Physics – Uspekhi **62** (6) 529 – 567 (2019)

REVIEWS OF TOPICAL PROBLEMS

PACS numbers: **04.50.+h**, 12.10.Kt, **95.36.+x**, **98.80.–k**

Cosmological acceleration

S I Blinnikov, A D Dolgov

DOI: https://doi.org/10.3367/UFNe.2018.10.038469

Contents

Abstract. **An overview is given of the current status of the theory and observations of the acceleration of the expansion of the observable part of the Universe.**

Keywords: accelerating expansion of the Universe, dark energy

1. Introduction

One of the most impressive discoveries made in astronomy in the last two decades was that cosmological expansion is not slowing down with time as would be natural to expect for matter that moves in its own gravitational field. Quite the opposite, the expansion rate is growing, and this process has begun quite recently in cosmological history.

S I Blinnikov [1,2,3,4,*], **A D Dolgov** [1,2,†]
[1] National Research Center Kurchatov Institute, Alikhanov Institute of Theoretical and Experimental Physics, ul. B. Cheremushkinskaya 25, 117218 Moscow, Russian Federation
[2] Novosibirsk State University, ul. Pirogova 2, 630090 Akademgorodok, Novosibirsk, Russian Federation
[3] Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, Kashiwa, 277-8583, Japan
[4] Dukhov Research Institute of Automatics, ul. Sushchevskaya 22, 127055 Moscow, Russian Federation
E-mail: * sergei.blinnikov@itep.ru, † dolgov@fe.infn.it

Received 21 March 2018, revised 12 July 2018
Uspekhi Fizicheskikh Nauk **189** (6) 561 – 602 (2019)
DOI: https://doi.org/10.3367/UFNr.2018.10.038469
Translated by M Zh Shmatikov; edited by A M Semikhatov

The Nobel Prize in physics was awarded in 2011 to three astronomers (Saul Perlmutter, Brian P Schmidt, and Adam G Riess) "for the discovery of the accelerating expansion of the Universe through observation of distant supernovae." This discovery, though not unexpected, is of great importance. Even though it has not closed the chapter on the discussion about the nature of cosmological expansion in the modern epoch, it has provided a very convincing argument in favor of the expansion that occurs with acceleration. Distant supernovae have proved to be dimmer than expected. More precisely, the radiation flux detected from those stars proved to be lower than expected for the measured redshift $z \sim 1$ and under the assumption that the luminosity of those supernovae at $z \sim 1$ is the same as at $z = 0$, i.e., that they are standard candles.

This observation implies that the supernovae are located farther than was assumed and hence the Universe is expanding faster than predicted by the Standard Cosmological Model. Strictly speaking, neither acceleration nor expansion have been measured directly in the sense that the available accuracy is not sufficient to detect that distances between galaxies change with the observer's time t. Actually, the distances are measured as a function of the galactic' redshifts, and the dependences obtained turned out to differ from those predicted by the Standard Cosmological Model. Possible pitfalls in interpreting these results and observation errors are discussed below.

It is of great importance that the accelerated expansion of the Universe is favored not only by the measured data on radiation fluxes and distances to supernovae but also by a number of other entirely independent astronomical observations. As we discuss below, the set of these observations includes data on the age of the Universe, analyses of its large-scale structure, and measured fluctuations of the cosmic microwave background (CMB) and baryon acoustic oscillations (BAOs). Based on this, we can confidently conclude that the accelerated expansion of the Universe is a well-established phenomenon.

To fully assess to what extent this discovery is unusual, we adhere to a simple, even if not quite accurate similarity between the cosmological expansion and the motion of a stone thrown vertically upward in Earth's gravitational field. If the initial kinetic energy of the stone is smaller than its potential energy, i.e., $E_{\text{kin}} < U$, then, having attained a certain height, the stone halts for a moment and falls back. Otherwise, if $E_{\text{kin}} > U$, it does not halt and escapes to infinity. In the intermediate, very special case where $E_{\text{kin}} = U$, the stone also flies infinitely far, but its speed vanishes at infinite distance. The stone's velocity decreases as it moves upward in each of these cases.

It had been assumed until recently that the cosmological expansion occurs in a way that is fully similar to the examples described above. The Universe's expansion can be regarded as the inertial motion triggered by an initial kick caused by gravitational repulsion (antigravitation) in the inflation epoch, which we discuss below. If the initial kick was not very strong, the Universe will cease to expand at some moment and will collapse back to a hot and dense singularity. If the kick was strong enough, the expansion will continue eternally, and the hazard of a hot 'bath' does not occur. It was assumed, similarly to the example with the stone, that the cosmological expansion rate decreases with time.

Recent discoveries have shown that this is not fully true. The normal expansion of the Universe with a decreasing rate changed at some time, quite recently on the cosmological scale, to accelerated motion. If we recall the analogy with the stone, the picture is as follows: the stone was initially flying in a usual way, gradually losing velocity; however, later, it started accelerating as if a rocket engine was activated or, using terms closer to the cosmological situation, Earth's gravitational field at large distances became antigravitational, resulting in repulsion instead of attraction.

The discovered accelerated expansion of the Universe can be very briefly described as follows: normal gravitational attraction that slows down the expansion turns into gravitational repulsion at a relatively late stage of cosmological evolution, and the expansion rate starts increasing with time. We note from the very beginning that such antigravitation is only possible in a relativistic theory of gravity, for example, Einstein's general relativity (GR) theory, as discussed in Section 2. Newton's theory only allows gravitational attraction. It is worth noting that the initial kick that resulted in the creation of our astronomically large Universe from a microscopically small and confined state and in the currently observed expansion of the Universe 'by inertia' was also a consequence of cosmic antigravitation. This stage is referred to as cosmological inflation, and the expansion during this stage also occurred with acceleration. We can assert that the large Universe, suitable for life, turned to be possible only in the GR framework. Later, the initial 'inflationary' antigravity either vanished or became negligibly small during almost all of the cosmological history. The reason antigravity started playing a noticeable role in cosmology again at the current stage remains unclear: moreover, at first glance, it is not needed at all.

To avoid misunderstanding, we note that antigravity is only possible in the GR for boundless systems. Any finite object with a positive energy density (and we can only consider such objects so as not to encounter problems of an unstable world) always creates only normal gravitational attraction. However, two such objects placed inside an environment with negative pressure no longer experience 'normal' attraction at rather long distances between them but are instead accelerated in opposite directions. It can be shown that gravitational attraction in the case of two galaxies with masses of the order of that of the Milky Way is compensated by gravitational repulsion of vacuum energy at distances of about two megaparsecs (see below).

It was believed quite recently that the ultimate fate of the Universe and the geometry of its 3D space are linked in a unique way. A closed universe that has the geometry of a 3D sphere cannot expand eternally. The expansion will halt at some moment in the distant future and turn into collapse, as in the example with the stone thrown with a small initial velocity. Expansion of an open universe with the geometry of a 3D hyperboloid will never stop. The expansion in the intermediate case of a flat 3D space will also continue eternally. The halt will only occur asymptotically at $t \to \infty$.

The last case is of special interest because observations show that the geometry of the Universe is very close to flat.

The most probable outcome in the case of accelerated expansion is eternal expansion for any geometry of the world. We note that this result is absolutely opposite to the conclusions made in the inflationary theory, according to which the most probable ultimate fate of the Universe in the case of unaccelerated expansion (as was believed earlier) at the current and later stages is to collapse back to a singularity. Thus the accelerated expansion of the Universe may save the world from this grim fate. Isn't it here that the need for it lies?

The statement that our Universe will end its life as a result of sufficiently long-term inflation in a hot singularity requires some comments. Inflation causes density perturbations, including those on the cosmological horizon scale. These perturbations are of stochastic nature, and the sign of a density fluctuation $\delta\rho$ can be both positive and negative. It is therefore natural to expect that at some time on a horizon scale $\delta\rho$ will turn out to be positive, and this lump of the Universe will then detach from the general cosmological expansion and collapse. The external observer will see it as our collapse into a black hole. However, as was noted above, given the accelerated expansion, this would not be our fate, if the density of dark energy that possibly causes the expansion decreases more slowly than the scale factor squared [see the discussion after Eqn (2.16) below].

The source that causes accelerated expansion is not known. Two options are primarily discussed. The first is so-called dark energy, i.e., a substance that has a negative pressure whose absolute value is larger than one third of its energy density, $|P| > \rho/3$ (see Section 2). A form of that dark energy could be vacuum energy or, equivalently, the cosmological constant for which $P = -\rho$. The other option for dark energy is a quasi-constant scalar field ϕ similar to the one that probably generated inflation. In that case, the difference between exponential expansion at the dawn of the world and in modern days only amounts to the difference, albeit giant, between the energy and time scales.

The vacuum energy density does not change in the process of expansion [see Eqn (2.7) below]. If the dark energy is the vacuum energy, accelerated expansion will continue eternally for any 3D geometry of the world, as we have already noted . But if the dark energy is the energy of a very light scalar field or a field with an almost flat potential, then in the very distant future when the Hubble parameter becomes comparable to the mass or slope of the potential of that field, the expansion rate will again start slowing down, and the field ϕ itself will vanish due to redshift and/or the production of massless particles. Eventually, the Universe's fate will again depend on its geometry, as was the case in good old Friedmannian cosmology.

The accelerated expansion can also be a consequence of some modifications of gravity at small curvatures. Instead of the standard GR whose Lagrangian is proportional to the curvature scalar R, theories have been considered with an additional term nonlinear in curvature, $R + f(R)$. In principle, additional terms that depend on the Ricci or Riemann tensor squared or more sophisticated invariants can be considered; however, no detailed analysis of these options has been done. Due to the nonlinear dependence of the action on curvature, gravitational field equations, generally speaking, are of a higher order than the usual second order. As a result, problems can occur that are related to ghosts, tachyons, and the stability of the theory. Thus, some constraints on the form of the $f(R)$ function can be imposed.

The choice between these two options (dark energy vs. modified gravity) is one of the central problems in the phenomenological description of the accelerated expansion of the Universe. The inflation mechanism also involves a similar problem because inflation can be caused, in addition to a scalar field (inflaton), by R^2 corrections to the GR action. There is, however, a significant difference from the acceleration mechanism at the late cosmological state, i.e., at current times or somewhat earlier. The R^2 inflation needs gravity to be modified at large curvatures, an effect that occurs rather naturally due to radiative corrections, while the phenomenological description of today's accelerated expansion requires modification of gravity at very small curvatures, the modifications being introduced *ad hoc* without any theoretical justification whatsoever.

There is also a much more basic theoretical problem that is closely related to the accelerated expansion, and this is the problem of vacuum energy: theory, and in some sense also experiment, show that the vacuum is not empty but features a colossal energy density that exceeds the observed bound or possibly the measured value by 50 to 100 orders of magnitude. Some mechanism is therefore needed to compensate these giant contributions. No mechanism has been found as yet, despite numerous attempts. The problem of vacuum energy compensation seems to be one of the most challenging ones in modern fundamental physics.

This review is organized as follows. In Section 2, we present (and in Section 9.1, deduce both at a simple and naive level and rather accurately) the main cosmological equations and the Friedmann equations, introduce the concept of the cosmological constant (vacuum energy), and explain how cosmic antigravity may emerge. The main cosmological parameters are also defined in this section. Section 3 contains a description of vacuum energy and possible approaches to its solution. In Section 4, we discuss the astronomical data that are indicative of the accelerated expansion of the Universe. An analysis of data on supernovae and baryon acoustic oscillations is presented in special Section 5 due to the great importance of that analysis and interesting prospects for future research. A phenomenological description of the accelerated expansion of the Universe owing to dark energy or modified gravity is contained in Sections 6 and 7. In Section 9.1, we present two versions of a simplified derivation of the Friedmann equations (one using the Newtonian limit and the other based on the variational principle); Section 9.2 contains the main cosmological parameters and a discussion of the approaches to measuring them. Section 9.3 is devoted to the scalar field in cosmology.

2. Friedmann equations and cosmological acceleration

Underlying modern cosmology are the Einstein equations [1, 2], which relate the space–time curvature to the matter energy–momentum tensor $T_{\mu\nu}$:

$$R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu} R = \frac{8\pi}{m_{\mathrm{Pl}}^2}\, T_{\mu\nu}\,. \tag{2.1}$$

Here, $g_{\mu\nu}$ is the metric tensor that determines the 4D space-time interval as

$$\mathrm{d}s^2 = g_{\mu\nu}\, \mathrm{d}x^{\mu}\, \mathrm{d}x^{\nu}\,, \tag{2.2}$$

and $R_{\mu\nu}$ and $R = g^{\mu\nu} R_{\mu\nu}$ are the Ricci tensor and scalar curvature. They can be expressed in the known way in terms of the metric tensor and its first and second derivatives, as described in any textbook on GR or Riemannian geometry. The value $m_{\mathrm{Pl}} = 1.2 \times 10^{19}$ GeV is referred to as the Planck mass. Here and below, we use the natural system of units where $c = k = h/(2\pi) = 1$. The Newton gravitational constant is $G_{\mathrm{N}} = 1/m_{\mathrm{Pl}}^2$ in these units.

Friedmann was the first to apply the Einstein equations to the cosmology of the real Universe [3], although Einstein himself attempted to develop a static model of the Universe (and introduced the lambda term for this), and de Sitter was developing a model of a vacuum universe. Friedmann assumed that the Universe is uniform and isotropic (at least at large scales) and therefore space is supposed to have a constant 3D curvature with the interval

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2\left[\mathrm{d}\chi^2 + \sin^2\chi^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2)\right]. \tag{2.3}$$

A somewhat different form can be obtained by setting $\sin\chi = r$; then $\mathrm{d}\chi^2 = \mathrm{d}r^2/(1 - r^2)$, and the interval can be represented as (with $k = 1$)

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1 - kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2)\right]. \tag{2.4}$$

The quantity $a(t)$, which is referred to as the scale factor, determines the distance between two events in the 3D space. This form of the Friedmann metric is taken from [3] and another metric of this from [4] in the representation by Robertson and Walker [5, 6]. Consequently, the metric of the uniform and isotropic universe was named the Friedmann–Robertson–Walker (FRW) metric.

If $k = +1$, the 3D space can be considered a 3D spherical surface embedded into a flat 4D space. This case is referred to as the closed Universe. In all other cases, the Universe is open. If $k = -1$, the 3D space is a hyperboloid, while if $k = 0$, our space is flat in three dimensions and is described by standard Euclidian geometry (studied in high school).

Substituting the metric tensor that corresponds to interval (2.4) in Eqns (2.1), we obtain the equation that drives the evolution of the scale factor from the 'creation' of the world until today, as far as the Universe can be considered uniform and isotropic:

$$\left(\frac{\dot a}{a}\right)^2 = \frac{8\pi}{3}\,\frac{\rho}{m_{\rm Pl}^2} - \frac{k}{a^2}\,, \tag{2.5}$$

where ρ is the matter energy density, i.e., the T_{tt} (or T_{00}) component of the energy–momentum tensor. We note that in accordance with the assumption that the Universe is uniform and isotropic, it is assumed that the energy–momentum tensor has a diagonal form with spatial components proportional to the pressure density, $T_i^j = -\delta_i^j P$.

The second Friedmann equation expresses acceleration in the process of expansion in terms of energy and pressure density of matter:

$$\frac{\ddot a}{a} = -\frac{4\pi}{3m_{\rm Pl}^2}\,(\rho + 3P)\,. \tag{2.6}$$

We note that if $P < -\rho/3$, the acceleration is positive, i.e., gravity becomes repulsive (antigravity). As was noted in the Introduction, this was the reason for the initial kick that resulted in the expansion of the Universe, and, most probably, it is also the reason for the accelerated expansion of the Universe that is currently observed.

We quote another equation that describes evolution of the energy density in Friedmann cosmology:

$$\dot\rho = -3H(\rho + P)\,. \tag{2.7}$$

This equation is a covariant conservation law for the energy–momentum tensor,

$$D_\mu T_\mu^\nu = 0\,, \tag{2.8}$$

where D_μ is the covariant derivative in the gravitational field. Equation (2.7) follows from two Friedmann equations (2.5) and (2.6) but is presented separately due to its importance.

Section 9.1 contains an elementary (albeit somewhat cheating) derivation of these equations virtually without using GR, allowing one to better feel their physical content. It also contains a more accurate but nonstandard derivation based on the variational principle used in GR.

Two independent equations of the three presented above, (2.5), (2.6), and (2.7), contain three unknown functions: $a(t)$, $\rho(t)$, and $P(t)$. To close the set of equations, one more equation is needed. The equation of state is typically used for this; it describes pressure density as a function of the energy density, $P = P(\rho)$. A linear relation holds in many cases that are of practical interest:

$$P = w\rho\,, \tag{2.9}$$

where w is usually some constant parameter. This assumption is not mandatory however, and versions of the theory with $w = w(t)$ are discussed not infrequently. This assumption is natural for a phenomenological description and analysis of observational data at various redshift values (i.e., at different moments of cosmological evolution). The dependence $w(t)$ also emerges in a natural way, for example, when gravitating matter is a dynamic field. The function $w(t)$ is then determined by the equations of motion of that field (see Section 6 below). We note that the relation $P = P(\rho)$ is, strictly speaking, not always true. Pressure can depend on the energy density via its time derivative or the integral over time or can depend on other thermodynamic variables (temperature, specific entropy, etc.). However, relation (2.9) obviously always holds: $w(t) = P(t)/\rho(t)$.

In simpler cases that are nevertheless of practical interest, $w = \text{const}$. For example, for nonrelativistic matter, $P \ll \rho$, and it is therefore assumed that $w = 0$. It is known that for relativistic matter, $w = 1/3$. The form of the cosmological expansion law is especially simple if $k = 0$ [see Eqn (2.5)] when the 3D space is a flat, Euclidian one. Observations show that this is the case with good accuracy, and it is realized in our world. The scale factor increases in the nonrelativistic case in accordance with the law

$$a_{\rm NR}(t) \sim t^{2/3}\,. \tag{2.10}$$

The expansion law in the relativistic case has the form

$$a_{\rm rel}(t) \sim t^{1/2}\,. \tag{2.11}$$

Friedmann found solutions of Eqns (2.5)–(2.7) for various expansion laws and predicted the expansion of the universe, which was later discovered in astronomical observations. This discovery is usually credited to Hubble [7], although it was made earlier by Lemaître [8] (see also [9]). It is therefore relevant to refer to the expansion law of the universe as the Friedmann–Lemaître–Hubble law.

Einstein was for some reason against a nonstationary universe and did not agree for a long time (until Hubble's results) with Friedmann's solutions. In an attempt to apply GR to cosmology, Einstein found that there are no stationary solutions, and to 'save' the stationary universe, he proposed introducing an additional term into Eqns (2.1) (see below), the so-called lambda term or, equivalently, the cosmological constant [10]:

$$R_{\mu\nu} - \frac{1}{2}\,g_{\mu\nu}R - \Lambda g_{\mu\nu} = \frac{8\pi}{m_{\rm Pl}^2}\,T_{\mu\nu}\,. \tag{2.12}$$

The added term does not violate covariant conservation of the left-hand and right-hand sides of this equation. The Einstein tensor is automatically conserved in the metric theory:

$$D_\mu\left(R_\nu^\mu - \frac{1}{2}\,g_\nu^\mu R\right) \equiv 0\,. \tag{2.13}$$

This is the so-called contracted Bianchi identity, which automatically holds in Riemannian geometry. The vanishing of the Einstein tensor derivative resembles the automatic vanishing of the derivative of the left-hand side of Maxwell's equations, $\partial_\mu\partial_\nu F^{\mu\nu} \equiv 0$. The right-hand side of Eqn (2.12) is covariantly conserved [see (2.8)] due to general covariance, i.e., invariance of the theory with respect to the choice of the reference frame. The covariant derivative of the metric tensor is zero by construction of the theory. This can be easily verified by expressing the covariant derivative of the metric in terms of the Christoffel symbols Γ:

$$\begin{aligned} g_{ik;m} &= \partial_m g_{ik} - \Gamma^j_{im} g_{jk} - \Gamma^j_{km} g_{ij} = \partial_m g_{ik} - \Gamma_{kim} - \Gamma_{ikm} \\ &= \underline{\partial_m g_{ik}} - \frac{1}{2}\left(\underline{\partial_m g_{ik}} + \utilde{\partial_i g_{mk}} - \underline{\underline{\partial_k g_{im}}}\right) \\ &\quad - \frac{1}{2}\left(\underline{\partial_m g_{ki}} + \underline{\underline{\partial_k g_{mi}}} - \utilde{\partial_i g_{km}}\right) = 0\,. \end{aligned} \tag{2.14}$$

Here, the terms underlined in the same way cancel each other if the symmetry $g_{ik} = g_{ki}$ is taken into account. Therefore, the covariant conservation of all components in Eqn (2.12) is not violated if $\Lambda = \text{const}$. It is for this reason that this quantity is referred to as the cosmological constant.

Einstein's idea was that the antigravity created by the lambda term could compensate the gravitational attraction of ordinary matter and thus ensure the stationarity of the Universe. Such a stationary solution indeed exists, but it is unstable and, most importantly, it disagrees with the expansion of the Universe discovered later. It is for this reason that Einstein later rejected the hypothesis regarding the existence of the lambda term, considering it the biggest blunder of his life (as quoted by G Gamow in his book *My World Line*).

It later became clear that the cosmological constant is equivalent to the vacuum energy with the energy–momentum tensor

$$T^{\text{vac}}_{\mu\nu} = g_{\mu\nu}\rho^{\text{vac}} \tag{2.15}$$

and the equation of state $P = -\rho$ (see, e.g., the references in [11]). If the vacuum energy dominates, the Universe must expand in an accelerated way with an exponentially growing scale factor:

$$a^{\text{vac}}(t) \sim \exp\left(H_{\text{vac}} t\right), \tag{2.16}$$

where $H^2_{\text{vac}} = 8\pi\rho^{\text{vac}}/3m^2_{\text{Pl}} \approx \text{const}$.

We also note that according to Eqn (2.7) the energy density of relativistic matter decreases as $1/a^4$, and that of nonrelativistic matter, as $1/a^3$. The cosmological energy density in the epoch when either of them dominates decreases as $1/t^2$. On the other hand, the vacuum energy density does not change in the process of either expansion or contraction, i.e., $\rho^{\text{vac}} = \text{const}$.

This behavior of the cosmological energy density explains the link between the geometry of space and the ultimate fate of the Universe. This fate can be determined in the simplest way in the absence of vacuum or dark energy. As follows from Eqn (2.5), if $k > 0$, i.e., in a closed universe, the term proportional to curvature, k/a^2, eventually becomes larger than the term proportional to the matter energy density, which decreases not slower than $1/a^3$. Hence, H vanishes, and expansion is replaced by contraction.

Cosmological solutions in the theory with the lambda term have been explored by Friedmann [3, 4] and in studies [12–15]. The authors of the latter research believed, in contrast to Einstein, that adding the cosmological constant as in Eqns (2.12) was a very important generalization of GR. On the other hand, the attitude of many renowned physicists to the cosmological constant was very negative. In particular, Gamow wrote in the mid-1960s in his autobiography *My World Line* regarding the astronomical data in favor of a nonzero cosmological constant: "Lambda rears its ugly head again and again and again." However, afterwards, this evidence disappeared, only to re-emerge later (see Section 4.1).

However, as we see in Section 3, quantum field theory not only requires the cosmological constant (or, equivalently, vacuum energy) to be nonzero but also predicts that it has a gigantic value.

3. Vacuum energy problem

The vacuum energy problem seems to be the most challenging one in modern fundamental physics. Theoretical estimates of various contributions to the vacuum energy yield a fantastically huge value. It would not be an exaggeration to say that the theory predicts $\rho^{\text{vac}} \approx \infty$. The disagreement between the theoretical values and astronomical data is of the order of $10^{50} - 10^{100}$. Dispersion of those values corresponds to various physical sources of the vacuum energy. The huge value of the theoretical contribution and the vanishingly small total result lead one to recall Feynman's statement regarding radiative corrections in quantum electrodynamics: "The corrections are infinite but small!"

The vacuum in quantum field theory is the lowest-energy state, and, generally speaking, its energy is not necessarily zero. It is worth recalling the well-known example of the quantum oscillator, for which the ground-state energy is $\omega/2$, where ω is the oscillator frequency.

If the vacuum energy is not zero, the energy–momentum tensor that corresponds to that energy and momentum must be proportional to the metric tensor [see Eqn (2.15)], because this tensor is the only symmetric 'invariant' second-order tensor, i.e., a tensor that does not change in passing from one reference frame to another, as would be natural to expect for the one vacuum.

A quantum field is an infinite set of oscillators with all possible frequencies. The energy density of vacuum quantum fluctuations for any bosonic field consequently turns out to be infinitely large (see Section 9.3):

$$\begin{aligned}\rho^{\text{vac}}_{\text{b}} = \langle\mathcal{H}_{\text{b}}\rangle_{\text{vac}} &= \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\,\frac{\omega_k}{2}\,\langle a^\dagger_k a_k + b_k b^\dagger_k\rangle_{\text{vac}} \\ &= \int \frac{\mathrm{d}^3 k}{2\,(2\pi)^3}\,\omega_k = \infty^4 .\end{aligned} \tag{3.1}$$

Here, $\langle\mathcal{H}_{\text{b}}\rangle_{\text{vac}}$ is the vacuum expectation value of the Hamiltonian of the bosonic field, $\omega_k = (k^2 + m^2)^{1/2}$ is the energy of the quantum with momentum k and mass m, and a_k, $a^\dagger_k$, and b_k, $b^\dagger_k$ are the operators of annihilation and creation of particles and antiparticles.

The calculated pressure of vacuum quantum fluctuations also turns out to be infinite. If we use Eqns (9.28) from Section 9.3 and formally cut off the integrals for ρ and P at the same upper limit, the vacuum averages calculated in this way fail to satisfy the condition $\rho^{\text{vac}} = -P^{\text{vac}}$. This observation was used in [16] to reach the conclusion that vacuum fluctuations violate the Lorentz invariance of the vacuum, and because the vacuum must be Lorentz invariant, we have to require that the most diverging terms that can violate the Lorentz invariance vanish. However, diverging integrals should be handled with care, and, in particular, a comparison of formally equal expressions for different quantities does not actually imply that they are indeed equal; below, we give an example where bosonic and fermionic contributions mutually cancel in a supersymmetric world where the obtained finite result (up to logarithmic divergence) satisfies the required Lorentz-invariance condition $\rho^{\text{vac}} = -P^{\text{vac}}$.

Life in a world with an infinite energy density is apparently impossible. Such a world would either expand with an 'infinitely' large rate if the energy density is positive or instantaneously collapse into a singularity if it is negative. Fortunately, because commutators are replaced with anticommutators in the quantization procedure, the contribution of fermionic fields has the same value but the opposite sign:

$$\rho_{\rm vac}^{(\rm f)} \equiv \langle \mathcal{H}_{\rm f} \rangle_{\rm vac} = \int \frac{{\rm d}^3 k}{(2\pi)^3} \frac{\omega_k}{2} \langle a_k^\dagger a_k - b_k b_k^\dagger \rangle_{\rm vac}$$
$$= - \int \frac{{\rm d}^3 k}{(2\pi)^3} \omega_k = -\infty^4 . \qquad (3.2)$$

Hence, if each boson had a fermionic partner with exactly the same mass, the vacuum energy of quantum fluctuation would vanish. This was first observed by Pauli [17] and independently by Zel'dovich [18] (see also [19], where Pauli's work is discussed in depth). Interestingly, the idea of cancelation of the vacuum energies of bosons and fermions emerged prior to the appearance of work that introduced supersymmetry [20–22], where the cancelation occurs in a natural way. However, supersymmetry, even if it exists, is not an exact symmetry, and the masses of bosons and fermions are significantly different. Fourth-power divergences should nevertheless cancel irrespective of the masses. Quadratic divergences also cancel in the case of so-called soft supersymmetry violation, but the final result proportional to the differences between masses of ordinary particles and their superpartners turns out to be colossal at cosmological scales:

$$\rho_{\rm SUSY}^{\rm vac} \sim m_{\rm SUSY}^4 > 10^{55} \rho_{\rm c} , \qquad (3.3)$$

where $\rho_{\rm c} \approx 4 \times 10^{-47}$ GeV^4 is the cosmological energy density in the present-day Universe (see Section 9.2) and $m_{\rm SUSY}$ is the supersymmetry violation mass scale, which should be larger than 100 GeV as follows from available experimental constraints. The absence of signals from supersymmetric particles at the Large Hadron Collider (LHC) at CERN seems to indicate that either $m_{\rm SUSY}$ must be significantly larger or supersymmetry does not exist at all.

We note that the total contribution of bosons and fermions whose masses are tuned such that the contributions diverging quartically and quadratically cancel yields a logarithmically diverging result that satisfies the condition $\rho = -P$, and therefore the Lorentz invariance is not violated, as noted above. It is clear that if the mass difference tends to infinity, the result diverges as a second or even fourth power of mass.

If exact supersymmetry yields zero vacuum energy, this is not necessarily true for broken supersymmetry. Moreover, the so-called global supersymmetry in its softly violated version, which does not destroy the renormalizability of the theory, requires a nonzero vacuum energy close to the value $m_{\rm SUSY}^4$ presented above. However, if the theory is extended to include gravity, i.e., a supergravity (SUGRA) theory is considered, mandatory nonzero vacuum energy is no longer required, and very fine tuning of parameters can be used to obtain $\rho_{\rm SUGRA}^{\rm vac} = 0$. However, the natural value of the vacuum energy in that theory is $\rho^{\rm vac} \sim m_{\rm Pl}^4$, and hence the accuracy of the tuning required to cancel the vacuum energy must be of the order of 10^{-123}, a value that looks quite unnatural.

It is of interest to note that the vacuum energy experiences colossal jumps in phase transitions from a symmetric phase to a phase with broken symmetry: underlying the modern elementary-particle theory is the concept of spontaneously broken (gauge) symmetry. As the universe cools, the vacuum state changes in that theory. The energy density jump is $\delta\rho^{\rm vac} \sim 10^{60}$ GeV^4 for the phase transition in the grand unification theory (GUT), $\delta\rho^{\rm vac} \sim 10^8$ GeV^4 in the electroweak theory, and $\delta\rho^{\rm vac} \sim 10^{-2}$ GeV^4 in quantum chromodynamics (QCD) in going from the confinement to the deconfinement phase.

The vacuum energy problem becomes especially challenging if the structure of the QCD vacuum is considered. A proton is known to consist of three light quarks: p $\sim$ uud, the mass of each being about 5 MeV. One might expect that the proton mass is quite small, $m_{\rm p} \sim (15 \text{ MeV} - E_{\rm B}) < 15$ MeV, where $E_{\rm B}$ is the binding energy of the quarks in the proton. But the obtained result is at bast 60 times smaller than the proton mass. The missing contribution to the mass comes from the nontrivial properties of the QCD vacuum. Despite intuitive expectations, this vacuum is not empty but filled with a condensate of quark [23] and gluon [24] fields:

$$\langle \bar{q} q \rangle \neq 0 , \quad \langle G_{\mu\nu} G^{\mu\nu} \rangle \neq 0 . \qquad (3.4)$$

The energy density of these condensates is negative and is about 1 GeV^4. This value, which is reasonable at elementary-particle physics scales, is huge for the cosmological standards of the present-day Universe:

$$\rho_{\rm QCD}^{\rm vac} \approx -10^{45} \rho_{\rm c} . \qquad (3.5)$$

Quarks destroy the gluon condensate inside the proton, thus increasing the proton mass to the required value:

$$m_{\rm p} = 2m_{\rm u} + m_{\rm d} - \rho^{\rm vac} l_{\rm p}^3 \sim 1 \text{ GeV} , \qquad (3.6)$$

where $l_{\rm p}$ is the proton size.

An almost mystical situation occurs. According to a theory that is perfectly established and agrees with experiment, the vacuum is not empty. It contains quark and gluon fields, whose energy density is 45 orders of magnitude larger than the cosmological value. Nevertheless, the total vacuum energy density is equal to the cosmological value. Something else 'resides' in the vacuum, and this 'something', possibly a new field, compensates the 45 orders of magnitude of $\rho_{\rm QCD}^{\rm vac}$. We note that the new field must be light; otherwise, it would not be able to be effective at cosmological scales. Moreover, it 'does not talk to' quarks or gluons. Otherwise, this field would interact with them, and this interaction would be experimentally detectable. The vacuum energy is apparently described by a single parameter, and all the described contributions are to be compensated using a single subtraction constant, but such a solution seems very unnatural.

The problem is aggravated by the fact that the vacuum (or vacuum-like) energy is not exactly zero, and for an unknown reason its value is close to the density of ordinary matter (including dark matter), although the cosmological evolution laws are quite different for $\rho^{\rm vac}$ and $\rho_{\rm m}$. As was noted above, $\rho^{\rm vac}$ remains unchanged as the universe expands, while $\rho_{\rm m}$ decreases as the scale factor a cubed in the case of nonrelativistic matter, and as $1/a^4$ in the case of relativistic matter. A number of attempts have been made to solve this problem or rather these problems, but none was particularly successful. We provide a brief description of those attempts below. More detailed reviews of these problems can be found in [25–38].

The first and simplest, but terribly inelegant, solution that was mentioned several lines above is that all those contributions to vacuum energy are compensated by some subtraction constant, whose value (selected with an extreme accuracy) equals the total of physical contributions to the vacuum energy with an accuracy of 10^{-45} – 10^{-127}. This is essentially the choice of zero on the energy scale. Because the value of only one parameter must be explained, the feasibility of that specific inelegant solution cannot be ruled out, although it seems quite unnatural. It is also desirable that all contributions to the vacuum energy originate from physical fields rather than from an arbitrarily chosen number.

Somewhat better seems to be the proposal to solve the vacuum energy problem based on the anthropic principle. Underlying this principle is the hypothesis that values of certain fundamental constants are set by the observer located in the Universe where conditions suitable for that observer's existence are available. A list of earlier studies of the anthropic principle, apparently incomplete, is contained in [39–42]; a more detailed discussion can be found, e.g., in books [43–46].

For the anthropic principle to be naturally realized, it is necessary that there be a nearly infinite (or infinite) number of universes with various values of fundamental constants and, in particular, with different bare values of the vacuum energy (i.e., with various subtraction constants). Cosmologically large universes with various physical properties emerge quite naturally in the inflationary models [47]. Chaotic inflation seems to be the best candidate in this sense [48]. The idea that the large number of universes solves the vacuum energy problem seems to have been first proposed in [49].

Strong support for the idea of a multiplicity of universes was lent in the context of superstring theory. These theories are formulated in a 10-dimensional space; it is assumed, however, that the six extra dimensions are compact and small. A realistic mechanism for compactifying the extra dimensions that is compatible with inflation has been proposed in [50]. It was quickly understood that there are a huge number of ways to compactify that 6-dimensional space. Estimates made in [51] show that the number of possible compactification types is about 10^{500} or, maybe, significantly more. Similar ideas regarding a large number of various universes have been developed in [52].

Relatively few (if any) universes among those multiple universes can be suitable for life. The absolute value of vacuum energy, whether positive or negative, in a universes suitable for life cannot be too large. If ρ^{vac} is large and negative, the universe would collapse prior to the formation of stars and planets, while if it is large and positive, expansion would be so fast that no celestial bodies would have enough time to form, because matter density would very rapidly become negligibly small. These arguments were presented in a quantitative form in [53–55]. It was shown that if $\rho^{\text{vac}} > 0$, the vacuum energy density cannot differ from the cosmological density of ordinary matter by more than 2 to 3 orders of magnitude. The anthropic principle was used in [56] to derive constraints on various cosmological parameters, including vacuum energy.

The problem of the anthropic constraint on vacuum energy was explored in detail in [57]. The author of [57] makes a significantly stronger statement that the anthropic principle predicts a nonzero vacuum energy, and its value is close to the observed one. The anthropic constraint on the (positive) vacuum energy can be derived from the requirement that the vacuum energy could only start dominating after the galaxy formation epoch. The energy density of ordinary matter decreases in the process of cosmological expansion as the scale factor cubed, while ρ^{vac} remains constant; therefore, the following condition must be satisfied:

$$\rho^{\text{vac}} < (1 + z_{\max})^3 \rho_{\text{m}}^0 , \tag{3.7}$$

where ρ_{m}^0 is the matter density in the present-day Universe and $z_{\max}$ is the maximum redshift at which the formation of galaxies commences.

According to [57], if $z_{\max} \sim 10$, the anthropic constraint on the vacuum energy density is $\rho^{\text{vac}} < 4000\rho_{\text{m}}^0$. This constraint was improved in [54] (see also [58]): $\rho^{\text{vac}} < 100\,\rho_{\text{m}}^0$, where it was noted for the first time that not all values of Λ are compatible with the presence of sentient life in the Universe.

The anthropic approach was applied in [59] to three time scales (galaxy formation time, moment when the cosmological constant dominates in the energy density of the Universe, and the age of the Universe) to reach the conclusion that the vacuum energy density is close to the observed one.

A similar conclusion regarding the smallness of vacuum energy was made in [60]; however, according to the authors, the zero vacuum energy is much more probable than the observed dark energy value.

A detailed analysis of the anthropic approach to the probability of a specific vacuum energy value was performed in [61, 62]. Although the authors agree that the anthropic considerations favor the largest probability of a small vacuum energy, their results slightly differ from each other. Given the uncertainty of estimates, we can conclude that they agree well with each other.

A negative vacuum energy density results in gravitational attraction, and to prevent early contraction of the universe prior to its reaching its current age, the condition $|\rho^{\text{vac}}| < \rho_{\text{m}}^0$ should be satisfied.

Despite obvious successes in understanding options for a natural realization of the anthropic principle, the theory that does enable calculating main parameters in a dynamical way looks unsatisfactory. In this relation, we recall the Friedmann cosmology problems whose solution would have required the anthropic principle if a nice and economic inflationary solution had not been found, which made a clear-cut prediction regarding the density perturbation spectrum. Criticism of the anthropic principle can be found in [63]. However, any approach can exist in physics as far as and to the extent that it does not disagree with experiment and well-established theories in their applicability areas.

It was hoped some time ago that there is a symmetry that would require a zero lambda term, as occurs in unbroken supersymmetry. These hopes seem to be futile. The required symmetry must be realized at rather low energies, below 100 MeV, otherwise it would be impossible to compensate the vacuum energy in QCD, Eqn (3.5), but physics at those energies has been well explored and any appropriate symmetry is definitely nonexistent there. Moreover, to obtain the observed vacuum (or vacuum-like) energy at a level of $(10^{-12}\ \text{GeV})^4$, we have to pass to an energy range that is significantly lower than 100 MeV.

In the opinion of at least one of the authors of this review, of most interest is the mechanism of dynamic compensation of vacuum energy proposed in 1982 in [64]. It is assumed that there is a new light or massless field ϕ coupled to gravity in such a way that if there is a vacuum energy, the condensate of that field emerges that compensates the initial vacuum energy. This mechanism resembles the well-known solution of the *CP*-violation problem by means of introducing an axion field. The very idea of the system responding to an external effect so as to reduce that effect is quite general. This mechanism, known in chemistry since the 19th century, is referred to as Le Chatelier's principle.

The original idea in [64] and a variety of subsequent proposals (see the discussion in reviews [25–38]) were based on introducing a scalar field; however, higher-spin fields, for example, vector [65] or tensor [66], cannot be ruled out either. Regardless of their specific realization, these predictions feature some general and rather attractive properties. First, owing to the contribution of the energy–momentum of ϕ to cosmological dynamics, the initial exponential expansion is replaced by a power-law expansion. Second, the field ϕ compensates the vacuum energy incompletely and only with an accuracy of the order of $\rho_c(t)$. Third, the 'uncompensated' vacuum energy can have an unusual equation of state that results, in particular, in accelerated expansion. Thus, such a mechanism not only provides a solution of the problem of vacuum energy compensation from a large value typical of elementary-particle physics to cosmologically small values but also enables resolving the so-called coincidence problem, i.e., explaining why the uncompensated dark energy ρ_{DE} is close to the total time-dependent energy density in the Universe $\rho_c(t)$, in full correspondence with what is observed in the skies. The existence of cosmological dark energy was predicted in this sense in 1982 [64], long before it was observed astronomically. Unfortunately, numerous attempts to find a realistic cosmological model that would include that mechanism have thus far failed.

Bronstein seems to have been the first to hypothesize that vacuum energy can be time dependent, and its value can be close to the cosmological energy density [67]. However, the models where $\Lambda = \Lambda(t)$, if taken literally, are far from being innocuous because covariant conservation of the left-hand side of the Einstein equation is known to imply the condition $\Lambda = \mathrm{const}$ [see a discussion of the problem immediately after Eqn (2.12)]. To compensate vacuum energy nonconservation, nonconservation of the matter energy–momentum tensor $T_{\mu\nu}$ is added, which has the same value but opposite sign. However, the theory then loses its predictive power because such models only determine the covariant divergence

$$D_\mu T^\mu_\nu = -\frac{m_{\mathrm{Pl}}^2}{8\pi}\,\partial_\nu \Lambda\,, \tag{3.8}$$

but are not able to say anything about the value of $T_{\mu\nu}$. It is for this reason that Landau strongly criticized Bronstein's approach.

It should be kept in mind that the energy–momentum tensor is calculated in GR as a functional derivative of the matter action with respect to the metric, and is therefore automatically conserved due to the principle of general covariance. Rejecting that principle would lead, in general, to a nonzero graviton mass, contradicting the observations. (See study [68], where modified theories of gravity are analyzed on the scale of stars and galaxies, including possible violation of the Vainshtein mechanism [69], which could be helpful in avoiding problems with a nonzero graviton mass.) An alternative option is to drastically modify the theory of gravity, for example by introducing nonmetric theories.

All of these problems can be avoided if a new light or massless field is introduced in the theory whose approximate equation of state is $P \approx -\rho$, which corresponds to a quasi-constant vacuum-like energy density.

The first model of the dynamic cancelation of vacuum energy proposed in [64] was based on a massless scalar field that is nonminimally coupled to gravity and satisfies the equation of motion

$$\ddot\phi + 3H\dot\phi + U'(\phi, R) = 0\,. \tag{3.9}$$

The simplest form of the potential $U(\phi, R)$ was chosen, $U = \xi R\phi^2/2$. It readily follows that if $\xi R < 0$, this equation in the de Sitter space has unstable solutions that exponentially grow with time because, for a curvature R, the effective mass squared of the ϕ field is negative. A similar situation occurs in the Higgs model, where long-wavelength states with $\phi = 0$ are unstable.

As ϕ grows, its effect on cosmological evolution cannot be disregarded, and it is easy to verify that the expansion that was initially exponential, $a(t) \sim \exp(H_v t)$, is asymptotically replaced with a power-law dependence:

$$\phi \sim t\,, \quad a(t) \sim t^\beta\,, \tag{3.10}$$

where β is some constant expressed in terms of the model parameters. Thus, due to the inverse effect of ϕ on the cosmological solution, the exponential expansion law transforms into Friedmann behavior, despite the presence of the initially nonzero vacuum energy.

A disadvantage of that simple model based on a scalar field is that the energy–momentum tensor of that field is not proportional to the metric tensor,

$$T_{\mu\nu}(\phi) \neq \tilde\Lambda g_{\mu\nu}\,, \tag{3.11}$$

and hence the vacuum energy does not vanish even asymptotically. The expansion regime changes due to the weakening of gravitational interaction, whose coupling constant decreases with time, first exponentially and then as time squared:

$$G_{\mathrm{N}} \sim \frac{1}{t^2}\,. \tag{3.12}$$

If this behavior of G_{N} were realized in the early universe and were somehow stabilized later, this mechanism could explain the hierarchy of gravitational and electroweak scales [70].

For successful predictions of primordial nucleosynthesis, the constant G_{N} must change by no more than 10% from the moment when the age of the Universe was about one second to modern times [71–74]. Even stronger constraints on the variability of G_{N} follow from the analysis of the arrival of signals from pulsar J1713 + 0747 (the so-called pulsar timing method): $\dot G_{\mathrm{N}}/G_{\mathrm{N}} = (-0.6 \pm 1.1) \times 10^{-12}$ year^{-1} (99.7% CL), a rate that is at least 30 times slower than the expansion rate of the Universe [75]. Stabilization of the G_{N} variation with time is therefore apparently needed. There are, however, some indications of some mystic variability of $G_{\mathrm{N}}(t)$ [76, 77].

Several dozen studies have been published by now where various mechanisms of the dynamic compensation of vacuum energy by a scalar field are discussed. Earlier studies presented in [78–90] and in [91–94] primarily explored a phenomenological description of dark energy in terms of a scalar field. References to later studies can be found in [25–38]. Unfortunately, none of the proposed mechanisms can be considered to be fully satisfactory.

Weinberg formulated a no-go theorem regarding natural compensation of the vacuum energy by a scalar field (see [25]). The basis of the theorem is that compensation requires two conditions: the total $\rho_{\mathrm{tot}} = \rho^{\mathrm{vac}} + \rho_\phi$ must be zero, and the derivative of the potential $U'(\phi, R)$ at that point must also vanish. The latter condition implies that the value of ϕ at which the total vacuum-like energy vanishes, $\rho_{\rm tot} = 0$, is at the same time a solution of equation of motion (3.9) with $\phi = \text{const}$. However, as occurs not infrequently in physics, no-go theorems can be circumvented by changing the conditions of the problem. For example, nonzero spin fields can be used (see below) or the form of the coupling of ϕ to gravity can be modified, as was done in [95] and analyzed in more detail in [96, 97]. The main idea of these studies is to modify the kinetic term of the scalar field by introducing a coefficient that is inversely proportional to the curvature squared:

$$A = \int \mathrm{d}^4 x \sqrt{-g} \times \left[-\frac{1}{2}(R + 2\Lambda) + F_1(R)\,\frac{D_\mu \phi D^\mu \phi}{2R^2} - U(\phi, R) \right]. \quad (3.13)$$

Here, a system of units is used in which $m_{\rm Pl}^2/8\pi = 1$. The corresponding equation of motion for ϕ has the form

$$D_\mu \left[D^\mu \phi \left(\frac{1}{R} \right)^2 \right] + U'(\phi) = 0\,, \quad (3.14)$$

which can be reduced in the FRW metric for a spatially uniform field $\phi(t)$ to the equation

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + 3H \right) \left(\frac{\dot\phi}{R^2} \right) + U'(\phi) = 0\,. \quad (3.15)$$

An additional term appears, in particular, in the Einstein equations that is proportional to higher derivatives of ϕ:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - 4C_1 \left(R_{\mu\nu} - \frac{1}{4} g_{\mu\nu} R + g_{\mu\nu} D^2 - D_\mu D_\nu \right) R - \frac{D_\mu \phi D_\nu \phi}{R^2} + \frac{(D_\alpha \phi)^2}{2R^2} \left(g_{\mu\nu} + \frac{4R_{\mu\nu}}{R} \right) - g_{\mu\nu} \left[U(\phi) + \rho_{\rm vac} \right] + 2(g_{\mu\nu} D^2 - D_\mu D_\nu)\,\frac{(D_\alpha \phi)^2}{R^3} = T_{\mu\nu}\,, \quad (3.16)$$

where $T_{\mu\nu}$ is the matter energy–momentum tensor, $(D\phi)^2 \equiv D_\alpha \phi D^\alpha \phi$, and we set $F_1(R) = C_1 R^2$ for simplicity.

Taking the trace over μ and ν yields

$$-R + 3\left(\frac{1}{R} \right)^2 (D_\alpha \phi)^2 - 4\left[U(\phi) + \rho^{\rm vac} \right] - 6D^2 \left[2C_1 R - \left(\frac{1}{R} \right)^2 \frac{(D_\alpha \phi)^2}{R} \right] = T^\mu_\mu\,. \quad (3.17)$$

The covariant d'Alembert operator that acts on the scalar has the form $D^2 = \mathrm{d}^2/\mathrm{d}t^2 + 3H\,\mathrm{d}/\mathrm{d}t$ in the spatially uniform case, and the Hubble parameter is related to the scalar curvature by the formula

$$R = -6(2H^2 + \dot H)\,. \quad (3.18)$$

Equations (3.15), (3.17), and (3.18) fully describe our system in the absence of ordinary matter. Studies [96, 97] showed that even if the vacuum energy is initially nonzero, these equations have the asymptotic solution

$$H = \frac{h}{t}\,, \qquad R = \frac{r}{t^2}\,, \qquad U(\phi) - \rho_{\rm vac} \sim \frac{1}{t^2}\,. \quad (3.19)$$

The amplitude of the field ϕ tends to a value at which the potential $U(\phi)$ compensates the initially nonzero vacuum energy, and the exponential expansion law is replaced with Friedmann behavior. This implies that this solution compensates the vacuum energy, thus enabling Weinberg's ban to be circumvented. This solution has other attractive properties; it can be shown, notably, that the Hubble parameter corresponding to it is $H = 1/(2t)$, as is the case in actual cosmology at the stage where relativistic matter dominates. However, this value of H is in no way related to the type of matter the universe contains, whereas in ordinary cosmology $H \sim \sqrt{\rho}$.

Moreover, such 'nice' solutions turn out to be unstable [97]. The last term in the left-hand side of Eqn (3.17), which contains D^2, turns out to be small and inessential. However, in studying the stability of a solution of the equation, terms with higher derivatives must be retained because they drastically affect stability. It was shown in [92] that the terms with higher derivatives result in a singular solution with R and H becoming infinitely large during a finite and short time. The curvature sign changes in the general solution from initially negative to positive, and the Hubble parameter tends to minus infinity, resulting in the collapse of the Universe, regardless of the presence of ordinary matter.

A more general action of the form

$$A = \int \mathrm{d}^4 x \sqrt{-g} \left[-\frac{m_{\rm Pl}^2}{16\pi}(R + 2\Lambda) + F_2(\phi, R)\, D_\mu \phi D^\mu \phi + F_3(\phi, R)\, D_\mu \phi D^\mu R - U(\phi, R) \right] \quad (3.20)$$

has not been explored yet. Terms depending on $R_{\mu\nu}$ or $R_{\mu\nu\alpha\beta}$ could in principle be added in an attempt to obtain a realistic model. However, the presence of higher curvature tensors, in addition to scalar ones, should generally result in the emergence of tachyons or ghosts in the theory of gravity, which are luckily avoided in GR or in theories whose action depends on R alone, despite the presence of second derivatives.

The Weinberg theorem can also be circumvented in a model of dynamic compensation involving two scalar fields [98–102]. One of the two scalar fields ϕ considered in the model in [101] has the usual kinetic term in the Lagrangian, while the second field χ has a kinetic term that is singular as $\phi \to 0$ (p is an integer),

$$L_{\rm kin} = \frac{1}{2}\,\partial_\mu \phi \partial^\mu \phi + \frac{\mu^{2p}}{2\phi^{2p}}\,\partial_\mu \chi \partial^\mu \chi\,, \quad (3.21)$$

which resembles the Lagrangian in the later study [95] in that it makes the ϕ field relax in the vicinity of zero. More accurately, the converse statement is valid: study [95] uses this idea from [101]. Consequently, these studies are very similar in regard to both their advantages and disadvantages.

Compensating fields with a nonzero spin can presumably solve the problem of vacuum energy more successfully than the scalar field does. A model of the compensating vector field V_μ was proposed in [65], with the Lagrangian

$$L_1 = \eta \left[\frac{F^{\mu\nu} F_{\mu\nu}}{4} + (V^\mu_{;\mu})^2 \right] + \xi R m^2 \ln \left(1 + \frac{V^2}{m^2} \right). \quad (3.22)$$

The time component of this field, V_t, turns out to be unstable in the de Sitter space and grows with time:

$$V_t \sim t + \frac{c}{t} \, . \tag{3.23}$$

The dominating term in the energy–momentum tensor that is responsible for this solution is proportional to the metric tensor and has the form

$$T_{\mu\nu}(V_t) = -\rho^{\rm vac} g_{\mu\nu} + \delta T_{\mu\nu} \, , \tag{3.24}$$

where $\delta T_{\mu\nu}$ asymptotically tends to zero.

The gravitational constant in this model depends on time but only logarithmically, a behavior that can be consistent with the constraints on variations of $G_{\rm N}$. However, this scheme is plagued with a standard disadvantage of known compensation mechanisms: there is no relation between the rate of cosmological expansion H and matter energy density in the Universe, as is the case in standard cosmology [see (2.5)].

Another interesting option is provided by a massless second-rank tensor field $S_{\mu\nu}$ [66], with the Lagrangian that contains only kinetic terms:

$$L_2 = \eta_1 S_{\alpha\beta;\gamma} S^{\alpha\gamma;\beta} + \eta_2 S^{\alpha}_{\beta;\alpha} S^{\gamma\beta}_{\;\;;\gamma} + \eta_3 S^{\alpha}_{\alpha;\beta} S^{\gamma;\beta}_{\gamma} \, . \tag{3.25}$$

In other words, this is a free field minimally coupled to gravity. It is known that to avoid the infrared catastrophe, the massless field must be coupled to a conserved source. On the other hand, the only conserved source of a second-rank tensor field is the energy–momentum tensor (see, e.g., [103]). A conclusion can therefore be made that the graviton can be the only spin-two massless field. However, this constraint is not applicable to free fields, and the massless rank-two tensor field is not ruled out. We also note that Lagrangian (3.25) depends only on field derivatives and is therefore invariant under a shift by a covariantly constant second-rank tensor. Owing to this, quantum radiative correction must not generate mass for $S_{\alpha\beta}$. This is an important difference from the scalar field, whose mass can emerge as a result of radiative corrections.

The components of the $S_{\alpha\beta}$ field in a flat FRW metric satisfy the equations

$$(\partial_t^2 + 3H\partial_t - 6H^2)\, S_{tt} - 2H^2 s_{jj} = 0 \, , \tag{3.26}$$

$$(\partial_t^2 + 3H\partial_t - 6H^2)\, s_{tj} = 0 \, , \tag{3.27}$$

$$(\partial_t^2 + 3H\partial_t - 2H^2)\, s_{ij} - 2H^2 \delta_{ij} S_{tt} = 0 \, , \tag{3.28}$$

where $s_{tj} = S_{tj}/a(t)$ and $s_{ij} = S_{ij}/a^2(t)$.

It is easy to verify that the time–time component S_{tt} and the isotropic space–space component $S_{ij} \sim \delta_{ij}$ are unstable and grow with time in the de Sitter metric. The energy–momentum tensor of those components tends to a constant proportional to the metric tensor and compensates the initial vacuum energy up to terms of the order of $m_{\rm Pl}^2/t^2$. However, although this is not obvious, the constant of gravitational coupling with matter varies with time, as follows from the arguments presented in [104].

A general comment is relevant here. In developing a theory of fields with spin, additional requirements are imposed that the components with lower spins vanish. For example, the time component of a vector, which is a 3D scalar, is eliminated from the vector-field theory. The 3D vector is eliminated in addition to the 3D scalar from the rank-two tensor field theory. However, the situation is quite the opposite in the example considered above: 3D scalar components of the 4D tensor remain physical variables, while the elimination of higher spins is required. It is not clear whether a theory of that kind can be developed in a self-consistent way; however, an example is published of a scalar gauge field theory that is described by the time component of a vector V_t, while spin-one components are eliminated [105].

Promising results have been obtained recently in a series of studies [106–109] for the mechanism of dynamic compensation of vacuum energy in a theory of several coupled vector fields whose time components increase proportionally to time, thus compensating the vacuum energy. This model seems to enable avoiding the problem of a growing gravitational constant described above, although the 'eternal' problem of the absence of a canonical dependence of the Hubble parameter on the energy density of ordinary matter in early times, for example, in the period of primordial nucleosynthesis, persists.

The mechanisms of dynamic compensation of vacuum energy described above have mostly been analyzed in the inflationary framework, which, as we have stressed, provided a successful prediction of the density perturbation spectrum. Nevertheless, inflation cannot be considered an established fact; it remains a hypothesis. Moreover, inflation itself can be generated in various ways, for example, with the mechanism described above [101]. Other scenarios in addition to inflation have been developed; they are not as comprehensive but claim to explain the small value of the present-day cosmological constant (see, e.g., [101], where the dynamic mechanism is included in the 'cyclic universe' scenario).

In conclusion, we note another cosmological problem related to vacuum energy. The current value of $\rho^{\rm vac}$ is very close to the matter energy density and differs from it by only a factor of two, in spite of significant differences between their evolutions in the process of expansion of the Universe. A natural explanation of why these values are close to each other is missing, as is the understanding of why the energy densities of baryons and dark matter are also close to each other.

We stress in concluding this section that the existence of a deep relation between the problems of compensation of vacuum energy and the physics of dark energy seems to be quite natural. Most probably, these problems can only be solved jointly.

4. Data in favor of cosmological acceleration

4.1 History

In about the 1920s, physicists started in depth discussions of GR-based models of the Universe, including those of a finite-volume universe. The first GR models of the Universe were formulated by Einstein, de Sitter, and Friedmann. It is amazing that the size of the Milky Way was not known at that time, even by the order of magnitude. It was certainly not known whether galaxies other than the Milky Way existed [111].

What was known to astronomers about the size of the Galaxy and our location in it by the 1920s? Accurate measurements of distances and the detection of proper motion of nearby stars (the concept of an immobile stellar sphere had been abandoned long ago) started as early as the 19th century. Counting weak stars in various directions shaped the picture where the Sun was located close to the Galaxy center; indeed, the densities of stars along the Milky Way strip are approximately the same. It was as late as 1920 that H Shapley correctly indicated that the Galaxy center is located in the direction of the Sagittarius constellation: globular star clusters bunched around this direction. Shapley moved the Sun far away from the Galaxy center. We now know that the distance to the center is about 8 kpc (about 25,000 light years) [112], and Shapley, moreover, greatly overestimated that distance. Most astronomers shared the opinion of Curtis, who made the following estimate: "Our Galaxy's diameter is, most probably, no more than 30,000 light years"; they also believed that the Sun is located close to the Galaxy center.

A historical dispute between Shapley and Curtis, 'the great debate', was held on April 26, 1920 at the US National Academy. Curtis supported a small scale of distances in the Galaxy, and Shapley a large one.

In those debates, Curtis quoted some estimates taken from Lundmark's article: "Wolf, about 14,000 light years in diameter; Eddington, about 15,000 light years; Shapley (1915), about 20,000 light years; Newcomb, no less than 7,000 light years and later possibly 30,000 light years in diameter and 5,000 in thickness; and Kapteyn, about 60,000 light years." Shapley's estimate as of 1920 was: "The Galaxy's diameter is about 300,000 light years."

The parties to the dispute failed to persuade each other at that time. It was as late as the 1930s that it became clear that Shapley was right in his statement about the Sun being located far away from the center. (Modern-day data on the Galaxy size: while Curtis's estimate was $\sim$ 30,000 light years and Shapley's estimate was $\sim$ 300,000 light years, the currently adopted value is $\sim$ 100,000 light years or $\approx$ 30 kpc. Shapley and Curtis both made about the same error but in the opposite direction.) Shapley's estimate was only confirmed to be correct when astronomers (Lindblad and Oort) understood the asymmetry of stellar motion around us, in the Galaxy's rotation.

However, Shapley made another mistake: he considered spiral galaxies such as M31 (Andromeda) to be small gas formations located inside the Galaxy or quite near to it. Curtis asserted that those spirals are stellar words similar to the Milky Way. He proved to be right (but underestimated the sizes of the galaxies.)

One can now compare the picture of that small stationary Universe, a model that Einstein had tried to develop before 1920, with the reality known to us nowadays. Not only did the observable scales change (by a factor of millions!): the concept of a nonstationary Universe has emerged.

The plot showing redshifts of objects as a function of their distances is referred to as the Hubble diagram (after his study was published in 1929 [113]). Hubble used his plot to establish the law of how distances to galaxies grow as the redshift increases.

German astronomer Carl Wirtz (1876–1939) was actually the first to work out that law eight years prior to Hubble. Alan Sandage (Baade's PhD student, who worked for Hubble from 1950 to 1953) called Wirtz a "European Hubble without a telescope." Wirtz, who was the real pioneer in observational cosmology, used data on 29 spiral galaxies in 1921 to discover that the further away a galaxy is located, the larger is its redshift [115] (published in June 1922). Unfortunately, due to the poor situation with the German economy, his studies failed to find any support. In 1936, when Hubble was already famous worldwide. Wirtz made a desperate attempt to point out his priority in *Zeitschrift für Astrophysik* [116], but in spite of that he has been virtually forgotten. Nevertheless, the authors of review [117] called Wirtz "a pioneer of cosmological dimensions."

An objective and brief history of how the Universe's expansion was discovered is presented in [118]. The role of observers such as Vesto Slipher is covered in [119].

Van den Bergh [118] asserts straightforwardly: "The myth that expansion of the Universe was discovered by Hubble was disseminated for the first time by Humason (1931) [120]. The actual nature of that discovery proved to be more intricate and interesting."

According to studies of the history of the discovery [121, 122], in 1927 Georges Henri-Joseph Edouard Lemaître (1894–1966) published a study under the title "Uniform Universe of constant mass with an increasing radius with consideration for the beam velocities of extragalactic nebulae" in the *Annals of the Scientific Society of Brussels* [8] (an English translation from the original publication in French appeared later [123, 124].) In that study, he established a linear dependence between the velocity at which a galaxy moves away and the distance to that galaxy. Lemaître was the first to calculate the numerical value of the 'Hubble' parameter in the relation between the velocity v and the distance to the galaxy D. This parameter is known today as H_0 in the formula $v = H_0 D$. Using the measured values of redshifts obtained by Vesto Melvin Slipher (1875–1969) and Gustaf Strömberg (1882–1962) and published in 1925 and the distances to 42 galaxies estimated by Hubble (1926) [125], Lemaître obtained the value $H_0 = 625$ km s^{-1} Mpc^{-1}.

When Wirtz, Lemaître, and Hubble discovered their redshift–distance relation, they did not have a reliable way to measure distances to galaxies. They only assumed how the distance depends on the galaxy size and the light flux; Wirtz only formulated the law in qualitative terms, and Lemaître and Hubble underestimated the actual distances to the galaxies by a factor of almost 10. We recall that Wirtz did not know whether spiral nebulae are located inside or outside the Milky Way and could have made even stronger mistakes. Nevertheless, Wirtz was right in establishing the phenomenon that redshift increases as the distance grows.

4.1.1 History of how hints about the lambda term emerged in observations. Prior to explorations of Ia supernovae, only those possibilities had been considered in depth where deceleration rather than acceleration occurs under the effect of mutual gravitation in the universe for all three types of the universes listed above. Most astronomers did not expect to encounter data that could be interpreted as acceleration of expansion, although such models of the Universe were developed long ago by de Sitter (the vacuum) and Friedmann (for a nonempty universe) with the lambda term. As early as 1922, Friedmann [3] analyzed in depth the fate of the Universe expanding with acceleration for a positive value of Λ. (In the same publication, he also explored cosmological models with a zero or negative value of Λ.)

Models with the lambda term occasionally became fashionable if some peculiarities emerged, for example, in the z-distribution of quasars. Such models were proposed by Shklovsky [126] and Kardashev [127] in 1967. However, those peculiarities later 'dissolved', and the fashion waned. Some convincing arguments in favor of the reality of the lambda term were put forward by Beatrice Tinsley and collaborators in 1975–1978 based on a comparison of the age of the Universe and that of globular stellar clusters [128, 129]. The globular clusters seemed at that time to be older than the Universe in models without acceleration. However, that disagreement has also dissolved.

An important role was played by Fukugita and coauthors [130–132]. First, they stressed that the parameter H_0 cannot be as small as 50 km s^{-1} Mpc^{-1}, the value obtained by Sandage and Tammann [133], and then the correct age of the Universe might be ensured by a positive Λ. It was also shown in [130] that number counts of faint galaxies require a 'sizable' value of Λ.

Many professional astronomers first considered those statements regarding H_0 and Λ as heresy until the Hubble diagram for type-Ia supernovae showed in 1998 that many supernovae are located a little further from us than follows from simple models. The acceleration of the expansion of the Universe (for example, due to the lambda term) could explain those increased distances, and this explanation has become universally adopted. In Section 5, we consider how supernovae are used to measure distances.

5. Data on supernovae and baryon acoustic oscillations

5.1 Cosmography primer: distances in the Universe

If wavelengths of various spectral lines in spectra of distant objects are correctly identified and their laboratory values are known, the redshift can be determined as the ratio of the frequency shift, $z = \omega_1/\omega_0 - 1$. Astronomers can do this with a very high accuracy for nearby objects and with an accuracy of several percent for distant ones. It is thus possible to determine to what extent the wavelength changed from the moment of emission t_1 to the moment of observation t_0. We can use this observation and the condition

$$\omega a(t) = \text{const} \tag{5.1}$$

to find the ratio of scale factors at those two moments of time:

$$\frac{a(t_0)}{a(t_1)} = 1 + z\,.$$

How are distances measured? This problem is more challenging.

We can determine the proper distance. Let the coordinate of the first observer be $r = 0$ and that of a distant galaxy $r = r_1$. Then

$$D_{\text{prop}}(t) = \int_0^{r_1} \sqrt{g_{rr}}\,\mathrm{d}r = a(t)\int_0^{r_1} \frac{\mathrm{d}r}{\sqrt{1-r^2}} \tag{5.2}$$

corresponds to the distance that would be measured by observers located rather closely in the expanding universe between $r = 0$ and $r = r_1$ at the same moment of cosmic time t. To perform such a measurement, they would have to make an agreement in advance. Direct use of D_{prop} is therefore not very helpful.

We now introduce an auxiliary variable

$$D_{\text{flat}}(t) \equiv a(t)\,r\,. \tag{5.3}$$

We refer to D_{flat} as the 'flat' distance. If curvature is zero, $k = 0$, the value of D_{flat} that would be measured at the moment of coordinate time t by a team of observers with rigid rulers between the points with radial coordinates 0 and r would coincide with the proper distance. (We recall that r is dimensionless!) We see below that $D_{\text{flat}}(t)$ is involved in measured distances, and, in general, $k \neq 0$.

Of importance in theoretical cosmology is the concept of the comoving distance

$$D_{\text{com}} = a(t_0)\,\chi\,,$$

where χ is the Friedmann radial coordinate involved in Eqn (2.3). Friedmann used χ for a closed world where $k = 1$. In the general case,

$$r = \begin{cases} \sin\chi\,, & k = 1\,,\\ \chi\,, & k = 0\,,\\ \sinh\chi\,, & k = -1\,. \end{cases} \tag{5.4}$$

We cannot directly measure the 'distances' D_{prop}, D_{flat}, and D_{com} to distant objects in the expanding universe using rulers or radio location. Used instead in cosmography are various definitions of the distances that depend on the measurements that can be actually made (for example, angular size distance based on measurement of the angular size across a standard ruler.) Details are nicely presented by Weinberg [134, 135]. We follow his work and also that of Carrol [136].

5.1.1 Angular size distance. We only present here a sketch of the definition (see the details in [134, 135, 137]).

If there is a source of a known size (a standard ruler), it can be used to determine distances by measuring its angular size and applying straightforward trigonometric formulas for Euclidian space. For example, if the line of sight is perpendicular to the standard ruler, the distance apparently doubles when the angular size halves.

If distances D are large, the angles are small, and a ruler of length R is seen in flat space at the angle $\theta = R/D$. This enables determining the angular diameter distance

$$D_{\text{A}} \equiv \frac{R}{\theta}\,, \tag{5.5}$$

(for small θ).

It can easily be shown that the distance in the expanding universe (both spatially flat and curved) is

$$D_{\text{A}} = \frac{a_0 r}{1+z}\,, \tag{5.6}$$

where r is the dimensionless radial coordinate in the FRW metric (2.4), because light was emitted when the ruler was located $1 + z$ times closer than it is at the moment of reception, when the scale factor is a_0.

The angular diameter distance is primarily used in cosmology in analyzing anisotropy of the cosmic microwave background (CMB). The almost ideal black-body spectrum received from the surface of last scattering is observed in the microwave range. Small perturbations of the average temperature $T \approx 2.7$ K, measured as a function of the angular diameter distance, reflect acoustic oscillation of the primordial plasma that occurred at that period. The maximum fluctuation value, δT, observed for an angular separation of one degree corresponds to the 'last' acoustic wave that entered below the horizon just at the moment of recombination. The length of this wave is known, and we can use the value of that angle to determine the distance to the last scattering surface. These data are evidence, in particular, that our Universe is spatially flat (see details in [137–139]). The angular diameter distance is also used in analyzing BAO (see Section 5.5 below).

5.2 Photometric distance

Of greater value for using supernovae data in cosmography is the so-called photometric distance,

$$D_{\mathrm{ph}} = \left(\frac{L}{4\pi F}\right)^{1/2}, \tag{5.7}$$

where L is the absolute luminosity (i.e., light power) of the source, a standard candle, and F is the flux measured by the observer (the energy that comes to the receiver's unit area per unit of time). This distance is sometimes denoted as D_{L} (luminosity distance). This definition corresponds to the observation that the flux from a source at a distance D in the flat space is $F = L/(4\pi D^2)$.

However, we cannot straightforwardly substitute $D_{\mathrm{flat}} = a_0 r$ from (5.3) in the Friedmann universe, where a_0 is the scale factor at the moment when photons were detected at the comoving coordinate r from the source; as r grows, the flux diminishes not only due to normal dilution (the area of the sphere increases as $4\pi D^2_{\mathrm{flat}}$) but also due to two effects: individual photons experience redshift by a factor of $1+z$, and the rate of arrival of protons is also reduced by a factor of $1+z$. We therefore have

$$F = \frac{L}{4\pi a_0^2 r^2 (1+z)^2},$$

or

$$D_{\mathrm{ph}} = a_0 r(1+z) = D_{\mathrm{flat}}(t_0)(1+z). \tag{5.8}$$

Thus, we obtain that the distances defined in Eqns (5.5) and (5.7) are in general not equal to each other, even in flat space, if the metric is not stationary. The equality $D_{\mathrm{A}} = D_{\mathrm{ph}}$ only holds in a static flat space, while in the expanding universe, $D_{\mathrm{ph}} = D_{\mathrm{A}}(1+z)^2$. This means that the difference between the two distances in the case of a hot plasma blob that we observe at the moment of recombination, when $z \approx 1000$, is one million-fold!

The photometric distance D_{ph} is a quantity that can be measured if we have an astrophysical source whose absolute luminosity L is known (standard candle). However, r in Eqn (5.8) cannot be measured directly, and we therefore have to eliminate it. On a null geodesic line along which photons propagate from a distant object to the observer, we can set $\mathrm{d}\theta = \mathrm{d}\varphi = 0$, whence

$$0 = \mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \frac{a^2}{1-k\bar{r}^2}\,\mathrm{d}\bar{r}^2, \tag{5.9}$$

or

$$\int_{t_1}^{t_0} \frac{c\,\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}\bar{r}}{(1-k\bar{r}^2)^{1/2}}. \tag{5.10}$$

The integral in the right-hand side of (5.10) is an elementary one:

$$\int_0^r \frac{\mathrm{d}\bar{r}}{(1-k\bar{r}^2)^{1/2}} = \begin{cases} \arcsin(r), & k=1, \\ r, & k=0, \\ \mathrm{arsinh}(r), & k=-1. \end{cases} \tag{5.11}$$

We transform the left-hand side of (5.10) as follows:

$$\int_{t_1}^{t_0} \frac{\mathrm{d}t}{a(t)} = \int_{a_1}^{a_0} \frac{\mathrm{d}t}{\mathrm{d}a}\frac{\mathrm{d}a}{a} = -\int_{a_0/a_1}^{1} \frac{a}{a_0}\frac{\mathrm{d}t}{\mathrm{d}a}\,\mathrm{d}\left(\frac{a_0}{a}\right)$$
$$= \int_1^{a_0/a_1} \frac{a}{a_0}\frac{\mathrm{d}t}{\mathrm{d}a}\,\mathrm{d}\left(\frac{a_0}{a}\right).$$

Hence,

$$a_0 \int_0^r \frac{\mathrm{d}\bar{r}}{(1-k\bar{r}^2)^{1/2}} = c\int_1^{z+1} \frac{\mathrm{d}(\bar{z}+1)}{H} = c\int_0^z \frac{\mathrm{d}\bar{z}}{H},$$

showing that this expression reduces to observable values like $1/H$ (equal to $a/\dot{a}$), z (also using $a_0/a_1 = z+1$), and so on. We can thus eliminate r in Eqn (5.8) for D_{ph} if r is expressed through the integral $\int_0^z \mathrm{d}\bar{z}/H$ and the elementary integral $\int_0^r (1-k\bar{r}^2)^{-1/2}\,\mathrm{d}\bar{r}$.

For this, we use Friedmann equation (2.5) and take into account that the vacuum energy can be nonzero:

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

an expression that is equivalent to

$$H^2 = H_0^2\left[\Omega_{\mathrm{m}}(1+z)^3 + \Omega_\Lambda + (1-\Omega_{\mathrm{m}}-\Omega_\Lambda)(1+z)^2\right], \tag{5.12}$$

where $\Omega_{\mathrm{m}} = \mathcal{E}_{\mathrm{m}}/\rho_{\mathrm{c}}c^2$ and $\Omega_\Lambda = \mathcal{E}_{\mathrm{DE}}/\rho_{\mathrm{c}}c^2$ are the density parameters defined above for nonrelativistic matter (whose density ρ_{m} changes in inverse proportion to the comoving volume, i.e., as $(1+z)^3$) and dark energy (this density is constant in the simplest case of the Λ term). We note that ρ_{m} contains the energy of both ordinary and invisible nonrelativistic matter (dark matter, or DM). Below, we use standard abbreviations for cosmological models; for example, CDM denotes the model with cold dark matter, ΛCDM is the same model with the Λ term included, etc.

Having substituted H in $\int_0^z \mathrm{d}\bar{z}/H$ and having expressed r through $\int_0^r (1-k\bar{r}^2)^{-1/2}\mathrm{d}\bar{r}$ in the case $k=-1$ (other cases of k are obtained automatically by analytic continuation), we arrive at a working formula for the photometric distance:

$$D_{\mathrm{ph}}(z) = \frac{c}{H_0}(1+z)\frac{1}{\sqrt{\Omega_k}}\sinh\left\{\sqrt{\Omega_k}\int_0^z \left[\Omega_{\mathrm{m}}(1+\bar{z})^3 \right.\right.$$
$$\left.\left. + \Omega_\Lambda + \Omega_k(1+\bar{z})^2\right]^{-1/2}\mathrm{d}\bar{z}\right\}. \tag{5.13}$$

Here, $\Omega_k \equiv 1 - \Omega_{\mathrm{m}} - \Omega_\Lambda$ and, if $\Omega_k < 0$, the hyperbolic sine (sinh) becomes the trigonometric sine (sin), and $\sqrt{\Omega_k}$ becomes $\sqrt{|\Omega_k|}$. If $\Omega_k \to 0$, the limit can easily be found; sinh disappears from the expression for D_{ph}, and only the integral $\int_0^z [\ldots]^{-1/2}\,\mathrm{d}\bar{z}$ remains.

The function $D_{\mathrm{ph}}(z)$ is now expressed in terms of cosmological parameters. We can see from (5.13) that D_{ph} is very simply related to both the 'flat' value D_{flat} in (5.3) and the Friedmann radial coordinate χ. We can easily derive a similar expression for $D_{\rm ph}(z)$ for a variable density of dark energy if $P = w(z)\,\mathcal{E}$.

If $P_{\rm DE} = w\mathcal{E}_{\rm DE}$, we have

$$H(z) = H_0\left[\Omega_{\rm m}(1+z)^3 + v(z)\,\Omega_{\rm DE} + \Omega_k(1+z)^2\right]^{1/2}, \tag{5.14}$$

where $\Omega_{\rm DE} = \mathcal{E}_{\rm DE}/\rho_{\rm c}$. The function $H^{-1}(z)$ in (5.14) should now be substituted instead of the terms in square brackets in (5.13). The density $\mathcal{E}_{\rm DE}$ is constant in the simplest case of the Λ term (where $w = -1$). The function $v(z)$ follows from the equation of state of dark energy $P(z) = w(z)\,\mathcal{E}_{\rm DE}(z)$,

$$v(z) = \exp\left[3\int_0^z \frac{1+w(z)}{1+z}\,{\rm d}z\right], \tag{5.15}$$

and is equal to unity for a constant Λ term, i.e., for $w = -1$.

We stress that the formulas for the photometric distance $D_{\rm ph}$ of type (5.8) that contain $D_{\rm flat}$ are applicable to any k and not only $k = 0$. The distance $D_{\rm ph}$ relates L and F via the area of the sphere through which the entire power L passes. However, the area of the sphere for FRW-type metrics is related to only $g_{22} \equiv g_{\theta\theta}$ and $g_{33} \equiv g_{\phi\phi}$, i.e., it does not contain $g_{11} = g_{rr}$, the component that contains the 3D curvature parameter k. We always obtain $4\pi a^2 r^2$ for the area of the sphere, and our value of $D_{\rm flat}$ sets the radius of that sphere. For this reason, we obtain a formula for $D_{\rm ph}$ that is also valid in curved 3D space. The 'true' distance $D_{\rm prop}$ includes $\int \sqrt{g_{rr}}\,{\rm d}r$.

Our formula is obviously not applicable in an anisotropic curved space where g_{22} and g_{33} do not have the Euclidian form.

5.3 Cosmic distance ladder. Supernovae

The ladder of cosmic distances is based on primary distance indicators such as trigonometric parallaxes of stars, nearby clusters of stars with common motion, and variable stars, in particular, cepheids (when the Baade–Wesselink method is applied to them).

There are numerous secondary distance indicators that use various gauges for the distances established on the previous ladder rung [140]. For example, the cepheids perform as secondary distance indicators if the period–luminosity relation calibrated using objects located at various distances is employed.

Supernovae, being astronomical objects with the largest luminosity L that can be observed at immense distances, play a very important role in testing cosmological models. We explain in this section how supernovae are used as distance indicators.

Astronomical classification of supernovae is based on their visible light spectra near the luminosity maximum. First, the presence of hydrogen lines is checked.

1. Type I supernovae: *no hydrogen lines* near the brightness maximum. A subtype Ia with the ionized silicon line is distinguished among them. Such supernovae are the brightest stars (i.e., exhibit a very high luminosity at the maximum).

2. Type II supernovae: *clearly pronounced hydrogen lines* in spectra both at the brightness maximum and a long time afterward.

To measure the flux F and luminosity L, astronomers traditionally use stellar magnitudes, which are a logarithmic measure of the flux. The difference between two stellar magnitudes is by definition

$$m_1 - m_0 = -2.5\,(\lg F_1 - \lg F_0) = -2.5\lg\left(\frac{F_1}{F_0}\right). \tag{5.16}$$

The factor -2.5 is chosen because traditionally the light flux of a star of zero stellar magnitude is exactly 100 times larger than that of the fifth stellar magnitude. The zero point m_0 in (5.16) should be set using a standard star (usually Vega, i.e., α Lyrae). The values m and F in (5.16) may have additional indices (for example, for the frequency ν in the case of monochromatic fluxes or U, B, V, etc. if various filters are used.)

The absolute stellar magnitude $\mathcal{M}$ is defined as the stellar magnitude at a standard distance D of 10 pc. Thus, $\mathcal{M} = -2.5\lg L + {\rm const}$, implying that the larger the is luminosity L, the smaller $\mathcal{M}$ is.

In Fig. 1, as an example, we display the absolute magnitude curves $\mathcal{M}(t)$ for several type-Ia supernovae measured with BVI filters (blue, visible, and infrared).

It is now firmly established that type-Ia supernovae are products of a thermonuclear explosion of degenerate stars. Reliable simulation data show that light, i.e., entropy in type-Ia supernovae, is generated in radioactive decays: ^{56}Ni transmutes into ^{56}Co, which transmutes into ^{56}Fe. This mechanism explains, notably, why the shapes of the Ia light curves in B and V filters are so similar (see Fig. 1). The gamma lines that accompany the decay $^{56}{\rm Co} \to {}^{56}{\rm Fe}$ have been recently detected in direct observations by the Integral spacecraft [141].

A large number of type-Ia supernovae can be observed up to $z \sim 1.7$ [142], and spectra can be used to identify three type-IIn supernovae at $z = 0.808$, 2.013, and 2.357 [143]. The redshift of the most distant known type-Ia supernovae is $z = 1.914$ [144].

It was discovered owing to Ia supernovae [145–147] that the cosmological lambda term in the Friedmann model is not zero, $\Omega_\Lambda > 0$, i.e., the Universe is now expanding with acceleration. This can be interpreted in a broader class of models as the presence of 'dark energy'. One of the most significant challenges to fundamental physics is currently to establish the reality and properties of dark energy (and dark matter). Supernovae visible at cosmological distances will keep playing a key role in fulfilling this goal. As we have noted, this discovery was awarded the 2011 Nobel Prize in physics.

The basics of using supernovae for measuring distances can be briefly formulated as follows:

(1) The theoretical dependence of photometric distance (5.7) on the redshift z, $D_{\rm ph\,theor}(z, \Omega_{\rm m}, \Omega_{\rm DE}, w(z), \ldots)$, is set by cosmological models, for example, in Eqns (5.12)–(5.15). The theoretical models can certainly be much richer than those used in the quoted equations; they can be based on non-Einsteinian gravity, extra space dimensions, etc. However, the parameters they contain are quite different.

(2) A comparison of the predicted dependence $D_{\rm ph\,theor}(z)$ with the 'observed' $D_{\rm ph\,obs}(z)$ yields the parameters best fitting observations such as $\Omega_{\rm m}, \Omega_{\rm DE}$, and $w(z)$.

Several methods have been proposed to use supernovae and their gas remnants in cosmography. These methods can be divided into two groups.

The first uses the concept of a standard candle, which we analyze in detail in Section 5.4. The standard candle needs calibration; it is based on the cosmic distance ladder. Supernovae perform here as secondary distance indicators.

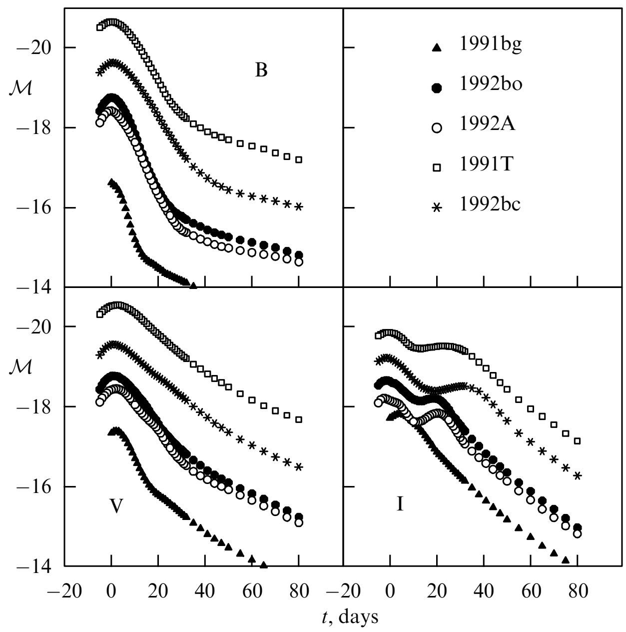

Figure 1. Set of dependences $\mathcal{M}(t)$ for type-Ia supernovae measured using BVI filters. Source: CTIO templates.

Supernovae in the second group of methods are primary distance indicators. These methods are reviewed in Section 5.9.

5.4 Variety of type-Ia supernovae light curves and their use in cosmography

Figure 1 shows that although the shapes of the light curves of type-Ia supernovae are very similar, their luminosities at the maximum are very different, and therefore they are not standard candles!

This diversity was not so obvious to many researchers three or four decades ago. It was almost universally believed at that time that the Ia supernovae are identical, i.e., are standard candles in the sense that absolute luminosity maxima (i.e., light power) of various supernovae are the same. It was found later [148, 149] that this is not true; however, procedures have been proposed that enable determining the absolute luminosity, i.e., standardize the candle [150, 151].

The history of how the applicability of supernovae to cosmology has been developing can be described in detail as follows. Some nonuniformity of the maximum luminosity of type-I supernovae was known long ago; however, the dispersion seemed to be much less than those of other rapidly varying astrophysical objects, for example, novae. Baade [152] noted in 1938 that the dispersion of the absolute stellar magnitude at the maximum is much smaller for supernovae (1.1^m according to his measurements) than for novae. He suggested employing them as distance indicators using the same absolute stellar magnitude for all supernovae. We note that his selection only contained 18 objects, and type-I supernovae had not been divided into subtypes. The idea to use supernovae for cosmography was immediately supported [153, 154]. Pskovskii [155] was the first to introduce the parameters that characterize the light curve (in particular, his famous β for the slope of the light curve 'tail'). He measured β for type-I supernovae in 1967 to show that all these values are very similar, confirming the conclusion that type-I supernovae are suitable for use as distance indicators. It was as late as 1977 that Pskovskii, using large supernovae selections, noticed a relation between the light curve slope (β) and the absolute stellar magnitude of supernovae [148]. He then derived a formula for that correlation.

Thus, Pskovskii was the first to relate the shape of supernovae light curves to luminosity at the maximum. A history of that discovery is described in [156]. Pskovskii was the first to discover not only the variety of type-Ia supernovae but also an important correlation between the peak luminosity $L_{\max}$ and the decline rate $L(t)$ [148, 149]. It is correlations of this kind that are used now to find the absolute luminosity, i.e., to standardize the candle. Details of various approaches to standardizing the candle are reviewed in [121, 122]. Studies [121, 122] found that an important role in Pskovskii's discovery was played by correspondence with B W Rust, a US astrophysicist, who never published his research results in accessible journals, however. Nevertheless, Rust's thesis [150], which was noticed by the classics of research on the expanding Universe [157, 158], played a role in establishing the correlation that enables standardizing the type-Ia supernovae candle.

This correlation can be formulated as follows: "Brighter type-Ia supernovae are slower." Here, the word 'brighter' means 'having high luminosity' and 'slower' refers to the flux decline rate after the maximum. This relation is qualitatively seen in Fig. 1; however, 40 years ago these data were not known. Phillips [151] discovered a similar correlation after Pskovskii. Such correlations are now referred to as the Pskovskii–Phillips (PP) relation, WLR (width luminosity relation), or BDR (brightness decline rate). An example of modern data for the PP correlation $M_{\rm B} - \Delta m_{15}$ is displayed in Fig. 2 (courtesy of the authors of [159]). Here, Δm_{15} denotes the luminosity variation expressed in stellar units 15 days after the maximum in the B-filter. Larger Δm_{15} implies faster decline; then, indeed, the supernova is weaker in its maximum.

If the decline rate of the observed flux $F(t)$ for a faraway supernova has been measured, then relations similar to those illustrated in Fig. 2, including the time dilation factor $(1+z)$, can be used to obtain L and the photometric distance $D_{\rm ph}$. In this way nonstandard candles are standardized.

Next, an equation similar to (5.13) is used and the least-square method is applied to find the cosmological parameters that best reproduce the observational data. Both standardization and determination of cosmological parameters are done in practice as part of a single global fit. Such an approach involves errors. This implies that χ^2 includes not only parameters related to cosmology but also those associated with the light curve. A nonzero Λ was obtained in [145–147] just in this way.

Similar activities are in progress, and the accuracy of data obtained is increasing; however, it cannot be said that it is to be very high. For example, the first results of the Legacy Survey (SNLS) group [160] yielded the best fit $\Omega_{\rm m} = 0.263 \pm 0.042\,(\text{stat.}) \pm 0.032\,(\text{syst.})$, for $D_{\rm ph}(z)$ implying $\Omega_\Lambda \approx 0.74$ for flat Λ-cosmology. Using Eqns (5.14) and (5.15) for dark energy, the SNLS group obtained $w = -1.023 \pm 0.090\,(\text{stat.}) \pm\ 0.054\,(\text{syst.})$ w in the equation of state $P = w\rho$ is constant. These results have been obtained by combining data on supernovae with those from the Sloan Digital Sky Survey (SDSS) for baryon acoustic oscillations (BAOs), i.e., data on the distribution of galaxies in space (see Section 5.5). The same French group [161] in a later and important study used a compilation of data from the collaborations SNLS, SDSS, Nearby SNe, and HST. The combined data on the CMB, BAOs, and SN Ia's yield,

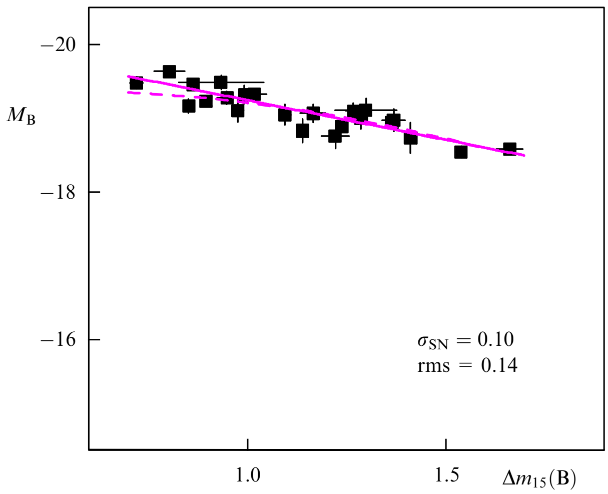

Figure 2. Pskovskii–Phillips relation for Carnegie project data [159]. The horizontal axis shows a flux decrease 15 days after the maximum, Δm_{15}, and the vertical axis, the absolute stellar magnitude at the luminosity maximum.

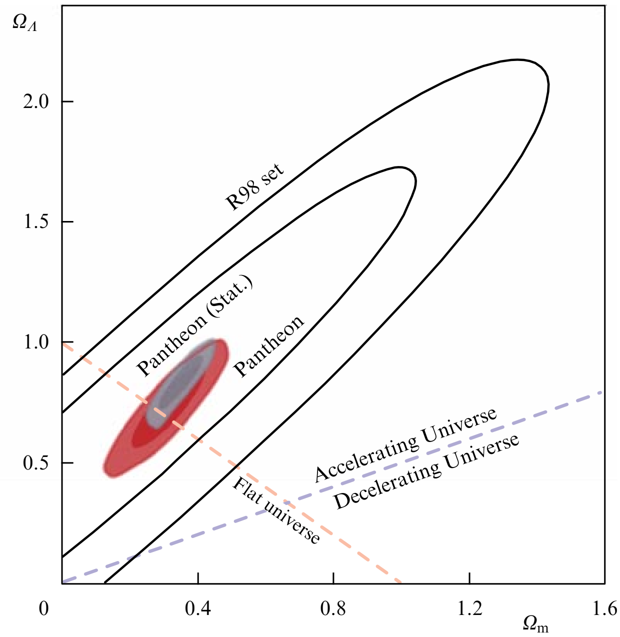

Figure 3. (Color online.) Evidence of the existence of dark energy based on data on type-Ia supernovae alone [163]. Confidence contours are shown for 68% and 95% levels for the cosmological parameters Ω_m and Ω_Λ in the CDM model with fixed $w = -1$ but variable Ω_k for both the first set of supernovae R98 [146] and the Pantheon set. The Pantheon constraints with systematic uncertainties are shown in red and purely statistical uncertainties are shown in grey.

according to [161], the Hubble parameter $H_0 = 68.50 \pm 1.27$ km $^{-1}$ Mpc^{-1}, a value that is somewhat lower than that obtained by Riess's group, $H_0 = 73.8 \pm 2.4$ km s^{-1} Mpc^{-1}, but agrees well with recent results of WMAP (Wilkinson Microwave Anisotropy Probe). Analysis [161] yielded the value $\Omega_m = 0.295 \pm 0.034$ (stat. + syst.) for flat ΛCDM cosmology, in good agreement with values obtained in measurements of CMB fluctuations performed in the Planck and WMAP experiments. The parameter $w = -1.018 \pm 0.057$ (stat. + syst.) was obtained for the flat universe with the CMB data.

In subsequent Section 5.5, we briefly describe the prospects provided by the BAO method for cosmological research. This method is an important complement to those based on supernovae properties. Figure 3 shows recent results obtained using measurements of supernovae alone [163], with 276 type-Ia supernovae ($0.03 < z < 0.65$) collected from observations made in the PanSTARRS1 project and useful estimated distances to type-Ia supernovae taken from the reviews by SDSS, SNLS, HST, etc. The largest united set of type-Ia supernovae was thus created containing 1,049 supernovae in total in the range $0.01 < z \lesssim 2$, which was named Pantheon Sample by the authors.

All results for cosmological parameters extracted from the above-mentioned observations of supernovae have been obtained using the approach standard for FRW cosmology models with dark energy, where the scale factor evolution is described using the energy density averaged over very large scales ($>$ 100 Mpc). Averaging nonlinear Einstein equations is actually a nontrivial problem. Deviations can occur from the solutions obtained by means of ordinary averaging (the so-called backreaction effect). Without entering into the disputes related to that problem, we only give references to recent studies where, in particular, data on type-Ia supernovae are used and this effect is discussed [164–166].

5.5 Baryon acoustic oscillations (BAOs)

Baryon acoustic oscillations (BAOs) is the name that is used for primordial acoustic waves propagating in a hot plasma of photons and baryons due to the pressure of photons in the early Universe. We here consider the main concepts of the BAO method as applied to determining cosmological parameters (including the proportion of dark energy). Details can be found in [167], a good review that also contains extensive useful information about other methods used for studying dark energy, including supernovae. Review [168] is dedicated to BAOs. The basics needed for understanding the BAO theory can also be found in older reviews on the development of perturbations in the Universe; for example, review [169] is of use in this regard.

Baryons and photons in the early Universe are strongly coupled, and perturbations propagate in plasma in the form of acoustic oscillations. After hydrogen recombines at the redshift $z_r \approx 10^3$, photons and baryons virtually cease interacting, and light pressure on baryons vanishes. This phenomenon is referred to as decoupling of photons and baryons. The waves in baryons, i.e., propagating sound waves, halt shortly after that due to cosmological expansion and leave a footprint in the distribution of matter on a scale that corresponds to the distance that the sound waves pass prior to that epoch (i.e., on the acoustic horizon scale).

The sound waves propagate until the recombination epoch $t \approx 4 \times 10^5$ years with a velocity $\approx c/\sqrt{3}$ (somewhat smaller before the recombination itself) and freeze on a scale that in the modern epoch yields a comoving size $l_{BAO} \approx 150$ Mpc (i.e., $l_{BAO} \approx 100\, h^{-1}$ Mpc). This value of the halt radius has been derived in quantitative terms on the basis of simplified models in [170, 171] (see also [172, Section 7.1.2]).

We now formulate these statements in a more accurate form. While baryons are strongly coupled to photons, we have the speed of sound (for $c \equiv 1$)

$$u_s = \left(\frac{dP}{d\rho}\right)_S^{1/2} \approx \frac{1}{\sqrt{3}}\,. \tag{5.17}$$

Here, the subscript s refers to sound and S denotes entropy (the speed of sound is adiabatic). Because $u_s = a(t)\,dr/dt$, where r is the Robertson–Walker radial coordinate in (2.4), we have the acoustic horizon coordinate

$$r_s = \int_0^{t_r} \frac{u_s\,dt}{a(t)}\,, \tag{5.18}$$

and the horizon size

$$l_s(t_r) = a(t_r)\,r_s = a(t_r)\int_0^{t_r} \frac{u_s\,dt}{a(t)}\,. \tag{5.19}$$

For any redshift z after recombination, the matter correlation function must have a singularity at the length

$$l_{\mathrm{BAO}}(z) = \frac{1+z_r}{1+z}\,l_s(t_r)\,. \tag{5.20}$$

If, prior to recombination, we approximately set $u_s \approx 1/\sqrt{3}$ (5.17) and $a(t) \propto t^{1/2}$ (this is only valid in the radiation-dominated epoch), we arrive at the estimate

$$l_s(t_r) = t_r^{1/2}\int_0^{t_r} \frac{dt}{\sqrt{3}t^{1/2}} = \frac{2}{\sqrt{3}}\,t_r \approx 135\ \mathrm{kpc}\,, \tag{5.21}$$

for the recombination time $t_r \approx 380{,}000$ years. For the nonrelativistic law $a(t) \propto t^{2/3}$, we obtain another estimate: $l_s = \sqrt{3}t_r \approx 200$ kpc. The expansion law is actually somewhere between these two limit cases. Also, the speed of sound in (5.17) differs from $1/\sqrt{3}$ due to the contribution of baryons to pressure:

$$u_s^2 = \frac{1}{3}\,\frac{1}{1+R_B}\,, \tag{5.22}$$

where

$$R_B = \frac{3\rho_b}{4\rho_\gamma} \approx \frac{3\times10^4\Omega_b h^2}{1+z}\left(\frac{T_0}{2.725\ \mathrm{K}}\right)^4 \tag{5.23}$$

(see Eqn (9.17) for $\Omega_b h^2$ below).

As a result, we obtain $l_s(t_r) \approx 150$ kpc for the recombination epoch: a CMB radiation spot of just that angular size determines the main peak in the power spectrum of temperature perturbations expanded over spherical harmonics. The same scale sets the present-day size of the correlation length of acoustic oscillations $l_{\mathrm{BAO}}(0) \approx 150$ Mpc, because $z_r \approx 10^3$.

The data collected by WMAP over five years of observations of CMB yield [173]

$$l_{\mathrm{BAO}}(0) = 153.3 \pm 2.0\ \mathrm{Mpc}\,, \quad z_d = 1020.5 \pm 1.6\,. \tag{5.24}$$

This value for the present-day BAO scale is obtained from the acoustic horizon at the moment when the 'epoch of dragging' photons by baryons ended. Equation (5.24) contains z_d instead of z_r because two different definitions are used in publications for the recombination epoch (this circumstance is of no importance for order-of-magnitude estimates but is essential for accurate calculations): the moment of the last photon scattering is distinguished from the end of the baryons-dragged-by-photons epoch. The last scattering is the moment when the optical depth in Compton scattering of photons to today's observer becomes less than unity, without taking the subsequent reionization into account.

The moment t_d when the drag epoch ends is defined as the time when baryons decouple from photons. Actually, baryons couple to electrons by electric fields, and the latter couple to photons owing to Compton scattering. The cross section of the process is the same, the Thomson cross section σ_T, as in the last photon scattering, but the number of photons is much larger than the number of baryons (and electrons) in the hot (determined by entropy) universe. The mean free path of photons $(n_e\sigma_T)^{-1}$ is therefore many orders of magnitude larger than that of baryons $(n_\gamma\sigma_T)^{-1}$ in the sea of photons (because $n_\gamma/n_b \sim 10^9$ and, prior to recombination, $n_\gamma/n_b \sim 10^9$).

Actually, for the analysis of how photon and baryons exit the strong coupling regime to be correct, we should compare the efficiency of their momentum exchange rather than mean free paths of the particles. We perform this analysis following study [174], which showed that using the hydrodynamic approximation for Compton scattering results in momentum exchange between fluids of baryons and radiation that tends to equalize the average velocities of the liquids u_b and u_γ. However, momentum densities $(\rho_\gamma + p_\gamma)\,u_\gamma = (4/3)\,\rho_\gamma u_\gamma$ and $(\rho_b + p_b)\,u_b \approx \rho_b u_b$ are not equal, even if the velocities are. Momentum conservation (Euler equation or, more accurately, Navier–Stokes equations) requires that the acceleration of the baryon fluid due to Compton drag be multiplied by $R_B^{-1} = (4/3)\,\rho_\gamma/\rho_b$ [cf. Eqn (5.23)] compared with the same value for the photon fluid. The process dynamics are thus governed not by the immense number n_γ/n_b but by R_B, which is close to unity for $z \approx 10^3$. A beautiful pedagogical presentation of this issue is offered in book [175]. It is also shown there that analytic theory [174] is fully confirmed by numerical simulations of the kinetic equations for photons and baryons in the recombination epoch of the Universe.

Because the hydrodynamic approximation for photons is invalid after recombination, the difference between the moment of last scattering and the end of the drag epoch can be described by paraphrasing Weinberg [135] (page 441 of the Russian edition): "Strictly speaking, the moment of the drag epoch end, t_d, should be chosen on the recombination stage when a typical electron stops exchanging appreciable momentum with the photons, rather than the slightly earlier time of the last scattering t_{ls} when a typical photon stops exchanging appreciable momentum with the electrons. Because R_B is not very different from unity, there is little difference between these times." The history of ionization determines to what extent the moment of drag epoch end t_d differs from the moment of last scattering (see approximating formulas in [174], Appendix E).

We must take into account that $t_d > t_{ls}$ only if $\Omega_b \ll 1$. If the inequality $\Omega_b h^2 > 0.03$ had been satisfied, the last scattering would have occurred later than the baryon drag epoch ended.

The characteristic BAO scale is thus the sound horizon scale at the end of the strong coupling epoch when photon pressure can no longer prevent gravitational instability in baryons (which occurs somewhat later than the last scattering of photons because the parameter $\Omega_b h^2 = 0.022$ is small.)

It is frequently said that BAOs leave a clear-cut footprint on the characteristic scale of matter clustering; however, we should take into account that this footprint is only exhibited in statistical terms as a local bump in the correlation function at $r \approx 150$ Mpc.

We now explain how BAOs are used for measuring cosmological distances. Galaxies or galaxy clusters are scanned in certain celestial areas within a rather large solid angle. The color indicators of galaxies are used to determine the so-called photometric redshift. Such scanning enables a BAO footprint to be identified by photometric redshift in the angular clustering of galaxies within rather narrow intervals. Let the angular diameter of such a footprint be α. Because the linear diameter $r \approx 150$ Mpc defines a standard ruler, we can measure the angular size distance $D_{\mathrm{A}}(z) = R/\alpha$ according to Eqn (5.5). Spectroscopic studies of the galaxies in the same volume enable measuring redshift values more accurately and observing BAOs along the line of sight and not by the angle alone. Owing to this, $D_{\mathrm{A}}(z)$ can be measured more accurately. Also, measured differences between velocities of the galaxies on the BAO scale can be used to directly determine the Hubble parameter as a function of redshift, $H(z)$. Other matter distribution indicators can also be used to measure BAOs. Because the BAO scale is known in absolute units (on the basis of simple physical calculations and the values of the parameters measured for the CMB), the $D(z)$ determined using the BAO method is expressed in absolute units, Mpc, rather than h^{-1} Mpc units, as a result of which BAO and measurements of distances to supernovae yield different values for the same redshift. Recent results for the cosmological parameters from the detailed Baryon Oscillation Spectroscopic Survey (BOSS) are reported in [176]. Data from the Planck and BOSS projects have been used to obtain $\Omega_k = 0.0003 \pm 0.0026$ and $w = -1.01 \pm 0.06$, in good agreement with the flat ΛCDM model. Adding data on type-Ia supernovae improves the constraint on the parameter in the equation of state of dark matter to $w = -1.01 \pm 0.04$ to yield the Hubble constant $H_0 = 67.9 \pm 0.9$ km s^{-1} Mpc^{-1}.

We can see BAOs directly in observational data on galaxy clusters as they are processed, for example, in [177] (see Figs 4 and 5). Figure 4 shows two areas on the celestial sphere, and points represent positions of galaxy clusters. The distribution of points is apparently random and fully uniform.

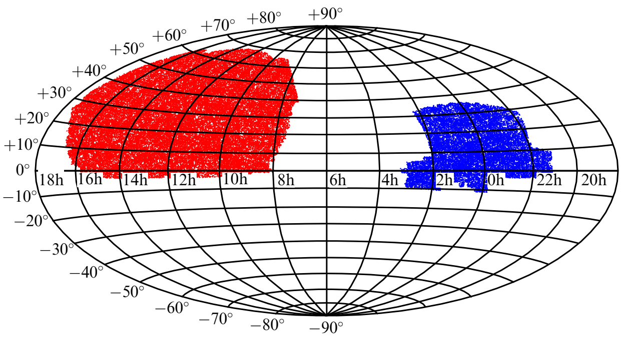

Figure 4. Set of galaxy clusters [177] in two celestial areas. The distribution of clusters seems to be uniform (cf. Fig. 6b), unless a detailed analysis is performed.

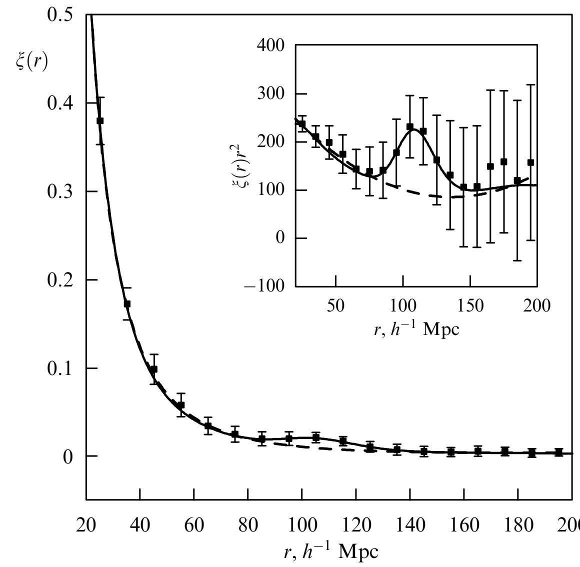

Figure 5. Correlation function for galaxy clusters [177] exhibiting an apparent maximum due to BAO.

The BAO footprint in the distribution of matter is exhibited in the distribution of both galaxy clusters and field galaxies [178]. Observations of that footprint proved to be a powerful and reliable probe of dark energy (see, e.g., [179]). The observable BAO scale determined in scans of galaxies can be compared with the actual physical scale of the sound horizon, which can be independently estimated using initial data on the CMB, to establish a correspondence between observational coordinates and physical coordinates; this correspondence is sensitive to the history of expansion and hence to dark matter properties (see, e.g., [180]). The accuracy with which the BAO correlation function can be measured and hence therefore constraints on dark energy obtained from BAO measurements depend on the BAO footprint strength and the accuracy with which the BAO footprint can be separated from broadband noise in the perturbation power spectrum. Due to the formation of a large-scale structure and the observable effect of spatial distortions caused by redshift, the BAO footprint is expected to gradually fade out, along with the evolution of the structure; this footprint is therefore much weaker at small redshifts, where expansion of the Universe is primarily driven by dark energy. We give an illustrative example of how the signal from a ring-shaped distribution of points on a 2D plane is blurred: when many random points are located on a small number of rings, the signal can be identified without any problem. But if the same number of points is scattered across a large number of rings (the number of points per ring is small), a special procedure is needed to process and identify correlations (see Fig. 6).

However, the main difficulty in reconstructing the density distribution is due not to the detection of the signal on the background of random distribution of points but rather to the fact that the density peak in the current epoch corresponds to a size of 150 Mpc; to determine the peak with an acceptable accuracy, a large space volume must therefore be explored. The reconstruction problem is described in detail in reviews [181, 182].

Measuring the characteristic scale of BAOs in the correlation function of various matter distribution indicators is an efficient tool for exploring cosmic expansion and a reliable method for obtaining cosmological parameters. The BAO peak in the correlation function for a redshift z emerges for the angular distance of objects $\Delta\theta = l_{\mathrm{d}}/[(1+z)\,D_{\mathrm{A}}(z)]$, where $D_{\mathrm{A}} = D_{\mathrm{ph}}/(1+z)^2$ is the angular-diameter-based distance and $l_{\mathrm{d}} = l_{\mathrm{s}}(z_{\mathrm{d}})$ is the sound horizon for the redshift of decoupling z_{d} (drag), i.e., the epoch when baryons decouple from photons. The BAO correlation function is also exhibited in the redshift $\Delta z = l_{\mathrm{d}}/D_H$ where $D_H \equiv c/H(z)$. The measured position of the BAO peak at

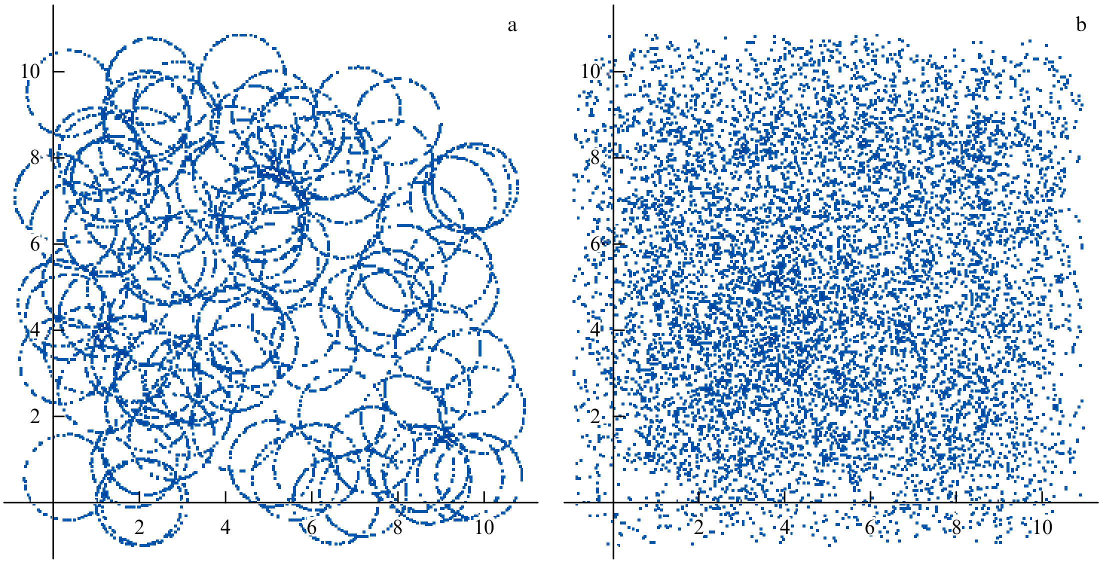

Figure 6. (a) 100 random points for each of 100 randomly located circles. (b) 1000 circles containing 10 points each. The algorithm is taken from http://mathematica.stackexchange.com/question/57938/how-to-generate-random-points-in-a-region.

some z sets therefore restricts combinations of the cosmological parameters that determine D_H/l_d and D_A/l_d at that redshift [183].

5.6 BAOs in the correlation function of galaxies

The BAO peak was primarily observed in the correlation function of a pair of galaxies obtained in redshift scans. Because of the low statistical significance, bounds on D_V/l_d alone could be obtained, where

$$D_V(z) \equiv \left\{z(1+z)^2 D_H D_A^2\right\}^{1/3}. \qquad (5.25)$$

Both the physics of and data on BAOs depend on the content of matter in the universe. They therefore a priori depend on the selected dynamic structure (see, e.g., review, [184]). The analysis in [185] showed that the ratio $D_V(z)/D_V(z_0)$ weakly depends on the dynamic structure; reliable constraints on cosmological parameters can therefore be obtained using such relations.

Results of the Planck collaboration [186] and study [161], where D_V/l_d was BAO-measured at $z = 0.106, 0.35$, and 0.57 on the basis of [187–189], and the results of publications [190] and [191] for $z = 0.35$ and $z = 0.2$ yield

$$\frac{D_V(z)}{D_V(z=0.35)} = (0.335 \pm 0.016,\ 0.576 \pm 0.022,\ 1.539 \pm 0.039) \qquad (5.26)$$

for the respective values $z = (0.106,\ 0.2,\ 0.57)$.

This approach to BAOs and publicly accessible data on type-Ia supernovae were used in [183] to obtain observational constraints on the class of modified gravity models that result at low redshifts in power cosmology (i.e., models in which the scale factor grows as a power of time). The spatially flat universe was shown to be well suited for describing BAOs and type-Ia supernovae if the expansion regime is $a(t) \propto t^{\beta}$ with β close to 3/2.

5.7 Summary of results on supernovae combined with BAOs

New results of the same SNLS group [192] yield the values $\Omega_m = 0.173^{+0.095}_{-0.098}$ and $w = -0.85^{+0.14}_{-0.20}$ or $\Omega_m = 0.214^{+0.072}_{-0.097}$ and $w = -0.95^{+0.17}_{-0.19}$ (all error intervals here are purely statistical) on the basis of richer statistics of observations for two procedures of fitting light curves to observational data (the observational points on the light curve are usually very sparse and therefore special procedures are needed to identify the maximum, values Δm_{15}, etc.). They obtained $w = -0.91^{+0.16}_{-0.20}$ (stat.)$^{+0.07}_{-0.14}$ (syst.) for the parameter in the equation of state for dark energy (which is considered to be constant to at least $z = 1.4$) in a flat universe on the basis of data on supernovae alone. These values of w agree with the cosmological constant, which requires $w \equiv -1$. The results reported of the CfA3 review published by the Harvard-Smithsonian Center for Astrophysics [193] (see Figs 7 and 8) combined with BAO data used as an a priori distribution of probabilities yield $1 + w = 0.013^{+0.066}_{-0.068}$(0.11syst.), also in good agreement with the cosmological constant. BAO data combined with their set of supernovae also yield $\Omega_m = 0.281^{+0.037}_{-0.016}$ and $\Omega_\Lambda = 0.718^{+0.062}_{-0.056}$. Although the results of various groups formally agree with each other, we can nevertheless see that the difference between them is larger than could be expected based on estimates of their own errors.

The cosmological parameters extracted from observations of supernovae and BAOs are continuously being refined. For example, study [161] is in a sense a continuation of paper [192] described above. Also, some inaccuracies of the earlier analysis have been corrected. Taking data on BAOs into consideration, the value $w = -1.027 \pm 0.055$ for the parameter in the equation of state of dark energy was reported in [161], in good agreement with their value with the CMB taken into account, $w = -1.018 \pm 0.057$ (stat. + syst.).

The combined results displayed in Fig. 9 are taken from study [163], which contains the most complete set of type-Ia supernovae as of the end of 2017.

Data on BAOs [194] combined with data on CMB temperature fluctuations measured by the Planck mission, polarization measured by WMAP-9, and a 6dG scan of BAOs failed to indicate any deviations from the flat ΛCDM model. The Hubble parameter in this model proves to be

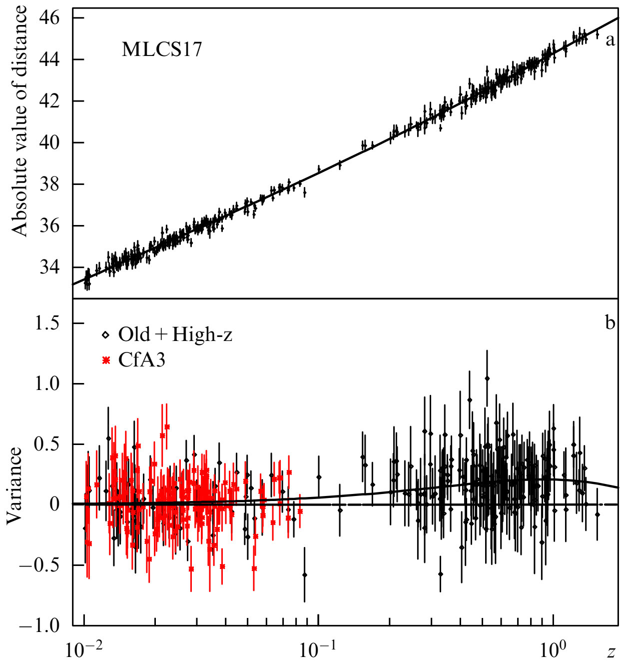

Figure 7. (Color online.) Hubble diagram for (a) type-Ia supernovae and (b) its variances from the CfA3 catalog [193]. New data are shown in red, and older data in black. The variances pertain to the universe without dark energy, $\Omega_m = 0.27$ and $\Omega_\Lambda = 0$. The best cosmological fit is shown in the panel of variances. See a more modern version of the Hubble diagram for type-Ia supernovae in Fig. 8 in [161].

$H_0 = 67.15 \pm 0.8$ km s^{-1}Mpc^{-1}. The variable in the equation of state of dark energy is $w_{DE} = -1.080 \pm 0.135$. The curvature in a nonflat ΛCDM model is $\Omega_k = -0.0043 \pm 0.0047$, compatible with zero.

5.8 Systematics and dependence on z

An important factor must be taken into account in passing from a description of the Pskovskii–Phillips relation to the development of the Hubble diagram. It was found long ago that the luminosity of type-Ia supernovae depends not only on the light curve shape (the Pskovskii–Phillips relation) but also on the supernova color. Moreover, the dependence on color is significant. The absolute stellar magnitude of type-Ia supernovae is defined in the most popular standardization model SALT2 [195] as: $M = M_0 + \alpha X_1 - \beta C$, where X_1 is an analog of Δm_{15} (responsible for the light curve shape) and C is the color. Tripp [196] seems to be the first to have introduced the color parameter, although Riess was also taking it into account in some way in his MLCS method.

It is also well established (although this correction has not yet been introduced into cosmological analysis) that the absolute stellar magnitude of type-Ia supernovae depends on the host galaxy. The following physical parameters are used to characterize the host: the division into subtypes (spiral, elliptic [197, 198]), the mass of the galaxy stellar component [161, 199, 200], and global and local star formation rates [201, 202] (the immediate surroundings of a supernova are explored [203]). The corrections related to the host galaxy properties are introduced into in the cosmological analysis literally by hand. For example (see Eqn (5) in [161]), a small term $10^{10} M_\odot$ is added to M_0 for supernovae whose galaxy stellar mass is more than Δ_M, i.e., some part of the Hubble diagram (corresponding to massive host galaxies) is simply shifted with respect to the other part (with low-mass galaxies).

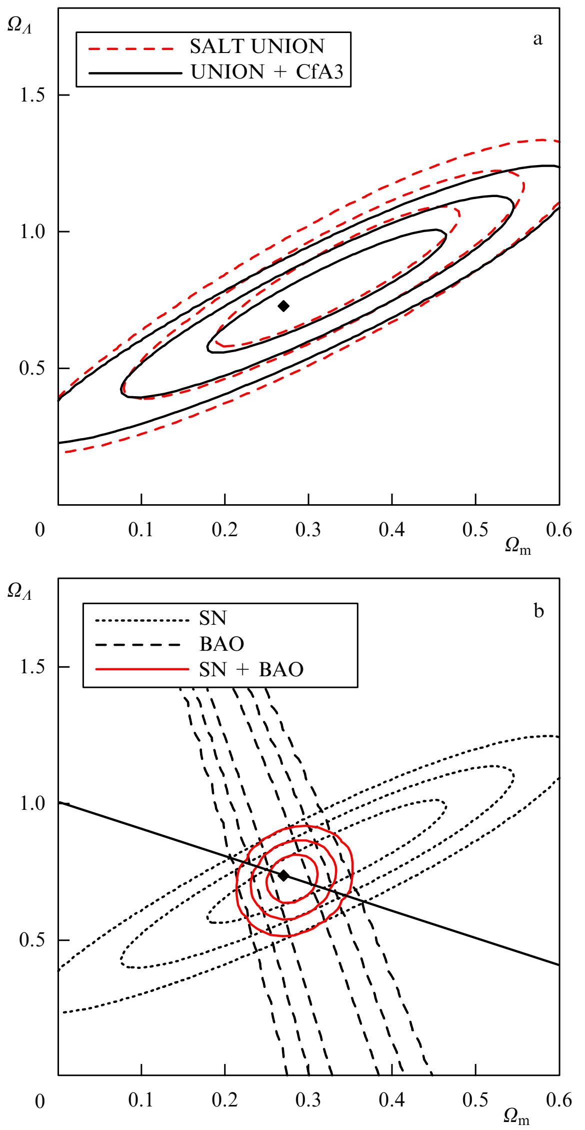

Figure 8. (Color online.) Parameters Ω from the review CfA3 [193]. Contours of allowed values of Ω_Λ and Ω_m at $1 + w = 0$ without a flat-space assumption. Concordance cosmology $\Omega_\Lambda = 0.73$ and $\Omega_m = 0.27$ is displayed with the dot. Figure a shows that adding the CfA3 set significantly narrows the contours along the Ω_Λ axis. Figure b shows a combination of the SN contours with BAO-based constraints, while the straight line describes the flat Universe, $\Omega_\Lambda + \Omega_m = 1$.

There are many factors that can affect the results obtained in cosmology using type-Ia supernovae calibrated by means of the Pskovskii–Phillips relation; for example, light absorption and scattering in the inter-galaxy space or in host galaxies (see studies [159, 204] and the references therein), changes in the metallicity of supernovae progenitors, and the relative role of various pre-supernovae as the Universe ages.

We try to cursorily explain the role of these factors in applications of supernovae in cosmology to the reader trained as a physicist.

The most difficult task is to determine how light emitted by a supernova is absorbed and scattered on its way to the observer. X-ray astronomy provides plentiful information about hot intracluster gas, but information about intercluster baryon gas is very scarce. This is the well-known missing baryon problem: cosmological nucleosynthesis predicts that

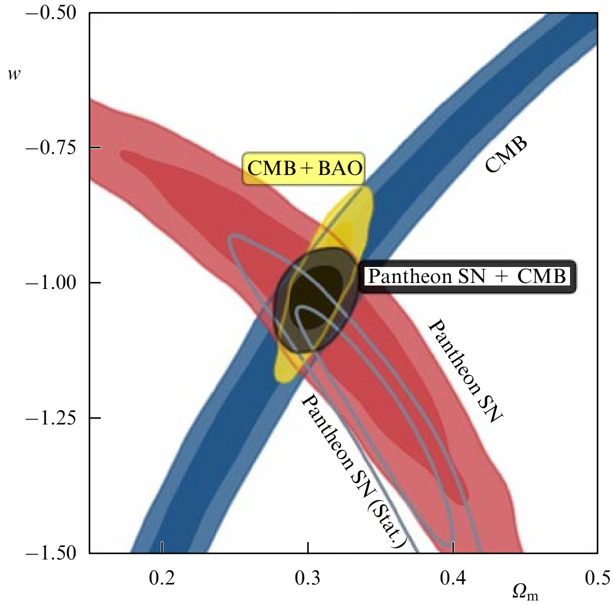

Figure 9. (Color online.) Confidence contours for 68% and 95% levels for Ω_m and w in CDM models based on the results in [163]. Constraints are shown based on the CMB (blue color tints), supernovae with systematic uncertainties (red color), supernovae with statistical uncertainties alone (grey lines), and supernovae + CMB (black lines).

about 4–5% of the energy density is localized in baryons, and about the same number follows from an analysis of the CMB. The total number of all observable baryons (in gas and stars) is at the same time significantly smaller. These hidden baryons can be in a state of a warm intercluster gas. The temperature of that gas can be too high to allow detecting it in the visible or near-ultraviolet range but too low for noticeable emission in the X-ray range. That gas can in principle contain an admixture of 'grey' dust (whose absorption does not depend on wavelength). As a result, we see the distant galaxies dimmed, this effect being possibly misinterpreted as an effect of accelerated cosmological expansion. We refer in this regard to [205], where grey dust and its effect on the Hubble diagram is explored. The authors also analyzed the effect of the evolution of mass of merging white dwarfs with Hubble time. However, this effect cannot occur at redshifts larger than $z \sim 1$: supernovae this distant do not exhibit the accelerated expansion effect.

The grey-dust hypothesis may seem to be absolutely artificial; nevertheless, it cannot be fully rejected until the problem of hidden baryons is resolved.

Very scarce information is also available about the so-called 'reddening' in supernovae's host galaxies. This is the same effect of light absorption/scattering but occurring in the medium that surrounds the exploding star. Properties of that medium, which can vary strongly over space, can be explored for nearby galaxies, but the problem becomes very difficult for distant objects.

Metallicity, i.e., the content of elements heavier than helium in pre-supernovae, can affect the behavior of supernovae's light curves, i.e., Pskovskii–Phillips relations, and hence the candle standardization.

Several quite permissible scenarios have been proposed for the birth of type-Ia supernovae in a very complicated and intricate evolution of binary stars (single degenerate stars, white dwarfs, cannot explode). It is difficult to estimate the relative role of these scenarios in the nearby universe (published data disagree with each other). Such an estimate becomes an even more challenging task in the distant young universe.

All these factors (light absorption, metallicity, the relative role of evolution scenarios) can vary as the age of the universe grows: stars produce an increasingly large amount of metals that pollute the interstellar environment. Systematic errors thus occur in determining distances and cosmological parameters by means of supernovae.

Another possible source of errors is related to incorrect classification and an admixture of unusual events of the type of Ia supernovae. For example, a peculiar subclass of type-Ia of supernovae, the subtype SN 2002cx, has been discovered. These supernovae are weak but slow (see Fig. 10 taken from [207]), i.e., they behave in a way quite opposite to that prescribed by the Pskovskii–Phillips relation used in cosmology; according to that relation, slowly declining type-Ia supernovae are the brightest. We now imagine that the number of SN 2002cx-type events increases as the cosmological redshift z grows. We then conclude based on the Pskovskii–Phillips relation established for nearby type-Ia supernovae, i.e., at $z = 0$, that the type-Ia supernovae at large z seem on average to be dimmer, and hence the photometric distance to them is larger than for the true Ω_Λ. Thus, a false contribution to dark energy may be obtained.

It is quite relevant to quote Conley et al. [192]: "Evolution in the absolute magnitude of SNe Ia with redshift is not constrainable without a detailed physical model because it can in principle mimic any cosmology." This does not imply that supernovae cannot be used for reliable cosmography; just new approaches to the problem should be developed. One such new approach to type-Ia supernovae is presented in [208]. The author of this study notes: "These results imply that at least 3/4 of the variance in the Hubble variances in current supernova cosmology analyses is due to SN Ia." Results like this confirm our apprehension about using type-Ia supernovae for analyzing dark energy as is done, for example, in [209, 210], because the properties of stars change with the age of the Universe. The authors of [208] use their observations in an attempt to estimate the quantitative measure of astrophysical uncertainty. They accurately proceed in the local Universe and improve standardization options; however, at larger distances unknown astrophysics

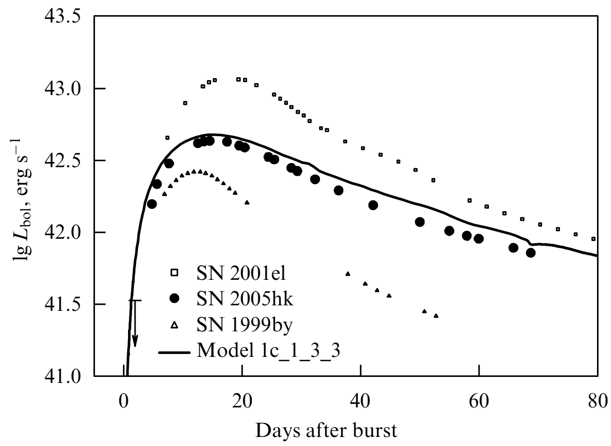

Figure 10. (Color online.) SN 2005hk in comparison with two 'normal' SN Ia's.

inevitably introduces unknown systematics because type-Ia supernovae are secondary distance indicators, and it remains unclear how their improved calibration at $z \sim 0$ would vary at large z (their estimate ''standardizing high-redshift supernovae to within 0.06 ± 0.07 magnitudes'' remains conceptually statistical.)

A new, direct dense shell method for measuring distances in cosmology has been proposed in [211–213] with the participation of the authors of this review. This method is discussed in the next section.

5.9 Supernovae as primary distance indicators

The 'standard candle' method requires knowledge of the distances to a large number of supernovae measured using another, independent method involving the cosmic distance ladder [133]. Otherwise, without large statistics on objects with a known distance having been accumulated, the candle standardization procedure cannot be calibrated (see, e.g., reviews [156, 214]). This implies that supernovae are used in this method as secondary distance indicators.

Another group of methods enables using supernovae as primary distance indicators. For example, luminosity and light curves of type-II supernovae (SN II) are very different, and, generally speaking, these supernovae are not suitable for the standard candle method. But they have a significant advantage: distances can be measured directly, for example, using the expanding photosphere method (EPM) [215]. This method needs neither candle standardization nor the cosmic distance ladder.

High-quality spectral data enable determining distances using the spectral-fitting expanding atmosphere method (SEAM) [216]. Unlike EPM, SEAM does not involve an assumption that supernovae radiate as black bodies.

We describe the development of a new method to measure distances in cosmology that is based partly on EPM and SEAM and partly on ESM (expanding shock front method) [217]. This method uses data on type-IIn supernovae with the highest luminosity, which have been recently not only discovered in large numbers but also explored in depth [218].

It was proposed to refer to this method as the dense shell method (DSM) because the luminosity of type-IIn supernovae is due to propagation of a thin dense layer in the surrounding environment. Supernovae SN 2006gy and SN 2009ip were used as an example to show that this method is valid: the distances to those supernovae have been determined without applying the standard distance calibration. The new method needs neither the standard candle approximation used for type-Ia supernovae nor the cosmic distance ladder.

Photons are produced in type-II supernovae in shock waves that propagate in the envelope (during a time of $< 10^4$ s in SN 1987A and up to $\sim 10^7$ s in type-IIn supernovae). The shock wave generates in ordinary type-II supernovae not only short-time bursts of hard radiation but also an entropy reservoir that ensures radiance at the 'plateau' stage for several months. It is the source of radiance in type-IIn supernovae, where the shock wave propagates in the surrounding medium for several months [219–222].

The shapes and amplitudes of light curves of type-II supernovae are very diverse, as a result of which they cannot be described by introducing a standard candle, i.e., a kind of unified light curve. The light curve heavily depends on the properties of the envelope that surrounds the supernova energy source, whether it is a collapsing star core or thermonuclear fusion in the core. Type-II supernovae are at the same time much less dependent on details of the burst owing to the envelope. A real photosphere is observed over several months in the envelope, which is manifested in the light curves as a classical plateau.

The idea of the expanding photosphere method (EPM) originates from Baade [223] and Wesselink [224], who developed it to measure distances to variable stars, cepheids.

Because a detailed model of type-IIn supernovae can be developed, a new direct method DSM can be created on the basis of the same idea that enables using the bright light of type-II supernovae for cosmology.

The concept of the method is as follows. If the photosphere velocity v is known, its radius changes during the time of measurement $\mathrm{d}t$ by $\mathrm{d}r = v\,\mathrm{d}t$, and the change in radius $\mathrm{d}r$ can be immediately determined without using any cosmic distance ladders. The measured radiation flux is

$$F = \frac{4\pi r^2 \sigma T^4}{D^2}\,, \tag{5.27}$$

where D is the photometric distance. The temperature T is measurable, as are $\mathrm{d}r$ and $\mathrm{d}F$, while D does not change. It is convenient to define $S = \sqrt{F}$:

$$S = \frac{2\sqrt{\pi\sigma} r T^2}{D}\,. \tag{5.28}$$

If T does not significantly change between two measurements, we arrive at

$$\mathrm{d}S = \frac{2\sqrt{\pi\sigma}\,\mathrm{d}r T^2}{D}\,. \tag{5.29}$$

The value $\mathrm{d}r$ can be directly measured in a number of cases. Namely, $\mathrm{d}r = v_{\mathrm{ph}}\,\mathrm{d}t$ if v_{ph} is the photosphere velocity. We then directly determine the distance D using measured $\mathrm{d}S$, $\mathrm{d}r$, and T.

The authors of [215] clearly understood that spectral lines show the velocity u of matter, and the photosphere itself moves with respect to matter (because the matter absorption coefficient diminishes in the process of expansion). Even the signs of u and v_{ph} can be opposite if the photosphere is shrinking. This is the main difficulty for EPM and SEAM: for these methods to be operative, it is necessary to assume that free expansion occurs, and the velocity of matter is $u = r/t$. This situation occurs if dense matter is absent for some time around the star. The case of type-IIn supernovae is quite the opposite: there is a large amount of matter around the star, and the shock wave cannot burst into the low-density environment for months or even years.

By contrast, as plots reported in [221] and [222] show, all matter behind the shock wave front is compressed in those supernovae into a cold dense shell. The photosphere is glued to that dense shell and exactly $u = v_{\mathrm{ph}}$, and this value can be measured. The described picture corresponds to Baade's idea [223] proposed as early as the 1920s.

Summarizing, we can formulate a new method, DSM, to determine cosmological distances using type-IIn supernovae. This method consists of the following stages:

- Narrow components of spectral lines is measured to assess the properties (densities and velocities) of the near-star envelope. This stage requires neither high accuracy of measurements nor simulation.
- Broad emission components of lines is measured and the velocity on the photosphere level is determined (with the highest accuracy possible).

• Although the law $u = r/t$ is not applicable to these supernovae, $v_{\rm ph}$ now corresponds to the true velocity of the photosphere radius (and not only to the matter motion velocity as in SN II-P).

• Baade's initial idea [223] to measure the increase in radius $dr = v_{\rm ph}\,dt$ by integrating over time can now be applied (definitely with necessary consideration for scattering, limb darkening/brightening, etc.). The obtained changes in the radius must be used in iterating the optimal model.

• The observed flux can now be used to determine the distance D as

$$D = \frac{2\sqrt{\pi\sigma}\,dr T^2}{dS}\,. \tag{5.30}$$

If T changes significantly, this simple approach fails, and a model should be developed that best reproduces the data of broadband photometry and the velocity $v_{\rm ph}$, which is controlled by observations of $dr(t)$. A model like this is needed to calculate the evolution of r and the correction factor ζ and, in reality, to make detailed predictions for the theoretical flux F_ν.

We now represent the main algorithm of that method on the basis of a black-body model with color temperature $T_{\rm c}$ with the correction factor ζ. We assume that observations are performed frequently enough to measure changes in the radius $dr = v_{\rm ph}\,dt$ for a set of points. Let the initial (and unknown) radius be R_0 and $R_i = R_0 + \Delta R_i$ for $i = 1, 2, 3, \ldots,$ where ΔR_i can be considered to be known from summing the measured dr. We then have

$$\zeta_i^2 (R_0 + \Delta R_i)^2 \pi B_\nu(T_{{\rm c}i}) = 10^{-0.4A_\nu} D^2 f_{\nu i}\,, \tag{5.31}$$

or, after taking the square root,

$$\zeta_i (R_0 + \Delta R_i)\sqrt{\pi B_\nu(T_{{\rm c}i})} = 10^{-0.2A_\nu} D\sqrt{f_{\nu i}}\,. \tag{5.32}$$

A quality model yields a set of ζ_i and $T_{{\rm c}i}$ for all observation points, and we can then use the measured $f_{\nu i}$ and ΔR_i to find the unknown R_0 and combination $10^{-0.4A_\nu}D^2$ by the least square method. To obtain the distance D, we need to either know the absorption A_ν from astronomical observations or obtain it from the same equations modified for various spectral filters. In this case, already knowing R_0, we arrive at a set of equations of the form

$$10^{-0.4A_\nu} D^2 = a_s\,, \tag{5.33}$$

where the subscript s refers to one of the UBVFRI filters, and a_s is a constant that depends on the selection of the filter. As a result, we obtain the differences $A_{s1} - A_{s2}$ and, knowing how extinction depends on frequency, find A_ν.

The DSM method described here was used in [211–213] to obtain distances to type-IIn supernovae SN 2006gy, SN 2009ip, and SN 2010jl. A comparison with the distances to these objects measured using alternative (indirect) methods shows the operability of the new method, enabling systematic errors introduced by the cosmic distance ladder to be ruled out.

5.10 Merging of neutron stars and the standard siren method

A breakthrough into the era of gravitation wave astronomy occurred in 2015: it was announced for the first time that the merging of a pair of massive black holes had been observed. Before, humans had only observed the Universe in electromagnetic and neutrino channels; now, we acquired a new 'vision': an absolutely new gravitational-wave channel for studying objects in the Universe, exploring their properties, and testing theoretical predictions. The first gravitational-wave signal was received on September 14, 2015 by two antennas of the LIGO collaboration [225]. An analysis showed that the signal originated from two massive black holes. It was followed by more discoveries of signals from merging black holes until August 14, 2017 (in which the upgraded VIRGO detector participated).

Three days later, on August 17, 2017, another important discovery was made: a gravitational-wave signal, GW170817, was detected [226] whose duration and shape evidenced the merging of not two black holes but two neutron stars in a close binary system. The time delay of the signal in the LIGO antennas with VIRGO data taken into account was used to localize the source within approximately 30 degrees squared on the celestial sphere. The Fermi [227] and Integral [228] space observatories recorded a weak short gamma-ray burst GRB170817A in the square spot where the gravitational-wave source was localized 1.7 s after the LIGO signal was lost.

The gravitational wave amplitude and the pattern of its frequency variation enable estimating the distance to the source. It proved to be approximately 40 Mpc with an error of several dozen percent; however, despite the large error, only few galaxies corresponded to a distance this large in the localization area on the celestial sphere. Ground telescopes (including the MASTER robot [229]) therefore rather rapidly identified the outburst of a weak supernova (the so-called 'kilonova') in the NGC 4993 galaxy located at a distance of 40 Mpc from Earth.

It is important to stress in relation to this outstanding discovery in world science that the phenomenon was first predicted in the 1980s in the work of one of the authors of this review with his collaborators [230]. It was predicted for the first time that when a pair of neutron stars merges, heavy elements should be ejected and an electromagnetic outburst occur, including a gamma-ray burst rather than radiation from a gravitational wave alone. Models had been proposed prior to that study that described the merging of a neutron star and a black hole but with an inefficient gamma-ray burst within our Galaxy [231], while Blinnikov et al. [230] explicitly predicted distances of about tens of megaparsecs and outburst energy of the order of supernova outbursts. That study was preceded by a good study [232] that contained a range of predictions regarding gravitational-wave and neutrino signals that originate from the merging of two neutron stars in a binary; however, nothing was said about the electromagnetic signal, which is a million times weaker than those two signals but can be detected much more easily. The scenario in [230] was later corroborated by a quantitative calculation [233] that yielded all the main characteristics of the weak gamma-ray burst GRB170817A (total energy and hardness of photons in the gamma-ray range and typical velocities of heavy-element ejection.)

The photometric distance to the gravitational wave source measured using the gravitational wave signal alone, the so-called standard siren method, makes it possible to measure the Hubble constant H_0 if the host galaxy redshift is known. Schutz [234] was the first to indicate this immediately after publication of [230], when it became clear that a strong gravitational-wave signal must be accompanied by a weak electromagnetic signal. The electromagnetic signal can be associated with a specific galaxy much more reliably than using gravitational waves, and this potential enhances the accuracy with which the Hubble parameter can be measured. The distance estimated on the basis of the gravitational wave signal inevitably contains errors related primarily to the uncertainty in the orbit plane tilt angle (the distance $43.8^{+2.9}_{-6.9}$ Mpc was obtained already taking kilonova's localization into account [235]). The Hubble parameter was determined using data for the gravitational wave from GW170817: $H_0 = 70.0^{+12.0}_{-8.0}$ km s^{-1} Mpc^{-1} (at a 68.3% confidence level [235]). This estimate is obviously compatible with the supernova measurements performed by the Riess group, 73.45 ± 1.66 km s^{-1} Mpc^{-1} [236, 237], and data on the CMB obtained by the Planck collaboration, $H_0 = 67.8 \pm 0.9$ [238] and a more recent value, $H_0 = 67.4 \pm 0.5$ [239]. The estimates of confidence intervals show that the H_0 values obtained by these two groups significantly disagree. The accuracy in measuring H_0 using this method can be improved if a large number of merging neutron star binaries is discovered. We note that currently our DSM method [211–213] described above is, owing to its accuracy, quite competitive with the standard siren method, and it can be developed at a lower cost than the latter method, because even most powerful telescopes are less expensive than facilities that detect gravitational waves.

6. Dark energy

Dark energy is the general name for the unknown substance that possibly causes cosmological acceleration. Any form of matter can be dark energy as long as it is described by the equation of state $P = w\rho$ (2.9) with $w < -1/3$ [see Eqn (2.6)]. It should be kept in mind that the parameter w is very close to -1 according to observations. A special particular case $w = -1$ corresponds to the vacuum energy described above and differs conceptually from any other case.

As was correctly noted in [11], due to the presence of vacuum energy in this equation of state, the distance between two hydrogen atoms in the absolute vacuum of our modern Universe, sufficient for their gravitational attraction to be compensated by gravitational repulsion of vacuum energy, is only one meter. Strictly speaking, repulsion at this distance would be unnoticeable on the background of the induced electric dipole interaction between atoms. This thought experiment is valid for ideally neutral particles with a mass of 1 GeV. This conclusion is valid in the real world for two galaxies with the mass of the Milky Way located at a distance of a couple of megaparsecs.

An alternative mechanism of accelerated cosmological expansion might be offered by gravity modified at large (cosmological) distances, considered in Section 7.

The existence of dark energy was predicted in [64] prior to the discovery of the accelerated expansion of the Universe. A conclusion was made in that study and in several subsequent ones [65, 66] that the compensation mechanism must have an uncompensated residue with the energy density close to the cosmological one and a nonstandard equation of state. It was noted, in particular, in [66] that an equation of state with $w < -1$ can emerge, which would result during a finite time in a cosmological singularity, which was later called a phantom [240].

It was noted in other earlier papers [91–94] that in some potentials the scalar field can mimic time-dependent vacuum-like energy. The description of dark energy significantly advanced when potentials were discovered that yield so-called tracking solutions, first found in [241, 242] and somewhat later in [243, 244]. These tracking solutions are of interest because the scalar field potential can be selected in such a way that the energy density of the field rather naturally turns out to be close to the current energy density of ordinary matter, thus solving the problem of ρ_{m} being close to ρ^{vac} in the present-day Universe. This option has rapidly become very fashionable [245–263] (see also reviews [264–266]).

Possible forms of dark energy models based on a scalar field can be divided into several classes. In particular, this could be a scalar field with the standard kinetic term (see Eqn (9.24) in Section 9.3), but with the potential $U(\phi)$ of a very unusual form, which results, at the very least, in a nonrenormalizable and essentially nonlinear theory. Caldwell et al. [248] proposed calling that field the 'quintessence'.

Other versions include the so-called K-essence theory, where the kinetic term has a nonstandard form:

$$A_{\mathrm{K}} = \int \mathrm{d}^4 x \, \sqrt{-g}\, P(\phi, X)\,, \tag{6.1}$$

where $X = (1/2)\, g^{\mu\nu}\, \partial_\mu \phi\, \partial_\nu \phi$ is the usual kinetic term, but the function P depends on it nonlinearly. P is often chosen in a factored form: $P(\phi, X) = f_1(\phi) f_2(X)$. The freedom of choice here is very broad, owing to which studies where this issue is explored are quite numerous.

Another way to describe dark energy is based on a scalar tachyon field, an example of whose action is given by the formula

$$A_{\mathrm{T}} = \int \mathrm{d}^4 x \, V(\phi)\, \sqrt{-\det\left[g_{\mu\nu} - \partial_\mu \phi\, \partial_\nu \phi\right]}\,, \tag{6.2}$$

where $V(\phi)$ is a potential function chosen ad hoc. We stress the unusual dimension of the field ϕ, which is inversely proportional to the first power of energy. This action resembles the Born–Infeld action [267], where $\partial_\mu \phi\, \partial_\nu \phi$ is substituted with the Maxwell tensor $F_{\mu\nu}$, or modified gravity [268, 269], where $R_{\mu\nu}$ is added to the action in the radicand. All those theories feature a number of similar properties.

Finally, it is worth mentioning the 'phantom' field whose action sign is opposite to the normal one. This field results in $w < -1$, as follows from the formula for energy and pressure density (9.28) if the sign in front of the ϕ field derivatives is reversed. The incorrect sign of the kinetic term apparently results in instability of high-frequency modes, a serious problem for such theories, which has already emerged on the classical level, to say nothing about quantum theory problems.

A more detailed discussion of these options for the phenomenological (effective, in the spirit of effective Lagrangians) description of dark energy and a vast list of references can be found in reviews [264–266]. It looks as if everything that was considered pathological in normal theories is now used to 'produce' accelerated expansion. We cannot, however, rule out that a solution to the problem may be found just on this path.

A straightforward candidate for the role of dark energy carrier is a scalar field ϕ with the canonical kinetic term and very small (or even zero) mass, $m < H_0$, where H_0 is the Hubble parameter in the modern epoch. This field satisfies the

Klein–Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a^2(t)} + U'\phi = 0\,, \tag{6.3}$$

where $U(\phi)$ is the potential of that field and $U' = \mathrm{d}U/\mathrm{d}\phi$.

If the Hubble parameter is large (specific conditions for the required value are presented below), the solution of the above equation is approximately constant. First, the spatial nonuniformity of ϕ is smoothed out due to the cosmological expansion caused by the factor $1/a^2$ in front of spatial derivatives. The second time derivative is at the same time negligibly small due to large Hubble friction. We note that Eqn (6.3) for a spatially uniform field $\phi = \phi(t)$ coincides with the equation of motion in Newton mechanics, where the role of coordinate is played by $\phi(t)$. Its solution can therefore be intuitively designed knowing the shape of the potential $U(\phi)$. The term $3H\dot{\phi}$ behaves as fluid friction in mechanics, and, if the friction is strong, the motion resembles floating in glycerin, where the velocity tends to a constant value under the effect of a constant force.

We thus assume that the second derivatives in Eqn (6.3) can be disregarded, find a solution, and check whether it corresponds to the initial assumptions regarding slow variation of ϕ. The field ϕ under the assumptions made satisfies the first-order equation

$$\dot{\phi} = -\frac{U'}{3H}\,. \tag{6.4}$$

This is nothing but the slow-roll approximation widely used in inflationary models.

If the main contribution to the cosmological energy density ρ comes from ϕ, then ρ is determined in the slow field variation limit by the potential $U(\phi)$ and, according to Eqn (2.5), the Hubble parameter is

$$H^2 = \frac{8\pi U}{3m_{\mathrm{Pl}}^2}\,. \tag{6.5}$$

The slow-roll approximation is only valid if $\ddot{\phi} \ll 3H\dot{\phi}$ and $\dot{\phi}^2 \ll 2U(\phi)$. For these conditions to hold, the following inequality must be satisfied:

$$\frac{U''}{U} \ll \frac{8\pi}{3m_{\mathrm{Pl}}^2}\,, \tag{6.6}$$

which requires a very large ϕ field. Notably, the conditions specified above are satisfied for a massive noninteracting field, i.e., a field with the harmonic potential $U = m^2\phi^2/2$, if

$$\phi^2 > \frac{4\pi}{3}\,m_{\mathrm{Pl}}^2\,. \tag{6.7}$$

If it is required that the energy density of the ϕ field, namely $\rho_\phi = m_\phi^2\phi^2$, be of the order of the modern cosmological energy density, its mass must be constrained by a very low value: $m_\phi < 1/t_{\mathrm{u}} \approx 10^{-42}$ eV.

Some information about the scalar field theory is presented in Section 9.3, which also contains a derivation of the scalar field energy–momentum tensor, which has form (9.27). In the case of a quasi-constant and quasi-uniform field, it becomes proportional to the metric tensor, $T_{\mu\nu} \propto g_{\mu\nu}$, thus implementing the vacuum-like equation of state (2.15), i.e., $w \approx -1$, which eventually results in an almost exponentially accelerated expansion. It is essential, however, that the field ϕ decreases, albeit slowly. It is assumed that at the equilibrium point where $\mathrm{d}U/\mathrm{d}\phi = 0$, the potential also vanishes, $U = 0$. (This requires that the vacuum energy be absent, and the fulfillment of this requirement is, generally speaking, optional.) A simple example of such potentials tending to zero at infinity are the power-law $U \sim 1/\phi^q$ or exponential $U \sim \exp(-\phi/\mu)$ potentials, where μ is a constant parameter with the dimension of mass [270]. These potentials were introduced for the purpose of a phenomenological description of accelerated expansion; however, substantiation of such potentials is rather weak.

The motion of $\phi(t)$ in such potentials is conceptually different from that in potentials that have a minimum at a finite ϕ, for example, $U(\phi) = m_\phi^2\phi^2/2$ or $U(\phi) = \lambda\phi^4/4$, and zero vacuum energy. We note that these two potentials are natural in quantum field theory because they correspond to renormalizable theories. The expansion regime changes drastically in such potentials if ϕ decreases to a value at which H^2 becomes comparable to m_ϕ^2 or $\lambda\phi^2$, i.e., ϕ decreases to a level significantly lower than the Planck value. The quasipotential regime of accelerated expansion is replaced at that moment with the ordinary decelerating Friedmann regime: either nonrelativistic ($w = 0$) for the quadratic potential or relativistic ($w = 1/3$) for the quartic potential. The field ϕ then starts oscillating about the minimum, generating massless particles, photons or gravitons.

If the potential does not have a minimum at finite ϕ, the field monotonically tends to zero, and the expansion is eternally accelerated; the parameter w in the slow-roll regime is constant and negative, thus ensuring accelerated expansion. For example, we consider the exponential potential [270, 271]

$$U(\phi) = U_0 \exp\left(-\frac{\phi}{\mu}\right). \tag{6.8}$$

The equation of motion for ϕ reduces under the assumption $\dot{\phi}^2 \ll U(\phi)$ to Eqn (6.4), which can be easily integrated if the potential term dominates in the energy density and the Hubble parameter is given by Eqn (6.5). The field ϕ logarithmically grows with time:

$$\phi(t) = 2\mu\,\ln\left[\sqrt{\frac{U_0}{96\pi\mu^4}}\,m_{\mathrm{Pl}}(t - t_0) + \exp\left(\frac{\phi_0}{2\mu}\right)\right], \tag{6.9}$$

where ϕ_0 is the value of the field at the initial moment t_0. For large t, we then have

$$U(\phi) \approx \frac{96\pi\mu^4}{(m_{\mathrm{Pl}}t)^2}\,, \quad (\dot{\phi})^2 \approx \frac{4\mu^2}{t^2}\,, \tag{6.10}$$

and the slow roll-condition is satisfied for $m_{\mathrm{Pl}} \lesssim \mu$. The Hubble parameter then decreases in inverse proportion to time:

$$H \approx \frac{16\pi\mu}{m_{\mathrm{Pl}}t}\,, \tag{6.11}$$

and hence the expansion is described by a power law:

$$a \sim t^{16\pi\mu/m_{\mathrm{Pl}}}\,. \tag{6.12}$$

If $16\pi\mu > m_{\mathrm{Pl}}$, expansion occurs with acceleration. If $m_{\mathrm{Pl}}/\mu \to 0$, the expansion law tends to be exponential, and $w \to -1$.

7. Modified gravity

An alternative hypothesis suggested for a phenomenological description of cosmological acceleration is the assumption that gravity is modified at small curvature. This approach is usually realized by adding some function of the curvature scalar R to the Hilbert–Einstein action:

$$A = \frac{m_{\rm Pl}^2}{16\pi}\int d^4x\sqrt{-g}\,[R + F(R)] + A_{\rm m}\,, \tag{7.1}$$

where $m_{\rm Pl} = 1.22\times 10^{19}$ GeV is the Planck mass, R is the scalar curvature, and $A_{\rm m}$ is the matter field action. The additional term $F(R)$ modifies gravity at large distances; therefore, this approach is referred to as an infrared modification of gravity. The nonlinear function $F(R)$ is chosen in such a way that the equation of motion has the solution $R = \text{const}$ in the absence of matter. We note that, in principle, more intricate versions of modifying gravity might be considered that contain functions not only of R but also of invariant combinations built from $R_{\mu\nu}R^{\mu\nu}$ or the Riemann tensor. However, theories with $F(R)$ alone are currently under consideration. This is dictated not only by their simplicity but also by problems with ghosts and tachyons that emerge in more sophisticated theories.

The Einstein equations for $F(R)$ gravity are modified to

$$(1 + F')\,R_{\mu\nu} - \frac{1}{2}(R + F)\,g_{\mu\nu} + (g_{\mu\nu}D_\alpha D^\alpha - D_\mu D_\nu)\,F' = \frac{8\pi T_{\mu\nu}^{(\rm m)}}{m_{\rm Pl}^2}\,, \tag{7.2}$$

where $F' = dF/dR$, D_μ is the covariant derivative in a FRW metric and $T_{\mu\nu}^{(\rm m)}$ is the energy–momentum tensor of matter. Taking the trace of this equation, $g_{\mu\nu}\delta A/\delta g_{\mu\nu}$, we obtain

$$3D^2F'_R - R + RF'_R - 2F = \frac{8\pi T_\mu^\mu}{m_{\rm Pl}^2}\,, \tag{7.3}$$

where $D^2 \equiv D_\mu D^\mu$ is the covariant d'Alembert operator. Exploration of the solutions of this equation is frequently sufficient for analyzing the $F(R)$ theory.

We note, however, that adding a term to an action that is nonlinear in curvature can result in serious pathologies in the theory, including violation of unitarity and the emergence of ghosts and/or tachyons. These problems are usually ignored for curvatures close to the Planck value (who knows what happens there), but noticeable deviations from ordinary gravity at small curvatures, i.e., in the weak-field limit, may result in disagreement with well-established observational facts. We do not consider the entire range of these problems, but limit our analysis to possible deviations from the standard GR at the level of solutions of the classical equations of motion.

In the pioneering studies of cosmological acceleration originating from modified gravity [272, 273], the function $F(R)$ was chosen in the form $F(R) = -\mu^4/R$, where μ is a small parameter with the dimension of mass. However, as was shown in [274], this choice of $F(R)$ results in very strong exponential instability of the theory in the presence of matter, such that the ordinary gravity theory would be heavily distorted. Indeed, we consider the equation that describes the evolution of curvature as a function of time (7.3) in the theory containing $F(R) = -\mu^4/R$:

$$D^2R - 3\,\frac{(D_\alpha R)\,(D^\alpha R)}{R} = \frac{R^2}{2} - \frac{R^4}{6\mu^4} - \frac{\tilde T R^3}{6\mu^4}\,. \tag{7.4}$$

Here, $\tilde T = 8\pi T_\nu^\nu/m_{\rm Pl}^2 > 0$. This equation has an obvious solution $R^2 = 3\mu^4$ in the absence of matter, which describes the de Sitter universe with a constant curvature scalar.

We now assume that an ordinary celestial body, for example Earth or the Sun, is the source of the gravitational field, and hence the created gravitational field is weak and the background space is 4D flat and has the Minkowski metric. We seek a solution of Eqn (7.4) in the perturbation theory, assuming a small deviation from standard GR. The curvature is algebraically expressed in the lowest order in terms of the trace of the energy–momentum tensor: $R_0 = -\tilde T$. The standard solution in the vacuum $R = 0$ is now an approximate one, because the function $F(R)$ is chosen, as was noted above, such that Eqn (7.3) has the solution $R = R_{\rm c} = \text{const}$ in the absence of matter. Here, $R_{\rm c}$ is equal to the observed cosmological constant and is negligibly small compared with the curvature inside any material body: it is as small as the ratio of the cosmological energy density and the energy density of that body. It can also be verified that the stationary solution outside a gravitating body rapidly decreases in the modified theory as the distance from that body increases. This observation enables making a conclusion that stationary solutions in modified gravity agree well with the Newtonian limit of the standard GR for a sufficiently small μ.

Introducing an additional term μ^4/R with a sufficiently small μ to the action seems at first glance to not result in significant deviations from the standard gravity theory; however, this is not fully true: in the modified theory, the scalar R becomes a dynamic variable whose evolution is driven by a second-order equation with a small coefficient in front of the higher time derivative. This results in the emergence of very strong instability of the solution in the presence of material bodies [274].

We now use Eqn (7.4) in a perturbative calculation of the gravitational field or, more accurately, of the scalar curvature within a celestial body, for example, the Sun, Earth, or a gas cloud in a galaxy whose energy density is time dependent. We seek a solution in the form $R = R_0 + R_1$, where $R_0 = -\tilde T$ is the standard GR solution. Assuming, as was noted above, that the background metric is flat, we find that the deviation from GR, R_1, satisfies the equation

$$\ddot R_1 - \Delta R_1 - \frac{6\dot{\tilde T}}{\tilde T}\,\dot R_1 + \frac{6\partial_j\tilde T}{\tilde T}\,\partial_j R_1 + R_1\left[\tilde T + 3\,\frac{(\partial_\alpha\tilde T)^2}{\tilde T^2} - \frac{\tilde T^3}{6\mu^4}\right] = \Delta\tilde T + \frac{\tilde T^2}{2} - \frac{3(\partial_\alpha\tilde T)^2}{\tilde T}\,, \tag{7.5}$$

where $(\partial_\alpha\tilde T)^2 = \dot{\tilde T}^2 - (\partial_j\tilde T)^2$.

The last term in square brackets in the left-hand side of Eqn (7.5) results in exponential instability of small fluctuations and an instability of the gravitational field produced by the mass/energy density of the body under consideration, which changes with time on a regular basis. The characteristic instability time turns out to be very short:

$$\tau_{\rm instab} = \frac{\sqrt6\mu^2}{T^{3/2}} \sim 10^{-26}\ {\rm s}\left(\frac{\rho_{\rm m}}{{\rm g\ cm^{-3}}}\right)^{-3/2}, \tag{7.6}$$

where $\rho_{\rm m}$ is the mass/energy density of the body, $\mu^{-1} \sim t_{\rm u} \approx 3 \times 10^{17}$ s, and $t_{\rm u}$ is the age of the Universe. Because μ is small, the term that contains μ^4 in the denominator in Eqn (7.5) is much larger than all other terms.

Spatial derivatives usually eliminate or suppress instability. For example, the Jeans instability is removed by the pressure that counteracts gravity. However, the term containing the Laplace operator, ΔR_1, which is inversely proportional to the system dimension squared, is in the case under consideration much less than the term that causes the instability. It can be easily verified that the instability can be suppressed on a scale that is smaller than the Compton wavelength of the proton, i.e., smaller than 10^{-14} cm.

Further modification of modified gravity has been proposed to solve the instability problem. We only consider the class of models proposed in [275–277]. Some other forms of the modification of gravity are reviewed in [278]. Various forms of the action studied in [275–277] are

$$F_{\rm HS}(R) = -\frac{R_{\rm vac}}{2}\,\frac{c(R/R_{\rm vac})^{2n}}{1 + c(R/R_{\rm vac})^{2n}}\,, \tag{7.7}$$

$$F_{\rm AB}(R) = \frac{\epsilon}{2}\,\lg\left[\frac{\cosh\,(R/\epsilon - b)}{\cosh b}\right] - \frac{R}{2}\,, \tag{7.8}$$

$$F_{\rm S}(R) = \lambda R_0\left[\left(1 + \frac{R^2}{R_0^2}\right)^{-n} - 1\right]. \tag{7.9}$$

Although these functions seem to look different, their implications are virtually identical. Below, we follow the analysis performed in [279] and use formula (7.9) in specific examples.

Introducing the notation $f = R + F(R)$, we represent the field equations in the form

$$f' R_\mu^{\nu} - \frac{f}{2}\,\delta_\mu^{\nu} + (\delta_\mu^{\nu}\Box - D_\mu D^{\nu})\, f' = \frac{8\pi\, T_\mu^{\nu}}{m_{\rm Pl}^2}\,. \tag{7.10}$$

Here and below, the prime on f or F denotes the derivative with respect to R.

Taking the trace over μ and ν, we arrive at a closed equation for R:

$$3\Box f'(R) + Rf'(R) - 2f(R) = 8\pi m_{\rm Pl}^{-2} T_\mu^{\mu}\,. \tag{7.11}$$

The theory modified in this way yields accelerated cosmological expansion if the equation

$$Rf'(R) - 2f(R) = 0 \tag{7.12}$$

has the solution $R = R_1 > 0$ with an (approximately) constant R_1.

To avoid possible pathologies in the theory, the following conditions must be satisfied:

(1) stability of cosmological solutions in the future:

$$\frac{F'(R_1)}{F''(R_1)} > R_1\,; \tag{7.13}$$

(2) classical and quantum stability (gravitational attraction and the absence of ghosts):

$$F'(R) > 0\,; \tag{7.14}$$

(3) the absence of instability in material bodies shown above:

$$F''(R) > 0\,. \tag{7.15}$$

These twice modified theories still result in significant problems, despite significant improvements. First, in the cosmological situation where the energy density decreases with time, a singularity had to exist in the near past when curvature was infinitely large, although energy density remained finite.

Moreover, astronomical systems with a growing energy density either have already ended in a singular state or will come to that state in the near future [280, 281]. Following [281], we consider version (7.9) of modified gravity in the limit $R \gg R_0$, where the following approximation can be used:

$$F(R) \approx -\lambda R_0\left[1 - \left(\frac{R_0}{R}\right)^{2n}\right]. \tag{7.16}$$

We analyze the evolution of R within massive astronomic objects with the mass/energy density $\rho \gg \rho_{\rm cosm}$ growing with time.

We conjecture as above that the gravitational field of the objects under study is weak, and therefore ordinary derivatives can be used instead of covariant ones. In this approximation, Eqn (7.11) has the form

$$\begin{aligned}(\partial_t^2 - \Delta)R &- (2n+2)\,\frac{\dot R^2 - (\nabla R)^2}{R} \\ &+ \frac{R^2}{3n(2n+1)}\left[\frac{R^{2n}}{R_0^{2n}} - (n+1)\right] \\ &- \frac{R^{2n+2}}{6n(2n+1)\,\lambda R_0^{2n+1}}\,(R + \tilde T) = 0\,.\end{aligned} \tag{7.17}$$

The equation is drastically simplified if R is replaced with another function $w \equiv F' = -2n\lambda(R_0/R)^{2n+1}$, which satisfies the equation

$$(\partial_t^2 - \Delta)w + \frac{{\rm d}U(w)}{{\rm d}w} = 0\,, \tag{7.18}$$

where the potential $U(w)$ is

$$\begin{aligned}U(w) &= \frac{1}{3}(\tilde T - 2\lambda R_0)\, w \\ &+ \frac{R_0}{3}\left[\frac{q^{\nu}}{2n\nu}\,w^{2n\nu} + \left(q^{\nu} + \frac{2\lambda}{q^{2n\nu}}\right)\frac{w^{1+2n\nu}}{1+2n\nu}\right]\end{aligned} \tag{7.19}$$

with the parameters $\nu = 1/(2n+1)$ and $q = 2n\lambda$. We here follow the notation introduced in [281] and hope that the use of the same notation w for both the new curvature function and the cosmological parameter that relates pressure and energy density does not lead to confusion.

We note that the singularity $R = \infty$ corresponds to $w = 0$. Therefore, in analyzing the singularity, we can safely disregard spatial derivatives in that equation. Indeed, spatial derivatives usually do not allow the function to grow; on the contrary, they 'drag' it to zero. If the Fourier transformation by coordinates is done, then $-\Delta w = k^2 w$, which means that this component behaves like the harmonic oscillator potential with a minimum at zero. Therefore, if a singularity occurs when the term Δw is disregarded, it certainly does so if this term is taken into account.

If only the leading terms in the small ratio R_0/R are retained and spatial derivatives are disregarded, Eqn (7.18) takes the form of the ordinary oscillator equation

$$\ddot{w} + \frac{\tilde{T}}{3} - \frac{q^{\nu}(-R_0)}{3w^{\nu}} = 0 \qquad (7.20)$$

with the potential that has the following form in this approximation:

$$U(w) = \frac{\tilde{T}w}{3} - \frac{q^{\nu}(-R_0)w^{1-\nu}}{3(1-\nu)} . \qquad (7.21)$$

The potential U is time dependent if the mass/energy density of the celestial body under consideration varies with time. As an example, we consider an object that consists of nonrelativistic matter whose mass density is growing linearly:

$$\tilde{T} \equiv \tilde{T}(t) = \tilde{T}_0\left(1 + \frac{t}{t_{\text{contr}}}\right), \qquad (7.22)$$

where t_{contr} is the characteristic time of system contraction. The linear approximation apparently loses its validity at some moment, but the results do not change in a qualitative way.

It is convenient in what follows to introduce dimensionless variables $t = \gamma\tau$ and $w = \beta\zeta$ such that

$$\gamma^2 = \frac{3q}{-R_0}\left(-\frac{R_0}{\tilde{T}_0}\right)^{2(n+1)},$$
$$\beta = \frac{\gamma^2\tilde{T}_0}{3} = q\left(-\frac{R_0}{\tilde{T}_0}\right)^{2n+1}. \qquad (7.23)$$

Equation (7.20) expressed in terms of these variables takes a very simple form:

$$z'' - z^{-\nu} + (1 + \kappa\tau) = 0 . \qquad (7.24)$$

The prime here denotes a derivative with respect to the dimensionless time τ.

It can be easily seen that the minimum of the potential shifts to $\zeta = 0$, and the depth of the potential in its minimum decreases. It seems to be quite evident that even if ζ was constant in the initial state and corresponded to a minimum of the potential, the growing mass density causes oscillations of ζ around the sliding minimum, and z attains zero in the process of oscillation build-up. Computations performed in [281] yield exactly that scenario.

For qualitative analysis of oscillations, it is convenient to use the integral law of energy evolution, which for the general form of the oscillatory equation

$$\zeta'' + \frac{\partial U}{\partial \zeta} = 0 \qquad (7.25)$$

with a potential that can explicitly depend on time, has the form

$$\frac{(\zeta')^2}{2} + U(\zeta,\tau) - \int d\tau\, \frac{\partial U}{\partial \tau} = \text{const} . \qquad (7.26)$$

Relation (7.26) also enables a general analysis of how the singularity of R emerges, as has been done in [282].

Infinite R values can be avoided if a term proportional to R^2 is added to the action:

$$F(R) \to F(R) - \frac{R^2}{6m^2} . \qquad (7.27)$$

Cosmological models with the action quadratic in curvature were explored for the first time in [285–287]. Leading terms in the curvature can emerge as a result of radiative corrections to the ordinary Einstein–Hilbert action if the vacuum expectation value of the energy–momentum tensor in a curved space is considered. We note that the radiative corrections generate not only R^2 but also $R_{\mu\nu}R_{\mu\nu}$, which are not as innocuous as the former term because they result in pathology in the form of tachyons and ghosts.

An interesting feature of the action containing terms quadratic in curvature is the early inflation stage, as is the case in the Starobinsky model [286]. Inflation naturally ends in this model due to production of particles by the new scalar gravitational degree of freedom, the curvature scalar, which becomes a dynamic variable owing to R^2 corrections. Heating of the Universe, a herald of the end of inflation, which occurs due to gravitational production of particles in the R^2 theory, was explored earlier in [287–291] and in more recent studies [292, 293].

Particle production in the late Universe due to high-frequency oscillations of the curvature scalar in the theory with action (7.9) was analyzed in [294, 295]. A conclusion was reached that the produced particles can make a significant contribution to the flux of high-energy cosmic particles. This result was criticized in [296, 297], where the high efficiency of particle production was questioned. However, it was shown in [298] that this criticism is not justified.

If the equation of motion that governs the evolution of $R(t)$ contains an R^2 term, it cannot be explicitly reduced to an oscillatory equation like (7.25), and takes a more involved form [281]:

$$\left(1 - \frac{R^{2n+2}}{6\lambda n(2n+1)\,R_0^{2n+1}m^2}\right)\ddot{R} - (2n+2)\,\frac{\dot{R}^2}{R}$$
$$- \frac{R^{2n+2}(R+T)}{6\lambda n(2n+1)\,R_0^{2n+1}} = 0 . \qquad (7.28)$$

We can nevertheless see that due to the presence of the second term in the coefficient in front of $\ddot{R}$, the curvature cannot grow to infinity. It is of interest that a naive estimate of the R cut-off at a value at which the second term is of the order of unity turns out to be too low. The growth of R stops at a significantly larger value, as was shown in [294]. This study explored the evolution of $R(t)$ in detail in infrared-modified theories with an action like (7.7)–(7.9), with the term $R^2/6m^2$ added for astronomical systems with growing energy density. An analytic solution was obtained using Eqn (7.26), as was a numerical one, and they agree well with each other.

The numerical solution is unstable at high frequencies and 'blows up' at relatively short times. It is therefore not possible to reliably advance to the asymptotic regime. However, the calculations are quite reliable at sufficiently small frequencies and agree well with the analytic result. On the other hand, the accuracy of analytic results increases as the frequency grows. Owing to this, the entire essential frequency range can be spanned.

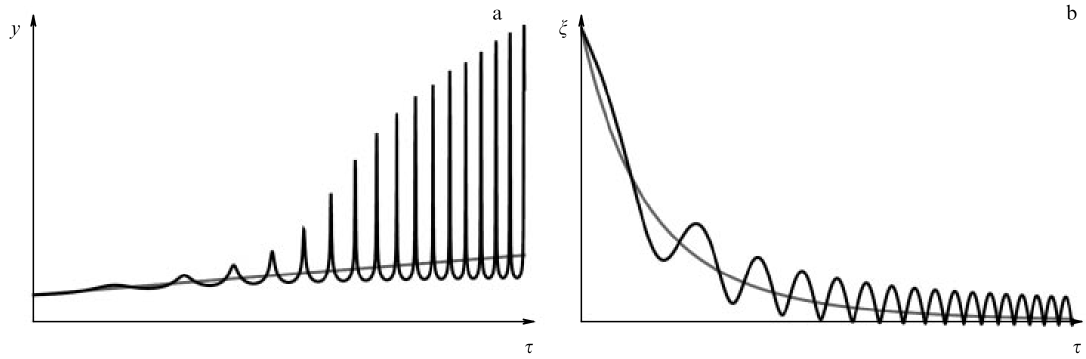

Figure 11. Curvature oscillation peaks. Results are presented for $n = 2$, $g = 0.001$, $\kappa = 0.04$, and $y_0' = \kappa/2$.

Equation of motion (7.28) was transformed for analysis into (7.25); but an explicit expression for the potential cannot be derived in that case. The obtained approximations nevertheless describe the solutions sufficiently well. The relations $|R_c| \ll |R| \ll m^2$ hold for the considered case of astronomical systems with growing density, and then F is approximately given by

$$F(R) \simeq -R_c\left[1 - \left(\frac{R_c}{R}\right)^{2n}\right] - \frac{R^2}{6m^2}\,. \tag{7.29}$$

We consider an approximation for the uniform distribution of matter with the growing mass density that remains sufficiently small (for example, a gas cloud in the process of production of a galaxy or star). The background metric can be considered in this case to be the Minkowski metric. In this approximation Eqn (7.3) takes the form

$$3\partial_t^2 F' - R - T = 0\,. \tag{7.30}$$

As above, we introduce dimensionless variables

$$\begin{aligned} &z \equiv \frac{T(t)}{T(t_{\rm in})} \equiv \frac{T}{T_0} = \frac{\rho_{\rm m}(t)}{\rho_{\rm m0}}\,, \quad y \equiv -\frac{R}{T_0}\,,\\ &g \equiv \frac{T_0^{2n+2}}{6n(-R_c)^{2n+1}m^2} = \frac{1}{6n(mt_{\rm u})^2}\left(\frac{\rho_{\rm m0}}{\rho_c}\right)^{2n+2},\\ &\tau \equiv m\sqrt{g}\,t\,, \end{aligned} \tag{7.31}$$

where $\rho_c \approx 10^{-29}$ g cm^{-3} is the cosmological energy density in the present-day Universe, $\rho_{\rm m0}$ is the initial value of energy density in the system under consideration, and $T_0 = 8\pi\rho_{\rm m0}/m_{\rm Pl}^2$.

We next introduce a new unknown function

$$\xi \equiv \frac{1}{2n}\left(\frac{T_0}{R_c}\right)^{2n+1} F_{,R} = \frac{1}{y^{2n+1}} - gy\,, \tag{7.32}$$

which can be used to rewrite Eqn (7.30) in the simple oscillatory form

$$\xi'' + z - y = 0\,, \tag{7.33}$$

where the prime on ξ denotes the derivative with respect to the dimensionless time τ. Change of variables (7.32) is similar but not identical to that done in [281] even if the R^2 term is absent. This is more convenient in the case under consideration for technical reasons.

The oscillator potential is evidently determined by the condition

$$\frac{\partial U}{\partial \xi} = z - y(\xi)\,. \tag{7.34}$$

The y variable is no longer expressed in terms of ξ in analytic form, and the explicit formula for the potential is unavailable. Rather accurate approximated formulas for $U(\xi)$ can nevertheless be derived, especially for small g, separately for positive and negative ξ. Details of this approach can be found in [294].

Oscillations of $\xi(t)$ are nearly harmonic, but the physical field $R(t)$ oscillates in a way that is very far from harmonic; it exhibits sharp and short peaks with large amplitude where $y \gg 1$, i.e., the solution strongly deviates from the standard GR form $R_{\rm GR} = -T_0$. These peaks correspond to unrealized singular points where curvature could become infinite but was cut off by the R^2 term. The behavior of curvature as a function of time is displayed in Fig. 11. As a result of the strong anharmonicity of oscillations of y or, equivalently, of R, the energy of low-frequency modes excited when the system slowly contracts is pumped to high-frequency modes.

The high-frequency large-amplitude oscillations of curvature result in the production of high-energy cosmic rays in the period when large-scale structures are formed in the Universe. Fluxes of such particles can be observable for a rather broad range of theory parameters.

We discussed this unusual behavior of curvature in detail also because these solutions violate the Jebsen–Birkhoff theorem; in particular, an amazing phenomenon of gravitational repulsion, antigravitation, occurs in finite-size systems [299]. The occurrence of gravitational repulsion in infinite systems, for example, accelerated cosmological expansion, does not disagree with GR. A discussion of the violation of the Jebsen–Birkhoff theorem in $F(R)$-modified theories can be found in [300, 301].

We now show how gravitational repulsion emerges using an example of a spherically symmetric distribution of matter with the metric given by the standard formula

$$ds^2 = A(r,t)\,dt^2 - B(r,t)\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2)\,. \tag{7.35}$$

A metric of this type for the $F(R)$ theories has been analyzed in [302, 303]; however, these studies did not discuss curvature oscillations for which the antigravitation effect originates.

We assume that the metric coefficients A and B are close to unity, i.e., that the metric is close to the flat Minkowski metric. It is under this assumption that the quantitative results about R oscillations have been obtained. Nonzero components of the Ricci tensor corresponding to metric (7.35) are

$$R_{00} = \frac{A'' - \ddot{B}}{2B} + \frac{(\dot{B})^2 - A'B'}{4B^2} + \frac{\dot{A}\dot{B} - (A')^2}{4AB} + \frac{A'}{rB}, \quad (7.36)$$

$$R_{rr} = \frac{\ddot{B} - A''}{2A} + \frac{(A')^2 - \dot{A}\dot{B}}{4A^2} + \frac{A'B' - (\dot{B})^2}{4AB} + \frac{B'}{rB}, \quad (7.37)$$

$$R_{\theta\theta} = -\frac{1}{B} + \frac{rB'}{2B^2} - \frac{rA'}{2AB} + 1, \quad (7.38)$$

$$R_{\varphi\varphi} = \left(-\frac{1}{B} + \frac{rB'}{2B^2} - \frac{rA'}{2AB} + 1\right)\sin^2\theta = R_{\theta\theta}\sin^2\theta, \quad (7.39)$$

$$R_{0r} = \frac{\dot{B}}{rB}. \quad (7.40)$$

The prime and dot on symbols denote derivatives with respect to r and t. The corresponding curvature scalar $R = g^{\mu\nu}R_{\mu\nu}$ is

$$\begin{aligned} R &= \frac{1}{A}R_{00} - \frac{1}{B}R_{rr} - \frac{1}{r^2}R_{\theta\theta} - \frac{1}{r^2\sin^2\theta}R_{\varphi\varphi} \\ &= \frac{A'' - \ddot{B}}{AB} + \frac{(\dot{B})^2 - A'B'}{2AB^2} + \frac{\dot{A}\dot{B} - (A')^2}{2A^2B} + \frac{2A'}{rAB} - \frac{2B'}{rB^2} \\ &+ \frac{2}{r^2B} - \frac{2}{r^2} = \frac{2}{A}R_{00} - \frac{2B'}{rB^2} + \frac{2}{r^2B} - \frac{2}{r^2}. \end{aligned} \quad (7.41)$$

Assuming, as was said above, that the metric insignificantly differs from the flat one, i.e.,

$$A_1 = A - 1 \ll 1, \quad B_1 = B - 1 \ll 1, \quad (7.42)$$

we check the self-consistency of that assumption for the oscillatory solutions found in [294], where R significantly exceeds its GR value. For this, it is convenient to use Eqn (7.2) in the form

$$R_{00} - \frac{R}{2} = \frac{\tilde{T}_{00} + \Delta F_{,R} + F/2 - RF_{,R}/2}{1 + F_{,R}}, \quad (7.43)$$

$$R_{rr} + \frac{R}{2} = \frac{\tilde{T}_{rr} + (\partial_t^2 + \partial_r^2 - \Delta)F_{,R} - F/2 + RF_{,R}/2}{1 + F_{,R}}, \quad (7.44)$$

because the left-hand side does not contain second derivatives of curvature. The derivatives of $A(r,t)$ and $B(r,t)$ squared in the limit of a weak gravitational field can be ignored, which leads to the following formulas for the components R_{00} and R_{rr} of the Ricci tensor and the curvature scalar R:

$$R_{00} \approx \frac{A'' - \ddot{B}}{2} + \frac{A'}{r}, \quad (7.45)$$

$$R_{rr} \approx \frac{\ddot{B} - A''}{2} + \frac{B'}{r}, \quad (7.46)$$

$$R \approx A'' - \ddot{B} + \frac{2A'}{r} - \frac{2B'}{r} + \frac{2(1-B)}{r^2}. \quad (7.47)$$

If the matter energy density inside the cloud, i.e., for $r < r_{\mathrm{m}}$, is much larger than the average cosmological energy density, the following relations hold:

$$F_{,R} \ll 1, \quad F \ll R. \quad (7.48)$$

The gravity modification effects for static solutions are not large in this limit, and the solution is close to the standard Schwarzschild solution, in agreement with the published results. As was noted above, study [296] showed that high-frequency oscillations of curvature are excited in systems with growing density. We can disregard spatial derivatives F'_R in such solutions in comparison to time derivatives because the characteristic time of oscillations is microscopically small and spatial variations are macroscopically large. It therefore follows from Eqn (7.3) that $(\partial_t^2 - \Delta)F_{,R} = (\tilde{T} + R)/3$, and we obtain

$$B_1' + \frac{B_1}{r} = r\tilde{T}_{00}, \quad (7.49)$$

$$\begin{aligned} A_1'' - \frac{A_1'}{r} &= -\frac{3B_1}{r^2} + \ddot{B}_1 + \tilde{T}_{00} - 2\tilde{T}_{rr} + \frac{\tilde{T}_{\theta\theta}}{r^2} \\ &+ \frac{\tilde{T}_{\varphi\varphi}}{r^2\sin^2\theta} \equiv S_A. \end{aligned} \quad (7.50)$$

Assuming deviations from the Minkowski metric to be small, we disregard the corrections to $T_{\mu\nu}$ that are caused by space–time curvature. We check below in which case this assumption is valid.

Equation (7.49) has the solution

$$B_1(r,t) = \frac{C_B(t)}{r} + \frac{1}{r}\int_0^r \mathrm{d}r' r'^2 \tilde{T}_{00}(r',t). \quad (7.51)$$

To avoid a singularity at $r = 0$, we must set $C_B(t) \equiv 0$. Then the formula for B_1 formally coincides with the standard Schwarzschild solution, and the solution for A_1 contains additional freedom:

$$A_1(r,t) = C_{1A}(t)\,r^2 + C_{2A}(t) + \int_r^{r_{\mathrm{m}}} \mathrm{d}r_1\, r_1 \int_{r_1}^{r_{\mathrm{m}}} \frac{\mathrm{d}r_2}{r_2} S_A(r_2,t). \quad (7.52)$$

Integration limits are chosen such that no singularity occurs at $r_2 = 0$.

Using Eqn (7.51) combined with $C_B = 0$, we present S_A in the form

$$\begin{aligned} S_A &= -\frac{3}{r^3}\int_0^r \mathrm{d}r' r'^2 \tilde{T}_{00}(r',t) + \frac{1}{r}\int_0^r \mathrm{d}r' r'^2 \ddot{\tilde{T}}_{00}(r',t) \\ &+ \tilde{T}_{00} - 2\tilde{T}_{rr} + \frac{\tilde{T}_{\theta\theta}}{r^2} + \frac{\tilde{T}_{\varphi\varphi}}{r^2\sin^2\theta}, \end{aligned} \quad (7.53)$$

to eventually obtain

$$\begin{aligned} A_1(r,t) &= C_{1A}(t)\,r^2 + C_{2A}(t) \\ &+ \int_r^{r_{\mathrm{m}}} \mathrm{d}r_1\, r_1 \int_{r_1}^{r_{\mathrm{m}}} \frac{\mathrm{d}r_2}{r_2}\left(\tilde{T}_{00}(r_2,t) - 2\tilde{T}_{rr}(r_2,t)\right. \\ &\left. + \frac{\tilde{T}_{\theta\theta}(r_2,t)}{r^2} + \frac{\tilde{T}_{\varphi\varphi}(r_2,t)}{r^2\sin^2\theta}\right) - \int_r^{r_{\mathrm{m}}} \mathrm{d}r_1\, r_1 \int_{r_1}^{r_{\mathrm{m}}} \frac{\mathrm{d}r_2}{r_2} \\ &\times \left(\frac{3}{r_2^3}\int_0^{r_2} \mathrm{d}r' r'^2 \tilde{T}_{00}(r',t) - \frac{1}{r_2}\int_0^{r_2} \mathrm{d}r' r'^2 \ddot{\tilde{T}}_{00}(r',t)\right). \end{aligned} \quad (7.54)$$

We first find the integration constants in the cases where the curvature does not contain an oscillating term. We define the mass inside a radius r standardly

$$M(r,t) = \int_0^r \mathrm{d}^3\bar{r}\, T_{00}(\bar{r},t) = 4\pi \int_0^r \mathrm{d}\bar{r}\,\bar{r}^2\, T_{00}(\bar{r},t)\,. \qquad (7.55)$$

If the entire matter is concentrated within a radius r_{m}, the total mass of the system is $M \equiv M(r_{\mathrm{m}})$, and, in correspondence with the classic Schwarzschild solution, it does not depend on time. Because $\tilde{T}_{00} = 8\pi T_{00}/m_{\mathrm{Pl}}^2$, for $r > r_{\mathrm{m}}$ we obtain $B_1 = r_{\mathrm{g}}/r$, where $r_{\mathrm{g}} = 2M/m_{\mathrm{Pl}}^2$ is the standard external Schwarzschild radius, in accordance with expectations.

We now calculate A_1 in (7.54). If $r > r_{\mathrm{m}}$, the first integral evidently vanishes because $r_2 > r_{\mathrm{m}}$ and, according to the assumption, $T_{\mu\nu} = 0$. The integral containing $\ddot{\tilde{T}}_{00}$ also vanishes because the external mass is constant. The remaining integral can be easily calculated:

$$\int_r^{r_{\mathrm{m}}} \mathrm{d}r_1\, r_1 \int_{r_1}^{r_{\mathrm{m}}} \frac{\mathrm{d}r_2}{r_2}\,\frac{3}{r_2^3} \int_0^{r_2} \mathrm{d}r' r'^2\, \tilde{T}_{00}(r',t)$$
$$= \frac{r_{\mathrm{g}}}{r} + \frac{r_{\mathrm{g}}\, r^2}{2r_{\mathrm{m}}^3} - \frac{3r_{\mathrm{g}}}{2r_{\mathrm{m}}}\,. \qquad (7.56)$$

The metric coefficient outside the source is therefore given by

$$A_1 = -\frac{r_{\mathrm{g}}}{r} + \left(C_{1A}(t) - \frac{r_{\mathrm{g}}}{2r_{\mathrm{m}}^3}\right) r^2 + \left(C_{2A}(t) + \frac{3r_{\mathrm{g}}}{2r_{\mathrm{m}}}\right). \qquad (7.57)$$

We set

$$C_{1A} = \frac{r_{\mathrm{g}}}{2r_{\mathrm{m}}^3}\,, \qquad C_{2A} = -\frac{3r_{\mathrm{g}}}{2r_{\mathrm{m}}}\,. \qquad (7.58)$$

The first condition is needed to remove the term proportional to r^2 at infinity, and the second condition is optional but can always be imposed by redefining time. We note that this choice of arbitrary constants is valid if the solution does not contain rapidly oscillating terms.

The internal solution in modified gravity has the same form (7.51) and (7.54); however, the coefficient C_{1A} can depend on time in a nontrivial way. This coefficient can be found from Eqn (7.47) if the scalar curvature $R(t)$ is known. Using Eqns (7.51) and (7.54) and comparing them with Eqn (7.47), we conclude that the dominant contribution to curvature comes from the sum $A'' + 2A'/r$ and hence $C_{1A}^{(\mathrm{osc})}(t) = R(t)/6$ with the curvature scalar calculated in [294]:

$$R(t) = R_{\mathrm{GR}}(r)\, y(t)\,, \qquad (7.59)$$

where $R_{\mathrm{GR}} = -8\pi T(r)/m_{\mathrm{Pl}}^2$ is the GR solution, and the rapidly oscillating function $y(t)$ can significantly exceed unity. According to [294], the maximum value of y in the peak region is

$$y(t) \sim 6n(2n+1)\, m t_{\mathrm{u}} \left(\frac{t_{\mathrm{u}}}{t_{\mathrm{contr}}}\right)$$
$$\times \left(\frac{\rho_{\mathrm{m}}(t)}{\rho_{\mathrm{m0}}}\right)^{(n+1)/2} \left(\frac{\rho_{\mathrm{c}}}{\rho_{\mathrm{m0}}}\right)^{2n+2}, \qquad (7.60)$$

where t_{u} is the age of the universe and t_{contr} is the characteristic time of system contraction; thus, the density of the contracting cloud behaves as $\rho_{\mathrm{m}}(t) = \rho_{\mathrm{m0}}(1 + t/t_{\mathrm{contr}})$ and ρ_{m0} and $\rho_{\mathrm{c}} = 10^{-29}$ g cm^{-3} are the initial density of cloud mass/energy and the present-day cosmological constant. Results in [293] show that the parameter m contained in Eqn (7.27) must exceed 10^5 GeV to avoid disagreement with primordial nucleosynthesis. The factor mt_{u} therefore acquires a giant value, $mt_{\mathrm{u}} \geqslant 10^{47}$, and y can be much greater than unity unless it is suppressed by a small ratio $(\rho_{\mathrm{c}}/\rho_{\mathrm{m0}})^{2n+2}$ with a large exponent n.

We note that the vacuum solutions in ordinary and modified gravity significantly differ. The term proportional to r^2 emerges in the standard case outside ($r > r_{\mathrm{m}}$) and inside ($r < r_{\mathrm{m}}$) the cloud with the same coefficient and must therefore vanish. On the other hand, there is no such condition in modified gravity, and the term $C_{1A} r^2$ can therefore be present at $r < r_{\mathrm{m}}$ and absent at $r \gg r_{\mathrm{m}}$.

The vacuum solution for R can be apparently written as $R \sim R_{\mathrm{c}}$, where R_{c} is the small cosmological curvature, plus a possible oscillating term. The metric functions within the cloud are therefore given by

$$B(r,t) = 1 + \frac{2M(r,t)}{m_{\mathrm{Pl}}^2 r} \equiv 1 + B_1^{(\mathrm{Sch})}\,, \qquad (7.61)$$

$$A(r,t) = 1 + \frac{R(t)\, r^2}{6} + A_1^{(\mathrm{Sch})}(r,t)\,. \qquad (7.62)$$

In other words, we have constructed a solution assuming that it consists of two terms: the Schwarzschild term and the oscillating one, and the rapidly oscillating part emerges in systems with an energy density that slowly varies with time. The formula for $A_1^{(\mathrm{Sch})}(r,t)$ is determined by integrals (7.54) with the constants $C_{A1} = r_{\mathrm{g}}/2r_{\mathrm{m}}^3$ and $C_{A2} = -3r_{\mathrm{g}}/r_{\mathrm{m}}$, as follows from (7.58).

As regards the integrals in (7.54) in the inner region, we calculate them under the assumption that matter is nonrelativistic and therefore the spatial components $T_{\mu\nu}$ are negligibly small compared to T_{00}. We also assume for simplicity that the mass/energy density $T_{00} \equiv \rho_{\mathrm{m}}(t)$ is constant across space but can depend on time. The first two integrals in Eqn (7.54) mutually cancel, and only the integral that is proportional to the second derivative of the mass density remains. As a result, we obtain

$$A_1^{(\mathrm{Sch})}(r,t) = \frac{r_{\mathrm{g}} r^2}{2r_{\mathrm{m}}^3} - \frac{3r_{\mathrm{g}}}{2r_{\mathrm{m}}} + \frac{\pi\ddot{\rho}_{\mathrm{m}}}{3m_{\mathrm{Pl}}^2}\,(r_{\mathrm{m}}^2 - r^2)^2\,. \qquad (7.63)$$

As we have noted, $R(t)$ in the modified theory is usually larger than its counterpart in GR, $|R_{\mathrm{GR}}| = 8\pi\rho_{\mathrm{m}}/m_{\mathrm{Pl}}^2$, and therefore the second term in the right-hand side of Eqn (7.62), $R(t)\, r^2/6$, makes the dominant contribution to A_1 at sufficiently large r. Indeed, $r^2 R(t) \sim r^2 y R_{\mathrm{GR}}$, and $y > 1$, while the canonical Schwarzschild contribution is of the order of $r_{\mathrm{g}}/r_{\mathrm{m}} \sim \rho_{\mathrm{m}} r_{\mathrm{m}}^2/m_{\mathrm{Pl}}^2 \lesssim r_{\mathrm{m}}^2 R_{\mathrm{GR}}$.

The equation of motion of a nonrelativistic probe (geodetic equation) in the lowest order in gravitational interaction has the form

$$\ddot{r} = -\frac{A'}{2} = -\frac{1}{2}\left(\frac{R(t)\, r}{3} + \frac{r_{\mathrm{g}} r}{r_{\mathrm{m}}^3}\right), \qquad (7.64)$$

where A is defined in Eqn (7.62). Because $R(t)$ is negative in the modified gravity considered here and its absolute value is large, gravitational repulsion emerges within the contracting cloud with $\rho > \rho_{\mathrm{c}}$, which turns out to be stronger than gravitational attraction if

$$\frac{|R|\,r_{\rm m}^3}{3r_{\rm g}} = \frac{|R|\,r_{\rm m}^3 m_{\rm Pl}^2}{6M} = \frac{|R|\,r_{\rm m}^3 m_{\rm Pl}^2}{8\pi\rho\,r_{\rm m}^3} = \frac{|R|}{\tilde{T}_{00}} \equiv y > 1\,. \tag{7.65}$$

Cellular structures, in particular, cosmic voids, can be formed as a result.

It is of interest that the gravitational repulsion force is manifested in this theory in the form of short but strong kicks, significantly stronger than ordinary gravitational attraction.

We note that although the result was obtained in the approximation of flat background space, it has a more general nature and also remains valid in the case of curved space–time. Moreover, it is natural to expect that antigravity also emerges in other similar versions of $F(R)$ theories.

8. Conclusion

After our review had been submitted for publication, a series of studies appeared that are directly related to the subject we discuss. Following the referee's proposal, we briefly describe some of the most recent publications. First of all, review [304] should be noted, where various types of modifications of gravity are considered rather than $F(R)$, as in our review. Review [304] contains a description of cosmological evolution from the inflationary epoch to our time and a model of bouncing from the singularity in the collapsing Universe.

Constraints were imposed in [305] on the parameters of scalar–tensor and $F(R)$ theories based on an analysis of observations of gravitational waves generated by merging neutron stars.

A detailed discussion of cosmological evolution in $F(R)$ theories is reported in study [306], which had not been published yet at the time our review was in preparation.

The authors of [307] make a strict distinction between dark energy and modified gravity on the basis of strong and weak equivalence principles. The phenomenology and typical observable characteristics of those two categories of models are analyzed.

Review [308] stresses the quantum nature of the dark energy problem and the necessity of a solution that circumvents the so-called Weinberg no-go theorem [25].

A critical analysis of a vast array of astronomical observations of different kinds is presented in [309]. In the opinion of its authors, the very reality of cosmological accelerations has been reliably established.

The issue of how reliably type-Ia supernovae can be used as standard candles is discussed in [310], given that gravitation may be modified. The authors conclude that type-Ia supernovae cannot be considered standard candles in theories where gravity is time dependent, and the conclusion regarding accelerated expansion may be invalid in general. However, the issue remains open regarding the extent to which the required time dependence of the gravitational constant is compatible with the complete set of data evidencing in favor of antigravitating dark energy.

A new trend has recently emerged in the description of the accelerated expansion of the Universe and dark energy. This trend, referred to as swampland [311, 312], is an alternative to the string-theory landscape. This term means an almost infinite number of vacuum states (more than $10^{100,000}$) when superstring models are compactified to a 4D space. The number of vacuum states this huge opens an avenue to the anthropic solution of the vacuum energy problem [see the discussion above between Eqns (3.6) and (3.7) in Section 3].

The authors of [311, 312] note that it is generally believed that all self-consistent effective quantum field theories coupled to gravity can be with a high probability obtained this way or another by compactifying string theory, thus making string theories not very helpful from the perspective of low-energy phenomenology. According to the opinion put forward in [311, 312], this is not true, and actually effective field theories may exist that do not result in a self-consistent theory of gravity under ultraviolet closure. The authors of [311, 312] propose calling the set of theories that do not result in string theory under ultraviolet closure the swampland, in contrast to landscape.

The cosmological effects that could emerge in the 'swampland' picture are analyzed in [313]. A critical analysis of these models and their implications is contained in [314].

The number of studies where 'swampland' issues are explored is now several dozen, and, a separate review would be needed to analyze them.

Acknowledgments

We are grateful to one of the referees for constructive comments, to Gaston Folatelli for the discussion of systematic errors in the observable parameters of type-Ia supernovae and to Maria Pruzhinskaya for the careful reading of the text and numerous corrections pertaining to the history and cosmological applications of supernovae.

S B is grateful to Shun Saito and Tomomi Sunayama for the valuable references to publications on BAOs and critical comments and to the Russian Science Foundation (grant no. 19-12-00229) for support of supernovae simulation studies. A D's study of the vacuum energy problem and mechanism of the accelerated expansion of the Universe owing to scalar field condensate, which is described in Sections 3 and 6, was supported by the RSF grant no. 19-42-02004.

9. Appendices

9.1 Derivation of the Friedmann equations

9.1.1 Elementary derivation from Newtonian theory. We consider a space filled with uniformly and isotropically distributed matter and choose a small sphere with a radius $ra(t)$ in it. Here, r is a fixed, so-called comoving coordinate, and possible change in the radius is defined by the function $a(t)$. The notation is chosen in correspondence with the cosmological notations used in the text.

We now consider the behavior of a material probe on the boundary of that sphere. The external layer of the ball in the spherically symmetrical case is known to have no effect on the probe particle. The sum of the kinetic and potential energies of that particle must be conserved:

$$\frac{r^2\,\dot{a}^2}{2} - \frac{4\pi}{3}\,\frac{\rho r^2 a^2}{m_{\rm Pl}^2} = {\rm const}\,. \tag{9.1}$$

This equation, after dividing by $r^2 a^2/2$, is identical to Eqn (2.5).

We next consider the balance of the medium energy density $\mathrm{d}E = -P\,\mathrm{d}V$, where $E = \rho V$ and hence $\mathrm{d}E = V\,\mathrm{d}\rho + 3(\mathrm{d}a/a)\,V\rho$. This evidently yields Eqn (2.7).

Having differentiated Eqn (2.5) and using Eqn (2.7), we obtain Eqn (2.6).

A natural question may arise at this stage: how is it possible to obtain the relativistic Einstein theory in which not only mass but also pressure gravitates starting with the Newtonian nonrelativistic theory? The answer is that we assumed that the source of gravitation is, at least partly, not mass density but energy density. This is the critical difference from the Newtonian nonrelativistic theory.

9.1.2 Derivation from the variational principle. The basic cosmology equations, i.e., Friedmann equations, are usually derived in GR textbooks under the assumption of uniform and isotropic metric (2.3) or (2.4). Indeed, Eqn (2.12) with $\Lambda = 0$ in mixed components yields the Einstein tensor G_μ^ν in the form

$$G_\mu^\nu \equiv R_\mu^\nu - \frac{1}{2} g_\mu^\nu R = \frac{8\pi}{m_{\mathrm{Pl}}^2} T_\mu^\nu , \tag{9.2}$$

whence, for FRW metric (2.4), we obtain

$$G_0^0 = R_0^0 - \frac{1}{2} R = \frac{8\pi}{m_{\mathrm{Pl}}^2} T_0^0 . \tag{9.3}$$

Calculating the Ricci and Einstein tensors, we arrive at

$$G_0^0 = \frac{3k + 3(\dot{a})^2}{a^2} . \tag{9.4}$$

Because $T_0^0 = \rho$, we immediately obtain Eqn (2.5).

We now show that this equation can be derived within GR not using the energy–momentum tensor T_i^k and Einstein equations but directly from the action S. The derivation from the variational principle only requires knowledge of the curvature scalar R. This derivation contains a number of instructive points.

The action S consists of two terms: the gravitational component and the action of matter S_{m} determined by a Lagrangian $\mathcal{L}_{\mathrm{m}}$:

$$S = S_H + S_{\mathrm{m}} = \varkappa \int R\sqrt{-g}\,\mathrm{d}^4x + \int \mathcal{L}_{\mathrm{m}}\sqrt{-g}\,\mathrm{d}^4x . \tag{9.5}$$

The constant $\varkappa$ in Eqn (9.5) is

$$\varkappa \equiv \frac{1}{16\pi G_N} \equiv \frac{m_{\mathrm{Pl}}^2}{16\pi} . \tag{9.6}$$

We consider the simple case of a perfect (ideal) fluid, where the Lagrangian density $\mathcal{L}_{\mathrm{m}}$ is just the energy per unit volume, ρ. This case is considered in detail in Fock's book [315].

Having substituted the expression for the curvature scalar R in terns of the scale factor and its derivative and the energy density ρ multiplied by $\sqrt{-g} = a^3$, we obtain the Lagrangian function for $a(t)$:

$$L[a(t)] = -6\varkappa(a^2\ddot{a} + a\dot{a}^2 + ka) + \rho a^3 . \tag{9.7}$$

We can eliminate the second time derivative $\ddot{a}$ by integrating the corresponding component in the action by parts:

$$\begin{aligned}\int_{t_{\mathrm{in}}}^t a^2\ddot{a}\,\mathrm{d}t &= \int_{t_{\mathrm{in}}}^t a^2\,\mathrm{d}\dot{a} = a^2\dot{a}\Big|_{t_{\mathrm{in}}}^t - \int_{t_{\mathrm{in}}}^t \dot{a}\,\mathrm{d}a^2 \\ &= \mathrm{const} - 2\int_{t_{\mathrm{in}}}^t \dot{a}^2 a\,\mathrm{d}t .\end{aligned} \tag{9.8}$$

We now discuss the limits for integration over time. It would be desirable to take the lower limit as zero, $t_{\mathrm{in}} = 0$ (the time that is referred to as the Big Bang moment); but we must not forget that the solution is singular at that zero in idealized models. The singularity structure cannot be explored within GR. It is also enticing to set the upper limit at infinity, but this cannot be done either because our model universe may be finite in time.

Having eliminated the second derivative, we obtain the Lagrangian in its standard form with first derivatives alone, but in contrast to the action extremum principle in classical mechanics, where coordinates alone are specified at trajectory ends, we now have to specify the velocity as well (here, it is $\dot{a}$). Only under the condition that $\delta a = 0$ and $\delta\dot{a} = 0$ at the integration region boundaries can we drop the term outside the integral. We then obtain the Lagrangian for the scale factor $a(t)$ as

$$L = -6\varkappa(-\dot{a}^2 a + ka) + \rho a^3 . \tag{9.9}$$

We can use any barotropic equation of state to relate ρ to the particle number density n. We can assume, in particular, that ρ depends only on a (via n), not on $\dot{a}$. We then obtain the 'Hamiltonian'

$$\mathcal{H} = \frac{\partial L}{\partial \dot{a}}\dot{a} - L = 6\varkappa(a\dot{a}^2 + ka) - \rho a^3 .$$

It does not contain any explicit dependence on time; hence, the 'energy' must be constant:

$$(a\dot{a}^2 + ka) - \frac{\rho a^3}{6\varkappa} = \mathrm{const} . \tag{9.10}$$

We can set the constant in the right-hand side to zero by adding the density of dustlike matter to the energy density: $\rho \to \rho + \Delta\rho$. In other words, the arbitrary constant can be absorbed into the definition of energy density by selecting $\rho_0 a_0^3$. No arbitrariness seems to be involved in deriving the same equation from the Einstein equations in the standard way, as described above. However, arbitrariness is still there, but the constant is determined by the correct choice of the zero level for matter energy.

The choice of zero for energy density in Eqn (9.2) only seems to be natural: if there is no curvature, the Riemann, Ricci, and Einstein tensors are all zero, and we also set $\rho = 0$. But if the true equations of gravity contain the lambda term, then naturalness fails: the lambda term can be added to the left-hand side, or nonzero vacuum energy can be added to the right-hand side if the equation of state is $P = -\rho^{\mathrm{vac}}$. Otherwise, an arbitrary fraction of the lambda term can be added to the left-hand side and an arbitrary fraction of vacuum energy added to the right-hand side to obtain equivalent dynamics of the universe.

Having set the constant in Eqn (9.10) to zero, after dividing by a^3 and using $\varkappa$ from Eqn (9.6), we thus obtain the formula

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} , \tag{9.11}$$

which yields the well-known first Friedmann equation (2.5).

We now return to the issue of boundary terms that emerge as a result of integration by parts in deriving the equations of motion. The relation between the Hilbert principle and

Einstein equations is far from trivial because of the terms outside the integral.

The zero values of variations of derivatives on the boundaries are usually mentioned in textbooks, only to immediately forget about this restriction. The only exception we know is Wald's book [316] (see also Eqn (4.7) in a more recent book by Poisson [317]; this issue is also discussed in a book on cosmology [318] (pages 451–464)). It is shown in these books that terms containing external curvature must be added to the Hilbert action. That action is usually referred to as the Gibbons–Hawking–York action [319], although York [320] seems to be the first to have added external curvature to the action in an explicit form.

The following form of the action for the metric g for general matter fields ϕ is proposed, for example, in one of the frequently quoted papers [321]:

$$I(g, \phi) = \int_M \left[\frac{R}{16\pi} + L_{\mathrm{m}}(g, \phi) \right] + \frac{1}{8\pi} \oint_{\partial M} K, \tag{9.12}$$

where R is the curvature scalar for the metric g, L_{m} is the matter Lagrangian, and K is the trace of the external curvature tensor of the boundary. Units are chosen in Eqn (9.12) such that $m_{\mathrm{Pl}} = 1$. We do not define the external curvature K_{ik} in a formal way. It can be found, for example, in book [322] (volume 2, page 153 of the Russian edition. In our example with the Friedmann (and Robertson–Walker) metric with our metric signature, we have

$$K_{ik} = 0.5\dot{g}_{ik} \tag{9.13}$$

(cf. Eqn (21.70) in [322]). Here, Latin indices take the values 1, 2, 3 (not as in Landau and Lifshitz's book [323]!)

It is not difficult to understand the meaning of that formula for the external curvature: the boundary of the manifold ∂M is in our case a set of cross sections taken at $t = \mathrm{const}$, i.e., our entire 3D space. The external curvature is built, according to [322], of vectors that point to the fourth dimension, which in our case is t. It is in this way that the time derivative appears in Eqn (9.13).

Taking the trace

$$K \equiv K_i^i ,$$

we see that for the choice of the coefficients made in Eqn (9.12), our term outside the integral in (9.8), $a^2\dot{a}|_{t_0}^{t}$, which contains just the same time derivative as in Eqn (9.13), disappears. We must not forget about other coefficients that have not been presented here. Our simple example thus clarifies how the action term added to the Hilbert action in Eqn (9.12) is chosen.

9.2 Cosmological parameters

We here quote the values of the main cosmological parameters and briefly describe the methods used to measure them. So as not to clutter the list of references, we only refer to publication [324], which contains references to original studies and brief reviews.

A value of importance in cosmology is the so-called critical energy density

$$\rho_{\mathrm{c}} = \frac{3H^2 m_{\mathrm{Pl}}^2}{8\pi} , \tag{9.14}$$

which is also referred to as the 'closure' density. This term is related to the fact that $\rho = \rho_{\mathrm{c}}$ is the boundary value at which transition from the open Universe to the closed one occurs. According to astronomical data, the total energy density of all forms of matter in the Universe, ρ_{tot}, is very close to the critical value (see Eqn (9.16) below).

The cosmological energy density of a particular form of matter is characterized by a dimensionless parameter

$$\Omega_a = \frac{\rho_a}{\rho_{\mathrm{c}}} . \tag{9.15}$$

It has been established that

$$\Omega_{\mathrm{tot}} = \frac{\rho_{\mathrm{tot}}}{\rho_{\mathrm{c}}} = 1.006 \pm 0.006 . \tag{9.16}$$

This result is primarily obtained from the analysis of the angular fluctuation spectrum of the CMB. The physical wavelength that corresponds to the first maximum at the hydrogen recombination moment ($z \approx 10^3$) is known, and the angle at which it is observed now and therefore the position of the first maximum depends on the geometry of the Universe. This observed position just corresponds to a flat, Euclidian universe.

The cosmological density of baryon matter can be determined in several independent ways: the observed abundance of light elements, helium ^{4}He, and deuterium $\mathrm{D} \equiv {}^2\mathrm{H}$ produced in the process of primordial nucleosynthesis; the ratio of the height of peaks in microwave background fluctuations; and the scale at which diffusive or Silk damping commences. All these methods yield close values:

$$\Omega_{\mathrm{b}} h^2 = 0.022 , \tag{9.17}$$

where h is the dimensionless Hubble constant normalized to 100: $h = H/100$ km s^{-1} Mpc^{-1}. The generally adopted value is [324]

$$h = 0.673 \pm 0.012 . \tag{9.18}$$

This result is obtained from the analysis of angular fluctuations of the CMB and the large-scale structure of the Universe [325]. However, it is worth noting the systematic and rather significant disagreement of this result with h measured using standard astronomical methods [326]:

$$h = 0.738 \pm 0.024 . \tag{9.19}$$

It is currently not clear whether the Hubble parameter depends on distance in an unusual way, and therefore the expansion law differs from that adopted in standard ΛCDM cosmology (see, e.g., [183], where possible variations of the expansion law are analyzed), or the disagreement can be explained by the presence of an unstable but long-lived component of dark matter [327, 328]. However, we cannot rule out a more prosaic explanation, which also seems to be more probable: this disagreement is due to systematic errors in determining the astronomical distance ladder. In relation to the last possibility, of significant interest now is the method for direct measurement of H proposed in [211], which is free from uncertainties of standard astronomical methods.

The baryon matter density is anyway close to

$$\Omega_{\mathrm{b}} = 0.05 . \tag{9.20}$$

We note that no less than half of that 5% contribution of baryons to the cosmological density is directly observed. It is not known where and in what objects the missing half is located.

The total density of baryons and dark matter is determined using a number of independent cosmological observations. Some historically earlier methods cannot compete with those developed recently, such as analyzing CMB fluctuations or BAOs, but we nevertheless mention them to show that the analysis of physically different phenomena in the Universe results in close values of densities of baryons and dark matter and the total density of dark energy. Modifications of dynamics or gravitational interaction are not infrequently proposed as an alternative to dark matter, but they fail to describe the entire set of available evidence of dark matter.

Summarizing, the very presence of dark matter and the cosmological density of normally gravitating (baryon plus invisible, dark) matter are determined using data on:

(1) flat rotation curves: velocities of gas particles or small satellite galaxies does not decrease with the distance from the radiating center of a (large) galaxy but rather tend to a constant value. This law holds up to distances of the order of 10 galactic radii within which radiating matter is located;

(2) gravitational lensing of distant objects, which enables estimating the total amount of gravitating matter along the line of sight;

(3) balance of hot gas in rich galactic clusters: the gravity needed to contain gas inside the cluster is about five times stronger than the gravitation force generated by the visible matter;

(4) evolution of galactic cluster formations as a function of redshift: the number of galactic clusters with $\Omega = 1$ in the spatially uniform flat Universe, i.e., in the Universe with $z \sim 1$, where all mass is contained in normally gravitating matter, would be one third of the observed value. But if the main part of matter is present in the form of antigravitating dark energy, the observed picture agrees with theory.

All of these methods yield the total density of clusterized normally gravitating matter value $\Omega_{\mathrm{DM}} + \Omega_{\mathrm{b}} \approx 0.3$. More accurate methods based on the analysis of BAOs and angular fluctuations of the CMB temperature result in the value (with the baryon contribution subtracted)

$$\Omega_{\mathrm{DM}} = 0.27\,. \tag{9.21}$$

The contribution that is missing to obtain unity is provided by dark energy:

$$\Omega_{\mathrm{DE}} = 0.68\,. \tag{9.22}$$

As noted above, indications of the antigravitational properties of dark energy followed from the discovery that distant supernovae appear dimmer than was expected for standard decelerating cosmological expansion. Accelerated expansion is also needed to resolve the crisis with the age of the universe. The universe without antigravitating dark energy would be one and a half times younger than the age that follows from the age of old star clusters and nuclear chronology. The accelerated expansion results in suppression of structure formation at large scales, in agreement with observations. If the equation of state of dark energy is parameterized in the form $P = w\rho$, then we have (see [324] and [329])

$$w = -1.01 \pm 0.04\,. \tag{9.23}$$

It is worth noting that the data do not exclude the phantom value $w < -1$ and even favor it. A conclusion was made in [330] based on data on BAOs that variation in dark energy possibly occurs in the process of cosmological expansion. The available data unfortunately do not enable identifying the reason for accelerated expansion, be it dark energy or infrared modification of gravity.

9.3 Scalar field

We here present a summary of the formulas for a scalar field required for cosmology. The action for a real-valued scalar field is usually chosen in the form

$$A[\phi] = \int \mathrm{d}^4x\,\sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\,\partial_\mu\phi\,\partial_\mu\phi - U(\phi)\right) \tag{9.24}$$

and results in the equation of motion

$$D^2\phi + U'(\phi) = 0\,, \tag{9.25}$$

where $U' = \mathrm{d}U/\mathrm{d}\phi$, $D^2 = g^{\mu\nu}D_\mu D_\nu$, and D_μ is the covariant derivative in an external gravitational field. This equation in FRW metric (2.4) with $k = 0$ becomes

$$\ddot{\phi} + 3H\dot{\phi} - \frac{1}{a^2}\nabla^2\phi + U' = 0\,. \tag{9.26}$$

The energy–momentum tensor that corresponds to this action is

$$T_{\mu\nu} = 2\,\frac{\delta A}{\delta g^{\mu\nu}} = \partial_\mu\phi\,\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}(\partial\phi)^2 - U(\phi)\right]. \tag{9.27}$$

The energy and pressure densities are consequently

$$\rho = \frac{\dot{\phi}^2 + (\nabla\phi)^2/a^2}{2} + U(\phi)\,,$$

$$P_{ij} = \delta_{ij}\left[\frac{\dot{\phi}^2 - (\nabla\phi)^2/a^2}{2} - U(\phi)\right] + \frac{\partial_i\phi\,\partial_i\phi}{a^2}\,. \tag{9.28}$$

We note that if $U > 0$, the average pressure is always smaller than the energy density, $P_i^i/3 < \rho$. The vacuum-like equation of state $P \approx -\rho$ is apparently approximately valid in the case of ϕ slowly varying in time and space. The maximally stiff equation of state, $P = \rho$, occurs in the opposite limit case of rapidly varying $\phi(t)$, when the time derivative dominates in $T_{\mu\nu}$ and the field ϕ is spatially uniform. The speed of sound for such an equation of state is equal to the speed of light. This equation could be realized in cosmological contraction.

So-called tachyon equations of state are sometimes considered. They emerge if $U < 0$, for example, for $U = -m^2\phi^2/2$, i.e., in theories with the mass squared negative. It is generally believed that such theories should support superluminal propagation of signals, but this conclusion is not true: signal velocity is the velocity of its front propagation, which is determined by the asymptotic behavior of the refractive index for the frequency/energy tending to infinity. The potential can be disregarded in this limit, and we arrive at normal propagation with the speed of light. The group velocity is apparently superluminal, but this only means that the wave is deformed, a phenomenon that is quite natural because the vacuum state is unstable if $m^2 < 0$. Anomalous-dispersion media are known to exhibit the same picture: the group velocity is be superluminal but the wave is deformed, transforming into a shock wave with a nonanalytic front.

We note that the equation of state $P = P(\rho)$ does not always exist. Although the parameter $w(t) \equiv P(t)/\rho(t)$ can always be introduced, the functional relation $P = P(\rho)$ may be absent. For example, the relation between P and ρ can be nonlocal in time. Nevertheless, despite the absence of the equation of state, the set of cosmological equations is complete because the equation of motion of the field is the missing equation (9.26).

The field is quantized according to the standard procedure. The field operator in uniform space is expanded in its spatial Fourier modes with time-dependent coefficients:

$$\phi(t, \mathbf{x}) = \int \frac{\mathrm{d}^3 k}{\sqrt{2E_k\,(2\pi)^3}} \left[a_k \exp(\mathrm{i}\mathbf{k}\mathbf{x}) f_k(t) + a_k^\dagger \exp(-\mathrm{i}\mathbf{k}\mathbf{x}) f_k^*(t) \right], \tag{9.29}$$

where $E_k = \sqrt{m^2 + k^2}$. The functions $f_k(t)$ are solutions of the Fourier-transformed equation (9.25) or Eqn (9.26) if the FRW metric is used. In particular, in the flat-space case we obtain the known result: $f_k(t) = \exp(-\mathrm{i}E_k t)$. The solution of Eqn (9.26) in the cosmological case for the FRW metric is also known in analytic form. It can be expressed in terms of Bessel functions.

The operators a_k and $a_k^\dagger$ are creation and annihilation operators of the field ϕ quanta (elementary particles). They satisfy the commutation relations

$$[a_k^\dagger, a_{k'}] = 2E_k\,(2\pi)^2\,\delta^3(\mathbf{k} - \mathbf{k}'). \tag{9.30}$$

The operator a_k acting on the vacuum annihilates it, $a_k|\mathrm{vac}\rangle = 0$, while the operator $a_k^\dagger$ creates a single-particle state with the momentum k: $a_k^\dagger|\mathrm{vac}\rangle = |k\rangle$. The factor $2E_k$ in Eqn (9.30) only appears in the Minkowski space. The Wronskian of the solutions of the equations of motion for $f_k(t)$ should be used in the general case of a curved space–time.

Fermion fields are quantized in a similar way, but the commutator of creation–annihilation operators is replaced with the anticommutator. This circumstance is of great importance for mutual cancelation of the diverging parts of the vacuum energies of bosons and fermions (see Section 3).

References

1. Einstein A *Sitzungsber. Preuß. Akad. Wiss. Berlin* **48** 844 (1915)
2. Einstein A *Ann. Physik* **49** 769 (1916)
3. Friedman A *Z. Phys.* **10** 377 (1922)
4. Friedmann A *Z. Phys.* **21** 326 (1924)
5. Robertson H P *Rev. Mod. Phys.* **5** 62 (1933)
6. Walker A G, McCrea W H *Mon. Not. R. Astron. Soc.* **94** 159 (1933)
7. Hubble E *Proc. Natl. Acad. Sci. USA* **15** 168 (1929)
8. Lemaître G *Ann. Soc. Sci. Bruxelles A* **47** 49 (1927)
9. Nussbaumer H, Bieri L, arXiv:1107.2281
10. Einstein A *Sitzgsber. Preuß. Acad. Wiss.* **1** 142 (1918)
11. Chernin A D *Phys. Usp.* **51** 253 (2008); *Usp. Fiz. Nauk* **178** 267 (2008)
12. de Sitter W *Proc. Kon. Ned. Acad. Wet.* **20** 229 (1917)
13. Lemaître G *Nature* **127** 706 (1931)
14. Lemaître G *Ann. Soc. Sci. Bruxelles A* **53** 51 (1933)
15. Eddington A S *Mon. Not. R. Astron. Soc.* **90** 668 (1930)
16. Koksma J F, Prokopec T, arXiv:1105.6296
17. Pauli W *Ausgewählte Kapitel aus der Feldquantisierung gehalten an der Eidg. Techn. Hochschule in Zürich, 1950 – 1951* (Ausgearb. von U Hochstrasser, M R Schafroth) (Zürich: H. Maag, 1951); Translated from English: *Selected Topics in Field Quantization* (Pauli Lectures on Physics, Vol. 6) (Cambridge, Mass.: MIT Press, 1973)
18. Zel'dovich Ya B *Sov. Phys. Usp.* **11** 381 (1968); *Usp. Fiz. Nauk* **95** 209 (1968)
19. Visser M *Particles* **1** 138 (2018)
20. Gol'fand Yu A, Likhtman E P *JETP Lett.* **13** 323 (1971); *Pis'ma Zh. Eksp. Teor. Fiz.* **13** 452 (1971)
21. Volkov D V, Akulov V P *JETP Lett.* **16** 438 (1972); *Pis'ma Zh. Eksp. Teor. Fiz.* **16** 621 (1972)
22. Wess J, Zumino B *Phys. Lett.* **49** 52 (1974)
23. Gell-Mann M, Oakes R J, Renner B *Phys. Rev.* **175** 2195 (1968)
24. Shifman M A, Vainshtein A I, Zakharov V I *Nucl. Phys. B* **147** 385 (1978)
25. Weinberg S *Rev. Mod. Phys.* **61** 1 (1989)
26. Dolgov A D, in *The Quest for the Fundamental Constants in Cosmology. Proc. of the 24th Rencontre de Moriond. 9th Moriond Astrophysics Meeting, Les Arcs, France, March 5 – 12, 1989* (Eds J Adouse, J Tran Thanh Van) (Gif-sur-Yvette: Editions Frontieres, 1990) p. 227
27. Dolgov A D, in *Phase Transitions in Cosmology, Fourth Paris Colloquium, Observatoire de Paris, 4 – 9 June, 1997* (Eds H J De Vega, N Sanchez) (Singapore: World Scientific, 1998) p. 161
28. Binétruy P *Int. J. Theor. Phys.* **39** 1859 (2000) lectures at Les Houches summer school "The Early Universe" and Peyresq 4 meeting, July 1999
29. Sahni V, Starobinsky A *Int. J. Mod. Phys. D* **9** 373 (2000)
30. Weinberg S, astro-ph/0005265; talk given at Dark Matter 2000, February 2000
31. Fujii Y *Grav. Cosmol.* **6** 107 (2000)
32. Vilenkin A , in *The Dark Universe. Matter, Energy and Gravity. Proc. of the Space Telescope Science Institute Symp., Baltimore, MD, USA, April 2 – 5, 2001* (Space Telescope Science Institute Symp. Ser., Vol. 15, Ed. M Livio) (Cambridge: Cambridge Univ. Press, 2003) p. 173; hep-th/0106083
33. Sahni V *Class. Quantum Grav.* **19** 3435 (2002); in *Proc. of the Early Universe and Cosmological Observations: a Critical Review, Cape Town, 23 – 25 July 2001*; astro-ph/0202076
34. Straumann N, astro-ph/0203330, Invited lecture at the first Séminaire Poincaré, Paris, March 2002
35. Peebles P J E, Ratra B *Rev. Mod. Phys.* **75** 559 (2003)
36. Kim J E *Mod. Phys. Lett.* **19** 1039 (2004); in *2003 Intern. Symp. on Cosmology and Particle Astophysics CosPA 2003, Taipei, Taiwan, November 13 – 15, 2003*; hep-ph/0402043
37. Burgess C P *Ann. Physics* **313** 383 (2014); hep-th/0402200, contribution to the Proc. of SUSY 2003, Univ. of Arizona, Tucson AZ, June 2003
38. Bousso R *Gen. Rel. Grav.* **40** 607 (2008), arXiv:0708.4231
39. Dicke R H *Nature* **192** 440 (1961)
40. Carter B, in *Confrontation of Cosmological Theories with Observational Data. Proc. of the Symp., Krakow, Poland, September 10 – 12, 1973* (IAU Symp. 63, Ed. M Longair) (Dordrecht: D. Reidel Publ. Co., 1974) p. 291
41. Carr B J, Rees M J *Nature* **278** 605 (1979)
42. Rozental' I L *Sov. Phys. Usp.* **23** 296 (1980); *Usp. Fiz. Nauk* **131** 239 (1980)
43. Rozental I L *Elementarnye Chastitsy and Structura Vselennoi* (Elementary Particles and Structure of Universe) (Moscow: Nauka, 1984)
44. Barrow J D, Tipler F J *The Anthropic Cosmological Principle* (Oxford: Oxford Univ. Press, 1986)
45. Rozental I L *Geometriya, Dinamika, Vselennaya* (Geometry, Dynamics, Universe) (Moscow: Nauka, 1987)
46. Rozental I L *Big Bang, Big Bounce. How Particles and Fields Drive Cosmic Evolution* (Berlin: Springer-Verlag, 1988)
47. Vilenkin A *Phys. Rev. D* **27** 2848 (1983)
48. Linde A D *Phys. Lett. B* **175** 395 (1986)
49. Sakharov A D *Sov. Phys. JETP* **60** 214 (1984); *Zh. Eksp. Teor. Fiz.* **87** 375 (1984)
50. Kachru S, Kallosh R, Linde A, Trivedi S P *Phys. Rev. D* **68** 046005 (2003)
51. Douglas R M *JHEP* **2003** (05) 046 (2003)
52. Susskind L, hep-th/0302219
53. Carter B, McCrea W H *Phil. Trans. R. Soc. London A* **310** 347 (1983)
54. Weinberg S *Phys. Rev. Lett.* **59** 2607 (1987)

55. Linde A, in *Three Hundred Years of Gravitation* (Eds S W Hawking, W Israel) (Cambridge: Cambridge Univ. Press, 1987) p. 604
56. Tegmark M, Rees M J *Astrophys. J.* **499** 526 (1998)
57. Vilenkin A, in *Universe or Multiverse*? (Ed. B Carr) (Cambridge: Cambridge Univ. Press, 2007) p. 163; astro-ph/0407586
58. Garriga J, Vilenkin A *Phys. Rev. D* **61** 083502 (2000)
59. Garriga J, Livio M, Vilenkin A *Phys. Rev. D* **61** 023503 (1999)
60. Mersini-Houghton L, Adams F C *Class. Quantum Grav.* **25** 165002 (2008)
61. Hong S E, Stewart E D, Zoe H *Phys. Rev. D* **85** 083510 (2012)
62. Hartle J, Hertog T *Phys. Rev. D* **88** 123516 (2013)
63. Kane G L, Perry M J, Zytkow A N *New Astron.* **7** 45 (2002); astro-ph/0001197
64. Dolgov A D, in *The Very Early Universe. Proc. of the Nuffield Workshop, Cambridge, 21 June to 9 July, 1982* (Eds B W Gibbons, S W Hawking, S T C Siklos) (Cambridge: Cambridge Univ. Press, 1983)
65. Dolgov A D *JETP Lett.* **41** 345 (1985); *Pis'ma Zh. Eksp. Teor. Fiz.* **41** 280 (1985)
66. Dolgov A D *Phys. Rev. D* **55** 5881 (1997)
67. Bronstein M *Phys. Z. Sowjetunion* **3** 73 (1933)
68. Koyama K, Sakstein J *Phys. Rev. D* **91** 124066 (2015); arXiv: 1502.06872
69. Vainshtein A I *Phys. Lett. B* **39** 393 (1972)
70. Biswas T, Notari A *Phys. Rev. D* **74** 043508 (2006)
71. Rothman T, Matzner R *Astrophys. J.* **257** 450 (1982)
72. Accetta F S, Krauss L M, Romanelli P *Phys. Lett. B* **248** 146 (1990)
73. Copi C J, Davis A N, Krauss L M *Phys. Rev. Lett.* **92** 171301 (2004)
74. Bambi C, Giannotti M, Villante F L *Phys. Rev. D* **71** 123524 (2005)
75. Zhu W W et al. *Astrophys. J.* **809** 41 (2015)
76. Anderson J D et al. *Europhys. Lett.* **110** 1002 (2015)
77. Schlamminger S, Gundlach J H, Newman R D P*hys. Rev. D* **91** 121101(R) (2015)
78. Wilczek F *Phys. Rep.* **104** 143 (1984)
79. Özer M, Taha M O *Phys. Lett. B* **171** 363 (1986)
80. Özer M, Taha M O *Nucl. Phys. B* **287** 776 (1987)
81. Fujii Y *Astropart. Phys.* **5** 133 (1996)
82. Peccei R D, Solà J, Wetterich C *Phys. Lett. B* **195** 183 (1987)
83. Ford L H *Phys. Rev. D* **35** 2339 (1987)
84. Freese K et al. *Nucl. Phys. B* **287** 797 (1987)
85. Gasperini M *Phys. Lett. B* **194** 347 (1987)
86. Reuter M, Wetterich C *Phys. Lett. B* **188** 38 (1987)
87. Barr S M, Hochberg D *Phys. Lett. B* **211** 49 (1988)
88. Weiss N *Phys. Lett. B* **197** 42 (1987)
89. Fujii Y, Nishioka T *Phys. Rev. D* **42** 361 (1990)
90. Fujii Y, Nishioka T *Phys. Lett. B* **254** 347 (1991)
91. Abbott L F *Phys. Lett. B* **150** 427 (1985)
92. Banks T *Nucl. Phys. B* **249** 332 (1985)
93. Bertolami O *Nuovo Cimento B* **93** 36 (1986)
94. Barr S M *Phys. Rev. D* **36** 1691 (1987)
95. Mukohyama S, Randall L P*hys. Rev. Lett.* **92** 211302 (2004)
96. Dolgov A D, Kawasaki M *ЯФ* **68** 860 (2005); *Phys. Atom. Nucl.* **68** 828 (2005); astro-ph/0307442
97. Dolgov A D, Kawasaki M, astro-ph/0310822
98. Fujii Y *Phys. Rev. D* **62** 064004 (2000)
99. Fujii Y, Nishioka T *Phys. Lett. B* **25** 347 (1991)
100. Fujii Y *Astropart. Phys.* **5** 133 (1996)
101. Rubakov V A *Phys. Rev. D* **61** 061501(R) (2000); hep-ph/9911305
102. Hebecker A, Wetterich C *Phys. Rev. Lett.* **85** 3339 (2000)
103. Weinberg S *The Quantum Theory of Fields* Vol. 1 (Cambridge: Cambridge Univ. Press, 1995) Sec. 5.9
104. Rubakov V A, Tinyakov P G *Phys. Rev. D* **61** 087503 (2000)
105. Ogievetsky V I, Polubarinov I V *Sov. J. Nucl. Phys.* **4** 156 (1967); *Yad. Fiz.* **4** 216 (1966)
106. Emelyanov V, Klinkhamer F R *Phys. Rev. D* **85** 063522 (2012)
107. Emelyanov V, Klinkhamer F R *Int. J. Mod. Phys. D* **21** 1250025 (2012); arXiv:1108.1995
108. Emelyanov V, Klinkhamer F R *Phys. Rev. D* **85** 103508 (2012); arXiv:1109.4915
109. Emelyanov V, Klinkhamer F R *Phys. Rev. D* **86** 027302 (2012)
110. Steinhardt P J, Turok N *Science* **312** 1180 (2006)
111. Shapley H, Curtis H D, Bull. Natl. Res. Council Vol. 2, Pt. 3, No. 11 171–217 (1921)
112. Boehle A et al. *Astrophys. J.* **830** 17 (2016)
113. Hubble E *Proc. Natl. Acad. Sci. USA* **15** 168 (1929)
114. Sandage A, in *Practical Cosmology. Inventing the Past. Saas-Fee Advanced Course 23. Lecture Notes 1993. Swiss Society for Astro-Physics and Astronomy, XIV* (Berlin: Springer-Verlag, 1995) p. 1–232
115. Wirtz C *Astron. Nachricht.* **215** 349 (1922)
116. Wirtz C *Z. Astrophys.* **11** 261 (1936)
117. Seitter W C, Duerbeck H W, in *Cosmic Distance Scales in a Post-Hipparcos Era* (ASP Conference Series, vol. 167. Eds D Egret, A Heck) (San Francisco, CA: Astronomical Society of the Pacific, 1999) p. 237
118. van den Bergh S *J. R. Astron. Soc. Canada* **105** 197 (2011)
119. O'Raifeartaigh C, arXiv:1212.5499
120. Humason M L *Astrophys. J.* **74** 35 (1931)
121. Pruzhinskaya M V, Lisakov S M *Priroda* (12) 36 (2015)
122. Pruzhinskaya M V, Lisakov S M *J. Astron. History Heritage* **19** (2) 203 (2016)
123. Lemaître A G *Mon. Not. R. Astron. Soc.* **91** 483 (1931)
124. Lemaître A G *Mon. Not. R. Astron. Soc.* **91** 490 (1931)
125. Hubble E P *Astrophys. J.* **64** 321 (1926)
126. Shklovsky J *Astrophys. J.* **150** L1 (1967)
127. Kardashev N *Astrophys. J.* **150** L135 (1967)
128. Gunn J E, Tinsley B M *Nature* **257** 454 (1975)
129. Tinsley B M *Nature* **273** 208 (1978)
130. Fukugita M et al. *Astrophys. J.* **361** L1 (1990)
131. Fukugita M, Hogan C *Nature* **347** 120 (1990)
132. Fukugita M, Hogan C J, Peebles P J E *Nature* **366** 309 (1993)
133. Sandage A, Tammann G A *Astrophys. J.* **194** 559 (1974)
134. Weinberg S *Gravitation and Cosmology* (New York: Wiley, 1972); Translation into Russian: *Gravitatsiya i Kosmologiya* (Moscow: Mir, 1975)
135. Weinberg S *Cosmology* (Oxford: Oxford Univ. Press, 2008); Translation into Russian: *Kosmologiya* (Moscow: URSS, 2013)
136. Carroll S *Spacetime and Geometry, An Introduction to General Relativity* (San Francisco: Addison Wesley, 2004); gr-qc/9712019
137. Gorbunov D S, Rubakov V A *Introduction to the Theory of the Early Universe. Hot Big Bang Theory* (Singapore: World Scientific, 2011); Translated from Russian: *Vvedenie v Teoriyu Rannei Vselennoi. Teoriya Goryachego Bol'shogo Vzryva* (Moscow: LKI, 2008)
138. Challinor A, in *Workshop on Cosmology and Gravitational Physics, 15–16 December 2005, Thessaloniki, Greece*, Invited Review; astro-ph/0606548
139. Rubakov V A, Vlasov A D *Phys. Atom. Nucl.* **75** 1123 (2012); *Yad. Fiz.* **75** 1190 (2012); arXiv:1008.1704
140. Keel W C "The extragalactic distance scale", http://www.astr.ua.edu/keel/galaxies/distance.html
141. Churazov E et al. *Nature* **512** 406 (2014)
142. Poznanski D et al. *Mon. Not. R. Astron. Soc.* **382** 1169 (2007)
143. Cooke J et al. *Nature* **460** 237 (2009)
144. Jones D O et al. *Astrophys. J.* **768** 166 (2013)
145. Schmidt B P et al. *Astrophys. J.* **507** 46 (1998)
146. Riess A G et al. *Astron. J.* **116** 1009 (1998)
147. Perlmutter S et al. *Astrophys. J.* **517** 565 (1999)
148. Pskovskii Yu P *Sov. Astron.* **21** 675 (1977); *Astron. Zh.* **54** 1188 (1977)
149. Pskovskii Yu P *Sov. Astron.* **28** 658 (1984); *Astron. Zh.* **61** 1125 (1984)
150. Rust B W "The use of supernovae light curves for testing the expansion hypothesis and other cosmological relations", Ph.D. Thesis (Oak Ridge, TN: Univ. of Illinois, Oak Ridge National Lab., 1974)
151. Phillips M M *Astrophys. J. Lett.* **413** L105 (1993)
152. Baade W *Astrophys. J.* **88** 285 (1938)
153. Wilson O C *Astrophys. J.* **90** 634 (1939)
154. Zwicky F *Phys. Rev.* **55** 726 (1939)
155. Pskovskii Yu P *Sov. Astron.* **11** 63 (1967); *Astron. Zh.* **44** 82 (1967)
156. Phillips M M, in *1604–2004*: *Supernovae as Cosmological Lighthouses* (Eds M Turatto et al.) (ASP Conf. Ser., Vol. 342, Eds M Turatto et al.) (San Francisco: Astronomical Society of the Pacific, 2005) p. 211
157. Branch D, Tammann G A *Annu. Rev. Astron. Astrophys.* **30** 359 (1992)

158. de Vaucouleurs G, Pence W D *Astrophys. J.* **209** 687 (1976)
159. Folatelli G et al. *Astron. J.* **139** 120 (2010); arXiv:0910.3317
160. Astier P et al. *Astron. Astrophys.* **447** 31 (2006); astro-ph/0510447
161. Betoule M et al. *Astron. Astrophys.* **568** A22 (2014); arXiv: 1401.4064
162. Riess A G et al. *Astrophys. J.* **730** 119 (2011)
163. Scolnic D M et al. *Astrophys. J.* **859** 101 (2018); arXiv:1710.00845
164. Dam L H, Heinesen A, Wiltshire D L *Mon. Not. R. Astron. Soc.* **472** 835 (2017)
165. Rácz G et al. *Mon. Not. R. Astron. Soc.* **469** L1 (2017)
166. Buchert T *Mon. Not. R. Astron. Soc.* **473** L46 (2018)
167. Weinberg D H et al. *Phys. Rep.* **530** 87 (2013); arXiv:1201.2434
168. Bassett B A, Hlozek R, in *Dark Energy. Observational and Theoretical Approaches* (Cambridge: Cambridge Univ. Press, 2010) Sec. 6
169. Kodama H, Sasaki M *Prog. Theor. Phys. Suppl.* **78** 1 (1984)
170. Giovannini M *Int. J. Mod. Phys. D* **14** 363 (2005); astro-ph/0412601
171. Giovannini M *A Primer on the Physics of the Cosmic Microwave Background, by Massimo Giovannini* (Singapore: World Scientific, 2008)
172. Gorbunov D S, Rubakov V A *Introduction to the Theory of the Early Universe. Cosmological Perturbations and Inflationary Theory* (Singapore: World Scientific, 2011); Translated from Russian: *Vvedenie v Teoriyu Rannei Vselennoi. Kosmologicheskie Vozmushcheniya. Inflyatsionnaya Teoriya* (Moscow: KRASAND, 2010)
173. Komatsu E et. al. *Astrophys. J. Suppl. Ser.* **180** 330 (2009)
174. Hu W, Sugiyama N *Astrophys. J.* **471** 542 (1996); astro-ph/9510117
175. Dodelson S *Modern Cosmology* (Amsterdam: Academic Press, 2003)
176. Alam Sh et al., arXiv:1607.03155
177. Hong T, Han J L, Wen Z L *Astrophys. J.* **826** 154 (2016); arXiv:1511.00392
178. Seo H-J, Beutler F, Ross A J, Saito S *Mon. Not. R. Astron. Soc.* **460** (2016); arXiv:1511.00663
179. Anderson L et al. *Mon. Not. R. Astron. Soc.* **441** 24 (2014)
180. Eisenstein D J, Hu W, Tegmark M *Astrophys. J.* **504** L57 (1998)
181. Eisenstein D J et al., arXiv:0607061
182. Padmanabhan N, White M, Cohn J D *Phys. Rev. D* **79** 063523 (2009); arXiv:0812.2905
183. Dolgov A, Halenka V, Tkachev I *J. Cosmol. Astropart. Phys.* **10** 047 (2014); arXiv:1406.2445
184. Bassett B A, Hlozek R, arXiv:0910.5224
185. Tegmark M et al. (SDSS Collab.) *Phys. Rev. D* **74** 123507 (2006); astro-ph/0608632
186. Ade P A R et al. (Planck Collab.) *Astron. Astrophys.* **571** A16 (2014); arXiv:1303.5076
187. Beutler F et al. *Mon. Not. R. Astron. Soc.* **416** 3017 (2011)
188. Padmanabhan N et al. *Mon. Not R. Astron. Soc.* **427** 2132 (2012)
189. Anderson L et al. *Mon. Not. R. Astron. Soc.* **427** 3435 (2012)
190. Xia J-Q et al. *Phys. Rev. D* **85** 043520 (2012); arXiv:1103.0378
191. Percival W J et al. *Mon. Not. R. Astron. Soc.* **401** 2148 (2010)
192. Conley A et al. *Astrophys. J. Supp. Ser.* **192** 1 (2011)
193. Hicken M et al. *Astrophys. J.* **700** 1097 (2009); arXiv:0901.4804
194. Kazin E A et al. *Mon. Not. R. Astron. Soc.* **441** 3524 (2014)
195. Guy J et al. *Astron. Astrophys.* **466** 11 (2007)
196. Tripp R, Branch D *Astrophys. J.* **525** 209 (1999)
197. Sullivan M et al. *Mon. Not. R. Astron. Soc.* **340** 1057 (2003)
198. Henne V et al. *New Astron.* **51** 43 (2017)
199. Sullivan M et al. *Mon. Not. R. Astron. Soc.* **406** 782 (2010)
200. Johansson J et al. *Mon. Not. R. Astron. Soc.* **435** 1680 (2013)
201. Sullivan M et al. *Astrophys. J.* **648** 868 (2006)
202. Neill J D et al. *Astrophys. J.* **707** 1449 (2009)
203. Rigault M et al. *Astrophys. J.* **802** 20 (2015)
204. Folatelli G et al. *Astrophys. J.* **773** 53 (2013)
205. Bogomazov A I, Tutukov A V *Astron. Rep.* **55** 497 (2011); *Astron. Zh.* **88** 541 (2011)
206. Li W et al. *Publ. Astron. Soc. Pacific* **115** 453 (2003); astro-ph/0301428
207. Phillips M M et al. *Publ. Astron. Soc. Pacific* **119** 360 (2007); astro-ph/0611295
208. Fakhouri et al. *Astrophys. J.* **815** 58 (2015); arXiv:1511.01102
209. Alam U, Sahni V, Starobinsky A A *J. Cosmol. Astropart. Phys.* **2** 011 (2007)
210. Shafieloo A, Sahni V, Starobinsky A A *Phys. Rev. D* **80** 101301 (2009)
211. Blinnikov S, Potashov M, Baklanov P, Dolgov A *JETP Lett.* **96** 153 (2012); *Pis'ma Zh. Eksp. Teor. Fiz.* **96** 167 (2012)
212. Potashov M, Blinnikov S, Baklanov P, Dolgov A *Mon. Not. R. Astron. Soc.* **431** L98 (2013)
213. Baklanov P V, Blinnikov S I, Potashov M S, Dolgov A D *JETP Lett.* **98** 432 (2013); *Pis'ma Zh. Eksp. Teor. Fiz.* **98** 489 (2013)
214. Leibundgut B *Nucl. Phys. A* **688** 1 (2001)
215. Kirshner R P, Kwan J *Astrophys. J.* **193** 27 (1974)
216. Baron E et al. *Astrophys. J. Lett.* **616** L91 (2004)
217. Bartel N et al. *Astrophys. J.* **668** 924 (2007)
218. Taddia F et al. *Astron. Astrophys.* **555** A10 (2013)
219. Grassberg E K, Imshennik V S, Nadyozhin D K *Astrophys. Space Sci.* **10** 28 (1971)
220. Grasberg E K, Nadezhin D K *Sov. Astron. Lett.* **12** 68 (1986); *Pis'ma Astron. Zh.* **12** 68 (1986)
221. Chugai N N et al. *Mon. Not. R. Astron. Soc.* **352** 1213 (2004)
222. Woosley S E, Blinnikov S, Heger A *Nature* **450** 390 (2007)
223. Baade W *Astron. Nachricht.* **228** 359 (1926)
224. Wesselink A J *Bull. Astron. Inst. Netherlands* **10** 91 (1946)
225. Abbott B P et. al. *Phys. Rev. Lett.* **116** 061102 (2016)
226. Abbott B P et al. *Phys. Rev. Lett.* **119** 161101 (2017)
227. Abbott B P et al. *Astrophys. J.* **848** L13 (2017)
228. Savchenko V et al. *Astrophys. J.* **848** L15 (2017)
229. Lipunov V M et al. *Astrophys. J.* **850** L1 (2017)
230. Blinnikov S I et al. *Sov. Astron. Lett.* **10** 177 (1984); *Pis'ma Astron. Zh.* **10** 422 (1984); arXiv:1808.05287
231. Lattimer J M, Schramm D N *Astrophys. J.* **210** 549 (1976)
232. Clark J P A, Eardley D M *Astrophys. J.* **215** 311 (1977)
233. Blinnikov S I et al. *Sov. Astron.* **34** 595 (1990); *Astron. Zh.* **67** 1181 (1990)
234. Schutz B F *Nature* **323** 310 (1986)
235. Abbott B P et al. (The LIGO Scientific Collab., The Virgo Collab., The 1M2H Collab., The Dark Energy Camera GW-EM Collab. and the DES Collab., The DLT40 Collab., The Las Cumbres Observatory Collab., The VINRO UGE Collab. & The MASTER Collab.) *Nature* **551** 85 (2017)
236. Riess A G et al. *Astrophys. J.* **826** 56 (2016)
237. Riess A G et al. *Astrophys. J.* **861** 126 (2018)
238. Ade P A R et al. (Planck Collab.) *Astron. Astrophys.* **594** A13 (2016)
239. Aghanim N et al. (Planck Collab.), arXiv:1807.06209
240. Caldwell R R *Phys. Lett. B* **545** 23 (2002)
241. Peebles P J E, Ratra B *Astrophys J.* **325** L17 (1988)
242. Ratra B, Peebles P J E *Phys. Rev. D* **37** 3406 (1988)
243. Wetterich C *Nucl. Phys. B* **302** 668 (1988)
244. Wetterich C *Nucl. Phys. B* **302** 645 (1988)
245. Zlatev I, Wang L, Steinhardt P J *Phys. Rev. Lett.* **82** 896 (1999)
246. Liddle A R, Scherrer R J *Phys. Rev. D* **59** 023509 (1999)
247. Chiba T, Sugiyama N, Nakamura T *Mon. Not. R. Astron. Soc.* **289** L5 (1997)
248. Caldwell R R, Dave R, Steinhardt P J *Phys. Rev. Lett.* **80** 1582 (1998)
249. Hu W, Eisenstein D J, Tegmark M *Phys. Rev. D* **59** 023512 (1998)
250. Huey G et al. *Phys. Rev. D* **59** 063005 (1999)
251. Ferreria P G, Joyce M *Phys. Rev. Lett.* **79** 4740 (1997)
252. Ferreria P G, Joyce M *Phys. Rev. D* **58** 023503 (1998)
253. Coble K, Dodelson S, Frieman J A *Phys. Rev. D* **55** 1851 (1997)
254. Turner M S, White M *Phys. Rev. D* **56** 4439 (1997)
255. Frieman J, Waga I *Phys. Rev. D* **57** 4642 (1998)
256. Carroll S M *Phys. Rev. Lett.* **81** 3067 (1998); astro-ph/9806099
257. Steinhardt P J, Wang L, Zlatev I *Phys. Rev.* **59** 123504 (1999)
258. Chiba T *Phys. Rev. D* **60** 083508 (1999); gr-qc/9903094
259. Bento M C, Bertolami O *Gen. Rel. Grav.* **31** 1461 (1999); gr-qc/9905075
260. Perrotta F, Baccigalupi C, Matarrese S *Phys. Rev. D* **61** 023507 (2000); astro-ph/9906066
261. Garriaga J, Livio M, Vilenkin A *Phys. Rev. D* **61** 023503 (2000); astro-ph/9906210
262. Bartolo N, Pietroni M *Phys. Rev. D* **61** 023518 (2000); hep-ph/9908521
263. Fujii Y *Phys. Rev. D* **62** 044011 (2000)

264. Copeland E J, Sami M, Tsujikawa S *Int. J. Mod. Phys. D* **15** 1753 (2006)
265. Bean R, arXiv:1003.4468
266. Mortonson M J, Weinberg D H, White M to appear as Chapter 25 of Particle Data Group 2014 Review of Particle Physics; arXiv:1401.0046
267. Born M, Infeld L *Proc. R. Soc. Lond. A* **144** 425 (1934)
268. Deser S, Gibbons G W *Class. Quantum Grav.* **15** L35 (1998)
269. Comelli D, Dolgov A *JHEP* **2004** (11) 062 (2004)
270. Lucchin F, Matarrese S *Phys. Rev. D* **32** 1316 (1985)
271. Sahni V, Feldman H, Stebbins A *Astrophys. J.* **385** 1 (1992)
272. Capozziello S, Carloni S, Troisi A *Recent Res. Dev. Astron. Astrophys.* **1** 625 (2003); astro-ph/0303041
273. Carroll S M et al. *Phys. Rev. D* **70** 043528 (2004); astro-ph/0306438
274. Dolgov A D, Kawasaki M *Phys. Lett. B* **573** 1 (2003)
275. Hu W, Sawicki I *Phys. Rev. D* **76** 064004 (2007)
276. Appleby A, Battye R *Phys. Lett. B* **654** 7 (2007)
277. Starobinsky A A *JETP Lett.* **86** 157 (2007); *Pis'ma Zh. Eksp. Teor. Fiz.* **86** 183 (2007)
278. Nojiri S, Odintsov S *Phys. Rep.* **505** 59 (2011)
279. Appleby S A, Battye R A, Starobinsky A A *JCAP* **2010** (06) 005 (2010)
280. Frolov A V *Phys. Rev. Lett.* **101** 061103 (2008)
281. Arbuzova E V, Dolgov A D *Phys. Lett. B* **700** 289 (2011)
282. Reverberi L *Phys. Rev. D* **87** 084005 (2013)
283. Gurovich V Ts, Starobinsky A A *Sov. Phys. JETP* **50** 844 (1979); *Zh. Eksp. Teor. Fiz.* **77** 1683 (1979)
284. Starobinsky A A *JETP Lett.* **30** 682 (1979); *Pis'ma Zh. Eksp. Teor. Fiz.* **30** 719 (1979)
285. Starobinsky A A, in *Proc. of the Second Seminar "Quantum Theory of Gravity"* (Moscow: INR Press, 1982) pp. 58–72; reprinted in: *Quantum Gravity* (Eds. M A Markov, P C West) (New York: Plenum Publ., 1984) pp. 103–128
286. Starobinsky A A *Phys. Lett. B* **91** 99 (1980)
287. Mamaev S G, Mostepanenko V M, Starobinskii A A *Sov. Phys. JETP* **43** 823 (1976); *Zh. Eksp. Teor. Fiz.* **70** 1577 (1976)
288. Zel'dovich Ya B, Starobinskii A A *JETP Lett.* **26** 252 (1977); *Pis'ma Zh. Eksp. Teor. Fiz.* **26** 373 (1977)
289. Vilenkin A *Phys. Rev. D* **32** 2511 (1985)
290. Mijic M B, Morris M S, Suen W-M *Phys. Rev. D* **34** 2934 (1986)
291. Suen W-M, Anderson P R *Phys. Rev. D* **35** 2940 (1987)
292. Gorbunov D S, Panin A G *Phys. Lett. B* **700** 157 (2011); arXiv:1009.2448
293. Arbuzova E V, Dolgov A D, Reverberi L *JCAP* **2012** (02) 049 (2012); arXiv:1112.4995
294. Arbuzova E V, Dolgov A D, Reverberi L *Eur. Phys. J. C* **72** 2247 (2012)
295. Arbuzova E V, Dolgov A D, Reverberi L *Phys. Rev. D* **88** 024035 (2013)
296. Gorbunov D, Tokareva A *JETP* **120** 528 (2015); *Zh. Eksp. Teor. Fiz.* **147** 599 (2015); arXiv:1412.3413
297. Gorbunov D, Tokareva A *Phys. Rev. D* **96** 103527 (2017); arXiv:1412.3770
298. Arbuzova E V, Dolgov A D, Reverberi L *Eur. Phys. J. C* **78** 481 (2018); arXiv:1707.02541
299. Arbuzova E V, Dolgov A D, Reverberi L *Astropart. Phys.* **54** 44 (2014)
300. Capozziello S, De Laurentis M *Phys. Rep.* **509** 167 (2011)
301. Capozziello S, Stabile A, Troisi A *Phys. Rev. D* **76** 1040 (2007)
302. de la Cruz-Dombriz A, Dobado A, Maroto A L *Phys. Rev. D* **80** 124011 (2009)
303. Cembranos J A R, de la Cruz-Dombriz A, Montes Nunez B *JCAP* **2012** (04) 021 (2012)
304. Nojiri S, Odintsov S D, Oikonomou V K *Phys. Rep.* **692** 1 (2017); arXiv:1705.11098
305. Nojiri S, Odintsov S D *Phys. Lett. B* **779** 425 (2018)
306. Capozziello S, Mantica C A, Molinari L G *Int. J. Geom. Meth. Mod. Phys.* **16** 1950008 (2018)
307. Joyce A, Lombriser L, Schmidt F *Annu. Rev. Nucl. Part. Sci.* **66** 95 (2016); arXiv:1601.06133
308. Brax P *Rep. Prog. Phys.* **81** 016902 (2018)
309. Rubin D, Hayden B *Astrophys. J.* **833** L30 (2016); arXiv:1610.08972
310. Wright B S, Li B *Phys. Rev. D* **97** 083505 (2018); arXiv:1710.07018
311. Brennan T D, Carta F, Vafa C, arXiv:1711.00864
312. Obied G et al., arXiv:1806.08362
313. Agrawal P et al. *Phys. Lett. B* **784** 271 (2018); arXiv:1806.09718
314. Akrami Y, Kallosh R, Linde A, Vardanyan V *Fortschr. Phys.* **67** 1800075 (2019)
315. Fock V A *Theory of Space, Time, and Gravitation* (New York: Pergamon Press, 1959); Translated from Russian: *Teoriya Prostranstva, Vremeni i Tyagoteniya* (Moscow: GITTL, 1955)
316. Wald R M *General Relativity* (Chicago: Univ. Chicago Press, 1984); Translation into Russian: *Obshchaya Teoriya Otnositel'nosti* (Moscow: RUDN, 2008)
317. Poisson E *A Relativist's Toolkit*: *The Mathematics of Black-Hole Mechanics* (Cambridge: Cambridge Univ. Press, 2004)
318. Kolb E W, Turner M S *The Early Universe* (Reading, Mass.: Addison-Wesley, 1990)
319. Gibbons G W, Hawking S W *Phys. Rev. D* **15** 2752 (1977)
320. York J W *Phys. Rev. Lett.* **28** 1082 (1972)
321. Hawking S W, Horowitz G T *Class. Quantum Grav.* **13** 1487 (1996); gr-qc/9501014
322. Misner C W, Thorne K S, Wheeler J A *Gravitation* (San Francisco: W.H. Freeman, 1973); Translation into Russian: *Gravitatsiya* (Moscow: Mir, 1977)
323. Landau L D, Lifshitz E M *The Classical Theory of Fields* (Oxford: Pergamon Press, 1975); Translated from Russian: *Teoriya Polya* (Moscow: Fizmatlit, 2003)
324. Tanabashi M et al. (Particle Data Group) *Phys. Rev. D* **98** 030001 (2018)
325. Ade P A R et al. (Planck Collab.) *Astron. Astrophys.* **571** A16 (2014); arXiv:1303.5076
326. Riess A G et al. *Astrophys. J.* **730** 119 (2011)
327. Berezhiani Z, Dolgov A D, Tkaåhev I I *Phys. Rev. D* **92** 061303(R) (2015); arXiv:1505.03644
328. Chudaykin A, Gorbunov D, Tkachev I *Phys. Rev. D* **94** 023528 (2016)
329. Alam S et al. *Mon. Not. R. Astron. Soc.* **470** 2617 (2017)
330. Sahni V, Shafieloo A, Starobinsky A A *Astrophys. J.* **793** L40 (2014); arXiv:1406.2209