14

Inertia and Cosmology in Einstein's Relativity

(An Attempt at a Synthesis of Newton, Mach, and Einstein)

J. V. NARLIKAR

Tata Institute of Fundamental Research
Bombay, India

Nature does not begin with elements, as we are obliged to begin with them. It is certainly fortunate for us, that we can, from time to time, turn aside our eyes from the overpowering unity of the All, and allow them to rest on individual details. But we should not omit, ultimately to complete and correct our views by a thorough consideration of the things which for the time being we left out of account.

Ernst Mach
The Science of Mechanics

1. NEWTON, MACH, AND EINSTEIN

The concept of inertia, which originated with Galileo, found a mathematical expression in Newton's laws of motion. According to the second law of motion, the force acting on a particle is proportional to the acceleration of the particle. The constant of proportionality measures the inertia of the particle and is called its inertial mass. Thus, if $\mathbf{P}$ is the force, $\mathbf{f}$ the acceleration, and m the inertial mass, we have

$$\mathbf{P} = m\mathbf{f} \tag{1.1}$$

However, when formulating this law, Newton was faced with a fundamental difficulty: How to fix the reference frame relative to which $\mathbf{f}$ is measured? Clearly, (1.1) cannot hold in *all* reference frames. If it is valid in frame $\mathscr{F}$ then in another frame $\mathscr{F}'$ with an acceleration $\mathbf{a}$ relative to $\mathscr{F}$, the law of motion becomes

$$\mathbf{P}' \equiv \mathbf{P} - m\mathbf{a} = m\mathbf{f}' \tag{1.2}$$

where $\mathbf{f}'$ is the acceleration of the particle with respect to $\mathscr{F}'$, and $\mathbf{P}'$ is the force measured in $\mathscr{F}'$. Thus the force is modified from $\mathbf{P}$ to $\mathbf{P} - m\mathbf{a}$. The extra term $-m\mathbf{a}$ that has to be added to $\mathbf{P}$ depends on the mass of the particle and is hence called the *inertial force*. The frames in which (1.1) holds *without* the necessity of such a modification are called inertial frames. All these frames (including $\mathscr{F}$) are unaccelerated relative to one another.

What distinguishes inertial frames from noninertial ones? A priori Newton could see no physical reason to make such a distinction. That such a distinction exists in nature can be seen or demonstrated in numerous ways. Newton has discussed the so-called bucket experiment. The rotation of a bucket is a relative term. If one observer sees the bucket rotating, another (e.g., one sitting on the bucket) can claim that it is nonrotating. In Newton's experiment the distinction can be made absolute. If the bucket is suspended by a string, the string is given a twist and then the bucket let go, it spins as the twisted string unwinds. If the bucket contains some water, its surface will become curved and dip toward the center. This absolute effect demonstrates the presence of inertial forces and distinguishes between inertial and noninertial frames.

Why such a distinction should exist, Newton could not understand. However he used this distinction to postulate so-called *absolute space*. This is, in effect, a specific inertial frame in which (1.1) holds. All accelerations relative to this are detectable by the inertial forces.

The advent of special relativity did not change the logical status of the second law of motion. It is true that the concept of absolute space was done away with by Einstein. But the special status of inertial frames remained. Einstein himself was well aware of this special status. Indeed, he has given an analogy (Einstein, 1949) with the apparently special role of the vertical direction at any point on the Earth. Men with very limited experience of their surroundings would think that the Earth is flat and that the vertical direction has some absolute status when compared to horizontal ones. A wider knowledge of the universe would correct this impression and clarify the true status of the vertical. In the same way, will the mystery behind the special status of inertial frames be cleared up by a better understanding of the structure of the universe?

Mach, in the last century, indicated that such a clue is indeed provided by the universe: a clue that was not available to Newton. Astronomical observations have shown that the frame of reference of the local observer, in which the distant parts of the universe appear to be nonrotating, is an inertial frame. This is a remarkable result! An observer can measure the rotation of the Earth relative to Newton's absolute space by observing the motion of a Foucault pendulum that is driven by the inertial forces. Or, he can measure the rotation relative to distant galaxies.[1] In either case he gets the same answer! In other words, cosmology appears to provide a handle on the question of why certain frames enjoy a special status.

This was the point made by Mach in his critique of Newtonian mechanics. As one who believed in formulating all scientific deductions on directly observable quantities (and not on abstract theories), Mach strongly attacked the concept of absolute space; and sought to replace it by a background space of the remote parts of the universe. He went even further. The Newtonian concept of *inertia* and its measure in terms of *mass* were to him unsatisfactory. If mass is the quantity of matter in a body how does one set about measuring it? To Mach, mass and inertia were not the intrinsic properties of the body but the consequences of the existence of the body in a universe containing other matter. To measure mass one has to use (1.1): measure the force and divide it by the acceleration produced. But (1.1) itself depends on the use of absolute space, which has now been identified with the background space of distant matter. So, according to Mach's reasoning, mass is somehow determined by the distant matter.

How? The answer was not given by Mach. His reasoning, given above, is usually known as Mach's principle. Even though it is not precisely formulated it holds a certain intellectual appeal. How did Einstein react to it? Coming as it did in the formative years of the theory of relativity, Mach's principle was bound to play an important role in Einstein's approach to space–time and gravitation.

Einstein expressed his dislike of absolute space in this way:

> "absolute space" as originally determinative was quite explicitly introduced by Newton as the omnipresent active participant in all mechanical events; by "absolute" he obviously means uninfluenced by the masses and by their motion. What makes this state of affairs appear particularly offensive is the fact that there are supposed to be infinitely many inertial systems, relative to each other in uniform translation, which are supposed to be distinguished among all other rigid systems . . . (Einstein, 1949, p. 27).

Einstein was no doubt impressed by Mach and has paid tribute to him in several places. At one time he hoped that general relativity could be shown to incorporate Mach's principle in some form or other. This hope was not realized—as we shall see in the following section. Indeed, Einstein later turned away from the Machian concepts. In his own words,

> Mach conjectures that in a truly rational theory inertia would have to depend upon the interaction of the masses, precisely as was true for Newton's other forces, a conception which for a long time I considered as in principle the correct one. It presupposes implicitly, however, that the basic theory should be of the general type of Newton's mechanics: masses and their interaction as the original concepts. The attempt at such a solution does not fit into a consistent field theory, as will be immediately recognized . . . (Einstein, 1949, p. 29).

Einstein was objecting here to the concept of action at a distance implicit in

[1] In Mach's time the existence of galaxies other than our own was not established. The rotation was measured against the background of distant stars in our galaxy. The background of galaxies is known to provide a better inertial frame than that of these stars, which are considerably nearer to us.

Mach's principle. Newtonian gravitation was an action-at a-distance theory. Electromagnetism began as an action-at-a-distance theory and proved inadequate to describe all the observed phenomena including radiation. Einstein was impressed by Maxwell's field theory, not only because it successfully described electromagnetism but also because it had internal beauty. The theory already contained the Lorentz invariance required by special relativity. Action at a distance, on the other hand, was *instantaneous* and hence inconsistent with special relativity. Although, unlike electromagnetism, there were no experimental data against Newtonian gravitation (with the esoteric exception of the perihelion precession of Mercury!) its logical inconsistency was already apparent. And Mach's principle seemed to imply a similar idea of action at a distance.

Nevertheless, it is worth asking now, on the birth centenary of Einstein, whether the Machian ideas are really irreconcilable with a theory of gravitation. Cannot the Newtonian action at a distance be reformulated to give expression to Mach's ideas without offending relativity? We shall try to answer this question in this chapter. We begin by discussing the areas of conflict between general relativity and Mach's principle. We then turn our attention to the action at a distance approach. First, we see how this approach can be made to work in electromagnetism. We then formulate a theory of gravity on similar lines, which also incorporates Mach's ideas, and demonstrate how this theory approaches general relativity in the case of many particle systems. Finally we discuss two areas where the two theories differ.

2. GENERAL RELATIVITY AND MACH'S PRINCIPLE

We consider here the extent to which Mach's principle has been accommodated in general relativity. The discussion is on two fronts: observational and conceptual.

2.1. Observational Aspects

One of the early successes of general relativity was in cosmology. Side by side with Hubble's discovery of nebular redshift (Hubble, 1929) came the models of relativistic cosmology (Friedmann, 1922) that could explain these observations in terms of the expanding universe. These models can be described by the Robertson–Walker line element

$$ds^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \tag{2.1}$$

Here (r, θ, ϕ) are the constant comoving coordinates of a typical galaxy (treated as a point!); t is the cosmic time, which serves also as the proper time of each galaxy; and $S(t)$ is the time-dependent scale factor. Thus, $S(t)$, increasing with t, implies expansion. The parameter k takes the values 0, 1, or -1 and denotes the sign of curvature of the subspaces t = const. Models with $k = +1$ describe "closed" universes, while models with $k = 0, -1$ are open-universe models. Einstein's field equations determine the form of $S(t)$.

The transformation given by

$$R = rS(t), \qquad T = F(x)$$
$$X = \int_a^r \frac{x\,dx}{1 - kx^2} + \int_b^t \frac{dx}{S(x)\,dS(x)/dx} \qquad (a, b = \text{const}) \tag{2.2}$$

changes (2.1) to

$$ds^2 = e^\nu dT^2 - e^\lambda dR^2 - R^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{2.3}$$

with

$$e^{-\lambda} = 1 - kr^2 - r^2\dot{S}^2, \quad e^\nu = (1 - kr^2)S^2\dot{S}^2F_1^{-2}e^\lambda \tag{2.4}$$

Here $\dot{S}(t) = dS/dt$, $F_1(x) = dF/dx$. Choosing the arbitrary function F so that at $R = 0$, $e^\nu = 1$, and making a power series expansion near $R = 0$ leads to

$$ds^2 \cong \left(1 - R^2\frac{\ddot{S}}{S}\right) dT^2 - \left(1 + \frac{kR^2}{S^2} + \frac{\dot{S}^2R^2}{S^2}\right) dR^2 - R^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{2.5}$$

The $\cong$ implies that powers of R higher than the second are neglected.

In any "local" experiment $R \to 0$ and the line element tends to that of special relativity. In other words, an observer at $R = 0$ can use this line element to represent a local inertial frame. However, if he surveys the universe using (2.3) he will discover that distant galaxies are nonrotating (i.e., have constant θ, ϕ). In other words, these cosmological models obtained from Einstein's equations meet the observational requirement that led Mach to propose his ideas.

Nevertheless this agreement does not mean that general relativity necessarily contains Mach's principle. In the collection of papers published in the Reviews of Modern Physics to commemorate Einstein's 70th birthday, a paper by Gödel (1949) denied just such a claim. By way of a counterexample, Gödel produced the following model of a homogeneous spinning universe:

$$ds^2 = dt^2 + 2e^{x^1}dt\,dx^2 - (dx^1)^2 + \tfrac{1}{2}e^{2x^1}(dx^2)^2 - (dx^3)^2 \tag{2.6}$$

It has density $\rho = (8\pi)^{-1}$ and a λ term $= \frac{1}{2}$ in geometric units. The significance of Godel's model was that it has distant galaxies (with constant x^1, x^2, x^3) as rotating with respect to any inertial frame used by a typical observer (with constant x^1, x^2, x^3). In this sense the solution is *anti-Machian*.

For a time it was thought that an anti-Machian solution like this was possible because of the nonzero λ term in Einstein's equations. The λ term is not universally accepted. Its introduction had been prompted by reasons that are no longer considered cogent. Assuming that it should not be there, should we be able to get rid of Gödel-type solutions? Alas, no! It has been possible to find solutions of Einstein's equations *without* the λ term that are anti-Machian in the above sense (Ruzmaikina and Ruzmaikin, 1969). So long as such solutions cannot be ruled out, it is incorrect to argue that general relativity uniquely reproduces the observational background to Mach's principle. In retrospect, it is the symmetry of the line element (2.1) that, above anything else, is responsible for the Machian result. A suggestion that this symmetry could be achieved by a continuous creation of matter was made by Hoyle and Narlikar (1963). We shall not discuss that approach here.

2.2. Conceptual Aspects

Returning to the second law of motion (1.1) consider what happens when we have a single particle moving under no forces in an otherwise empty universe. We then have

$$m\mathbf{f} = 0 \qquad (2.7)$$

If we follow the Newtonian prescription for inertia we deduce from this

$$\mathbf{f} = 0 \qquad (2.8)$$

i.e., the particle moves with a uniform velocity. Assuming that the absolute space of Newton is not really an ad hoc concept but has some physical origin, in the present case it becomes difficult to understand (2.8). How is this frame of reference determined in an empty space through which the particle is moving? Rather, we should have expected that in the absence of any physical background the motion of the particle should be indeterminate.

Just such a conclusion can in fact be drawn from (2.7) by taking the alternative solution:

$$m = 0 \qquad (2.9)$$

Then $\mathbf{f}$ is indeterminate, as required. However, (2.9) has more serious implications. It means the abandonment of the Newtonian concept that inertia is the property of matter and mass a measure of it. Somehow, m must now depend on the background, in such a way as to become zero when the background is nonexistent. *This* is the Machian standpoint.

In general relativity, an empty universe is described by the Einstein field equations

$$R_{ik} = 0 \qquad (2.10)$$

It is possible to find solutions of these equations that are everywhere well behaved. Timelike geodesics in such solutions correspond to (2.8). In the absence of a background such solutions also become meaningless. Here again it would be tempting to look for the other alternative represented by (2.9). However, the framework of general relativity does not allow it.

A second conceptual conflict concerns measurement. If masses do depend on the background, i.e., on the large-scale structure of the universe, then the possibility of their variation from one space–time point to another cannot be denied. Now, masses play a fundamental role in determining units. For example, if we put c (the speed of light) and $\hbar$ (= Planck constant/2π) equal to 1 in deference to relativity and quantum theory, all units can be expressed as powers of mass. Thus length is $\sim m^{-1}$, the constant of gravitation G is $\sim m^{-2}$, the electric field is $\sim m^{2}$, and so on. We may, for instance, choose the mass of a stable elementary particle like the proton as fixing the base unit. However, we cannot guarantee that this unit will be the same everywhere, once we admit the variation of mass from point to point. To take account of this variation we have to proceed in the following way.

Suppose at one space–time point $P(x^i)$ we construct the line element

$$ds^2 = g_{ik}(P)\, dx^i\, dx^k \tag{2.11}$$

We cannot attach an absolute sense to $|ds|$ at P, since we cannot be sure of an absolute value of our base unit. We can compare lengths in different directions at P without difficulty, however. Thus we determine ratios of $g_{ik}(P)$ rather than their absolute values. Thus, instead of (2.11) we may have

$$d\bar{s}^2 = \Omega^2(P) g_{ik}(P) dx^i\, dx^k \tag{2.12}$$

without altering these ratios, where $\Omega(P) \neq 0$. It seems more reasonable therefore to formulate a physical theory that only depends on the ratios instead of the absolute values of the metric tensor. Such a theory is said to be *conformally invariant*.

That is, a conformally invariant theory does not change its structure if the space–time metric undergoes a conformal transformation:

$$g_{ik} \rightarrow \Omega^2 g_{ik} = \bar{g}_{ik}, \qquad 0 < \Omega < \infty \tag{2.13}$$

where Ω is an arbitrary (suitably well behaved) function of the space–time coordinates.

The Maxwell field equations are conformally invariant. However, general relativity is not. The constancy of mass and the absolute character of the line element (2.11) in this theory are therefore in conflict with Mach's principle.

On the face of it this conflict seems insurmountable, and suggests that a rapprochement between general relativity and Mach's principle is impossible. In the following sections we discuss how, with suitable modifications, a theory of gravitation can be constructed from a Machian origin, which leads to general relativity for most practical purposes. The approach to this theory is via action at a distance, which we first describe in the context of electromagnetic theory.

3. THE ABSORBER THEORY OF RADIATION

In a letter to Weber on March 19, 1845, Gauss wrote:

> I would doubtless have published my researches long since were it not that at the time I gave them up I had failed to find what I regarded as the keystone, *Nil actum reputans si quid superesset agendum*, namely, the derivation of the additional forces—to be added to the interaction of electrical charges at rest, when they are both in motion—from an action which is propagated not instantaneously but in time as is the case with light. . . .

Gauss's attempts came some three decades before the Maxwellian field theory and six decades before special relativity. The success of these two theories shifted the emphasis from action at a distance to fields and it was not until well into the present century that the problem posed by Gauss was solved.

A beginning was made by Schwarzschild (1903), Tetrode (1922), and Fokker (1929a,b, 1932), who independently formulated the concept of *delayed* action at a distance. The action principle as formulated by Fokker may be written in the following form:

$$\mathscr{A} = -\sum_a \int m_a\, da - \sum_{a<b}\sum \int\int e_a e_b\, \delta(s_{AB}^2)\eta_{ik}\, da^i\, db^k \tag{3.1}$$

In the above expression the charged particles are labeled $a, b, \ldots$ with e_a and m_a the charge and mass of particle a. The worldline of a is given by the coordinate functions $a^i(a)$ of the proper time a. The space–time is Minkowskian, so that

$$da^2 = \eta_{ik}\, da^i\, da^k \tag{3.2}$$

with $\eta_{ik} = \mathrm{diag}(-1, -1, -1, 1)$. The first term of $\mathscr{A}$ therefore describes the inertial term. The second term describes the electromagnetic interaction between the worldlines of a typical pair of particles $\{a, b\}$. The delta function shows that the interaction is effective only when s_{AB}^2, the invariant square of distance between typical world points A, B on the worldlines of a and b, vanishes. This implies delayed action: $s_{AB}^2 = 0$ means that A and B are connected by a light ray.

Although this formulation met the requirement of relativistic invariance it gave rise to other difficulties. The major difficulty is as follows (see Fig. 1). For a typical point A on the worldline of a there are *two* points B_+ and B_- on the worldline of b for which $s_{AB}^2 = 0$. The effect of A is felt at B_+ (at a later time) and at B_- (at an earlier time). Similarly, since the action principle guarantees the equality of action and reaction, the reaction from B_+ and B_- is felt at A. Thus there are influences propagating with the speed of light, not only into the future *but also into the past*. This led to a conflict with the principle of causality, which seems to hold in everyday life. The other difficulties were of a less serious nature although not ig-

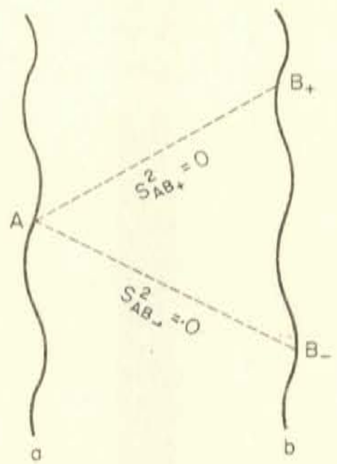

Fig. 1. A on the worldline of particle a receives a retarded reaction from B_- and an advanced reaction from B_+.

norable. For example, there was no "self-action" ($a = b$ is avoided in the double sum) and so there did not appear to be any obvious way of accounting for radiation damping.

These difficulties were removed by Wheeler and Feynman (1945) by bringing into the discussion the important role of the *absorber*.[2] In Fig. 1, the reactions from B_- arrive at A instantaneously, whatever the spatial separation of a and b. So it becomes necessary to take into account the reaction from the entire universe to A. Although the remote particles are expected to contribute less, their total number is large enough to make the calculation nontrivial. The essence of the argument given by Wheeler and Feynman is described below.

To begin with, define the 4-potential at X due to particle b by

$$A_i^{(b)}{}_{(X)} = e_b \int \delta(s_{XB}^2)\eta_{ik}\, db^k \tag{3.3}$$

and the corresponding direct-particle field by

$$F_{ik}^{(b)} = A_{k;i}^{(b)} - A_{i;k}^{(b)} \tag{3.4}$$

A direct-particle field is different from an ordinary field because it does not have any independent degrees of freedom. The 4-potential *identically* satisfies the relations

$$A^{(b)k}{}_{;k} = 0, \qquad \Box A^{(b)k} = \eta^{mn} A^{(b)k}{}_{;mn} = 4\pi j^{(b)k} \tag{3.5}$$

where

$$j^{(b)k}{}_{(X)} = e_b \int \delta_4(X, B)\, db^k \tag{3.6}$$

is the current density vector. Thus although (3.5) resembles the Maxwell wave equation (and the gauge condition) it represents identities.

The equation of motion of a typical charge a is obtained by varying its world-line and requiring $\delta\mathscr{A} = 0$. We get

$$m_a \frac{d^2 a^i}{da^2} = e_a \sum_{b \neq a} F^{(b)i}{}_k \frac{da^k}{da} \tag{3.7}$$

Thus charge a is acted on by all *other* charges in the universe.

We now turn to the difficulty introduced by the time symmetry of this formulation. Instead of being the retarded solution of (3.5), (3.3) is the time-symmetric half-advanced and half-retarded solution. The same applies to the direct-particle fields. Suppressing the indices i, k, we may write

$$F^{(b)} = \tfrac{1}{2}\{F_{\text{ret}}^{(b)} + F_{\text{adv}}^{(b)}\} \tag{3.8}$$

As shown in Fig. 2, this field is present in the past as well as the future light cone of B.

Wheeler and Feynman argued in the following way. If we move the charge b,

[2] The formal aspects of this work were discussed by the authors in 1949 in an article in *Reviews of Modern Physics*, written on the occasion of Einstein's 70th birthday (Wheeler and Feynmann, 1949).

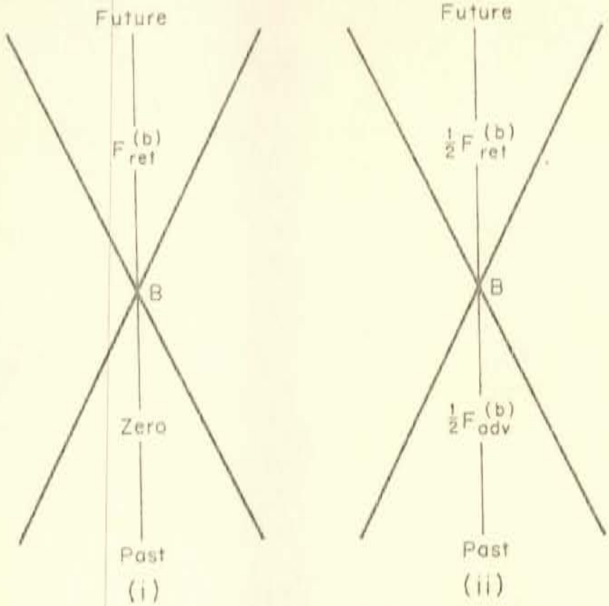

Fig. 2. (i) The full retarded field that is observed in practice. There is zero field in the past light cone of B. (ii) The time-symmetric field of the direct-particle theory is shown.

it generates a disturbance that affects all other charges in the universe. Their reaction arrives back instantaneously. Wheeler and Feynman showed how to calculate such a reaction in a universe of static Minkowski type with a uniform distribution of electric charges. They showed that the reaction to the motion of charge b can be calculated in a consistent fashion and comes out to be

$$R^{(b)} = \tfrac{1}{2}\{F_{\rm ret}^{(b)} - F_{\rm adv}^{(b)}\} \tag{3.9}$$

Thus a test particle in the neighborhood of charge b experiences a net "field"

$$F_{\rm tot}^{(b)} = F^{(b)} + R^{(b)} = F_{\rm ret}^{(b)} \tag{3.10}$$

This is the pure retarded field observed in real life! The self-consistency of the argument follows from the fact that the reaction $R^{(b)}$ has been calculated by adding the $\frac{1}{2}F_{\rm adv}^{(a)}$ fields of all particles $a \neq b$ that have been excited by this total field $F_{\rm ret}^{(b)}$. Thus only the future light cone of B comes into play (see Fig. 3). The reaction from the future cancels the advanced component of $F^{(b)}$ and doubles its retarded component.

Also, according to (3.7), $R^{(b)}$ is the force experienced by the particle b. This is nothing but the radiative reaction to the motion of b as obtained earlier by Dirac (1938) on empirical grounds. Thus the theory not only gets round the problem of causality but it also accounts for the radiation damping formula.

Physically, what happens is the following. To the motion of b the future half of the universe acts as an absorber. It "absorbs" all the "energy" radiated by b, and in this process sends the reaction $R^{(b)}$, which does the trick! For this reason

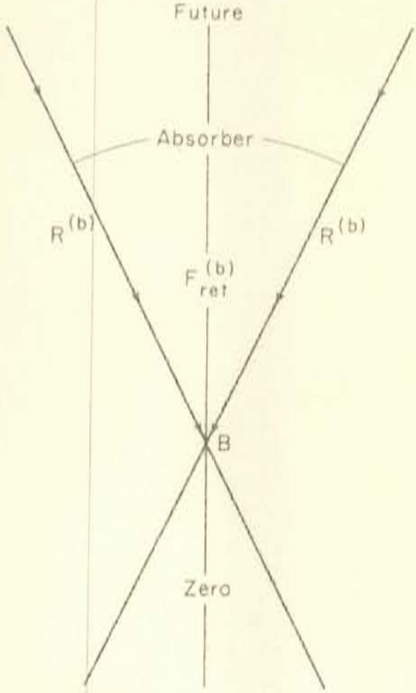

Fig. 3. In the Wheeler–Feynman theory the reaction from the future $R^{(b)}$ is generated by the absorber in response to the full retarded field $F^{(b)}_{ret}$. This field is the sum of $\frac{1}{2}F^{(b)}_{ret}$ of b and $\frac{1}{2}F^{(b)}_{ret}$ from $R^{(b)}$. The advanced part from $R^{(b)}$ cancels the $\frac{1}{2}F^{(b)}_{adv}$ field of b.

Wheeler and Feynman called this theory the *absorber theory of radiation*. The presence of the absorber is essential for the calculation to work.

3.1. The Response of the Universe

In the above self-consistent derivation there was one defect: it was not unique. Another self-consistent picture was possible in which the next field near every particle was the pure *advanced* field and the radiative reaction was of opposite sign to that of (3.9).

In Fig. 4 the two solutions are compared. In (a) we have the retarded solution while in (b) we have the advanced solution. In (a) the absorption in the future light cone is responsible while in (b) it is the absorption in the past that plays the crucial role. The important role of the absorbers is that they convert the time-symmetric situation of Fig. 2 to time-asymmetric ones of Fig. 4. It is, however, not possible to distinguish between Fig. 4a and Fig. 4b unless some other time asymmetry is involved. Wheeler and Feynman realized this and linked the choice of (a) to thermodynamics. Given the usual thermodynamic time asymmetry, they argued that the situation (b) would be highly unlikely (under the probability arguments of statis-

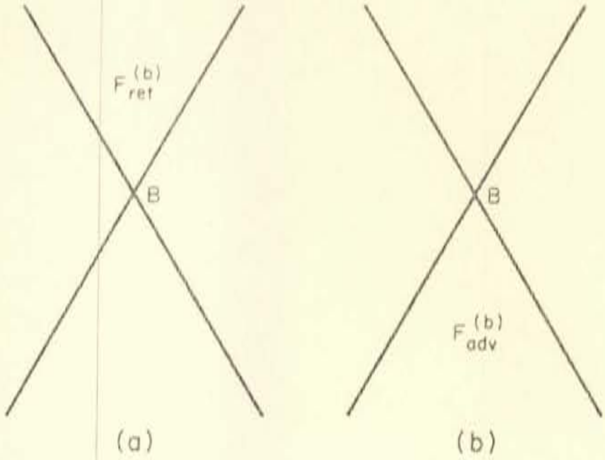

Fig. 4. The Wheeler–Feynman theory could not distinguish between (a) and (b) in a static universe, except through thermodynamics. The cosmological arrow of time is needed to make this distinction.

tical mechanics) and that the usual asymmetry of initial conditions will favor (a) to (b).

It was pointed out by Hogarth (1962) that it is not necessary to bring thermodynamics into the picture at all. If one takes account of the fact that the universe is expanding, its past and future are naturally different. The reaction from the absorbing particles in the future light cone (designated by Hogarth collectively as the *future absorber*) does not automatically come out equal and opposite to the reaction from the past absorber. Thus the two pictures (a) and (b) do not always follow in an expanding universe. Hogarth found that for (a) to hold but not (b), the future absorber must be *perfect* and the past absorber *imperfect;* and vice versa for (b) to hold but not (a).

An absorber is perfect if it entirely absorbs the radiation emitted by a typical charge. In the static universe discussed by Wheeler and Feynman both the past and future absorbers are perfect; and this leads to the ambiguity mentioned earlier. However, Hogarth found that the ambiguity is resolved if the cosmological time asymmetry is taken into account. He found, for example, that in most big-bang models (Hogarth, 1962) which expand forever, (b) is valid and not (a). In the steady-state model (Bondi and Gold, 1948; Hoyle, 1948), (a) is valid and not (b). In the big-bang models that expand and contract both the absorbers are perfect and the outcome is ambiguous.

Later Hoyle and Narlikar (1963) completed this work by first rewriting the Fokker action (3.1) in curved space as is necessary for cosmological discussion. They also deduced conclusions similar to Hogarth's under more general assumptions. Finally *they extended the entire picture to quantum theory*. Thus it is possible to discuss the entire range of phenomena of quantum electrodynamics without recourse to field theory (Hoyle and Narlikar, 1969, 1971). This therefore removes any possible objection to the concept of action at a distance in so far as it is applicable to electrodynamics.

The crucial role played in the whole calculation is that of the *response* of the universe. In the classical calculation the steady-state universe generates the "correct" response so that the local electric charges interact through retarded signals. The response from the big-bang models is of the wrong type. We are thus able to distinguish between the different cosmological models and decide on their validity or otherwise on the basis of the Wheeler–Feynman theory. We also see why charges interact through retarded signals: they do so because of the response of the universe. In the Maxwell field theory the choice of retarded solutions of Maxwell's equations is by an arbitrary fiat.

In the quantum calculation also it can be shown that the asymmetric phenomenon with respect to time, like the spontaneous downward transition of an atomic electron, is caused by the response of the universe. By contrast, the quantization of the Maxwell electromagnetic field ascribes these asymmetries to the so-called *vacuum* and to the rules of quantization.

The direct-particle approach therefore achieves for electrodynamics what Mach sought to achieve for inertia. By bringing in the response of the universe to a local experiment in electrodynamics we have essentially incorporated Mach's principle into electromagnetic theory. Mach's quotation at the beginning of this chapter applies to the present case: given the correct response of the universe, we can almost decouple our local system from it, although strictly speaking the theory would not be possible without the universe.

Can the same prescription be applied to inertia and gravitation? We discuss this problem in the following section.

4. INERTIA AS A DIRECT-PARTICLE FIELD

We now return to the problem of achieving a "reconciliation" between general relativity and Mach's principle. To this end we shall look for a theory with the following properties:

(a) It has Mach's principle built into one of its postulates.

(b) It is conformally invariant.

(c) It does not have the conceptual difficulties associated with the case of a single particle in an otherwise empty universe.

(d) For a universe containing many particles the theory reduces to general relativity for most physical situations.

We begin by a second look at the Fokker action for electrodynamics, this time rewritten in a curved Riemannian space–time:

$$\mathcal{A} = -\sum_a \int m_a \, da - \sum_{a<b}\sum 4\pi e_a e_b \iint \bar{G}_{i_a i_b} da^{i_a} db^{i_b} \tag{4.1}$$

Here, in going from (3.1) to (4.1) the first term of $\mathcal{A}$ needs a trivial modification: *da* is now computed with a Riemannian metric. The modification of the second term of $\mathscr{A}$ requires considerable thought. The $\delta(s_{AB}^2)\eta_{ik}$ is now replaced by $\bar{G}_{i_A i_B}$, a bivector propagator between A and B. It is the symmetric Green's function for the wave equation

$$\Box \bar{G}_{nl_B} + R_n^l \bar{G}_{ll_B} = [-\bar{g}(X, B)]^{-1/2} \bar{g}_{nl_B} \delta_4(X, B) \tag{4.2}$$

Here $\bar{G}_{nl_B}$ behaves as a vector at X and B, respectively, with the indices n and l_B (the subscript X on n is suppressed for the convenience of writing). $\bar{g}_{nl_B}$ is the parallel propagator between X and B [see Synge (1960) for details] and $\bar{g}(X, B)$ its determinant. In the limit $g_{ik} \to \eta_{ik}$, $\bar{G}_{i_A k_B} \to \delta(s_{AB}^2)\eta_{ik}$. The detailed structure of this propagator has been studied by DeWitt and Brehme (1960).

The electromagnetic part of $\mathscr{A}$ is conformally invariant but the mechanical part (the first term) is not. We now compare (4.1) with the action for field theory of Maxwell and for general relativity. This action, denoted by $\mathscr{A}^{(F)}$ is given by

$$\begin{aligned}\mathscr{A}^{(F)} = \frac{1}{16\pi G}\int R(-g)^{1/2}\, d^4x - \sum_a \int m_a\, da \\ - \frac{1}{16\pi}\int F_{lm}F^{lm}(-g)^{1/2}\, d^4x - \sum_a e_a \int A_i\, da^i.\end{aligned} \tag{4.3}$$

The third and fourth term of $\mathscr{A}^{(F)}$ represent, respectively, the free-field term and the field–particle interaction term.

In the direct-particle theory the second term of $\mathscr{A}$ replaces these two terms of the field theory. The fields as such lose their independent status and are replaced by propagators connecting particle world lines. What can we do about the first two terms of (4.3)? The second term already exists in (4.1) and it is tempting to simply insert the first term into (4.1) as representing gravity.

This procedure, however, is contrary to the direct-particle picture. The first term of (4.3), although containing geometrical information, has also the character of a field. Hence it is out of place. We have already commented on the non-Machian character of the second term of (4.3). For these reasons the approach suggested above is not desirable.

The clue to the correct procedure that needs to be adopted is provided by a comparison of the last term of (4.3) with the second term of (4.1). If in the former we replace the potential A_i by a sum over the direct-particle potentials defined by a relation analogous to (3.3) for a curved space, we shall recover something that looks like the latter! In the same way we now replace the masses m_a by direct-particle fields defined in the following manner:

$$m^{(b)}(X) = \int \lambda_b \bar{G}(X, B)\, db, \qquad \lambda_b = \text{coupling const} \tag{4.4}$$

$$m_a(A) = \lambda_a \sum_{b \neq a} m^{(b)}(A), \qquad \lambda_a = \text{coupling const} \tag{4.5}$$

The propagator $\bar{G}(X, B)$ has to be *biscalar* since masses are scalars and we wish to preserve a symmetry between X and B. The action (4.1) is now changed to

$$\mathscr{A} = -\sum_{a<b}\sum \iint \lambda_a \lambda_b \tilde{G}(A, B)\, da\, db$$

$$- \sum_{a<b}\sum 4\pi \iint e_a e_b \bar{G}_{i_A i_B}\, da^{i_A} db^{i_B} \tag{4.6}$$

What should be the exact form of $G(A, B)$? Taking a clue from electromagnetism, we expect it to be a symmetric Green's function of a scalar wave equation. However, we also want $\mathscr{A}$ to be conformally invariant. These two requirements fix the form of the scalar propagator *uniquely to within a multiplicative factor*. We shall take $\tilde{G}(X, B)$ to satisfy the scalar wave equation

$$\Box \tilde{G}(X, B) + \tfrac{1}{6} R(X)\tilde{G}(X, B) = [-g(X, B)]^{-1/2}\delta_4(X, B) \tag{4.7}$$

The wave operator is uniquely fixed by the requirement of conformal invariance.

Turning from these purely formal aspects to those of interpretation we note that (4.4) and (4.5) are essentially Machian ideas on inertia expressed mathematically. The mass m_a of particle a at its world point A is the sum of the contributions of all other particles in the universe. Thus requirement (a) has been met. Requirement (c) is also met, because for a single particle in an otherwise empty universe there is no action! The minimum number of particles required to define $\mathscr{A}$ is two. Thus for each of the two particles the other provides the "background" in the Machian sense. The requirement of conformal invariance is also met by our choice of the propagator. It therefore remains to examine requirement (d).

So far we have concentrated on inertia and ignored gravity. The action (4.6) does not contain the gravitational term

$$\frac{1}{16\pi G}\int R(-g)^{1/2}\, d^4x$$

explicitly. Yet, as we shall see in the following section, the theory is fully capable of describing gravitational phenomena.

5. CONFORMAL GRAVITY

Returning to the action (4.3) we note that when we try to derive the Einstein field equations by the Hilbert action principle, we get the Einstein tensor from the first term. This term does not exist any more in the direct particle action (4.6). Shall we get any gravitational term at all from (4.6) if we sought to perform the metric variation $g_{ik} \to g_{ik} + \delta g_{ik}$? A look at the electromagnetic part of (4.3) does not inspire confidence that the answer to this question should be in the affirmative. There it is the third rather than the fourth term that contributes the energy tensor of electrodynamics, and it is the fourth term that was used in going over to (4.6). Nevertheless, a closer examination shows that the terms in (4.3) do give nontrivial answers when the metric variation is performed.

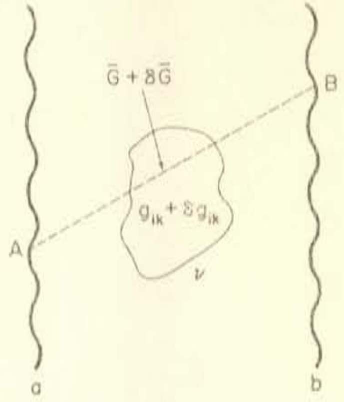

Fig. 5. The propagator $\bar{G}$ (indices suppressed) changes to $\bar{G} + \delta\bar{G}$ because the space–time in $\mathcal{V}$ has a changed geometry: $g_{ik} \to g_{ik} + \delta g_{ik}$.

The reason for this is illustrated with the example of Fig. 5. Here we have the electromagnetic propagator $\bar{G}_{i_A i_B}$ connecting A and B, respectively, on the world lines of a and b. Suppose we perform a variation in the space–time metric of a compact region $\mathcal{V}$. Since the propagator is a global property of space–time structure, it will change because of this change in structure of $\mathcal{V}$. The change in $\bar{G}_{i_A i_B}$ is therefore expressible, in a first-order calculation, as a volume integral over $\mathcal{V}$.

In the electromagnetic case the answer may be expressed in the following form:

$$-\delta \sum_{a<b}\sum 4\pi \iint e_a e_b \bar{G}_{i_A i_B}\, da^{i_A} db^{i_B} = -\tfrac{1}{2} \int T^{ik} \delta g_{ik} (-g)^{1/2} d^4x \qquad (5.1)$$

where

$$T^{ik} = \frac{1}{8\pi} \sum_{a<b} \left\{ \frac{1}{2} g^{ik} F^{(a)mn}_{\ \ \text{ret}} F^{(b)}_{mn\ \text{adv}} - F^{(a)i}{}_{l\ \text{ret}} F^{(b)kl}_{\ \ \text{adv}} - F^{(b)i}{}_{l\ \text{ret}} F^{(a)kl}_{\ \ \text{adv}} \right\} \qquad (5.2)$$

The details of this derivation are given by Narlikar (1974).

It is interesting to note in passing that this derivation resolves an ambiguity about the energy tensors of direct-particle electrodynamics. Wheeler and Feynman (1949) had discussed two tensors for this theory. Of these one was the canonical tensor given above by (5.2) and the other was the Frenkel tensor:

$$T^{ik} = \frac{1}{8\pi} \sum_{a \neq b}\sum \left\{ \frac{1}{2} g^{ik} F^{(a)mn} F^{(b)}{}_{mn} - F^{(a)il} F^{(b)k}{}_{l} - F^{(b)il} F^{(a)k}{}_{l} \right\} \qquad (5.3)$$

By $F^{(u)}_{mn}$, etc., we mean here the time-symmetric (half-advanced + half-retarded) direct-particle field defined by (3.4). Both the tensors (5.2) and (5.3) are manifestly time symmetric and both reproduce the same electrodynamic forces when their divergence is taken. Wheeler and Feynman had concluded:

> From the standpoint of pure electrodynamics it is not possible to choose between the two tensors. The difference is of course significant for the general theory of relativity, where energy has associated with it a gravitational mass. So far we have not attempted to discriminate between the two possibilities by way of this higher standard.

As mentioned above, the usual prescription of metric variation uniquely yields the canonical tensor. The facts that we could get a nontrivial answer to the variational problem and that this resolves a long-standing ambiguity reinforce our belief that we are proceeding along the correct path toward a theory of gravitation.

We now consider the variation of the first term of (4.6) as $g_{ik} \to g_{ik} + \delta g_{ik}$. We shall ignore the second term and concentrate on gravitation alone. Also for simplicity we begin by putting $\lambda_a = 1$ for all a. Later we shall return to different coupling constants.

The method is similar to that adopted for electromagnetism. We compute the change in the propagator $\bar{G}(A, B)$ as the geometry changes in any compact region $\mathcal{V}$. The details of this somewhat lengthy calculation are given elsewhere [see Hoyle and Narlikar (1974)]. We simply quote the result. The field equations turn out to be

$$\tfrac{1}{2}\phi(R_{ik} - \tfrac{1}{2}g_{ik}R) = -T_{ik} + \tfrac{1}{6}(g_{ik}\Box\phi - \phi_{;ik}) + \tfrac{1}{2}[m_i^{(\text{ret})}m_k^{(\text{adv})} + m_k^{(\text{ret})}m_i^{(\text{adv})} - g_{ik}m^{l(\text{ret})}m_l^{(\text{adv})}] \tag{5.4}$$

where $\Box$ is the covariant wave operator,

$$m(X) = \sum_a \int \bar{G}(X, A)\, da, \qquad m_i = \partial m/\partial x^i \tag{5.5}$$

$$\phi(X) = m^{(\text{ret})}(X)m^{(\text{adv})}(X) \tag{5.6}$$

and $m^{(\text{ret})}$ and $m^{(\text{adv})}$ denote twice the retarded and advanced parts of $m(X)$, respectively. The energy tensor T_{ik} is the familiar energy tensor for a system of particles $a, b, \ldots$ with masses $m_a, m_b, \ldots$ as defined by the *Machian* prescription (4.4) and (4.5). Note that the masses are time symmetric. The function $m(X)$ satisfies the conformally invariant wave equation

$$\Box m + \tfrac{1}{6}Rm = N \tag{5.7}$$

where

$$N(X) = \sum_a \int \delta_4(X, A)[-\bar{g}(X, A)]^{-1/2} da \tag{5.8}$$

is the invariant particle number density.

There are 10 equations in (5.4) and one equation (5.7) for the 11 unknowns g_{ik} and m. However, the divergence and trace of (5.4) identically vanish, showing that *there are in fact five fewer independent equations*. This is hardly surprising since four of these five are due to the general coordinate invariance (as in general relativity) while the fifth identity (the vanishing of trace) is due to *conformal* invariance. It is easy to verify that if $\{g_{ik}, m\}$ is a solution of these equations then so is $\{\Omega^2 g_{ik}, \Omega^{-1} m\}$ for an arbitrary well-behaved (i.e., of type C^2) *nonvanishing* finite function Ω. This arbitrary function is nothing but the expression of the arbitrariness of mass-dependent units discussed in Section 2.

Suppose now that it is possible to choose Ω such that

$$m^{(\mathrm{ret})}\Omega^{-1} = m_0, \qquad m_0 = \mathrm{const} \tag{5.9}$$

Suppose also that the response of the universe is such as to cancel all advanced components and double the retarded ones so that the effective mass function is

$$m + \tfrac{1}{2}m^{(\mathrm{ret})} - \tfrac{1}{2}m^{(\mathrm{adv})} = m^{(\mathrm{ret})} \tag{5.10}$$

Later we shall discuss this condition and its implications. Then the field equations are simplified to

$$R_{ik} - \tfrac{1}{2}g_{ik}R = -\kappa T_{ik}, \qquad \kappa = 6/m_0^2 \tag{5.11}$$

These are the familiar equations of general relativity! The conformal frame for which (5.9) and (5.10) hold will be called the *Einstein frame*. We have thus completed the remaining part of the programme outlined at the beginning of Section 4.

The following points are worth emphasizing in the above derivation of Einstein's equations, which is so radically different from the standard ones (Einstein, 1915; Hilbert, 1915).

(I) The approach to Einstein's equations is via the wider framework of a conformally invariant gravitation theory. Only in the limit of many particles in a suitably responding universe do we arrive at Einstein's equations. In the other limit of zero or no particles there is no theory! Thus it brings out the reason why the Machian paradox of one particle in an empty universe is not valid in the context of Einstein's equations. This reason does not emerge in the standard derivations of Einstein's equations.

(II) It is significant that the coupling constant $\kappa = 8\pi G/c^4 = 6/m_0^2$ is positive in this approach. This conclusion is unaffected by the change of sign of the coupling constants $\lambda_a, \lambda_b, \ldots$ (taken here as unity); nor is it affected by the choice of signature (i.e., $- - - +$ instead of $+ + + -$) of the space–time metric. The choice of the conformally invariant scalar propagator leads to the coupling constant being positive, i.e., to gravity being "attractive." In the standard deviations the coupling constant is fixed (in sign as well as magnitude) by a comparison with Newtonian gravity.

(III) A considerable discussion has gone on regarding the admissibility of the so-called λ term in Einstein's equations. This is because this term could be accommodated in Einstein's heuristic derivation or in Hilbert's action principle. If at all, this term is not likely to be important except in cosmology. The present cosmological picture—on observational grounds—does not force us to accept this term. It is worth emphasizing that the direct-particle approach to gravity given here *does not permit the* λ *term*. Future cosmological observations will tell us whether this definitive conclusion is valid.

(IV) The condition (5.9) that leads to Einstein's equations needs to be reexamined carefully under two special circumstances. Near a typical particle a, we expect the mass function $m^{(a)}(X)$ to "blow up;" so that $m(X) \to \infty$ as $X \to A$ on the worldline of a. In order to make $\Omega^{-1}m(X)$ finite at A, we therefore require $\Omega \to \infty$ as $X \to A$. However, we have already ruled out such conformal functions by restricting Ω to finite values. Thus the transition to Einstein's equations is not valid as we tend to any typical source particle. The nature of the equations and their solutions near a particle in this theory have been discussed by Hoyle and Narlikar (1966) and by Islam (1968).

The second, more interesting situation could arise if there existed hypersurfaces on which $m = 0$. For a transition to Einstein's equations via (5.9) we then need $\Omega = 0$ on these surfaces. But we have also excluded the $\Omega = 0$ cases from conformal transformations. Thus the Einstein equations break down on $m = 0$ surfaces.

What is the physical significance of zero mass surfaces? How do they arise in the universe? We discuss these questions in Section 7.

6. THE FRIEDMANN UNIVERSES

Let us consider the simplest Friedmann cosmological model, the so-called Einstein–de Sitter model. It is given by the line element

$$ds^2 = dt^2 - \left(\frac{3Ht}{2}\right)^{2/3} (dx^2 + dy^2 + dz^2) \tag{6.1}$$

where H is the present measured value of the Hubble constant and $t = 2/3H$ is the present cosmic epoch. The t coordinate measures cosmic time and serves as a synchronous coordinate for all "fundamental" observers with constant coordinates (x, y, z). According to the Weyl postulate we may identify these observers with galaxies in an expanding universe. The density of matter at epoch t is given by

$$\rho = (18\pi G t^2)^{-1} \tag{6.2}$$

This model is a solution of Einstein's field equations. It has a space–time singularity at $t = 0$ when $\rho = \infty$. It is a coordinate-independent singularity, unlike the one that seems to occur in the Schwarzschild solution at the horizon. The space–time structure breaks down at $t = 0$. In cosmological jargon this epoch is identified with the creation of the universe with a "big bang."

Let us examine this solution using a different conformal frame in the conformal theory. Assuming that a suitable response is produced by the universe, we see that we may replace ϕ by m^2 and $m^{(\mathrm{ret})}$ and $m^{(\mathrm{adv})}$ by m in (5.4) with the result that (5.4) reduces to

$$\tfrac{1}{6}m^2(R_{ik} - \tfrac{1}{2}g_{ik}R) = -T_{ik} + \tfrac{1}{6}\{g_{ik}\Box m^2 - m^2_{;ik} + \{m_i m_k - \tfrac{1}{2}g_{ik}m^l m_l\} \tag{6.3}$$

A little calculation shows that (6.3) and (5.7) together are satisfied by the simple solution

$$g_{ik} = \eta_{ik}, \qquad m = A\tau^2, \qquad A = \text{const} \tag{6.4}$$

where the Minkowski metric η_{ik} applies to coordinates $x^i \equiv (x, y, z, \tau)$. It is easy to see that the line element

$$ds^2 = d\tau^2 - dx^2 - dy^2 - dz^2 \tag{6.5}$$

is related to (6.1) by the conformal transformation

$$d\bar{s} = \Omega\, ds, \qquad \Omega = \left(\frac{3Ht}{2}\right)^{2/3}, \qquad \tau = 3\left(\frac{3H}{2}\right)^{-2/3} t^{1/3} \tag{6.6}$$

This is hardly surprising since in going from (6.4) to (6.1) the mass will transform to

$$\bar{m} = m\Omega^{-1} = A\tau^2\left(\frac{3Ht}{2}\right)^{-2/3} = \frac{4A}{H^2} \equiv m_0 = \text{const} \tag{6.7}$$

as required by the Einstein conformal frame. As a further check note that if we substitute (6.7) into (5.7) in the Einstein frame, we have from (5.11)

$$\bar{N} = \Box\bar{m} + \tfrac{1}{6}\bar{R}\bar{m} = \tfrac{1}{6}\bar{R}m_0 = \tfrac{1}{6}\kappa m_0\rho$$

Hence in Minkowski space–time we have from (6.2) and (6.6)

$$N = \bar{N}\Omega^3 = \tfrac{1}{2}m_0 H^2 = 2A$$

In other words,

$$m = \tfrac{1}{2}N\tau^2 \tag{6.8}$$

But this is precisely the full retarded contribution from all particles in the universe, as shown in Fig. 6. Thus the solution forms a self-consistent scheme provided we can convince ourselves that the response of the universe is indeed such as to give the full retarded solution.

In flat Minkowski space–time the propagator $\bar{G}(A, B)$ is given by

$$\bar{G}(A, B) = \frac{1}{4\pi}\,\delta(s^2_{AB}) \tag{6.9}$$

Hence if we compute the function

$$m(X) = \sum_a \int \bar{G}(X, A)\, da \tag{6.10}$$

we find that the retarded contribution is given by $N\tau^2/4$, i.e., by half the value (6.8), while the advanced contribution is *infinite*. This was pointed out by Hawking (1965), who suggested that a finite value might be obtained by having positive and *negative* contributions to $m(X)$, as in electrodynamics. Although an

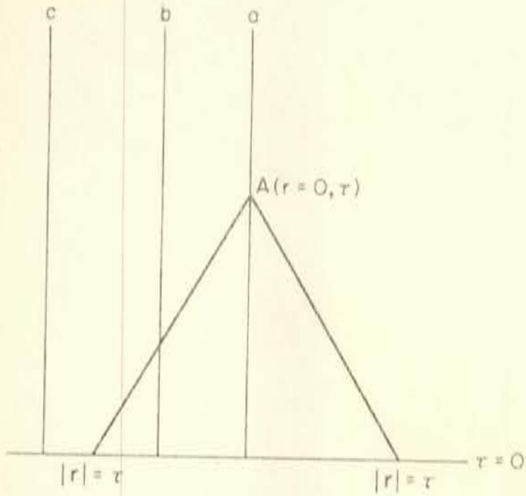

Fig. 6. The past light cone from a particle a at A (at spatial origin and epoch τ) intersects the hyperplane $\tau = 0$ in a sphere of radius τ. Only particles (like b) lying in this sphere make the retarded contribution. Particles outside (like c) do not contribute.

exactly equal number of positive and negative contributions will lead to a zero $m(X)$, as in the overall electrostatic contribution of all charges in the universe, this device does not lead to a satisfactory resolution of the difficulty. Hoyle and Narlikar (1972) have discussed how infinite contributions from the future can lead to an interesting interpretation of the Dirac equation for a free particle. Here we describe a subtle modification that can resolve the difficulty at the classical level.

6.1. Negative λ

We go back to the action (4.6) and take into account the possibility that λ_a and λ_b switch signs as the $\tau = 0$ surface is crossed. That is, we define

$$\lambda_a = \begin{cases} +1, & \text{if } A \text{ lies to the future of } \tau = 0 \\ -1, & \text{if } A \text{ lies to the past of } \tau = 0 \end{cases} \tag{6.11}$$

This prescription implies two important departures from the picture developed so far. In the Friedmann cosmology $t \geq 0$; and correspondingly we took $\tau \geq 0$ for Minkowski space–time. However, if we had not been prejudiced by the Friedmann picture (in the Einstein frame) and had instead worked with Minkowski space–time we would not have imposed the artificial cutoff by $\tau = 0$. In other words, the space–time manifold and the particle worldlines could have continued back to $\tau = -\infty$.

The second point of difference arises for the computation of $m(X)$ for the new prescription. This is illustrated by Fig. 7. A_+ is a point lying in the future of $\tau = 0$ while A_- is a similar point to the past of $\tau = 0$. $B_\pm$ and $C_\pm$ are similarly typical points on the world lines of b and c, such that A_+B_+, A_+B_-, A_+C_+, A_+C_- are null

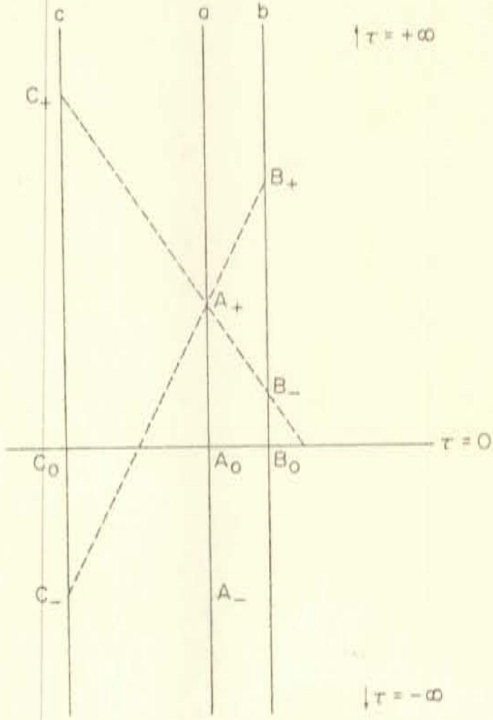

Fig. 7. The contributions, advanced and retarded, are equal from particle b while they are equal and opposite from particle c.

rays. When computing the contribution from b to $m(A_+)$ we notice that since $B_\pm$ both lie to the future of $\tau = 0$ their advanced and retarded contributions are equal, and hence

$$m^{(b)}(A_+) = \tfrac{1}{2}m^{(b)(\mathrm{ret})}(A_+) + \tfrac{1}{2}m^{(b)(\mathrm{adv})}(A_+) = m^{(b)(\mathrm{ret})}(A_+) \qquad (6.12)$$

However, as far as c is concerned, C_+ lies to the future of $\tau = 0$ while C_- lies to the past of it. Here the advanced and retarded contributions are equal and opposite so that

$$m^{(c)}(A_+) = \tfrac{1}{2}m^{(c)(\mathrm{ret})}(A_+) + \tfrac{1}{2}m^{(c)(\mathrm{adv})}(A_+) = 0 \qquad (6.13)$$

What is the difference between b and c in their relation to a? As is evident from Fig. 7, the world line b intersects the surface $\tau = 0$ at a point B_0 that lies *inside* the past light cone of A_+. The similar point C_0 on the world line of c lies *outside* the past light cone of A_+. Thus only the world lines of former type contribute to A_+ and their contribution agrees with (6.8).

We may perform a similar exercise for A_-. In this case we come up with the same answer for the mass of a particle at A_-. (Although the contributions from b-type particles are negative, they are now multiplied by a negative value of λ_a.)

We therefore have a completely symmetric "other half" added on to the Friedmann universe, in a fashion that was not possible in the Einstein frame.

6.2. Observational Cosmology

Although we have discussed the Einstein–de Sitter cosmology in some detail, that discussion and what follows now can be easily extended to other Friedmann models [see Hoyle and Narlikar (1974) for details]. We shall continue with the Einstein–de Sitter model.

Looking at the model from Minkowski space–time it is interesting to reinterpret the nebular redshift observed by Hubble. In this nonexpanding space–time, a redshift cannot arise from a non-Euclidean geometry in the Einstein frame. It arises instead from the change in the mass function.

Suppose a light signal was sent by a remote galaxy at epoch $\tau = \tau_1$ and it was received by us at $\tau = \tau_0$, where $\tau_0 > \tau_1$. Since all spectral frequencies depend in direct proportionality on the masses of atomic particles, the frequency of emission ν_1 is related to the frequency reception ν_0 (of any given atomic or molecular process) by

$$\frac{\nu_1}{\nu_0} = \frac{m(\tau_1)}{m(\tau_0)} = \frac{\tau_1^2}{\tau_0^2} = \left(\frac{t_1}{t_0}\right)^{2/3} \tag{6.14}$$

The observer in our laboratory would expect the spectral frequency to be ν_0 (wavelength λ_0); instead he receives a frequency ν_1 (wavelength λ_1) from the remote galaxy. He therefore associates a *redshift* to the light of an amount z, where

$$1 + z = \frac{\lambda_1}{\lambda_0} = \left(\frac{t_0}{t_1}\right)^{2/3} \tag{6.15}$$

This is the familiar redshift formula of Einstein–de Sitter cosmology. Indeed, all the standard formulas of cosmology receive reinterpretation in a considerably simpler form in this new way of looking at the universe! The observable results are no different from those obtained by using the standard Einstein frame. This is to be expected in view of our earlier remark that this new way of looking at gravitation does not differ from that of general relativity (except in a few special situations mentioned in Section 5).

There are certain advantages in extending space–time to $\tau < 0$. It was pointed out by Hoyle (1975) that the microwave background near $\tau = 0$ can be created by the thermalization of starlight in $\tau < 0$. Because $m \to 0$ as $\tau \to 0$, the photons get thermalized by low-mass electrons close to $\tau = 0$. In this way it is possible to understand the origin of the microwave background and to see why its energy density is comparable to the starlight energy density at the present epoch. In the standard big-bang picture the universe starts at $t = 0$ ($\tau = 0$) and the existence and intensity of the presently observed microwave background have to be ascribed to initial conditions.

In standard cosmology the space–time singularity prevents any discussion of the physics at $t = 0$. We now see that this, so far as the Friedmann model is concerned, is the result of using the "wrong" conformal frame. Since at $\tau = 0$, $m = 0$, we cannot use the Einstein frame to describe the universe in a time interval including $\tau = 0$ (since it involves $\Omega = 0$). Instead we should use a conformal frame like the Minkowski one, which enables us to have a nonsingular space–time manifold. The price paid for this is the admission of a variable mass that vanishes at some epoch. But this is a "lesser evil" than the breakdown of space–time structure in the Einstein frame.

7. ZERO MASS SURFACES

The existence of positive and negative coupling constants λ implies in general the existence of hypersurfaces $m = 0$. The Friedmann cosmology (and its reinterpretation in the previous section) is an example of this. In an arbitrary situation, i.e., without the postulates of homogeneity and isotropy, the zero mass surfaces may not have such symmetries or regularities as in the Friedmann case. However, we may look for one common feature between the Friedmann case and the general case: Do these $m = 0$ hypersurfaces *always* correspond to space–time singularities in the Einstein conformal frame?

Considerable work has been done recently on this problem. The results, described below briefly, indicate that the answer to the above question is in the affirmative. These investigations are based on the following type of calculation (Narlikar and Kembhavi, 1977). For a specific singular manifold $\mathcal{M}_0$ in the Einstein frame, it is shown that (i) there exists a nonsingular manifold $\mathcal{M}$ as a solution of (6.3), which contains a zero mass surface Σ; (ii) there exists a conformal transformation $\mathcal{M} \to \mathcal{M}_0$ such that the conformal function vanishes on Σ; and (iii) the surface Σ goes into the singularity of $\mathcal{M}_0$ under this transformation.

(I) Consider the Bianchi type I universe with shear (Hawking and Ellis, 1973) given by the line element

$$ds_0^2 = dt^2 - \left(\frac{9M}{2}\right)^{-4/3} t^2(t + \Sigma)^{-2/3}dx^2 - \left(\frac{9M}{2}\right)^{4/3} (t + \Sigma)^{4/3}(dy^2 + dz^2) \quad (7.1)$$

where M and Σ are positive constants. This is obtained as the solution of Einstein's equations for homogeneous dust and has a pancake singularity at $t = 0$. It can be shown that for this manifold $\mathcal{M}_0$ the corresponding $\mathcal{M}$ is given by the conformal function Ω, where

$$ds_0 = \Omega\, ds, \qquad \Omega = \frac{t}{t^2 + 1} \quad (7.2)$$

Here ds represents the line element of $\mathcal{M}$. The manifold $\mathcal{M}$ is nonsingular: it is geodesically complete and has finite curvature invariants at $t = 0$. Thus the singularity at $t = 0$ of $\mathcal{M}_0$ is due to the existence of the zero mass surface at $t = 0$ in $\mathcal{M}$.

(II) A few years ago Belinsky *et al.* (1970) obtained what they claim to be the *most general* type of cosmological singularity in general relativity. By choosing a synchronous time coordinate t, they showed that the singularity developes in an oscillatory manner as $t \to 0$. In this case, corresponding to $\mathcal{M}_0$, (the most general *singular* cosmological manifold), it is possible to find $\mathcal{M}$ that is nonsingular in the manner described above. The singularity of $\mathcal{M}_0$ is again identified with the zero mass surface of $\mathcal{M}$. This has been discussed in detail by Kembhavi (1978).

(III) The singularities in (I) and (II) are of the type where the curvature invariants diverge in $\mathcal{M}_0$. Kembhavi (1978) has also considered another type of singularity, where instead of the divergence of curvature invariants the geodesic world lines (null or timelike) cannot be continued indefinitely in terms of their affine parameters. In this case the singularity arises from geodesic incompleteness, and an example of this is Taub–NUT space–time (1969). Here also the geodesically incomplete Taub manifold $\mathcal{M}_0$ can be shown to be a conformal transform of the nonsingular geodesically complete manifold $\mathcal{M}$, obtained as a solution of (6.3), such that the variable mass function vanishes on two hypersurfaces. Again, it is these hypersurfaces, which after a conformal transformation, go over into the two singular hypersurfaces of Taub space–time $\mathcal{M}_0$.

Apart from these examples of cosmological singularities it is also possible to deal in a similar way with compact singularities of the Schwarzschild type. These examples generate confidence that *all* space–time singularities that occur in general relativity in a physical context (as opposed to mathematical or pathological cases) are in fact the outcome of trying to describe the universe in the Einstein conformal frame in the neighborhood of an $m = 0$ hypersurface.

It is interesting to speculate here whether this conjecture will turn out to be correct. If it does, it will have thrown some light on the singularity question in general relativity. For, in conformal gravity the appearance of the $m = 0$ hypersurfaces seems inevitable. Hence if we insist on using the Einstein frame, the appearance of space–time singularities in general relativity is also seen to be inevitable. We shall consider certain physical aspects of the zero mass surfaces in the following sections.

In conclusion, we point out a certain analogy with the case of coordinate transformations. In a covariant theory we often discard one set of coordinates if the physical description in terms of these coordinates seems to diverge, although space–time is nonsingular. A familiar example is the unsuitability of the Schwarzschild coordinate near the Schwarzschild event horizon. In this case other more suitable coordinates like the Kruskal–Szekeres coordinates are used. In the same way, in a conformally invariant theory, one particular conformal frame may be unsuitable to describe physics in certain cases. In this case other, more suitable conformal frames can be used. In the above discussion it is argued that near $m = 0$ hypersurfaces the Einstein conformal frame is unsuitable, that other conformal frames are more suitable in that they generate nonsingular manifolds.

8. BLACK HOLES AND WHITE HOLES

The Friedmann cosmological case considered in Section 6 is a highly special case because of its many symmetries. In general the zero mass surfaces are not likely to be flat homogeneous isotropic hypersurfaces. Also, they need not be open. In Fig. 8 is shown a closed zero mass hypersurface Σ.

Consider the world line of particle *a* intersecting Σ at two points A_1 and A_2. Let $P_+(P_-)$ and $Q_-(Q_+)$ be two points on this world line close to A_1 and A_2, respectively, and lying inside (outside) Σ. For P_+ sufficiently close to A_1, i.e., for the arc length A_1P_+ small compared to ρ_1, a linear dimension characteristic of the radius of curvature of Σ at A_1, the hypersurface Σ would appear to an observer at P_+ like a surface stretching out to infinity. For the information available to P_+ about Σ is confined to the stretch of Σ within the past light cone of P_+. The experience of P_+ would therefore be broadly similar to that of an observer in a Friedmann model. Conversely, our own cosmological experience has a characteristic dimension $\sim H^{-1}$ (H = Hubble's constant), so that for

$$\rho \gg H^{-1} \tag{8.1}$$

the Friedmann model represents an approximation to a zero mass closed hypersurface of "radius" ρ.

The observer freely and radially falling into a Schwarzschild black hole has a trajectory similar to that of a fundamental observer in a *contracting* Friedmann model. In general relativity this observer falls into a singularity and his world line is abruptly terminated at a finite affine parameter (i.e., his proper time). It has been conjectured that the observer emerges into another universe in the form of an ejection from a *white hole*. However, in the framework of orthodox general rel-

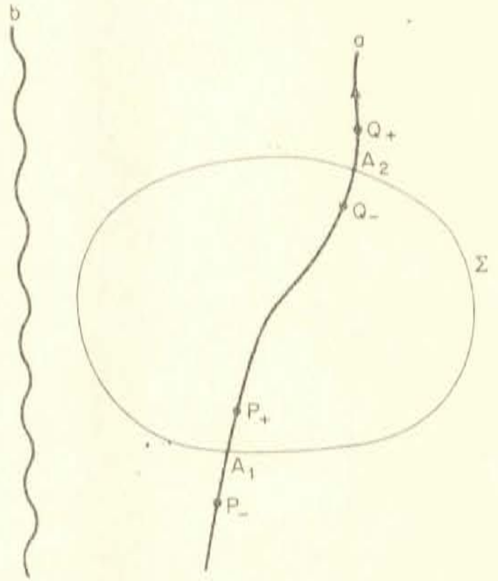

Fig. 8. A closed zero mass surface corresponds to two black hole–white hole systems. An external observer at *b*, using the Einstein frame will see a black hole forming at A_1 and a white hole erupting at A_2. He will not see the white hole at A_1 and the black hole at A_2.

ativity it is not possible to make such a continuation across the Schwarzschild space–time singularity.

In the conformal theory proposed in the previous sections, this difficulty is seen as arising from the choice of the Einstein frame. If a different conformal frame is used there is no space–time singularity at A_1 in Fig. 8. We can in this frame continue the worldline of a across Σ from P_- to P_+. Thus we have the logical outcome of a white hole. Continuing further the worldline meets Σ again at A_2, so that we have another black hole/white hole combination. An external observer like b (see Fig. 8.) may see only the first black hole and the second white hole and since they may be spatially separated, he may not be able to link one with the other. Thus the possibility emerges here that a black hole in one part of the universe is connected with a white hole in another part well separated from it!

9. ANOMALOUS REDSHIFTS

In recent years considerable data have accumulated on the so-called *anomalous redshifts* of quasi-stellar objects (QSO). The word anomalous implies a departure from the conventional cosmological explanation of the redshifts. In the cosmological explanation, two neighboring objects A and B when seen by a remote observer should exhibit the same redshift since they are seen at the same epoch. Anomaly arises if A and B have substantially different redshifts.

The best example of anomalous redshifts to date is that of the pair of QSOs identified with the radio sources 4C-1150 (a) and (b), which are separated by only ~4.5 arc seconds. Of these one QSO has a redshift of 1.9 while the other has a redshift of 0.44. A chance juxtaposition of two QSOs at different line of sight distances has a very small probability—less than 10^{-3}. This case has been discussed by Wampler *et al.* (1973). There are many more cases of such anomalous redshifts, mainly between a galaxy and a QSO [see Arp (1977) for an excellent review]. The anomaly is too large to be explained by the Doppler shift between the two objects. A part of the excess redshift could be due to gravitation if the objects are highly compact (Das, 1976). However, in many cases the excess redshift objects do not show any high degree of compactness. It is of course possible, as many would like to argue (Rees, 1977) that all cases of anomalous redshifts reported so far are optical selection effects or misidentifications. With the steadily increasing number of such cases this point of view is becoming more difficult to sustain.

A new explanation of the anomalous redshifts is provided within the framework of conformal gravity. The details of this are given elsewhere [see Narlikar (1977)]. The idea is briefly described below.

In (6.14) and (6.15) we obtained the redshift formula by comparing the masses at epochs τ_0 and τ_1. This was on the assumption that the zero mass surface was given by $\tau = 0$. Suppose instead of this the surface has a kink as shown in Fig. 9. Let a and b be two neighboring particles with world lines intersecting the zero mass surface Σ at A_0 and B_0, respectively. Of these A_0 lies on $\tau = 0$ while B_0 lies on $\tau = \tau_b > 0$, i.e., in a kink. The action can be rewritten to take account of such

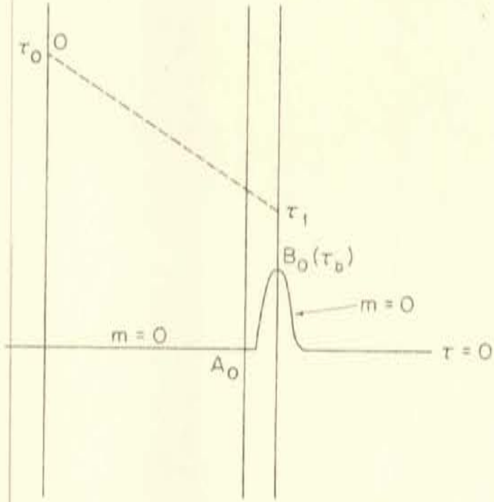

Fig. 9. A kink in the zero mass surface can produce anomalous redshifts as discussed in the text.

kinks (Narlikar, 1977) and the masses computed. If the kinks are few and far between and do not affect the Friedmann cosmological picture significantly then the following result can be derived.

Let τ_0 be the epoch of the point of observation O. The null cone from O into the past intersects the world lines of a and b at A and B, respectively, as shown in Fig. 9. Let the epoch of A and B be τ_1 (being the same since A and B are close to each other). Then the masses at O, A, and B are given by

$$m(O) = A\tau_0^2, \qquad m(A) = A\tau_1^2, \qquad m(B) = A(\tau_1 - \tau_b)^2 \tag{9.1}$$

where A is a constant. The redshifts of a and b as seen by O are then given, respectively, by

$$z_a = \frac{\tau_0^2}{\tau_1^2} - 1, \qquad z_b = \frac{\tau_0^2}{(\tau_1 - \tau_b)^2} - 1 \tag{9.2}$$

Note that $z_b > z_a$ and that $z_b - z_a$ increases as τ_b approaches τ_1.

In the Einstein frame the epoch on the zero mass surface is identified with the epoch of creation. Thus we may say that in this frame the particle b was created later than particle a ($\tau_b > 0$). The above result indicates that a younger object will have a larger (and hence anomalous) redshift. It is interesting that Arp's empirical findings do show this effect. It needs to be determined whether this effect is seen quantitatively. For such a determination it will be necessary to be able to estimate the ages of QSOs and galaxies more accurately than has been possible hitherto.

10. CONCLUDING REMARKS

We began our discussion of inertia and cosmology with Newton, Mach, and Einstein. Our end product is a theory that combines the important concepts of all three of them! We have used the concept of action at a distance first introduced in a mathematical form by Newton, although we had to revise it to make the action propagate along null lines. This concept was used to describe the inertia of a particle in terms of the contributions from other particles in the universe. This is a mathematical expression of Mach's principle. Finally, the space–time structure in which this mass interaction propagates is taken to be a Riemannian 3 + 1 manifold in which gravitational effects manifest themselves through the non-Minkowskian geometry. This is the underlying idea behind the general theory of relativity and the resulting "field" equations do become the Einstein field equations under certain general conditions.

Although the Einstein field equations so obtained serve as the simplest set of equations of gravitation, their use is seen to be invalid near a typical particle or near a zero mass surface. The insistence on the use of the Einstein conformal frame leads inevitably to space–time singularities on such surfaces. Inhomogeneities of such a cosmological zero mass surface can lead to the phenomenon of anomalous redshifts as was discussed in the last section.

Since Einstein's great work on gravitation, general relativity has been subjected to much scrutiny on a conceptual level. The approach described here should be considered as one of them. It brings out the elegance of Einstein's equations, their unique character, and their essential simplicity. It resolves the lacuna of their relationship to Mach's principle and prompts a look at the space–time singularity from the wider perspective of conformal invariance.

A look at the direct particle action (4.6) raises the tantalizing possibility that the propagators $\bar{G}_{i_k i_m}$ and $\bar{G}(A, B)$ are parts of some "master propagator." If such a propagator were to exist it would unify gravity and electromagnetism. However, the answer may not be as straightforward as indicated by (4.6) since a "supertheory" is expected to unify all basic interactions of science and not just these two. Any approach at such a unification must therefore be incomplete, for at no stage can a scientist boast of having put together all the pieces of the jigsaw of the laws of nature. Hence, the scientist cannot afford to wait until he has all pieces before beginning to connect them.

Nevertheless, as this chapter has tried to show, any attempt to understand physics in purely local terms ignores the important aspect of the large-scale struc-

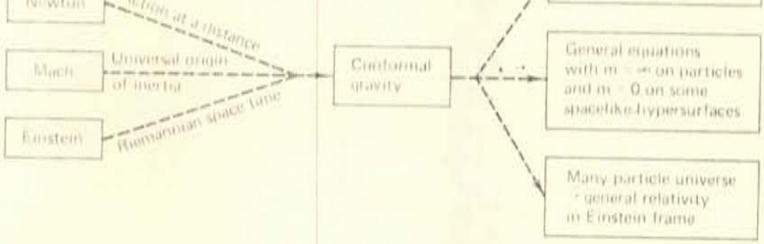

Fig. 10. A flow diagram of the theory of conformal gravity.