rev. Lett.* **102** (2009) 191302, doi:10.1103/PhysRevLett.102.191302, arXiv:0902.2833 [hep-th];
T.R. Dulaney and M.I. Gresham, "Primordial power spectra from anisotropic inflation", *Phys. Rev.* **D81** (2010) 103532, doi:10.1103/PhysRevD.81.103532, arXiv:1001.2301 [astro-ph.CO];
A.E. Gumrukcuoglu, B. Himmetoglu and M. Peloso, "Scalar-scalar, scalar-tensor, and tensor-tensor correlators from anisotropic inflation", *Phys. Rev.* **D81** (2010) 063528, doi:10.1103/PhysRevD.81.063528, arXiv:1001.4088 [astro-ph.CO].

[104] M. Libanov and V. Rubakov, "Cosmological density perturbations from conformal scalar field: infrared properties and statistical anisotropy", *JCAP* **11** (2010) 045, doi:10.1088/1475-7516/2010/11/045, arXiv:1007.4949 [hep-th];
M. Libanov, S. Ramazanov and V. Rubakov, "Scalar perturbations in conformal rolling scenario with intermediate stage", *JCAP* **06** (2011) 010, doi:10.1088/1475-7516/2011/06/010, arXiv:1102.1390 [hep-th].

[105] L. Ackerman, S.M. Carroll and M.B. Wise, "Imprints of a primordial preferred direction on the microwave background", *Phys. Rev.* **D75** (2007) 083502, doi:10.1103/PhysRevD.75.083502, Erratum: *Phys. Rev.* **D80** (2009) 069901, doi:10.1103/PhysRevD.80.069901, arXiv:astro-ph/0701357;
A.R. Pullen and M. Kamionkowski, "Cosmic microwave background statistics for a direction-dependent primordial power spectrum", *Phys. Rev.* **D76** (2007) 103529, doi:10.1103/PhysRevD.76.103529, arXiv:0709.1144 [astro-ph].

[106] J. Kim and E. Komatsu, "Limits on anisotropic inflation from the Planck data", *Phys. Rev.* **D88** (2013) 101301, doi:https:10.1103/PhysRevD.88.101301, arXiv:1310.1605 [astro-ph.CO];
G.I. Rubtsov and S.R. Ramazanov, "Revisiting constraints on the (pseudo)conformal universe with Planck data", *Phys. Rev.* **D91** (2015) 043514, doi:10.1103/PhysRevD.91.043514, arXiv:1406.7722 [astro-ph.CO];
S. Ramazanov *et al.*, "General quadrupolar statistical anisotropy: Planck limits", *JCAP* **03** (2017) 039, doi:10.1088/1475-7516/2017/03/039, arXiv:1612.02347 [astro-ph.CO].

[107] V.A. Rubakov, M.V. Sazhin and A.V. Veryaskin, "Graviton creation in the inflationary universe and the grand unification scale", *Phys. Lett.* **B115** (1982) 189, doi:10.1016/0370-2693(82)90641-4;

R. Fabbri and M.D. Pollock, "The effect of primordially produced gravitons upon the anisotropy of the cosmological microwave background radiation", *Phys. Lett.* **B125** (1983) 445, doi:10.1016/0370-2693(83)91322-9;

L.F. Abbott and M.B. Wise, "Constraints on generalized inflationary cosmologies", *Nucl. Phys.* **B244** (1984) 541, doi:10.1016/0550-3213(84)90329-8;

A.A. Starobinsky, "Cosmic background anisotropy induced by isotropic flat-spectrum gravitational-wave perturbations", *Pis'ma Astron. Zh.* **11** (1985) 323, English transl. publ. in *Sov. Astron. Lett.* **11** (1985) 133, https://ui.adsabs.harvard.edu/abs/1985SvAL...11..133S/abstract.

[108] M. Kamionkowski, A. Kosowsky and A. Stebbins, "A probe of primordial gravity waves and vorticity", *Phys. Rev. Lett.* **78** (1997) 2058, doi:https://doi.org/10.1103/PhysRevLett.78.2058, arXiv:astro-ph/9609132;

U. Seljak and M. Zaldarriaga, "Signature of gravity waves in polarization of the microwave background", *Phys. Rev. Lett.* **78** (1997) 2054, doi:10.1103/PhysRevLett.78.2054, arXiv:astro-ph/9609169.