PHYSICS TODAY VOL 20 NO 11

Relativistic Cosmology

Questions asked by cosmologists today concern the large-scale structure of the universe. How do intergalactic distances expand? Is the universe isotropic and homogeneous? Flat or curved? Finite or infinite? Did it all start with a "big bang?"

by Wolfgang Rindler

Some of the most exciting recent developments in physics have been in the realms of the very small and the very large. Chief among the latter was the discovery of "quasars,"[1] which has quadrupled our depth penetration of the universe (*if* quasars are where most astronomers think they are), and which has also stimulated the possibly irrelevant but intrinsically fruitful investigations into gravitational collapse. Then came the discovery of the 3°K background radiation,[2] which seems to have taken us back a long way in time to an apparent "big-bang" origin of the universe. Both bode ill for the steady-state theory to which many of us had been attracted. Most recently, Robert H. Dicke's speculations and observations[3] on the oblate sun have battered at the edifice of general relativity, and thus at the very foundations of modern cosmology. Even advances in the realm of the very small affect our knowledge of the very large; for, after all, most of the phenomena we see in the sky are manifestations of either gravitational or nuclear activity.

In consequence of all this there has been a marked return of interest in the whole field of cosmology, as could be observed, for example, at recent American Physical Society meetings as well as at the series of Texas Symposia on Relativistic Astrophysics that were begun in 1963. There may not be many more answers today than there were five years ago, but there certainly seem to be more questions and more activity. It is agreed that we need more researchers, more big optical telescopes (as has been eloquently pleaded by Margaret Burbidge, Engelbert Schücking, and others) and early extraterrestrial observatories. This may therefore be a good moment to pause and take stock once more of just where we stand in cosmology.

The problems

Cosmology, of course, is the study of the large-scale structure of the universe. Its "fundamental particles" are the galaxies. Its concern is with such questions as how intergalactic distances expand; whether the universe is isotropic and homogeneous or not, flat or curved, finite or infinite, bounded or unbounded; whether it started with a big bang or oscillates or is in a steady state or whether perhaps it has a "hierarchical" structure of clusters, superclusters, supersuperclusters and so on ad infinitum; beyond this there are the deep problems of cosmogony, such as the origin of the elements and the origin and evolution of stars, galaxies and clusters.

The tools

The traditional tools of cosmology are telescopes—earth-bound optical and radio telescopes at present, but x-ray, neutrino and extraterrestrial telescopes perhaps in the future. In addition, some quite simple nontelescopic observations seem to have profound cosmological implications: The darkness of the night sky points to an expanding universe[4] as does, perhaps, the thermodynamic "arrow of time."[5] Again, the apparent local isotropy of inertia seems to imply, through Mach's principle, the isotropy of the whole universe. (This implication has been corroborated by the telescopic evidence, and, to an apparently much higher degree of accuracy—about 1%—by the isotropy of the 3°K radiation.[6]) Finally, when it comes to the details of cosmogony,

---

Wolfgang Rindler is a professor in the division of mathematics and mathematical physics at the Southwest Center for Advanced Studies in Dallas, Texas. He is author of one of the standard texts on special relativity, a well known paper on cosmological horizons, and numerous others on relativity and cosmology.

much of "ordinary" physics is, of course, highly relevant.

The facts

To fix our ideas, here are some of the relevant facts. Stars are distributed very sparsely within galaxies. In a model in which stars are represented by pinheads, these would be 50 km apart, and the solar system would be a 20-meter circle centered on a pinhead sun. A typical galaxy contains 10^{11} stars, has a radius of 3×10^4 light years and is 3×10^6 light years from its nearest neighbor. Like a dime, it has a width only about a tenth of its radius, and dimes, spaced about a meter apart make a good model of the galactic distribution. On this scale, dimes 1.5 km away would represent the farthest visible galaxies, and objects 6 km away would represent the farthest known quasars. However, relatively few galaxies are single. Most belong to clusters of from 2 to 1000, to even 10 000, apparently bound by gravity. The density of visible matter in the universe is judged to be about 10^{-30} gm/cm^3, and some galaxies appear to be about 10^{10} years old.

Cosmological principle

Perhaps the most basic task of theoretical cosmology, and that to which the present report is almost entirely devoted, is to construct a kinematic model of the universe, that is, to establish the space and motion pattern of the aggregate of galaxies. That there *is* motion, and in fact expansion, has been almost universally accepted since Edwin P. Hubble's discovery of the red shift in the optical spectra of distant galaxies.

As a simplicity postulate, supported to some extent by observation, it is generally assumed that the universe is homogeneous and isotropic (*cosmological principle*). Homogeneity, in A. Geoffrey Walker's formulation, means that what one can learn from the totality of observations of the universe in one galaxy can be learned in any other galaxy. It also means that all "sufficiently large" spatial regions of the universe are equivalent. The *fundamental particles* of cosmology, whose homogeneous distribution is thus assumed, used to be identified with galaxies but have more recently been identified with clusters of galaxies. It really makes no difference, provided there is *some* stage at which homogeneity begins. (A hierarchical universe is thereby excluded, although Gerard H. de Vaucouleurs,[7] among others, has urged its further consideration in the light of evidence for the existence of superclusters.) Isotropy, of course, means that there are no systematically preferred directions at any galaxy. (And this excludes, for example, "island universes," whose outermost galaxies are clearly not centers of spherical symmetry.)

Walker[8] has established a number of theorems that make it appear possible that homogeneity and global isotropy are actually theoretical consequences of *local* isotropy everywhere. A much weaker assumption, equivalent to the cosmological principle, would then be: Each particle always sees an isotropic distribution of particles in its neighborhood. But repeated assertions in the literature aside, this equivalence has not been proved.

Robertson-Walker model

An important theoretical discovery was made in 1935: namely that the cosmological principle (applied to the basic congruence of galaxy world lines, with uniquely connecting finite-speed light paths) implies a kinematic model of the universe that is unique apart from an arbitrary *expansion function* $R(t)$ and a *curvature index* k that can be $+1$, 0 or -1. This is the Robertson-Walker (R-W) model, named after its independent discoverers. It was, of course, Einstein's general relativity that first led investigators to contemplate curved and expanding world models and that served as the original framework of modern cosmology. But the significance of the R-W model is that it is *independent* of general relativity and applies equally to all cosmological theories that accept the cosmological principle—for example, the steady-state theory or the "pseudo-Newtonian" theory.

The four main features of the R-W model are the following:

1. Since occurrences in the neighborhood of one particle are duplicated, by hypothesis, at all other particles, the changing aspect of the universe at each particle (for example, the density) acts as a clock and defines a sequence of simultaneities, called *cosmic time* t. The calibration of t can be achieved by some microscopic phenomenon like the vibrations of a cesium atom. (In steady-state or static theories, this procedure for setting up a cosmic time must be modified.)

2. Existence of cosmic time allows construction of cosmic space sections. Each of these is a *composite* or patchwork of *local* space maps, all made at the same cosmic instant t. It is found that each such section must be a three-dimensional Riemannian space of constant curvature, which apart from magnitude can be positive, zero, or negative, corresponding to $k = +1$, 0 or -1; and for topological reasons this curvature index cannot change with time. Suppressing one spatial dimension, we can illustrate (see figure 1) the three possible types of section by a sphere, an infinite plane and (locally) a saddle. Particles (galaxies or clusters of galaxies) correspond to points distributed homogeneously over the sections, and are indicated by dots.

3. The mode of expansion (or possibly contraction) of the sections is determined by a function $R(t)$. In the cases $k = \pm 1$, $R(t)$ is chosen so that $k/R^2(t)$ is the Gaussian curvature of the section at cosmic time t. In all cases the distance between fixed particles is a constant times $R(t)$. Theories in which $R(t)$ becomes zero in the finite past are called big-bang theories. It may seem that in such theories the early universe necessarily has small total volume. But this is not so: R-W models with $R(0) = 0$ and with $k = 0$ or -1 have infinite space sections at all finite cosmic times, unless topological identifications are made. Such models become pretty unrealistic when pushed back far enough in time. Even when R is so small that all the galaxies touch, the total volume, no matter how it is measured, must be infinite since there are infinitely many galaxies. Hence the big bang is infinitely extended. Only for "point" galaxies can the big bang then be localized in some sense. For example, E. A. Milne's universe is simply a shower of nongravitating point galaxies shot out isotropically from a point-like big bang in the global space-time of special relativity. All uniform speeds short of the speed of light are attained by these galaxies. There are no "outermost" galaxies: beyond each there are others. Although

in this description the model is insular, nevertheless if the galaxies are suitably distributed, Milne's universe is equivalent to an R-W model with $R(t) = ct$, and $k = -1$ (!). Such is the effect of using cosmic time.

4. Given that light has a finite speed, the hypotheses imply that light propagates along geodesics in the space sections, and at constant local speed. In figure 1 these geodesics correspond to great circles on the sphere and straight lines in the plane. For example, in the case of $k = 1$, one can usefully think of photons as bugs crawling along great circles over the material of a rubber balloon on which the particles are permanently marked by ink dots; the balloon is inflated or deflated in accordance with a prescribed function $R(t)$, but independently of $R(t)$ the bugs always crawl at constant speed over the surface.

Cosmological horizons

We can use this model to illustrate briefly two of the *horizon* concepts in cosmology.[9] For definiteness we consider a universe of positive curvature though the argument applies equally in all three cases. In the first diagram of figure 1 I have marked our own galaxy and a photon on its way to us along a geodesic. It can happen that "the balloon is blown up" at such a rate that this photon never gets to us. As Sir Arthur Eddington has put it, light is then like a runner on an expanding track, with the winning post (us) receding forever from him. In such a case there will be two classes of photons on every geodesic through us: those that reach us at a finite time and those that do not. They are separated by the aggregate of photons that reach us exactly at $t = \infty$—shown in the diagram as a dashed circle—but in the full model these photons constitute a spherical light front converging on us. This light front is called our *"event horizon"* and its existence and motion depend on the form of $R(t)$. Events occurring beyond this horizon are forever beyond our possible powers of observation (that is, if we remain on our own galaxy). It is sometimes loosely said that at the horizon galaxies stream away from us at the speed of light, in violation of special relativity. But it must be remembered that special relativity need not apply on the cosmological scale and that we and our horizon are certainly not contained in a common inertial frame.

The same diagram can also be made to illustrate the concept of a *particle horizon*. Suppose the very first photons emitted by our own galaxy at a big-bang creation event are still around, and now let the dashed circle in the diagram represent their present position. As this light front moves outward over more and more galaxies, these galaxies see us for the very first time. By symmetry, however, at the very cosmic instant when a galaxy sees *us* for the first time, we see *it* for the first time. Hence at any cosmic instant this light front, called the "particle horizon," divides all galaxies into two classes relative to us: those already in our view and all others.

Red shift

As another simple use of the model we can deduce the red-shift formula

$$1 + z = R(t_0)/R(t_1) \qquad (1)$$

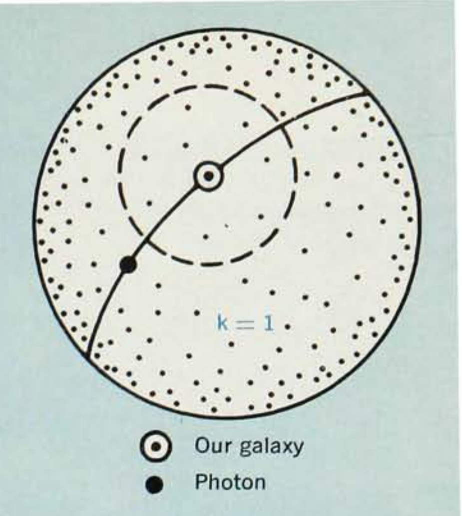
where $z = \Delta\lambda/\lambda$, λ being the wavelength of light emitted by a distant galaxy at cosmic time t_1 and received by us at t_0 with wavelength $\lambda + \Delta\lambda$. If two closely successive "bugs" crawl over a nonexpanding track, they arrive as far apart as when they left. But if the track expands—or contracts—proportionally to $R(t)$, then their distances apart at reception and emission will be in the ratio $R(t_0)/R(t_1)$. Equating the distance between the bugs with a wavelength, we get equation 1! Note that the cosmological red shift is really an *expansion* effect rather than a *velocity* effect.

Narrowing the model

We next turn to the question of how to determine the form of the expansion function $R(t)$ and the value of the curvature index k applicable to our own universe. Observations can go a certain way toward this goal. For this purpose one must develop from the model some relations between directly observable quantities. Consider, first, Hubble's discovery that

$$z \approx Hx \qquad (2)$$

where z is the red shift, H is Hubble's parameter (which must be allowed to be a function of t) and x is distance.



THE POSSIBLE UNIVERSES that are isotropic, with one spatial dimension suppressed. Dots are "fundamental particles" (galaxies). Top drawing illustrates the horizon concept. —FIG. 1

In classical optics the Doppler effect at velocity v is given by $z = v/c$ or

$$z = v \qquad (3)$$

if we choose units such that $c = 1$ as we shall in the sequel. From this, Hubble deduced the now familiar velocity-distance proportionality $v \approx Hx$, which is easily understood from the balloon-type expansion of the R-W model. But the classical equation 3 is only a first approximation to the cosmological equation 1. To compare equation 1 with observations, we first expand it in powers of the "light distance" $\tau = t_0 - t_1$, and then, again using the model, we convert τ to the "uncorrected luminosity distance," which is defined as the square root of the intrinsic brightness B of a galaxy at emission divided by its apparent brightness b as observed here and now. In this way we find

$$z = \frac{\dot{R}}{R}\left(\frac{B}{b}\right)^{1/2} - \frac{1}{2}\left(\frac{\dot{R}^2}{R^2} + \frac{\ddot{R}}{R}\right)\left(\frac{B}{b}\right) + \frac{1}{6}\left(4\frac{\dot{R}^3}{R^3} + 7\frac{\dot{R}\ddot{R}}{R^2} + \frac{\dddot{R}}{R} + \frac{k\dot{R}}{R^3}\right)\left(\frac{B}{b}\right)^{3/2} + \dots \qquad (4)$$

where R and its derivatives are evaluated at t_0. Since for nearby galaxies all distance definitions become equivalent, comparison of equations 2 and 4 shows

$$H = \dot{R}/R \qquad (5)$$

The present value of this ratio, as well as of the ratios $\ddot{R}/R$, $\dddot{R}/R$, etc., and of k/R^2 (and thus of R if $k \neq 0$) can, in principle, be found by comparing the theoretical relation 4 with an accurate observational z–b curve. In practice, however, only the first two terms in the expansion of z have yielded any information so far: H is about 100 km/sec/Mpc (Mpc = megaparsec) and the dimensionless *deceleration parameter* q defined by

$$q = -R\ddot{R}/\dot{R}^2 = -\ddot{R}/RH^2 \qquad (6)$$

seems to be about unity. The latter would indicate that the universe is decelerating.

Difficulties

On the other hand, formula 4 is usually compared with observation on the assumption that all observed galaxies have the same intrinsic brightness at emission, an assumption that is questionable when we consider that by present observations we see neighboring galaxies essentially *now* but distant galaxies as they were some 10^9 years ago. Any systematic evolution of galaxies could invalidate equation 4 with B constant beyond the first term. In a recent very careful study, using various plausible models of galactic evolution, Beatrice Tinsley[10] concluded that the value of q as defined by equation 6 is almost totally unrestricted by the *observed z-b* curve. On the other hand, Alan Sandage is known to disagree with this.

Formula 4 also ignores all possible intergalactic absorption of light, another questionable assumption. Another difficulty is the "k" correction": b is calculated on the assumption that the total light energy received is actually measured, but photographic plates are more sensitive to blue than to red and the Doppler shift moves different parts of the galactic spectra into the photographic range. If the galactic spectra were accurately known, due allowance could be made for this effect. But the earth's atmosphere scatters ultraviolet light, thus not only adding another possible error but also preventing us from having complete knowledge of galactic spectra. The introduction of extraterrestrial observatories will help this situation as will the use of modern photographic emulsions with greater sensitivity in the red.

The z–b relation discussed above is by no means the only theoretical relation between observables. Among others I will only mention the N–m and N–z relations (numbers of particles in a given area of sky with apparent magnitude greater than m or red shift less than z) and the D–m and D–z relations (angular diameters of galaxies, or clusters of galaxies, against their apparent magnitude or red shift). The above-mentioned difficulties of the unknown galactic evolution pattern and intergalactic absorption would seem to be absent, for example, in the N–z relation,

$$N \propto \frac{1}{3}\frac{R^3}{\dot{R}^3}z^3 - \frac{1}{2}\left(\frac{R^3}{\dot{R}^3} - \frac{\ddot{R}R^4}{R^5}\right)z^4 + \dots$$

from which we might hope to deduce q. But, unfortunately, for faint galaxies it is hard to determine z; moreover, there are unknown statistical fluctuations in the brightnesses of galaxies of the *same* age (and therefore presumably at the same z), and it is the bright ones that tend to be counted and the faint ones that tend to be missed. This selective *Scott effect*[11] again vitiates the determination of q. None of the other possible relations between observables are at present exempt from similar difficulties. For example, in the D–z relation one of the difficulties in the case of galaxies is to observe their apparent diameters accurately, and in the case of galactic clusters we are again faced with an unknown factor, namely the dynamic evolution of clusters. Although Howard P. Robertson[12] has suggested how the N–m relation might be used to yield a μ term (combined effect of galactic evolution and intergalactic absorption) that could then be substituted in the z–m relation to get q, this procedure has not proved practicable.

The age problem

It is worth discussing briefly one consequence of a possible positive q, that is, of a decelerating universe. If the universe had been expanding linearly at its present rate, its present age would be $T_0 = R_0/\dot{R}_0 = 1/H_0$ (see figure 2) where the suffix zero here and in the sequel indicates present values. In general relativity it so happens that an R-W model whose expansion is declerating now has been decelerating always (see equation 7 below), and this would imply that the universe is younger than T_0. The present estimates of H_0 actually vary between about 60 and 100 km/sec/Mpc,[13] corresponding to $T_0 = 16.2$ and 9.7×10^9 years, respectively. Thus the high estimates of H_0 in conjunction with $q_0 > 0$ lead to conflicts with some currently estimated galactic ages of about 12×10^9 years,[13] and these conflicts can be referred to as the *"age problem."* No such problem exists with the recent low estimates both of H_0 and of the galactic ages.[13] The uncertainty in H_0 is, of course, simply a reflection of our uncertainty about the absolute cosmic distance scale.

Theory to the rescue

So far, then, the observations have yielded only the present value of $\dot{R}/R$ to any degree of certainty. Potentially, they can at best provide us with the present values of R and the first few of its derivatives, and with the sign of k. Thus in any case, we need other methods to restrict the model further. This can be done either by an additional symmetry postulate (as in the steady-state theory) or by a theory of gravitation, that is, a dynamics that restricts the kinematics.

Steady-state theory

The additional symmetry postulated by the steady-state theory is the so-called *"perfect cosmological principle,"* which asserts that every particle always sees the same density and isotropic motion pattern. Hence, in particular, H is a constant, and it is nonzero if we accept that there *is* motion. Then, from equation 5, $R \propto \exp(Ht)$; and, furthermore, since the Gaussian curvature k/R^2 of the space sections must be constant, k can only be zero. This determines the model uniquely. Of course in this theory there must be continual creation of matter to keep the density constant. For a while the steady-state theory enjoyed great popularity, but a number of observational discrepancies lately led to its decline.

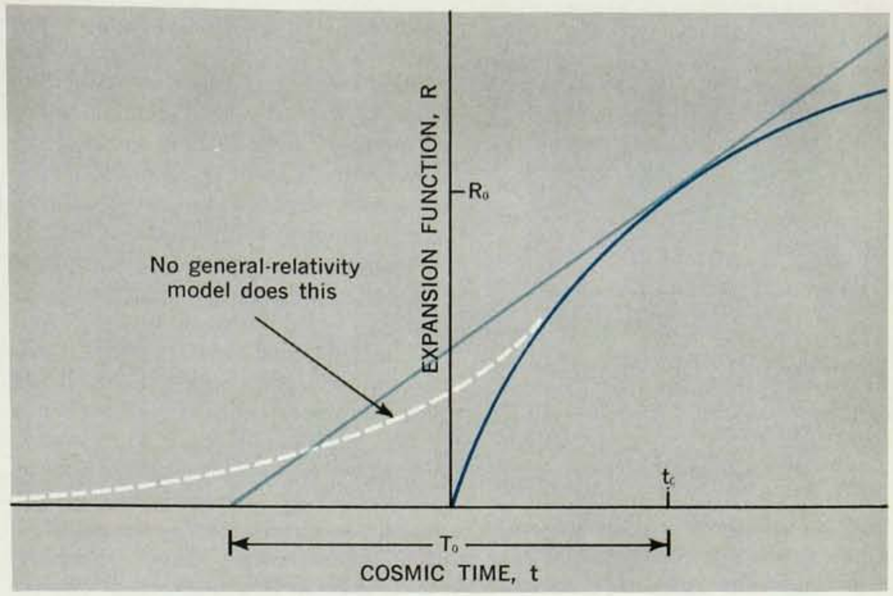
The first of these was the apparently positive value of q_0 obtained from the observed z–m curve, whereas $R \propto \exp(Ht)$ implies $q = -1$. Unlike other theories, the steady-state theory cannot appeal to galactic evolution to disclaim formula 4. Nevertheless E. L. Scott showed that the selective tendency to observe brighter rather than fainter galaxies[11] could invalidate the q objection. (Intergalactic absorption, on the other hand, would make it worse.) A second difficulty came from the number counts of radio sources, which yielded a steeper slope of the log N–log b curve (b is apparent brightness) than can be explained by *any* theory in which sources do not systematically evolve. The situation could still be saved, however, by arguing that some of the counted sources were not galaxies.[14]

A third difficulty arose from the discovery of the 3°K radiation already mentioned, which is most readily explainable as the result of a primordial big bang. However, since the characteristic maximum of the Planck distribution curve of this radiation has not yet been established (all the presently known six points are on the ascending limb[6]) other explanations still appear possible.[15] The most recent counter evidence comes from the N–z observations for quasars. These show so much spread that at first sight no conclusion seems possible. Yet this very spread was ingeniously exploited by Dennis W. Sciama and Martin J. Rees,[16] who based their argument on the fact that according to the steady-state theory the overall quasar distribution must be constant.

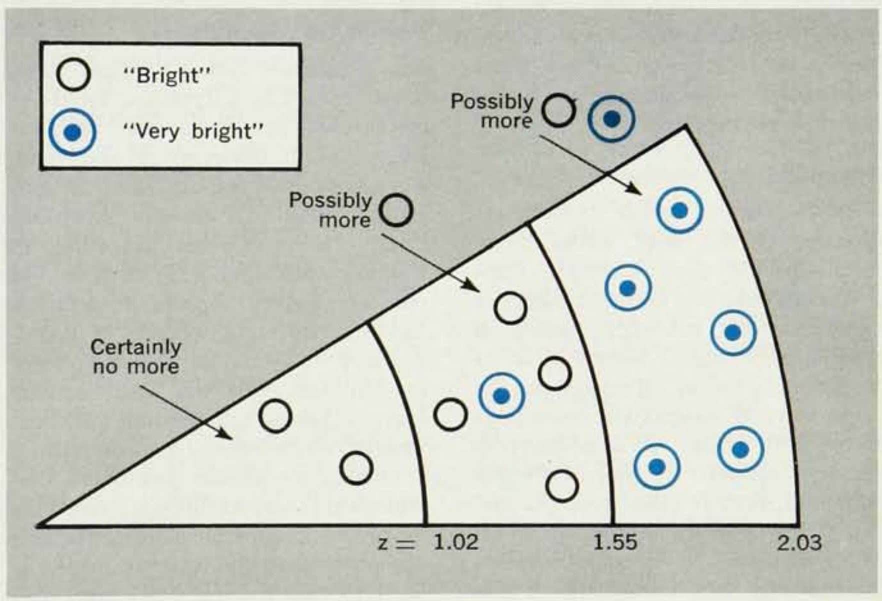
They divided the portion of the sky covered by the revised 3C catalog into three depth zones, which, according to the theory, should contain equal numbers of sources of any given intrinsic brightness (see figure 3). The authors consider quasars in two brightness ranges that we shall simply call "bright" and "very bright." They can utilize only the 12 intrinsically brightest quasars among the 35 whose red shifts are listed in the catalog. In zone I are listed only two bright quasars, and since they are bright and "near," none could have been missed.

In zone II there is one very bright quasar (again, none could have been missed) and four or more bright ones —more, because the arbitrary Cambridge cutoff at nine flux units can already affect bright quasars in this zone.

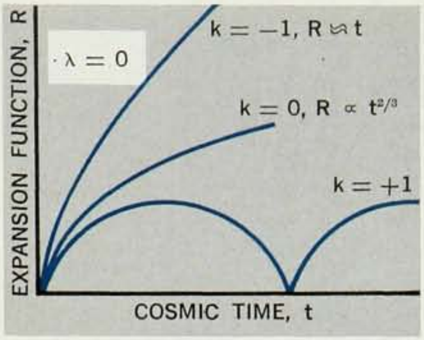
In zone III there are listed five very bright ones, but because of the cutoff there may be more of these as well as of the bright ones. Thus the three



DECELERATING RELATIVISTIC UNIVERSE, has age t_0 less than the reciprocal T_0 of the Hubble parameter. —FIG. 2



"SIDE VIEW" OF SOLID ANGLE OF SKY covered by the revised 3C catalog. According to steady-state cosmology there should be equal numbers of sources of given power in each depth zone. —FIG. 3



FRIEDMAN MODELS shown above are the three possible with $\lambda = 0$ and with $\rho \neq 0$. —FIG. 4

zones do *not* contain equal numbers of bright quasars, or of very bright ones, or of bright *and* very bright ones; on the contrary, the more distant zones appear to contain more quasars. And this discrepancy appears to be statistically significant. Of course the crucial assumption here is that quasars are indeed as far from us as their red shifts indicate. But astronomers are by no means agreed on this.[17]

Dynamics for cosmology

Next we turn to possible dynamic restrictions of the R-W model. Undoubtedly the most accepted, complete, consistent and aesthetic theory of gravitation for cosmology today is general relativity. Nevertheless, a "pseudo-Newtonian" theory[18] can be made to yield comparable restrictions on the R-W model. Its chief utility is that it translates the relativistic equations into more familiar concepts and that it serves to discuss certain problems that would be too complicated in the full relativistic formalism. It also yields a satisfactory dynamics for the steady-state theory.[19] At the other end of the spectrum, certain post-Einsteinian relativistic theories (for example, Pascual Jordan's and Dicke's[3]) have also been used in cosmology.

Although in this article I shall accept general relativity, one should keep an open mind, remembering that *no* physical theory is sacrosanct. A theory is simply a model of nature, and this model, like any other analogy, may have its limitations. Thus over the vast time spans involved in cosmology, general relativity may not be fully correct. It is based, after all, on the strong assumption (equivalence principle) that what goes on inside a sufficiently small, freely falling, nonrotating box is *totally* independent of the rest of the universe. One does not have to be a Machian to envisage the possibility that in a very dense universe the physics in the box might well be different from the usual, at least in its numerical content—for example, c and the constant of gravitation G might be different. Jordan's and Dicke's theories envisage such possibilities though at present they are hardly forced on us. Schücking[20] has truly said about general relativity, "Imagine the analogous situation in electromagnetic theory in which people should accept Maxwell's equations before Volta, Oersted, Ampère, Faraday and Galvani had done their experiments, magnetism had not yet been observed in one single instance and there was only Coulomb's $1/r^2$ law!"

Pseudo-Newtonian theory

The difficulties facing classical Newtonian theory as a dynamics for R-W-model universes deal mainly with existence of absolute space and time. For example, if the expansion is nonuniform, and two galaxies accelerate relative to each other, according to Newton, *both* can not be inertial. Yet our own galaxy appears to define a local inertial frame, whence, by homogeneity, we would expect *all* galaxies to coincide with the local inertial frame. Taking into account such difficulties, cosmologists have constructed a satisfactory *pseudo*-Newtonian cosmological theory as follows: (a) One drops the idea of absolute space and accepts instead the relativistic idea of local inertial frames. (b) One uses the inverse-square law only in its local form, that is, Poisson's equation. (c) One assumes $k = 0$ in the R-W model since Newton's is essentially a flat-space theory. To these assumptions one can add a space expansion ("λ-term") and the equivalence of mass and energy so as to parallel the results of general relativity even more closely.

Consider now any two specific galaxies; let their distance apart as a function of cosmic time be $R(t)$, and let this be the choice of expansion function in the R-W model. As a consequence of Poisson's equation, the second galaxy is attracted to the first by a force proportional to the mass M *inside* a sphere of radius R centered on the first galaxy, and inversely proportional to R^2:

$$\ddot{R} = -\frac{GM}{R^2} + \frac{\lambda c^2 R}{3} - \frac{4\pi G p R}{c^2} \qquad (7)$$

The last two terms are optional; they are certainly extremely small. (In this equation only, we have retained c so as better to indicate magnitudes.) The λ term (λ is the cosmological constant) represents an extra repulsion directly proportional to distance but totally independent of mass! It can be regarded as a "space-expansion" term, not caused by the cosmic matter and thus philosophically somewhat suspect. The p term (p is pressure) is due to the gravitational attraction of the mass equivalent of the pressure energy inside the sphere ($\propto pR^3/R^2$). Note that $M = 4\pi R^3 \rho/3$ is a constant (ρ is the density) since the number of galaxies inside the possibly expanding or contracting sphere remains constant. Multiplying equation 7 (on the assumption $p = 0$) by $2\dot{R}$ and integrating, we find

$$\dot{R}^2 = \frac{C}{R} + \frac{\lambda R^2}{3} - k \qquad (8)$$

where we have written C for $2GM$, and where k appears simply as a constant of integration. If $k \neq 0$, we can assure $k = \pm 1$ by a scale change in $R(t)$. Equation 8 is then formally identical with Friedman's differential equation, which governs the expansion of the model in the full relativistic theory with $p = 0$, except that there the constant k is the curvature index of the model. I will discuss later how this equation restricts the model.

General relativity

But let us first turn to general relativistic dynamics. Since the equations of motion in general relativity are geometric (they specify geodesics in space-time), the field equations must link the geometry to the matter content of the universe. For simplicity, cosmological matter is usually regarded as a pressureless, strainless, uniform "dust" distribution fully specified by its average density $\rho(t)$. This assumption may have to be modified during the early stages of a big-bang universe, but quite soon afterwards it could be essentially correct. The proper motions of the galaxies and the possible intergalactic presence of undetected neutrinos, magnetic fields, cosmic rays, quanta (for example, gravitons) and hydrogen are not usually considered to add a significant pressure.

The relativistic field equations for a "dust" filled R-W model are:

$$\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} - \frac{\lambda}{3} = \frac{8\pi G\rho}{3} \qquad (9)$$

$$\frac{2\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} - \lambda = 0$$
$$(= -8\pi Gp) \qquad (10)$$

In equation 10 we have parenthetically exhibited a pressure term on the right-hand side instead of zero to indicate the only modification to equations 9 and 10 for nonvanishing pressure. Multiplying the left-hand side of equation 9 by R^3, and differentiating, yields $\dot{R}R^2$ times the left-hand side of equation 10, and thus zero. Consequently

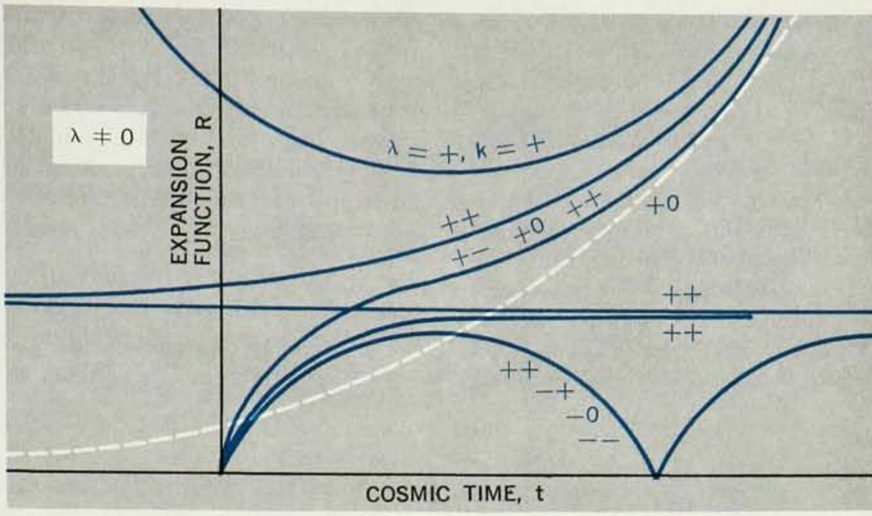
$$\frac{8\pi G\rho R^3}{3} = C \qquad (11)$$

where C is a constant, evidently positive. Substituting equation 11 in 9 yields precisely the Friedman differential equation 8, while the difference between equations 10 and 9 is formally identical with equation 7. Considerable calculation is needed to get equations 9 and 10 from the general field equations, in contrast to the trivial effort in the psuedo-Newtonian theory.

Pressure

Before proceeding, I shall make a remark about the effect of possible pressure. Although the p term in equation 7 is ad hoc from a Newtonian point of view, it is totally rigorous from the relativistic point of view from equations 9 and 10. It is sometimes said that pressure as such never produces expansion, and that only a pressure gradient does. But this is obviously disproved by the contraction of an extended balloon under the negative "pressure" of its material (say, in vacuum). However, it appears from equation 7 that in relativistic cosmology a pressure would only act by virtue of its mass equivalent. Thus a positive pressure would not produce *more* expansion but *less*!

William H. McCrea[19] has suggested an ingenious dynamics for the steady-state theory based on this idea and on equation 7. He postulates $\lambda = 0$ and



FRIEDMAN MODELS with $\lambda \neq 0$ and $\rho \neq 0$. The labels refer to the signs of λ, k. The dashed curve represents the de Sitter model ($\rho = 0$). All models allow translation and reversal in time. —FIG. 5

$p = -c^2\rho =$ constant. Although this ensures $R \propto \exp(Ht)$, $H =$ constant, the vast negative pressure remains unexplained. Still, whereas a positive pressure in an expanding universe *does* work, a negative pressure *absorbs* work. The energy so absorbed by a sphere of fixed proper radius is continually converted into mass and escapes with the expanding universe, leading to conservation of energy! The negative pressure is equivalent to negative mass, thereby producing the gravitational repulsion that drives the expansion.

Let us note here, parenthetically, that to date there has appeared no satisfactory dynamics for a hierarchical universe.

Friedman-equation solutions

We now turn to the integration of equation 8.[21,22] The three constants C, λ, k (and, to some extent, an initial value of R) determine the model uniquely, apart from an arbitrary time translation and a time reversal. For certain triplets C, λ, k, different choices of initial R can lead to two essentially different models, one with finite and one with infinite past. We shall assume $C \neq 0$ (that is, $\rho \neq 0$). If $\lambda = 0$, there are only three solutions, corresponding respectively to $k = 0$, 1 or -1 (see figure 4).

If $\lambda \neq 0$ the variety of possible models becomes much larger. This fact is shown in figure 5, where the sign pairs indicate the signs of λ and k, in this order, leading to models of the general type shown. Each model can, of course, be translated or reversed in time. The dashed line, contrary to our premise that $C \neq 0$, shows the empty relativistic de Sitter universe $R \propto \exp(Ht)$; it is included here because it also characterizes the steady-state universe and because all accelerating relativistic models resemble it asymptotically. The horizontal line corresponds to Einstein's static model.

Philosophy to the rescue?

Thus, despite the dynamic restrictions on the R-W model, we are still faced with a choice from among a discrete multitude of possibilities. *Philosophically,* there might be a preferred choice: The values $\lambda = 0$ and $k = 1$ seem, from some points of view, most attractive. Since from the "action-at-a-distance" point of view λ plays the role of an "uncaused" repulsion, one is tempted to reject it; also, one would reject it if one desired Einstein's field equations to reduce *exactly* to special relativity in the absence of gravitation and otherwise to Newton's theory as a first approximation. Yet if we simply look at the relativistic field equations in the *formal* way in which they originated—equating the matter tensor to a pseudo-Laplacian geometric tensor that shares its algebraic and differential properties, that geometric tensor undoubtedly contains a λ. Hence it might be premature to reject λ, especially since, as we shall see, it can in principle be found from observations. Again the choice $k = 1$ might appear desirable. It implies closed space sections that would, in some sense, validate Mach's principle according to which the totality of matter in the universe *and nothing else* determines the local inertial frames. For this reason Einstein in his early work favored the choice $k = 1$. Nevertheless in 1932 even he, in collaboration with de Sitter, put forward the model with $k = 0$ and $\lambda = 0$, remarking that k is, in principle, a matter of observation and not of choice. In any case, it is now well known that by making suitable topological identifications, nonsingular closed space sections are possible *even if $k = 0$ or -1*, and C and λ are arbitrary.

Observational evidence

Let us therefore turn again to the observational evidence. In addition to the parameters H and q it is useful to define the dimensionless *density parameter* σ thus

$$\sigma = 4\pi G\rho/3H^2 = C/2H^2R^3 = 2.66 \times 10^{28}\rho/h^2 \quad (12)$$

where ρ is in gm/cm³ and $h = H/100$ in km/sec/Mpc. In terms of these parameters we may write the difference of equations 9 and 10, and of three times 9 and 10, respectively, in the following forms:

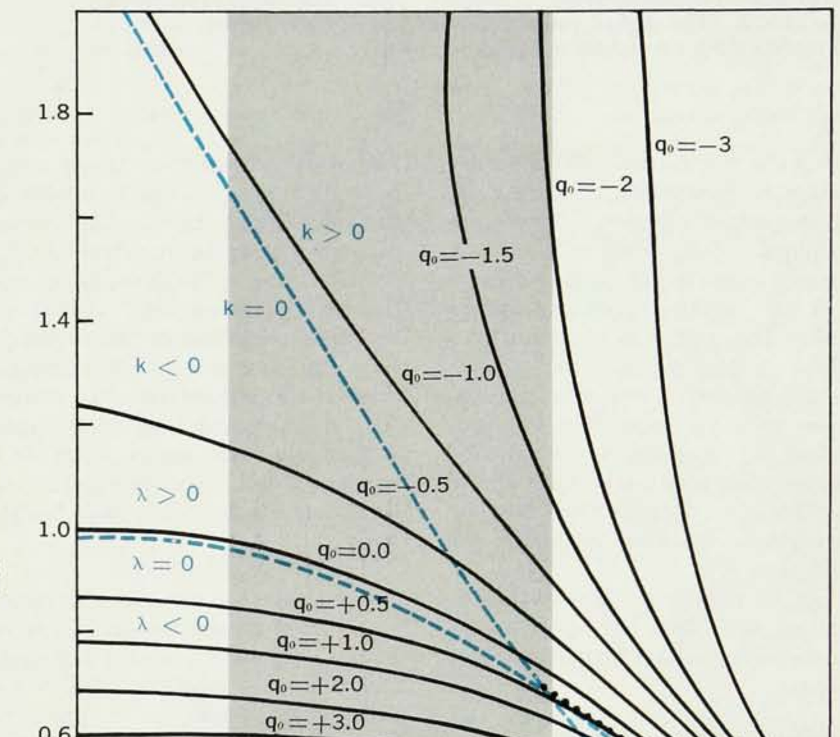
$$\sigma - q = \frac{\lambda}{3H^2} \quad (13)$$

$$3\sigma - q - 1 = \frac{k}{H^2R^2} \quad (14)$$

For later reference we substitute from equations 12, 13 and 14 (evaluated *now*) into equation 8 and so obtain[22]

$$\dot{y}^2 = H_0^2\left[\frac{2\sigma_0}{y} + (\sigma_0 - q_0)y^2 + 1 + q_0 - 3\sigma_0\right], y = \frac{R}{R_0} \quad (15)$$

(Equations 7–15 are *all* shared by relativistic and pseudo-Newtonian dynamics.) Now, in principle, an empirical knowledge of H_0, q_0, and σ_0 allows us to find λ from equation 13 and k/R_0^2 from equation 14. For example, the value $q_0 = 1.8 \pm 0.7$ recently given by Allan Sandage[23] and the very liberal restriction[24] $0.015 < \sigma_0 < 0.50$ would imply $-7.5 < \lambda/H_0^2 < -1.8$ and $k < 0$. This, however, cannot be taken too seriously because of the already mentioned evolutionary uncertainty in q_0. Suppose, therefore, we ignore the evidence of q_0. If we assume for philosophical reasons that $\lambda = 0$, then, by equation 13, $\sigma = q$; by equation 14, the other philosophical assumption $k = 1$ can now be maintained only if σ_0 is pushed beyond its already improbable maximum value of 0.5. Moreover, such a model is uncomfortably young since $\lambda = 0$ and $k = 1$ imply $t_0 < 2T_0/3$.



ROBERTSON DIAGRAM for Friedman models with finite age t_0. Dashed lines separate models with $\lambda > 0$ and < 0, $k > 0$ and < 0. Dotted line ($\sigma_0 \geq 1/2$) and $\lambda = 0$ line ($\sigma_0 < 1/2$) separate oscillating from nonoscillating models. Full lines are for constant q_0. Shaded area corresponds to $H_0t_0 > 0.56$ and $0.015 < \sigma < 0.50$. —FIG. 6

The Robertson diagram

In fact, the only really *certain* limitations on the model at present are on its age and density: The universe must be at least as old as any of its parts and at least as dense as its averaged *visible* contents. In addition, we can probably have confidence in the presently determined value of H_0 up to a factor of 2 or so. Robertson[12] pointed out that if H_0 is known, the present age t_0 and present density ρ_0 fully determine the solution of the Friedman equation, provided we leave out of consideration models to which no finite age can be assigned. (By integrating equation 15 one gets $H_0 t_0$ in terms of σ_0, q_0; equations 12–14 then yield the other parameters.)

Robertson constructed a diagram with log ρ_0 and t_0 as coördinates in which, assuming the value of H_0, every big-bang Friedman model corresponds to a unique "point." The limitations on ρ_0 and t_0 then single out a zone in which all such permissible models must occur. However, since H_0 has been subject to repeated revisions and since the observational determination of ρ_0 (at least the contribution from galaxies) involves a knowledge of H_0 in such a way that the limitations are really on σ_0 rather than ρ_0,[24] Beatrice Tinsley[10] has redrawn the Robertson diagram with log ρ_0 and $H_0 t_0$ as coordinates (see figure 6). In this diagram there is a one-to-one correspondence between points and Friedman big-bang models, up to a common scale change in R and t. Robertson drew the demarkation line between models having $\lambda > 0$ and $\lambda < 0$ and also the one between those with $k > 0$ and $k < 0$.

I have superimposed on the diagram some lines of constant q_0.[22,25] The diagram also permits a strict separation between oscillating and nonoscillating big-bang models: All the oscillating ones lie *on and below* the curve $\lambda = 0$ for $\sigma_0 < 1/2$, and for $\sigma_0 \geqslant 1/2$ they lie strictly *below* the curve[22]

$$\sigma_0^2 - (q_0 + 1)^2\sigma_0/3 + (q_0 + 1)^3/27 = 0$$

The second curve is indicated in the diagram by the dotted line that corresponds to models asymptotically approaching the static Einstein model. The shaded area represents the limitations $0.015 < \sigma < 0.5$ (though the most likely value seems to be around 0.05)[24] and $H_0 t_0 > 0.56$, which corresponds, for example, to $t_0 > 9 \times 10^9$ years and $H_0 > 60$ km/sec/Mpc.[13]

Unfortunately, the only definite conclusion from these limitations at present seems to be that $\lambda = 0$ would entail $k = -1$. The upper density limit, however, may be in error if there are significant contributions from undetected intergalactic ionized hydrogen[15] (upper limits can be placed on *neutral* hydrogen from the observations of quasars) or possibly from stars or galaxies that have collapsed within their Schwarzschild radius.

An original line of investigation using the "gravitational-lens effect" is being pursued by Sjur Refsdal,[26] and this may lead to an independent observational determination of H_0 and perhaps of other parameters. Important theoretical studies of the effects of possible inhomogeneities and anisotropies have been initiated by J. Kristian and R. K. Sachs[27] and Sachs and Wolfe.[6] Of course, a great number of exact cosmological solutions of Einstein's field equations satisfying homogeneity but not isotropy are known (for example, the famous Gödel universe), but their status is still that of theoretical curiosities rather than serious substitutes for the R-W models.

In conclusion, I recommend to the interested reader the other surveys by Robertson,[12,21] McCrea,[28] Heckmann and Schücking,[18] Sciama,[29] Bondi,[4] Sandage,[25] Zeldovich,[30] Glanfield,[31] Stabell and Refsdal,[22] and Novikov and Zeldovich.[32]

* * *

This article is based on an invited talk at the APS Southwestern Meeting (Austin, Feb. 1967). The work was supported, in part, by NASA and AFOAR grants.

References

1. H.-Y. Chiu, PHYSICS TODAY **17**, no. 5, 21 (1964); H. J. Smith, Appl. Optics **5**, 1701 (1966); E. M. Burbidge, p. 399 in *Annual Review of Astronomy and Astrophysics*, Vol. 5, Annual Reviews, Palo Alto, Calif. (1967).
2. R. H. Dicke, P. J. E. Peebles, P. G. Roll, D. T. Wilkinson, Astrophys. J. **142**, 414 (1965); A. A. Penzias, R. W. Wilson, Astrophys. J. **142**, 419 (1965).
3. R. H. Dicke, PHYSICS TODAY **20**, no. 1, 55 (1967); I. W. Roxburgh, Nature **213**, 1077 (1967).
4. H. Bondi, *Cosmology*, Cambridge U. Press, London (1961), p. 19.
5. F. Hoyle, 1963 Moscow lecture, unpublished.
6. R. B. Partridge, D. T. Wilkinson, paper at Jan. 1967 APS Annual Meeting, New York; R. K. Sachs, A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
7. G. de Vaucouleurs, Soviet Astron—AJ **3**, 897 (1960).
8. A. G. Walker, J. Lond. Math. Soc. **19**, 219, 227 (1944).
9. W. Rindler, Monthly Notices of the Royal Astronomical Society **116**, 662 (1956).
10. B. M. Tinsley, PhD thesis, University of Texas (1967).
11. E. L. Scott, Astron. J. **62**, 248 (1957).
12. H. P. Robertson, Publications of the Astronomical Society of the Pacific **67**, 82 (1955).
13. P. W. Hodge, G. Wallerstein, Pub. Astr. Soc. Pac. **78**, 411 (1966); O. C. Wilson, Pub. Astr. Soc. Pac. **79**, 46 (1967); P. A. Wayman, ibid. 156.
14. D. W. Sciama, M. N. Roy. Astr. Soc. **128**, 49 (1964); D. W. Sciama, W. C. Saslaw, Nature **210**, 348 (1966); P. Veron, Nature **211**, 724 (1966).
15. M. Kaufman, Nature **207**, 736 (1965); P. J. E. Peebles, R. H. Dicke, Nature **211**, 574 (1966).
16. D. W. Sciama, M. J. Rees, Nature **211**, 1283 (1966); P. Kafka, Nature **213**, 346 (1967).
17. G. R. Burbidge, E. M. Burbidge, Astrophys. J. **148**, L107 (1967).
18. W. H. McCrea, E. A. Milne, Quarterly Journal of Mathematics **5**, 73 (1934); O. Heckmann, E. Schücking, p. 489 in *Handbuch der Physik, Vol.* 53, Springer-Verlag, Berlin (1959).
19. W. H. McCrea, Proc. Roy. Soc. **A206**, 562 (1951).
20. E. L. Schücking, AMS Summer Seminar at Cornell (1965).
21. H. P. Robertson, Rev. Mod. Phys. **5**, 62 (1933).
22. R. Stabell, S. Refsdal, M. N. Roy. Astr. Soc. **132**, 379 (1966).
23. A. Sandage, paper at 1967 (January) Texas Symposium on Relativistic Astrophysics in New York.
24. G. O. Abell, p. 1 in *Annual Review of Astronomy and Astrophysics, Vol.* 3, Annual Reviews, Palo Alto, Calif. (1965); P. J. E. Peebles, Astrophys. J. **147**, 859 (1967).
25. A. Sandage, Astrophys. J. **133**, 355 (1961).
26. S. Refsdal, M. N. Roy. Astr. Soc. **128**, 295, 307 (1964); **132**, 101 (1966); **134**, 315 (1966).
27. J. Kristian, R. K. Sachs, Astrophys. J. **143**, 379 (1966).
28. W. H. McCrea, p. 321 in *Reports on Progress in Physics, Vol. 16,* The Institute of Physics and the Physical Society, London (1953).
29. D. W. Sciama, p. 311 in *Vistas in Astronomy, Vol. 3,* Pergamon Press, Oxford (1960).
30. Ya. B. Zeldovich, p. 241 in *Advances in Astronomy and Astrophysics, Vol.* 3, Academic Press, New York (1965).
31. J. R. Glanfield, M. N. Roy. Astr. Soc. **131**, 271 (1966).
32. I. D. Novikov, Ya. B. Zeldovich, op. cit. ref. 1 (Ann. Rev.), p. 627. □