Mon. Not. R. astr. Soc. (1972) **155**, 305–321.

COSMOLOGICAL MODELS IN A CONFORMALLY INVARIANT GRAVITATIONAL THEORY—I

THE FRIEDMANN MODELS

F. Hoyle and J. V. Narlikar

(Received 1971 July 27)

SUMMARY

The present paper discusses the formulation of the Friedmann cosmological models in terms of a conformally invariant gravitational theory. This theory is Machian in the sense that the mass of a particle arises from the interaction of the particle with a mass field $m(X)$ generated by other particles. In cosmology the mass field $m(X)$ at any particular space–time point X arises predominantly from particles at great distance from X.

The Friedmann models are usually discussed in terms of the Robertson–Walker line element. It is known that this line element is conformal to the Minkowski line element $ds^2 = d\tau^2 - d^2 - r^2 (d\theta^2 + \sin^2 \theta \, d\phi^2)$. Cosmological space-time can therefore be transformed to Minkowski space-time by a suitable conformal transformation. It is not possible in the usual expositions to take advantage of this geometrical simplification because Einstein's gravitational equations are not conformally invariant. However, the present theory is conformally invariant so that transformation to Minkowski space is possible not only for the geometry but also for the physics.

The three Friedmann cases $k = 0, \pm 1$ are discussed in detail from this point of view. Although the cases $k = \pm 1$ are spatially homogeneous in the Robertson–Walker frame they are not similarly homogeneous in the Minkowski frame, where they can be seen to represent only local clouds that happen to be symmetrically distributed with respect to an observer at $r = 0$. This lack of homogeneity is not shared by the $k = 0$ case, which emerges from the analysis as the only model consistent with homogeneity in both frames, Robertson–Walker and Minkowski.

The conformal transformation function between these two frames is singular at $\tau = 0$. It is this mathematical breakdown of the transformation function which introduces the well-known singularity of the Friedmann models with respect to the Robertson–Walker frame—the singularity usually referred to as the origin of the Universe. From the present point of view this so-called origin does not arise physically at all. It turns out that the Universe possesses an opposite half, $\tau < 0$ in the Minkowski frame, which connects smoothly with ' our ' half, $\tau > 0$. Both halves of the Universe contribute to the mass function $m(X)$, and are therefore connected physically. Indeed the appropriate form for $m(X)$ appears to demand that both halves of the Universe be present. The half $\tau < 0$ is missed when the Robertson–Walker frame is used.

1. INTRODUCTION

All physical measurements are in terms of dimensionless numbers. The observer may assert that such and such an object has a flux of S W m^{-2} Hz^{-1}, but his actual observation was for a telescope of explicit aperture, was made over an explicit time interval and over an explicit bandwidth and on a receiver with an explicit response characteristic. All the dimensionalities appearing in

S have been introduced by the observer himself—he has divided out by the aperture of his telescope, by his bandwidth, and so on. The procedure has the valid purpose of permitting different observers with different instruments to compare their results.

We can introduce a unit for any quantity arrived at in this way. For example we can take 10^{-26} W m^{-2} Hz^{-1} as the unit of S, or 10^{-24} cm^2 as the unit of nuclear cross-section. Having done this, we can forget the unit we have introduced for our quantity provided in our calculations we always arrive at dimensionless numbers *for each such quantity taken separately*. Usually we measure time and spatial length in different units and this forces us to keep a separate accounting of the time unit and the length unit in our calculations. But if we note that time and space are connected through

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \tag{1}$$

and if we elect to use the same unit for both, the need for separate accounting is avoided. The situation is that the more units we introduce the harder it becomes to keep our accounts straight. We could elect to use a different unit for each of the three spatial Cartesian coordinates x, y, z, as well as for t, but this would force us to keep four separate accounts instead of the one account needed if we use the same unit for all x, y, z, t. Nobody has suggested the latter procedure but the system of units frequently used in electromagnetic theory is even more absurd. Wishing to get rid of the velocity of light in calculation, and not understanding that c should be set equal to unity by measuring length and time in the same unit, the velocity of light is absorbed misleadingly into the unit of an electromagnetic quantity. At the best this sets up three separate accounting procedures where a single account would have sufficed. At the worst it obscures the importance of the Minkowski metric (1).

The least objectionable of the multiunit systems in common use is the c.g.s. system. But we have already seen that the need for both the centimetre and the second is artificial. Can mass also be related to length? To see that it can, we turn to an aspect of quantum mechanics just as important to physics as the Minkowski metric. The amplitude for a particle to go from one point to another, say from point 1 to point 2, is given by summing the phase factor $\exp(iS)$ for every possible path from 1 to 2. Here S must be dimensionless. Now S is constructed from

$$S = -\int_1^2 m\,da, \tag{2}$$

where m is the inertial mass of the particle and da is the element of proper time along the particular path under consideration. This shows that m must be the reciprocal of a length. If one wishes to have separate units for mass and length it is necessary artificially to introduce a scale factor $\hbar$, replacing $\exp(iS)$ by $\exp(iS/\hbar)$. This is analogous to the need for introducing c as the velocity of light into (1) when we choose separate units for length and time. Both c and $\hbar$ are cumbersome and unnecessary.†

† The 4π which appears in physical equations cannot be similarly removed. The 4π is a dimensionless number equal to the ratio of the proper area of a sphere in Euclidean 3-space to the square of the radius. No artifice of any kind can conceal the remarkable fact that this number is transcendental. If 4π is suppressed in one place it inevitably bursts out somewhere else.

Proceeding in this way all physical quantities can be given dimensionalities that are some power L^n of the length unit L. For example

$$\text{mass} \sim L^{-1}, \quad \text{frequency} \sim L^{-1}, \quad \text{charge} \sim L^0, \quad \text{gravitational 'constant'} \sim L^2. \tag{3}$$

In any calculation of an observable result all we need do in combining various quantities is to make sure all the exponents add to zero.

Since in our calculations we are seeking to calculate dimensionless numbers which are then to be compared with observed dimensionless numbers it does not matter if we change the unit of length by any constant factor. But how if we change the unit by a different factor at different points, say by some well-behaved function of position $\Omega(x, y, z, t)$? The answer is that still nothing will be changed provided that in calculating a dimensionless number appropriate to a particular point x, y, z, t we combine quantities with dimensionality in L all taken at x, y, z, t. But if we combine quantities with dimensionality some taken at x_1, y_1, z_1, t_1, some taken at a different point x_2, y_2, z_2, t_2, things will be changed. A theory which never does this, and which transforms all physical quantities of dimensionality L^n by Ω^n, is said to be conformally invariant.

The electromagnetic theory of Maxwell is conformally invariant. The derivative terms in Dirac's equation are conformally invariant but the mass term is not. Einstein's general relativity is not conformally invariant. The situation in the well-known physical theories is therefore mixed. Should this be so? It turns out that lack of conformal invariance is always connected with mass. As long as mass is taken to be a fixed quantity belonging autonomously to a particle there is no possibility of making all physical theories conformally invariant. For conformal invariance we require mass to change by Ω^{-1}. The anomaly is that we agree to make this change when Ω is taken to be a constant everywhere the same but we balk at it when Ω is a function of position. We find it hard to rid ourselves of the obsession that a proton at x_2, y_2, z_2, t_2 should have the same mass as a proton at x_1, y_1, z_1, t_1. The differing values of Ω at the two points would destroy equality.

The situation can be seen from (2). In order that the amplitude for a particle to go from 1 to 2 shall be unaffected by a conformal transformation it is necessary that m change along the path from 1 to 2, in order to compensate for the change in da produced by the conformal transformation, $da \to \Omega da$. We shall show in the next section how mass can be given the property $m \to \Omega^{-1}.m$ and will then see how a conformally invariant gravitational theory can be obtained.

2. CONFORMAL GRAVITATION AND THE NATURE OF MASS

The work of the present section has been given in detail elsewhere (1, 2), so we shall describe only those considerations that are relevant to later sections and we shall do so briefly.

A conformal transformation changes the Riemannian line element

$$ds^2 = g_{ik}\, dx^i\, dx^k \tag{4}$$

through

$$g_{ik}^* = \Omega^2 g_{ik} \tag{5}$$

to

$$ds^{*2} = g_{ik}^*\, dx^i\, dx^k = \Omega^2\, ds^2. \tag{6}$$

Here $\Omega(x^i)$ must be a real non-zero function—we cannot have zero or infinity in the scaling of our length unit. We require a theory in which mass is changed by (5) in accordance with

$$m^* = \Omega^{-1} m, \tag{7}$$

otherwise a dimensionless number whose calculation involves mass will be affected by the conformal transformation, for example S in (2). How is this to be done? Only by accepting a Machian view of the nature of mass. The mass appearing in (2) for a particular path of particle a must arise from the other particles $b, c, \ldots$ in the Universe, through an expression of the form

$$m_a(A) = \sum_{b \neq a} \int P(A, B)\, db, \tag{8}$$

in which we have added a subscript to m_a to denote the mass of particle a. $P(A, B)$ is a propagation function that goes from point B on the path of particle b to point A on the path of particle a. If we sum the action (2) for all particles we obtain

$$\sum_a S_a = -\sum_a \int m_a da = -\sum\sum_{b \neq a} \int\int P(A, B)\, da\, db \tag{9}$$

calculated for a particular path of each particle. Classically, it is sufficient to consider one path for each particle. In quantum theory all paths must be considered. Here we work in terms of classical theory.

So far as (9) is concerned we can put

$$P(A, B) = P(B, A), \tag{10}$$

since an antisymmetric part in $P(A, B)$ makes no contribution to (9). We require that under the conformal transformation (5) $P(A, B)$ shall change to

$$P^*(A, B) = \Omega(A)^{-1}.\Omega(B)^{-1}.P(A, B). \tag{11}$$

The m_a given by (8), and $m_a{}^*$ given by

$$m_a{}^* = \sum_{b \neq a} \int P^*(A, B)\, db^*, \tag{12}$$

then satisfy the required relation (7). There is only one symmetrical scalar propagator $\tilde{G}(A, B)$ with the property (11). It satisfies (*cf* 1)

$$\Box_X \tilde{G}(X, B) + \tfrac{1}{6} R(X)\, \tilde{G}(X, B) = [-g(B)]^{-1/2}\, \delta_4(X, B), \tag{13}$$

where $\delta_4(X, B)$, the 4-dimensional delta function, represents a source of unit strength at point B. In Minkowski space the scalar Riemannian curvature $R(X)$ is zero and the solution of (13) is

$$\tilde{G}(X, B) = \frac{1}{4\pi}\, \delta(s_{XB}{}^2). \tag{14}$$

The only possibility for $P(X, B)$ is therefore

$$P(X, B) = \lambda^2 \tilde{G}(X, B), \tag{15}$$

where λ^2 is a dimensionless coupling constant. Not until a much later stage shall we need λ^2, so for the present we shall simplify writing by putting $\lambda^2 = 1$. The coupling constant can easily be reintroduced when it is required.

The gravitational equations are obtained in the usual way, by making a small *general* change in the metric tensor $g_{ik} \rightarrow g_{ik} + \delta g_{ik}$, not simply a small conformal change, and by requiring that the action (9) shall remain unchanged in first order. The outcome after a somewhat lengthy calculation is found to be (1):

$$F(R_{ik} - \tfrac{1}{2} g_{ik} R) = -3(T_{ik} + \Phi_{ik}) + (g_{ik} \Box F - F_{;ik}). \tag{16}$$

The Ricci tensor appears here as it does in Einstein's equations—it arises because of the $R(X)$ term in (13). The other expressions in (16) are defined in terms of mass fields $m^{(a)}(X)$, $m^{(b)}(X)$, ... at the field point X where (16) is taken. These fields are generated by the particles a, b, ... in accordance with

$$m^{(a)}(X) = \int \tilde{G}(X, A)\, da. \tag{17}$$

Thus

$$F = \tfrac{1}{2} \sum_{a \neq b}\sum m^{(a)} m^{(b)}, \tag{18}$$

$$\Phi_{ik} = -\tfrac{1}{2} \sum_{a \neq b}\sum [m_i^{(a)} m_k^{(b)} + m_k^{(a)} m_i^{(b)} - g_{ik} m_l^{(a)} m^{(b)l}], \tag{19}$$

$$T^{ik}(X) = \sum_a \int \delta_4(X, A)[-g(A)]^{-1/2} m_a(A) \frac{da^i}{da} \frac{da^k}{da}\, da, \tag{20}$$

$$m_a(A) = \sum_{b \neq a} m^{(b)}(A). \tag{21}$$

In (19), $m_i^{(a)} = \partial m^{(a)}/\partial x^i$, $m_k^{(b)} = \partial m^{(b)}/\partial x^k$.

The world line of particle a can be determined either by taking the divergence of (16) or by varying the path in (2). The result is

$$\frac{d}{da}\left[m_a \frac{da^i}{da}\right] + m_a \Gamma_{kl}^i \frac{da^k}{da} \frac{da^l}{da} - g^{ik} \frac{\partial m_a}{\partial a^k} = 0. \tag{22}$$

Variations of $m_a(X)$ along the trajectory have no effect but variations normal to it contribute mass gradient terms in (22).

In problems involving many particles it often happens that F, Φ_{ik} can be approximated by

$$F \simeq \tfrac{1}{2}\left[\sum_a m^{(a)}\right]^2, \tag{23}$$

$$\Phi_{ik} = -\left[\sum_a m^{(a)}\right]_i \left[\sum_b m^{(b)}\right]_k + \tfrac{1}{2} g_{ik} g^{lp} \left[\sum_a m^{(a)}\right]_l \left[\sum_b m^{(b)}\right]_p. \tag{24}$$

For (23) to hold it is necessary that cross-product terms

$$\sum_{a \neq b}\sum m^{(a)}\, m^{(b)}$$

shall overwhelm the sum of squares,

$$\tfrac{1}{2} \sum_a [m^{(a)}]^2.$$

This will certainly be the case for a many-particle system provided the mass contributions $m^{(a)}$, $m^{(b)}$, ... are comparable with each other in magnitude and are of the same sign. The latter conditions are satisfied in cosmology. Similar considerations apply to (24).

Now because the theory is conformally invariant we can choose a particular conformal frame in which F is constant. This is because the masses in (23) all transform as Ω^{-1}, so that

$$F^* = \Omega^{-2}F. \tag{25}$$

If F is not constant to begin with we can evidently make F^* constant by taking $\Omega \propto F^{1/2}$. We see from (23) that

$$m = \sum_a m^{(a)} \tag{26}$$

must be constant in the particular conformal frame in which F is constant. All the terms on the right-hand side of (16) except the T_{ik} term now disappear. Defining G by

$$8\pi G = \frac{3}{F} > 0, \tag{27}$$

the gravitational equations are

$$R_{ik} - \tfrac{1}{2} g_{ik}R = -8\pi G T_{ik}, \tag{28}$$

the same as Einstein's.

Since we have come at last to the Einstein equations it may be asked: what has the above procedure gained us? This: because the conformally invariant theory leads to (28) when the approximations (23), (24) hold good—as they do in cosmology—and because the usual cosmological models satisfy (28), the usual models satisfy the conformally invariant theory. We can therefore make conformal transformations on them, and we can argue when we have done so that the new conformal frame is every bit as good from a physical point of view as the old frame. This last step cannot be made in Einstein's theory.

It turns out that this procedure throws a great deal of light on the usual cosmological models, in particular it removes the so-called 'origin' of the Friedmann models.

3. CONFORMAL TRANSFORMATION OF THE FRIEDMANN MODELS

The Friedmann models are discussed with respect to the Robertson–Walker line element†

$$ds^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right], \qquad k = 0, \pm 1. \tag{29}$$

We pause to verify in relation to (29) that the approximation (23), (24) holds good, and hence that (16) leads to (28). The Friedmann models are obtained by taking (28), (29) together.

The three Robertson–Walker spaces (29) are each conformal to Minkowski space (3). Transform to Minkowski space using $\Omega > 0$. Then (14) holds and all mass fields can be seen to be positive. Yet the signs of the mass fields were not changed by the conformal transformation. This, together with the circumstance that

$$m = \sum_a m^{(a)}$$

† The expansion function S here is obviously not to be confused with the action function (2).

is dominated by a very large number of terms of comparable magnitude, permits (23), (24) to be used. We shall be closely involved with the calculation of m in suceeeding sections, and the way m is made up will then become clearer.

We consider the Robertson–Walker space with $k = +1$ first. Defining $\sin R = r$,

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

$$= dt^2 - S^2(t)[dR^2 + \sin^2 R(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{30}$$

The gravitational equations (28) lead to

$$\dot{S}^2 = \frac{1-S}{S}, \tag{31}$$

which can be solved in terms of T, defined by

$$T = \int_0^t \frac{dt}{S}, \tag{32}$$

to give

$$S = \tfrac{1}{2}(1 - \cos T), \tag{33}$$

$$t = \tfrac{1}{2}(T - \sin T). \tag{34}$$

In terms of T, (30) becomes

$$ds^2 = S^2[dT^2 - dR^2 - \sin^2 R(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{35}$$

Next, it can be verified that the *coordinate* transformations

$$\begin{aligned} \xi &= \tfrac{1}{2}(T+R), & X &= \tan\xi, & 2\tau &= X+Y, \\ \eta &= \tfrac{1}{2}(T-R), & Y &= \tan\eta, & 2\rho &= X-Y, \end{aligned} \tag{36}$$

map the interior of the triangle formed in the R, T plane by the points (0, 0), (π, 0), (0, π), into the quadrant $0 \leqslant \rho < \infty$, $0 \leqslant \tau < \infty$ in the ρ, τ plane. By carrying through the transformations it can also be shown (4) that

$$ds^2 = \frac{4S^2}{(1+X^2)(1+Y^2)}[d\tau^2 - d\rho^2 - \rho^2(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{37}$$

Therefore the *conformal* transformation

$$\Omega = (1+X^2)^{1/2}(1+Y^2)^{1/2}/2S \tag{38}$$

maps the original spherical space into half of Minkowski space

$$ds^{*2} = d\tau^2 - d\rho^2 - \rho^2(d\theta^2 + \sin^2\theta\, d\phi^2), \qquad 0 \leqslant \rho < \infty. \tag{39}$$

Since the total mass

$$m = \sum_b m^{(b)}$$

was constant in the Robertson–Walker space the mass m^* in the Minkowski space is variable and is given by

$$m^* = \frac{2Sm}{(1+X^2)^{1/2}(1+Y^2)^{1/2}} = \frac{[1+(\tau+\rho)^2]^{1/2}[1+(\tau-\rho)^2]^{1/2} + \tau^2 - \rho^2 - 1}{[1+(\tau+\rho)^2][1+(\tau-\rho)^2]}\, m. \tag{40}$$

The particle trajectories are R = constant in the Robertson–Walker frame and are given by

$$\frac{2\rho}{1+\tau^2-\rho^2} = \text{constant, } A^{-1} \text{ say,} \tag{41}$$

in the Minkowski frame. In the Robertson–Walker frame the particle density is independent of both R, T. In the Minkowski frame, on the other hand, the density depends on ρ, τ. At $\tau = 0$ it has the form

$$(1+\rho^2)^{-3}. \tag{42}$$

Thereafter it changes according to the trajectories (41). In terms of the constant A the radial coordinate along a trajectory is

$$\rho = (A^2+\tau^2+1)^{1/2}-A, \tag{43}$$

from which we see that $d\rho/d\tau = 0$ at $\tau = 0$, and that $\rho \to \tau$ as $\tau \to \infty$.

How shall we interpret these results? At $\tau = 0$ we have a spherically symmetric cloud with a radial density distribution (42). Initially, at $\tau = 0$, the cloud is at rest, but as τ increases the cloud moves radially outwards and ultimately dissipates itself. It does so, not because of gravitational forces which are zero in this frame, the space being flat, but because $\partial m^*/\partial\tau$, $\partial m^*/\partial\rho$ are non-zero and these mass gradient components appear in the equations of motion (22), where we now put $m_a = m^*$.

It follows from (40) that the mass m^* is zero at $\tau = 0$. m^* increases to a maximum as τ increases and then tends again to zero as $\tau \to \infty$. We interpret this result by saying that m^* is initially zero because mass interactions do not precede $\tau = 0$ and by saying that m^* is finally zero at $\tau \to \infty$ because the cloud dissipates itself and the mass interactions tend to zero.

It comes as no little surprise to find the cloud dispersing to infinity. In the Robertson–Walker frame we think of the universe in the $k = +1$ case as expanding from a singularity to a maximum of the function $S(t)$ and of it then collapsing back into the singularity. We think of the collapse as being a reverse of the expansion, expansion and contraction being determined by the two square roots of (31). What we now see is that these square roots represent a two-one mapping of the Minkowski frame on to the Robertson–Walker frame.

Passing on now to the $k = -1$ case, again we shall find an unexpected result in the Minkowski frame. Now we define $\sinh R = r$. In place of (30) we have

$$ds^2 = dt^2 - S^2(t)[dR^2 + \sinh^2 R(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{44}$$

and in place of (31)

$$\dot{S}^2 = \frac{1+S}{S}. \tag{45}$$

T is again defined by (32), but leading now to

$$S = \tfrac{1}{2}(\cosh T - 1), \tag{46}$$

$$t = \tfrac{1}{2}(\sinh T - T). \tag{47}$$

The appropriate coordinate transformations differ from (36) in the definitions of X, Y,

$$X = \tanh\xi, \qquad Y = \tanh\eta, \tag{48}$$

but are otherwise the same. The coordinate transformations now map the quadrant $0 \leqslant R < \infty$, $0 \leqslant T < \infty$ in the R, T plane into the interior of the triangle (0, 0), (1, 0), (0, 1) in the ρ, T plane. Corresponding to (37) we have

$$ds^2 = \frac{4S^2}{(1-X^2)(1-Y^2)}\,[d\tau^2 - d\rho^2 - \rho^2(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{49}$$

The function required for conformal transformation to Minkowski space is therefore

$$\Omega = (1-X^2)^{1/2}\,(1-Y^2)^{1/2}/2S, \tag{50}$$

$$ds^{*2} = d\tau^2 - d\rho^2 - \rho^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{51}$$

Corresponding to (40) and (41) we have

$$\begin{aligned} m^* &= \frac{2Sm}{(1-X^2)^{1/2}\,(1-Y^2)^{1/2}} \\ &= \frac{1+\tau^2-\rho^2-[1-(\tau+\rho)^2]^{1/2}\,[1-(\tau-\rho)^2]^{1/2}}{[1-(\tau+\rho)^2][1-(\tau-\rho)^2]}\, m, \end{aligned} \tag{52}$$

$$\frac{2\rho}{1+\rho^2-\tau^2} = \text{constant, } A^{-1} \text{ say,} \tag{53}$$

and corresponding to (42), (43),

$$(1-\rho^2)^{-3}, \tag{54}$$

$$\rho = A - (A^2 + \tau^2 - 1)^{1/2}. \tag{55}$$

We again have $d\rho/d\tau = 0$ at $\tau = 0$, but the radial coordinate ρ now tends to zero as $\tau \rightarrow 1$.

The $k = -1$ Friedmann model is therefore to be interpreted as an initially stationary cloud confined to $\rho \leqslant 1$. It collapses to zero radius at $\tau = 1$, and it does so because of an initial singularity in the density of matter (54) as the radial coordinate $\rightarrow 1$. Once again the collapse occurs because of mass gradient components in the equations of motion.

The $k = 0$ model is easily discussed. Beginning once more with the Robertson–Walker space,

$$ds^2 = dt^2 - S^2(t)[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{56}$$

the gravitational equations (28) lead to

$$\dot{S}^2 = \frac{1}{S}. \tag{57}$$

Still defining T by (32),

$$S = \tfrac{1}{4}T^2, \tag{58}$$

and

$$ds^2 = (\tfrac{1}{2}T)^4\,(dT^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2)]. \tag{59}$$

The conformal transformation

$$\Omega = \left(\frac{2}{\tau}\right)^2, \qquad \tau = T, \tag{60}$$

maps the Robertson–Walker space into

$$ds^{*2} = d\tau^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \qquad 0 \leqslant r < \infty,\ 0 \leqslant \tau < \infty, \tag{61}$$

The mass is given by

$$m^* = \tfrac{1}{4}\tau^2 m. \tag{62}$$

The interesting point now emerges that because in the $k = 0$ case the radial coordinate r is not changed matter is still distributed homogeneously in the Minkowski frame. The cases $k = \pm 1$ in this frame correspond only to local clouds, in the $k = -1$ case to a local cloud with a singular density distribution. We emphasize once again in relation to these remarks that the Minkowski frame is equivalent physically to the Robertson–Walker frame (except in one important respect which we shall discuss in the next section in which the Minkowski frame is superior). Hence it is just as appropriate to judge the question of homogeneity in the Minkowski frame as in the Robertson–Walker frame.

4. THE OPPOSITE HALF OF THE UNIVERSE

From here on we work with the Einstein–de Sitter case. The Robertson–Walker $k = 0$ space was transformed in the preceding section into the half $\tau \geqslant 0$ of Minkowski space. This naturally raises the question of the status of the other half of Minkowski space. Suppose instead of starting in Robertson–Walker space we *start* in Minkowski space, and suppose we start with the *whole* of Minkowski space, $-\infty < \tau < \infty$. Now we try to go back to the Robertson–Walker frame using the conformal transformation

$$\Omega = \tfrac{1}{4}T^2. \tag{63}$$

We then arrive in Robertson–Walker space without difficulty except at $T = 0$. Here we have $\Omega = 0$ and this is invalid. We obtain a singularity in Robertson–Walker space because of a zero in the transformation function not from any inherent peculiarity in the physics. We have no difficulty in joining $\tau < 0$, $\tau > 0$ in the Minkowski frame.

The situation we have now encountered recalls the horizon difficulty of the de Sitter cosmology. Instead of working with the Robertson–Walker space, given by putting $S = \exp Ht$, $k = 0$ in (29), de Sitter originally obtained his model in the form

$$ds^2 = (1 - H^2R^2)\, dT^2 - \frac{dR^2}{1 - H^2R^2} - R^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{64}$$

where $H^2 = \frac{1}{3}\Lambda$, Λ being the cosmical constant taken in this case > 0. The *coordinate* transformation relating (29) and (64) is

$$t = T + \frac{1}{2H}\ln(1 - H^2R^2), \qquad R = r\, e^{Ht}. \tag{65}$$

The form (64) is limited in the R, T plane to the strip $R < H^{-1}$, which maps into the region of the r, t plane confined between the curve

$$r = H^{-1}\, e^{-Ht} \tag{66}$$

and the t-axis. Yet we do not restrict the r, t coordinates to such a region when the Robertson–Walker space (29) is used. The r, t coordinates are permitted to take any values in the half plane $-\infty < t < \infty$, $0 \leqslant r$. We pass to the R, T coordinates back through the transformations (65), noticing that the transformation is singular on the curve (66), and taking the view that the restriction on R is only a mathematical construct. Such coordinate singularities are well known in general relativity.

In a similar way we take the view that the singularity at $t = 0$ in a Friedmann model is only a mathematical construct, due to an unfortunate choice of the conformal frame. There is a second half to the Universe which appears without difficulty when the Minkowski frame is used.

5. THE EINSTEIN–DE SITTER MODEL

An important issue remains to be cleared up. The above discussion assumed m constant in the Robertson–Walker frame. We then obtained $m^* = m\tau^2/4$ for the Einstein–de Sitter model in the Minkowski frame. But is this dependence of m^* on τ consistent with

$$m^*(X) = \sum_a m^{*(a)}(X) = \sum_a \int \tilde{G}^*(X, A)\, da^*, \tag{67}$$

i.e. were we justified in assuming m constant in the Robertson–Walker frame?

Since we prefer to work from here on in the Minkowski frame, it will be convenient to reverse the starred and unstarred quantities—unstarred quantities now refer to the Minkowski frame.

First we show that

$$2 \sum_a \int_{\tau_X \geqslant \tau_A \geqslant 0} \tilde{G}(X, A)\, da \tag{68}$$

gives the required answer for $m(X)$. It will then be necessary to relate (68) to (67).

By an appropriate choice of the unit of length we can arrange for the particle density to be unity. A particle a at spatial distance $r \leqslant \tau$ from the field point $X(\mathbf{x}, \tau)$ contributes $1/4\pi r$ to (68), as can be seen by choosing X as the origin of spherical polars and by integrating $2\tilde{G}(X, A) = \delta(s_{XA}{}^2)/2\pi$ along the portion of the particle path $r =$ constant, $\theta =$ constant, $\phi =$ constant, for which $\tau_A \leqslant \tau_X$. All particles with $r \leqslant \tau$ contribute

$$\frac{1}{4\pi} \int_0^\pi \sin\theta\, d\theta \int_0^{2\pi} d\phi \int_0^\tau r\, dr = \tfrac{1}{2}\tau^2. \tag{69}$$

We have to compare (69) with (62). Noting that (62) was based on (57), it can be shown that (62) must be multiplied by $8\pi Gm/3$ in order that τ in (62) be in the same unit as τ in (69). Since $8\pi G = 3/F = 6/m^2$ we have to multiply (62) by $2/m$ to make the unit the same as in (69). Evidently (62) then has the same constant, $\frac{1}{2}$, multiplying τ^2. We write

$$m(X) = \tfrac{1}{2}\tau^2 L^{-3}, \tag{70}$$

where L is the present length unit.

Now we have to enquire into the significance of (68). The situation is clarified by noticing that

$$2 \sum_a \int_{0 \leqslant \tau_A \leqslant \tau_X} \tilde{G}(X, A)\, da = \sum_{\tau_a \leqslant \tau_X} \int \tilde{G}(X, A)\, da, \tag{71}$$

and by interpreting the two sides of (71) in accordance with Fig. 1. The left-hand side is strikingly analogous to the situation in the absorber theory of electrodynamics. Only time retarded contributions to $\tilde{G}(X, A)$ appear in the evaluation because of the condition $\tau_A \leqslant \tau_X$. In the absorber theory, advanced contributions to the electromagnetic potential equal the retarded contributions, which are therefore doubled. This is the situation here also. However, in electrodynamics this doubling depends on a certain perfect absorber condition being satisfied and it is difficult to see how an analogous condition could apply in the present case. The difficulty here is that all particles $a, \ldots$ are being regarded as contributing to $m(X)$ in the same way whereas in electrodynamics the Universe is taken to be made up of equal number densities of positive and negative charges. We do not think the absorber condition could be obtained in electrodynamics if all charges were taken to have the same sign.

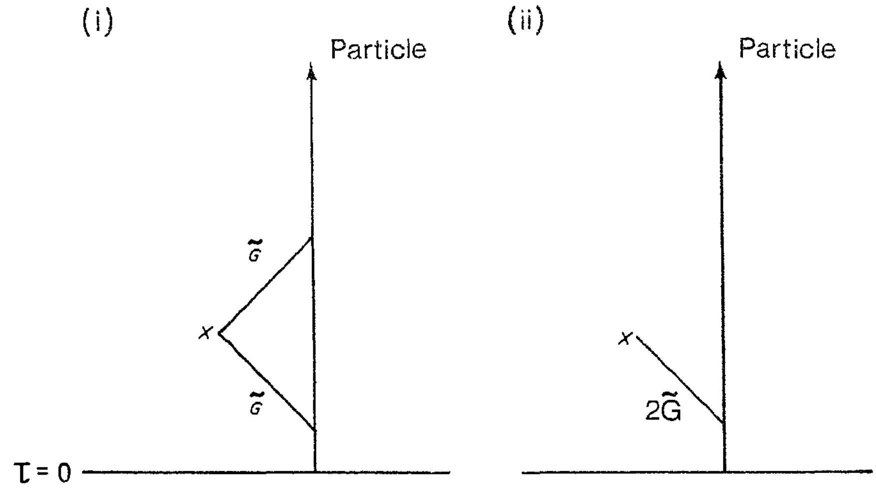

FIG. 1. *(i) illustrates right-hand side of equation* (71), *(ii) illustrates left-hand side. Particle is distant less than* τ_X *from X.*

When m is constructed from the right-hand side of (71) advanced contributions ($\tau_A > \tau_X$) arise as well as regarded contributions. The restriction of the summation to particles $a, \ldots$ such that $r_a \leqslant \tau_X$ requires interpretation. Here r_a is the spatial distance of X from the particle. This condition implies that the past half of the light cone from X intersects the world line of the particle at $\tau \geqslant 0$. Only particles for which this is the case appear in the summation. The condition $\tau_A \geqslant 0$ on the left-hand side of (71) implies a similar restriction. No interaction with the half of the Universe with $\tau < 0$ appears either on the left-hand side or the right-hand side of (71). In spite of what was said in the previous section we could understand this absence of interaction by arguing that the other half of the Universe does not exist. If the Universe begins at $\tau = 0$ then *retarded* interactions with $r_a > \tau_X$ must certainly be absent, but why not advanced interactions with $r_a > \tau_X$? Why are these also absent on the right-hand side of (71)? It does not seem possible to answer this question except through an explicit consideration of the half of the Universe with $\tau < 0$ and through the following modification of the definition of $m(X)$.

Consider the definition

$$m(X) = \sum_a m^{(a)}(X) = \sum_a \int \epsilon_A \tilde{G}(X, A)\, da, \tag{72}$$

where $\epsilon_A = \pm 1$ according as $\tau_A > 0$, $\tau_A < 0$. This is an analogous concept to the positive and negative charges in electrodynamics. We now have the situation illustrated in Fig. 2. There are always two contributions to $m^{(a)}(X)$, from points A, $\tilde{A}$ ($\tau_A > \tau_{\tilde{A}}$) where the light cone† of X intersects a. When $r_a < \tau_X$ both τ_A, $\tau_{\tilde{A}}$ are positive and $\epsilon_A = \epsilon_{\tilde{A}} = 1$, so that the advanced and retarded contributions to $m^{(a)}(X)$ augment each other. But when $r_a > \tau_X$ we have $\tau_A > 0$, $\epsilon_{\tilde{A}} = 1$; $\tau_A < 0$, $\epsilon_{\tilde{A}} = -1$. The advanced and retarded contributions then cancel each other. It follows that only particles $a, \ldots$ such that $r_a < \tau_X$ contribute to $m^{(a)}(X)$, which is just the restriction on the right-hand side of (71).

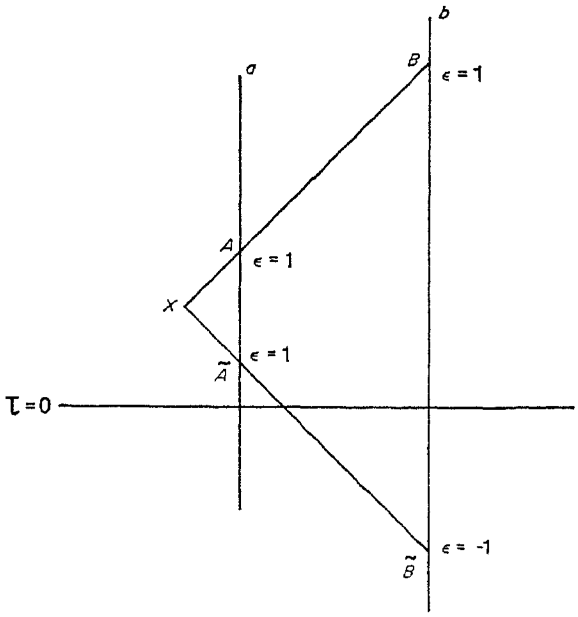

FIG. 2. *Advanced and retarded signals augment for particle* a *(distant* $< \tau_X$*), whereas signals cancel for particle* b *(distant* $> \tau_X$*), because retarded signal in latter case comes from* $\tau < 0$.

It is important to notice that an ϵ factor must also be introduced into the definition of the action in order that (16), (18), (19), (21) continue to hold

$$S = -\sum_a \int \epsilon_A m_a\, da = -\lambda^2 \sum_{a \neq b}\sum \int\int \epsilon_A \epsilon_B \tilde{G}(A, B)\, da\, db. \tag{73}$$

This preserves symmetry with respect to particle pairs. We shall return in the following paper to a consideration of the calculation of $m(X)$. In the Appendix to the present paper we also consider how the above concepts can be incorporated into quantum mechanics.

† Since $\tilde{G}(A, B) = \delta(s_{AB}^2)/4\pi$ in the Minkowski frame the function $\tilde{G}$ propagates along the light cone. The light cone is invariant with respect to conformal transformations.

Since we are working with a conformally invariant theory all relations with observational cosmology in the Minkowski frame must be the same as in the Robertson–Walker frame. It is worthwhile considering an explicit example. We choose the classic Hubble relation between magnitude and redshift. By a standard galaxy we mean one with a specified number of stars each containing a specified number of particles. Light from a galaxy with coordinate r must start on its journey at time $\tau - r$ in order to reach an observer at the space point $r = 0$ at time τ. Write $L(\tau - r)$ for the intrinsic luminosity of the galaxy at that time. The observed flux is just the Euclidean value

$$\frac{L(\tau-r)}{4\pi r^2}. \tag{74}$$

Let the stars in question be taken for simplicity to have opacity from electron scattering. Then it can be shown that the luminosity of a standard star varies with epoch as†

$$G^4 M^3 m^7. \tag{75}$$

The star mass M varies as the mass of its constituent particles, i.e. as m. Since G varies as m^{-2} it follows that (75) varies with epoch as m^2, i.e. as $(\tau - r)^4$ for a star at epoch $\tau - r$. Hence the flux (74) varies with r like

$$\frac{(\tau-r)^4}{r^2}. \tag{76}$$

The redshift $1 + z$ is equal to the ratio of m at the time of observation to m at the time of emission,

$$1+z = \frac{m(\tau)}{m(\tau-r)} = \left(\frac{\tau}{\tau-r}\right)^2. \tag{77}$$

Using (77) to eliminate r in (76), it is seen that the flux from our standard galaxy depends on its red shift according to

$$\frac{1}{1+z}\cdot\frac{1}{[\sqrt{1+z}-1]^2}, \tag{78}$$

which is the usual Hubble relation for the Einstein–de Sitter model. Although the concept of m, G varying with time is unfamiliar, it can hardly be doubted that we have arrived at (78) in an easier way than in the usual treatment. The luminosity dependence (75) can be verified from considerations of stellar structure. More immediately, all we need do is to notice that luminosity has dimensionality L^{-2} and hence must behave as m^2.

Institute of Theoretical Astronomy, Madingley Road, Cambridge

REFERENCES

(1) Hoyle, F. & Narlikar, J. V., 1964. *Proc. R. Soc.* A, **282**, 191.
(2) Hoyle, F. & Narlikar, J. V., 1966. *Proc. R. Soc.* A., **294**, 138.
(3) Infeld, L. & Schild, A., 1945. *Phys. Rev.*, **68**, 250.
(4) Roe, P. E., 1969. *Mon. Not. R. astr. Soc.*, **144**, 219.
(5) Hawking, S. W., 1965. *Proc. R. Soc.* A, **286**, 313.

† The equation of mechanical equilibrium shows that the temperature T behaves like $m.GM/R$, while the radiative transfer equation with electron scattering shows that L behaves like $m^3T^4R^4/M$. The dependence (75) follows by eliminating T.

APPENDIX

In the notation of Section 5 the mass $m(X)$ is given by

$$m(X) = \sum_{\tau_a \leqslant \tau_X} \int \tilde{G}(X, A)\, da. \tag{A1}$$

Define

$$N_\nu(X) = \sum_{\tau_X < \tau_a \leqslant \nu} \int_{\tau_A > 0} \tilde{G}(X, A)\, da. \tag{A2}$$

The mass $m(X)$ is determined by the sum of the advanced and retarded contributions of particles at distances $\leqslant \tau_X$ from the field point X. The quantity $N_\nu(X)$ is a mass given by the advanced potentials of all particles with distances between τ_X and ν. The retarded potentials of the latter particles, viz.

$$\sum_{\tau_X < a \leqslant \nu} \int_{\tau_A < 0} \tilde{G}(X, A)\, da, \tag{A3}$$

is also $N_\nu(X)$, but (A3) contributes with a minus sign so that such particles have a null contribution. And since, trivially,

$$\lim \nu \to \infty\, [N_\nu(X) - N_\nu(X)] = 0 \tag{A4}$$

we regard the effect of the portions $\tau > 2\tau_X$, $\tau < 0$ of the Universe, when taken together, as being null. This avoids the difficulty, pointed out by Hawking (5), that $\lim \nu \to \infty\, N_\nu$ is infinite.

This is the classical situation, the quantity N_ν plays no role. From a quantum point of view, on the other hand, N_ν is important, as will now be seen.

The non-relativistic propagator for a particle to go from a point $X_1(\mathbf{x}_1, \tau_1)$ to a point $X_2(\mathbf{x}_2, \tau_2)$ is of the form

$$F(X_2, X_1) \exp\left[\frac{im(\mathbf{x}_2 - \mathbf{x}_1)^2}{2(\tau_2 - \tau_1)}\right], \tag{A5}$$

where $F(X_2, X_1)$ is a weight function. When X_1, X_2 are infinitesimally separated the exponential factor in (A5) can be written as

$$\exp\left[\frac{i}{2}\, m\dot{\mathbf{x}}^2 . d\tau\right] \tag{A6}$$

where $d\tau = \tau_2 - \tau_1$ and $\frac{1}{2}m\dot{\mathbf{x}}^2$ is the non-relativistic kinetic energy for direct classical motion from X_1 to X_2. The exponent in (A6) is i times the usual action contribution of a classical path element from X_1 to X_2.

It is natural to attempt to generalize (A6) to the relativistic case simply by writing $-mda$ in place of $\frac{1}{2}m\dot{\mathbf{x}}^2 . d\tau$, where da is the element of proper time from X_1 to X_2. This suggests

$$F(X_2, X_1) \exp[-imda] \tag{A7}$$

for the infinitesimal propagator. Given the infinitesimal propagator it is readily possible to build the finite propagator by a path integral method.

However, (A7) does not contain spin and therefore does not lead to the correct propagator for Dirac particles. But spin can be introduced without difficulty by noticing that

$$da^2 = \eta_{ik}\, dx^i\, dx^k = \gamma_i \gamma_k\, dx^i\, dx^k = (\gamma_i\, dx^i)^2, \tag{A8}$$

where $\eta_{ik} = \mathrm{diag}\,(-1, -1, -1, +1)$ is the Minkowski tensor and γ_i ($i = 1, 2, 3, 4$) are the Dirac matrices satisfying

$$\gamma_i\gamma_k + \gamma_k\gamma_i = 2\eta_{ik}. \tag{A9}$$

A unit spin matrix is to be understood as multiplying η_{ik} in both (A8) and (A9), and also da^2 in (A8). Usually we understand da as being given by $(\eta_{ik}\, dx^i\, dx^k)^{1/2}$. This is an ugly and awkward quantity to be entering physics in any fundamental way. More elegantly, we can regard $\gamma_i\, dx^i$ as the square root of (A8). We then have

$$\begin{aligned} &F(X_2, X_1) \exp\,[-imq], \\ &q \equiv \gamma_i(x_2{}^i - x_1{}^i), \end{aligned} \tag{A10}$$

in place of (A7).

The obstacle to obtaining the relativistic propagator by the path integral methods which work so conveniently and elegantly in the non-relativistic case has been that (A10), even though it contains spin, does not give the correct infinitesimal propagator. This turns out to be due to the emission of N_ν. Instead of a single exponential factor, suppose that the two halves of the Universe contribute separate factors,

$$\exp\,[iN_\nu q] - \exp\,[-i(N_\nu + m)q], \tag{A11}$$

the first exponential coming from $\tau < 0$ and the second exponential from $\tau > 0$. Instead of a direct subtraction of the mass contributions, as in the classical discussion of Section 5, we are now subtracting the exponential amplitudes.

It is not hard to show that (A11) can be written as

$$\exp\left[-\frac{i}{2}mq\right]\{\exp\,[i(N_\nu + \tfrac{1}{2}m)\, q] - \exp\,[-i(N_\nu + \tfrac{1}{2}m)\, q]\}. \tag{A12}$$

By expanding the exponentials in the curly brackets in power series this becomes

$$2i\,\frac{\mathbf{q}}{q}.\sin\,[(N_\nu + \tfrac{1}{2}m)\, q].\exp\left[-\frac{i}{2}m\mathbf{q}\right], \tag{A13}$$

where $q^2 = \mathbf{q}^2$ is the invariant 4-dimensional distance from X_1 to X_2. In the form (A13) we are assuming the displacement from X_1 to X_2 to be time-like, so that $q^2 > 0$ and q is real. To deal with the case of a space-like displacement we note that for N_ν very large

$$\frac{1}{q}\sin\,[(N_\nu + \tfrac{1}{2}m)\, q], \qquad \frac{2}{q}\sin\,[(N_\nu + \tfrac{1}{2}m)\, q^2] \tag{A14}$$

are the same function. Hence, since eventually we shall take $\lim N_\nu \to \infty$, we can equally well write

$$4i\,\frac{\mathbf{q}}{q}\sin\,[(N_\nu + \tfrac{1}{2}m)\, q^2].\exp\left[-\frac{i}{2}m\mathbf{q}\right] \tag{A15}$$

as (A13). The form (A15) can be used for space-like displacements, $q^2 < 0$.

Next we notice that the weight function $F(X_2, X_1)$ has dimensionality L^{-3}. Since this function must be invariant with respect to coordinate transformations it has the form

$$F(X_2, X_1) = \frac{\text{constant}}{q^3}. \tag{A16}$$

Multiplying (A15) by (A16), and avoiding $q^2 = 0$ by taking a principal part, we have a quantity of the form

$$(\text{constant}).P.P.\left\{\frac{1}{q^3}\frac{\mathbf{q}}{q}\sin\left[(N_\nu+\tfrac{1}{2}m)\,q^2\right]\exp\left[-\frac{i}{2}m\mathbf{q}\right]\right\}. \tag{A17}$$

The function (A17) is the same in the limit $N_\nu \to \infty$ as

$$\text{constant}.\,\mathbf{q}\delta'(q^2).\exp\left[-\frac{i}{2}m\mathbf{q}\right], \tag{A18}$$

where $\delta'(q^2) = d\delta(q^2)/dq^2$. Since the displacement $\mathbf{q}$ is infinitesimal

$$\exp\left[-\frac{i}{2}m\mathbf{q}\right] \simeq 1-\frac{i}{2}m\mathbf{q}, \tag{A19}$$

and (A18) has the form

$$\text{constant}.\left[\mathbf{q}\delta'(q^2)+\frac{i}{2}\,m\delta(q^2)\right]. \tag{A20}$$

Here we have used $q^2\delta'(q^2) = -\delta(q^2)$. With a suitable choice of the normalizing constant and of the sign of the γ_i matrices (A20) is the usual infinitesimal propagator for Dirac particles.

From a classical point of view the quantity N_ν is an irrelevant nuisance but from the point of view of relativistic quantum mechanics N_ν plays an essential and crucial role. It is encouraging to this point of view that the factor $\frac{1}{2}$ in (A20) has appeared in the step from (A11) to (A12). Although a detail, the appearance of this factor is indicative of a correct concept.