THE ASTRONOMICAL ASPECT OF THE THEORY OF RELATIVITY*

BY

W. DE SITTER

I

1. INTRODUCTION. GRAVITATION AND INERTIA

The old Greek philosophy, which in Europe in the later middle ages was synonymous with the works of Aristotle, considered motion as a thing for which a cause must be found: a velocity required a force to produce and to maintain it. The great discovery of Galileo was that not velocity, but acceleration requires a force. This is the law of inertia of which the real content is: *the natural phenomena are described by differential equations of the second order*. The science of mechanics as based on this law of inertia was made into a consistent system by Newton. Newton also discovered the law of gravitation, that force which causes bodies on earth to fall, the moon to move in its orbit around the earth, and the planets around the sun. Both the law of inertia and the law of gravitation contain a numerical factor or a constant belonging to matter, which is called *mass*. We have thus two definitions of mass; one by the law of inertia: mass is the ratio between force and acceleration. We may call the mass thus defined the inertial or *passive* mass, as it is a measure of the resistance offered by matter to a force acting on it. The second is defined by the law of gravitation, and might be called the gravitational or *active* mass, being a measure of the force exerted by one material body on another. The fact that these two constants or coefficients are the same is, in Newton's system, to be considered as a most remarkable accidental coincidence and was decidedly felt as such by Newton himself. He made experiments to determine the equality of the two masses by swinging a pendulum, of which the bob was hollow and could be filled up with different materials. The force acting on the pendulum is proportional to its active mass, its inertia is proportional to its passive mass, so that the period will depend on the ratio of the passive and the active mass. Consequently the fact that the period of all these different pendulums was the same, proves that this ratio is a constant, and can be made equal to unity by a suitable choice of units, i.e., the inertial and the gravitational mass are the same. These experiments have been repeated in the nineteenth century by Bessel, and in our own times by Eötvös and Zeeman, and the identity of the inertial and the gravitational mass is one of the best ascertained empirical facts in physics—perhaps the best. It follows that the so-called fictitious forces introduced by a motion of the body of reference, such as a rotation, are indistinguishable from real forces. Thus, for example, the force acting on Newton's famous apple is the difference of the gravitational attraction between the earth and the apple,

* Consisting, in part, of the lectures delivered on the Hitchcock Foundation, 1932.

which is a "real" force, and the centrifugal force, due to the rotation of the earth, which in the classical system of mechanics is a "fictitious" force, since it arises from the inertia of the apple. This distinction between real and fictitious forces, however, is an artificial or formal one, introduced by the theory; there is no essential difference between the two. In Einstein's general theory of relativity there is also no formal theoretical difference, as there was in Newton's system. Inertia and gravitation are identical, the equality of inertial and gravitational mass is no longer an accidental coincidence, but a necessity.

2. The Restricted Principle of Relativity

The physical world has three spacial dimensions and one time dimension. Why this is so, and what is the meaning of it, is a difficult metaphysical or psychological problem. For our present purpose it may be simply accepted as an empirical fact. The position of a material particle m at a certain time t is thus defined by three space coordinates x, y, z. The complex of these five data, m, x, y, z, t, may be called an *event.* The different events are located in a four-dimensional continuity which is characterized by the "interval," of which the expression, in the simplest case, can be taken to be:

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2.$$

In this four-dimensional continuum, transformations of coordinates can be performed bringing the interval into the general form:

$$ds^2 = \sum_{\alpha,\beta} g_{\alpha\beta}\, dx_\alpha\, dx_\beta,$$

where α and β take the values from 1 to 4.

The laws of classical, or Newtonian, mechanics are invariant for orthogonal transformations of the three space coordinates x, y, z, and for linear transformations defining a velocity, together with a change of units, i.e., for transformations of the form

$$x' = ax + bt + c.$$

The equations of the electromagnetic theory are invariant for these same transformations, but also for the so-called *Lorentz-transformation,* which is an orthogonal transformation of the *four* coordinates x, y, z, and it. In the system of classical mechanics the continuum is not really a four-dimensional continuum but a linear series of three-dimensional continua; the time has a different character from the three space coordinates. In the electromagnetic theory the four coordinates x, y, z, and it are, as Einstein has shown, formally entirely equivalent. Thus, from about 1904 to about 1914, physicists were in a dilemma; as Sir William Bragg said: on Mondays, Wednesdays, and Fridays they believed in one system of physics and on Tuesdays, Thursdays, and Saturdays in a quite different one. In classical Newtonian mechanics space and time are absolute, have a real existence apart from the material phenomena. The independent existence of absolute space and absolute time has been specially postulated by Newton at the beginning of his great work. About twenty years ago this Newtonian system was still accepted by many physicists when they were discussing mechanical phenomena on Monday. On Tuesday, however, when they were thinking about electromagnetic phenomena, light, etc., they had the choice between three different systems. Lorentz still believed in the absolute space, which he called "æther," and the absolute time of Newton, and in his theory the motion of material systems and electrons through the æther affected the dimensions and other physical properties of these bodies, e.g. by the well-known Lorentz-contraction. The velocity of propagation of light, having nothing to do with matter, but being purely relative to the æther, was a constant. Einstein had some years previously shown that Lorentz's theory could be presented in a different form, abolishing the absoluteness of space and time, and putting the constancy of the velocity of light at the beginning of the theory as a postulate or axiom. This is the so-called restricted theory of relativity in which the postulate of absolute space and time is replaced by the constancy of the velocity of light. There was still a third theory, that of Ritz, who denied both the absoluteness of space and time, and the constancy of the velocity of light, returning thus, in a way, to Newton's emission theory of light.

This was the position about the end of the year 1912. The theories of Lorentz and Einstein are only two different interpretations of the same set of formulae and are consequently really the same theory. There is no *experimentum crucis* which can distinguish between the two; whether we accept the one or the other is a question of taste. On the other hand, between these two and the theory of Ritz an experimentum crucis is quite possible, and it was pointed out early in 1913[1] that the experiment had already been made hundreds of times. The existence of spectroscopic binaries and the possibility of representing the observed relative radial velocities by the ordinary Keplerian laws, provide a conclusive proof of the constancy of the velocity of light. We were thus left (on Tuesdays) with only one theory, the restricted theory of relativity, either in the form of Lorentz or of Einstein. In this theory the laws of nature are invariant for Lorentz-transformations, as has been pointed out. Several physicists—Lorentz, Poincaré, and others, have tried to fit the mechanical laws—the Monday theory—into the new Tuesday scheme, but of course this could not be done without some adjustment. Gravitation in the new system was still a force, like any other force, requiring its own particular law. Newton's law of gravitation, not being invariant for a general Lorentz-transformation, but only for the transformations of classical mechanics, required a slight emendation to fit it into the system of the restricted theory of relativity, which seriously impaired its beautiful simplicity and elegance, and the identity of gravitational and inertial mass remained an accidental coincidence or a miracle, as before.

3. The General Theory of Relativity. Field Equations and Equations of the Geodesic

In January, 1914, Einstein published the first draft of his general theory of relativity, not completed until November, 1915. In this theory the laws of nature are invariant, not only for Lorentz-transformations but for *any* arbitrary transformation of the four coordinates x, y, z, t, within certain restrictions of continuity, etc. If we make the assumption—which in the light of the modern developments of quantum theory, wave mechanics, and the like, might, however, appear somewhat

[1] de Sitter, *Proceedings*, R. Acad. Sci. Amsterdam, 15: 1297.

dangerous—that a material particle or electron has individuality, so that it makes sense to speak of different positions at different times of the same particle—if we make that assumption, the sequence of different positions of the same particle at different times forms a one-dimensional continuum in the four-dimensional space-time, which is called the *world-line* of the particle. All that physical experiments or observations can teach us refers to intersections of world-lines of different material particles, light-pulsations, etc., and how the course of the world-line is between these points of intersection is entirely irrelevant and outside the domain of physics. The system of intersecting world-lines can thus be bent and twisted about at will, so long as no points of intersection are destroyed or created and their order is not changed. This is the meaning of the invariancy for arbitrary transformations. The metrical properties of the four-dimensional continuum are described by the ten coefficients $g_{\alpha\beta}$ appearing in the expression for the interval ds in terms of dx, dy, dz, dt. The law of inertia requires that these potentials, as they are often called, $g_{\alpha\beta}$ shall be determined by differential equations of the second order. This naturally leads to the introduction of a certain tensor of the second order of which the components $G_{\alpha\beta}$ are made up of the $g_{\alpha\beta}$ and their first and second derivatives, and which has the identical property that:

$$\operatorname{div}(G_{\alpha\beta}-\frac{1}{2}g_{\alpha\beta}G)=0.$$

The physical state of matter and energy can be described by the so-called material energy tensor, of which the components are:

$$T_{\alpha\beta}=\rho\sum\nolimits_{\mu,\nu} g_{\alpha\mu}\, g_{\beta\nu}\, \frac{dx_\mu}{ds}\,\frac{dx_\nu}{ds}.$$

The laws of conservation of energy and matter are expressed by the equation:

$$\operatorname{div} T_{\alpha\beta}=0.$$

Also

$$\operatorname{div} g_{\alpha\beta}=0$$

is an identity.

The vanishing of the divergence means inherent permanency. It is thus natural to adopt for the relation connecting the metrical properties of the four-dimensional continuum with the physical properties of matter and energy, which forms the contents of this continuum, the identity of the two inherently permanent tensors, viz.: the metrical and the material tensor. The fundamental equation of the general theory of relativity is thus:

$$G_{\alpha\beta}-\frac{1}{2}g_{\alpha\beta}\,G+\lambda g_{\alpha\beta}+\kappa T_{\alpha\beta}=0, \tag{I}$$

λ and κ being two numerical constants. Calling the left member of this equation $K_{\alpha\beta}$, we have, of course, identically div $K_{\alpha\beta}=0$, which is equivalent to four conditions corresponding to the four laws of the conservation of energy (matter) and of momentum.

There are ten coefficients $g_{\alpha\beta}$ and ten equations (I) but there are four identities, so that the determination of the $g_{\alpha\beta}$ by (I) is not complete; there remains a fourfold indeterminacy. This is essential, because otherwise transformations of coordinates would no longer be possible.

The world-lines of material particles and light quanta are the geodesics in the four-dimensional continuum defined by the solutions $g_{\alpha\beta}$ of the field equations (I). The equations of the geodesic:

$$\text{(II)} \qquad \frac{d^2x_\nu}{ds^2}+\sum_{\alpha,\beta}\left\{ {\alpha\beta \atop \nu} \right\}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds}=0$$

thus are equivalent to the equations of motion of mechanics. When we come to solve the field equations (I) and substitute the solutions in the equations of motion (II), we find that, in the first approximation, that is for small material velocities (small as compared with the velocity of light), these equations of motion are the same as those resulting from Newton's theory of gravitation. Thus the distinction between gravitation and inertia has disappeared; gravitation is an intrinsic property of the four-dimensional continuum. A body, when not subjected to an extraneous force, describes a geodesic in the continuum, just as it described a geodesic in the absolute space of Newton under the influence of inertia alone.

Of the two constants λ and κ in the field equations, κ, which appears as a factor multiplying the material tensor, corresponds to the constant of gravitation in Newton's law and may thus be made equal to unity (or to 8π as is often done) by a suitable choice of the unit of mass; λ appears as a multiplier of the $g_{\alpha\beta}$ defining the metric, and consequently $\lambda=1$ may be taken to be equivalent to a choice of the unit of length.

It should be noted, however, that the equation (I) also makes sense if the term $\lambda g_{\alpha\beta}$ is omitted altogether, i.e. λ can be taken equal to zero. It can also be negative. The interpretation of λ as defining a material unit of length (which is favored by Eddington) is thus not imperative. The unit of length may be left free, and λ interpreted as meaning something else. We will return to the part played by λ later.

The unit of time has already been fixed by making the velocity of light, c, equal to one.

4. General Character of the Theory of Relativity

Two points should be specially emphasized in connection with the general theory of relativity.

First, it is a purely *physical* theory, invented to explain empirical physical facts, especially the identity of gravitational and inertial mass, and to coordinate and harmonize different chapters of physical theory, especially mechanics and electromagnetic theory. It has nothing metaphysical about it. Its importance from a metaphysical or philosophical point of view is that it aids us to distinguish in the observed phenomena what is absolute, or due to the reality behind the phenomena, from what is relative, i.e. due to the observer.

Second, it is a pure generalization, or abstraction, like Newton's system of mechanics and law of gravitation. It contains *no hypothesis*, as contrasted with the atomic theory or the theory of quanta, which are based on hypothesis. It may be considered as the logical sequence and completion of Newton's *Principia*. The science of mechanics was founded by Archimedes, who had a clear conception of the relativity of motion, and may be called the first relativist. Galileo, who was inspired by the reading of the works of Archimedes, took the subject up where his great predecessor had left it. His fundamental discovery is the law of inertia, which is the backbone of Newton's classical system of mechanics, and retains the same central position in Einstein's relativistic system. Thus one continuous line of thought can be traced through the development of our insight into the mechanical processes of nature, of which the different stages may be characterized by the sequence of these four great names: Archimedes, Galileo, Newton, Einstein.

It may be helpful to a good understanding of the conception of the physical universe implied by the general theory of relativity, to consider the different definitions of a straight line.

What are the possible physical realizations of a straight line? In the old mechanics there are four of these, viz.:

(1) a ray of light,
(2) the track of a material particle not subject to any forces,
(3) a stretched cord,
(4) an axis of rotation.

The fourth definition is the one favored by the great mathematician Henri Poincaré.

In classical mechanics these four physical representations of a straight line are identical. Are they still identical in the theory of relativity?

The definitions 1 and 2 define the straight line as the projection on the three-dimensional space x, y, z of a geodesic in the four-dimensional space-time continuum. This projection will be a geodesic in three-dimensional space only under very special conditions. In the general case the two projections will differ from each other, and neither of them will be a geodesic. Also the projection may be a geodesic in one system of coordinates but not in another.

The stretched cord is by definition a geodesic in the three-dimensional space. As a rule, this will not be a geodesic in the four-dimensional continuum. The rotation axis is also by definition a line in three-dimensional space. The definition, however, presupposes the possibility of the rotation of a rigid body, which would be possible only in a homogeneous, isotropic, and statical field, i.e., in a world without any material bodies (rotating or otherwise) in it, which by their gravitational field would upset the isotropy. The definition is thus meaningless in the general theory of relativity.

II

5. Integration of the Field Equations to the First Order

We must now consider more closely the two fundamental equations (I) and (II). It is, of course, not possible to do this without a certain amount of actual mathematical handling of the formulae. I do not intend, however, in these lectures to go into the detail of all the computations. I will, on the contrary, assume a general knowledge of the theory and the notations, and only call attention to those relations and formulae which are of special astronomical interest.[2]

[2] The best presentation of the general theory is still Eddington's book of 1923, *The Mathematical Theory of Relativity*. For the planetary motion and the motion of the moon, see: de Sitter, "On Einstein's theory of gravitation and its astronomical consequences," *Monthly Notices*, R. Astr. Soc. London, 76:699; 77:155. The mathematical foundation, the calculus of tensors, is given very completely in Eddington's book. For an exhaustive treatment see: Levi-Civita, *The Absolute Differential Calculus*, translated by Dr. E. Persico (1927).

The line element ds is determined by the potentials $g_{\alpha\beta}$, which must be found from the integration of the field equations (I). These contain the material energy tensor $T_{\alpha\beta}$ which depends on the velocities dx_α/ds. These latter are determined by the equations of motion; i.e. the equations of the geodesic (II). These contain the Christoffel symbols $\{\alpha\beta, \mu\}$, which are functions of the $g_{\alpha\beta}$ and their first differential quotients. Thus, rigorously, the treatment of the equations (I) and (II) must be simultaneous, and the problem is of a complication which surpasses our mathematical powers. We must proceed by successive approximations, and we will as a first approximation suppose the $g_{\alpha\beta}$ to differ only by small quantities from their so-called galilean values:

$$\begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{vmatrix}$$

corresponding to the special theory of relativity.

We will provisionally consider statical fields only so that the g_{i4} and the g_{4i} are zero and the others are independent of the fourth variable t. The line element can then be written:

(1) $$ds^2 = -a d\sigma^2 + f c^2 dt^2$$

$d\sigma$ being the three-dimensional line element, which we will suppose to have spherical symmetry:

(2) $$d\sigma^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2(d\psi^2 + \sin^2\psi\, d\theta^2).$$

As we are considering only small deviations from the galilean values we put:

$$a = 1 + \alpha, \quad f = 1 + \gamma$$

α and γ being small quantities of the first order. The equations of motion (II) then become to the first order:

(3) $$\frac{d^2 x_i}{c^2 dt^2} = -\frac{1}{2}\frac{\partial\gamma}{\partial x_i}.$$

Comparing with the ordinary Newtonian equations of motion:

$$\frac{d^2 x_i}{dt^2} = -\frac{\partial V}{\partial x_i}$$

we see that in first approximation γ is equivalent to the potential:

(4) $$\gamma = \gamma_1 = \frac{2V}{c^2}.$$

In the classical theory the potential V is determined by Poisson's equation:

$$\sum_k \frac{\partial^2 V}{\partial x_k^2} = 4\pi G \rho.$$

The equations replacing this in the new theory are the field equations (I) which, when developed to the first order, are found to be:

$$(5)\qquad \begin{aligned} &(1)\qquad \sum_\mu \frac{\partial^2 \alpha}{\partial x_\mu^2} = -\kappa\rho - \lambda \\ &(2)\qquad \sum_\mu \frac{\partial^2 \gamma}{\partial x_\mu^2} = \kappa\rho - 2\lambda. \end{aligned}$$

I have written down the formulae including λ. The numerical value of λ is entirely unknown, but it is certainly a small quantity of at least the second order of magnitude and can in the present approximation be neglected.

According to equation (3) the value of α is not required in the present approximation. Comparing (5, 2) with Poisson's equation we find that

$$\kappa = \frac{8\pi G}{c^2} = 1{\cdot}860\ .\ 10^{-27} \mathrm{gr}^{-1}\mathrm{cm},$$

λ being neglected.

To the first approximation we can thus take for γ the ordinary Newtonian potential (4).

The first equation (5) gives, neglecting λ:

$$\alpha = -\gamma.$$

The line element thus becomes:

$$(6)\qquad ds^2 = -(1-\gamma)d\sigma^2 + (1+\gamma)c^2dt^2.$$

6. The "Crucial Phenomena." Red Shift

Consider a fixed point in three-dimensional space so that $d\sigma = 0$ and consequently $\frac{dt}{ds} = \frac{1}{c\sqrt{1+\gamma}}$. The measure of time thus depends on the gravitational potential and is different at different places in the gravitational field; therefore the frequency of a periodic phenomenon, which is constant when expressed in the natural measure, or "proper time" ds, is variable when expressed in coordinate time t.

Consequently the spectral lines originating in a strong gravitational field will, to an observer placed in a weaker field, appear to be displaced toward the red, and inversely. The ratio of the observed and emitted wave lengths will be $1/\sqrt{1+\gamma}$, or, with sufficient approximation, $1 - \frac{1}{2}\gamma$. For a point in the gravitational field of the sun the potential is $V = -GM/r$, M being the sun's mass; therefore, for $r = R$ the radius of the sun, by (4):

$$1 - \frac{1}{2}\gamma = 1 + \frac{\kappa M}{8\pi R} = 1{\cdot}00000212.$$

The displacement toward the red of lines in the solar spectrum will thus be the same as would, according to Doppler's principle, correspond to a radial velocity of .00000212c or 0·634 km/sec. It has taken the solar physicists a long time to disentangle this small displacement (corresponding to 0·013 A for λ5000) from the many other anomalies observed in the solar spectrum, but there seems to be no doubt at present regarding its reality in the sun.

For a star with mass M and radius R expressed in those of the sun as units, the displacement will be:

$$0{\cdot}634\frac{M}{R}\ \text{km/sec.}$$

For the different spectral types we can expect the values given in the following short table. For the giants the absolute magnitude has been taken equal to −1 throughout.

Spectra	Main series	Giants
B	0·9	1·2
A	0·7	0·6
F	0·6	0·4
G	0·6	0·2
K	0·4	0·1
M	0·2	0·05

It is well known that the B-stars have a systematic displacement toward the red, the so-called K-term, and a part of this may be due to this cause. For a white dwarf, of course, the effect becomes very large, and we all remember the sensational announcement of the successful measuring of the displacement corresponding to 20 km/sec by Adams in the spectrum of the companion of Sirius.

In this computation only the gravitational field of the star itself is taken into account, and the general field of the galactic system is neglected. It is certain that the effect of this is entirely negligible.

7. Bending of Rays of Light

For a ray of light $ds=0$. The ray of light is the projection on the three-dimensional space of a geodesic in the four-dimensional continuum and can be determined from the ordinary condition for a geodesic, i.e. the equation (II). This contains the coefficients $g_{\alpha\beta}$. It is evident, therefore, that the ray will, in general, be curved and its curvature will depend on the gravitational field. Thus a ray of light passing near the sun will be bent round it. Computation shows that the displacement is inversely proportional to the minimum distance of the ray of light from the center of the sun, and equal to $4GM/c^2a$, a being this minimum distance. This would give 1·″75 at the sun's limb. As is well known, this displacement was observed by the English eclipse expeditions of 1919 (29th of May) to Brazil and Principe and again on the occasion of the eclipse of 21 September 1922 at Wallal, by the expedition of the Lick Observatory.[3]

[3] Dr. Freundlich's criticisms of the results derived from the observations of the Lick Observatory appear to me to be unfounded. Dr. Freundlich's own results from his observations in Sumatra, giving a much larger deflection (of 2·″2) must probably be explained as the effect of the insufficient accuracy of the determination of the plate constants, especially the scale value and the position of the optical center. The field of stars was very unsymmetrical, and the determination is necessarily weak.

8. Motion of Perihelia

So far it has not been necessary to go beyond the first approximation. As regards the planetary motions we have seen that, to the first approximation, the equations of motion are the same as those of the classical theory. In the second approximation we require second order terms in γ. It is found that we can retain the expression (6) for the line element, if we take $\gamma=\gamma_1+\frac{1}{2}\gamma_1^2$ retaining the value (4) of γ_1. Also the equations of the geodesic must be developed to the second order. I will not go into the details of this development but will only state the results.

For the equation of the orbit we find, instead of the ordinary equation of the ellipse

$$\frac{1}{r}=\frac{1+e\cos(\theta-\tilde{\omega})}{p}, \tag{7}$$

the similar equation

$$\frac{1}{r}=\frac{1+e\cos(g\theta-\tilde{\omega})}{p}, \tag{7'}$$

the value of g being $g=1-\dfrac{3GM}{c^2p}$.

The difference between (7′) and (7) is thus a motion of the perihelion amounting to

$$\frac{d\tilde{\omega}}{dt}=\frac{3GM}{c^2p}\cdot\frac{d\theta}{dt}.$$

This secular motion of the perihelia is the only observable effect in the planetary theory. For the case of Mercury, it is nearly equal to the well-known discrepancy between theory and observation, first discovered by Leverrier, which has baffled all attempts at explanation for over half a century. It is well known that the observed secular variations of the elements of the four inner planets could not be represented by theoretical values depending on a system of masses consistent with the observed periodic perturbations. The principal discordance is now removed by Einstein's correction of the motion of the perihelion of Mercury. The remaining discrepancies are not very disquieting. They are:

	$e\dfrac{d\tilde{\omega}}{dt}$	$i\dfrac{d\Omega}{dt}$	$\dfrac{de}{dt}$	$\dfrac{di}{dt}$
Mercury	$-0''.78\pm0''.43$*	$+0''.79\pm0''.47$	$-0''.90\pm0''.80$	$+0''.39\pm0''.80$
Venus	$-0''.18\pm0''.25$	$+0''.76\pm0''.15$	$+0''.11\pm0''.33$	$+0''.44\pm0''.34$
Earth	$-0''.08\pm0''.13$	....	$0''.00\pm0''.09$	$-0''.03\pm0''.16$
Mars	$+0''.48\pm0''.35$	$+0''.14\pm0''.12$	$+0''.29\pm0''.27$	$-0''.09\pm0''.20$

* [Adopting Dr. Jackson's new determination of the motion of the perihelion of Mercury from Hornsby's observations (*M. N.* 93: 126, Dec. 1932) we find for the residual 0″.41].

The theoretical motions have been taken from Newcomb, but they have been reduced to improved values of the masses. The probable errors contain those of the theoretical values, corresponding to the uncertainty of the masses. For the earth di/dt is the secular variation of the inclination of the ecliptic, for which the adopted observed value is a weighted mean of that derived by Spencer Jones in his *Revision* *of Newcomb's Occultation Memoir* and by Newcomb from the sun, Mercury, Venus, and Mars. The other observed values are Newcomb's, but reduced to the value 50.″2486 (1850) of the precession. The node of Venus presents the only serious discrepancy, which may or may not be real. The others are not larger than would be expected from the accidental errors. Out of 15 residuals, 8 are smaller than their probable error, and only one (the node of Venus) exceeds twice the probable error.

9. Motion of the Moon

The motion of the moon must be referred to a system of coordinates attached to and moving with the earth through the gravitational field of the sun, moon, and planets. Through the influence of this gravitational field a precession is imparted to these coordinate axes, the so-called geodesic precession,[4] amounting to 1.″917 per century. This appears in the motion of the moon as observed from the earth as a positive motion of the perigee and node. With reference to these moving axes, the motion of the moon is governed by the same equations as the motion of the planets around the sun, the only addition to Newtonian theory being thus a motion of the perigee, amounting to 0.″060 per century. The uncertainty of the observed motions of the lunar perigee and node is of the order of 1″ per century, and that of the theoretical motions as computed by Brown from the ordinary Newtonian theory is of the order of 3″ or 4″ per century. The differences between the observed and theoretical values (the latter including the new term) are $+2''\pm4''$ and $-8''\pm4''$ (probable errors) respectively. The added terms are thus too small to be verified by observation.

10. Precession

We can thus say that all effects of the relativity theory of gravitation have either been verified by observation or are too small for such verification to be possible. It is not probable that any effects have been overlooked—except possibly as regards the precession. It has so far been assumed that the motion of the earth around its center of gravity, as referred to the inertial frame attached to the earth (and thus affected by the geodesic precession with reference to a system attached to the sun), is adequately described by the usual formulae. It seems certain that this is true for the gravitational *field*. In the theory of relativity, however, a rigid body cannot exist, and a special definition is required to define what is meant by the axis of rotation and its motion in space. It might be that the equations for the motion of the axis of rotation thus defined in accordance with the new theory might differ from those for a rigid body according to the Newtonian theory. This has, so far as I know, not yet been investigated. The point is worth a careful scrutiny especially since, as was recently pointed out by Dr. Jackson,[5] the observed constants of precession and nutation cannot be reconciled with the adopted mass of the moon by the existing theoretical formulae. It does not seem probable, however, that appreciable differences in the motion of the axis of rotation (or what corresponds to it in the complete relativistic mechanics) will be found.

[4] See A. D. Fokker, *Proceedings* R. Acad. Sci. Amsterdam, 23: 729 (1921).

[5] *Monthly Notices* 90: 742 (June 1930).

III

11. INERTIAL FIELD OF THE UNIVERSE

In the general theory of relativity the difference between gravitation and inertia has disappeared, and only the general field described by the $g_{\alpha\beta}$ remains, which comprises both gravitation and inertia. It is, however, convenient to continue to make a difference and to call that part of the field which is produced by the presence of material bodies, i.e., by the deviations from homogeneity of the distribution of matter, gravitation, and the rest inertia. The question then arises: What is the inertial field of the universe? Or, in other words, what would be the field of the $g_{\alpha\beta}$ if all matter were either absent or distributed homogeneously and isotropically over space? We know that in our immediate neighborhood, say within the galactic system, the $g_{\alpha\beta}$ corresponding to this ideal condition are with great approximation the galilean values:

$$\begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{vmatrix}$$

But can we go beyond that and get any knowledge of the field of $g_{\alpha\beta}$ for the "universe"?

We know by actual observation only a comparatively small part of the whole universe. I will call this "our neighborhood." Even within the confines of this province our knowledge decreases very rapidly as we get away from our own particular position in space and time. It is only within the solar system that our empirical knowledge extends to the second order of small quantities (and that only for g_{44} and not for the other $g_{\alpha\beta}$), the first order corresponding to about 10^{-8}. How the $g_{\alpha\beta}$ outside our neighborhood are, we do not know, and how they are at infinity of space or time we shall never know. Infinity is not a physical but a mathematical concept, introduced to make our equations more symmetrical and elegant. From the physical point of view everything that is outside our neighborhood is pure extrapolation, and we are entirely free to make this extrapolation as we please to suit our philosophical or aesthetical predilections—or prejudices. It is true that some of these prejudices are so deeply rooted that we can hardly avoid believing them to be above any possible suspicion of doubt, but this belief is not founded on any physical basis. One of these convictions, on which extrapolation is naturally based, is that the particular part of the universe where we happen to be, is in no way exceptional or privileged; in other words, that the universe, when considered on a large enough scale, is isotropic and homogeneous.

12. OBSERVED DENSITY AND EXPANSION OF THE UNIVERSE

During the last few years the limits of our "neighborhood" have been enormously extended by the observations of extragalactic nebulae, made chiefly with the 100-inch telescope at Mount Wilson. These wonderful observations have enabled us to make more or less reliable estimates of the distances of these objects, and hence of their distribution over space. By assuming a plausible value for their average mass, we can make a rough guess at the density of matter in space. It is, at the present moment, hardly more than a guess, but the enormous increase in our knowledge during the last four or five years entitles us to the hope that in the near future we may be able to arrive at a real determination.

One of the most remarkable observational results of the last years is the systematic positive radial velocity of the extragalactic nebulae. This is found to be, within the errors of the determination, proportional to the distance:

$$\frac{dr}{dt}=hr.$$

This means that the whole universe is *expanding*, while remaining similar to itself (apart from the peculiar motions of the individual nebulae, which are small, and can be neglected, compared with the systematic motion of recession). We can thus represent the line element of three-dimensional space by $Rd\sigma$, R being a factor increasing with the time, and $d\sigma$ remaining the same. The four-dimensional line element consequently can be taken to be

$$ds^2=-R^2d\sigma^2+fdt^2 \tag{8}$$

with[6]

$$d\sigma^2=\Sigma_{p,q}\ \gamma_{pq}\ d\xi_p\ d\xi_q.$$

If the radius vector in the three-dimensional space of which $d\sigma$ is the line element be denoted by χ, then the radius vector in natural measure is

$$r=R\chi$$

and, the value of χ being subject only to small random motions different for each individual object, we have systematically

$$\frac{1}{r}\ \frac{dr}{dt}=\frac{1}{R}\ \frac{dR}{dt}=h.$$

The coefficient of expansion h is very large. Its actual value is still subject to considerable uncertainty, owing to the uncertainty of the scale of distances, but we can as a good estimate take about 460 km/sec per million parsecs.

It follows that, if the expansion goes on at the same rate, the universe doubles its size in about fifteen hundred million years.

The possibility has been suggested that the observed shift of the spectral lines toward the red might not indicate a receding motion of the spiral nebulae, but might be accounted for in some other way. In fact, all that the observations tell us is that light coming from great distances—and which therefore has been a long time on the way—is redder when it arrives than when it left its source. *Light is reddened by age:* traveling through space, it loses its energy as it gets older. Or, expressed mathematically: the wave length of light is proportional to a certain quantity R, which increases with the passing of time. By the general equations of the theory of relativity, the naturally measured distances in a homogeneous and isotropic world

[6] Throughout this paper the convention is made that Roman indices take the values 1, 2, 3 only, while Greek indices run from 1 to 4.

are then necessarily proportional to the same quantity R, unless some extraneous cause for the increase in wave length, or the loss of energy, were present. By extraneous, I mean foreign to the theory of relativity and the conception of the nature of light consistent with that theory. Moreover, this hypothetical cause should have no other observable consequences; in particular, it should produce loss of energy without any concomitant dispersion, which would blur the images and make the faint nebulae unobservable. It would require an hypothesis *ad hoc*, and a very carefully framed one, too, so as not to overshoot the mark. No such hypothesis deserving serious consideration has yet been forthcoming.

13. General Form of the Line Element

We will thus have to investigate the possibility of constructing a universe with the line element (8), in which R is a function increasing with the time t.

For the material energy tensor $T_{\alpha\beta}$ we can, on account of the homogeneity and isotropy, take:

$$T_{pq} = -g_{pq}\, p = R^2\gamma_{pq}\, p, \quad T_{4p} = T_{p4} = 0, \quad T_{44} = g_{44}\rho = g_{44}\,(\rho_0 + 3p), \tag{9}$$

ρ_0 being the material, or "invariant" density and ρ the "relative" density. The pressure p consists of the material pressure p_m, representing the random motions of the particles of matter, i.e. of the galactic systems, and the pressure of radiation p_s.

If we form the field equations (I) corresponding to the line element (8) and the energy tensor (9) it is found that the equation for G_{14} becomes:

$$G_{14} = -\frac{1}{Rf} \cdot \frac{dR}{dt} \cdot \frac{\partial f}{\partial \chi} = 0,$$

χ being the radius vector in the three-dimensional space with the line element $d\sigma$. Therefore either dR/dt or $\partial f/\partial\chi$ must be zero. In the second case f is a pure function of the time, and can be taken equal to unity without loss of generality. Consequently there are only two possible kinds of solutions, viz.: *static solutions* in which R is a constant and f is a function of the space coordinates, i.e., on account of the spherical symmetry, of the radius vector, independent of the time, and *non-static solutions* in which f is a constant while R is a function of the time.

14. Static Solutions

We know now, because of the observed expansion, that the actual universe must correspond to one of the non-static solutions. Historically, however, the static solutions were discovered first. In 1917 Einstein introduced into the field equations the term with λ and two solutions were found, which I have been in the habit of calling the solutions A and B. They are generally referred to in current literature as "Einstein's universe" and "de Sitter's universe" respectively. The line elements in the two cases are:

$$ds^2 = -R^2\left[d\chi^2 + \sin^2\chi(d\psi^2 + \sin^2\psi\, d\theta^2)\right] + dt^2, \tag{10A}$$

$$ds^2 = -R^2\left[d\chi^2 + \sin^2\chi(d\psi^2 + \sin^2\psi\, d\theta^2)\right] + \cos^2\chi\, dt^2. \tag{10B}$$

Thus in both cases the curvature of three-dimensional space is positive, and equal to $1/R^2$, and it depends on the value of λ by the conditions

$$\text{(11A)} \qquad \lambda+\kappa\rho=3/R^2, \qquad \kappa(\rho+p)=2/R^2,$$

$$\text{(11B)} \qquad \lambda+\kappa\rho=3/R^2, \qquad \kappa(\rho+p)=0.$$

Consequently in both cases λ is positive. The density has a finite value in the case A and is zero in the case B.

Of course we also had the solution without λ, i.e. the line element of the restricted theory of relativity:

$$\text{(10N)} \qquad ds^2=-R^2[d\chi^2+\chi^2(d\psi^2+\sin^2\psi d\theta^2)]+dt^2,$$

where R is an arbitrary constant, and

$$\text{(11N)} \qquad \lambda+\kappa\rho=0, \qquad \kappa(\rho+p)=0.$$

The universe A is truly static: material particles in it can have no *systematic* motion, but only random motions, corresponding to the pressure p. In the universe B there are no material particles, but if we put in one particle and one observer, the latter will see the particle moving away from him with a velocity which, if random motions are neglected, is given by[7]

$$\frac{1}{r}\frac{dr}{cdt}=\frac{1}{R}.$$

The universe B is thus not really static. It can only be made to appear so in consequence of its emptiness.

The universe A has density but no expansion: the universe B has expansion but no density.

It is convenient to express both the coefficient of expansion and the density by quantities of the dimension of a length. Thus

$$\text{(12)} \qquad h=\frac{1}{r}\frac{dr}{cdt}=\frac{1}{R_B}, \quad \kappa\rho=\frac{2}{R_A^2},$$

and we may add

$$\text{(12')} \qquad |\lambda|=\frac{3}{R_C^2}.$$

The observed values are rather uncertain, but we can adopt the following upper and lower limits[8] (expressed in cm):

$$\text{(13)} \qquad \begin{aligned} 10^{27} &< R_B < 4.10^{27} \\ 3.10^{26} &< R_A < 10^{29}. \end{aligned}$$

The value of R_C cannot be determined from astronomical observations. The two quantities R_A and R_B are thus in the actual universe of the same order of magnitude,

---

[7] See Appendix.

[8] *Proceedings* R. Acad. Sci. Amsterdam, 35: 602, 603 (1932). (The lower limit of the density ρ is 10^{-31}, which is two-thirds of the lower limit given by Hubble; the upper limit 10^{-26} is derived by Menzel from the absence of appreciable absorption).

while R_C is entirely unknown. The two universes A and B require, however, if we neglect the pressure p,

(A): $R_A = R_C/\sqrt{3} = R, \qquad R_B = \infty,$

(B): $R_B = R_C = R, \qquad R_A = \infty.$

In the case N we have, of course,

(N): $R_A = R_B = R_C = \infty.$

Consequently neither A nor B can be a good approximation to the actual universe. N (Newton's absolute space and time) might be a good approximation so long as we only wish to consider small distances and times, compared with which 10^{27} and ∞ are practically equivalent.

15. Relativity of Inertia

In 1917 this difficulty was not realized. The value of the density was still entirely unknown, and the expansion had not yet been discovered. The reason why there was felt a need to displace (10 N) by (10 A) or (10 B) was to achieve what at that time used to be called the "relativity of inertia"—a somewhat vague phrase to which various meanings were attached.

We set out to find a grand-scale model of the universe, which shall be homogeneous and isotropic. We know only a limited part of the universe, viz., "our neighborhood." In that neighborhood the distribution of matter is neither homogeneous nor isotropic: it consists almost entirely of emptiness, the matter being conglomerated into stars and galactic systems at large mutual distances; but if considered on a large enough scale it has a certain finite average density. In our large-scale model, which takes account of inertia only and leaves gravitation out of consideration, the condensations are neglected. We can thus either take as our approximation a homogeneous universe in which the density is the average density of the actual universe, into which we must then later, as a second approximation, introduce the effect of the condensations of matter into galactic systems; or we can take an empty universe and put in the galactic systems later. What we may call the "material postulate of relativity of inertia" is the assertion that inertia cannot exist without matter; therefore we must choose the first-mentioned method of approximation, i.e. the solution A or any other solution having a finite value of R_A. But no other solution satisfying this condition was then known.

The potentials $g_{\alpha\beta}$ defining the line element are given by differential equations. Consequently, they are only determined apart from constants of integration, or boundary conditions at infinity. Of course we know nothing about infinity as has already been pointed out. The real condition determining the constants of integration is that they shall represent the observed phenomena in "our neighborhood." They are only put into the form of boundary conditions at infinity for reasons of mathematical convenience. It follows that the values of the $g_{\alpha\beta}$ at infinity will be different in different (but equivalent) systems of coordinates. This leads to what may be called the "mathematical postulate of relativity of inertia," which requires the $g_{\alpha\beta}$ at infinity to be zero, so as to be the same in all systems of coordinates. Solution A satisfies this postulate for the g_{pq} of three-dimensional space, and solution

B for all $g_{\alpha\beta}$. The vanishing of g_{pq} at infinity is equivalent to the finiteness of space, i.e. to a positive curvature. The boundary conditions at infinity are abolished by abolishing infinity.

It can be proved that the solutions A and B are the only possible static, homogeneous, and isotropic solutions with positive curvature.[9] Since the discovery of the expansion of the universe we know that we must choose our grand-scale model among the non-static solutions, and the solutions A and B are only of historic interest. We will therefore now concentrate our attention on the non-static case.[10]

16. Non-static Solutions

The non-static solutions were discovered by Friedmann[11] in 1922, and independently in 1927 by Lemaître,[12] who worked out the astronomical consequences in considerable detail. The papers in which these authors communicated their discoveries, however, were discovered by the astronomical world at large only in the spring of 1930, and since then the theory of these expanding universes has been the object of constant interest and much discussion. Friedmann discusses the solutions of the field equations for different values of λ. Lemaître considers only a positive λ. Both authors consider a positive curvature of space only. The fact that both λ and the curvature may as well be negative or zero was only pointed out by Dr. Heckmann[13] in July 1931.

We take the line element

(8) $$ds^2 = -R^2 d\sigma^2 + d(ct)^2$$

with $$d\sigma^2 = \Sigma_{p,q}\, \gamma_{pq}\, d\xi_p\, d\xi_q, \qquad g_{pq} = -R^2\gamma_{pq}.$$

R is a function of t only, and the γ_{pq} are independent of t.

We have then

$$G_{pq} = {}_{(3)}G_{pq} - (R\ddot{R} + 2\dot{R}^2)\gamma_{pq},$$

$$G_{p4} = G_{4p} = 0, \qquad G_{44} = 3\,\frac{\ddot{R}}{R},$$

dots denoting differential quotients d/cdt, and ${}_{(3)}G_{pq}$ being the contracted Riemann tensor corresponding to the three-dimensional line element $d\sigma$.

For the material tensor we take

(9) $$T_{pq} = -g_{pq}\, p, \qquad T_{p4} = T_{4p} = 0, \qquad T_{44} = \rho = \rho_0 + 3p, \qquad T = \rho_0.$$

The field equations

(I) $$G_{\alpha\beta} - \lambda g_{\alpha\beta} + \kappa\left(T_{\alpha\beta} - \frac{1}{2}\, g_{\alpha\beta}\, T\right) = 0$$

then become

(14) $$_{(3)}G_{pq} + 2k\gamma_{pq} = 0$$

---

[9] de Sitter, *Proceedings* R. Acad. Sci. Amsterdam, 20: 1311 (1918), also Tolman, *Proceedings* Nat. Acad. Sci. Washington, 15: 297 (1929) and Robertson, *ibid.*: 822 (1929).

[10] See, however, the Appendix.

[11] *Zeitschr. für Physik*, 10: 377.

[12] Ann. Soc. Scient. de Bruxelles, 47 A, : 49; also translated in *M. N.* 91: 483 (1931).

[13] *Göttinger Nachrichten*, 1931: 127.

with

(15) $$2k = R^2(\lambda + \kappa p + \frac{1}{2}\kappa\rho_0) - R\ddot{R} - 2\dot{R}^2$$

and

(16) $$3\frac{\ddot{R}}{R} - \lambda + \kappa(3p + \frac{1}{2}\rho_0) = 0.$$

17. Three-Dimensional Space of Constant Curvature

The equation (14) means that the three-dimensional space with the line element $d\sigma$ has the constant curvature k. The value of k is given by (15). It is independent of the space coordinates, since R, p and ρ_0 are independent of the space coordinates, and independent of the time on account of (14), in which the time does not enter. It is no loss of generality if we restrict the possible values of k to $+1$, 0, and -1. The line element $d\sigma$, then, is that of a space of *unit* curvature, and it has one of the three standard forms

$$\begin{aligned} k = +1 &: \quad d\sigma^2 = d\chi^2 + \sin^2\chi(d\psi^2 + \sin^2\psi\, d\theta^2) \\ k = 0 &: \quad d\sigma^2 = d\chi^2 + \chi^2(d\psi^2 + \sin^2\psi\, d\theta^2) \\ k = -1 &: \quad d\sigma^2 = d\chi^2 + \sinh^2\chi(d\psi^2 + \sin^2\psi\, d\theta^2). \end{aligned}$$

The curvature of the actual three-dimensional space (line element $Rd\sigma$) is then

$$\epsilon = \frac{k}{R^2}.$$

In each of the two cases $k = \pm 1$ the curved space can be projected on an euclidean space; thus, if we put for brevity

$$d\varphi^2 = d\psi^2 + \sin^2\psi\, d\theta^2$$

(17)
$$\begin{aligned} k = +1: \quad R^2 d\sigma^2 &= R^2(d\chi^2 + \sin^2\chi d\varphi^2) \\ &= \frac{dr^2 + r^2 d\varphi^2}{\left(1 + \frac{1}{4}\frac{r^2}{R^2}\right)^2} && r = 2R\tan\frac{1}{2}\chi \\ &= \frac{d\mathrm{r}^2}{\left(1 + \frac{\mathrm{r}^2}{R^2}\right)^2} + \frac{\mathrm{r}^2 d\varphi^2}{1 + \frac{\mathrm{r}^2}{R^2}} && \mathrm{r} = R\tan\chi = \frac{r}{1 - \frac{1}{4}\frac{r^2}{R^2}}, \\ k = -1: \quad R^2 d\sigma^2 &= R^2(d\chi^2 + \sinh^2\chi\, d\varphi^2) \\ &= \frac{dr^2 + r^2\, d\varphi^2}{\left(1 - \frac{1}{4}\frac{r^2}{R^2}\right)^2} && r = 2R\tanh\frac{1}{2}\chi \\ &= \frac{d\mathrm{r}^2}{1 + \frac{\mathrm{r}^2}{R^2}} + \mathrm{r}^2 d\varphi^2 && \mathrm{r} = R\sinh\chi = \frac{r}{1 - \frac{1}{4}\frac{r^2}{R^2}}. \end{aligned}$$

By the first transformation, the so-called "stereographic projection," transforming χ to r the curved space is in both cases projected on the inside of the sphere $r=2R$ in euclidean space. In the second case, the transformation to r, the projection fills the whole of euclidean space.

In the case $k=+1$ we must of the two possible cases choose the "elliptical" space, in which χ can take only the values from zero to $\frac{1}{2}\pi$, and not the "spherical" space in which the maximum value of χ is π. The elliptical space is the one of which our ordinary euclidean geometry is the limiting case for $R=\infty$. In our common geometry a plane has a line (and not a point) at infinity; two straight lines have one point of intersection (and not two); if we go to infinity along one branch of a hyperbola, we return along the other, on the other (and not on the same) side of the asymptote. All these are properties of the elliptical, as contrasted with the spherical space. The spherical space is, in fact, an entirely unnecessary, and therefore physically meaningless, reduplication of the elliptical space.[14] Moreover, the spherical space gives rise to discontinuities without physical meaning at the antipodal points of material particles.

18. Fundamental Equations of the Expanding Universe

From the equations (15) and (16) we find easily the fundamental equations of the expanding universe:

$$2\frac{\ddot{R}}{R}+\frac{\dot{R}^2}{R^2}+\frac{k}{R^2}=\lambda-\kappa p\ , \tag{18}$$

$$\frac{\dot{R}^2}{R^2}+\frac{k}{R^2}=\frac{1}{3}(\lambda+\kappa\rho)\ . \tag{19}$$

Since $\dot{R}/R=h$, the equations (18) and (19) can be brought to the form

$$\begin{aligned}\lambda+\kappa\rho&=3(\epsilon+h^2)\\ \kappa(\rho+p)&=2(\epsilon-\dot{h})\ ,\end{aligned} \tag{20}$$

which can be compared with (11A) and (11N) for the static universes (10A) and (10N).[15]

Since $\dot{h}$ is entirely unknown these equations are insufficient to determine λ and ϵ from the observational data, even supposing that not only h but also ρ and p were accurately known. Even the signs of λ and of ϵ remain indeterminate.

The equation of energy $\operatorname{div}T_{\alpha\beta}=0$ gives

$$\dot{\rho}+3\frac{\dot{R}}{R}(\rho+p)=0. \tag{21}$$

The equation (21) can also be derived from (18) and (19). The three equations (18), (19), and (21) therefore give only two conditions for the three unknowns,

---

[14] See also Eddington, *The Mathematical Theory of Relativity*, pp. 157-159.

[15] The quasi-static universe (10B) has $g_{44}=\cos^2\chi$, and is thus not directly comparable with the non-static universes like (10A) and (10N) which have $g_{44}=1$, since h is not invariant. If (10B) is transformed to a line element with $g_{44}=1$, the equations (20) are found to be satisfied. See Appendix.

R, ρ, and p. They must be supplemented by an "equation of state" giving a relation between ρ and p, or between ρ_0 and p. The pressure p is the sum of the material pressure p_m and the pressure of radiation p_s. The invariant mass of radiation is zero. Therefore:

$$\rho_m = \rho_0 + 3p_m, \qquad \rho_s = 3p_s, \qquad \rho = \rho_0 + 3p_m + 3p_s. \tag{22}$$

The material pressure represents the random motions of material particles, i.e. in our case of the galactic systems. It is easily found[16] that

$$\frac{3p_m}{\rho_0} = R^2\left(\frac{d\sigma}{ds}\right)^2 = \frac{\varphi_0^2}{R^2}, \tag{23}$$

where φ_0^2 is a measure of the average random velocities. The galactic systems are continually sending out energy of radiation, by which their mass is diminished. We can measure this rate of transformation of matter into energy against the rate of expansion of the universe, putting

$$\frac{\dot{M}}{M} = -\gamma\frac{\dot{R}}{R},$$

where $M = R^3\rho_0$. In the case $k = +1$, $\pi^2 M$ is thus the total mass of the universe. In the cases $k=0$ and $k=-1$, M is just another variable replacing ρ_0 and is introduced in order to separate the change of density due to the transformation of matter from that due to the expansion. The total change of ρ_0 is, of course, given by

$$\frac{\dot{\rho}_0}{\rho_0} = \frac{\dot{M}}{M} - 3\frac{\dot{R}}{R}.$$

From the known magnitudes and masses of the spiral nebulae we can estimate the rate of conversion $\dot{M}/M$, which comes out about the same as that of a dwarf star of somewhat later type than the sun. We find in this way that γ is of the order of magnitude of 10^{-6}. The change of ρ from this cause is thus negligible compared with the change of density by the expansion given by (21), and we can with sufficient approximation neglect the interaction between matter and radiation. Our "equation of state" then consists of two equations, viz.: the second of (22) for radiation and (23) for matter, while the equation of energy (21) is also split up into two equations:

$$\begin{aligned} &\frac{\dot{\rho}_0}{\rho_0} + \frac{3R^2+2\varphi_0^2}{R^2+\varphi_0^2}\cdot\frac{\dot{R}}{R} = 0 \\ &\frac{\dot{p}_s}{p_s} + 4\frac{\dot{R}}{R} = 0, \end{aligned} \tag{24}$$

from which we find at once

$$\kappa\rho_0 = \frac{3R_1}{R^2\sqrt{R^2+\varphi_0^2}}, \qquad \kappa\rho_m = \frac{3R_1\sqrt{R^2+\varphi_0^2}}{R^4}, \tag{25}$$

$$\kappa p_s = \frac{\beta_1}{R^4}, \tag{26}$$

R_1 and β_1 being constants of integration.

[16] Lemaître, *B. A. N.* V, 200: 273 (1930), and Heckmann, *Göttinger Nachrichten* 1931: 130. See also de Sitter, *B. A. N.* V, 193: 217 (1930), and art. 33 of the present paper.

Introducing this into equation (19) we have the equation for the expanding universe:

$$\dot{R}^2 = l\frac{R^2}{R_C^2} - k + \frac{R_1\sqrt{R^2+\varphi_0^2}+\beta_1}{R^2}\,. \qquad (k,\ l=+1,\ 0,\ -1) \tag{27}$$

Both k and l can have the values $+1$, 0, or -1, and they determine the sign of the curvature and of λ respectively. R_1 and β_1 are necessarily positive.

19. Discussion of the Pressure of Radiation

The energy of radiation consists of the ordinary radiation of the stars and the cosmic rays:

$$p_s = p_* + p_c$$

As to the first, Eddington[17] estimates the energy-density as

$$3p_* = 7{\cdot}67\ .\ 10^{-13}\ \text{ergs/cm}^3 = 8{\cdot}5\ .\ 10^{-34}\ \text{gr. cm}^{-3}.$$

This is inside the galactic system, in the neighborhood of the sun, where the material density is of the order of 10^{-23}. We can thus take approximately

$$\frac{3p_*}{\rho_0} = 10^{-10}, \tag{28}$$

and the same ratio may be taken to hold over the whole universe. As to the cosmic radiation, Millikan and Cameron[18] give for the total energy received by a square centimeter outside the earth's atmosphere $3{\cdot}07\ .\ 10^{-4}$ ergs. This gives for the density

$$3p_c = \frac{3{\cdot}07\ .\ 10^{-4}}{\pi c}\ \text{ergs/cm}^3 = 3{\cdot}6\ .\ 10^{-36}\ \text{gr cm}^{-3}. \tag{29}$$

This density is probably the same all over space. With a view to the great uncertainty of the average material density of the universe, of which the adopted limiting values are 10^{-26} and 10^{-31}, the ratio $3\ p_c/\rho_0$ remains very uncertain, but it is certainly small, and we can safely assume that the ratio $3\ p_c/\rho_0$ is of the order of magnitude of 10^{-6}, or 10^{-5} at the utmost. We can thus with sufficient approximation neglect the radiation pressure altogether and take $\beta_1=0$ in (27).*

* [A. H. Compton (*Phys. Rev.* 41:681, 1932) and J. Clay (*Proceedings* R. Acad. Sci. Amst. 35: 1282, Dec. 1932) have recently published observations from which it seems to follow that the cosmic radiation is corpuscular, instead of electromagnetic. It is easily verified that this makes no difference in the formulae, the "invariant" density corresponding to the observed p_c being practically zero on account of the enormous velocity, so that we have still $\rho_c=3p_c$ as for electromagnetic radiation.]

20. Transformation of Matter into Radiation

In deriving the equation (27) we neglected the interaction between matter and radiation. This is certainly sufficiently exact for all practical purposes. I will, however, in connection with the transformation of matter into energy by the radiation

---

[17] *The Internal Constitution of the Stars*, p. 371.

[18] *Phys. Review* 31: 930 (1928). [Regener recently found $5{\cdot}2.10^{-3}$ ergs, see *Nature*, 1933, Jan. 28].

of the stars, draw attention to a remarkable result which can be derived from the energy equation (21). If we put

$$M=R^3\rho_0, \quad E=3R^3p, \quad p_m=0,$$

the equation (21) becomes

$$R(\dot{M}+\dot{E})+E\dot{R}=0. \tag{30}$$

If we take as before $\frac{\dot{M}}{M}=-\gamma\frac{\dot{R}}{R}$, $M=M_0\,R^{-\gamma}$,

we find at once from (30)

$$E=E_0R^{-\gamma}$$

with $E_0=\gamma M_0/(1-\gamma)$. This shows that, notwithstanding the conversion of matter into energy, the total amount of energy in the universe is *decreasing*. In fact, the loss of energy corresponding to the red shift that results from the receding motion, exceeds the gain by the conversion of matter into energy. The old question what becomes of the energy that is continually being poured out into space by the radiation of the stars, thus finds an unexpected solution. It is used up, and more than used up, by the work done in the expansion of the universe. Nevertheless, it would be wrong to say that the expansion is caused *by* the pressure of radiation. The universe would expand just the same if there were no radiation at all. It expands simply because it cannot remain of the same size throughout, a static universe being unstable.

We have here neglected the material pressure p_m, but the result remains the same if it is duly taken into account.[19]

21. Rays of Light

The path of a ray of light is the geodesic given by $ds=0$, or

$$d\sigma=\frac{cdt}{R}.$$

Therefore if two successive light pulses leave a source at a time interval δt, they will be observed at a time interval δt_1, and we have, the distance σ between the source and the observer remaining the same,

$$\frac{\delta t_1}{R_1}-\frac{\delta t}{R}=0.$$

Consequently

$$\frac{\delta t_1}{\delta t}=\frac{\nu}{\nu_1}=\frac{\lambda_1}{\lambda}=\frac{R_1}{R}.$$

The observed and the emitted wave lengths are in the same ratio as the values of R at the times of observation and emission. Since R increases with the time, the wave length also increases, and the energy decreases, as the light travels through space. This is observed as the red shift of the lines in the spectra of the extragalactic nebulae.

---

[19] See de Sitter, *B. A. N.* V, 193: 216, 217 and 200: 274, 275 (1930).

22. Balance of Gravitation and Expansive Force

The equations that we have discussed show that in the grand-scale model of the universe, in which the effect of the local deviations from homogeneity is disregarded, i.e. the inertial field alone is considered, neglecting gravitation, there is an inherent tendency for the universe to change its scale, which at the present moment results in an expansion, but may perhaps at other times have been or may become a contraction. If we come to put in the details, the stars and the galactic systems, or, mathematically expressed, the singularities of the field, we find that there is also a tendency, called gravitation, to diminish the mutual distances between these singularities. We know that for small distances, within the solar system, the second tendency is by far the strongest, and the effect of the expansion is entirely negligible. On the other hand, for very large distances, such as those from one galactic system to the next, the expansive force is stronger than the gravitational attraction. There must be a limiting value of the mutual distances, i.e. of the density, for which the two forces balance.

Consider the motion of a test-body in the gravitational field inside a homogeneous sphere of material density ρ. The acceleration by the ordinary gravitation of this field is given by

$$\frac{1}{r}\left(\frac{d^2r}{c^2dt^2}\right)_1 = -\frac{\mathrm{G}}{c^2}\rho = -\frac{\kappa\rho}{8\pi} = -B.$$

The acceleration by the expansive force, on the other hand, is, with sufficient approximation

$$\frac{1}{r}\left(\frac{d^2r}{c^2dt^2}\right)_2 = \frac{1}{R}\,\frac{d^2R}{c^2dt^2} = \frac{\ddot{R}}{R} = \frac{1}{3}\lambda - \frac{1}{6}\kappa\rho_0 = A,$$

by (18) and (19), ρ_0 being the general average density of the universe (the pressure being neglected). Therefore the total acceleration is

$$\frac{d^2r}{c^2dt^2} = -(B-A)r.$$

So long as $B-A$ is positive the path of the body will be a closed curve, in fact an ellipse, it will not leave the system to which it belongs, and the system will not expand. We find easily

$$B-A = \left(\frac{3}{4\pi}\,\frac{\rho}{\rho_0}+1\right)\frac{1}{R_A{}^2} - \lambda\,.$$

The condition that gravitation shall be stronger than the expansive force thus is

$$(31)\qquad \frac{3}{4\pi}\,\frac{\rho}{\rho_0} > R_A{}^2\,\lambda - 1.$$

Now R_A is very imperfectly known, and λ not at all. If λ were negative or zero, there would be no lower limit to the density; systems of even the smallest density would have sufficient gravitational force to keep their members together, provided, of course, that the velocities did not exceed the velocity of escape. If λ is positive there is a definite lower limit for the density. Taking $\lambda = +10^{-54}$, $R_A = 10^{28}$ (corresponding to $\rho_0=10^{-29}$), which are not implausible values, the condition becomes roughly

$$\frac{\rho}{\rho_0}>400.$$

In our galactic system, and presumably in all other galaxies as well, this condition is, of course, amply fulfilled, the ratio ρ/ρ_0 near the sun being, with the adopted value of ρ_0, about 10^6. For the Coma cluster of nebulae, Hubble finds $\rho/\rho_0=250$, which is below the limit. The mutual gravitation of the nebulae belonging to this cluster thus apparently is not sufficient to keep the cluster together. The observed range of velocities in the Coma cluster is about 1100 km/sec, of which certainly 700 or 800 km/sec must be real. The velocities of the separate members of the cluster relative to the center of gravity consequently are so large that they certainly exceed the velocity of escape, even if the gravitational field of the cluster were strong enough to keep its slow-moving members in control. It would appear that the existing clusters of nebulae are not real clusters, bound together by the gravitational attraction of their members, but just accidental and temporary irregularities in the homogeneous distribution of the galaxies over space.

Both R_A and ρ_0 in (31) change with the time. We can write (31) in the form

$$\frac{3\kappa}{4\pi}\rho>2\lambda-\frac{3R_1}{R^3}, \tag{32}$$

where the only variable quantity is R. Thus for very small values of R even the smallest density is sufficient to keep a system together against the expansive tendency of the universe, but as R increases the limit of the density needed to withstand the disintegration by expansion increases, and for very large values of R it approaches the limit

$$\rho>\frac{8\pi}{3\kappa}\lambda=4{\cdot}5\ .\ 10^{27}\lambda.$$

With all plausible values of λ this limit is so low that the continued existence of our own and all the other galaxies as individual finite systems of practically constant size in the expanding universe appears quite certain.

IV

23. Transformation of the Fundamental Equation. Different Types of Solution

We will now consider more in detail the different solutions of the fundamental equation (27). If we put

$$y=\frac{R}{R_1}, \qquad \tau=\frac{ct}{R_1}, \qquad \gamma=l\frac{R_1^2}{R_C^2}=\frac{1}{3}R_1^2\lambda,$$

the equation becomes

$$\left(\frac{dy}{d\tau}\right)^2=\frac{\beta+\sqrt{y^2+\eta_0^2}}{y^2}-k+\gamma y^2, \qquad (k=+1,\ 0,\ -1) \tag{33}$$

with

(33*) $$\beta=\frac{\beta_1}{R_1^2}, \quad \eta_0^2=\frac{\varphi_0^2}{R_1^2},$$

while γ can take all values, positive, negative or zero.

Except for very small values of y, we can in (33) safely neglect β and η_0^2. It then becomes

(34) $$\left(\frac{dy}{d\tau}\right)^2=\frac{1}{y}-k+\gamma y^2=\frac{P}{y},$$

where

(35) $$P=1-ky+\gamma y^3.$$

The quantity P varies during the expansion. Since

$$\frac{1}{R_B}=h=\frac{\dot{R}}{R}=\frac{1}{R_1 y}\frac{dy}{d\tau}, \qquad \frac{2}{R_A^2}=\kappa\rho_0=\frac{3R_1}{R^3},$$

its value at any time is

(36) $$P=\frac{3}{2}\frac{R_A^2}{R_B^2}=\frac{3h^2}{\kappa\rho_0}.$$

Taking the limits given by (13) we find that at the present moment it must be enclosed within the rather wide limits:

(37) $$0{\cdot}01<P<15{,}000.$$

If we took no account of the uncertainty of h, we would have

(36*) $$P=\frac{4{\cdot}03\ .\ 10^{-28}}{\rho_0},$$

and with the limits sometimes adopted for ρ_0, viz.: $10^{-28}>\rho_0>10^{-30}$ we would have

(37*) $$4<P<400.$$

I think, however, that (37) represents the present state of our knowledge better than (37*).

Since y is necessarily positive, real solutions of (34) are possible only for positive values of P. In the figure 1, in which the coordinates are y and γ, the curves $P=0$ have been represented by full lines. The curves approach asymptotically to the negative axis of γ for $y=0$, and to the positive axis of y for $y=\infty$. For $k=0$ and $k=-1$ the curves do not intersect the axis of y. The curve for $k=+1$ intersects the axis of y at $y=1$ and has a maximum for $y=1{\cdot}5$, $\gamma=\gamma_1=+4/27$. P is positive above the curves and negative below them; the real solutions thus correspond to the part of the semi-plane above these curves. It is seen by inspection of the diagram that there are three possible types of solution, which may be called the oscillating universes, and the expanding universes of the first and of the second kind.

In the oscillating solutions the value of y oscillates between zero and a maximum value y_1. In the expanding solutions of the first kind it increases from zero to infinity, and in those of the second kind it increases from a certain minimum value y_2 to

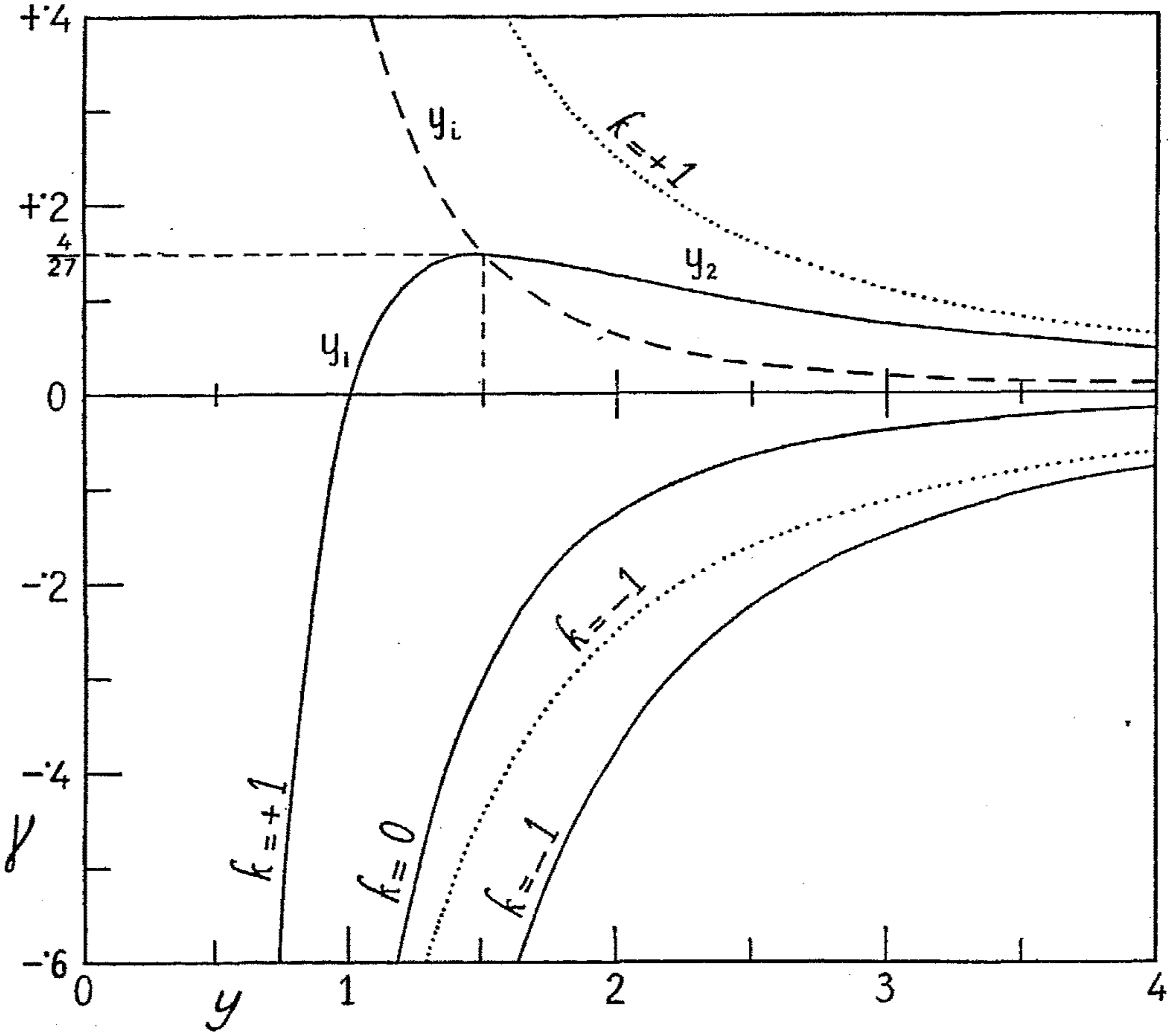

Figure 1
The full lines are the curves $P=0$, the dotted lines are $P=1$. The broken line is the curve on which the points of inflexion lie.

infinity. It is clear from the diagram that the occurrence of the different solutions depends on the values of k and γ as indicated in the following table, where also the values of P have been given.

(38) OCCURRENCE OF TYPES OF SOLUTIONS

	$k=+1$	$k=0$	$k=-1$
$\gamma>\gamma_1$	Exp. I, $P>1-\frac{2}{3}y_m$	Exp. I, $P>1$	Exp. I, $P>1$
$\gamma_1\geqq\gamma>0$	Exp. II, $P>0$ Osc., $P<1$	Exp. I, $P>1$	Exp. I, $P>1$
$\gamma=0$	Osc., $P<1$	Exp. I, $P=1$	Exp. I, $P>1$
$\gamma<0$	Osc., $P<1$	Osc., $P<1$	Osc., $P<1+\frac{2}{3}y_m$

In the expanding universes of the second kind P increases continually with y from $P=0$ for $y=y_2$ to $P=\infty$ for $y=\infty$. In the expanding universes of the first kind, with the exception of those for $k=+1$, $\gamma>\gamma_1$, P increases from $P=1$ for $y=0$ to $P=\infty$ for $y=\infty$. In the case $k=+1$, $\gamma>\gamma_1$, P begins by decreasing to a minimum $P_{min}=1-2y_m/3$ for $y=y_m=(3\gamma)^{-\frac{1}{2}}$ and from there increases to $P=\infty$ as in the other cases. In the oscillating universes, with the exception of those for

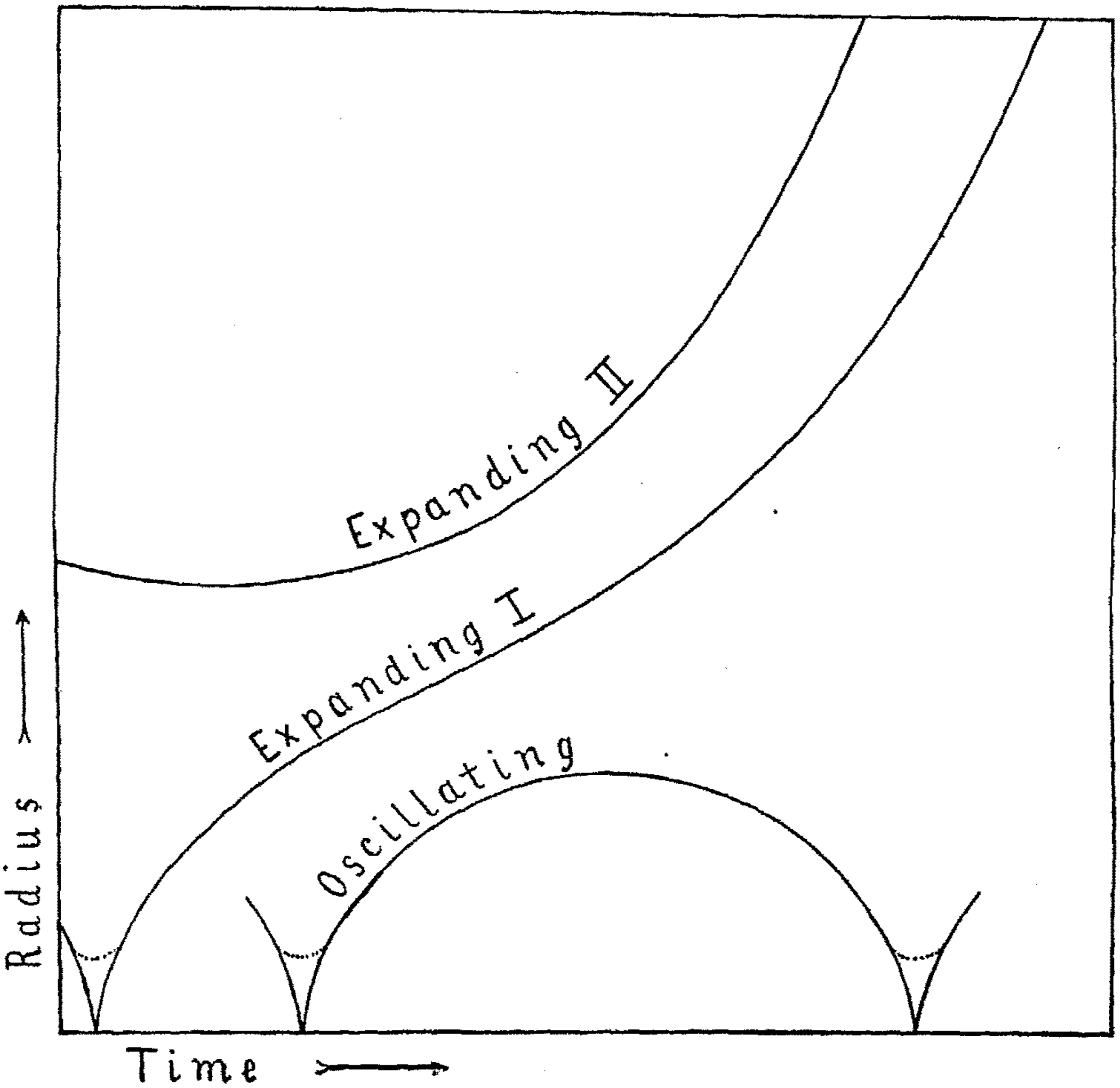

Figure 2
The different types of non-static universes.

$k=-1$, P decreases from $P=1$ for $y=0$ to $P=0$ for $y=y_1$. In the case $k=-1$, $\gamma<0$, P begins by increasing to a maximum $P_{max}=1+2y_m/3$, for $y=y_m=(-3\gamma)^{-\frac{1}{2}}$. In the case $k=0$, $\gamma=0$, P has the constant value $P=1$.

The curves $P=1$ have also been entered in the diagram in dotted lines. The axis of $\gamma(y=0)$ belongs to this curve for all values of k.

The general type of the variation of y with τ in the different cases is represented in the figure 2. For $y=0$ we have $P=1$ and therefore $\frac{dy}{d\tau}=\infty$: all solutions leave the axis of y perpendicularly: the expansion in the case of the oscillating universes and the expanding universes of the first kind begins with an explosion. The actual value $y=0$ is, of course, impossible in nature, and if y becomes very small, i.e. the density extremely large, the equations cease to be applicable. Presumably in the actual universe, if it is one of these two types, there will be a minimum value of y as shown by the dotted lines in figure 2. In the expanding universes of the second kind there is a minimum value y_2. These solutions exist only for the limited range of values of γ between zero and $\gamma_1=+4/27$.

The expanding solutions of the first kind have a point of inflection (except in the case $\gamma=0$), for a value of $y=y_i$ given by

(39)
$$\frac{d^2y}{d\tau^2}=-\frac{1}{2y^2}+\gamma y=0,$$

from which

$$y_i=(2\gamma)^{-\frac{1}{3}}.$$

This curve is also given in the diagram of figure 1, as a broken line. It is, in fact, the curve $P=3/2$ for $k=0$.

24. Integrable Cases

The cases $\gamma=\gamma_1$, $k=+1$ and $\gamma=0$, $k=+1$, 0, or -1 are the only ones in which the equation (34) can be integrated by elementary functions. For $k=+1$, $\gamma=\gamma_1$ the solutions are

$$(\gamma_1=4/27, \qquad y_1=1{\cdot}5)$$

(40)

(C)
$$(\tau-\tau_0)\sqrt{\gamma_1}=\cosh^{-1}\frac{y+y_1}{y_1}-\frac{1}{\sqrt{3}}\cosh^{-1}\frac{2y+y_1}{y-y_1},$$

(F)
$$(\tau-\tau_0)\sqrt{\gamma_1}=-\cosh^{-1}\frac{y+y_1}{y_1}+\frac{1}{\sqrt{3}}\cosh^{-1}\frac{2y+y_1}{y_1-y},$$

and, of course, also

(40*) (A)
$$y=\text{constant}=y_1.$$

The solution (C) is "Lemaître's universe" in which $\tau-\tau_0$ becomes $-\infty$ (logarithmic infinity) for $y=y_1$, while for $\tau-\tau_0=+\infty$, y becomes infinite of the order of e^τ. (F) is the limiting member of the family of oscillating universes for $k=+1$ in which y approaches asymptotically to the limiting value $y=y_1$ for $\tau=\infty$. (A) is our old friend "Einstein's universe." This latter, however, is unstable, since if we give y a small increment δy, the corresponding increment of the acceleration is

$$\delta\frac{d^2y}{d\tau^2}=\left(\frac{1}{y_1{}^3}+\gamma\right)\delta y=+\frac{4}{9}\delta y.$$

The acceleration thus has the same sign as δy, and the small increment goes on increasing.

It should be noted that (A) is the only possible case where $dy/d\tau$ and $d^2y/d\tau^2$ can be zero at the same time. From (34) and (39) we find from the condition $\frac{dy}{d\tau}=\frac{d^2y}{d\tau^2}=0$:

$$y=\frac{3}{2k},$$

which gives a finite positive value for y only if $k=+1$. In fact, the two solutions (C) and (F) form two branches of one and the same curve having a point of inflection at infinity, while (A) is the tangent at that point.

The curves (C), (F) and (A) are represented in figure 3, which will be explained later.

The solutions for the case $\gamma=0$ are:

$$\text{(41)}\qquad \begin{array}{lll} \text{(H)} & k=+1,\ y_1=1,\ P<1: & y=\sin^2\psi,\ \tau-\tau_0=\psi-\frac{1}{2}\sin 2\psi, \\ \text{(Q)} & k=0, & P=1:\ \tau-\tau_0=\frac{2}{3}y^{\frac{3}{2}} \\ \text{(L)} & k=-1, & P>1:\ y=\sinh^2\psi,\ \tau-\tau_0=\frac{1}{2}\sinh 2\psi-\psi. \end{array}$$

The solutions (Q) and (L) are expanding universes of the first kind, but without a point of inflection. (Q) is of a parabolical nature, the limiting value of $dy/d\tau$ for $y=\infty$ being zero; (L) is of hyperbolical character (without an asymptote, however) the limiting value being $dy/d\tau=1$ for $y=\infty$; (H) is an oscillating universe, the curve being similar to a half ellipse. The curves (H), (Q), and (L) are also given in figure 3.

In all other cases the integration of (34) leads to elliptic functions. The curves given in figure 3 for these cases have been computed by numerical integration, as will be explained later.

25. Empty Universes

By the introduction of y and τ instead of R and t we have, of course, excluded from consideration the case $R_1=0$, i.e. $\rho_0=0$, or universes containing no matter. In order to treat these we put

$$y'=\frac{R}{R_C},\quad \tau'=\frac{ct}{R_C},\quad \alpha=\frac{R_1}{R_C}.$$

The equation (27) then becomes

$$\text{(42)}\qquad \left(\frac{dy'}{d\tau'}\right)^2=\frac{\beta'+\alpha\sqrt{y'^2+\eta_0'^2}}{y'^2}-k+l\,y'^2,\qquad (k,\ l,=+1,\ 0,\ -1)$$

with

$$\text{(42*)}\qquad \beta'=\frac{\beta_1}{R_C^2},\quad \eta_0'^2=\frac{\varphi_0^2}{R_C^2},$$

while α can take all positive values or zero.

If we neglect again β' and $\eta_0'^2$ the equation becomes

$$\text{(43)}\qquad \left(\frac{dy'}{d\tau'}\right)^2=\frac{\alpha}{y'}-k+l\,y'^2.$$

The solutions of (43) with α different from zero are the same as those of (34), the correspondence being given by $\gamma=l\alpha^2$, $y'=\alpha y$, $\tau'=\alpha\tau$.

For $\alpha=0$ the solutions of (43) give the different "empty universes." The integration can be carried out in all cases. We find

$$
\begin{array}{llll}
l=+1: & k=+1:(B_+): & y'=\cosh\,(\tau'-\tau_0') \\
 & k=\ \ 0:(B_0): & y'=\frac{1}{2}e^{\tau'-\tau_0'} \\
 & k=-1:(B_-): & y'=\sinh\,(\tau'-\tau'_0) \\
l=\ \ 0: & k=+1: \text{no real solution} & \\
 & k=\ \ 0:(N_0): & y'=\text{constant} \\
 & k=-1:(N_-): & y'=\tau'-\tau_0' \\
l=-1: & k=+1: \text{no real solution} & \\
 & k=\ \ 0: \text{no real solution} & \\
 & k=-1\ \ (S): & y'=\sin(\tau'-\tau_0').
\end{array} \tag{44}
$$

The solutions (B_+) and (B_0) are of the expanding type of the second kind, (B_+) having a minimum $y'=1$, while (B_0) approaches asymptotically to $y'=0$ for $\tau'=-\infty$. (B_-) and (N_-) are expanding universes of the first kind, while (S) is oscillating. The solution (N_0), $R=\text{constant}$, is Newton's absolute euclidean space, independent of the time. All these solutions, with the exception of (N_0), are shown in figure 3.

All these empty universes can be transformed into (quasi-) static ones, as will be shown in the appendix. (B_+), (B_-), and (B_0) are equivalent to the static solution B.

26. Introduction of the Pressure

So far we have neglected the pressure, i.e. in (33) we have neglected β and η_0^2, and in (42) β' and $\eta_0'^2$. The solutions of (42) with $\alpha=0$ but β' different from zero have been investigated by Dr. Heckmann[20] and diagrams are given by him for these solutions for all nine combinations: $k=+1, 0, -1$, and $l=+1, 0, -1$, and for different values of β'. The curves are of the same character as those for the case $\beta'=0$, $\alpha\neq0$, i.e. the solutions of (43) or (34), which have been treated above. Since in the actual universe the value of β' is extremely small, these solutions do not correspond to any physical reality, and it has not been thought necessary to reproduce Dr. Heckmann's figures.

If we do not neglect β and η_0^2 in (33) the equation can be simplified by putting

$$z^2=y^2+\eta_0^2. \tag{45}$$

The equation then becomes

$$\left(\frac{dz}{d\tau}\right)^2=\frac{Q}{z^2}, \quad Q=\delta+z-Az^2+\gamma z^4, \tag{46}$$

with

$$
\begin{aligned}
A&=k+2\gamma\eta_0^2 \\
\delta&=\beta+k\eta_0^2+\gamma\eta_0^4.
\end{aligned} \qquad (k=+1,\ 0,\ -1) \tag{47}
$$

[20] *Göttinger Nachrichten*, February 1932: 181.

Real solutions, again, are possible only if Q is positive. The curves $Q=0$ are, for small values of β and η_0^2, i.e. if A is nearly equal to k and δ is small, very similar to the curves $P=0$ given in figure 1. There is added a branch

$$z_0 = -\delta + A\delta^2 - 2A^2\delta^3 + \ldots\ldots, \tag{48}$$

which does not, however, correspond to a real value of y. Since β and η_0^2 are of the same order of magnitude, z_0 is also of the same order, and z_0^2 is of a smaller order of magnitude. Consequently $z_0^2-\eta_0^2$ is negative, and to $z=z_0$ corresponds an imaginary value of y.

The ordinary roots of $Q=0$ differ very little from those of $P=0$. The small corrections can easily be found numerically, when required, or can be computed by the following developments in series. If y_{10} is the root of $P=0$ for a given value of γ, then the root z_1 of $Q=0$ for the same γ is found from

$$\begin{aligned} z_1 &= y_{10}+\zeta \\ \zeta &= \frac{\epsilon}{a}-\frac{b}{a^3}\epsilon^2+\ldots\ldots \\ \epsilon &= \delta-2\gamma\, y_{10}^2\,\eta_0^2, \qquad a=3-2Ay_{10}, \qquad b=A-6\gamma\, y_{10}^2, \end{aligned} \tag{49}$$

and the value y_1 of y for which $Q=0$ is then found from (45):

$$y_1^2 = z_1^2-\eta_0^2\,.$$

In the case $k=1$ the curve $P=0$ has a maximum for $\gamma=\gamma_1=+4/27$ and $y=y_1=1{\cdot}5$. The corresponding values for $Q=0$ are found to be

$$\begin{aligned} \gamma_1 &= \frac{4}{27}A^3\left(1-\frac{4}{3}A\,\delta+\frac{64}{27}A^2\delta^2-\ldots\ldots\right), \\ z_1 &= \frac{3}{2A}+\frac{4}{3}\delta-\frac{32}{27}A\,\delta^2+\ldots\ldots. \end{aligned} \tag{50}$$

The value of γ must be found by successive approximations, since A depends on γ, but the approximations converge extremely rapidly. The corresponding value of y_1 is again found from $y_1^2=z_1^2-\eta_0^2$.

27. Integrable Cases

The solutions for the case $\gamma=\gamma_1$ now become

$$\text{(C)}\quad (\tau-\tau_0)\sqrt{\gamma_1}=\cosh^{-1}\left(\frac{z+z_1}{z_1\sqrt{1-\frac{\delta}{z_1^4}}}\right)-\frac{1}{\sqrt{3+\frac{\delta}{z_1^4}}}\cosh^{-1}\left(\frac{2z+z_1+\frac{\delta}{z_1^3}}{(z-z_1)\sqrt{1-\frac{\delta}{z_1^4}}}\right),$$

(51)

$$\text{(F)}\quad (\tau-\tau_0)\sqrt{\gamma_1}=-\cosh^{-1}\left(\frac{z+z_1}{z_1\sqrt{1-\frac{\delta}{z_1^4}}}\right)+\frac{1}{\sqrt{3+\frac{\delta}{z_1^4}}}\cosh^{-1}\left(\frac{2z+z_1+\frac{\delta}{z_1^3}}{(z_1-z)\sqrt{1-\frac{\delta}{z_1^4}}}\right),$$

where γ_1 and z_1 are now taken from (50) and y is derived from z by $y^2=z^2-\eta_0^2$, and

(51*) (A) $$z=z_1,$$

and consequently $y=y_1$ as before.

For the case $\gamma=0$ we introduce the auxiliary variable x by

$$z=\frac{x-b}{1-2kb},\quad b=\delta-3k\delta^2+10k^2\delta^3-\ldots\ldots$$

Then we find instead of (41)

(52)
$$\begin{aligned}
&(H) \qquad x=\sin^2\psi,\quad \tau-\tau_0=\psi-\frac{\sin 2\psi}{2(1-2b)}\\
&(Q) \qquad \tau-\tau_0=\frac{2}{3}x^{\frac{3}{2}}-2bx^{\frac{1}{2}},\\
&(L) \qquad x=\sinh^2\psi,\quad \tau-\tau_0=\frac{\sinh 2\psi}{2(1+2b)}-\psi.
\end{aligned}$$

If β and η_0 are neglected the oscillating solution (H), as given by (41), starts at $\psi=0$ and its period is π. If β and η_0 are taken into account we must use (52), but the initial value now is not $\psi=0$, but $\psi=\psi_0$, corresponding to $y=0$, and the period consequently is $\pi-2\psi_0$. For $y=0$ we have $z=\eta_0$ and ψ_0 is thus determined by

$$\sin^2\psi_0=(1-2b)\eta_0+b.$$

Similarly the solution (L) does not start with $\psi=0$ but with $\psi=\psi_0$ given by

$$\sinh^2\psi_0=(1+2b)\eta_0+b.$$

In the solution (Q), $k=0$, we have $b=\delta=\beta$, $x=z+\beta$, so the starting value of x is $x_0=\eta_0+\beta$. If we wish to have $\tau=0$ for $y=0$, account must be taken of these values of ψ_0 and x_0 in determining the constant of integration τ_0.

The maximum value of x in solution (H) is $x_1=1$, from which

$$z_1=1+\delta-\delta^2+\ldots\ldots$$

$$y_1=1+\frac{1}{2}\eta_0^2+\beta+\ldots\ldots$$

instead of $y_1=1$. The same value of z_1 is found from (50).

28. Numerical Value of the Pressure Terms

From (33*), (25), and (26) we have (neglecting φ_0^2)

$$\beta=\frac{3p_s}{\rho_0}y\,,$$

where $p_s=p_*+p_c$. In (28) we found $3p_*/\rho_0=10^{-10}$, consequently the ordinary radiation is entirely negligible. For the cosmic radiation we found in (29) $3p_c=3\cdot6\,.\,10^{-36}$ gr. cm^{-3} and from (36*) we have $\frac{1}{\rho_0}=2\cdot5\,.\,10^{27}\,P$, consequently

$$\beta=\frac{3p_c}{\rho_0}\,y=0\cdot9\,.\,10^{-8}\,Py.$$

This is also probably negligible, although it might be appreciable if P were near the upper limit given by (37). I have, however, neglected it, and I have taken throughout

$$\beta=0.$$

For the random radial velocities of the extragalactic nebulae we can take $V_0 = 750$ km/sec $= c/400$. Assuming that these random velocities have no preference for any direction, we have

$$\varphi_0^2 = 3R^2 V_0^2/c^2$$

and consequently by (33*)

$$\eta_0^2 = \frac{3y^2 V_0^2}{c^2}$$

or roughly

$$\eta_0 = 0{\cdot}004\, y.$$

I have adopted the following values:

$$(53)\quad \begin{aligned} &\text{Expanding universes except (L):} && \eta_0 = {\cdot}02 \\ &\text{(L):} && \eta_0 = {\cdot}04 \\ &\text{Oscillating universes for } k = +1: && \eta_0 = {\cdot}004 \\ &\text{Oscillating universes for } k = 0, -1: && \eta_0 = {\cdot}01 \end{aligned}$$

29. Numerical Integration and Developments in Series

The curves for the cases (C), (F), (H), (Q), (L), given in figure 3, were computed by the formulae (51) and (52), using the values (53) of η_0, and $\beta = 0$. For the other cases numerical integration was used, by the formula

$$d\tau = \sqrt{\frac{y}{P}}\, dy$$

neglecting β and η_0^2. At the beginning of the curves, near $y = 0$, for the expanding universes of the first kind and for the oscillating universes, a development in series, however, was used. The series employed, of which the derivation may be omitted, is:

$$(54)\quad \begin{aligned} u &= z - z_0 \\ \tau - \tau_0 &= (Bu)^{\frac{1}{2}}(2z_0 + \sum_p a_p u^p) \\ B &= (1 - 2Az_0 + 4\gamma z_0^3)^{-1} \\ a_1 &= \frac{2}{3} + \frac{1}{3} ABz_0 \\ a_2 &= \frac{1}{5}AB + \frac{3}{20}A^2B^2z_0 - \frac{1}{5}B\gamma z_0^2 \\ a_3 &= \frac{3}{28}A^2B^2 + \frac{5}{56}A^3B^3z_0 - \frac{5}{7}B\gamma z_0 - \frac{6}{7}AB^2z_0^2 \\ a_4 &= \frac{5}{72}A^3B^3 - \frac{1}{9}B\gamma + \frac{35}{576}A^4B^4z_0 - \frac{5}{6}AB^2\gamma z_0 \\ a_5 &= \frac{35}{704}A^4B^4 - \frac{3}{22}AB^2\gamma + \frac{63}{1408}A^5B^5z_0 - \frac{75}{88}A^2B^3\gamma z_0 \\ a_6 &= \frac{63}{1664}A^5B^5 - \frac{15}{104}A^2B^3\gamma \\ a_7 &= \frac{77}{2560}A^6B^6 - \frac{7}{48}A^3B^4\gamma \\ &\cdots\cdots\cdots \end{aligned}$$

where z_0^3 has been neglected throughout, z_0^2 for a_4 and beyond, and z_0 for a_6 and a_7. A is given by (47), z_0 by (48), and z, as always, by (45). If the pressure is neglected we have $u=z=y$, $z_0=0$, $A=k$, $B=1$, and the formulae are much simplified.

Near the maximum of the oscillating universes and the minimum of the expanding universes of the first kind the numerical integration breaks down, and recourse must again be had to development in series. We put

$$z=z_1(1+v),\ \tau'=\frac{\tau-\tau_1}{2z_1},$$

z_1 being given by (49) and we find:

$$
\begin{aligned}
v &= C\tau'^2(1+\Sigma_p b_p \tau'^{2p}) \\
C &= 2A-\frac{3}{z_1}-\frac{4\delta}{z_1^3} \\
b_1 &= \frac{1}{3}A+\frac{2}{3}\frac{\delta}{z_1^2} \\
b_2 &= \frac{2}{5}b_1^2+\frac{1}{5}C\epsilon_1 \qquad (55) \\
b_3 &= \frac{3}{35}b_1^3+\frac{9}{35}Cb_1\epsilon_1+\frac{1}{7}C^2\epsilon_2 \\
b_4 &= \frac{2}{175}b_1^4+\frac{27}{175}b_1^2\epsilon_1+\frac{2}{75}C\epsilon_1^2+\frac{2}{7}Cb_1\epsilon_2+\frac{1}{9}C^3\epsilon_3 \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

$$\epsilon_i=(-1)^i\left[\frac{1}{z_1}+(i+3)\frac{\delta}{z_1^2}\right].$$

If the pressure is neglected the formulae are again much simplified. We have then $z=y$, $z_1=y_1$, $A=k$, $\delta=0$, $\epsilon_i=(-1)^i/y_1$.

30. EXAMPLES

The solutions that have been computed and are represented in figure 3 are the following:

	Value of γ					Empty universes		
	+0·20	γ_1	+0·10	0	−0·10	$l=+1$	$l=0$	$l=-1$
$k=+1$		C	D					
	E.	F	G	H	J	B_+		
$k=0$			P	Q	R	B_0	[N_0]	
$k=-1$			K	L	M	B_-	N_-	S

Further "Einstein's universe" (A), $y=1{\cdot}5$, is also shown. The solution (N_0), $R=$constant, is omitted in the diagram. The computations for the solutions (C) and (F) were made by the formulae (51), and for (H), (Q), and (L) by (52); for the parts of the curves near $y=0$ and near the maxima and minima of y the formulae (54) and (55) were used. These all include the pressure terms, using the values (53). The influence of the pressure terms is, however, entirely negligible in all cases on the scale of the diagram. The empty universes were computed by the formulae (44).

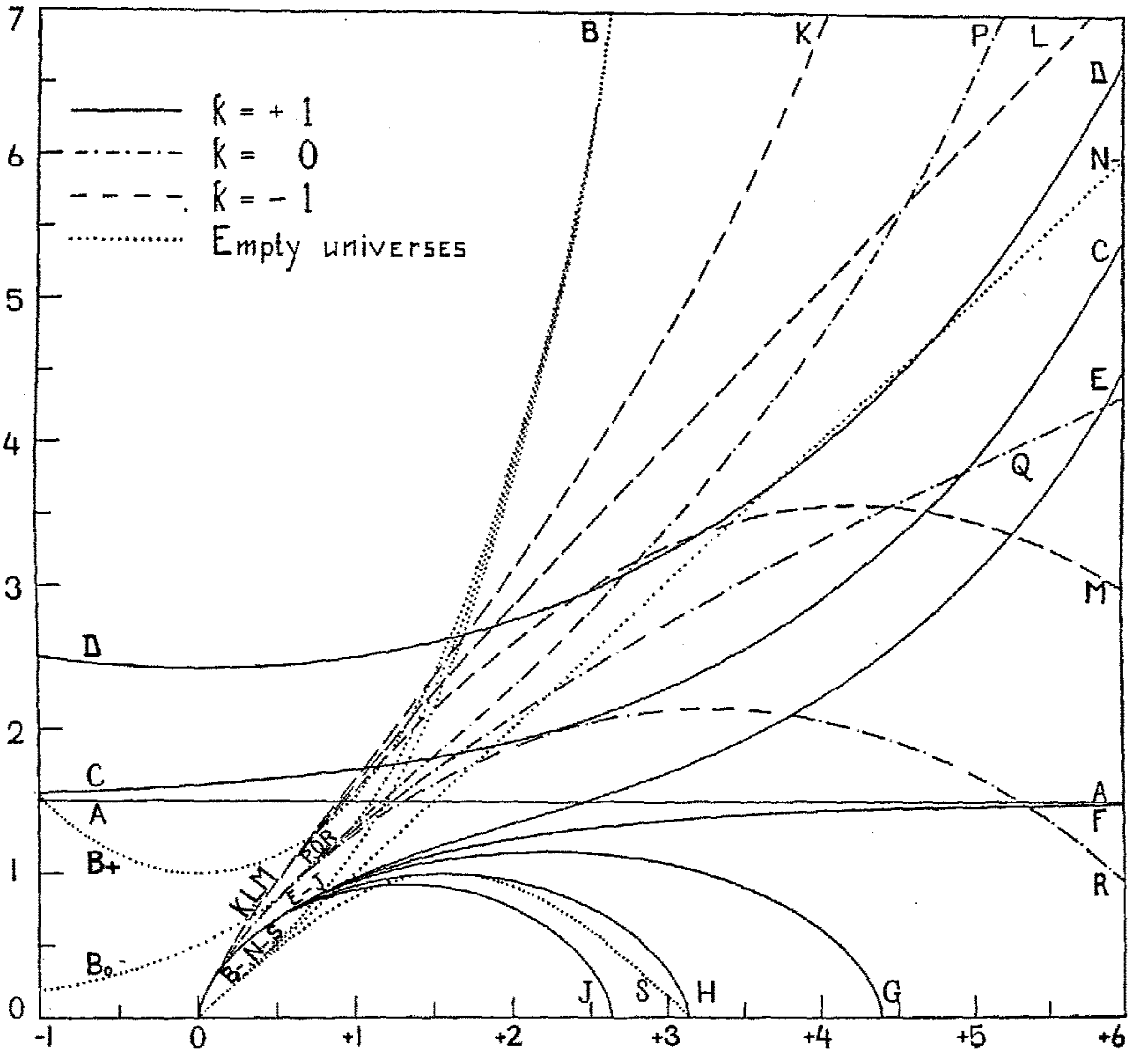

Figure 3

Expanding universes for different values of k and γ, and empty universes. The horizontal coordinate is $\tau-\tau_0$, the vertical coordinate is y ($\tau'-\tau_0'$ and y' for the empty universes).

The resulting computed numbers are given in the tables A and B.

TABLE A

Values of y and $\tau-\tau_0$ for Different Universes

y	(C) $\tau-\tau_0$	(D) $\tau-\tau_0$	(E) $\tau-\tau_0$	(F) $\tau-\tau_0$	(G) $\tau-\tau_0$	(H) $\tau-\tau_0$	(J) $\tau-\tau_0$	(P) $\tau-\tau_0$	(Q) $\tau-\tau_0$	(R) $\tau-\tau_0$	(K) $\tau-\tau_0$	(L) $\tau-\tau_0$	(M) $\tau-\tau_0$
0·0			0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000
·1			·0204	·0216	·0216	·0216	·0216	·0198	·0198	·0206	·0193	·0178	·0200
·2			·0621	·0632	·0635	·0635	·0635	·0582	·0582	·0591	·0550	·0530	·0559
·3			·1196	·1211	·1212	·1212	·1213	·1081	·1081	·1090	·0996	·0974	·1005
·4			·1925	·1942	·1946	·1946	·1949	·1669	·1671	·1682	·1503	·1482	·1514
·5			·2814	·2835	·2841	·2852	·2863	·2336	·2341	·2356	·2058	·2039	·2073
·6			·3882	·3912	·3928	·3960	·3992	·3071	·3082	·3104	·2650	·2632	·2671
·7			·5152	·5204	·5243	·5327	·5419	·3866	·3888	·3921	·3272	·3259	·3303
·8			·6650	·6756	·6850	·7070	·7344	·4714	·4753	·4805	·3918	·3913	·3966
·9			·8419	·8641	·8864	·9488	1·0699	·5609	·5675	·5757	·4584	·4586	·4655
1·0			1·049	1·097	1·154			·655	·665	·678	·527	·529	·537
1·2			1·562	1·785				·851	·875	·903	·667	·674	·686
1·4			2·168	3·361				1·058	1·103	1·162	·811	·824	·845
1·6	−0·299		2·751					1·269	1·348	1·464	·956	·979	1·011
1·8	+1·414		3·242					1·483	1·608	1·831	1·102	1·138	1·187
2·0	+2·239		3·644					1·695	1·884	2·334	1·247	1·299	1·372
2·2	2·800		3·979					1·904	2·174		1·392	1·464	1·569
2·4	3·230		4·264					2·107	2·477		1·534	1·631	1·779
2·6	3·581	1·475	4·513					2·305	2·793		1·675	1·795	2·006
2·8	3·880	2·122	4·733					2·496	3·122		1·813	1·971	2·256
3·0	4·141	2·589	4·930					2·679	3·462		1·948	2·144	2·539
3·2	4·373	2·966	5·110					2·856	3·814		2·081	2·319	2·873
3·4	4·582	3·286	5·274					3·025	4·178		2·210	2·492	3·313
3·6	4·774	3·566	5·426					3·188	4·552		2·336	2·669	
3·8	4·949	3·816	5·567					3·344	4·937		2·459	2·846	
4·0	5·112	4·043	5·699					3·494	5·332		2·579	3·024	
4·4	5·406	4·444	5·940					3·777	6·151		2·810	3·384	
4·8	5·667	4·791	6·155					4·039	7·009		3·029	3·746	
5·2	5·903	5·098	6·354					4·283	7·903		3·237	4·112	
5·6	6·116	5·374	6·529					4·510	8·833		3·435	4·480	
6·0	6·313	5·625	6·694					4·722	9·796		3·623	4·848	
8·0	7·111	6·631	7·368					5·618	15·08		4·439	6·720	
10·0	7·714	7·380	7·882					6·318	21·08		5·100	8·616	

TABLE B

Values of y and τ Near Maxima and Minima of y

(C)		(D)		(F)		(G)		(H)		(J)		(R)		(M)	
y	$\tau-\tau_0$	$\tau-\tau_0$	y	y	$\tau-\tau_0$	$\tau_1-\tau$	y	y	$\tau-\tau_0$	$\tau_1-\tau$	y	$\tau_1-\tau$	y	$\tau_1-\tau$	y
1·51	−3·801	0·0	2·4236	1·0	1·097	0·0	1·1535	0·90	0·9488	0·0	0·9217	0·0	2·1544	0·0	3·5771
1·52	−2·748	·1	2·4244	1·1	1·392	·1	1·1522	·92	1·0012	·1	·9183	·1	2·1528	·1	3·5751
1·53	−2·133	·2	2·4267	1·2	1·785	·2	1·1482	·94	1·0784	·2	·9080	·2	2·1480	·2	3·5692
1·54	−1·697	·3	2·4307	1·3	2·356	·3	1·1417	·96	1·1693	·3	·8908	·3	2·1399	·3	3·5593
1·55	−1·357	·4	2·4362	1·4	3·361	·4	1·1324	·98	1·2875	·4	·8664	·4	2·1286	·4	3·5454
1·56	−1·080	·5	2·4433	1·42	3·688	·5	1·1204	·99	1·3707	·5	·8344	·5	2·1140	·5	3·5276
1·57	− ·845	·6	2·4520	1·44	4·114	·6	1·1055	1·00	1·5704	·6	·7944	·6	2·0963	·6	3·5058
1·58	− ·641	·7	2·4624	1·45	4·385	·7	1·0876			·7	·7458	·7	2·0753	·7	3·4802
1·59	− ·461	·8	2·4744	1·46	4·716	·8	1·0666			·8	·6872	·8	2·0510	·8	3·4507
1·60	− ·299	·9	2·4880	1·47	5·144	·9	1·0422			·9	·6178	·9	2·0235	·9	3·4172
1·7	+ ·774	1·0	2·5033	1·48	5·745	1·0	1·0143			1·0	·5353	1·0	1·9928	1·0	3·3800
1·8	+1·414	1·1	2·5204	1·49	6·784							1·1	1·9588	1·1	3·3389
1·9	+1·874	1·2	2·5391	1·50	16·78							1·2	1·9216	1·2	3·2940
2·0	+2·239											1·3	1·8810	1·3	3·2454
		y	$\tau-\tau_0$												
		2·55	1·2541												
		2·60	1·4754												
		2·65	1·6648												
		2·70	1·8324												
$y_2=1·500123$				$y_1=1·500009$		$\tau_1=2·2003$		$\tau_1=1·57063$ $y_1=1·000008$		$\tau_1=1·3215$		$\tau_1=3·3110$		$\tau_1=4·2606$	

There is no observational datum which would enable us to choose between these different solutions. The data of observation are the coefficient of expansion, or R_B, for which we can provisionally adopt $R_B = 2 \,.\, 10^{27}$ cm, and the density, which is still very uncertain. Instead of the density we can take R_A, or, in connection with R_B, $P = 3R_A{}^2/2R_B{}^2 = 3h^2/\kappa\rho_0$. If once we have determined our choice of k and γ, the table (38) shows the values of P that are admissible.

When we have fixed our choice on a value of P, y is derived from

(35) $$P = 1 - ky + \gamma y^3.$$

Then R_1, which gives the scale, is found from

(56) $$R_1{}^2 = \frac{P}{y^3} R_B{}^2 = R_B{}^2 \left(\gamma - \frac{k}{y^2} + \frac{1}{y^3} \right).$$

It should be noticed that unless y is very small, the value of R_1 depends practically on R_B and γ alone.

In the solution (Q), $k = 0$, $\gamma = 0$ we have $P = 1$ and consequently $R_1{}^2 y^3 = R_B{}^2$, but R_1 and y are indeterminate. Three-dimensional space is in this case euclidean. The fact that R_1 and y, and therefore $R = R_1 y$, are indeterminate does not therefore represent any indeterminateness of the instantaneous three-dimensional space. On the other hand, we have $\tau - \tau_0 = 2y^{\frac{3}{2}}/3$ and therefore $R_1{}^2(\tau - \tau_0)^2 = 4R_B{}^2/9$, or

(57) $$c(t - t_0) = \frac{2}{3} R_B.$$

Adopting $R_B = 2 \,.\, 10^{27}$ cm this would give $t - t_0 = 1{\cdot}41 \,.\, 10^9$ years. This solution (Q) is the one of which the compatibility with our present day observational data was recently pointed out by Einstein and de Sitter[21]. The coefficient of expansion in this system is given by

$$R_B{}^2 = R_1{}^2 y^3 = \frac{9}{4} c^2 (t - t_0)^2,$$

and the density by

(58) $$\kappa\rho_0 = \frac{3}{R_B^2} = \frac{4}{3c^2(t - t_0)^2}$$

or

(58′) $$\rho_0 = \frac{0{\cdot}717 \,.\, 10^{27}}{c^2(t - t_0)^2}.$$

The value of γ_1 is $+\,4/27 = {\cdot}148148$ if the pressure is neglected. Taking the values (53) for the pressure terms we find from (50):

(C) $\eta_0 = {\cdot}02$ $\gamma_1 = {\cdot}148016$ [22]

(F) $\eta_0 = {\cdot}004$ $\gamma_1 = {\cdot}148148$.

In the expanding universes of the second kind P increases continually with the time from zero to infinity, in those of the first kind from unity to infinity, except in the case (E) where it passes through a minimum:

(E) $$P_{min} = 0{\cdot}139, \quad y_{min} = 1{\cdot}291, \quad (\tau - \tau_0)_{min} = 1{\cdot}833.$$

[21] *Proceedings* Nat. Acad. Sci. Washington, 18: 51 (1932).

[22] This value was used for the computation by (51) of the data given in the tables A and B above. The diagram, however, was drawn from a computation based on $\gamma_1 = {\cdot}15153$. The difference is negligible on the scale of figure 3.

In the oscillating solutions P decreases from unity to zero, except in the case (M) where it passes through a maximum:

(M) $$P_{max}=2·217, \quad y_{max}=1·826, \quad (\tau-\tau_0)_{max}=1·210.$$

EXPANDING UNIVERSES OF THE SECOND KIND

	k	γ	y_2	P	y	$\tau-\tau_0$	$R_1.10^{-27}$	$\frac{t-t_0}{10^9 \text{ years}}$
(C)	+1	γ_1	1·5	10	4·5	∞	0·663	∞
(D)	+1	+·10	2·4236	10	5·220	5·126	0·530	2·92

EXPANDING UNIVERSES OF THE FIRST KIND

	k	γ	P	y	$\tau-\tau_0$	$R_1.10^{-27}$	$\frac{t-t_0}{10^9 \text{ years}}$
(E)	+1	+·20	10	4·023	5·713	0·784	4·82
(P)	0	+·10	10	4·481	3·831	2/3	2·70
(Q)	0	0	1	See formulae (57), (58), (58′).			
(K)	−1	+·10	10	3·746	2·426	0·872	2·24
(L)	−1	0	10	9	7·663	0·234	1·89

OSCILLATING UNIVERSES

	k	γ	y_1	Period	P	y	$\tau-\tau_0$	$R_1.10^{-27}$	$\frac{t-t_0}{10^9 \text{ years}}$
(F)	+1	γ_1	1·5	∞	0·1	1·094	1·371	0·553	0·80
					0·5	0·521	0·304	3·762	1·21
(G)	+1	+·10	1·153	4·401	0·1	1	1·154	0·632	0·77
					0·5	0·514	0·298	3·842	1·21
(H)	+1	0	1	π	0·1	0·9	0·949	0·740	0·74
					0·5	0·5	0·285	4	1·21
(J)	+1	−·10	0·922	2·643	0·1	0·841	0·824	0·820	0·72
					0·5	0·488	0·274	4·149	1·21
(R)	0	−·10	2·154	6·622	0·1	2·080	2·632	0·211	0·56
					0·5	1·710	1·655	0·632	1·11
(M)	−1	−·10	3·577	8·521	0·1	3·541	3·834	0·095	0·38
					0·5	3·388	3·281	0·227	0·78
					1	3·162	2·809	0·358	1·06
					2	2·424	1·805	0·750	1·43
						1·153	0·651	2·284	1·58

The above table gives the data for the different solutions given in figure 3. I have taken $P=10$ for all expanding universes with the exception of (Q), of course, which has $P=1$, and for the oscillating universes I have made the computations for two values of P viz.: $P=0·1$ and $P=0·5$. In the solution (M) I have also added $P=1$ and $P=2$. The latter gives, of course, two values of y with corresponding values of $\tau-\tau_0$ and R_1, on both sides of the maximum. The corresponding densities are given by

$$(36^*) \qquad \rho_0=\frac{4\text{·}03}{P}\ .\ 10^{-28}\ \text{gr. cm}^{-3}.$$

In order to get $R=R_1y$ and $ct=R_1\tau$ in centimeters, we must multiply y and τ by R_1. It may be mentioned that 10^{27} cm $=1\text{·}058\ .\ 10^9$ light years $=3\text{·}244\ .\ 10^8$ parsecs.

31. Indeterminacy of Solution

It has already been pointed out that there is no observational evidence available which would enable us to decide which of the several possible solutions represents the actual universe. This is not because the data are not sufficiently accurate, but because they are deficient in number. In order to define any particular solution we require three things: the curve on which it lies, which is determined by k and γ, the position on the curve, given by y (or $\tau-\tau_0$), and the scale, which is determined by R_1. Astronomical observations give us only two data: the coefficient of expansion and the density, R_B and R_A or R_B and $P=3R_A{}^2/2R_B{}^2$. If P were known accurately, instead of being indeterminate within the wide limits (37), some restriction would be placed on the choice of the solution, as is evident from the table (38). Thus if it were certain that P exceeded unity, i.e. if the density were small, all oscillating universes would be excluded, except those for negative curvature and negative γ, and if P were smaller than 1, the expanding solutions of the first kind would be impossible, except for positive curvature and $\gamma>\gamma_1$. But the choice of the signs of the curvature and of λ, i.e. k and l, would still be free, and also the numerical value of γ or λ would be undeterminable.

Sir Arthur Eddington[23] has recently published a remarkable formula, linking up the numerical data referring to the universe with those referring to the electron. As published by Eddington the formula reads

$$(59) \qquad \frac{\sqrt{N}}{R}=\frac{mc^2}{e^2}=3\text{·}54\ .\ 10^{12},$$

where N is the number of protons, i.e. the number of hydrogen atoms, in the universe (or the number of protons and electrons, i.e. twice this number, Sir Arthur is not quite sure on this point), and R is the "radius of curvature of the empty space between the particles of matter." For this latter he takes the radius of the "empty universe," solution B, i.e. he takes $R=R_C$. As to N, it is taken equal to M/H or to $2M/\text{H}$, M being the total mass of the universe, and H the mass of the hydrogen atom. By $M=\pi^2R^3\rho=3\pi^2R_1/\kappa$ (the pressure being neglected) we have then

$$N=\frac{3\pi^2R_1}{\kappa\text{H}}=9\text{·}578\ .\ 10^{51}R_1,$$

or twice this value if the protons and electrons must both be counted. Eddington's formula thus becomes

$$(60) \qquad \frac{R_1}{R_C{}^2}=\zeta=1\text{·}307\ .\ 10^{-27}$$

or one-half of this value.

[23] *Proceedings* Royal Society. A. 133: 605 (1931); M. N. 92: 3 (1931). See also the Observatory, 55, : 206 (1932).

The derivation of Eddington's formula supposes, of course, a finite universe, and therefore $k=+1$. It is conceivable that only a finite number of electrons[24] (presumably depending on the number of degrees of freedom of the equations defining what an electron is) would be distinguishable one from another. If this were so there would be a physical basis for the finiteness of the universe—though not an *observational* basis, but one depending on the structure of our theory of the electron. Whether this theory will admit such a conclusion to be drawn from it I am unable to judge, but until this has been shown to be so, the assertion that the universe is finite is a pure a priori assumption, which can be based only on philosophical or metaphysical grounds.

It might also be that the formula (60), or a similar one, could be derived without involving the hypothesis of the finiteness of three-dimensional space. In that case k would remain indeterminate.

If the formula (60) were accepted, then the observed density and the coefficient of expansion, or P and R_B, would be sufficient to determine all required characteristics of the universe, *if the values of l and k, i.e. the signs of λ and the curvature, were given a priori*. We would then have given R_B, P, and ζ, and from $R_1^2y^3=PR_B^2$ we find

$$\zeta PR_B^2=lR_1\gamma y^3.$$

The equation (35) then gives, if we put $x=\sqrt{P/y}$,

$$\frac{P-1}{P}x^3+kx-l\zeta PR_B=0, \qquad (k,\ l=+1,\ 0,\ -1) \tag{61}$$

from which x could be determined, and hence y, after which R_1 would be found by (56) and then γ would be given by

$$\gamma=\frac{lR_1^2}{R_C^2}=lR_1\zeta. \tag{62}$$

The values of l and k would remain indeterminate. The value $l=0$, however, is excluded by (60), since it requires $R_C^2=\infty$.

It is easily verified that the equation (61) gives exactly the same variety of solutions as (35), according to the different values of P, k, and l. Since P can be derived only from observational data within the very wide limits given by (37), the value of x derived from (61), and hence that of y, are extremely uncertain, and the same is true of R_1 and γ. Even the relation (60) would thus, in the present state of our knowledge, not enable us to make a definite choice between the different solutions.

If we make an arbitrary decision regarding γ, then R_1 and R_C can be determined from (60) and (62). Then the coefficient of expansion is given by (19), or (34), or (56), or

$$\frac{R_1^2}{R_B^2}=R_1^2h^2=\gamma-\frac{k}{y^2}+\frac{1}{y^3}. \tag{63}$$

[24] This was suggested by Sir Arthur Eddington in a recent conversation with the writer.

Omitting the last two terms we have $h_\infty = \gamma^{\frac{1}{2}}/R_1 = 1/R_C = \zeta/\gamma^{\frac{1}{2}}$, and, except for very small values of y, the actual value of h will not differ much from h_∞. If we take, with Sir Arthur Eddington, $\gamma = +4/27$, the value for Lemaître's universe, this gives:

$$R_1 = 1\cdot133 \,.\, 10^{26},\ R_C = 2\cdot944 \,.\, 10^{26},\ h_\infty = 3140 \text{ km/sec}/10^6 \text{ ps}$$

for the value (60) of ζ, and

$$R_1 = 2\cdot266 \,.\, 10^{26},\ R_C = 5\cdot888 \,.\, 10^{26},\ h_\infty = 1570 \text{ km/sec}/10^6 \text{ ps}$$

if one-half of this value must be taken. This would be the limiting value of the coefficient of expansion for $y = \infty$ (or very large) in Lemaître's universe, if Eddington's equation (60) is adopted. For $y = 3$, i.e. twice the minimum value in Lemaître's universe, the value of h is about two-thirds of the limiting value. With other values of γ we would, of course, get different results. The observed value of h would require a considerably smaller value of h_∞ (unless y were *very* small), i.e. a larger value of γ, i.e. an expanding universe of the first kind.

I have dwelt rather long on the consequences of Eddington's formula (59), because, although at first sight it might seem to make the problem determinate by adding one more datum, on closer investigation it appears that even if it be adopted, we can decide which of the several possible solutions represents our actual universe only by making an a priori hypothesis, which is practically equivalent to the choice of a particular solution.

If Eddington's equation is not used, we have nothing to guide us, so we assume a certain γ, our choice being determined merely by personal preference. Then, since the observed density is extremely uncertain, we practically make a rough guess at it, i.e. at P, and then find y from (35) and R_1^2 from (56). The value of R_1, if y is not too small, is largely independent of the adopted value of P, and depends practically on the observed expansion and the assumed γ alone. If, on the other hand, we believe in Eddington's formula, we find that in order to be able to use it we must, on account of the same uncertainty of P, again practically assume a value of γ; R_1 then again depends on this assumed value of γ and on ζ, and P and y are found afterwards. The only thing that we have gained is that the determination of P and y now depends on Eddington's equation, instead of on the observed density.

32. The Time Scale

It will be noticed that the values of $\tau - \tau_0$ are in all cases, with the exception of (C), of the same order as y. If we multiply $\tau - \tau_0$ by $1\cdot058 \,.\, 10^{-27} R_1$ we get the time elapsed since y had its minimum value (either y_2 or zero) expressed in units of a thousand million years. This time is extremely short, being of the order of the age of the oldest rocks of the earth. In the case (C) the interval $\tau - \tau_0$ is infinite, but the infinity is only logarithmic: the time elapsed since y was exactly y_2 is infinite but the time elapsed since y was, say $1\cdot1\, y_2$ or even $1\cdot01\, y_2$, is again of the same order as in the other cases. The whole past history of the universe since it passed through the minimum of y is compressed into a very short compass. And not only the past; the same is true of the future. In the case of the oscillating universes, this is at once evident from the shortness of the periods. But also in the expanding cases, with the exception of (Q) and (L), for which $\gamma=0$, $\tau-\tau_0$ increases as log y for very large values of y. Thus, e.g. for the solutions (C), (D), and (E) we find that the values of $c(t-t_0)$ expressed in units of 10^9 years for $y=10, 100$, and 1000 are as follows:

y	(C)	(D)	(E)
10	5·41	4·14	6·53
100	9·60	8·23	10·83
1000	13·81	12·33	15·11

This shortness of the time scale is rather startling at first sight. As soon as the theory of the expanding universe became generally known, the beginning of the expansion was identified with the "beginning of the world," or, since that phrase has no definite physical meaning, with the beginning of the evolution of the stars and stellar systems. Now this identification is entirely gratuitous. Suppose the universe were one of the oscillating kind, say (M) with $P=1$ (but any other case will do just as well for my argument). Then the beginning of the expansion happened $1{\cdot}06 \,.\, 10^9$ years ago, and the maximum value of y will be reached in $0{\cdot}55 \,.\, 10^9$ years. At that epoch y will be stationary, and after that it will begin to decrease again, first slowly and then more rapidly. There is no reason to call this stationary point the "end of the world." Nor is there any reason to call the stationary point of the expanding solutions of the second kind, when y is a minimum, the "beginning of the world." It appears to me that there is, at least in these cases, no indication whatever of a direct connection between the expansion of the universe and the evolution of stars and stellar systems. The two processes are going on simultaneously, but mainly independently of each other. The question becomes more complicated for the universes of the oscillating type or of the expanding type of the first kind in the neighborhood of $y=0$. The density and the pressure then become enormous; as y approaches zero all material velocities approach the velocity of light, as will be shown in the next article, and it is impossible to say what will happen, since evidently in the limit, for $y=0$, the equations are no longer applicable.

We can, of course, easily relegate the catastrophe to the time minus infinity, by introducing another time variable, e.g. $c\mathrm{T}=\kappa \log y$ which will make $y=0$ for $\mathrm{T}=-\infty$. If for κ we take the present value of R_B we will have at the present moment $d\mathrm{T}/dt=1$. There is nothing in our experience of the physical world that would enable us to distinguish between the times T and t. We do not know which of these times it is that we use as independent variable in the equations of celestial mechanics, or by which we measure the rate of progress of radio active disintegration, or of the evolution of a star, or of any other physical process. But the infinity would again be only logarithmic, and, if the universe were of the oscillating type, or of the expanding type of the first kind, we would still have enormous densities, pressures, and material velocities at times only a few thousand million years in the past. The only conclusion we can draw from these considerations is that we are not able, with the means at present at our disposal, either theoretical or observational, to extrapolate with certainty farther back into the past (or forward into the future) than a few *hundred* million years.

33. MOTION OF MATERIAL PARTICLES

The equations of motion of a particle are the differential equations of the geodesic (II). Taking the line element (8) or

$$ds^2 = R_1^2(-y^2 d\sigma^2 + d\tau^2)$$

we find easily[25]

$$\frac{d}{ds}\left(\frac{d\sigma}{ds}\right)^2 + \frac{4}{y}\left(\frac{d\sigma}{ds}\right)^2 \frac{dy}{d\tau}\,\frac{d\tau}{ds} = 0,$$

$$\frac{d}{ds}\left(\chi^2 \mathbf{S}^2 \frac{d\theta}{ds}\right) + \frac{2}{y}\left(\chi^2 \mathbf{S}^2 \frac{d\theta}{ds}\right)\frac{dy}{d\tau}\,\frac{d\tau}{ds} = 0,$$

where $\mathbf{S} = \sin \chi\sqrt{k}/\chi\sqrt{k}$, so that $\chi^2\mathbf{S}^2 d\theta/ds$ is the expression for the "area." These equations can be integrated at once, and they give:

$$\frac{d\sigma}{ds} = \frac{\eta}{R_1 y^2}, \qquad \chi^2\mathbf{S}^2\frac{d\theta}{ds} = \frac{\omega}{R_1^2 y^2}, \tag{64}$$

η and ω being constants of integration. (The denominators R_1 and R_1^2 are introduced for later convenience.) Eliminating ds from (64) we find the differential equation of the orbit, which proves to be the equation of a geodesic in the three-dimensional space with the line element $d\sigma$. We have thus in the three cases,

$$\begin{array}{lll} k = +1: & \tan \chi = \tan \chi_0 \sec \theta, & \tan \sigma = \sin \chi_0 \tan \theta \\ k = 0: & \chi = \chi_0 \sec \theta, & \sigma = \chi_0 \tan \theta \\ k = -1: & \sinh \chi = \sinh \sigma \operatorname{cosec} \theta & \tanh \sigma = \sinh \chi_0 \tan \theta, \end{array} \tag{65}$$

where χ_0 is the minimum radius vector.

The velocity in this path, however, is not constant, as it would be if y were constant (i.e., in a static universe), but is given by the first of (64), from which, since

$$\left(\frac{d\tau}{ds}\right)^2 = \frac{1}{R_1^2} + y^2\left(\frac{d\sigma}{ds}\right)^2,$$

we find

$$\frac{d\sigma}{d\tau} = \frac{\eta}{y\sqrt{y^2+\eta^2}}\,. \tag{66}$$

Consequently $d\sigma/d\tau$ becomes infinite for $y=0$. The space of which the line element is $d\sigma$ is, however, only a mathematical abstraction, introduced to make the equations tractable and to bring out clearly the expansion. The line element of the true physical space is $Rd\sigma = R_1 y d\sigma$. Remembering that $R_1 d\tau = cdt$ we have

$$R\frac{d\sigma}{dt} = \frac{c\eta}{\sqrt{y^2+\eta^2}}, \tag{66*}$$

which for $y=0$ becomes

$$\left(\frac{Rd\sigma}{dt}\right)_0 = c\,. \tag{67}$$

If y approaches zero, all material velocities approach the velocity of light.

[25] See de Sitter, *B. A. N.* V, 193: 217 (1930), where only the case $k=+1$ is considered. For $k=0$ and $k=-1$ however the results are the same.

If we wish to investigate the actual track of a material particle, we must, however, not use (66*) but (66), which has the disadvantage that $d\sigma/d\tau$ becomes infinite for $y=0$. Combining with (33) we find, however,

$$\frac{d\sigma}{dy}=\frac{\eta}{\sqrt{y^2+\eta^2}}\cdot\frac{1}{\sqrt{Q}}, \tag{68}$$

where

$$Q=\beta+\sqrt{y^2+\eta_0{}^2}-ky^2+\gamma y^4$$

remains finite for $y=0$. The pressure quantity $\eta_0{}^2$ in Q is an average of the values of η^2 for all material particles in the universe. For $y=0$ we have thus

$$\left(\frac{d\sigma}{dy}\right)_0=\frac{1}{\sqrt{\beta+\eta_0}},$$

which, though large, is not infinite. It should be noticed that it is independent of the individual η's, and consequently the same for all particles, whether they are slow- or quick-moving for ordinary values of y. We can thus carry out the integration of (68) which must in all cases be done numerically. Then, choosing an arbitrary value for χ_0, and taking $\sigma=0$ for $\chi=\chi_0$, we can construct the track by means of (65). Projecting on euclidean space by means of the formulae (17), and introducing rectangular coordinates, we have for the three cases

$$(69)\qquad \begin{aligned}
k=+1:&\quad \begin{aligned}\mathrm{x}&=\mathrm{r}\cos\theta=R_1\,y\tan\chi_0\\ \mathrm{y}&=\mathrm{r}\sin\theta=R_1\,y\sec\chi_0\tan\sigma\end{aligned}\\
k=\ \ 0:&\quad \begin{aligned}\mathrm{x}&=\mathrm{r}\cos\theta=R_1\,y\,\chi_0\\ \mathrm{y}&=\mathrm{r}\sin\theta=R_1\,y\,\sigma\end{aligned}\\
k=-1:&\quad \begin{aligned}\mathrm{x}&=\mathrm{r}\cos\theta=R_1\,y\sinh\chi_0\cosh\sigma\\ \mathrm{y}&=\mathrm{r}\sin\theta=R_1\,y\sinh\sigma\end{aligned}
\end{aligned}$$

Some examples have been computed and are represented in figure 4. I have taken rather large values of χ_0 and large peculiar velocities η, so as to bring out the characteristic features of the trajectories near $y=0$. The adopted data are:

I :	Universe (L), $k=-1$, $\eta_0=\cdot04$	$\eta=\cdot02$, $\sinh\chi_0=0\cdot1$ $\sigma=0$ for $y=0\cdot10$
II :	Universe (E), $k=+1$, $\eta_0=\cdot02$	$\eta=\cdot02$, $\tan\chi_0=0\cdot1$ $\sigma=0$ for $y=0\cdot01$
III :	Universe (Q), $k=0$, $\eta_0=\cdot02$	$\eta=\cdot02$, $\chi_0=-0\cdot2$ $\sigma=0$ for $y=0$
IV :	Universe (Q), $k=0$, $\eta_0=\cdot02$	$\eta=\cdot04$, $\chi_0=0\cdot1$ $\sigma=0$ for $y=0\cdot10$
V :	Universe (D), $k=+1$, $\eta_0=\cdot02$	$\eta=\cdot02$, $\tan\chi_0=0\cdot005$ $\sigma=0$ for $y=2\cdot50$

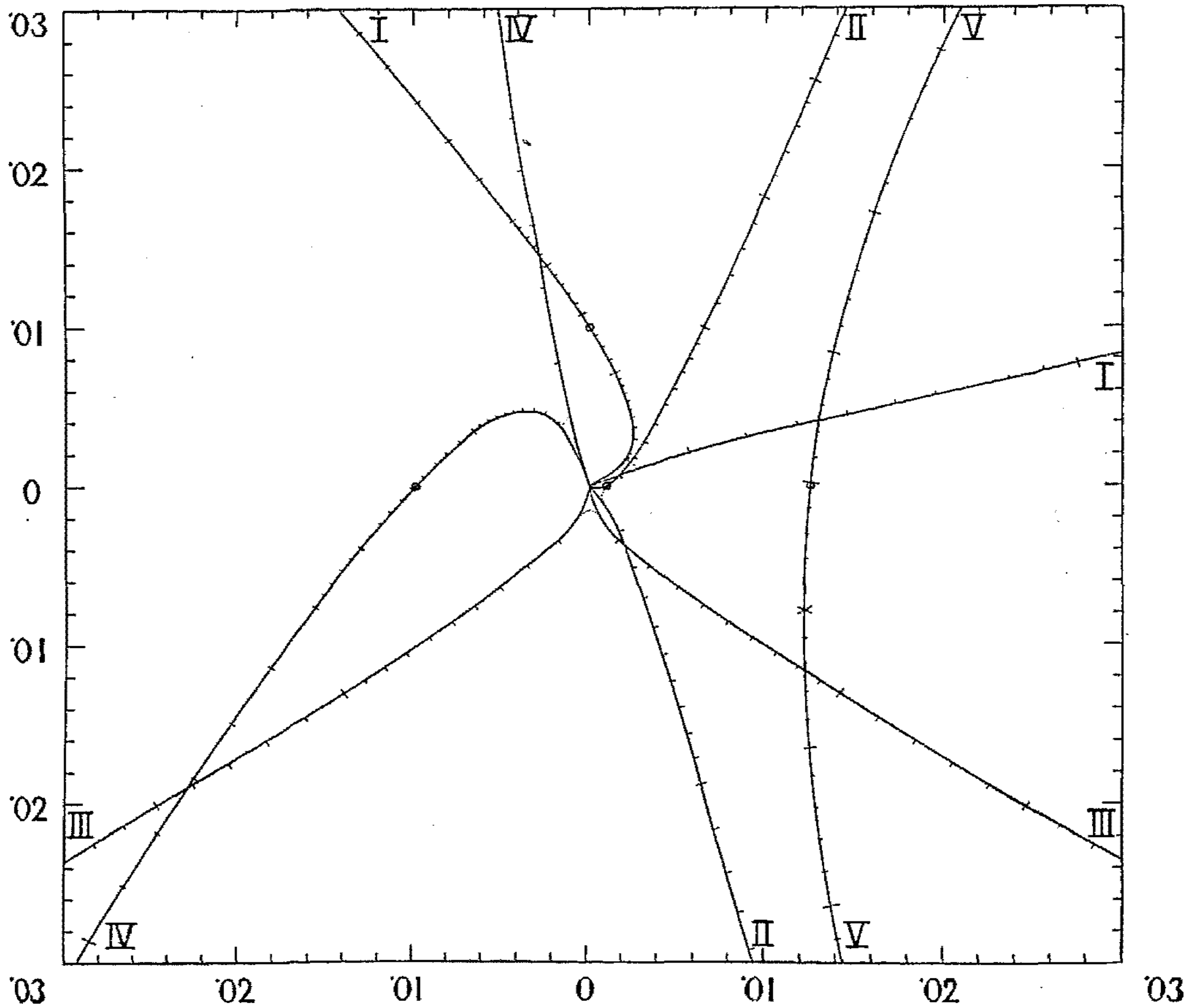

Figure 4

Some typical trajectories of material points in expanding universes of the first kind: I in (L), II in (E), III and IV in (Q), and of the second kind: V in (D). The unit of length and time is the value of R_1 for each universe. The time marks along the curves I to IV are at a distance of 0·001 up to 0·010, then at 0·002 up to 0·040, and from there at 0·01 up to 0·10. The time $\tau=\tau_0$ is at the origin in each of these cases. In curve V the time marks are at a distance of 0·20 throughout, and the point $\tau=\tau_0$ is marked by a cross. The points $\sigma=0$ $(\chi=\chi_0)$ are marked by small circles

The trajectories have been drawn to scale, the value of R_1 being taken as unit in each case, and the scale is marked on all four sides of the diagram. The orientation, i.e. the zero of position angle $\theta=0$, was differently chosen for each curve. The zero point of σ, i.e. the point where $\chi=\chi_0$, is indicated on each curve by a small circle, except in the case III where it was taken in the origin, i.e. coinciding with the moment when $y=0$. The values of the time $ct=R_1\tau$ have also been indicated along the curves, the points marked by short bars being, for the cases I to IV those for $\pm(\tau-\tau_0)=0$, ·001, ·002, . . . ·009, ·010, ·012, ·014, . . . ·038, ·040, ·05, ·06 . . . ·09 and ·10. The multiples of $\pm$·010 are distinguished by the bar extending to both sides of the curve. The increase of the velocity as y approaches zero is very well shown by these time marks. Of course all trajectories are, so to say, sucked into the origin at the time $\tau=\tau_0$ when $y=0$.

In case V the times marked are $\pm(\tau-\tau_0)=0$, 0·2, 0·4 . . . 4·0, 4·2. The time $\tau=\tau_0$ is marked by a cross instead of a bar. It will be seen that in this case the increase of the velocity near the minimum value of y is not noticeable,[26] and neither the point $\tau=\tau_0$ nor the point $\sigma=0$, $\chi=\chi_0$ is at all different from any other points on the trajectory.

We have throughout used the line element

$$ds^2=-R^2d\sigma^2+c^2dt^2 \tag{8}$$

with

$$d\sigma^2=\Sigma_{p,q}\,\gamma_{pq}\,d\xi_p\,d\xi_q,$$

because it is mathematically convenient for the description of the observed expansion. But the coordinates in natural measure, corresponding to the galilean line element that an observer naturally uses to describe the phenomena in his neighborhood, are not ξ_i but $x_i=R\xi_i$[27], or their projections on euclidean space x and y as given by (69). We need only consider the radius vector, so we can take, for the sake of argument, $d\chi=d\sigma$. The observed radial velocity is thus

$$\frac{dr}{cdt}=y\frac{d\sigma}{d\tau}+r\frac{\dot R}{R}=\frac{\eta}{\sqrt{y^2+\eta^2}}+rh. \tag{70}$$

The second term is by far the preponderating term, and it represents the observed recession of the spiral nebulae. The formula (70) is the most convenient for the purpose of comparing the values of dr/dt of different objects at the same time, i.e. for the same values of y and h, but different values of r and η, but it is not convenient for the integration. If we wish to determine the track of one particle, it is better to perform the transformation to galilean coordinates after the integration, as has been done in the present article. In the case of the expanding universes of the second kind these trajectories are curves of a hyperbolic nature, such as V in the diagram, very similar to the hyperbola described by a test body in the quasi-static solution B.[28] It should be noted that the point of nearest approach to the origin in this case is not necessarily the point corresponding to $\tau=\tau_0$, for which y has its minimum value y_2, but depends upon χ_0 and η, and will be reached at different times by different objects.

The statement that all material bodies, if abstraction is made of their mutual interaction, describe trajectories of this nature, is equivalent to saying that the universe expands in the manner required by the particular solution to which these trajectories correspond. If this solution is an expanding solution of the second kind, nothing very exceptional will happen at the time of minimum y. If it is an expanding solution of the first kind, or one of the oscillating universes, the parts of the trajectories nearest the origin are of the nature of the curves I to IV of figure

---

[26] In fact there is a slight *decrease* of the velocity in the projection on euclidean space, as the effect of the projection in magnifying the scale away from the origin exceeds the effect of the change in the real velocity.

[27] Strictly speaking we should say $dx_i=Rd\xi_i$. Transformation to galilean coordinates is possible only locally, and $Rd\xi_i$ is not a complete differential.

[28] See Appendix.

4, all reaching the origin at the same time $\tau=\tau_0$. But evidently in this case the mutual interactions cannot be neglected, and in actual nature the parts of the orbits near the zero point must be very different from those represented in the figure.

It should be noted, however, that, although in the idealized case—neglecting the mutual interaction—all bodies reach the origin at the same epoch, they approach it with very different velocities, and the perturbations of the orbits by the mutual gravitation between them will begin to be appreciable at very different times for different pairs of bodies. The simultaneity will thus be destroyed: the shortest mutual distances between different pairs of bodies will not all be reached at the same time, nor will the bodies pass exactly through the origin. It is of course impossible to say exactly what will happen, or has happened, but evidently it cannot have been entirely without influence on the development of the stars and galactic systems. I think the effects of this influence can still be traced.

The spirals and our own galactic system are all rotating with periods of the order of a few hundred million years. They are all very unhomogeneous in structure, consisting of condensations, or star clouds, separated by regions of smaller density. If the rotation had been going on undisturbed for a great many revolutions, this unhomogeneity could not subsist. But if only a small number of revolutions (of the order of ten) has been completed since a strong perturbation occurred, the unhomogeneity is of comparatively recent date, and has not yet had time to be smoothed out. Also the spiral structure itself is most readily explained as an effect of tidal forces resulting from a near approach. If, however, we compute the frequency of near approaches of spiral nebulae on the basis of their average peculiar random motions, and average distances apart at the present time, taking no account of the change of size of the universe, we find that they should be very rare, the time between encounters being more nearly of the order of 10^{12} years, instead of 10^9.

Also it is a significant coincidence that the minimum value of R occurred at about the date of the birth of the planetary system. Modern theories ascribe the origin of the planets to a near approach, or even a collision, of the sun and another star. Evidently the chances that such a collision should occur were very much larger at the epoch of minimum size of the universe than they are now.

It does not follow, however, that the minimum value of y in our universe must have been zero. It is quite possible that the density at the minimum of an expanding universe of the second kind will be large enough to increase the frequency of encounters sufficiently to produce these effects. There is no escape from the conclusion that we do not know which of the possible solutions represents the actual universe.

APPENDIX

ON STATIC AND QUASI-STATIC (EMPTY) UNIVERSES

Although the static universes are, so to say, of only academic interest, it is perhaps worth while briefly to consider their chief characteristic features.

34. Existence of Static Universes and Transformation to Non-static Ones

We take again the line element

$$ds^2 = -R^2 d\sigma^2 + v^2 dt^2 \tag{8}$$

where now we suppose R to be a constant and v a function of the space coordinates only. We suppose the three-dimensional line element $d\sigma$ to have spherical symmetry

$$d\sigma^2 = d\chi^2 + b(d\psi^2 + \sin^2\psi\, d\theta^2), \tag{71}$$

where b is a function of χ only, and, on account of the spherical symmetry, v is also a function of χ only. Since we do not wish the origin of coordinates to be an exceptional point, we have to add the further condition that the three-dimensional space shall have the constant curvature k, and we can limit the values of k to $+1$, 0, and -1. Then we have

$$\begin{aligned} k=+1 &: \quad b = \sin^2\chi \quad \frac{b'}{b} = 2\cot\chi \\ k= 0 &: \quad b = \chi^2 \quad \frac{b'}{b} = \frac{2}{\chi} \\ k=-1 &: \quad b = \sinh^2\chi \quad \frac{b'}{b} = 2\coth\chi. \end{aligned} \tag{72}$$

In each case the invariant G of the line element $d\sigma$ is

$$_{(3)}G = -\frac{2}{b} - \frac{1}{2}\frac{b'^2}{b^2} + \frac{2b''}{b} = -6k,$$

differential quotients $d/d\chi$ being denoted by accents.

Further, we have

$$G_{11} = {}_{(3)}G_{11} + \frac{v''}{v}, \quad G_{22} = {}_{(3)}G_{22} + \frac{1}{2}b'\frac{v'}{v}, \quad G_{33} = G_{22}\sin^2\psi,$$

$$G_{44} = -\frac{vv''}{R^2} - \frac{vv'}{R^2}\frac{b'}{b}.$$

We take again the energy tensor

$$T_{pq} = -g_{pq}p, \quad T_{p4} = T_{4p} = 0, \quad T_{44} = g_{44}\rho, \quad T = \rho_0 = \rho - 3p. \tag{9}$$

Then the field equations

$$G_{\alpha\beta}-\lambda g_{\alpha\beta}+\kappa(T_{\alpha\beta}-\frac{1}{2}g_{\alpha\beta}T)=0 \tag{I}$$

become

$$\begin{aligned} &{}_{(3)}G_{11}+\frac{v''}{v}+R^2[\lambda+\frac{1}{2}\kappa(\rho-p)]=0 \\ &{}_{(3)}G_{22}+\frac{1}{2}b'\frac{v'}{v}+R^2b[\lambda+\frac{1}{2}\kappa(\rho-p)]=0, \end{aligned} \tag{73}$$

$$\frac{v''}{v}+\frac{b'}{b}\cdot\frac{v'}{v}+R^2[\lambda-\frac{1}{2}\kappa(\rho+3p)]=0. \tag{74}$$

Since the three-dimensional space $d\sigma$ must have the constant curvature k we must have

$${}_{(3)}G_{pq}+2k\gamma_{pq}=0,$$

from which we find, by comparison with (73) the conditions

$$\begin{aligned} &\frac{v''}{v}=\frac{1}{2}\frac{b'}{b}\frac{v'}{v}=q \\ &q+R^2[\lambda+\frac{1}{2}\kappa(\rho-p)]=2k, \end{aligned} \tag{75}$$

q being a constant. Substituting in (74) we find

$$q=-\frac{1}{3}\,R^2[\lambda-\frac{1}{2}\kappa(\rho+3p)],$$

and consequently we have

$$\begin{aligned} &\lambda+\kappa\rho=\frac{3k}{R^2} \\ &\kappa(\rho+p)=\frac{2(k+q)}{R^2}; \end{aligned} \tag{76}$$

which can be compared with the equations (20) for the non-static universes, and of which (11A) and (11N) are special cases.

The equations (75) determine v. In the different cases we have by (72)

$k=+1$: $\quad \frac{v'}{v}=q\tan\chi, \qquad \frac{v''}{v}=q+(q^2+q)\tan^2\chi.$

Since also $v''/v=q$, we must have $q^2+q=0$. Consequently there are two possibilities:

$q=\ \ 0$: $v=c$ (solution A) $\qquad \kappa(\rho+p)=\frac{2}{R^2},$

$q=-1$: $v=c\cos\chi$ (solution B) $\qquad \kappa(\rho+p)=0.$

$k=0$: $\quad \frac{v'}{v}=q\chi, \qquad \frac{v''}{v}=q+q^2\chi^2.$

The only possibility is

$$q=0 : v=c \text{(solution N)} \qquad \kappa(\rho+p)=0$$

$$k=-1 : \qquad \frac{v'}{v}=q \tanh \chi, \quad \frac{v''}{v}=q+(q^2-q)\tanh^2\chi.$$

We have again two possibilities, viz.:

$$q=1 : v=c\cosh\chi \quad \text{(solution S)} \quad \kappa(\rho+p)=0,$$

$$q=0 : v=c \qquad \kappa(\rho+p)=-\frac{2}{R^2}.$$

This last solution thus gives a negative density. It can therefore be dismissed as impossible in nature, and the solution (A) is the only possible static solution with a finite positive density. The solutions (B), (N), and (S) have $\rho+p=0$, and are thus "empty universes." All of these are only quasi-static, and can be transformed to non-static solutions, in which R is a function of the time, while $v=c$ is constant.

If we denote by $d\sigma_k$ the line element of a three-dimensional space of constant curvature k ($k=+1$, 0, or -1), the condition that a transformation shall be possible such as

$$-d\sigma_k^2+v^2dt^2=-y^2d\sigma_l^2+d\tau^2,$$

v being a function of the radius vector χ only, and y a function of the time τ only, is easily found to be (a dot denoting differentiation with respect to τ)

$$\dot{y}^2-ky^2+l=0, \tag{77}$$

which gives for y the values of the non-static solutions

(B_+), (B_0), (B_-), (N_0), (N_-), and (S).

For v we find then, independent of l,

$$k=+1 : \quad v=\cos\chi$$

$$k=0 : \quad v=1$$

$$k=-1 : \quad v=\cosh\chi,$$

agreeing with the static forms (B), (N), and (S) of the solutions. Comparing (77) with the equation for the non-static universe (in which we interchange l and k)

$$\dot{y}^2-ky^2+l+\frac{a}{y}=0 \tag{43}$$

we find that the transformation is possible for empty universes $\alpha=0$ only. This is equivalent to saying that the transformation of a static into a non-static solution is possible only if the *four*-dimensional space-time has constant curvature. This requires $G_{\alpha\beta}+3\epsilon' g_{\alpha\beta}=0$, from which by the equations given above we find

$$3\epsilon'=-\left[\lambda+\frac{1}{2}\kappa(\rho-p)\right]=-\left[\lambda-\frac{1}{2}\kappa(\rho+3p)\right],$$

giving $\kappa(\rho+p)=0$, i.e. an empty universe. The curvature of four-dimensional space-time is found to be

$$\epsilon'=-\frac{1}{3}(\lambda+\kappa\rho)=-\frac{k}{R^2}.$$

The actual transformations are easily found. We put, for brevity,

$$d\varphi^2=d\psi^2+\sin^2\psi\, d\theta^2.$$

Then

(B): $$ds^2=-(d\chi^2+\sin^2\chi\, d\varphi^2)+\cos^2\chi\, dt^2$$

is transformed as follows:

$$\sin\chi=\cosh\tau\sin\rho,\qquad \tanh t=\tanh\tau\sec\rho$$

(B_+) $$ds^2=-\cosh^2\tau(d\rho^2+\sin^2\rho\, d\varphi^2)+d\tau^2,$$

$$\sin\chi=r\, e^u,\qquad t=u+\lg\sec\chi$$

(B_0) $$ds^2=-e^{2u}(dr^2+r^2\, d\varphi^2)+du^2,$$

$$\sin\chi=\sinh\nu\sinh\omega,\ \tanh t=\tanh\nu\cosh\omega$$

(B_-) $$ds^2=-\sinh^2\nu(d\omega^2+\sinh^2\omega\, d\varphi^2)+d\nu^2.$$

(N): $$ds^2=-(d\chi^2+\chi^2\, d\varphi^2)+dt^2$$

gives

$$\chi=r,\qquad t=u$$

(N_0) $$ds^2=-(dr^2+r^2d\varphi^2)+du^2,$$

$$\chi=\nu\sinh\omega,\qquad t=\nu\cosh\omega$$

(N_-) $$ds^2=-\nu^2(d\omega^2+\sinh^2\omega\, d\varphi^2)+d\nu^2.$$

(S) : $$ds^2=-(d\chi^2+\sinh^2\chi\, d\varphi^2)+\cosh^2\chi\, dt^2$$

is transformed as follows:

$$\sinh\chi=\sin\nu\sinh\omega,\qquad \tan t=\tan\nu\cosh\omega$$

(S) $$ds^2=-\sin^2\nu\,(d\omega^2+\sinh^2\omega\, d\varphi^2)+d\nu^2.$$

We find thus again all the empty universes of art. 25, and no others.

35. Motion of Material Particles

In order to investigate the motion of a material particle in the different static universes A, B, N, and S, we have to construct the equations of the geodesic. The integration of the first three of these gives the equation of energy and the equation of areas. Then the equation of the track is found by elimination of ds from these two. The fourth equation of the geodesic, giving d^2t/ds^2, is in all cases easily integrated, enabling us to replace ds by dt in the equation for the velocity or for the area, and, by integration of the resulting equation, to derive the expression for the coordinates as functions of the time.

For the cases A and B the results have been worked out long ago.[29] In A the track is a geodesic in the space $d\sigma$ which is described with uniform velocity.

In B it is best to use the projection on euclidean space by the formulae (17), taking $r=R\tan\chi$. The line element is

$$ds^2=-\frac{dr^2}{\left(1+\frac{r^2}{R^2}\right)^2}-\frac{r^2(d\psi^2+\sin^2\psi\, d\theta^2)}{1+\frac{r^2}{R^2}}+\frac{dt^2}{1+\frac{r^2}{R^2}}. \tag{78}$$

The equation for the track is found to be

$$\left(\frac{dr}{d\theta}\right)^2=\frac{r^2(r^2-a^2)(r^2+b^2)}{a^2b^2},$$

which gives the hyperbola

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1,$$

and we have

$$x=a\cosh u,\quad y=b\sinh u,\quad u=R(t-t_0).$$

The coordinates x and y thus increase continually with the time: the universe is expanding.

The radial velocity is given by

$$\left(\frac{dr}{dt}\right)^2=\frac{r^2}{R^2}\left(1-\frac{a^2}{r^2}\right)\left(1+\frac{b^2}{r^2}\right). \tag{79}$$

The semi-axes of the hyperbola are $a=r_0$, $b=Rv_0$, r_0 being the minimum distance from the origin, and $v_0=r_0(d\theta/dt)_0$ the velocity at that point.

Since in the line element (78) g_{44} differs from unity, there will be a displacement of all spectral lines toward the red amounting to

$$\left(\frac{\lambda_1-\lambda}{\lambda}\right)_1=\frac{1}{2}\,\frac{r^2}{R^2}. \tag{80}$$

Superimposed on this will be the Doppler effect due to the velocity (79). In 1917, when the solution B was first discovered, it was not realized that the velocity dr/dt, of which only the square is determined by (79), would always be positive, and it was thought that this Doppler effect would not be systematic, the red shift (80) being the only systematic effect. The velocities are, however, all positive; in other words, all observable bodies are on the receding branches of their respective hyperbolas, having passed the apex long ago, so that none remain on the approaching branches. The Doppler effect is a first-order effect, being proportional to r/R. Since (80) is a second-order term, we must compute the Doppler effect by the rigorous formula correct to the second order. We then find for a velocity q:

$$\left(\frac{\lambda_1-\lambda}{\lambda}\right)_2=q-\frac{1}{2}q^2,$$

[29] de Sitter, M. N. lxxviii: 14-19 (1917).

and, taking $q = dr/dt = \mathrm{r}/R$, we find that the red shift (80) is exactly cancelled, leaving only the linear effect

$$\frac{\lambda_1 - \lambda}{\lambda} = \frac{\mathrm{r}}{R},$$

in exact agreement with the result from the theory of the expanding universe.

In the case N we have the ordinary Newtonian mechanics: the inertial track of a particle is a geodesic in euclidean space described with constant velocity.

Finally, in the case S, of which the line element, if we use again the projection on euclidean space, is

$$ds^2 = -\frac{d\mathrm{r}}{\left(1+\frac{\mathrm{r}^2}{R^2}\right)} - \mathrm{r}^2(d\psi^2 + \sin^2\psi\, d\theta^2) + \left(1+\frac{\mathrm{r}^2}{R^2}\right)dt^2,$$

we find the equation of the track

$$\left(\frac{d\mathrm{r}}{d\theta}\right)^2 = \frac{\mathrm{r}^2(\mathrm{r}^2 - a^2)(b^2 - \mathrm{r}^2)}{a^2 b^2},$$

which is the differential equation of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

and we have again $a = \mathrm{r}_0$, $b = R\mathrm{v}_0$.

The equation of areas, $\mathrm{r}^2 d\theta/ds = \text{constant}$, is satisfied, as in all other cases. The expression for the radial velocity becomes rather complicated. It is

$$\left(\frac{d\mathrm{r}}{dt}\right)^2 = \frac{\mathrm{r}^2}{R^2}\left(1-\frac{a^2}{\mathrm{r}^2}\right)\left(\frac{b^2}{\mathrm{r}^2}-1\right)\frac{\left(1+\frac{\mathrm{r}^2}{R^2}\right)^2}{\left(1+\frac{a^2}{R^2}\right)\left(1+\frac{b^2}{R^2}\right)}.$$

The orbit being an ellipse the universe is of the oscillating kind, which is verified by the transformation to the non-static form given above.