Soviet Physics Uspekhi

SPECIAL ISSUE

THE COSMOLOGICAL CONSTANT AND THE THEORY OF ELEMENTARY PARTICLES

530.12 : 531.51

THE COSMOLOGICAL CONSTANT AND THE THEORY OF ELEMENTARY PARTICLES

Ya. B. ZEL'DOVICH

Institute of Applied Mathematics, USSR Academy of Sciences

Usp. Fiz. Nauk 95, 209–230 (May, 1968)

INTEREST in gravitation theory with a cosmological constant was revived in 1967. Three papers were published, by Petrosian, Salpeter, and Szekeres in the USA[1] and by Shklovskiĭ[2] and Kardashev[3] in the USSR, in which universe evolution models in such a theory (the Λ models) are considered. The stimulus for the revival of the theory was provided by new observational data on remote quasistellar sources (quasars and quasags, QSR and QSG in the English-language literature)* It turned out, first of all, that for these objects the connection between the brightness and the red shift does not fit the simple models without a cosmological constant (and without assumptions concerning the evolution of the quasars!). In addition, as noted by the Burbidges[4], in ten quasars whose spectra have revealed absorption lines the red shift of these lines $z = (\lambda - \lambda_0)/\lambda_0$ lies in the narrow range $1.94 < z < 1.96$ or even $1.945 < z < 1.955$. This phenomenon will henceforth be referred to briefly as $a = 1.95$.

The Λ models were introduced in[1] to explain the observed relation between the red shift and the brightness; the explanation of $z = 1.95$ in the absorption spectrum was touched upon casually. References 2 and 3 are devoted entirely to the explanation of $z = 1.95$: the absorption lines are ascribed to galaxies lying along the path of the light ray arriving from the quasar. The predominant appearance of one value of z is attributed by the authors to the fact that in this case the expansion of the universe was greatly slowed down both compared with the preceding period ($z > 1.95$) and compared with the succeeding period ($z < 1.95$ up to $z = 0$, corresponding to the present time). The slowed-down expansion leads to an increase of the path traversed by the ray in the corresponding interval of z, and takes into account the probability of encounter between the light ray from the quasar and the galaxy since that absorption lines with precisely this value of z, i.e., with z close to 1.95, are recorded.

An expansion law with a sharp deceleration at a definite value of z is possible only for the Λ models; it is necessary here to satisfy with great accuracy the relation between the total amount of matter in the universe and the value of the cosmological constant Λ. The discussed model is closed in its three dimensional geometrical structure. As shown by Kardashev[3], the assumption of a decelerated expansion at a definite value of z (together with the known value of the Hubble constant) yields perfectly defined values of the density of matter and of the radius of the world at the present time.

At first glance such an explanation is on the whole unlikely. It must be borne in mind, however, that other attempts at explaining the predominant absorption with $z = 1.95$ are at present no less far-fetched and artificial. In a paper at the 13th Congress of the International Astronomical Union in Prague (August 1967), Burbidge spoke of $z = 1.95$ as an argument in favor of the local theory of quasars. According to the local theory, the distance from us to the quasars is less than 100 Mpsec, and the red shift of the emission and absorption lines is of gravitational origin and is connected with the work function of the quanta from the gravitational field of the quasar[5]*. However, no concrete model which yields precisely $z = 1.95$, or at least an equal value of z for the quasars with different masses during different stages of evolution, was proposed by Burbidge or anyone else.

Thus, the predominant appearance of $z = 1.95$ in the absorption is really an argument in favor of the Λ model of the universe. At the same time, it is still impossible to regard this argument as final. The Λ model proposed in[3] raises also unanswered questions (pertaining to the formation of galaxies) and simply difficulties connected with the observed distribution of the quasars with respect to the red shift of the emission line z_{em}. This distribution does not reveal at $z_{em} = 1.95$ the concentration that could naturally be expected in the Λ model. Nor does this Λ model agree with the rather crude estimates obtained for the law of expansion in the nearest region with $z < 0.5$ from Sandage's observations of various galaxies[7]. Even the initial statement itself concerning the predominant value $z = 1.95$ for the absorption lines should be refined and verified for a large number of quasars. Thus, the question of the concrete Λ model with a perfectly defined value of Λ remains open at the

*The term Λ model will henceforth be used to designate the solution of the equations of an expanding universe, in which it is assumed that the cosmological constant is $\Lambda \neq 0$ (see Appendices II and III). Quasars are quasistellar (i.e., pointlike) radio sources, and quasags are quasistellar galaxies, similar to quasars in their optical properties and in particular having large z, but having no noticeable radio emission.

*The universally accepted assumption of the cosmological origin of the red shift, connecting it with the over-all expansion of the universe, yields $R = cz/H = 4000z$ Mpsec for nearby quasars with small z; at large values of z, the definition of R is not unique, but it is clear that we are always dealing with distances larger by tens and hundreds of times than in the local theory. A very recent sensational communication[29] concerns a report by Matthews, that an appreciable change of the optical picture of the quasar 3C-287 with $z = 1.055$ was observed over the period of one year, from 1965 through 1966. This fact is interpreted by Matthews as favoring of the local theory. The conclusion, however, is ambiguous: in accordance with calculations by Rees[6], the particles ejected with relativistic velocity by the explosion, can change of angular dimension at a rate $d\theta/dt = c\beta/R\sqrt{1-\beta^2}$, corresponding to an apparent linear velocity $v/\sqrt{1-\beta^2} > c$, $\beta = v/c$; these considerations have been further developed and analyzed recently by I. S. Shklovskiĭ. Thus, when $1-\beta \cong 10^{-5}$ we can reconcile Matthews' observations with the cosmological hypothesis concerning a large distance to the quasars.

present time, and much work must be still done for its solution.

However, we can already raise even now another question: to what extent was the assumption $\Lambda \equiv 0$, which was frequently made recently, (for example, the text book of Landau and Lifshitz[8] or the authors reviews[9] justified? In this connection, the arguments advanced were either esthetic (the theory with $\Lambda \equiv 0$ is more beautiful, simpler, the formulas are more compact, there exists a particular solution—the flat empty world of Minkowski) or else the arguments were reminiscent of the principle of economy of thought (why introduce an extra parameter Λ so long as it is not really necessary?). Once papers in which the authors (see [1-3]) are strongly interested in $\Lambda \neq$ appear, the arguments presented above lose their attraction and conviction. It turns out that many authors[10] always considered precisely the scheme with $\Lambda \neq 0$ as more beautiful, by virtue of its greater generality.

The history of the question of the cosmological constant is inseparably connected with the name of Einstein. In the first paper on the application of the general theory of relativity to cosmology, Einstein's aim was to construct a static universe with a finite average density of matter[11], and reached the conclusion that to this end it is necessary to introduce into the equations an additional term, namely the cosmological constant. Following the papers of A. A. Friedmann[12], who considered nonstationary solutions*, and particularly after the red shift was observed, Einstein wrote: "Under these circumstances, it is necessary to raise the question: Is it possible to describe the experimental facts without introducing the Λ term, which is clearly unsatisfactory from the theoretical point of view?"[13]. However, the "unsatisfactory" nature is not explained further at all. Obviously, this question must be answered on the basis of objective data.

This leads to a new formulation of the problem: What is known reliably concerning the quantity Λ? What limits can be assigned to this quantity with assurance at the present time? What kind of experiments or observations can refine the value of Λ? The genie has been let out of the bottle, and it is no longer easy to force it back in. Even if $\Lambda = 0$ exactly, it is now necessary to arrive at this answer with great difficulty, slowly, gradually, by decreasing the limits: today perhaps $-10^{-55}\ \mathrm{cm}^{-2} < \Lambda < 10^{-55}\ \mathrm{cm}^{-2}$, in ten years perhaps $-10^{-56} < \Lambda < 10^{-56}$. Even if it is shown that the value of Λ is sufficiently small and does not influence noticeably the cosmological evolution (unlike the aforementioned hypothesis[1-3]), the question still remains whether Λ actually does vanish exactly and identically. In our opinion, a new field of activity arises, namely the determination of Λ. But first let us answer the following question: how is it possible to visualize the meaning of the cosmological constant? Why is its definition interesting for physics as a whole? One approach to this quantity was already noted above, and is suggested by the dimsionality of Λ, namely cm^{-2}. This is the curvature of empty space. But the theory of g gravitation connects the curvature with energy, momentum, and pressure of matter. By transferring in the gravitation equation the terms with Λ to the right hand side, we obtain $R_{ik} - \frac{1}{2}g_{ik}R = \frac{8\pi G}{c^4}T_{ik} - g_{ik}\Lambda$. The assumption $\Lambda \neq 0$ denotes that the empty space produces the same gravitational field as when the space contains matter with mass density $\rho_\Lambda = c^2\Lambda/8\pi G$, energy density $\epsilon_\Lambda = c^2\Lambda/8\pi G$, and pressure $P_\Lambda = -\epsilon_\Lambda$. In this sense we can speak of an energy density of the vacuum and a pressure (stress tensor) of vacuum.

We note that the assumptions concerning ϵ_Λ and P_Λ were formulated in such a way that the relativistic invariance of the theory is not violated: ϵ_Λ and P_Λ are the same for all coordinate systems (Lorentz-transformed moving relative to one another). These quantities ϵ_Λ and P_Λ never appear in experiments with elementary particles, nor in atomic or molecular physics: the vacuum energy of the vessel in which the experiment is performed plays the role of the constant term that cancels out in the energy-conservation law (for details see Appendices I and II).

The only type of phenomena in which ϵ_Λ and P_Λ appear are gravitational phenomena. In this case ϵ_Λ and P_Λ "work" not only in vacuum space: as seen from the formula, they enter as full-fledged terms also in the presence of ordinary matter. This means that in principle it would be possible to determine and measure ϵ_Λ and the corresponding $\rho_\Lambda = \epsilon_\Lambda/c^2$ in the Cavendish experiment: the attraction of a lead sphere depends on the sum of the density of the lead ($11\ \mathrm{g/cm}^3$) and the density of vacuum ($|\rho_\Lambda|$ smaller than $10^{-28}\ \mathrm{g/cm}^3$) in the investigated volume.

In practice it is impossible to measure the influence of ρ_Λ and ϵ_Λ in either a laboratory experiment or even in observations of the motion of planets and the solar system or the motion of stars in the galaxy: in fact, the average density of matter in the solar system is $\bar{\rho} = 10^{-7}$ in a sphere whose radius equals the distance from the earth to the sun. The average density in the galaxy is of the order of $10^{-24}\ \mathrm{g/cm}^3$. The influence of ρ_Λ is strong only in the largest scale—in the scale of the entire universe, i.e., in cosmology. It is precisely from cosmological considerations that we can now impose the limits $|\rho_\Lambda| < 5 \times 10^{-28}\ \mathrm{g/cm}^3$, corresponding to $|\Lambda| < 10^{-54}\ \mathrm{cm}^{-2}$.

Scientific predictions are always risky; nevertheless we can propose that the concrete model[3] with $\rho_\Lambda = 5 \times 10^{-29}$ and with a halt at $z = 1.95$ will be accepted or rejected within 2–3 years. But if this model is rejected, then further progress, namely proof that $|\Lambda|$ is smaller than a certain value, calls for much more time. Actually, the same factors which limit the maximum value of the density of any type of matter[20] enter into consideration at large negative values of ρ_Λ: we can write for the time elapsed from the instant of the singularity $\rho = \infty$ to the present time the inequality $T < \sqrt{3\pi/32G|\rho_\Lambda|}$. But it is obvious that $T > 3 \times 10^9$ years, since even the geological age of the earth is of the order of 5×10^9 years. From this we get $\rho_\Lambda < 5 \times 10^{-28}\ \mathrm{g/cm}^3$.

The situation is different in the case of positive $\rho_\Lambda > 0$. Given the density of ordinary matter at the present time ρ_0 and given the rate of expansion of the

*In modern language these solutions are called self-similar: the world expands and remains similar to itself. We note that Friedmann considered equations both with $\Lambda = 0$ and $\Lambda \neq 0$.

universe (the Hubble constant H_0), large positive $\rho_\Lambda > 2(3H_0^2/8\pi G) + \rho_0$ leads to cosmological solutions in which the universe had never passed through a high-density period (the presented expression is approximate; for details see Appendix VII). The presence in the universe of thermal radiation corresponding to 3° K makes such a solution unlikely (see, for example,[9]).

In principle, exact measurement of the brightness and of the red shift of several remote bodies (not fewer than two) with exactly known absolute luminosities would be sufficient to determine both the Hubble constant and the difference $\rho_0 - 2\rho_\Lambda$, in terms of which the so-called parameter of acceleration of the cosmological model is expressed. In practice it is necessary to perform a large statistical investigation of far bodies, since their properties can be equated to the properties of nearby bodies only in the mean. This raises new kinds of difficulties connected with the need for taking into account the evolution of galaxies and quasars, i.e., the difference between the properties (average, disregarding individual fluctuations) of far objects which have emitted light long ago, and the properties of nearby objects at the present time. It is necessary furthermore to refine the magnitude of the average density of all types of matter (stars, intergalactic gas, quanta, neutrinos, gravitons) in the universe.

We now turn to a different aspect of the situation, namely to the close connection between the question of Λ and the theory of elementary particles. The very first attempts of quantizing the electromagnetic field led to the paradoxical conclusion that vacuum energy has infinite density. Vacuum was thus defined as the lowest energy state of the considered system whose characteristics are given by Maxwell's equations. The particles—in this case photons—are elementary excitations of the system. In the analogous problem of quantum theory, concerning the motion of atomic nuclei in a crystal lattice, the situation is similar: there exist elementary excitations—phonons (quanta of sound), and there exists a zero-point energy of a state in which there is not a single phonon, at absolute zero temperature, i.e., a state that can be likened to vacuum.

In the case of a crystal, the zero-point energy has a fully defined finite value and can be measured. In particular, the difference between the zero-point energies of different isotopes of the same element leads to a dependence of the heat of evaporation of the crystal on the atomic weight of the isotope. In field theory—in the simplest variant—the zero-point energy is infinite. It is, however, possible to reformulate the theory in such way that the zero-point energy of the free field is exactly equal to zero. In Maxwell's classical theory, the energy density is $(E^2 + H^2)/8\pi$, where E is the electric field and H the magnetic field. As emphasized in[14], there is no formulation of quantum electrodynamics in which the mean value $\overline{E^2}$ or $\overline{H^2}$ in vacuum vanishes (far from charges or in the absence of real quanta). Consequently, when formulating the theory (with the aid of normal products of operators*) in such a way that in vacuum we have identically $\epsilon \equiv 0$, we pay for this by losing the classical relation between ϵ and the fields.

A second source of vacuum energy arose in the electron theory developed by Dirac: the concept of filled levels with negative energy leads literally to infinite negative energy density. In this case, too, the theory was soon reformulated in such a way that ϵ was identically equal to zero for the "vacuum of noninteracting particles." This, however, did not guarantee at all that the energy of vacuum remains equal to zero when account is taken of particle interaction. A feature of modern theory is that the particle interaction comes into play not only in the presence of the real particles that should take part in the interaction.

It must be recalled first that the very term "interaction" is not used in the sense of classical physics. In school we speak of interaction between two colliding bodies, or of interaction (Coulomb) between a proton and an electron. In quantum field theory we speak, in particular, of four-fermion interaction when the neutron decays and is transformed into a proton, electron, and neutrino, or we speak of an interaction between an electron and a quantum, when the electron emits (produces) a quantum.

A free electron traveling by inertia cannot emit a real quantum that can be seen or registered far from the electron. But it can be said that a free electron emits quanta and immediately absorbs them—and this changes its properties (for example, mass, magnetic moment). The change of mass cannot be observed, since there is no experiment capable of given the mass of an electron that does not juggle with quanta. However, the change of the magnetic moment of the electron has been confirmed with all the accuracy of modern experiment. But in this very sense there can occur in vacuum the processes of creation of the triplet e^+, e^-, γ and annihilation of these particles, and many similar processes. In modern theory, the question of the state and properties of vacuum is not as simple or as obvious as in the pre-quantum times of Newton or Maxwell.

It is possible to distinguish here between several possible points of view. The first consists in the assumption that the energy of the vacuum is identically equal to zero so long as we do not take into account any fields or interactions. When these are taken into account, the energy of the vacuum is not equal to zero, but when we consider processes with real particles, the energy of vacuum enters as an additive constant. The problem of particle theory is formulated as a calculation of observable processes with real particles, and the technique of calculation should be such that the answer does not depend on the unknown or undetermined or even infinite energy of the vacuum. This is how the problem was formulated by Feynman, who successfully performed this program. During the course of the calculation there are performed, for example, the following operations: the amplitude A_{12} of the transition (particles in the first state + vacuum) $\rightarrow$ (particles in the second state vacuum) is divided by the amplitude A_{vac} of the transition (vacuum) $\rightarrow$ (vacuum), and only the ratio A_{12}/A_{vac} is a quantity that is real and pertains to the particles.

This way of getting around the question of the energy

*The definition of the normal product is given in numerous books on the theory of quantized fields ([14] and earlier).

of vacuum* is satisfactory everywhere except in the theory of gravitation! The energy density of vacuum, as already mentioned, appears in the gravitation problem as a real observable quantity that does not cancel out. In the theory of elementary particles, there exists another, so-called axiomatic trend.

It is assumed by the way of an axiom that the energy density of vacuum is identically equal to zero in accordance with the definition of the vacuum ([28], Sec. 3.1). When such a statement is proclaimed openly as one of the possibilities, there are no objections. However, frequently one encounters the statement that this assumption is necessary, and that it is the only one that agrees with the relativistic invariance of the theory. Such a statement is simply in error. We have already noted above that the characteristic relaxation between the pressure and the energy density $P_\Lambda = -\epsilon_\Lambda$ is relativistically invariant. We shall demonstate below with a concrete example how the theory of particles, given a definite choice of formulation of the theory, yields a nonzero ϵ_Λ, with relativistic invariance strictly observed.

In[28] they consider the energy and momentum of the vacuum as a whole. We could, by specifying a definite normalization volume V, speak roughly of an energy $E = V\epsilon$. The (three-dimensional) momentum p of the vacuum, obviously vanishes, since there is no preferred direction for it. The energy and momentum form a four-dimensional vector $\{E, \mathbf{p}\}$, in this case $\{E, 0\}$. Obviously, such a combination is not invariant and yields $\mathbf{p} \neq 0$ in another system of coordinates, provided we do not put $E = 0$ (meaning also $\epsilon = 0$).

The error in this reasoning lies in the fact that a definite volume was taken) thereby violating the invariance. A medium of infinite extent, and particularly vacuum, it is characterized by just an energy density, which represents T_{00}, a component of a second-rank tensor—the energy-momentum tensor. The entire tensor includes components of the type $T_{0\alpha} = T_{\alpha 0}$ (where $\alpha = 1, 2, 3$ labels the spatial axes), characterizing the energy fluxes and simultaneously the momentum density in space.

Finally, the components $T_{\alpha\beta}$ determine the stress tensor, in principle the same as in elasticity theory. In the particular case of a gas or a liquid (without account of shear stresses) $T_{\alpha\beta} = \delta_{\alpha\beta} P$.

These generally known facts are repeated here only to emphasize that the question is not whether vacuum has an energy-momentum vector, but whether vacuum has an energy-momentum-stress tensor. A relativistically-invariant vector does not exist (equals zero), but a nonzero relativistically-invariant tensor is quite possible; it has the form

$$\text{const}\cdot\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix},$$

and this is precisely the tensor referred to in the case $\Lambda \neq 0$. It cannot be excluded a priori. The following questions remain:

1) Are there any other principles by virtue of which it is necessary to put $\Lambda \neq 0$?

2) Is it necessary to regard $\Lambda \neq 0$ as a new independent world constant?

3) Is it possible to construct a likely value of Λ (at least in order of magnitude) from the known world constants?

We shall attempt below to answer just the third question, leaving the first two unanswered. If the observations confirm $\Lambda \neq 0$, then this answer (based only on dimensionality theory and on a comparison of orders of magnitude of quantities) will perhaps be useful in the construction of a genuine logically-consistent theory.

We use here the remarks of Eddington[15], Dirac[16], and other authors concerning the curious numerical relations in cosmology. At the same time, it is possible to impart these relations a new meaning, and to eliminate the contradiction with general theory of relativity. The relations of the aforementioned authors are constructed in the following manner. We take the ratio of the world's radius R to the characteristic length from the theory of elementary particles $\hbar/mc$ or the ratio of the age of the world (from the instant of the singularity) $T \cong 1/H \cong R/c$ to the characteristic time $\hbar/mc^2$. These ratios are of the order of 10^{42}. On the other hand, the dimensionless quantity characterizing the gravitational interaction is $\hbar c/Gm^2 \sim 2 \times 10^{38}$ (m—proton mass). The logarithms of the two dimensionless numbers coincide within less than 10%.* It is assumed that this agreement is not an accident. However, in an evolving world, the former ratio, which contains the radius or the age of the world, does not remain constant. It was concluded from this[16] that the second ratio $\hbar c/Gm^2$ also varies, or specifically that the gravitational constant varies in inverse proportion to the world time T. Dirac notes clearly and distinctly that the variability of G does not agree with general relativity theory (GRT), but the physical significance of the agreement between the large numbers appears to him more significant than the logical harmony of GRT.

How does the situation change with the coincidences of the large numbers in a theory with a cosmological constant, i.e., in the Λ model of the universe?

Let us write down the analogous relation, replacing the world radius R by the quantity $\Lambda^{-1/2}$, which has the same dimensionality. This yields the ratio[18]

$$\Lambda^{-1/2} : \frac{\hbar}{mc} = \frac{\hbar c}{Gm^2}, \quad \text{or} \quad \Lambda \cong \frac{G^2 m^6}{\hbar^4}; \qquad (1)$$
$$\rho_\Lambda = \frac{Gm^2}{\hbar c}\, m \left(\frac{mc}{\hbar}\right)^3 .$$

On the other hand, let us assume that the world radius is of the order of $\Lambda^{-1/2}$:

$$\frac{R}{\Lambda^{-1/2}} = n \sim 1. \qquad (2)$$

Taken together, these two premises contain the same magic feature of large numbers, which struck Dirac.

*See, for example, the good old text book [27], p. 48: "H_0 is called the zero-point energy of the field; it is infinite. . . , but as an additive constant H_0 has no physical meaning."

*It is possible to consider the ratios $e^2/Gm^2 = 1.2 \times 10^{36}$ and T/τ 3×10^{37}, where $\tau = e^2/m_0 c^3$, In this connection, Gamow[17] advanced the hypothesis that the change e and the dimensionless quantity $e^2/\hbar c$ are variables. Soon after this hypothesis was advanced, concrete objections were raised[24–26].

But let us consider these assumptions in greater detail, each separately. The first assumption (1) has the character of a "external" relation between the world constants Λ, G, m, $\hbar$, and c. In principle it could be verified by laboratory experiments and, what is most important, it agrees with the constancy of all the quantities and with GRT.

The second hypothesis has an entirely different character, and pertains to evolutional astronomy.

The world's radius R_0 during the halt time (corresponding to z = 1.95) is simply connected with $\Lambda^{-1/2}$ in the Λ model. It can be assumed that it is precisely in this period that the majority of the galaxies were produced. In order for them not to evolve too far it is necessary to have $R/R_0 = n$ not too large; in the concrete Λ model it is assumed that $n = 1 + z = 2.95$, but it is not assumed that this number is constant! During the course of the further expansion, n should increase, for example to $n = 3.3$ after 10^9 years. The relations in which the present-day R enters are treated as approximate, therefore the variability of R does not contradict the constancy of G and of other world constants.

We note, however, that the exact relation between the total amount of matter and the cosmological constant, which is necessary in order to realize a Λ model with a prolonged halt, still remains a puzzle. This puzzle has no bearing on the variability of the constants. It concerns those initial conditions with which it is necessary to supplement the equations of cosmology in order to obtain a definite solution. It is possible to search for an evolutional approach to the resolution of the puzzle.

Let us dwell also briefly on the differences between the theory with a cosmological constant $\Lambda \neq 0$ and the hypothesis[19] of the presence of a definite concentration of weakly interacting particles (neutrinos or gravitons). Such particles, by virtue of their large penetrating ability, should fill space practically uniformly and to produce an energy density ϵ_1 that does not depend on the spatial coordinates. Thereby, however, ends the similarity. The energy density ϵ_Λ, by definition, does not depend on the time, and ϵ_1 decreases like R^{-4} during the course of the expansion of the universe. The quantity ϵ_Λ corresponds to $P_\Lambda = -\epsilon_\Lambda$, but ϵ_1 corresponds to $P_1 = +\epsilon_1/3$. Thus, the particles (neutrinos, gravitons) define a definite rest system in which they move on the average chaotically. It is easy to verify that when $\Lambda = 0$, by virtue of the connection between P_1 and ϵ_1, the particles do not lead to cosmological solutions with a halt in the expansion (see the Appendix below).

Finally, and most importantly, no matter how weak the interaction between the particles and ordinary matter, in principle the presence of particles in vacuum can be observed.

Yet in the theory with the cosmological constant, $\epsilon_\Lambda \neq 0$ is ascribed precisely to the lower energy state—vacuum. Thus, the indicated two hypotheses (ϵ_Λ and ϵ_1) actually differ greatly.

In the foregoing review we considered in most general form the phenomena that call for a review of the Einstein equations with a cosmological term, as well as the questions faced by observational astronomy and theoretical physics in this connection.

In the Appendix we consider in greater detail, and with a large number of formulas, individual questions touched upon in this general exposition.

APPENDIX

I. VARIATIONAL PRINCIPLE AND GRT EQUATIONS

To take into account the cosmological term, formulas (93.1) and (95.5) of "Field Theory" by Landau and Lifshitz* should be replaced by

$$\delta S_g = -\frac{c^3}{16\pi G}\,\delta\left[\int R\sqrt{-g}\,d\Omega + \int 2\Lambda\sqrt{-g}\,d\Omega\right], \qquad (I.1)$$

$$R_{ik} - \frac{1}{2}g_{ik}R - \Lambda g_{ik} = \frac{8\pi G}{c^4}T_{ik}. \qquad (I.2)$$

In the local-Euclidean (Minkowski) metric with $g_{00} = 1$, $g_{\alpha\beta} = -\delta_{\alpha\beta}$, introduction of Λ is equivalent to an addition to the material tensor T_{ik} such that

$$T'_{ik} = T_{ik} + \frac{c^4}{8\pi G}\Lambda\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}, \qquad (I.3)$$

so that $T'_{00} = T_{00} + \epsilon_\Lambda$ and $T'_{\alpha\beta} = T_{\alpha\beta} + \delta_{\alpha\beta}P_\Lambda$, where $\epsilon_\Lambda = -P_\Lambda = c^4\Lambda/8\pi G$.

II. PROPERTIES OF ϵ_Λ AND P_Λ

We shall prove the relativistic invariance of the combination ϵ_Λ and P_Λ. For ordinary matter at rest (a liquid with isotropic Pascal pressure P and density ρ, $\epsilon = \rho c^2$, $T_{\alpha\beta} = \delta_{\alpha\beta}P$), on going over to a coordinate system in which the liquid moves with a velocity $V_x = \beta c$, we get

$$\varepsilon' = \frac{\varepsilon + \beta^2 P}{1-\beta^2}, \quad T'_{0x} = \frac{\beta(\varepsilon + P)}{\sqrt{1-\beta^2}}, \quad T'_{xx} = \frac{P + \beta^2\varepsilon}{1-\beta^2},$$
$$T'_{yy} = T'_{zz} = P, \quad T'_{0y} = T'_{0z} = T'_{xy} = T'_{xz} = T'_{yz} = 0. \qquad (II.1)$$

Substituting $P_\Lambda = -\epsilon_\Lambda$, we verify that in the new system $\epsilon'_\Lambda = \epsilon_\Lambda$, $P'_\Lambda = P_\Lambda = -\epsilon_\Lambda$, and T_{ik} is diagonal as before. The vacuum can be regarded as a "substance" with given ϵ_Λ and P_Λ also in the sense that the general relation

$$dE = -P\,dV. \qquad (II.2)$$

is satisfied. In fact, if $E = \epsilon_\Lambda V$ and $\epsilon_\Lambda = \text{const}$, then

$$dE = \varepsilon_\Lambda\,dV \equiv -P_\Lambda\,dV \text{ when } P_\Lambda = -\varepsilon_\Lambda. \qquad (II.3)$$

III. HOMOGENEOUS AND ISOTROPIC Λ MODEL. EQUATIONS

Let us consider the cosmology of the Λ model, which is determined by specifying the metric

$$ds^2 = c^2\,dt^2 - a^2(t)\,[d\chi^2 + \sin^2\chi\,(\sin^2\theta\,d\varphi^2 + d\theta^2)]; \qquad (III.1)$$

we call a the radius of the world, and the volume of the world is $V_0 = 2\pi^2a^3$. The equations for a(t) have

*I am citing the Fifth Edition (1967). The gravitational constant is denoted G in place of k, and Λ is the cosmological constant; it must not be confused with the Lagrangian of the physical system in formula (94.1).

the following form (a dot denotes differentiation with respect to time)

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right)a, \qquad (III.2)$$

$$\frac{\dot{a}^2}{2} = \frac{4\pi G}{3}\rho a^2 - \frac{c^2}{2}. \qquad (III.3)$$

The distance between any pair of points moving together with the surrounding matter (having no "random" velocities) is proportional to a, i.e., $r = r_{12} = ka$. The first equation can be regarded as a "Newtonian" equation for the gravitational action of a sphere of arbitrary radius r_{12}: point 1 is the center, and point 2 is on the surface; the matter surrounding the sphere on the outside is distributed symmetrically and therefore makes no contribution to the acceleration $\ddot{r} = -GM/r^2$. It turns out here that the role of the mass is played by $(4\pi/3)r^3[\rho + (3P/c^2)]$; the pressure also has weight. Substituting $r_{12} = ka$ and cancelling k, we obtain (III.2).

The second equation fixes the absolute value of the radius of curvature of the world a, provided we know the relative rate of expansion $\dot{a}/a = H$ (H—Hubble constant) and the density of matter. When account is taken of the cosmological constant, it is necessary to use in place of ρ and P for the matter the quantities ρ' and P', which include the density and the pressure of the vacuum, i.e., ρ_Λ and $P_\Lambda = -\rho_\Lambda c^2$. Thus, taking ρ and P to mean the density and pressure of the matter, we obtain

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2} - 2\rho_\Lambda\right)a, \qquad (III.4)$$

$$\frac{\dot{a}^2}{2} = \frac{4\pi G}{3}(\rho + \rho_\Lambda)a^2 - \frac{c^2}{2}. \qquad (III.5)$$

As is well-known, the GRT equations of the gravitational field include also the equations of motion of the matter that produces this field. The geometrical identities pertaining to the curvature of space lead to the conservation laws. It was therefore not surprising that the two GRT equations obtained from a consideration of two different components of the tensor equation $R_{in} = \frac{1}{2}g_{in}R = \kappa T_{in}$ contain also a thermodynamic identity describing the energy and the pressure.

We denote by E the energy contained in a given comoving volume V; we put $V = ha^3$. This volume, which varies in proportion to the total volume of the universe, contains a constant number of conserved particles n. It is possible to choose n such as to make V the volume per nucleon, and then E is the energy per nucleon; then

$$\rho = \frac{\varepsilon}{c^2} = \frac{E}{Vc^2} = \frac{E}{ha^3c^2}. \qquad (III.6)$$

We substitute this equation in (III, 3):

$$\frac{\dot{a}^2}{2} = \frac{4\pi G}{3c^2h}\frac{E}{a} - \frac{c^2}{2}. \qquad (III.7)$$

We take the derivative with respect to a:

$$\frac{1}{2}\frac{d\dot{a}^2}{da} = \frac{1}{2}\left(\frac{d}{dt}\dot{a}^2\right) : \frac{da}{dt} = \ddot{a} = -\frac{4\pi G}{3c^2h}\frac{E}{a^2} + \frac{4\pi G}{3c^2h}\frac{1}{a}\frac{dE}{da}. \qquad (III.8)$$

Comparing with (III.1) we get

$$\frac{dE}{da} = -3ha^2P, \quad dE = -P\,d(ha^3) = -P\,dV. \qquad (III.9)$$

Thus, the thermodynamic equation (III.3), i.e., the first law of thermodynamics, the energy conservation law, follows from two GRT equations, namely, (III.1) and (III.2). This statement can be reversed: if we specify one of the GRT equations and the energy conservation law (III.9), then the second GRT equation is obtained as a corollary. It need not be considered in explicit form. As shown in Appendix II, ρ_Λ, ϵ_Λ and P_Λ satisfy the thermodynamic equation (III.9), and therefore everything stated above is valid also in the theory with a cosmological constant.

We note, finally, that the equations are valid also for an open, hyperbolic model. If the metric is

$$dS^2 = c^2\,dt^2 - b^2(t)\,[d\chi^2 + \mathrm{sh}^2\chi\,(\sin^2\theta\,d\varphi^2 + d\theta^2)] \qquad (III.10)$$

then the GRT equations are obtained from (III.1), (III.2) or (III.3), (III.4) by replacing a with ib ($i = \sqrt{-1}$). In this case (III.1) remains unchanged:

$$\ddot{b} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2} - 2\rho_\Lambda\right)b, \qquad (III.11)$$

and the sign of c^2 in Eq. (III.3) is reversed:

$$\frac{\dot{b}^2}{2} = \frac{4\pi G}{3}(\rho + \rho_\Lambda)b^2 + \frac{c^2}{2}. \qquad (III.12)$$

IV. CLASSIFICATION OF SOLUTIONS OF THE Λ MODEL

A complete investigation and classification of the equations of cosmology with the Λ term can be found in a number of papers and reviews. We note, in particular, the article by A. L. Zel'manov[10]. However, from pedagogical considerations it is useful to present simple and intuitive considerations that make it possible to understand the qualitative properties of the solution with a minimum number of mathematical transformations.

Let us consider the most interesting case $\rho_\Lambda > 0$ and a closed (spherical) world. We take as the basis Eq. (III.5). We start with the simplest case, when the matter consists of resting non-interacting particles:

$$\rho = nm_0, \quad \varepsilon = nm_0c^2, \quad P = 0, \qquad (IV.1)$$

where n—particle density and m_0—their rest mass.

We denote by N the total number of particles in the universe, $n = N/V_0 = N/2\pi^2a^3$. Substituting in (III.5), we get

$$\frac{\dot{a}^2}{2} = \frac{4\pi G}{3}\frac{Nm_0}{2\pi^2 a} + \frac{4\pi G}{3}\rho_\Lambda a^2 - \frac{c^2}{2} = \frac{f(a)}{2} - \frac{c^2}{2}, \qquad (IV.2)$$

where

$$f(a) = \frac{\alpha N}{a} + \beta a^2; \qquad (IV.3)$$

The values of the constants α and β are clear from the foregoing. The function f(a) goes off to infinity both when $a \to 0$ and when $a \to \infty$. It has a minimum at $\alpha N/a = 2\beta a^2$; $a_m = (\alpha N/2\beta)^{1/3}$.

The character of the solution depends essentially on whether this minimum lies above or below c^2 (Fig. 1).

Regarding not only G but also Λ (and consequently also ρ_Λ, β, and α) as world constants, we are left with one parameter N, on which the situation depends. Thus, different curves on the figure correspond to different N. By increasing N, we obviously go over from the lower curves to the upper ones. Since the square of the velocity $\dot{a}^2$ is equal to $f(a) - c^2$, obviously the

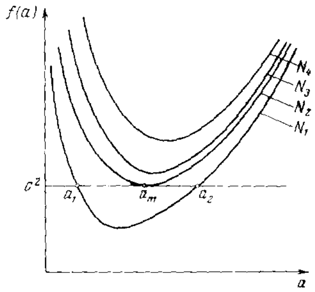

FIG. 1.

only possible values of a are those for which $f(a) > c^2$. Intersection of $f(a)$ corresponds to an instantaneous halt with a reversal of the sign of the rate of expansion (or contraction) of the universe.

Consequently, the $a(t)$ dependence for N_1 (lower curve of Fig. 1) can be of two types—Fig. 2a and Fig. 2b, comparison of a_1 and a_2 is shown on Figs. 1 and Figs. 2a, b. In this case there is no solution in which a could move smoothly from 0 to ∞. Such solutions occur at large values of N, for example, N_4 (cf. Fig. 1 and Figs. 2c, d). The square of the velocity is specified; therefore either each solution separately is symmetrical with respect to the replacement of t by $-t$ with change from compression to expansion, or else the solution describing the expansion (Fig. 2c) corresponds to another solution describing compression (Fig. 2d).

The intersection $f(a) = c^2$ at a definite angle ($f'(a) \neq 0$) corresponds to a halt, i.e., $\dot{a} = 0$, but in this case the acceleration $\ddot{a}$ does not vanish.

In the degenerate case $N = N_2$ (Fig. 1), when $f(a)$ is tangent to the horizontal c^2, it is easy to verify that $\dot{a} = 0$, $\ddot{a} = 0, \ldots$ when $a = a_m$. Thus, there exists a formal solution $a = a_m = \text{const}(t)$. In addition to this solution there are solutions that approach $a = a_m$ asymptotically from the left or from the right (Figs. 3a, b). In the solution of the type of Fig. 3b, the deviation from the stationary solution ($a = a_m$) increases exponentially with time: $a = a_m + \text{const} \cdot e^{wt}$, where $w \sim \sqrt{d^2f/da^2}$. In this sense we can speak of instability of the stationary solution with respect to small perturbations that leave the universe homogeneous and isotropic.*

Finally, the Λ model proposed by Kardashev[3] corresponds to a case close to the degenerate one (N_3 in Fig. 1). At a definite $a = a_m$ the rate of expansion, while not equal to zero, is still quite small (Fig. 4). An equation for a_m is shown in Fig. 4 near the plateau.

Let $N_3 = N_2(1 + \gamma)$, where γ is a small quantity. It is easy to see that at the critical value $N = N_2$ we have

$$\frac{\alpha N_2}{a_m} = \frac{2c^2}{3}, \qquad \beta a_m^2 = \frac{c^2}{3}. \tag{IV.4}$$

Near the critical state, for N_3, we get

$$\dot{a}^2 = \frac{2c^2}{3}\frac{a_m}{a} + \frac{2c^2}{3}\gamma\frac{a_m}{a} + \frac{c^2}{3}\frac{a^2}{a_m^2} - c^2 = \frac{2c^2}{3}\gamma + (a - a_m)^2\frac{c^2}{a_m^2}. \tag{IV.5}$$

The solution of this equation, in which we put $t = t_m$

*We disregard the question of homogeneity perturbations.

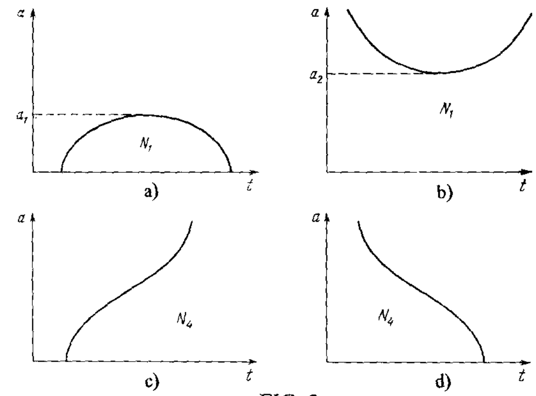

FIG. 2.

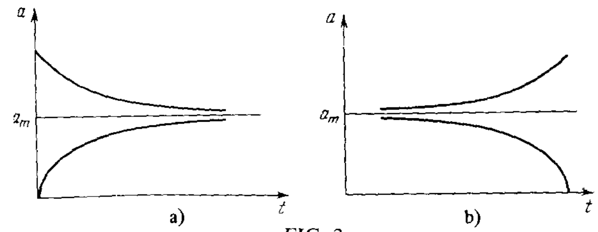

FIG. 3.

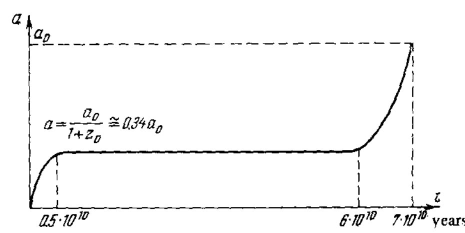

FIG. 4.

and $a = a_m$, is

$$t - t_m = \frac{a_m}{c}\,\text{arcsh}\sqrt{\frac{3}{2\gamma}}\left(\frac{a}{a_m} - 1\right) \cong \frac{a_m}{c}\ln\sqrt{\frac{6}{\gamma}}\left(\frac{a}{a_m} - 1\right), \quad \frac{a}{a_m} > 1. \tag{IV.6}$$

In order to obtain a long stay (compared with the characteristic time a_m/c) near a_m, it is necessary to choose γ exponentially small, i.e., it is necessary to have $(-\ln\gamma)$ large!

In the example considered in[3] ($a_m = 5 \times 10^9$ light years), it is assumed that the stay from $a = 0.9a_m$ to $a = 1.1a_m$ lasts 6×10^{10} years; it is necessary to have here $\gamma \cong 10^{-5}$.

It is precisely in connection with the smallness of γ, i.e., in connection with the fact that we assume that N is quite close to the critical N_2 but not equal to it, that referred to the proposed solution as a puzzle at the end of the main part of the article.

V. GEOMETRICAL PROPERTIES AND EVOLUTION OF Λ MODELS

In cosmological models with $\Lambda \cong 0$ there was a simple connection between the geometrical properties of the model (closed or open space) and its evolution.

These properties can be readily understood from the point of view of Fig. 1: Let us put $\Lambda = 0$ and $f(a) = \alpha N/a \to 0$ as $a \to \infty$, meaning that in the case of a closed world there must be such an a for which $f(a) = c^2$ and $f(a) < c^2$ when $a > a_m$; a closed world should go over from expansion to contraction in accordance with Fig. 2a. In exactly the same manner, when $\Lambda \equiv 0$ an open world must evolve monotonically to $a = \infty$, say expand without limits, in accordance with Fig. 2c.

In the presence of the cosmological term $\Lambda \neq 0$, there is no longer such a simple connection: we have seen in the preceding Appendix IV that when $\rho_\Lambda > 0$ a closed world can either evolve in accordance with Fig. 2a, or expand without limit, depending on the number of nucleons N. If $\rho_\Lambda > 0$, then the open world evolves monotonically, as in Figs. 2c, d. But if $\rho_\Lambda < 0$ (incidently, the astronomical data give no hint of such a possibility), then the expansion must give way to contraction in both open and closed worlds. Thus, when $\Lambda \neq 0$ the simple connection between the geometrical and evolutional properties of the world disappears.

In a closed model close to critical, during the time of the slow expansion, at a radius close to critical, light has time to traverse the entire universe several times. The same remote astronomical object can be seen several times. In the ideal case we shall see it from the earth in one direction on the rays traversing* the distances χ_0, $2\pi + \chi_0$, $4\pi + \chi_0, \ldots$, and in the opposite direction on the celestial sphere in rays traversing the paths $2\pi - \chi_0$, $4\pi - \chi_0, \ldots$ Different paths correspond here to different times of passage of the rays, and consequently, we shall see the same object at different ages, at different instants of its existence. For this reason, if the object glows brightly for only a small fraction of the time, we shall see the object in only one of the rays: this is probably the situation with quasars. Therefore the absence of quasar twins (visible from the opposite points on the celestial sphere) cannot be regarded as a contradiction of the closed cosmological model with decelerated expansion. The ratio of the visible brightness of the object to its absolute luminosity at the instant of the emergence of the ray does not decrease with increasing path covered by the ray. This ratio is maximal for bodies located in the "anti-center" of our galaxy, i.e., for $\chi_0 = \pi$. Petrosian and Salpeter[21] present a subtle analysis of the question of defocusing of rays as a result of the inhomogeneity of the universe, connected with the existence of separate galaxies and their clusters and with the gravitational deflection of light by these inhomogeneities.

VI. EVOLUTION WITH PASSAGE THROUGH A SINGULARITY WITH ACCUMULATION OF ENTROPY

The question of the possibility of the passage of the cosmological solution through a singular state with $\rho = \infty$ remains open at present. We assume that it is possible to join the solutions in which contraction takes place to the solutions with expansion. We can imagine that certain corrections to the GRT equations[23], which are insignificant under ordinary conditions, limit the maximum density and the minimum radius (Fig. 5). In addition to the general physical laws (baryon conservation, growth of entropy), it is necessary to have one more assumption, namely the smoothing of the inhomogeneities. This latter assumption cannot be regarded as convincing. If we disregard it, then it is necessary to solve the usual asymmetrical and inhomogeneous problem of motion, and furthermore in modified GRT equations. Bearing in mind the exceptional complexity of this problem, we shall disregard the question of inhomogeneities. It is known, that when homogeneous matter goes through a singularity at an observable entropy, matter of any composition is transformed into a standard mixture of 70–75% He + 30–25% He^4 (by weight). The question of antibaryons and of the excess of baryons near the singularity is solved in natural manner: formation of antibaryons is a consequence of the increase of the temperature upon compression of the system which initially contained only baryons. For more details concerning these questions see the 1965 review[9]. Let us turn to the picture of the evolution as a whole. We assume that the universe is closed, the total number of the baryons is smaller than the critical value, and the entropy is small. According to Appendix IV, the evolution proceeds in this case cyclically. However, the entropy increases from one cycle to the other. Taking the entropy into account, it is possible to show that

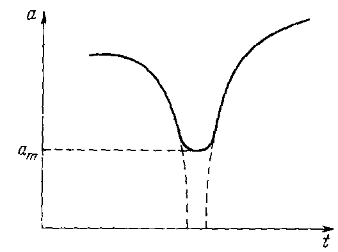

FIG. 5.

$$\rho = \frac{Nm_0}{2\pi^2 a^3} + \delta \frac{\hbar}{c} \frac{N^{4/3} S^{4/3}}{a^4}, \quad \text{(VI.1)}$$

where δ is a number on the order of unity and S is the dimensionless specific entropy per baryon. Consequently, buildup of system motion takes place with increasing S; at a definite value of S, a transition takes place from the cyclic regime to the unbounded expansion (Fig. 6).

From this point of view, the system goes over into the characteristic regime of expansion with a delay when $N < N_2$ just the same.

In the general case, however, if N is not specially close to N_2, the radiation density (corresponding to the entropy term $\delta\hbar N^{4/3} S^{4/3}/ca^4$) is small compared with the density of the baryons. The puzzle referred to above assumes the following formulation: why is the

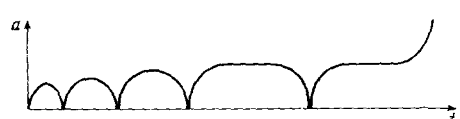

FIG. 6.

*Here the distance is given by the coordinate χ, the definition of which is given by the metric (III, 1). Near the halting point, the unit of χ is $a_m = a_0/(1 + z_0)$ units of length.

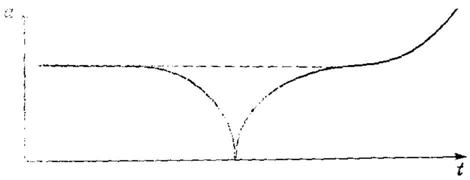

FIG. 7.

number of baryons such that the transition to the unlimited expansion occurs at a small ratio of the radiation density to the density at rest?

The unlimited increase of the radiation energy during the course of the cycles should not frighten anyone, since it does not contradict the law of energy conservation: the mass and the energy of the entire closed world as a whole are identically equal to zero! (See Landau and Lifshitz[8].).

It can be stated illustratively that the growing radiation energy is exactly offset by the growing absolute value of the negative energy of the gravitational interaction between the particles and masses of the universe. At first glance such a remark seems promising: the excess of density, necessary for a decelerated expansion in the Λ models, is precisely of the order of density of the radiation in our world. This coincidence can be understood by referring to the evolution scheme of Fig. 7.

At $t = -\infty$ we take cold baryons in the critical amount: $S = 0$, $N = N_2$. Then, formation of stars takes place in this system under the influence of the fluctuations, and nuclear reactions begin, while the total energy remains unchanged. The nuclear energy is transformed into the energy of the quanta and the neutrinos, and a nonzero pressure arises. In the language of the diagram of Fig. 8 (of the type of Fig. 1), we obtain in place of the curve 1—1 the curve 2—2, which passes through the same point, but with a finite slope. This leads to contraction of the system. A similar conclusion is arrived at also by a direct examination of the equation for $\ddot{a}$: the appearance of the pressure $P > 0$ while the density is conserved leads to $\ddot{a} < 0$. Further contraction and expansion lead to an increase of the entropy, and after the singularity the characteristic curve $f(a)$ is given by the line 3—3 of Fig. 8, which thus leads to a right-hand branch of the type shown on the top of Fig. 7 and postulated in modern Λ models.

What makes this scheme attractive is the fact that it establishes a natural connection between the density of the radiant energy in the critical state (at $z = 1.95$, $T_r = 3°(1 + 1.95) \cong 9°$, $\epsilon_r = 6 \times 10^{-11}$ erg/cm^3, $\rho_r = \epsilon_r/c^2 = 6.5 \times 10^{-32}$ g/cm^3 at $\rho_m = 5 \times 10^{-29}$ g/cm^3) and the duration of the delay or the duration of the stay near the critical state. Besides the general difficulties involved in any model that includes a transition to a singularity, this model has the following shortcomings:

1) it is necessary to have $N \equiv N_2$;

2) taking into account the finite fluctuations in any state, it is impossible to assume that the state with $N \equiv N_2$ and small S could have existed for an infinitely long time (a remark made by A. D. Sakharov). The time of development of the inhomogeneities is finite albeit large, $\sim 100\, a_m/c$.

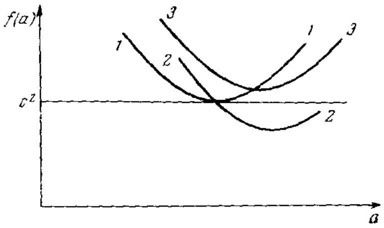

FIG. 8.

On the whole, it must be admitted that the Λ models, while resolving one difficulty of modern astronomy ($z = 1.95!$) at the same time raise new unsolved fundamental questions. The rather rough considerations advanced above, of course, cannot be regarded as any "advance in the physical (astronomical) sciences" (I am hinting at the name of the journal). They are more readily aimed at attracting the reader's attention to an unsatisfactory condition in an important branch of physics.

VII. CLASSIFICATION OF Λ MODELS IN ACCORDANCE WITH THE OBSERVABLE QUANTITIES

Observations in the vicinity of our galaxy (more accurately, the cluster or supercluster in which it is contained) make it possible to determine the Hubble constant $H = \dot{r}/r$, the acceleration parameter $q = -\ddot{r}/rH^2$, and the average density of matter $\bar{\rho}$.

The difficulties involved in the physical determination of these quantities are already indicated by the fact that the quantity H was revised many times: 500 km/sec-Mpsec (1 Mpsec $\cong 3 \times 10^{24}$ cm) in 1929, ~ 200 km/sec-Mpsec in 1950, 75 km/sec-Mpsec in 1957, ~ 100 Km/sec-Mp in 1962; at the present time it is assumed that $75 < H < 125$ in the same units. The parameter q is expressed in terms of the relative acceleration r of a remote object (at a distance r), in other words, from $v = \dot{r} = Hr$, $\ddot{r} = \dot{v} = \dot{H}r + H\dot{r} = \dot{H}r + Hv = \dot{H}r + H^2 r$ we get $q = -(\dot{H}/H^2) - 1$. We recall that when speaking of the Hubble constant, we have in mind the dependence of H on the coordinates (on the distance); this does not exclude variability of H as a function of the time. The determination of q is quite difficult. The latest published estimates by Sandage[7] give $q = +1 \pm 0.5$, but the estimate of the error can hardly be regarded as objective. As to the density $\bar{\rho}$, the part of the problem pertaining to the density of matter in the galaxies was solved by Oort in 1958. These estimates give $\bar{\rho}_g = 3 \times 10^{-31}$ g/cm^3 for the distance scale corresponding to $H = 75$ km/sec-H = 75 km/sec-Mpsec. The problem of determining the density of the intergalactic gas has come to be considered only in recent years (see, for example,[22]), but so far there is only a rough upper limit $\bar{\rho} < 3 \times 10^{-29}$. Thus, with respect to q and $\bar{\rho}$, one should seek more readily not of "observable" quantities but of "quantities principally accessible to observation." But it is precisely to demonatrate the importance of the actual performance of these observations that we shall demonstrate below the dependence of the most general properties of the universe on the quantities H, q, and $\bar{\rho}$. The results that follow pertain to homogeneous isotropic solutions with a cosmological constant $\Lambda \neq 0$. We note that the isotropy of the universe—the equivalence of all directions—was confirmed recently by measurements at centimeter wavelengths with accuracy better than 0.1%.

The isotropy of the world is indirectly confirmed also by its homogeneity: in an essentially inhomogeneous world, the radiation would be isotropic only for an observer who occupies specially (accidentally) the center of a spherically-symmetrical inhomogeneous distribution of matter. Giordano Bruno was not burned at the stake in 1600 in order that the idea that the earth (our galaxy) occupies a central position be resurrected in 1967!

Thus, we shall assume the following three quantities to be known: H, q, and $\bar{\rho}$. From $\bar{\rho}$ and H we make up the dimensionless quantity

$$\Omega=\bar{\rho}:\rho_c=\bar{\rho}:\frac{3H^2}{8\pi G}=\bar{\rho}:2\cdot 10^{-29}\ \mathrm{g/cm^3} \tag{VII.1}$$

(at H = 100 km/sec-Mpsec). All the properties of the solution depend on two dimensionless quantities Ω and q, and it is known reliably that $\Omega > 0$. We shall not consider all the possible solutions, but only those that can be candidates for a description of reality, i.e., our presently existing universe and its past and future. We stipulate here that an expansion (and not a contraction) takes place at the present time, and that the radiation density is many times smaller than the density of ordinary matter. We can then neglect, back to a very remote epoch, the pressure of matter (in particular, of neutrinos and quanta) in the equations.

We shall consider the plane of the variables q (abscissa) and Ω (ordinate) (Fig. 9). To each point in this plane there corresponds one cosmological solution satisfying the conditions formulated above. The problem consists of outlining the regions and lines characterizing the different solutions on this plane (half-plane, since $\Omega > 0$).

Each point corresponds to a definite value of Λ. The lines with constant dimensionless ratio

$$\frac{\Lambda c^2}{3H^2}=\frac{\rho_\Lambda}{\rho_c}=\lambda \tag{VII.2}$$

are* straight in the (q, Ω) plane:

$$\lambda=\frac{1}{2}\Omega-q. \tag{VII.3}$$

Several such lines (λ = −0.5, 0, + 0.5, +1) are drawn in Fig. 9. In particular, the line $\lambda = \Lambda = 0$ (the Milne model) and through the point A(Ω = 1, q = ½) corresponding to the Einstein—de Sitter flat model.

The topology of the universe as a whole also depends on q and Ω:

world ordinate and infinite

$$\Omega<\frac{2}{3}(1+q), \tag{VII.4}$$

*We derive this on the basis of (III, 5):

$$\ddot{a}=-\frac{4\pi G}{3}\left(\rho+\frac{3P}{c^2}+\rho_\Lambda+\frac{3P_\Lambda}{c^2}\right)a,$$

$$P=0,\quad \frac{P_\Lambda}{c^2}=-\rho_\Lambda,\quad -\frac{\ddot{a}}{aH^2}=q=\frac{4\pi G}{3H^2}(\rho-2\rho_\Lambda)$$

$$\frac{8\pi G}{3H^2}=\rho_c,\quad q=\frac{1}{2}\frac{\rho-2\rho_\Lambda}{\rho_c}=\frac{1}{2}\Omega-\lambda.$$

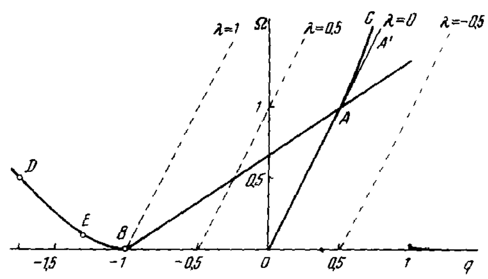

FIG. 9.

world closed and finite* at

$$\Omega>\frac{2}{3}(1+q). \tag{VII.5}$$

The separating line $\Omega = 2(1+q)/3$ passes through the point B(q = −1, Ω = 0, λ = 1) and through the point A(q = +0.5, Ω = 1, λ = 0); it corresponds to flat worlds.†

Let us return to the question of the future of the universe. It can be shown that the expansion will continue without limit in the region lying to the left of the straight-line segment OA and the line AC specified by the parametric equations ($\alpha > 1$)

$$q=\frac{\alpha^2+\alpha+1}{(\alpha-1)(2\alpha+1)},$$

$$\Omega=\frac{2\alpha^3}{(\alpha-1)^2(2\alpha+1)}. \tag{VII.6}$$

On the segment OA we deal with an open world; in such a world, the expansion must give way to contraction when $\lambda < 0$, and when $\lambda > 0$ the expansion continues without limit, as can be seen from the fact that in the equation

$$\frac{\dot{b}^2}{2}=\frac{4\pi G}{3}(\rho+\rho_\Lambda)b^2+\frac{c^2}{2} \tag{VII.7}$$

all the terms are always positive when $\lambda > 0$ and $\rho_\Lambda > 0$. The condition for the halt of the expansion of an open world coincides with the condition $\lambda > 0$, and near this boundary, but on the right of it (i.e., at $\rho_\Lambda < 0$, $\lambda < 0$, but $|\lambda| \ll 1$), the halt takes place at very large b, and in the limit on the line AC as $b \to \infty$.

A closed world with $\lambda < 0$ (region to the right of the continuation of the line OA upward, segment AA′) will be halted all the more. But a closed world can be halted also when $\lambda > 0$, if the parameters correspond to the region between AA′ and AC. An asymptotic halt occurs on the line AC, the parameter α has the meaning of the ratio of the radius of the world at the instant of the halt to the present-day radius, $\alpha = a_{\mathrm{halt}}/a_0 > 1$. Slightly above AC, at $a = \alpha a_0$, a sharp deceleration of the expansion occurs. In the entire region to the left of the line OAC the future of the universe constitutes an unbounded never-stopping expansion.

*We derive this on the basis of (III, 5):

$$\frac{\dot{a}^2}{2}=\frac{4\pi G}{3}(\rho+\rho_\Lambda)a^2-\frac{c^2}{2},\quad \frac{\dot{a}^2}{a^2}=H^2=\frac{8\pi G}{3}(\rho+\rho_\Lambda)-\frac{c^2}{a^2}<\frac{8\pi G}{3}(\rho+\rho_\Lambda),$$

$$1<\frac{8\pi G}{3H^2}(\rho+\rho_\Lambda)=\Omega+\lambda=\Omega+\frac{1}{2}\Omega-q,\quad \frac{3}{2}\Omega>1+q.$$

The preceding formula is obtained analogously from (III, 10).

*A topologically flat world is similar to a hyperbolic one.

Let us turn to the past. In the plane of Fig. 9 we can draw a line BD, the equation of which is given by the same parametric expressions (VII.6), but with $\alpha < 1$. To the right of the line BD, the cosmological solutions would evolve without stopping, starting with the singular state $\rho = \infty$.

The solutions to the left and below BD, between BD and the abscissa axis (including the solutions that expand at the present time), were contracting at $t = -\infty$ and changed over from contraction to expansion not through a singularity, but at fully defined finite values of the world radius and the maximum density*—in accordance with Fig. 2b.

Finally, the line BD itself corresponds to a universe that emerges asymptotically from the state of rest and (in accordance with the conditions that all the considered solutions must satisfy) expands at the present time with given H, ρ_1 and Λ. The parameter $\alpha < 1$ has in this case the meaning of the ratio of the radius of the world in the initial state of rest to the radius of the world at the present time.

The existence of cosmic radio emission corresponding to a temperature 3°K ("hot universe") apparently signifies that the universe was in a singular superdense state—in a state such that thermodynamic equilibrium was established at high density. This means that we can expect that the world is actually in a state above the line BD on Fig. 9 (and with respect to the abscissa axis, to the right of BD). Near the line BD (but above it) is situated a state in which a delay of the expansion took place in the past at the corresponding α. In particular, the concrete solution proposed by Kardashev[3] lies quite close to the point E:

$$\alpha = \frac{1}{1+z_{0c}} = \frac{1}{1+1.95} = 0.34, \quad \Omega = 0.105, \quad q = -1.31, \qquad \text{(VII.8)}$$

which is noted on Fig. 9.

It is also possible to indicate on the diagram other lines: the line of constant age of the universe, and the lines corresponding to single, double etc. survey of the entire closed universe within the time elapsed from the singularity. These lines condense near BD, since the line BD itself corresponds to infinite delay of the expansion. In order not to clutter up the figure, we do not show these lines.

VIII. COSMOLOGICAL CONSTANT IN ELEMENTARY PARTICLE THEORY. REGULARIZATION OF THE DENSITY AND PRESSURE

Let us take the expression for the energy density of the vacuum of scalar particles, obtained with allowance for the zero-point oscillations:

$$\varepsilon = \frac{1}{2}\frac{1}{(2\pi h)^3}\int_0^\infty c\sqrt{p^2+\mu^2}\,4\pi p^2\,dp = K\int_0^\infty \sqrt{p^2+\mu^2}\,p^2\,dp = KI(\mu), \qquad \text{(VIII.1)}$$

where $\mu = m_0 c$; the meanings of K and I follow from the formula (VIII.1). The corresponding expression for the pressure is

$$\left.\begin{aligned} T_{xx} = P = \frac{1}{2}\frac{1}{(2\pi h)^3}\int u_x p_x 4\pi p^2\,dp, \quad \mathbf{u} = \frac{c\mathbf{p}}{\sqrt{p^2+\mu^2}}, \\ \overline{u_x p_x} = \frac{1}{3}(\mathbf{up}), \quad P = K\cdot\frac{1}{3}\int\frac{p^2}{\sqrt{p^2+\mu^2}}\,p^2\,dp = K\,F(\mu). \end{aligned}\right\} \qquad \text{(VIII.2)}$$

For fermions occupying the negative-energy states (or, formally, performing anticommutation in the field-theory expression prior to the transition to the formal product), we obtain

$$\varepsilon = -4KI(\mu), \qquad P = -4KF(\mu) \qquad \text{(VIII.3)}$$

with its own value of μ. The coefficient 4 is obtained from a comparison of g = 2 for particles with spin $1/2$ and the factor $1/2$ in the zero-point energy of the bosons; this coefficient does not play a role in what follows.

Thus, from the consideration of the bosons and fermions, we obtain the expressions

$$\varepsilon = \sum_i C_i I(\mu_i), \qquad P = \sum_i C_i F(\mu_i), \qquad \text{(VIII.4)}$$

wh3re the coefficients C_i can have either sign. Generalizing further, we write

$$\varepsilon = \int f(\mu)\,I(\mu)\,d\mu, \qquad P = \int f(\mu)\,F(\mu)\,d\mu, \qquad \text{(VIII.5)}$$

where I and F are diverging integrals.

The last expressions can be regarded as a result of regularization, according to Pauli and Villars, of the expressions for ϵ and P, without considering so illustratively such individual terms as the contributions of the bosons and fermions.

We shall obtain below the conditions that must be satisfied by the regularizing function $f(\mu)$ in order that ϵ and P be finite. Inasmuch as the regularization is carried out in a relativistically-invariant manner, the result is also relativistically invariant. As noted in Appendix II, we should have here $P = -\epsilon$; indeed, it will be shown concretely that any f that gives finite values of ϵ and P satisfies this condition. To prove this, we shall consider first the finite quantities

$$I(\mu, p_0) = \int_0^{p_0}\sqrt{p^2+\mu^2}\,p^2\,dp \qquad \text{(VIII.6)}$$

and will take the limit as $p_0 \to \infty$ only at the end. We break up the integral

$$I(\mu, p_0) = \int_0^{r\mu} + \int_{r\mu}^{p_0}, \qquad r > 1, \qquad \text{(VIII.7)}$$

in order to expand in the second integral in terms of $\mu/p < 1/r < 1$:

$$\sqrt{p^2+\mu^2} = p + \frac{1}{2}\frac{\mu^2}{p} - \frac{1}{8}\frac{\mu^4}{p^3} + \ldots$$

we then obtain after integration ($C_1\mu^4 = \int_0^{r\mu}$)

$$I(\mu, p_0) = C_1\mu^4 + \frac{1}{4}p_0^4 + \frac{1}{4}\mu^2 p_0^2 - \frac{1}{8}\mu^4\ln\left(\frac{p_0}{\mu}\right) + O\left(\frac{\mu^6}{p_0^2}\right). \qquad \text{(VIII.8)}$$

Analogously we obtain

$$F(\mu, p_0) = C_2\mu^4 + \frac{1}{12}p_0^4 - \frac{1}{12}\mu^2 p_0^2 + \frac{1}{8}\mu^4\ln\frac{p_0}{\mu} + O\left(\frac{\mu^6}{p_0^2}\right). \qquad \text{(VIII.9)}$$

We substitute these expressions in the regularized

*We note that all the solutions near BD (both to the left and to the right of this line) correspond to a closed world. An open world had to emerge from a singularity in the past.

integrals

$$\varepsilon = \frac{1}{4} p_0^4 \int f(\mu)\, d\mu + \frac{1}{4} p_0^2 \int f(\mu)\, \mu^2\, d\mu$$
$$+ \left(C_1 - \frac{1}{8} \ln p_0\right) \int f(\mu)\, \mu^4\, d\mu + \frac{1}{8} \int f(\mu)\, \mu^4 \ln \mu\, d\mu + \frac{C_3}{p_0^2} + \ldots, \quad \text{(VIII.10)}$$

$$P = -\frac{1}{12} p_0^4 \int f(\mu)\, d\mu - \frac{1}{12} p_0^2 \int f(\mu)\, \mu^2\, d\mu$$
$$+ \left(C_2 + \frac{1}{8} \ln p_0\right) \int f(\mu)\, \mu^4\, d\mu - \frac{1}{8} \int f(\mu)\, \mu^4 \ln \mu\, d\mu + \frac{C_4}{p_0^4} + \ldots \quad \text{(VIII.11)}$$

Let us consider now the limits of these expressions as $p_0 \to \infty$. The conditions $\epsilon \neq \infty$ and $P \neq \infty$ are satisfied simultaneously if we impose of f the conditions

$$\int f(\mu)\, d\mu = \int f(\mu)\, \mu^2\, d\mu = \int f(\mu)\, \mu^4\, d\mu \equiv 0. \quad \text{(VIII.12)}$$

In this case the first three terms in ϵ and P drop out. On the other hand, when $p_0 \to \infty$ all the terms with p_0^{-2} and the succeeding ones drop out, too. As a result we are left with*

$$\varepsilon = +\frac{1}{8} \int f(\mu)\, \mu^4 \ln \mu\, d\mu,$$
$$P = -\frac{1}{8} \int f(\mu)\, \mu^4 \ln \mu\, d\mu = -\varepsilon, \quad \text{(VIII.13)}$$

q.e.d.

Thus, we have presented with this example a constructive proof that the field theory with relativistically-invariant regularization does not require at all a zero vacuum energy and, to the contrary, it leads naturally to the situation characterized by a cosmological constant.

IX. NUMERICAL VALUE OF Λ

The reasoning of the preceding section leads to a correct tensor form of the vacuum contribution to the energy and the pressure. However, an estimate of the order of magnitude of the obtained expression yields

$$\rho_\Lambda \sim m \left(\frac{mc}{\hbar}\right)^3 \sim 10^{17}\ \text{g/cm}^3, \quad \Lambda = 10^{-10}\ \text{cm}^{-2}. \quad \text{(IX.1)}$$

and in this case m is taken equal to the proton mass, we have left out the dimensionless factors, and the logarithms in (VIII.13) were replaced by 1.

It is clear that such an estimate has nothing in common with reality. In essence, it is just this discrepancy between the value of Λ (IX.1), which can be obtained from elementary-particle theory, and the value that is admissible from astronomical considerations, $|\Lambda| < 10^{-54}$ cm^{-2} and $|\rho_\Lambda| < 5 \times 10^{-28}$ g/cm^3, which served as the reason why many physicists assumed $\Lambda = 0$, once it became impossible to assume the value $|\Lambda| = 10^{-10}$ that follows from dimensionality considerations from the values of the constants c, $\hbar$, and m. Eddington and Dirac noted that the theory of gravitation, together with particle theory, gives a dimensionless quantity which differs very greatly from unity. Eddington introduced the ratio of the gravitational interaction of the electron with the proton to the electrostatic interaction:

$$\frac{Gm_p m_e}{r} : \frac{e^2}{r} = \frac{Gm_p m_e}{e^2} = 5 \cdot 10^{-40}. \quad \text{(IX.2)}$$

From the present-day point of view, the constants $\hbar$ and c are more fundamental than the electron charge. In addition, for uniformity, we shall take the proton mass wherever a particle mass is involved. We therefore choose as the quantity characterizing the smallness of the gravitational interaction

$$\frac{Gm_p^2}{\hbar c} = 1.8 \cdot 10^{-38}. \quad \text{(IX.3)}$$

In the note[18] there was advanced the hypothesis that

$$\rho_\Lambda \sim \frac{Gm^2}{\hbar c} m \left(\frac{mc}{\hbar}\right)^3 = \frac{Gm^6 c^2}{\hbar^4}, \quad \Lambda \sim \frac{G^2 m^6}{\hbar^4}. \quad \text{(IX.4)}$$

This quantity is still 10^7 times larger than the permissible value ($\rho_\Lambda = 2 \times 10^{-38} \times 10^{17} = 2 \times 10^{-11}$ in place of 5×10^{-28}). Numerical agreement could be obtained by replacing m_p^6 with $m_p^4 m_e^2$, or by choosing other powers and replacing $\hbar c$ with e^2; this is essentially what Dirac and Eddington did*. However, even a discrepancy of "only" 10^7 times is an accomplishment compared with the discrepancy of the estimates by a factor 10^{46}.

The expression (IX.4) can be intuitively interpreted as follows: virtual particles with mass m, the distance between which is $\lambda = \hbar/mc$, are produced in the vacuum; their self-energy is exactly equal to zero, but the gravitational interaction of neighboring particles causes the energy density of vacuum to be

$$\varepsilon_{\text{vac}} = \frac{Gm^2}{\lambda} \frac{1}{\lambda^3} = \frac{Gm^6 c^4}{\hbar^4}, \quad \text{(IX.5)}$$

corresponding to (IX.4).

Recently A. D. Sakharov proposed a gravitational theory, or more accurately, a justification of the equations of general relativity theory, based on the consideration of vacuum fluctuation[23].

An important role is played in this theory by the hypothesis that there exists a certain elementary length L or a corresponding limiting momentum $p_0 = \hbar/L$. The theory is not applicable at smaller lengths or at larger momenta. Sakharov obtains an expression for the gravitational constant G in terms of L or p_0:

$$G = \frac{c^3 L^2}{\hbar} = \frac{\hbar c^3}{p_0^2}. \quad \text{(IX.6)}$$

This expression has been known from Planck's time, but it was read "from right to left": the gravitation determines the length L and the momentum p_0. According to Sakharov, L. and p_0 are primary. We substitute (IX.6) in (IX.4) and obtain

$$\rho_\Lambda = \frac{m^6 c^5}{p_0^2 \hbar^3}, \quad \varepsilon_{\text{vac}} = \frac{m^6 c^7}{p_0^2 \hbar^3}. \quad \text{(IX.7)}$$

These are precisely the first discarded terms (as $p_0 \to \infty$) in formulas (VIII.10) and (VIII.11). Thus, we can propose the following interpretations of the cosmological constant: there exists a theory of elementary particles which would give (in accordance with a mechanism which is still undisclosed at present) and identically vanishing energy, provided that this theory were

*I. M. Khalatnikov notes that by integrating I (μ, p_0) by parts we obtain $-F(\mu, p_0)$; however, in this case the function in the upper limit is infinite, and therefore the longer procedure presented above seems to be also more convincing.

*A hypothesis is advanced in [8] that there enters also a small factor $\sim 10^{-5}$, which is characteristic of the weak interaction.

applicable without limit, up to arbitrarily large momenta; there exists a momentum p_0, beyond which the theory is not valid*; besides other consequences, a modification of the theory gives a nonzero vacuum energy; general considerations make it probable that the effect is proportional to p_0^{-2}.

A clarification of the question of the existence and magnitude of the cosmological constant will be of tremendous fundamental significance also for the theory of elementary particles.

Note added in proof. In Appendix III the author follows closely H. Bondi's book "Cosmology" (Cambridge University Press, 1961), and in Appendices IV and VII he follows the paper by R. Stabell and S. Refsdal, (Month. Not. Astron. Soci., 132, 379 (1966)).

*Understandably, p_0 enters in such a way as not to violate relativistic invariance of the theory, unlike p_0 in the formulas of Appendix VIII.

[1] V. Petrosian, E. Salpeter, and P. Szekeres, Astrophys. J. **147**, 1222 (1967).

[2] I. S. Shklovskii, Astron. tsirkulyar No. 429 (1967); Astrophys. J. **150** (1), L1 (1967).

[3] N. S. Kardashev, Astron. tsirkulyar No. 430 (1967); Astrophys. J. **150**, L135 (1967).

[4] G. R. Burbidge and E. M. Burbidge, Astrophys. J. 148 (2), L107 (1967).

[5] F. Hoyle, G. R. Burbidge, and W. L. W. Sargent, Nature **209** (No. 5025), 751 (1966).

[6] M. J. Rees, Month. Not. Roy. Astron. Soc. **135** (4), 345 (1967).

[7] A. Sandage, Astrophys. J. **133**, 355 (1961).

[8] L. D. Landau and E. M. Lifshitz, Teoriya Polya (Field Theory), Nauka, 1967 [Addison-Wesley, 1962].

[9] Ya. B. Zel'dovich, Advances Astron and Astrophys. 3, 24 (1965); see also Ya. B. Zel'dovich and I. D. Novikov, Relyativistskaya astrofizika (Relativistic Astrophysics), Nauka, 1968.

[10] A. L. Zel'manov, Proc. 6th Conf. on Cosmogony Problems, M., 1959; Physics Encyclopedic Dictionary, v. 2, art. "Cosmogony" Sov. Entsiklopediya, 1962.

[11] A. Einstein, Sitzungsber. Preuss. Akad., 142 (1917).

[12] A. A. Friedmann, Collected Works, Nauka 1966, pp. 229 and 238, or Z. Physik, **11**, 377 (1922) and **21**, 326 (1924).

[13] A. Einstein, Sitzungsber. Preuss. Akad., 235 (1931).

[14] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ. Kvantovaya mekhanika (Quantum Mechanics), v. 2, Relativistic Theory, Nauka, 1968.

[15] A. Eddington, Mathematical Theory of Relativity, Cambridge Univ. Press.

[16] P. A. M. Dirac, Nature **139**, 323 (1937); Proc. Roy. Soc. **A165**, 198 (1938).

[17] G. Gamow, Phys. Rev. Lett. **19**, 759 (1967).

[18] Ya. B. Zel'dovich, ZhETF Pis. Red. **6**, 922 (1967) [JETP Lett. **6**, 1345 (1967)].

[19] K. P. Stanyukovich, Gravitatsionnoe pole i élementarnye chastitsy (Gravitational Field and Elementary Particles), M., Nauka, 1965:

[20] Ya. B. Zel'dovich and Ya. A. Smorodinskiĭ, Zh. Eksp. Teor. Fiz. **41**, 907 (1961) [Sov. Phys.-JETP **14**, 647 (1962)].

[21] V. Petrosian, and E. Salpeter, Preprint, 1967.

[22] V. G. Kurt and R. A. Syunyaev, Kosm. issled. **5**, 573 (1967).

[23] A. D. Sakharov, Dokl. Akad. Nauk SSSR **170**, 70 (1967). [sic!]

[24] F. T. Dyson, Phys. Rev. Lett. **19**, 129 (1967).

[25] A. Peres, Phys. Rev. Lett. **19**, 1293 (1967).

[26] T. N. Bahcall and M. Schmidt, Phys. Rev. Lett. **19**, 1294 (1967).

[27] G. Wentzel, Quantum Theory of Fields, Interscience,

[28] R. F. Streater and A. S. Wightman, PCT, Spin, Statistics, and All That, Benjamin, 1964.

[29] Science News, 23 December 1967.

Translated by J. G. Adashko