A SINGULARITY-FREE EMPTY UNIVERSE

C. W. MISNER*

Department of Applied Mathematics and Theoretical Physics, Cambridge Univ., Cambridge England

and

A. H. TAUB

Computer Center, University of California, Berkeley, California

Submitted December 18, 1967

Zh. Eksp. Teor. Fiz. 55, 233–255 (July, 1968)

A solution of the empty space Einstein equations $R_{\mu\nu} = 0$ is studied. It is a cosmological model in the sense that it incorporates a metric which has closed homogeneous space-like hypersurfaces which expand anisotropically. It is everywhere analytic (C^{ω}) with Lorentz +++− signature and is shown to be non-singular in the sense of "distant boundaries" in that every geodesic are which approaches the boundaries (i.e., which cannot be contained in any compact set) has infinite affine length. However it also incorporates the Newman, Unti and Tamburino metric (NUT metric) which contains closed time-like and null lines. The analytic continuation from the cosmological model to NUT space can be done in (at least) two inequivalent ways. The physical evolution toward NUT space is unstable and short wavelength perturbations (gravitons, photons, etc.) are accelerated to disruptive energies before the Cauchy horizon separating the cosmological and the NUT regions is attained. This instability is shown by the same behavior of time-like and null geodesics which shows that this space-time is not geodesically complete, and that no analytic continuation of it can be. This space-time is also considered here in connection with the topic of the prevalence and nature of singularities in general cosmological solutions of the Einstein equations.

THIS paper studies a particular solution of the empty-space Einstein equations, discusses some ways in which it is singular and is non-singular, and describes all its geodesics.

The metric is presented in Sec. I, its singularities and their relationship to broader questions about singularities in general relativity and cosmology are discussed in Sec. II, with essential features concerning the behavior of the geodesics quoted from Sec. III where they are established.

I. THE METRIC

We wish to define ds^2 in such a way that it provides an analytic metric of Lorentz +++− signature everywhere on $R \times S^3$. To begin we obtain the space $R \times S^3$ with the desired structure as an analytic manifold[1)] by removing the point at the origin from the space R^4 of quadruples of real numbers (w, x, y, z) where the idea of an analytic function is clear. In particular the two functions t and $|q|^{-2}$ whose definitions follow from

$$e^t = |q|^2 = w^2 + x^2 + y^2 + z^2, \tag{1}$$

are then analytic everywhere (on $R \times S^3$). Consequently the following differential forms[2)] are also analytic (everywhere on $R \times S^3$):

$$\begin{aligned}
\sigma_x &= 2|q|^{-2}(xdw - wdx - zdy + ydz),\\
\sigma_y &= 2|q|^{-2}(ydw + zdx - wdy - xdz),\\
\sigma_z &= 2|q|^{-2}(zdw - ydx + xdy - wdz),\\
dt &= 2|q|^{-2}(wdw + xdx + ydy + zdz).
\end{aligned} \tag{2}$$

For any function $U(t)$ analytic on $-\infty < t < +\infty$, and any positive real constant l then, the tensor field defined by

$$ds^2 = (t^2 + l^2)(\sigma_x^2 + \sigma_y^2) + U(t)(2l)^2\sigma_z^2 + 2(2l)\sigma_z dt, \tag{3}$$

is analytic. To show that ds^2 is a metric we must check its signature. Now the differential forms of Eq. (2) are linearly independent as one verifies for instance from

$$\sigma_x \wedge \sigma_y \wedge \sigma_z \wedge dt = 2^4|q|^{-4}dw \wedge dx \wedge dy \wedge dz \neq 0. \tag{4}$$

We therefore choose them as basis vectors, in which case the corresponding components $g_{\mu\nu}$ of ds^2 can be read from Eq. (3) as

$$g_{\mu\nu} = \begin{pmatrix} t^2+l^2 & \cdot & \cdot & \cdot \\ \cdot & t^2+l^2 & \cdot & \cdot \\ \cdot & \cdot & 4l^2U(t) & 2l \\ \cdot & \cdot & 2l & \cdot \end{pmatrix}. \tag{5}$$

*Permanent address: Department of Physics and Astronomy, University of Maryland, College Park, Md.

1)Some of the questions considered in this paper are global, rather than local, and thus require elementary concepts from differential geometry in precise forms not detailed in the classical texts on general relativity. The definitions we require may be found in the first seventeen page of [1] or in [2]. Sections 5.1 through 5.4 of [3] are lighter reading, and Chapter VI of [4] is not difficult.

2)We identify the differential form $\lambda = \lambda_\mu dx^\mu$ whose coefficients are λ_μ with the covariant vector whose components with respect to the x^μ coordinate basis are λ_μ. Thus Eqs. (2) define a tetrad of covariant vectors σ^α (with $\sigma^1 = \sigma_x$, etc., $\sigma^4 = dt$) and writing $\lambda = \lambda_\alpha\sigma^\alpha$ would display the components of λ with respect to this tetrad. Similarly by $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ we identify the quadratic form ds^2 with the tensor whose coordinate components are $g_{\mu\nu}$. Later we will make use of contravariant vectors and identify them with corresponding differential operators $v = v^\nu(\partial/\partial x^\mu)$. See [1], [2], [3], or [4].

From this matrix it is clear that $g = \det g_{\mu\nu}$ is negative and that the signature is $+++-$. One has, in fact,

$$(-g)^{1/2} = 2l(t^2 + l^2). \tag{6}$$

when $l > 0$, which we shall always require. Therefore ds^2 is an analytic metric on $R \times S^3$. With the choice

$$U(t) = -1 + \frac{2(mt + l^2)}{t^2 + l^2} = \frac{(t - t_1)(t_2 - t)}{t^2 + l^2} \tag{7}$$

where

$$t_1 = m - (m^2 + l^2)^{1/2}, \quad t_2 = m + (m^2 + l^2)^{1/2}, \tag{8}$$

this metric satisfies $R_{\mu\nu} = 0$.

Although it is clear from the foregoing paragraph that the components of ds^2 in the wxyz coordinate system are analytic functions and form a matrix of the desired signature, it is not clear that any benefit would derive from writing them out explicitly. Many computations are possible using only the formula $d\sigma_z = \sigma_x \wedge \sigma_y$ (and its cyclic permutations) for the exterior derivatives of the σ_a. When a coordinate basis is desirable for other computations, it is usually better to sacrifice manifest analyticity[3)] and use the coordinates $t\psi\theta\varphi$ defined by

$$w = e^{t/2}\cos\frac{\theta}{2}\cos\frac{\psi+\varphi}{2}, \quad x = e^{t/2}\sin\frac{\theta}{2}\sin\frac{\psi-\varphi}{2},$$

$$y = e^{t/2}\sin\frac{\theta}{2}\cos\frac{\psi-\varphi}{2}, \quad z = e^{t/2}\cos\frac{\theta}{2}\sin\frac{\psi+\varphi}{2}. \tag{9}$$

One then finds for ds^2 the expansion[4)]

$$ds^2 = (t^2 + l^2)(d\theta^2 + \sin^2\theta d\varphi^2) + U(t)(2l)^2(d\psi + \cos\theta d\varphi)^2 - 2(2l)(d\varphi + \cos\theta d\varphi)dt \tag{10}$$

in which the metric components are rather simple. The angles $\psi\theta\varphi$ introduced through Eqs. (9) are the classical Euler angle coordinates [7] on the rotation group SO(3), but since we are considering the simply connected covering space S^3 instead of SO(3), we allow ψ to have a fundamental range of $0 \le \psi < 4\pi$ while θ and φ have their usual ranges of $0 \le \theta < \pi$ and $0 \le \varphi < 2\pi$.

From Eq. (3) or (10) it is easy to see the signature of the metric induced on any $t = \text{const}$ hypersurface. For each t in the range

$$t_1 < t < t_2, \tag{11}$$

where $U(t) < 0$ the hypersurface is space-like. The two hypersurfaces where $U(t) = 0$ are null, and for all other values of t the hypersurfaces are time-like. By Eq. (1) the condition $t = \text{const}$ defines a closed three-sphere S^3 in each case, so it is not hard to find closed time-like curves when $U(t) < 0$. Each curve on which t, θ, and φ are constant while ψ varies from 0 to 4π is closed, and is time-like when $U(t) < 0$. That this is an analytic curve can be seen by expressing the transformation (9) in quaternion [7] form:

$$q = w + ix + jy + kz = e^{t/2}e^{k\varphi/2}e^{j\theta/2}e^{k\psi/2}. \tag{12}$$

For any given constant quaternion p with $|p| = 1$ the transformation L_p: $q \to L_p q = pq$ of $R \times S^3$ into itself leaves each of the forms (2) and the metric (3) invariant [6], and these transformations form a simply transitive group of motions on each $t = \text{const}$ hypersurface. The transformation $q \to qe^{-k\alpha/2}$, where α is a real constant, is a further isometry. These additional isometries form a one-parameter subgroup of the group of right translation $R_p: q \to qp^{-1}$, consisting of those where p has the form $p(\alpha) = e^{k\alpha/2}$. This restriction on the p for which R_p is an isometry contrasts with the more highly symmetric Robertson-Walker metrics of positive curvature where all the R_p, as well as all L_p, are isometries. It was a requirement of homogeneous space-like hypersurfaces that originally led Taub [8] to a metric locally equivalent to Eq. (10), but in a form valid only for the region $U(t) > 0$ where the hypersurfaces are in fact space-like. This space has been discussed as a cosmological model first by Heckmann and Schucking,[9] and then by Behr,[10] by Brill,[11,12] and by Wheeler [13]. It has closed space-like sections topologically identical to those of the positive curvature Robertson-Walker or Friedmann models, but the expansion rate is not the same in all directions, and it satisfies the Einstein equations with neglect of the matter density, Joseph [14] in 1957 has already pointed out that the curvature tensor remained regular at the limits [corresponding to (11)] of the coordinate systems then available. Lifshitz and Khalatnikov [15] also studied the nature of the metric at these limits (example 4 in their Appendix H) and found the singularity there to be fictitious. A metric form valid in the regions where $U(t) < 0$ was obtained for quite unrelated reasons by Newman, Unti, and Tamburino [16] (NUT) in 1963, in a form which reduced to the Schwarzschild metric in the limit $l \to 0$. Misner [6] recognized that the NUT metric contained arbitrary removable singularities unless one let the $t = \text{const}$ hypersurfaces be closed, and Brill [6] suggested its relationship to the Taub cosmology, after which the present authors[5)] obtained the forms presented above which show the $U(t) < 0$, NUT-space, regions joined analytically to the $U(t) > 0$ Taub cosmology. A generalization in which space-time is not empty, but contains an electromagnetic field has been given by Brill.[11] In that case Eq. (7) is to be replaced by

$$U_B(t) = -1 + 2(mt + l^2 - {}^1/_8\Phi^2)(t^2 + l^2)^{-1}. \tag{13}$$

A similar generalization, maintaining the same high symmetry, to include a neutrino field is not possible.[17]

II: SINGULARITIES

For metrics with the Lorentz signature it is not clear how one should define the words "singular" and "non-singular." A discussion of the difficulties is given in the introduction of [6]. It may be that different definitions will be appropriate for different purposes, or that some definition will eventually appear to be overwhelmingly most suitable for physically interesting situations. At present we are content to consider several alternate viewpoints.

As a preliminary to this discussion let us define: A space-time U is a pair $\mathcal{U} \equiv (U, ds^2)$ where U is a connected differentiable 4-manifold and ds^2 is a metric tensor field of Lorentz $(+++-)$ signature defined and differentiable everywhere on U. In this definition and throughout the paper we take differentiable to mean C^∞ for simplicity; we frequently also consider the analytic C^ω case. The symbol M we re-

3) For some exercises on recognising differentiability when it is not manifest, see the article "Taub-NUT Space as a Counterexample to Almost Anything" by C. W. Misner in [5].

4) This computation is simplified by the use of quaternion notation for the forms of Eq. (2) as in Eqs. (52) of [6] and Eq. (12) below.

5) See a note added in proof to [6] for the coordinate transformation relating Eq. (10) above to previously used coordinate systems.

serve for the specific space-time $M \equiv (R \times S^3, ds^2)$ where ds^2 is the metric field on $R \times S^3$ defined by Eqs. (2) and (3). To avoid misunderstanding we emphasize Eq. (10) defines the same metric ds^2 by displaying its components with respect to another of the many local coordinate systems compatible with the differentiable structure[1] of $R \times S^3$. On the other hand, by interchanging σ_y and σ_z wherever they appear in Eq. (3) one would define a different metric $d\tilde{s}^2$ (i.e., $\tilde{g}_{\mu\nu} \neq g_{\mu\nu}$, where the components of both metics are referred to the same coordinate system), and we maintain the mathematical distinction $M \neq \tilde{M} \equiv (R \times S^3, d\tilde{s}^2)$. Of course M and $\tilde{M}$ are physically equivalent since the mapping $\mu: M \rightarrow M$ defined by $(w, x, y, z) \rightarrow (w, -x, z, y)$ is an isometry of M onto $\tilde{M}$, that is $\mu_* d\tilde{s}^2 = ds^2$. In fact for any global diffeomorphism μ of $R \times S^3$ onto itself, the map $\mu: R \times S^3 \rightarrow M$ defines a new space-time $M_\mu \equiv (R \times S^3, \mu_* ds^2)$ which is geometrically and physically equivalent to M_μ. But $M = M_\mu$ is satisfied only by the 4-parameter family of maps μ which define the symmetry of M.

A. Geometrically Non-singular Space-time

Since the metric of Eq. (3) is analytic everywhere as presented, the most obvious place to look for singularities is at $t = \pm\infty$. A computation of the Riemann tensor (see [6], Appendix A), however, shows that all its components in the orthonormal frame used in Eq. (28) below vanish as t^{-3} for $|t| \rightarrow \infty$, so we have no basis for asserting that there is a singularity at $|t| = \infty$. The example $ds^2 = (1 + \zeta^2)^{-2} d\zeta^2$ from one dimension reminds us, however, that although nothing unusual is happening at the bounaries $\zeta = + \infty$, there could still be a lot of space left unexplored beyond them. This is shown by the alternate form $ds^2 = d\xi^2$ obtained with $\zeta = \tan \xi$, for then the previous boundaries became $|\xi| = \pi/2$. We must therefore ask whether our space-time M (consisting of the manifold $R \times S^3$ with the metric of Eq. (3) assigned) could be a part of some larger (and possibly singular) space-time M^+. A negative answer is provided by the proposition that M is maximal.

The statement of this proposition presumes the definition: A space-time U is maximal if and only if there exists no space-time U^+ with the property that U can be identified with a proper open metric submanifold of U^+. The proof of the proposition follows: suppose M were not maximal, then we could assume $M \subset M^+$ and the boundary of M in M^+ would not be empty. Consider then, for a boundary point $p \in M^+ - M$, an arbitrarily small neighborhood defined by $\Sigma(z_\mu)^2 < \epsilon$ where z_μ are a Riemann normal coordinate system (cf. Sec. 6.3 of [1]) at p. Then this neighborhood of p contains some point $z_\mu = a_\mu$ of M (else p could not be a boundary point), and $z_\mu(\lambda) = (1 - \lambda) a_\mu$ for $0 \leq \lambda \leq 1$ is a geodesic in M^+ from this point in M to p. Since p is not in M there exists a geodesic in M (a part of this one) which approaches the boundary of M while its affine parameter remains bounded. The preceding sentence is a statement referring only to M and we must therefore be able to find the geodesic it refers to by studying all the geodesics of M. This study is carried out in Sec. III, where we find to the contrary that for every geodesic arc on which the affine parameter λ is bounded, so is the coordinate t. Alternatively, we find that one can only have a geodesic on which $|t| \rightarrow \infty$ by allowing $|\lambda| \rightarrow \infty$. Thus M is regular at every point, and is not part of any larger space-time in which we could search further for singularities so we assert that M is geometrically non-singular.

The above arguments can be stated more generally by defining: A space-time U has distant boundaries if every finite geodesic arc (given for some finite interval $a < \lambda < b$ of its affine parameter λ) can be contained in a compact subset of U. Then by the arguments just given it follows that a space-time with distant boundaries is always maximal. The converse is not true, as is shown by the Friedmann universes, or the Kruskal extension [18] of the Schwarschild solution. These space-times are maximal, but they contain geodesic arcs with λ bounded on which some curvature scalars are unbounded. But the curvature scalars (i.e., scalar polynomials in the curvature tensor and its covariant derivatives) are differentiable functions, and must therefore remain bounded on every compact set. Consequently, the finite geodesic arcs on which the curvature is unbounded cannot be contained in any compact set, so that maximal space-times do not have distant boundaries.

It seems geometrically reasonable to accept a space-time with distant boundaries as being non-singular, because this definition has several desirable properties: [6] (a) if p is some point of a space-time U, then U − {p} does not have distant boundaries, (b) if any curvature scalar can approach infinity along any finite geodesic arc, then the space-time does not have distant boundaries, (c) A compact space-time has distant boundaries, (d) a geodesically complete space-time has distant boundaries.

B. Geodesically Incomplete Space-time

The requirement (geodesic completeness) that all geodesics can be continued to infinite positive and negative values of the affine path parameter is stronger than the requirement of distant boundaries since there are [6] closed (compact) analytic space-times which are not geodesically complete. From a mathematical-geometrical point of view this may suggest that completeness is too strong a specification of non-singular behavior. For the analyst studying hyperbolic differential equations, or for a physicist, however, it will appear that "distant boundaries" is too weak a requirement, as it is compatible with pathological behaviors such as those we shall see in our space-time M.

In Fig. 1 we have sketched the light cones in M. The sketch is based on the fact, easily verified from Eq. (10), that the vectors **k** and ***l*** defined by[2]

$$\mathbf{k} = \frac{\partial}{\partial t}, \quad \mathbf{l} = U\frac{\partial}{\partial t} + \frac{1}{l}\frac{\partial}{\partial \psi}, \tag{14}$$

are null vectors. Since they lie in surfaces of constant θ and φ, they can be drawn on a two dimensional diagram as in Fig. 1 to describe the (projected) edges of the null cone at any point. In wxyz coordinates one has

$$\begin{aligned} \partial_t &= {}^1\!/_2 (w\partial_w + x\partial_x + y\partial_y + z\partial_z), \\ \partial_\psi &= {}^1\!/_2 (-z\partial_w + y\partial_x - x\partial_y + w\partial_z), \end{aligned} \tag{15}$$

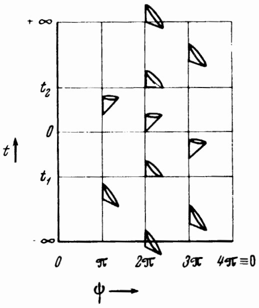


FIG. 1. The future null cones are shown projected onto a $t\psi$ surface. One sees how the t = const. hypersurfaces change continuously from time-like at $t = \pm\infty$ to space-like at t = 0 with null hypersurfaces intervening at $t = t_1$ and $t = t_2$.

where $\partial_w = \partial/\partial_w$, etc. These formulae show that ∂_t and ∂_ψ are linearly independent analytic vector fields everywhere on $R \times S^3$, so the same is true of **k** and l; also

$$\mathbf{v} = {}^1\!/_2(\mathbf{k} + \mathbf{l}) \tag{16}$$

is an analytic time-like unit vector:

$$\mathbf{k}^2 = 0 = \mathbf{l}^2, \quad \mathbf{k}\cdot\mathbf{l} = -2, \quad \mathbf{v}^2 = -1. \tag{17}$$

At any point the null cone which contains **v**, **k**, and l will be called the forward or future null cone; we thus define a time orientation on M. We will always assume that any time-like or null geodesic on M is parameterized so that its tangent vector lies in the forward light cone. Now one might expect that the two forward null vectors **k** and l whose directions are uniquely selected by the symmetry of this space-time, would be parallel to families of null geodesics. In Sec. III we verify this. However only **k** satisfies the geodesic equation $k^\mu_{;\nu}k^\nu = 0$, while l requires a renormalization. We find that $q = U^{-1}l$ does satisfy the geodesic equation $q^\mu_{;\nu}q^\nu = 0$. But from

$$\mathbf{q} = \partial_t + (lU)^{-1}\partial_\psi \tag{18}$$

and the fact that ∂_t and ∂_ψ are non-vanishing analytic vector fields, we see that **q** is singular when $U(t) = 0$. From either Eq. (18) or Fig. 1 it follows that a geodesic tangent to **q** on the t = 0 hypersurface (and hence everywhere) cannot leave the region where $U(t) > 0$, and from Eq. (18) these geodesic equations may be written

$$dt/d\lambda = 1, \quad d\psi/d\lambda = 1/lU. \tag{19}$$

It is then clear that the geodesic can be extended for a range of values of λ exactly equal to its range in t, namely $\lambda_{max} - \lambda_{min} = 2(m^2 + l^2)^{1/2}$ which is finite, so M is not geodesically complete. At the ends of this range one sees that ψ becomes logarithmically infinite. Thus this geodesic circles around the space infinitely many times as it approaches the horizons [19,20,21] where $U(t) = 0$, and does this while λ remains bounded.

Many other time-like, null, and space-like geodesics will be found to behave in this way in Sec. III; the null case was presented here for simplicity. But not all geodesics have this wild behavior near $U(t) = 0$, as is shown by the example of the null geodesics tangent to **k** whose paths are given by $t = \lambda$, ψ = const, θ = const, φ = const, and $-\infty < \lambda < +\infty$.

C. Non-unique Analytic Continuation

It is not reasonable to demand analyticity (in real configuration space) in solutions of the hyperbolic partial differential equations of physics, because then a function would be determined uniquely at any point not only by Cauchy data in the past light cone of this point, but equally well by analytic Cauchy data in a region space-like from the point in question. We have nevertheless emphasized the analyticity of the space-time under study here, because this is the only way we have of asserting a relationship between the regions where $U(t) > 0$ and those where $U(t) < 0$. The uniqueness theorem [22] for solutions of the Einstein equations shows that the entire region where $U(t) > 0$ is determined uniquely by the induced metric and the second fundamental form of, say, the hypersurface t = 0. But since the hypersurfaces where $U(t) = 0$ are null, the metric beyond them is not uniquely related to the t = 0 initial data. We are about to see that even analyticity, which introduces non-causal relationships, does not succeed in determining the continuation from $U(t) > 0$ to $U(t) < 0$.

Consider a second copy of $R \times S^3$ on which the fundamental analytic coordinates corresponding to wxyz will be distinguished by primes. We make this analytic manifold into a space-time M′ by defining on it a metric

$$(ds^2)' = (t'^2 + l^2)(\sigma_x'^2 + \sigma_y'^2) + U(t')(2l)^2\sigma_z'^2 - 2(2l)\sigma_z'dt', \tag{20}$$

where the differential forms σ_a' and dt' are defined by primed copies of Eqs. (2). Apart from the primes, the only difference between Eqs. (20) and (3) is a sign in the last term. In the region $U(t) > 0$ we can relate these two space-times by the isometry described in quaternions (c.f. Eq. 12) by

$$q' = q \exp\left\{-k\int_0^t (2lU)^{-1}dt\right\} \tag{21}$$

or in Euler angle coordinates by[6)]

$$t' = t, \quad \varphi' = \varphi, \quad \theta' = \theta, \quad \psi' = \psi - l^{-1}\int_0^t U^{-1}(t)\,dt. \tag{22}$$

From Eq. (21) it is evident that this mapping is analytic in the region where $U(t) > 0$, and from Eqs. (22) and (10) one easily verifies that it is an isometry. Now apply the differential operators ∂_t and ∂_ψ to Eqs. (22) to compute $\partial_\psi = \partial/\partial_{\psi'}$ and $\partial_t = (\partial/\partial_{t'}) - (lU)^{-1}(\partial/\partial_{\psi'})$, so the vector fields on M′ identified under this isometry with **k** and **q** on M are

$$\mathbf{k} = \frac{\partial}{\partial t'} - \frac{1}{lU}\frac{\partial}{\partial\psi'}, \quad \mathbf{q} = \frac{\partial}{\partial t'}. \tag{23}$$

Thus on M′ it is **q** which is analytic, and **k** which has a singularity where $U(t') = 0$. The null geodesics tangent to **q** which could not be extended beyond a parameter interval of $2(m^2 + l^2)^{1/2}$ in M can now be defined on M′ for all values of λ and extend into the region of M′ where $U(t') < 0$. But the geodesics tangent to **k**, which were well behaved in M now show the infinite spiraling at finite parameter values as they approach the $U(t') = 0$ horizons in M′. Thus, although the parts of M and M′ defined by $U(t(> 0)$

---

6)Note that Eq. (22) gives a coordinate description of a mapping between two different space-times M and M′ and is not a coordinate transformation. These $t\phi\theta\psi$ are coordinates on M but $t'\phi'\theta'\psi'$ are coordinates on M′. We do not find it helpful to avail ourselves of the possibility of also using Eq. (22) to define sets of coordinates $t'\phi'\theta'\psi'$ on parts of M.

are identical, this identification does not extend to the full analytic space-times M and M′.

Most geodesics behave either like the null geodesics tangent to **k**, or else like those tangent to **q**, when they cross one of the horizons for the first time, in that they can be continued somewhat farther in one or the other of the space-times M, M′. Some geodesics, however, can be continued in neither. Consider the vector field

$$\mathbf{n} = U^{1/2}(\mathbf{k} + \mathbf{q})/2, \tag{24}$$

which is a time-like vector, $\mathbf{n}\cdot\mathbf{n} = -1$. From

$$\mathbf{n} = U^{1/2}\left[\frac{\partial}{\partial t} + \frac{1}{2lU}\frac{\partial}{\partial\psi}\right] = U^{1/2}\left[\frac{\partial}{\partial t'} - \frac{1}{2lU}\frac{\partial}{\partial\psi'}\right] \tag{25}$$

one can verify that n is normal to the hypersurfaces t = const. In fact, n is geodesic with $n^{\mu}_{;\nu}n^{\nu} = 0$. It is also clear from Eqs. (25) that these geodesics require only a finite proper time to go from one to the other of the two null hypersurfaces $U(t) = 0$, but cannot be continued across them in either of the two space-times M nor M′.

We are not aware of any proof that no other analytic maximal continuation of the $U(t) > 0$ region exists, in addition to M and M′.

D. Stability

Stability is a particularly important property for a cosmological model for such a model usually involves various idealizations which may be only approximately satisfied. Thus an empty cosmological model in order to be acceptable must approximate one in which the stress-energy tensor is not precisely zero but merely negligibly small if for no other reason but to allow for some observers in the universe. In order to show that a given model is stable one must show that all perturbations of the metric tensor created by small changes in the initial defining data of the model remain small. If one can show that a single initial perturbation does not remain small one can prove the instability of a given model.

In the following we shall show that the space-time M is unstable by embedding it into a family of non-empty space-times M(e) each of which has all the symmetry properties of M and show that for some of these space-times small initial defining data does not remain small. This family is defined as follows: Under the transformation of coordinates

$$\bar{t} = t,\quad \bar{\varphi} = \varphi,\quad \bar{\theta} = \theta,\quad \bar{\psi} = \psi - (2l)^{-1}\int_0^t U^{-1}(t)\,dt$$

the line element given by equation (10) becomes

$$ds^2 = -U^{-1}dt^2 + (2l)^2 U(d\bar{\psi} + \cos\theta\, d\varphi)^2 + W(d\theta^2 + \sin^2\theta\, d\varphi^2), \tag{26}$$

with $W = t^2 + l^2$ and $U(t)$ given by equation (7)

The family of space-times M(e) is defined by specifying that each member be associated with the same 4-dimensional connected manifold as M and have the metric tensor defined by equation (26) with U and W functions of t which are to be determined from the field equations for non-empty space-time. It is evident that each of the space-times in this family admits the same four Killing vectors as does the space-time M.

It may be verified that the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{R}{2}g_{\mu\nu}$$

is given by

$$G_{00} = -\frac{1}{2W}\left[\dot{U}\dot{W} + \frac{1}{2}U\frac{\dot{W}^2}{W} + 2 - 2l^2\frac{U}{W}\right],$$

$$G_{11} = \frac{1}{W}\left[U\ddot{W} + \frac{1}{2}\dot{U}\dot{W} - 2l^2\frac{U}{W} - \frac{1}{4}U\frac{\dot{W}}{W} + 1\right],$$

$$G_{22} = \frac{1}{2W}\left[W\ddot{U} + U\ddot{W} + \dot{U}\dot{W} + l^2\frac{U}{W} - \frac{1}{2}U\frac{\dot{W}^2}{W}\right]. \tag{27}$$

where we have used the notation $\dot{f} = df/dt$, and the indices on $G_{\mu\nu}$ refer not to the coordinates but to the orthonormal bases ω^{μ} defined by

$$\omega^0 = U^{-1/2}dt,\quad \omega^1 = 2lU^{1/2}(d\bar{\psi} + \cos\theta\, d\varphi),$$
$$\omega^2 = W^{1/2}d\theta,\quad \omega^3 = W^{1/2}\sin\theta\, d\varphi. \tag{28}$$

That is, in this basis, the Einstein tensor is diagonal and satisfies the condition $G_{22} = G_{33}$. Further, its components relative to the tetrad ω^{μ} are functions only of the variable t. These facts are consequences of the symmetry requirement imposed on the space-times M(e).

The Einstein equations are

$$G_{\mu\nu} = 8\pi T_{\mu\nu},$$

since we use units in which G = c = 1. Hence $T_{\mu\nu}$ must have the same properties as $G_{\mu\nu}$. Thus if it is the stress-energy tensor of a perfect fluid, the energy density w and pressure p of this fluid must each be functions of t alone. Hence there will exist an equation of state for the fluid, an equation of the form

$$p = p(w).$$

will hold. As is well known it then follows from the conservation equations

$$T^{\mu\nu}_{;\nu} = 0$$

and the form of the stress-energy tensor

$$T^{\mu\nu} = (w + p)u^{\mu}u^{\nu} + pg^{\mu\nu}$$

that

$$(\sigma u^{\nu})_{;\nu} = 0, \tag{29}$$

where $\sigma(p)$ (or $\sigma(w)$) is such that

$$d\sigma/dw = \sigma/(w + p).$$

Thus there is a property of the fluid represented by the function σ which is conserved during the motion and the evolution of the space-time. In case p = 0, the material present is dust and one has $\sigma = w = \rho$ the matter density.

An equation of the form of equation (29) must often be required in addition to the conservation equations $T^{\mu\nu}_{;\nu} = 0$ in order to obtain a complete description of a model. For example when the stress-energy tensor describes a fluid with heat conductivity or viscosity, an equation of state of the form $w = w(p)$ need not exist. However in such a model one requires that matter be conserved in which case one must have

$$(\rho u^{\mu})_{;\mu} = 0,$$

and

$$w = \rho(c^2 + \varepsilon),$$

with ϵ the specific internal energy of the matter.

The equations for $G_{\mu\nu}$ enable us to conclude that if the stress-energy tensor in a space-time of the family $M(e)$ is that of a perfect fluid then the four-velocity of that fluid $\mathbf{u}$ is given by

$$\mathbf{u} = U^{1/2}\partial_t.$$

In case $M(e) = M(U(t)$ is given by equation (7)) we have

$$\mathbf{u} = \mathbf{n}.$$

Further, the equation of motion which may be written as

$$(w + p)\, u^{\mu}_{;\sigma} u^{\sigma} = - p_{,\sigma}(g^{\sigma\mu} + u^{\sigma}u^{\mu}),$$

reduce to

$$u^{\mu}_{;\sigma} u^{\sigma} = 0,$$

since $p = p(t)$.

Whenever an equation of the form of equation (29) holds there is a conserved integral

$$m = \int \sigma u^t (-g)^{1/2} d^3x = \sigma(t)\, u^t (2l)\, W (4\pi)^2 = \text{const.} \qquad (30)$$

which we have evaluated on a t = constant hypersurface using the metric of equation (26) for which $(-g)^{1/2} = W(2l)\sin\theta$. This equation gives the behavior of the scalar function σ

$$\sigma(t) = m \,/\, U^{1/2}(t)\, W(t)\, (2l)\, (4\pi)^2.$$

Therefore, even arbitrarily small initial values $\sigma(0)$, corresponding to arbitrarily small values of the constant M will lead to large values of σ as $t \to t_2$ if $U(t_2) = 0$.

By choosing M small enough we may make the components of $T_{\mu\nu}$ small compared to the components of the curvature tensor of the space-time M. In that case, the space-time M is one approximating $M(e)$ and $U(t)$ is given by equation (7) and $W = t^2 + l^2$. It may be verified that these functions satisfy the equations $G_{\mu\nu} = 0$ where $G_{\mu\nu}$ is given as above. For this $U(t)$ we have $U(t_2) = 0$ and hence the space-time M is not stable for the following reason. If one perturbs M by introducing a small amount of a perfect fluid (or any other material for which a conserved integral exists), in such a way that the symmetry of M is preserved, and thus creates a space-time $M(e)$, then at a finite time t less than or equal $t = t_2$ the space-time M will no longer be a reasonable approximation to the space-time $M(e)$.

In addition arguments based on Raychaudhuri's equation [9,23,24] and the computations of Behr [10,25] show that for the space-times $M(e)$ which contain only dust, an infinity in ρ, the density of the dust, must occur within a finite time.

The instability of the space-time M against perturbations due to the introduction of a small amount of matter represented by a stress energy tensor of a perfect fluid with an equation of state is traceable to the following facts: A vector conservation law holds, that is a conserved integral exists in which the time component of the vector field occurs. The world-lines of the matter in $M(e)$ may be approximated by the time-like geodesics in M with tangent vectors $\mathbf{n}$. That is, $\mathbf{n}$ is a proper vector of the perturbing stress-energy tensor $T^{\mu\nu}$. These geodesics approach tangency to the null-hypersurface $t = t_2$.

It will be shown in the next section, section IIE, that the Einstein field equations imply that any space-time $M(e)$ with the same symmetry properties as M, the vector field normal to the closed three-spaces left invariant by the symmetries of $M(e)$ is a proper vector of any stress-energy that may exist. This result depends on the assumption that the spaces $M(e)$ admit a four parameter group of motions, the three parameter group generated by L_p to which the transformation $q \to qe^{-k\alpha/2}$ is adjoined (cf. section I). If the space-time $M(e)$ is only required to admit the three parameter group of symmetries described by the transformation L_p, this result need not hold. It may then be possible to find perturbations due to the introduction of a perfect fluid with an equation of state against which the space-time M is stable.

Another type of perturbation of M which develops unstably in such a way as to destroy M as a plausible approximation to the physical situation in the neighborhood of the t_1 and t_2 hypersurfaces is the following one: Imagine that, at $t = 0$, M was filled with thermal electromagnetic radiation at some low temperature, or with any other distribution of photons which included both some photons following rays tangent to $\mathbf{k}$, and others following rays tangent to $\mathbf{q}$. The contribution of these two groups of photons to the stress-energy tensor is

$$T^{\mu\nu} = \rho_k k^{\mu}k^{\nu} + \rho_q q^{\mu}q^{\nu}, \qquad (31)$$

and the intensity in each group is governed by the equations [26,27]

$$(\rho_k k^{\mu})_{;\mu} = 0 = (\rho_q q^{\mu})_{;\mu}, \qquad (32)$$

from which follow equations isomorphic to (30). Thus, since $k^t = 1 = q^t$, one has

$$\rho_k(t) = N_k [2l(t^2 + l^2)(4\pi)^2]^{-1} \qquad (33)$$

and similarly for $\rho_q(t)$. Here N_k and N_q are constants proportional to the (conserved) number of photons in each group. It is a consequence of the result proved in section IIE, i.e., $n_{\mu}T^{\mu}_{\nu} = \lambda n_{\nu}$ that $\rho_k = \rho_q = \rho(t)$ so $N_k = N_q = N$. Although $\rho(t)$ is analytic, $T^{\mu\nu}$ is not, because q^{μ} is singular (in M). If we try M′ instead, then $T^{\mu\nu}$ has a singularity from k^{μ}, while q^{μ} remains regular. In either case $T^{\mu\nu}$ becomes large and, through Einstein's equations, will require significant modifications in the metric before t_1 or t_2 can be reached. To see that no other extension of the initial metric (t near 0) can keep $T^{\mu\nu}$ small we note that the scalar

$$T^{\mu\nu}T_{\mu\nu} = 2\rho^2(\mathbf{k}\cdot\mathbf{q})^2 = 8\rho^2 U^{-2}$$

becomes arbitrarily large as $t \to t_2$.

Another way of seeing the physically singular nature of M near t_1 and t_2 requires us to consider only two photons, again tangent respectively to $\mathbf{k}$ and $\mathbf{q}$, starting out from the same point at $t = 0$. Then $\mathbf{k}$ stays at a fixed value of $\psi = \psi_0$, while for the $\mathbf{q}$ photon ψ increases, and every time ψ increases by 4π the two are once again at the same point. The center of momentum energy W for the collisions of these two photons is given by

$$W^2 = -(\mathbf{k}+\mathbf{q})^2 = 4U^{-1}(t),$$

so as $t \to t_2$, the energy goes up without limit.

If one wished to make a mathematical analysis of the stability of M as a solution of $R_{\mu\nu} = 0$, thereby excluding the physical viewpoint we have taken which regards the introduction of various kinds of fields and matter as being among the admissible perturbations in the stability analysis, the previous arguments would apply none-the-less. For as Wheeler [28] has often illustrated, gravitational fields can quite effectively pretend to be matter. Isaacson's [29] analysis of the geometrical optics limit of gravitational radiation, following Brill and Hartle [30], shows that short wavelength perturbations to a gravitational field such as M propagate exactly as photons do, and influence the overall metric (averaging out the short wavelengths) through an effective stress tensor like that for photons, so that Eqs. (31) and (32), as well as the geodesic equations for **k** and q, can equally well be interpreted as applying to "gravitons", i.e. to small perturbations in $g_{\mu\nu}$ whose wavelengths are extremely small compared to l (which is a characteristic radius of curvature of the space-time M).

E. Machian Limitations

It is evident from equation (30) that if a perfect-fluid with an equation of state could be introduced into M with a sufficiently small initial value of σ, and with a velocity with components u^μ such that $u^t(t)$ is finite for all t, then σ would remain small for any required finite time, in particular for a time long enough to encompass the horizons at t_1 and t_2. As was remarked earlier and as will be proved in this section, the Einstein field equations imply that $\mathbf{u} = \mathbf{n}$, for reasons which can perhaps best be interpreted as some aspects of Mach's principle as it is incorporated in general relativity. This principle's objection to a velocity field $\mathbf{u} \neq \mathbf{n}$ as the 4-velocity of the average or dominant matter in the universe is that it implies a non-zero relative velocity between the large scale motions of matter and the average motion of space-time itself, while the Mach idea wants the basic properties of the space-time to result from and be tied to the existence and motion of the bulk of the matter in the universe. Because of the vagueness and flexibility of Mach's principle, the arguments of the previous sentence are nearly pointless unless, as here, they serve primarily to dress up a computation. In the computation the vector **n** normal to the t = const hypersurface (which are defined by the symmetries of M) plays the role of the "4-velocity of space." The differential form on the initial hypersurface

$$s \equiv s_i dx^i = (-n_\mu T^\mu{}_\nu dx^\nu)_{dt=0}, \tag{34}$$

which gives the Poynting vector, or momentum density, seen by an observer with 4-velocity **n**, is then a measure of the relative velocity of the matter and the space. In particular if $T^{\mu\nu}$ corresponds to a perfect fluid with 4-velocity **n**, then s = 0. On any hypersurface $t = t_0$, the Einstein equations involving $n_\mu T^{\mu\nu}$ read [31,32]

$$2\pi^{ij}{}_{|j} = ({}^3g)^{1/2} s^i, \tag{35}$$

where π^{ij} are components of a symmetric tensor density in the hypersurface, simply related to its second fundamental form, ${}^3g$ is the determinant of the induced metric $(ds^2)_{dt=0}$ in this hypersurface, and all index raising, covariant differentiation, etc. is done with this metric. (The definition [32] of π^{ij} in terms of the metric and its first time derivatives will be irrelevant to us.) If ξ^i is any Killing vector in this surface then we deduce from Eq. (35) that

$$\int s_i \xi^i ({}^3g)^{1/2} d^3x = 2\int \xi_i \pi^{ij}_{|j} d^3x = -\int (\xi_{i|j} + \xi_{j|i})\pi^{ij} d^3x = 0. \tag{36}$$

where the last two steps assumed that the t = const hypersurface was closed (no boundary integral when integrating by parts) and used the Killing equation $\xi_{i|j} + \xi_{j|i} = 0$. Thus the first integral in Eq. (36) defines some average momentum on the hypersurface $t = t_0$, which must vanish. Therefore, $s_i = \xi_i$, $\xi_i \neq 0$, in particular, is prohibited since this choice for s_i makes the average positive definite. Now if the Cauchy data are to have the same symmetry as M i.e. the same four Killing vectors (eqs. (43) below), apart from a choice of a preferred direction ($+\partial/\partial\psi$ vs. $-\partial/\partial\psi$), then s^i must point in the distinguished invariant direction, and $s^i(\partial/\partial x^i) = \alpha(\partial/\partial\psi)$, where α may depend on t_0 but must be constant on the hypersurface $t = t_0$. But $\xi = (\partial/\partial\psi)$ is a Killing vector in the initial hypersurface, so Eq. (36) gives $\alpha = 0$ resulting in s = 0. Thus the only symmetry preserving perturbations which may be admitted are those which have zero net momentum for an observer with 4-velocity **n**. As was pointed out earlier it is, for instance, not consistent to set $\rho_q = 0$ in Eq. (31) and obtain a non-singular photon population throughout M. One is required by the Einstein equations to choose $\rho_k = \rho_q$ in Eq. (31), with the consequence that the U(t) = 0 neighborhoods are significantly modified by the perturbation and with the suggestion that an infinite energy-density type singularity will be produced instead of the acausal peculiarities (closed null curves) shown in M.

F. Properties of M

The space-time M can be used as an illustration in connection with the general discussion about the role and prevalence of singularities in relativistic cosmological models. The arguments of Lifshitz and Khalatnikov [15] on this topic were based on local calculations concerning the behavior of the metric components in single coordinate patches, and their conclusion that a metric equivalent to M in the $U(t) > 0$ region could be extended in a non-singular way across t_1 and t_2 has been fully borne out by the metric presented in Section I here. Not only does this metric show no local behavior which could be called a singularity, but there is a globally meaningful sense, described in II A. above, in which it is nonsingular. At the same time, there can be other definitions of what, in a global sense, one means by non-singular, such as geodesic completeness. That definition figures in the theorems of Penrose [33,34] and Hawking [19,20,35] which show that in many circumstances the existence of a singularity in a solution of the Einstein equations is inevitable as a consequence certain properties which can be identified in the Cauchy data specifying the solution. Thus the behavior of M is consistent with all these theorems since M is neither complete, nor does it show the desirable causal properties which Penrose and Hawking sometimes demand.

The peculiar things which can happen near the horizons t_1 and t_2 in M have surprised every physicist we know who has studied them. The behavior of normal congruences of geodesics, from which synchronous coordinate systems are constructed, may be added to the list of peculiarities. The vector field n of Eq. (25) defines a geodesic congruence normal to the hypersurfaces of constant t. Starting from t = 0 one expects these geodesics to intersect each other within a finite time; they do not. None of them can be continued until it intersects the surface $t = t_2$. One could nevertheless define a caustic hypersurface, not as the locus of the intersection, but as the envelope of the geodesic congruences; this is then $t = t_2$. As the time-like geodesics approach this caustic they also approach tangency to it. Nevertheless the caustic is not a time-like hypersurface, it is a null hypersurface.

We should also consider not just the space-time M but also perturbations from it, in connection with the general singularity discussions. Hawking's Theorem 1 in [20] shows, if we take for an initial hypersurface in M any t = const surface where $U(t) > 0$ and where

$$n^{\mu}_{;\mu} = (t^2 + l^2)^{-1}\partial_t[(t^2 + l^2)U^{1/2}] \tag{37}$$

does not vanish [e.g. t = 0 when $m \neq 0$ cf. Eq. (7)], and if we then make <u>any</u> perturbation of the independent Cauchy data on <u>this</u> initial hypersurface, subject only to the restriction that the perturbation be small enough that $n^{\mu}_{;\mu}$ remains non-zero at every point of the initial hypersurface, then the result will again be a space-time which, like M, is not geodesically complete and will contain maximal time-like geodesics which extend for only a finite interval of proper time. This extremely strong result shows that the incompleteness singularity in M is a stable property. The admissible perturbations in applying Hawking's Theorem include not only metric perturbations preserving $R_{\mu\nu} = 0$, but also the introduction of any reasonable type of matter or fields into the space-time. The Machian limitations on admissible perturbations, discussed above, are of course imposed since they follow from the Einstein equations and serve to limit the independent Cauchy data to the correct number of functions.

In apparent striking contrast to the above stability of the incompleteness singularity in M is the lack of stability of true singularities at the limits of synchronous coordinate patches shown by Lifshitz and Khalatnidov [15]. This would be easily understandable if, like M, every small (i.e. initially small) perturbation of M, had distant boundaries and so showed no local singular behavior. For then every synchronous coordinate patch in small perturbations of M would, as in M, terminate in a fictitious singularity. Although we have no proof that the distant boundaries property is unstable, the indications from our own stability studies (Sec. II D above) are that it is, and that all the perturbations we analyzed (which include all perturbations of sufficiently small wavelength to be described in a geometrical optics limit where the behavior of geodesics is the controlling element in the linear analysis [29]) would lead to infinite energy densities and infinite curvature. If this is true, the behavior of the singularities bounding synchronous coordinate patches demands a more subtle explanation. We can think of two possibilities. The first possibility would seem to be required if a typical singularity involved infinite curvature everywhere (i.e. encountered by nearly every time-like geodesic) and would consist in an influence of high curvature on geodesic congruences resulting in almost all synchronous coordinate systems failing before they reached the singularity. A suggestion that this behavior might be possible is contained in the Raychaudhuri equation[9]

$$\frac{d^2L}{d\tau^2} + \frac{1}{3}(2\sigma^2 + R_{\mu\nu}u^{\mu}u^{\nu})L = 0 \tag{38}$$

governing the spatial volume element L^3 in a synchronous coordinate system. [If u^{μ} is the normal congruence of time-like geodesics, then $\dot{L} = \frac{1}{3}\theta L$ where $\theta = u^{\mu}_{;\mu}$, and $2\sigma^2 = u_{(\mu;\nu)}u^{\mu;\nu} - \frac{1}{3}\theta^2$.] Suppose the curvature term here could have the form $(\frac{1}{4} + \alpha^2)\tau^{-2}$; then, neglecting the σ^2 term, the equation would be

$$\ddot{L} + (\tfrac{1}{4} + \alpha^2)\tau^{-2}L = 0, \tag{39}$$

whose solutions are

$$L = \tau^{1/2}\sin[\alpha \ln(\tau/\tau_0)]. \tag{40}$$

Including the σ^2 terms gives even more closely spaced zeros of $L(\tau)$, so this type of infinite curvature would mean that <u>every</u> synchronous coordinate patch terminates in a caustic located a finite distance from the true singularity. However, we know of no example of a metric whose curvature gives the required behavior. The isotropic, radiation filled universe with $ds^2 = -dt^2 + 2t(dx^2 + dy^2 + dz^2)$ does lead (asymptotically as $t \to 0$) to a Raychaudhuri equation of the form of Eq. (39) along most time-like geodesics, but with $\alpha^2 = 0$ or $\alpha^2 = -(1/36)$ which allow the caustic to lie at the true singularity. If there is some general geometrical reason why there cannot exist metrics with curvature singularities giving $\alpha^2 > 0$ in Eq. (39) (or some other behavior resulting in infinitely many zeros of $L(\tau)$ on every time-like geodesic leading to the singularity), then it would be difficult to imagine that singularities involving infinite curvature everywhere, as in the Friedman universes, were typical. We must then turn to the second possibility for understanding how all initially small perturbations of many cosmological models can be time-like incomplete, while almost all synchronous coordinate patches can terminate in a fictitious singularity. This possibility consists in assuming that the singularity is "nearly hidden", i.e., that almost all time-like geodesics avoid the singularity. In this case evidently almost all synchronous coordinate patches will also either avoid the singularity completely or else touch it only at some irregular point (cusp) of their bounding caustic, leaving almost all points on the caustic distinct from the true singularity. Hawking and Carter (private communication) have emphasized the possibility that singularities need not extend throughout space, and Carter has shown [36] that the singularity in the Kerr metric [37,38,39] is nearly hidden in this sense. Carter finds all geodesics in the Kerr metric, and finds that from almost all points no time-like geodesic leads to the singularity. From the exceptional points on the equatorial plane of symmetry almost all geodesics avoid the singularity, but some (with a unique critical direction for their initial velocity) do hit it. Thus the Kerr space-time is not time-like complete, yet it contains no synchronous coordinate patch whose bounding caustic lies on the infinite curvature singularity.

III GEODESICS

The geodesics in M can be obtained by quadratures because of the multitude of constants of motion. In addition to the constant

$$g_{\mu\nu}v^\mu v^\nu = \varepsilon = \pm 1, 0, \tag{41}$$

where

$$v^\mu = dx^\mu / d\lambda = \dot{x}^\mu, \tag{42}$$

there is, for each Killing vector ξ, a constant $\mathbf{v}\cdot\xi$. This one verifies in the well-known way

$$d(v^\mu \xi_\mu)/d\lambda = v^\mu_{;\nu} v^\nu \xi_\mu + \xi_{\mu;\nu} v^\mu v^\nu = 0,$$

using the geodesic equation $v^\mu_{;\nu} v^\nu = 0$ and the Killing equation $\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0$. The Killing vectors of M are known [8,16,6]; in Euler angle coordinates they are

$$\begin{aligned} \xi_x &= -\sin\theta\partial_\theta - \cos\varphi(\mathrm{ctg}\,\theta\partial_\varphi - \mathrm{cosec}\,\theta\partial_\psi), \\ \xi_y &= \cos\varphi\partial_\theta - \sin\varphi(\mathrm{ctg}\,\theta\partial_\varphi - \mathrm{cosec}\,\theta\partial_\psi), \\ \xi_z &= \partial_\varphi, \quad \eta_z = -\partial_\psi. \end{aligned} \tag{43}$$

The corresponding constants of the motion will be designated as

$$p_a = \mathbf{v}\cdot\xi_a, \quad p_\| = -\mathbf{v}\cdot\eta_z, \tag{44}$$

where a = x, y, z. We will also write

$$p^2 = p_a p_a = p_x^2 + p_y^2 + p_z^2,$$

and

$$p_\perp^2 = p^2 - p_\|^2. \tag{45}$$

In their action on θ and φ (i.e. ignoring terms in ∂_ψ) the Killing vectors ξ_a are the standard generators of infinitesimal rotation of the sphere S^2. We can therefore use the invariance of M under these transformations to assume that the pole of the spherical coordinates θ, φ is located for our convenience. If we describe a transformation of the θ, φ coordinates by $n_a \to A_{ab} n_b$ where A_{ab} is a constant 3×3 orthogonal matrix and n_a are homogeneous coordinates defined by

$$(n_x, n_y, n_z) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta), \tag{46}$$

then Eqs. (43) will continue to hold if we replace the ξ_a by an equivalent set of Killing fields according to $\xi_a \to A_{ab}\xi_b$. This changes the constants p_a by the same rotation, $p_a \to A_{ab}p_b$. We will therefore choose A_{ab}, in studying any one given geodesic, in such a way that

$$p_x = 0 = p_y, \tag{47}$$

and

$$p_z = p \geqslant 0.$$

The Killing fields (43) satisfy the relation

$$n_a\xi_a \equiv \sin\theta\cos\varphi\,\xi_x + \sin\theta\sin\varphi\,\xi_y + \cos\theta\,\xi_z = -\eta_z, \tag{48}$$

which can be contracted with the geodesic tangent $\mathbf{v}$ to give

$$n_a p_a = p_\|. \tag{49}$$

For the convenient choice of θ, φ coordinates which gives Eqs. (46), this relationship reads

$$p\cos\theta = p_\| \tag{50}$$

and implies that θ is constant along the geodesic since p and $p_\|$ are. This equation is the analogue of the $\theta = \pi/2$ (motion in a plane through the origin) condition which appears in the more familiar (spherical) type of symmetry under the rotation group.

When the equation $p_\| = -\mathbf{v}\cdot\eta_z = \mathbf{v}\cdot\partial_\psi$ is written out explicitly using Eq. (10) it reads

$$(p_\| / 2l) = 2lU(\dot\psi + \cos\theta\cdot\dot\varphi) - \dot t, \tag{51}$$

where $\dot x^\mu = dx^\mu/d\lambda$. Similarly, Eq. (50) can be written

$$p_z\cos\theta - p_\| = \mathbf{v}\cdot(\xi_z\cos\theta + \eta_z) = \mathbf{v}\cdot(\cos\theta\partial_\varphi - \partial_\psi) = 0 \tag{52}$$

and, after simplifying with Eq. (51), reads

$$\dot\varphi = p(t^2 + l^2)^{-1}. \tag{53}$$

Although this equation cannot be integrated until we know $t(\lambda)$, it already shows that $\dot\varphi = d\varphi/d\lambda$ is bounded, $0 \le \dot\varphi \le p/l^2$, and that $\dot\varphi$ never vanishes when $p \neq 0$.

To determine $t(\lambda)$ we eliminate $\dot\varphi$ and $\dot\psi$ (and $\dot\theta = 0$) from Eq. (41) using Eqs. (53) and (51) (and Eq. (50)) above with the result

$$(p_\| / 2l)^2 = \dot t^2 + U[\varepsilon - p^2(t^2 + l^2)^{-1}]. \tag{54}$$

From this equation $t(\lambda)$ follows by quadrature, but it is most useful to discuss the solution by analogy to the differential equation for a Newtonian particle moving as one dimension in a potential $V(x)$, namely

$$E = \tfrac{1}{2}m\dot x^2 + V(x). \tag{55}$$

In our case $(p_\|/2l)^2$, which plays the role of E, cannot be negative. With this in mind we obtain a detailed qualitative description of $t(\lambda)$ by inspecting the effective potential in Eq. (54), namely

$$V(t) = U(t)[\varepsilon - p_\perp^2(t^2 + l^2)^{-1}]. \tag{56}$$

Several representative graphs of $V(t)$ are shown in Figs 2-4. By studying these graphs in relation to Eq. (54) the qualitative behavior of its solutions can be discussed. Thus this set of graphs is a basic reference tool for any discussion of the geodesics in M.

The first thing to notice about $V(t)$ is that it is bounded. Thus $\dot t$ is also bounded and consequently $|t| \to \infty$ is impossible except when $|\lambda| \to \infty$. This establishes for M the "distant boundaries" property discussed in Section II.

Next we consider time-like geodesics ($\epsilon = -1$) on which $(p_\| /2l)^2 < 1$. For these (and some other) geodesics, $t(\lambda)$ is bounded; it oscillates back and forth across an interval which includes all t where $U(t) > 0$, and has a finite oscillation period Λ so $t(\lambda) = t(\lambda + \Lambda)$. Evidently $\dot t$ changes sign each half cycle. But from Fig. 1 this would mean that $\mathbf{v}$ changed from the forward to the backward light cone each half

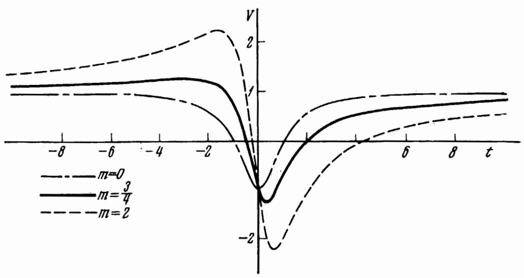


FIG. 2. The effective potential V(t) from Eq.(56) is plotted here for the case $p_\perp = 0$ where $V(t) = \bar{\epsilon}U(t)$. The curves actually shown are appropriate for time-like geodesics ($\epsilon = -1$), but for space-like geodesics the potential V is simply the negative of the quantity given in the Figure. We have chosen units so $l = 1$. The zeros of V here are the values of t_1 and t_2, so for m/l = 0, 3/4, 2 one has correspondingly $(t_1/l, t_2/l)$ = (-1, +1), (-1/2, 2), and (-0.236, 4,236).

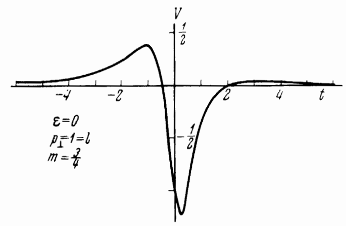


FIG. 3. Any null geodesic with $p_\perp \neq 0$ can, by a scale change in the geodesic parameter, be reduced to one for which $p_\perp/l = 1$. Thus the effective potential V(t) plotted here gives $dt/d\lambda = [(p_{\|}/2l)^2 - V(t)]^{1/2}$ for all these null geodesics in the metric with $m/l = 3/4$. [The other null geodesics with $p_\perp = 0$ have $V(t) = 0$ and $dt/d\lambda$ = const.] The plots are similar for other values of m, varying from a potential with two maxima symmetric about t = 0 when m = 0 to ones with even greater asymmetry than that shown here when m/l is than 3/4.

cycle. As it is impossible for a geodesic to change its time orientation on a time-orientable space-time (such as M), we conclude that these geodesics cannot be continued in M through a cycle, showing that M is not geodesically complete.

The misbehavior which best displays the incompleteness of M occurs in the coordinate ψ. By eliminating $\dot{\varphi}$ from Eq. (51) we find that

$$\dot{\psi} = \frac{1}{2lU}\left(\dot{t} + \frac{p_{\|}}{2l}\right) - \frac{p_{\|}}{t^2 + l^2}. \tag{57}$$

Consider then a time-like geodesic on which $\dot{t}$ and $p_{\|}$ are positive when t = 0. Then as $t \to t_2$ one has $U^{-1}(t) \to \infty$ and $\dot{t} + (p_{\|}/2l) \to p_{\|}/l$ so $\dot{\psi} \to \infty$. In contrast, if at t = 0 one had $\dot{t} > 0$ and $p_{\|} < 0$, then as $t \to t_2$, $\dot{t} \to |p_{\|}/2l| = -p_{\|}/2l$ and the first term in Eq. (57) is indeterminate. But Eq. (54) in the form

$$\left(\dot{t} + \frac{p_{\|}}{2l}\right)\left(\dot{t} - \frac{p_{\|}}{2l}\right) = U(t)[p_\perp^2(t^2 + l^2)^{-1} - \varepsilon] \tag{58}$$

allows us to rewrite Eq. (57) as

$$2l\dot{\psi} = \frac{p_\perp^2(t^2 + l^2)^{-1} - \varepsilon}{\dot{t} - (p_{\|}/2l)} - \frac{p_{\|}}{t^2 + l^2}. \tag{59}$$

In this form it is clear that $\dot{\psi}$ remains finite when $\dot{t}$ and $p_{\|}$ have opposite signs. The preceding considerations show that no geodesic can cross the horizon $t = t_2$ (resp. $t = t_1$) twice, for if $\dot{t}$ had the correct sign to keep $\dot{\psi}$ finite on the first pass, it will have the opposite sign, leading to $\dot{\psi} \to \infty$ on the second approach. See Fig. 5.

Let us study a bad geodesic with $p_{\|} > 0$ somewhat more closely. From Eq. (57) with $\dot{t} \to p_{\|}/2l$ as $t \to t_2$ we see that one has

$$d\psi / dt = \dot{\psi} / \dot{t} \sim 1 / lU \tag{60}$$

as $t \to t_2$. The remarkable thing about this asymptotic form is that all the parameters (p_a, ϵ) identifying the particular geodesic have dropped out, so this equation suggests that the coordinate transformation

$$\psi' = \psi - \int (lU)^{-1} dt$$

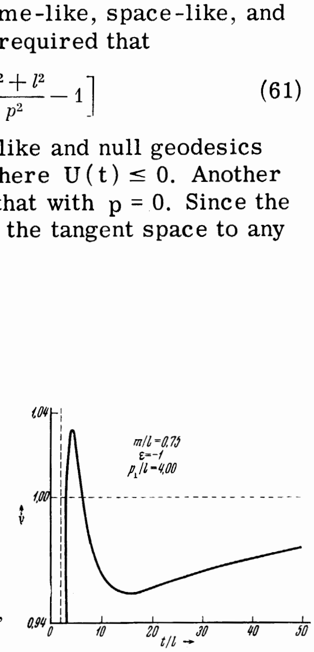
will eliminate the singular behavior in this whole class of geodesics. It was in fact this hint from the geodesic equations which led the authors both to the alternate analytic continuation described in Eqs. (20)—(22), and to the similar transformation which relates the analytic metric form of Eq. (10) to the singular form previously used in [6].

Let us point out some special classes of geodesics. If $p_{\|}$ is sufficiently negative, with $(p_{\|}/2l)^2 > V_{max}$, one has geodesics which extend from $t = -\infty$ to $t = +\infty$, and these exist for any value of $p_\perp^2$, and for space-like, time-like, and null geodesics. At the various positive maxima and minima of V(t) one can choose $(p_{\|}/2l)^2$ to give geodesics with $\dot{t} = 0$ which lie in a single hypersurface t = const. This behavior is stable (with respect to small changes in $p_\perp$, $p_{\|}$) at minima, and unstable at maxima. Figure 3 shows that there are unstable null geodesics of this type, while Fig. 4 illustrates that this behavior can be stable, as well as unstable, for time-like geodesics. In general these geodesics are ergodic on a surface of constant $t = t_c$ and $\theta = \tan^{-1}(p_\perp/p_{\|})$. For appropriate choices of p and θ, however, one may arrange that the linear functions $\psi(\lambda)$ and $\varphi(\lambda)$ have commensurate derivatives and thus obtain closed time-like, space-like, and null geodesics. For this it is required that

$$\frac{\dot{\psi}}{\dot{\varphi}} = \frac{1}{\cos\theta}\left[\varepsilon\frac{t_c^2 + l^2}{p^2} - 1\right] \tag{61}$$

be rational. The closed time-like and null geodesics all occur in the NUT region where $U(t) \leq 0$. Another special class of geodesics is that with p = 0. Since the three Killing vectors ξ_a span the tangent space to any

FIG. 4. For the effective potential of time-like geodesics, a sufficiently large value of p^2 can introduce an additional maximum and minimum in V(t) beyond those shown in the $p_\perp = 0$ curves of Fig. 2. [For m = 0 it is two maxima introduced when $p_\perp^2/l^2 > 2$.] The scale of this figure is chosen to show this additional structure and does not allow the other features similar to Fig. 2 to be shown. In particular the zeros at $t_1/l = 1/2$ and at $t_2/l = 2$, and another maximum at negative $t - t_1$, lie off scale. Since the local minimum shown here has a positive value, $V = V_{min} > 0$, one may choose $(p_{\|}/2l)^2 = V_{min}$ and thus determine a stable time-like geodesic on which t = const.



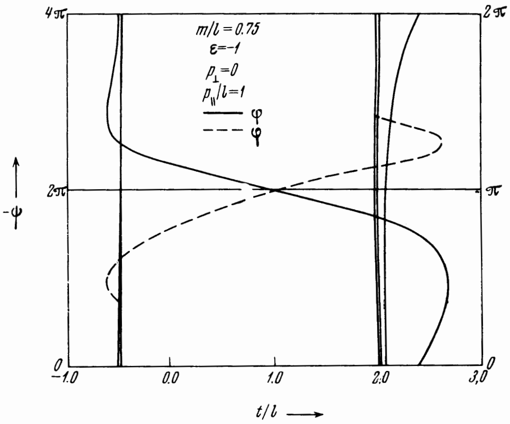


FIG. 5. A time-like geodesic is shown by plotting $\psi(t)$ and $\phi(t)$ in a case where $\theta(t) = 0$. Note how, for the parameters chosen here, ψ is well-behaved near $t_1/l = \frac{1}{2}$ and $t_2/l = 2$ as the geodesic first crosses these limits starting from $t = 0$. After reaching the turning points [determined by $V(t) = (p_\parallel/2)^2 = \frac{1}{4}$] and upon approaching t_1 or t_2 a second time, however, the geodesic shows a rapid (singular) change in ψ preventing the continuation of the geodesic back again to the causal region $t_1 < t < t_2$.

hypersurface $t = t_0$, the equations $p_a \equiv \mathbf{v}\cdot\xi_a = 0$ imply that $\mathbf{v}$ is normal to this hypersurface. This fixes $\mathbf{v}$ uniquely (up to a sign) when t_0 is time-like or space-like and gives, for instance, Eq. (25). The normal to the null hypersurfaces $t = t_1$ or $t = t_2$ is, however, in the unique null direction $\boldsymbol{l}$ lying in that hypersurface, so in these cases $\mathbf{v} = \dot{\psi}\partial_\psi$. In this case Eq. (51) is consistent with any value for $\dot{\psi}$, and the first integrals of the geodesic equation fail to determine the solution uniquely. We must therefore turn to the geodesic differential equation itself, namely

$$(g_{\mu\nu}\dot{x}^\nu)^\cdot - \tfrac{1}{2}\dot{x}^\alpha\dot{x}^\beta(\partial g_{\alpha\beta}/\partial x^\mu) = 0. \tag{62}$$

These equations are easy to examine since we know that t, θ and φ are constant along the geodesic of interest. We thus find, at $t = t_2$,

$$\ddot{\psi} + lU'(t_2)\dot{\psi}^2 = 0. \tag{63}$$

with the solution on the future null-cone given by

$$\psi = -\beta\ln(-\lambda), \quad -\infty < \lambda < 0,$$
$$\beta = -[lU'(t_2)]^{-1} = (2l)^{-1}(t_2{}^2 + l^2)(m^2 + l^2)^{-1/2}. \tag{64}$$

On this geodesic, after ψ increases by 4π the geodesic has returned to its starting point, but its tangent vector has increased in length by a factor $e^{4\pi/\beta}$. After many revolutions $\dot{\psi} \to +\infty$ as $\lambda \to 0^-$, and the geodesic cannot be continued to positive values of λ.

Up to this point we have always used a coordinate system adapted to the geodesic of interest so that we could assume θ = const along the geodesic. However most of the results can be freed from this limitation by expressing them in group invariant form. The result for the tangent to a geodesic family invariant under the translations generated by the ξ_a is

$$v = \dot{t}\partial_t - \frac{1}{2lU}\left(\dot{t} + \frac{p_\parallel}{2l}\right)\eta_z + \frac{p}{t^2+l^2}[\eta_x\cos(\alpha+f) - \eta_y\sin(\alpha+f)]. \tag{65}$$

Here $\dot{t}$ means the positive solution of Eq. (54), α is a constant, and $f(t)$ is the solution of

$$\frac{df}{dt} = \frac{1}{2lU} + \frac{p_\parallel}{\dot{t}}\left[\frac{1}{4l^2} - \frac{1}{t^2+l^2}\right], \tag{66}$$

satisfying $f(0(0) = 0$. The vector fields η_a (a = x, y, z) are defined in Eqs. (B6) of ref. [6] and, taken together with ∂_t, form the dual basis to the basis of differential forms given in Eq. (2). The geodesics congruences we used in Section II all correspond to special cases of Eq. (65) in which $p_\perp = 0$ so the indefinite integral in Eq. (66) need not be evaluated.

ACKNOWLEDGEMENTS

A large part of the work reported here was done in 1962–63 when one of us (C.W.M.) was at Palmer Physical Laboratory of Princeton University and the other (A.H.T.) was on leave from the University of Illinois and in residence at the Institute for Advanced Study in Princeton, New Jersey supported in part by a National Science Foundation grant. The computations on which the figures are based were done at the University of Maryland's Computer Science Center in 1963 with support from NASA. More recently this research has had support from the U. S. National Aeronautics and Space Administration (grant NS6-436 to the University of Maryland), the U. K. Science Research Council, a Senior Postdoctoral Fellowship award to C.W.M. from the U. S. National Science Foundation; A.H.T. was supported in part by the United States Atomic Energy Commission under contract number AT(11-1)-34, Project Agreement No. 125 while on sabbatical leave from the University of California, Berkeley, and in residence at the College de France, Paris. The authors are also grateful to Professor E. M. Lifshitz for arranging to have the article translated into Russian, and to the anonymous translator.

[1] R. L. Bishop and R. J. Crittenden, "Geometry of Manifolds" (Academic Press, New York, 1964).

[2] L. Auslander and R. E. MacKenzie, "Introduction to Differentiable Manifolds" (McGraw-Hill Book Co., New York, 1963).

[3] H. Flanders, "Differential Forms..." (Academic Press, New York, 1963).

[4] T. J. Willmore, "An Introduction to Differential Geometry" (Oxford University Press, London, 1959).

[5] "Lectures in Applied Mathematics", vol. 8, J. Ehlers, ed. (Proceedings of the Am. Math. Society's Summer Seminar on Relativity and Astrophysics at Cornell, 1965), pub. by American Mathematical Society, Providence, R. I., 1967.

[6] C. W. Misner, J. Math. Phys. **4** (1963), 924–937.

[7] H. C. Corben and P. Stehle, "Classical Mechanics", 2nd ed. (John Wiley and Sons, Inc., New York, 1960). Appendix IV.

[8] A. H. Taub, Ann. Math. **53** (1951), 472.

[9] O. Heckmann and E. Schücking, Chapter 11 of "Gravitation: an introduction to current research, L. Witten, ed. (John Wiley and Sons, Inc., New York, 1962).

[10] C. Behr, Zeit. fur Astrophysik **54** (1962), 268–286.

[11] D. R. Brill, Phys. Rev. **133** (1964), B845–B848.

[12] D. R. Brill, Supplemento al Nuovo Cimento (1) **2** (1964), 3–56.

[13] J. A. Wheeler, Rev. Mod. Phys. **34** (1962) 874–892, especially p. 884.
[14] V. Joseph, Proc. Cambridge Phil. Soc. **53** (1957), 836–842.
[15] E. M. Lifshitz and I. M. Khalatnikov, Advances in Physics (Philosophical Magazine Supplement) **12** (1963), 185–249.
[16] E. Newman, L. Tamburino, and T. Unit, J. Math. Phys. **4** (1963), 915–923.
[17] D. R. Brill and J. M. Cohen, J. Math. Phys. **7** (1966), 238–243.
[18] M. Kruskal, Phys. Rev. **119** (1960), 1743.
[19] S. W. Hawking, Proc. Roy. Soc. **A294** (1966), 511–521.
[20] S. W. Hawking, Proc. Roy. Soc. (1967), 187–201.
[21] B. Carter, ''Horizons, Orthogonal Transitivity. . . '' (preprint, D.A.M.T.P. Cambridge, 1966).
[22] K. Stellmacher, Math. Ann. **115** (1938), 136–152.
[23] A. Raychaudhuri, Phys. Rev. **98** (1955), 1123.
[24] A. Komar, Phys. Rev. **104** (1956), 544.
[25] C. Behr. Zeit. fur Astrophysik **60** (1965), 286–287.
[26] J. Kristian and R. K. Sachs, Ap. J. **143** (1966), 379–399.
[27] D. M. Zipoy, Phys. Rev. **142** (1966), 835–838.
[28] J. A. Wheeler, ''Geometrodynamics'' (Academic Press, London, 1962).
[29] R. A. Isaacson, Ph. D. Dissertation (Physics Department, University of Maryland, 1967) and Phys. Rev. **166** (1968) 1263–1271, 1272–1280.
[30] D. R. Brill and J. B. Hartle, Phys. Rev. **135** (1964), B271–B278.
[31] Y. Bruhat, Chapter 4 in ''Gravitation: an Introduction to Current Research'', L. Witten, ed. (John Wiley and Sons, Inc., New York, 1962).
[32] R. Arnowitt, S. Deser, and C. W. Misner, Annals of Physics **11** (1960), 116–121.
[33] R. Penrose, Phys. Rev. Letters **14** (1965), 57.
[34] R. Penrose, Adams Prize Competition Essay, December 1966.
[35] S. W. Hawking, Proc. Roy. Soc. **A295** (1966), 490–493.
[36] B. Carter, ''Global Properties of the Kerr Class of Solutions''. (preprint, D:A.M.T.P. Cambridge, 1966).
[37] R. P. Kerr, Phys. Rev. Letters **11** (1963), 238.
[38] B. Carter, Phys. Rev. **141** (1966), 1242–1247.
[39] R. H. Boyer and R. W. Lindquist, J. Math. Phys. 8, (1967), 265–281.