ANNALS OF PHYSICS **108**, 1–36 (1977)

Singularities and Causality Violation*,†

FRANK J. TIPLER

Department of Mathematics, University of California, Berkeley, California 94720

Received November 30, 1976

A number of important theorems in General Relativity have required a causality assumption; for example, the Geroch topology change theorem, and most of the Hawking–Penrose–Geroch singularity theorems. It is shown in this paper that the causality condition can be replaced by weaker causality conditions, and in some cases removed altogether. In particular, (a) it is shown that if the Einstein equations (and the weak energy condition) hold on the "topology-changing" space-time considered by Geroch, then topology change *cannot* occur. No causality assumption is needed in the proof. Furthermore, it is shown that if topology change occurs within a finite region, then this change of topology must be accompanied by singularities. (b) It is shown that causality violation cannot prevent the Hawking–Penrose–Geroch singularities unless the causality violation begins "at infinity"—a region which is free of matter and gravitational radiation—and this seems very unlikely.

1. INTRODUCTION

It has long been known that many solutions to the Einstein Equations possess causal anomalies in the form of closed timelike lines (CTL). In fact, all known asymptotically flat vacuum solutions with nonzero angular momentum contain such anomalies [1, 2], and there are indications that a very rapidly rotating star would have them also [3]. These considerations do not *guarantee* that our universe actually contains such a causality violating region, however, and the question of whether or not such a region exists in the real universe is clearly an important question to answer. Furthermore, we would like to know if it would be possible to *manufacture* such a region, say by speeding up the rotation of a star.

I shall provide a partial answer to the second question in this paper; I shall show that, in general, any attempt to evolve CTL from regular initial data will cause singularities to form in space-time. Thus if by the word "manufacture" we mean "construct using only ordinary materials *everywhere*," then the theorems of this paper will conclusively demonstrate that a CTL-containing region cannot be manufactured. For a singularity is a region where the matter density becomes infinite [4], and matter with arbitrarily large density clearly cannot be considered "ordinary material." It does not, of course, follow from this result that causality violating regions cannot exist, for the entire universe could be a causality violating set. In this case the notion

* Part of a Ph.D. thesis submitted to the Physics Faculty of the University of Maryland.

† Work supported in part by the National Science Foundation under grant No. MCS-76-21525.

ISSN 0003-4916

of "regular initial data" could not even be defined. Nor does it follow that it is impossible to create CTL. After all, singularities are expected to occur in black hole explosions in any case [5]. Thus the first question, the question concerning the existence of causality violating regions, is still open.

The proofs contained in this paper will be based on the "Global Techniques" developed by Hawking, Penrose, and Geroch. In order to make this paper largely self-contained, the basic lemmas from Global Techniques will be given in Section 2. The proofs of these lemmas can be found in such works as Geroch [6], Penrose [7, 8], and Hawking and Ellis [9]. I shall also include in this section some necessary lemmas from the theory of ordinary differential equations.

The next five sections will discuss the relationship between CTL and singularities from several points of view. Section 3 will begin with a summary of a theorem (published elesewhere [10]) which shows that a time machine cannot be manufactured in an asymptotically flat space-time without the formation of singularities. This section will conclude with a theorem showing that any causality violation which begins in a small region and expands must be accompanied by the formation of singularities.

The attempt to define which is meant by the phrase "small region" leads to a consideration of the relationship between CTL and the topological notion of compactness, and this relationship is discussed in Sections 4, 5, 6. Section 4 is devoted to the statement and proof of a very general theorem which states that causality violation which begins in a finite region with compact boundary must result in the formation of singularities. Section 5 and 6 will apply this theorem (and certain modifications of it) to various problems in General Relativity. In particular, Section 5 will apply this theorem to the problem of topology change. It will be shown that topology change *cannot* occur in a compact region if the weak energy condition holds. Also included in this section is a proof that if topology change occurs in a finite (but not compact) region, the change must be accompanied with singularities. Previous topology change theorems [11] have assumed the nonexistence of CTL; no such assumption is necessary in the above-mentioned theorems. Section 6 will be devoted to compact space-times. It will be shown that causality is violated at *every* point in a generic compact space-time in which the weak energy condition holds.

These results suggest that it might be possible to prove a conjecture made by Geroch [12] and Hawking and Ellis [9, p. 272]: Causality violation cannot prevent the formation of the singularities predicted by the Hawking–Penrose–Geroch singularity theorems. (It is important to *prove* that causality violation cannot prevent these singularities; one should not merely dismiss the possibility with word play as in "their [CTL] existence in itself would be singularity" [13], or as in "...collapse presumably produces singularities—or a violation of causality, which is also a rather singular occurrence!" [14, p. 935]. This conjecture is dealt with in Section 7, and to some extent it is proven. It is shown that causality violation can prevent singularities only if the causality violation begins at infinity.

Finally, Section 8 concludes the paper with an attempt to answer the question: Are CTL possible?

The notation, terminology, and conventions used in this paper are the same as those of Hawking and Ellis [HE] [9], unless otherwise noted. (However, it will be assumed that $\Lambda = 0$; the Einstein equations are $R_{ab} - \frac{1}{2}g_{ab}R = 8\pi T_{ab}$.)

2. PRELIMINARIES

The object of study in this paper is a *space-time*, which is a pair (M, g) where M is a real, four-dimensional, connected C^∞ Hausdorff manifold and g is a C^∞ Lorentz metric. (M, g) is C^∞ inextendible, orientable, and space and time orientable.

The Global Techniques used to analyze the structure of space-time are naturally divided into two parts, the first being the study of a space-time's causal structure, and the second being the analysis of conjugate points along a geodesic. The two basic relationships of the first are given in the

DEFINITION. A point p will be said to *chronologically precede* q (written $p \ll q$) if and only if there is a smooth (C^∞) future-directed nondegenerate timelike curve from p to q. (A degenerate curve is a curve which consists of only one point. Thus $p \ll p$ will mean that there is a CTL passing through p.) Similarly, we will say that p *causally* precedes q (written $p \prec q$) if and only if there is a future-directed causal (timelike or null) curve from p to q. We will allow the possibility that the causal curve is degenerate, so we have $p \prec p$ for all p in M. A closed causal curve is signified by the existence of a pair of distinct points p, q such that $p \prec q$ and $q \prec p$. The relations $\ll$ and $\prec$ have the following properties.

LEMMA 1.

$$
\begin{aligned}
&a \ll b \text{ implies } a \prec b;\\
&a \ll b,\ b \ll c \text{ implies } a \ll c;\\
&a \prec b \prec c \text{ implies } a \prec c;\\
&a \prec b \ll c \text{ implies } a \ll c;\\
&a \ll b \prec c \text{ implies } a \ll c.
\end{aligned}
$$

We can use the relations $\ll$ and $\prec$ to construct sets.

DEFINITION. The set $I^+(p) \equiv \{q \in M \mid p \ll q\}$ is called the *chronological future* of p; $I^-(p) \equiv \{q \in M \mid q \ll p\}$ is the *chronological past* of q. We also have $J^+(p) \equiv \{q \in M \mid p \prec q\}$, the *causal future* of p and $J^-(p) \equiv \{q \in M \mid q \prec p\}$ the *causal past* of p. The chronological or causal future of a set S is defined by $I^+(S) \equiv \bigcup_{p \in S} I^+(p)$, $J^+(S) \equiv \bigcup_{p \in S} J^+(p)$, respectively, and similarly for the pasts of S. $I^+(S)$ is open but $J^+(S)$ is not necessarily closed. The boundary of $I^+(S)$ will be denoted $\dot{I}^+(S)$.

DEFINITION. A set S will be called *achronal* if there are no points $p, q \in S$ such that $p \ll q$. A set S is said to be *acausal* if there are no points $p, q \in S$ for which $p \prec q$. The *edge* of an achronal closed set S (denoted edge (S)) is the set of points p in S such that if $r \ll p \ll q$, with γ a timelike curve from r to q, $p \in \gamma$, then every neighborhood of γ contains a timelike curve from r to q not meeting S.

LEMMA 2. *Every point p in the achronal set $\dot{I}^+(q) - q$ is the future endpoint of a null geodesic on $\dot{I}^+(q)$ which can be extended into the past on $\dot{I}^+(q)$ either indefinitely, or until it meets q.*

Another way to state this lemma is as follows. $\dot{I}^+(q)$ is generated by null geodesics which either have no past endpoints or have past endpoints at q. Generators which have no past endpoints never intersect as they are extended into the past.

As noted earlier, the existence of a CTL through a point p is denoted by $p \ll p$. We shall say that a space-time (M, g) satisfies the *chronology condition* if there are no points $p \in M$ such that $p \ll p$. Similarly, a space-time will be said to satisfy the *causality condition* if there are no closed causal curves. The set of points at which the chronology condition does not hold is the disjoint union of sets of the form $I^+(q) \cap I^-(q) \neq \varnothing$, $q \in M$. The set of points at which the causality condition does not hold is the disjoint union of sets of the form $J^+(q) \cap J^-(q) \neq \{q\}$, $q \in M$. Another causality requirement that is often imposed on a space-time is given by the following.

DEFINITION. The *strong causality condition* is said to hold at p if every neighborhood of p contains a neighborhood of p which no non-spacelike curve intersects more than once. Thus if a space-time satisfies the strong causality condition at every point, then not only are there no closed causal curves, but there are also no "almost" closed curves; that is, no causal curve from p can ever return to the vicinity of p. It is not possible for a causal curve to be "imprisoned" inside a compact set if strong causality holds. The notion of "imprisonment" is made precise by the following

DEFINITION. A causal curve γ that is future-inextendible is said to be *totally future imprisoned in S* if it enters and remains within a compact set S. A future-inextendible causal curve which does not remain within any compact set but continually reenters a compact set S is said to be *partially future imprisoned in S*.

We want to prove that causality violation which evolves from regular initial data must be accompanied by singularities; to do this we must put some restrictions on what is meant by "regular initial data." The set upon which it is most natural to give initial data is a *partial Cauchy surface* which is defined to be a closed acausal set S with no edge (i.e., edge $(S) = \varnothing$). A partial Cauchy surface is thus a spacelike hypersurface which no causal curve intersects more than once. We shall take the existence of such a set as a necessary condition for the existence of "regular initial data." For any partial Cauchy surface S, there are sets within which events are completely determined by data on S. These sets are given by

DEFINITION. The *future domain of dependence* or *Cauchy development* of S is defined by

$$\tilde{D}^+(S) \equiv \{p \mid p \in M \text{ and every past inextendible timelike curve containing } p \text{ intersects } S\}.$$

Similarly for the past history of S, we define

$$\tilde{D}^-(S) \equiv \{p \mid p \in M \text{ and every future inextendible timelike curve containing } p \text{ intersects } S\}$$

as the *past domain of dependence* of S, and we call $\tilde{D}(S) \equiv \tilde{D}^+(S) \cup \tilde{D}^-(S)$ the *total domain of dependence* of S. The future boundary of $\tilde{D}^+(S)$, that is, the limit of the region that can be predicted from knowledge of data on S, is defined by

$$\begin{aligned} H^+(S) &= \{p \mid p \in \tilde{D}^+(S), I^+(p) \cap \tilde{D}^-(S) = \emptyset\} \\ &= \tilde{D}^+(S) - I^-(\tilde{D}^+(S)). \end{aligned}$$

$H^+(S)$ is called the *future Cauchy horizon*; $H^-(S)$, the *past Cauchy horizon*, and $H(S)$, the *total Cauchy horizon* are correspondingly defined. $H^+(S)$ and $\tilde{D}^+(S)$ have the properties (for S closed and achronal)

(a) $\tilde{D}^+(S)$ is closed,

(b) $H^+(S)$ is achronal and closed,

(c) $\partial\tilde{D}^+(S) = H^+(S) \cup S$,

(d) edge $S =$ edge $H^+(S)$.

(The above statements are still true if $(+)$ is replaced by $(-)$.)

If $D(S) = M$, then the entire future and past can be predicted from data on S; in this case we call S a *Cauchy surface*. If (M, g) contains a Cauchy surface S, then $H(S) = \emptyset$. Some other important properties of Cauchy surfaces and partial Cauchy surfaces are given in the following five lemmas.

LEMMA 3. *If S is a partial Cauchy surface, then $H^+(S)$, if nonempty, is generated by null geodesic segments which have no past endpoints; these segments never intersect as they are extended into the past.*

LEMMA 4. *The strong causality condition holds on* int $\tilde{D}^+(S)$, *for S achronal and closed.*

LEMMA 5. *If S is closed and achronal, then for each set of points $q \ll p$ with $p \in \text{int } \tilde{D}^+(S)$ and $q \in S$, there is a timelike geodesic from q to p which attains the maximum length for timelike curves connecting q and p. Furthermore, to each*

$p \in \operatorname{int} \tilde{D}^+(S)$ there is a timelike geodesic orthogonal to S which attains the maximum length for timelike curves connecting p and S. (The term "length" of a causal curve denotes its proper time integral.)

LEMMA 6. *Let S be a Cauchy surface for the space-time (M, g). Then M is homeomorphic to $R^1 \times S$, and for each $a \in R^1$, $\{a\} \times S$ is a Cauchy surface for (M, g). Thus topology change is impossible if the space-time contains a Cauchy surface.*

LEMMA 7. *If $p \in \operatorname{int} \tilde{D}^+(S)$, then $J^+(S) \cap J^-(p)$ is compact, provided S is a partial Cauchy surface.*

The second part of Global Techniques is the study of conjugate points and the relationship of these points to the causal structure. Roughly speaking, two points p and q on a causal geodesic $\gamma(t)$ are said to be *conjugate along* $\gamma(t)$ if a geodesic which is infinitesimally close to $\gamma(t)$ intersects $\gamma(t)$(at p and q. More precisely, p and q will be conjugate along $\gamma(t)$ if the *expansion* θ of a geodesic congruence containing $\gamma(t)$ becomes infinite at p and q [HE, pp. 96–97; 100–101]. The expansion θ satisfies the equation

$$d\theta/dt = -R_{ab}K^aK^b - 2\sigma^2 - (1/n)\,\theta^2 \tag{2.1}$$

where K^a is the tangent vector to the geodesic, t is an affine parameter along $\gamma(t)$, and $n = 3$ for timelike geodesics, and $n = 2$ for null geodesics. The function σ^2 is called the *shear* of the congruence and is positive definite. For null geodesics it satisfies the equation

$$d\sigma_{mn}/dt = -C_{manb}K^aK^b - \theta\sigma_{mn} \tag{2.2}$$

where $2\sigma^2 \equiv \sigma_{mn}\sigma^{mn}$ and $m, n = 1, 2$ label the two spacelike directions of a pseudo-orthonormal frame parallel propagated along $\gamma(t)$ [HE, p. 86]. It can be shown [HE, pp. 97, 100] that p and q are conjugate along $\gamma(t)$ if and only if a function y, defined by $\theta \equiv (1/y)\,dy/dt$, satisfies $y = 0$ at q and p. If we define a new function z by the relation $z^n = y$, then $\theta = (n/z)\,dz/dt$ and (2.1) becomes

$$(d^2z/dt^2) + H(t)\,z = 0 \tag{2.3}$$

where

$$H(t) \equiv (1/n)(R_{ab}K^aK^b + 2\sigma^2). \tag{2.4}$$

Since $z^n = y$, y will be zero at p and q if and only if $z = 0$ and p at q. Thus we have reduced the problem of finding conjugate points to the problem of discovering the location of zeros in solutions to Eq. (2.3). This, fortunately, is a well-known problem in the theory of ordinary differential equations. This problem has an extensive literature from which we will take the following two lemmas.

LEMMA 8. *Let $H(t)$ be continuous and nonnegative in the interval $(a, +\infty)$. If*

$$\liminf_{t\to+\infty} t^2 H(t) > \tfrac{1}{4}. \tag{2.5}$$

Then every solution to (2.3) *has infinitely many zeros on* $(a, +\infty)$.

The proof can be found in Ref. [15].

LEMMA 9. *Let $H(t)$ be continuous on $(a, +\infty)$. If*

$$\int_a^\infty H(t)\, dt = +\infty$$

then all solutions to (2.3) *have infinitely many zeros on* $(a, +\infty)$.

The proof can be found in Ref. [16].

We shall also need the following proposition, a result which is *not* in the literature.

PROPOSITION 1. *A sufficient condition for the existence of a solution to* (2.3) *with at least two zeros in the interval $[t_0, \infty)$ is that there exist numbers t_1, t_2 with $t_0 < t_1 < t_2$ such that*

$$\frac{1}{t_1 - t_0} < \int_{t_1}^{t_2} H(t)\, dt \tag{2.6}$$

assuming that $H(t)$ is continuous and $H(t) \geqslant 0$ in $[t_0, \infty)$.

Proof. We can always find a solution $z(t)$ which has a zero at t_0. We will show that this solution has another zero in (t_0, ∞); for assume it does not. Then without loss of generality we can assume $z(t) > 0$ and $dz/dt \geqslant 0$ in (t_0, ∞), since if $dz/dt < 0$ at any point in (t_0, ∞), we would have a zero in (t_0, ∞) by the condition $H(t) \geqslant 0$. From (2.3) we obtain

$$\left.\frac{dz}{dt}\right|_{t=t_2} = \left.\frac{dz}{dt}\right|_{t=t_1} - \int_{t_1}^{t_2} z(t)\, H(t)\, dt.$$

Since $dz/dt \geqslant 0$, we have $z(t) \geqslant z(t_1)$ for any $t > t_1$. Since $H(t) \geqslant 0$, we have

$$\frac{z(t_1)}{t_1 - t_0} \geqslant \left.\frac{dz}{dt}\right|_{t=t_1} \qquad \text{or} \qquad z(t_1) \geqslant (t_1 - t_0) \left.\frac{dz}{dt}\right|_{t=t_1}.$$

Thus

$$\begin{aligned} \left.\frac{dz}{dt}\right|_{t=t_1} \int_{t_1}^{t_2} z(t)\, H(t)\, dt &\leqslant \left.\frac{dz}{dt}\right|_{t=t_1} - \int_{t_1}^{t_2} z(t_1)\, H(t)\, dt \\ &\leqslant \left.\frac{dz}{dt}\right|_{t=t_1} - (t_1 - t_0) \left.\frac{dz}{dt}\right|_{t=t_1} \int_{t_1}^{t_2} H(t)\, dt \\ &= \left.\frac{dz}{dt}\right|_{t=t_1} \left[1 - (t_1 - t_0) \int_{t_1}^{t_2} H(t)\, dt\right]. \end{aligned}$$

By hypothesis, the factor in brackets is negative. If $dz/dt\,|_{t=t_1} > 0$, then $dz/dt\,|_{t=t_2} < 0$, implying a zero of $z(t)$ in (t_2, ∞). If $dz/dt\,|_{t=t_1} = 0$, then $dz/dt\,|_{t=t_2} < 0$ since $\int_{t_1}^{t_2} z(t)\,H(t)\,dt > 0$ by assumption. In either case, $z(t)$ must have a zero in (t_0, ∞). This contradicts the assumption that $z(t)$ has no zeros in (t_0, ∞). Thus $z(t)$ must have at least one zero in (t_0, ∞), and hence two zeros in $[t_0, \infty)$. ∎

Proposition 1 concludes our digression into the theory of ordinary differential equations. Now we will connect this mathematical interlude with physics. Recall that most of the theorems required the assumption $H(t) \geqslant 0$. From Eq. (2.4) we see that for causal geodesics this corresponds to $R_{ab}K^aK^b + 2\sigma^2 \geqslant 0$, or since σ^2 is intrinsically nonnegative, to $R_{ab}K^aK^b \geqslant 0$. The Einstein equations tell us that when K^a is null, $R_{ab} = 8\pi T_{ab}K^aK^b$, so the above inequality will hold whenever $T_{ab}K^aK^b \geqslant 0$. This is assured by the *Weak Energy Condition*, which says that the energy-momentum tensor at each point p in M obeys the inequality $T_{ab}W^aW^b \geqslant 0$ for any timelike vector W^a in T_p, where T_p is the space of all tangent vectors to M at p. By continuity we will have $T_{ab}K^aK^b \geqslant 0$ for any null vector K^a in T_p. Physically, this condition says that the energy density as measured by any observer is nonnegative, and this holds for all known forms of matter (see [HE, pages 89–91] for a more detailed discussion of this point; see, however, Epstein *et al.* [17]). But the weak energy condition insures $R_{ab}K^aK^b \geqslant 0$ only for *null* geodesics. For timelike geodesics, we need a different condition to insure this.

DEFINITION. We shall say that the energy-momentum tensor satisfies the *Strong Energy Condition* at all $p \in M$ if for every timelike $W^a \in T_p$, we have

$$T_{ab}W^aW^b \geqslant \tfrac{1}{2}W^aW_aT.$$

Combining the Einstein equations and the strong energy condition, we get $R_{ab}W^aW^b \geqslant 0$ for all *causal* vectors W^a. Physically, the strong energy condition says that gravitation is always attractive, and this is true for all known forms of matter (see [HE, p. 95]).

The final locally defined energy condition we will need is the following

DEFINITION. We will say that the *Ubiquitous Energy Condition* holds on a set S if the energy-momentum tensor at each point $p \in S$ satisfies $T_{ab}K^aK^b > 0$ for all non-spacelike vectors $K^a \in T_p$. (If S is not specified, $S = M$ will be implied.) Physically, the ubiquitous energy condition says that the energy density is nonzero for every observer at every point of S. Furthermore, for all observed forms of matter, the condition $T_{ab}U^aU^b > 0$ for all timelike vectors U^a implies $T_{ab}K^aK^b > 0$ for all null vectors K^a also; the ubiquitous energy condition assumes this to be a property of all physically reasonable T_{ab}. Thus matter consisting entirely of radiation moving in *one* special direction—Type II matter in the notation of Hawking and Ellis [9, p. 89]—is ruled out by the ubiquitous energy condition even though $T_{ab} \neq 0$. However, such an energy-momentum tensor is extremely unlikely.

The ubiquitous energy condition was apparently originally proposed by Aristotle

(Nature abhors a vacuum), and later defended by numerous authors, among them G. W. Leibniz, who supported it with an argument which is cogent even in the world view of General Relativity: At any point in space-time we expect there will be a little randomly oriented radiation present, even in what would otherwise be a perfect vacuum. The microwave background radiation, for example, is expected to be present everywhere in space-time, except perhaps where there is matter to shield it out. This random background radiation would be sufficient to satisfy the ubiquitous energy condition. Even in radiation shielded regions there would be quantum mechanical zero-point radiation (Sakurai [18, p. 33]; see however, Epstein *et al.* [17]) which would in itself be sufficient to satisfy the condition. Thus the ubiquitous energy condition seems to be an eminently reasonable condition to impose on the whole of space-time, though for our purposes we will need to impose it only on certain compact sets. The ubiquitous energy condition has been imposed (in effect) on a compact set by Hawking [19] in one of his early singularity theorems.

DEFINITION. A space-time will be said to satisfy the *generic condition* if every timelike or null geodesic contains at least one point p at which

$$K^a K^b K_{[c} R_{d]ab[e} K_{f]} \neq 0 \tag{2.7}$$

where K^a is the tangent vector to the geodesic at p.

If (2.7) is satisfied at some point p, then the tidal force $R_{\alpha 4 \beta 4}$ (or R_{m4n4}) is nonzero at p along a timelike (or null) geodesic [HE, p. 101]. Thus physically this condition says that every geodesic feels tidal force at one point in its history at the very least. To see that this condition is reasonable, see [HE, p. 101]. In the theorems of this paper the generic condition and the ubiquitous energy condition will be used for the same purpose: to focus geodesic congruences.

The relationship between the causal structure and conjugate points lies in the following two lemmas.

LEMMA 10. *A timelike geodesic curve γ from q to p is of maximal proper time length if and only if there is no point conjugate to q along γ in the interval (q, p).*

LEMMA 11. *If there is a point r in (q, p) conjugate to q along a causal geodesic γ, then there is a timelike curve from q to p.*

3. FINITELY VICIOUS SPACE-TIMES

One of the major purposes of this paper is to show that it is not possible to manufacture a time machine—which we will define to be a region of high curvature generating a chronology violating set V such that V intersects the earth's world line—without the formation of singularities. One way of showing this is to note that a manufactured time machine would have to embody the following features. The time machine would have to be constructed in some localized region, for it is beyond our power to manipulate all the matter in the universe. Thus we would expect the gravitational field generated by the time machine to decrease in strength as we move away from it, eventually becoming negligible at large distances. That is, the time machine is formed in asymptotically flat space. (In the first approximation, complications introduced by cosmology are ignored.) The time machine would be built from normal matter (matter satisfying the weak energy condition) in a universe which is initially free of CTL; that is, the time machine evolves from regular initial data in the asymptotically flat space-time. Thus we will require the existence of a partial Cauchy surface S in the space-time so that we can define "initial data." We will further require (M, g) to be "partially asymptotically predictable" from S, which essentially means that $\tilde{D}^+(S)$ comes to an end because of the formation of CTL or singularities and *not* because of the choice of S [20]. Otherwise, we could not say that the CTL *evolved* from S. Finally, in order to use the time machine, it must be possible for an observer initially far away from the time machine, on earth, say, to travel to the time machine, go backwards in time, and return to earth before he left. In order for this to happen, the time machine must not be shielded from outside observers by an event horizon; in symbols, $J^-(\mathscr{I}^+) \cap V$ is nonempty, where V is the chronology-violating set. (Note that this condition does not preclude the existence of an event horizon. It merely says that *some* CTL are to be found outside of an event horizon.)

The following theorem shows that such a time machine cannot be constructed without the formation of singularities; CTL cannot arise from regular initial data in any asymptotically flat, geodesically complete space-time.

THEOREM 1. *An asymptotically flat space-time* (M, g) *cannot be null geodesically complete if*

(1) $R_{ab}K^aK^b \geqslant 0$ *for all null vectors* K^a;

(2) *the generic condition holds on* (M, g);

(3) (M, g) *is partially asymptotically predictable from a partial Cauchy surface* S;

(4) *the chronology condition is violated in*

$$J^+(S) \cap J^-(\mathscr{I}^+).$$

(Note that condition (1) follows from the Einstein equations and the weak energy condition.)

The *idea* behind the proof is quite simple; we first show that under the above conditions there exists a null geodesic which never leaves $H^+(S)$. This geodesic cannot be complete, for it can be shown that (1) and (2) imply that all complete null geodesics have a pair of conjugate points. This would be impossible, by Lemma 11 and the achronality of $H^+(S)$. The rigorous proof has been published elsewhere [10].

The theorem above has two major weaknesses from the physical point of view. First of all, the proof depends on the existence and structure of asymptotic infinity; the theorems ignore complications due to cosmology. This objection is perhaps not *too* serious, for the condition of asymptotic flatness is used only to show that at least one generator of $H^+(S)$ could be continued into the future for infinite affine parameter length while remaining in $H^+(S)$. This state of affairs would be expected even in a nonasymptotically flat universe, unless there were many causality violating regions whose Cauchy horizons intersected, as in Fig. 1. The other weakness is the fact that

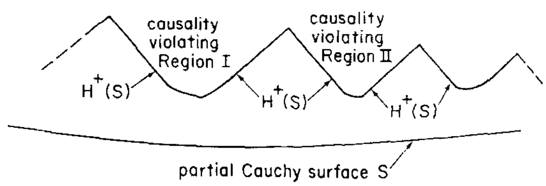

FIG. 1. Multiregion causality violation.

although we know singularities must occur with the formation of CTL in asymptotically flat space-times, we do not know *where* they occur. In this paper it will be shown that if causality violation arises in a *finite* region from regular initial data, then singularities must occur, and further they must occur in the finite region. The word "finite" can be made more precise in several ways; however, note that this word is usually associated in some manner with the mathematical notation of *compactness* (in standard cosmology, for example, a *finite* universe is one whose spacelike sections are topologically compact). Thus one way of making "finite" more precise is given in the following

DEFINITION. A space-time (M, g) will be called *finitely vicious* if it contains a partial Cauchy surface S; if it violates the causality condition in a subset of $J^+(S)$; and if it has a hypersurface slicing $S(\tau)$ with the properties

(i) S is one of the slices with $S(0) = S$;

(ii) there is a closed interval $[\tau_1, \tau_2]$ of the slice parameter such that if $\tau \in [\tau_1, \tau_2]$, then $S(\tau) \cap \tilde{D}^+(S)$ is spacelike, and $S(\tau) \cap H^+(S)$ is compact. Also, if τ_3, τ_4 are any two numbers in $[\tau_1, \tau_2]$ with $\tau_4 > \tau_3$, then $S(\tau_4) \cap \tilde{D}^+(S)$ lies to the future of $S(\tau_3) \cap \tilde{D}^+(S)$;

(iii) let B be the region of space-time in $\tilde{D}^+(S)$ between $S(\tau_1)$ and $S(\tau_2)$ inclusive, and let γ be any segment of a generator of $H^+(S)$ with $\gamma \cap S(\tau_2) \neq \varnothing$ and $\gamma \subset B$. Then γ can be extended in $H^+(S) \cap B$ such that the extension intersects each $S(\tau)$ for $\tau \in [\tau_1, \tau_2]$ exactly once. (The adverb "finitely" comes from the compactness of $H^+(S) \cap S(\tau)$; "vicious" is the adjective Carter [21] applies to any set containing CTL.)

The precise definition of a finitely vicious space-time is somewhat complicated, but the physical situation it is meant to model is actually very simple, and is illustrated in Fig. 2. If causality violation occurs somewhere in the future of a partial Cauchy surface S, a Cauchy horizon, $H^+(S)$, must develop to separate the acausal region from the interior of $\tilde{D}^+(S)$, since by Lemma 4, the strong causality condition holds on int $\tilde{D}^+(S)$. In the region where causality violation begins, $H^+(S)$ may have a very strange structure—indeed, the structure of the entire region may be very strange—but far away from this region we would expect space-time to take on familiar features. For example, we would expect that it would be possible to slice the space-time with a sequence of hypersurfaces, such that these hypersurfaces become spacelike far away from the region where causality violation begins. Furthermore, if we continue suffi-

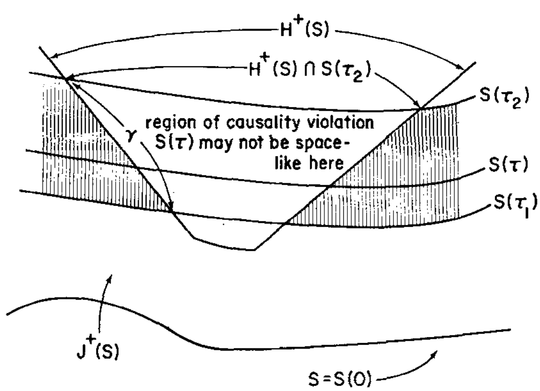

FIG. 2. A finitely vicious space-time. The shaded region is region B, and S is a partial Cauchy surface.

ciently far into the future along the hypersurface slicing (increasing τ), we should, if the strong gravitational fields which give rise to the causality violation are "localized" in space, come at last to a region of space-time where the light ray trajectories have the familiar property of intersecting a spacelike hypersurface once and only once. This region I have denoted with the letter "B" in the definition of "finitely vicious." Condition (iii) in this definition has been stated so as to allow the possibility that a generator of $H^+(S)$ is closed or almost closed; it is possible that a generator of $H^+(S)$ could intersect the region B several times. However, such a generator must leave B, enter the strong field region and reenter B; it *cannot* close or almost close entirely in B. The portions of the hypersurfaces in B have many of the properties of Cauchy surfaces. (Condition (iii) focuses attention on those generators of $H^+(S)$ which *end* on $S(\tau_2)$, the future boundary of the region B, because it is possible for null geodesics to enter $H^+(S)$ as generators as the geodesics move into the past. However, Lemma 3 assures us that a generator of $H^+(S)$ will never leave $H^+(S)$ in the past direction; if a null geodesic is a generator of $H^+(S)$ when it hits $S(\tau_2)$, it will be a generator of $H^+(S)$ when it hits $S(\tau_1)$, and when it hits any hypersurface $S(\tau)$ in between.)

We want to consider at present only those space-times in which the causality violating region is "localized in space," so we must find some way of making "localization in space" precise. One way of accomplishing this is to require $H^+(S) \cap S(\tau)$ to be compact since $H^+(S)$ is the boundary between the acausal region and the causal region. The remainder of condition (ii) is devoted to making certain that the word "space" has a meaning.

If a time machine were manufactured in our universe, our space-time would probably be finitely vicious. For it should be possible to find a sequence of hypersurfaces through the world tube of our galaxy which would be spacelike except in the immediate vicinity of the time machine. The boundary between the causal and acausal region would begin in the strong gravitational field near the machine, and expand until it finally intersected the world tube of the earth, where the properties of null geodesics are that required by condition (iii). We would expect the boundary to expand in all directions into regions of the galaxy where the "local" causal behavior is similar to that of the earth. For a certain period, the period between τ_1 and τ_2, say, the boundary $H^+(S)$ would be confined within our galaxy and so $S(\tau) \cap H^+(S)$ would be contained within a compact set. (If $H^+(S)$ expands in all directions away from the strong field region, then there should be a neighborhood of $H^+(S) \cap S(\tau)$ with compact closure.) Since both $H^+(S)$ and $S(\tau)$ are closed, it follows that $H^+(S) \cap S(\tau)$ is *itself* compact.

The expansion of the Cauchy horizon is an essential feature of a time machine, for the region of causality violation must expand to encompass the earth. This expansion also guarantees the occurrence of singularities, as shown by

THEOREM 2. *Suppose that a space-time (M, g) is finitely vicious, and suppose that the area of $H^+(S) \cap S(\tau_2)$ is strictly greater than the area of $H^+(S) \cap S(\tau_1)$. Then if the weak energy condition and Einstein equations hold, the space-time (M, g) is null geodesically incomplete.*

(The area of $H^+(S) \cap S(\tau)$ is defined in the same manner as black hole area: see [HE, p. 318].)

Proof. The proof is essentially the time reverse of the proof of Hawking's well-known Black Hole Area Law [HE, pp. 318–319]. Since any genrator of $H^+(S)$ which intersects $S(\tau_2)$ intersects $S(\tau_1)$ exactly once, the area of $H^+(S) \cap S(\tau)$ can increase from $S(\tau_1)$ to $S(\tau_2)$ only if the expansion θ of some of the generators of $H^+(S)$ is positive somewhere in B. This means that $\theta < 0$ along at least one generator γ as we move into the past. If γ were geodesically complete, then by the weak energy condition, the Einstein equations, and Eq. (2.1), some of the generators of $H^+(S)$ would intersect to the past of $H^+(S) \cap S(\tau_2)$. This is impossible by Lemma 3. Thus γ cannot be past complete; the space-time (M, g) is null geodesically incomplete. ∎

Notice that Theorem 2 provides us with information about the location of the singularity in a finitely vicious space-time with an expanding Cauchy horizon. A null generator of $H^+(S)$ ends in a singularity somewhere between $S(\tau_1)$ and $S = S(0)$.

The concept of finite viciousness was introduced as one method of making the notion of "finite region" precise. One could prove that the construction of a time machine would generate singularities without using this concept. First note that $\theta > 0$ somewhere along at least *one* null geodesic generator of $H^+(S)$, for the region of causality violation must expand to include the earth. We then infer from the proof of Theorem 2 that the condition $\theta > 0$ is a sufficient condition for the occurrence of singularities.

4. Compact Bounded Vicious Space-Times

Another way we can give meaning to the expression "causality violation begins in a finite region" is to say "causality violation begins in a region of space-time with *compact* boundary."

Definition. A space-time (M, g) will be called *compact bounded vicious* (c.b. vicious) if it contains a partial Cauchy surface S and a compact hypersurface A with the properties

(i) A is the boundary of a closed set B which has nonempty intersection with a causality violating region. A is not required to be connected;

(ii) $A \cap \tilde{D}^+(S)$ is nonempty, but $B \cap I^-(S) = \varnothing$;

(iii) if $H^+(S) \cap A$ is nonempty, then $U \cap A$ is spacelike and connected, where U is some neighborhood of $H^+(S) \cap A$.

Figure 3 gives an example of a c.b. vicious space-time. The definition is straightforward; only a few clarifying remarks need be made. First of all, the first part of

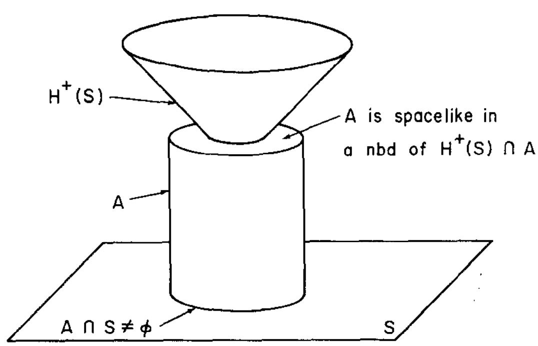

Fig. 3. A c.b. vicious space-time. Causality violation occurs somewhere inside the cone $H^+(S)$. A is the boundary of the cylinder; B is the entire cylindrical solid, including A.

condition (ii) is necessary in order to avoid situations like the one depicted in Fig. 4, where the compactness of C gives us absolutely no restriction on the size of the causality violating region. The second part of condition (ii) tells us that B is located entirely to the future of S. We state this requirement as $B \cap I^-(S) = \varnothing$ rather than $B \subset J^+(S)$ in order to include space-times like the one in Fig. 5.

Since causality is violated in B, but int $\tilde{D}^+(S) \cap B \neq \varnothing$, we know that $H^+(S) \cap B$

is nonempty. However, in some c.b. vicious space-times it is possible to choose A such that $H^+(S) \cap A$ is empty; it is possible to do this, for example, in the Taub–NUT universe, as is shown in Fig. 6. Also, the set A pictured in this figure is an example of a set A which is not connected. The fact that A is not required to be cpnnected constitutes a slight modification in the definition of hypersurface given in HE, where a hypersurface is said to be an orientable imbedded paracompact C^∞ connected Hausdorff three-dimensional submanifold of M without boundary [HE, pp. 14, 44]. But the hypersurface A possesses all of the other properties listed in the preceding sentence.

If the region where causality violation begins is "localized"—say it is confined to a small region in our galaxy—it should be possible to find an A sufficiently large so that A is spacelike in a neighborhood U of $H^+(S) \cap A$. The justification of this possibility

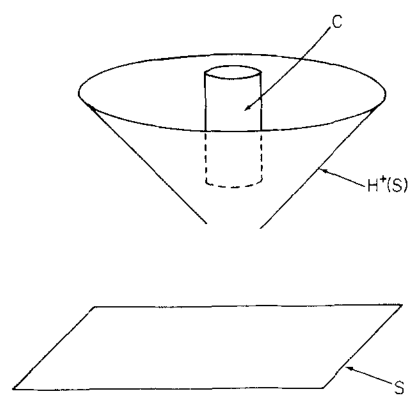

FIG. 4. A set C which has all the properties of the set B in a c.b. vicious space-time except (ii). Causality is violated everywhere in $J^+(S) - \tilde{D}^+(S)$; everywhere inside the cone forming $H^+(S)$.

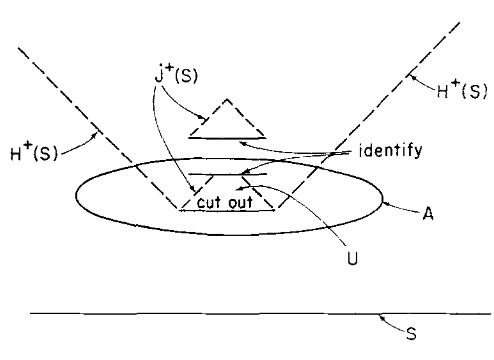

FIG. 5. A c.b. vicious space-time in which $B \cap I^-(S) = \phi$, but $B \not\subset J^+(S)$. Except for the cuts and identifications, the space-time is Minkowskian. The ellipse is the set A. B is the region enclosed by A. By Theorem 3, the set B contains singularities. We cannot eliminate a priori the possibility that a singularity will "shield" some portion U from a causal curve with past endpoints on S; that is, $U \cap J^+(S) = \phi$, but $U \subset B$.

has already been given in the discussion of finitely vicious space-times. $U \cap A$ is required to be connected in order to avoid "A's" like that pictured in Fig. 7.

The reader may wonder why the *boundary* of B was the set required to be compact rather than B itself. In other words, why not define a "compact vicious" space-time rather than a "compact bounded vicious" space-time? The answer to this question is simple. It will be shown in the next theorem that B *cannot* be compact in a physically realistic space-time. This fact makes it difficult to prove the occurrence of incomplete causal geodesics in B, for it is possible that the noncompactness of B results from causal curves in B "going off to infinity" rather than terminating in singularities.

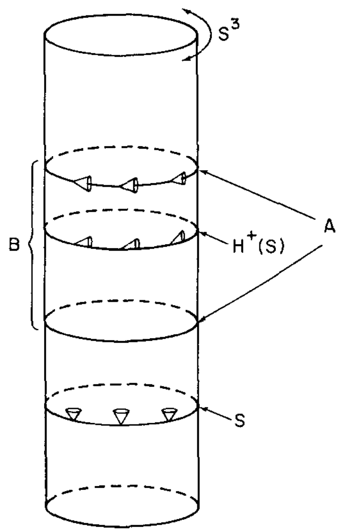

FIG. 6. The Taub–NUT universe, a c.b. vicious space-time. CTL occur to the future of $H^+(S)$.

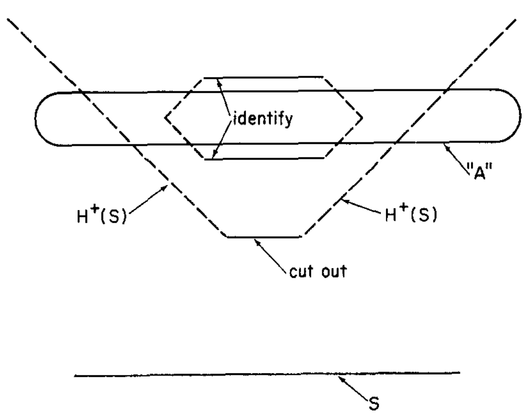

FIG. 7. A set "A" which has all the properties of the set A in a c.b. vicious space-time except $U \cap$ "A" is not spacelike and connected in a neighborhood U of $H^+(S) \cap$ "A".

In the following theorem, we shall eliminate this possibility by making use of the property of int $\tilde{D}^+(S)$ given by Lemma 5. To each point $p \in$ int $\tilde{D}^+(S)$, there is a timelike geodesic from S to p of maximal length. We shall say that the part of B in int $\tilde{D}^+(S)$ is *finite* if there is an upper bound to the lengths of causal curves from S to points in [int $\tilde{D}^+(S)$] $\cap B$. We can express this in the notation of HE by definining $d(p, q)$ for points $p, q \in M$ to be zero if $q \notin J^+(p)$ and otherwise to be the least upper bound of the lengths of future-directed piecewise non-spacelike curves from p to q. For sets S and U, we define $d(S, U)$ to be the least upper bound of $d(p, q)$, $p \in S$, $q \in U$ [HE, *p.* 215]. Thus [int $\tilde{D}^+(S)$] $\cap B$ is finite if and only if $d(S$, [int $\tilde{D}^+(S)] \cap B)$ is finite; it is finite if no causal curve from S can remain in [int $\tilde{D}^+(S)$] $\cap B$ forever. We can now prove that any physically realistic finite c.b. vicious space-time is singular.

THEOREM 3. *If the ubiquitous energy condition and the Einstein equations hold on the set B of a* c.b. *vicious space-time, then B is noncompact. Furthermore, if $d(S$,* [int $\tilde{D}^+(S)$] $\cap$ *$B)$ is finite, then the space-time is timelike geodesically incomplete: there must be an incomplete timelike geodesic in B.*

Scholium. The above theorem is still true if we replace the assumption that the ubiquitous energy condition holds on B with two much weaker assumptions (1) the weak energy condition, and (2) the assumption that there is at least *one* point $q \in H^+(S) \cap B$ such that $K^a K^b K_{[c} R_{d]ab[e} K_{f]} \neq 0$ at q (K^a is the tangent vector to a generator of $H^+(S)$ through q). Thus B will be noncompact if at least *one* generator of $H^+(S)$ in B feels tidal force at least *one* point in its history in B. (Condition (2) can be replaced by the generic condition.)

We will need the following two propositions in the proof of Theorem 3.

PROPOSITION 2. *Let S be a partial Cauchy surface, and let K be a compact set which is contained in* int $\tilde{D}^+(S) \cup S$. *Then $\overline{J^-(K) \cap S}$ is compact.*

Proof. Since the open sets $I^-(p)$, $p \in$ int $\tilde{D}^+(S)$, cover [int $\tilde{D}^+(S)] \cup S$, and since K is compact, there are a finite number of points $p_i \in$ int $\tilde{D}^+(S)$ such that $K \subset \bigcup_i I^-(p_i)$. Since $I^-(p_i) \subset J^-(p_i)$, we have $K \subset \bigcup_i J^-(p_i)$. Thus $J^-(K) \subset \bigcup_i J^-(p_i)$, which gives

$$J^-(K) \cap S \subset J^-(K) \cap J^-(S) \subset \left[\bigcup_i J^-(p_i)\right] \cap J^+(S) = \bigcup_i [J^-(p_i) \cap J^+(S)].$$

By Lemma 7, $J^-(p_i) \cap J^+(S)$ is compact, so $\bigcup_i [J^-(p_i) \cap J^+(S)]$ is compact. Now, since by definition S is closed, $\{\bigcup_i [J^-(p_i) \cap J^+(S)]\} \cap S$ is compact. But

$$\overline{J^-(K) \cap S} = \overline{J^-(K) \cap S} \cap S \subset \left\{\bigcup_i [J^-(p_i) \cap J^+(S)]\right\} \cap S$$

which implies $\overline{J^-(K) \cap S}$ is contained in a compact set. Thus $\overline{J^-(K) \cap S}$ is compact. ∎

PROPOSITION 3. *Let S be a partial Cauchy surface. If $H^+(S)$ is nonempty, then* *any generator of $H^+(S)$ which is totally past imprisoned inside a compact set B is geodesically complete in the past direction.*

The proof, like the proposition, is an obvious generalization of Lemma 8.5.5 of [HE, pages 295–297]. However, it should be noted that the HE proof of Lemma 8.5.5 contains a few algebraic errors; see my Ph.D. thesis for details [10].

Proof of Theorem 3. Suppose that B is compact. Since the causality condition is violated in B and $B \cap \tilde{D}^+(S) \neq \varnothing$, $H^+(S)$ must be nonempty, with some generators of $H^+(S)$ intersecting A in the past direction unless $H^+(S)$ is entirely contained in B. If $H^+(S) \subset B$, then $H^+(S)$ is compact since $H^+(S)$ is closed and B is compact. If $H^+(S) \cap A \neq \varnothing$, then any generator γ of $H^+(S)$ which once enters B can never leave B. For if γ did leave B it would have to intersect A at least once more. But this is impossible since $A \cap U$ is spacelike and connected in some neighborhood U of $H^+(S) \cap A$, and the initial intersection of the generator γ with A defines the future side of the spacelike region of A around $H^+(S) \cap A$ as facing outward from int B. Thus if the generator γ were to leave B, it would have to be future directed upon leaving B; this cannot occur since the space-time is time orientable.

By Proposition 3 the generators of $H^+(S)$ which are totally past imprisoned in B are geodesically complete in the past direction. Consider the expansion θ of the tangent vectors $\partial/\partial\lambda$ to the null geodesic generators of $H^+(S) \cap B$. Suppose that $\theta > 0$ at some point p on a generator γ of $H^+(S) \cap B$. Then by the ubiquitous (or weak) energy condition, the Einstein equations, and Eq. (2.1), some of the generators of $H^+(S)$ would intersect to the past of p. This is impossible by Lemma 3. Therefore, $\theta \leqslant 0$ on $H^+(S) \cap B$.

Now consider the family of differentiable maps

$$\beta_z : H^+(S) \cap B \to H^+(S) \cap B$$

defined by taking a point $q \in H^+(S) \cap B$ a distance z to the past along the null geodesic generator through q, where z is the proper distance in the metric g'_{ab}. The positive definite metric g'_{ab} is defined by introducing a future-directed vector field $\mathbf{V}$ which is geodesic in a neighborhood U of $H^+(S) \cap B$ with compact closure. Then $g'(\mathbf{X}, \mathbf{Y}) = g(\mathbf{X}, \mathbf{Y}) + 2g(\mathbf{X}, \mathbf{V})\, g(\mathbf{Y}, \mathbf{V})$. Let dA be the area measured in the metric g'_{ab} of a small element of $H^+(S) \cap B$. Under the map β_z,

$$(d/dz)\, dA = -\theta\, dA.$$

Now since $H^+(S) \cap B$ is compact, the integral

$$\int_{H^+(S)\cap B} dA$$

must be finite. Since $d/dz(dA) \geqslant 0$, and since z is not bounded above, this is possible only if every geodesic generator γ of $H^+(S) \cap B$ is a closed null geodesic which is entirely contained in B, for in this case z would be cyclic. Furthermore, it is possible for z to be cyclic only if $\theta = 0$ on $H^+(S) \cap B$. But by Eq. (2.1), $\theta = 0$ on $H^+(S) \cap B$ only if $-R_{ab}K^aK^b - 2\sigma^2 \equiv 0$ on $H^+(S) \cap B$. Using the Einstein equations, together with either the ubiquitous energy condition or the weak energy condition and the generic condition, we see that this is impossible. (The Einstein equations and the ubiquitous energy condition imply $R_{ab}K^aK^b > 0$ everywhere on $H^+(S) \cap B$. The Einstein equations, the weak energy condition, and the generic condition (or condition (2) of the Scholium to Theorem 3) imply that $R_{ab}K^aK^b + 2\sigma^2 > 0$ at least one point on some γ.) We have a contradiction and thus $H^+(S) \cap B$ must be noncompact. This implies that B is also noncompact, since $H^+(S)$ is closed.

Furthermore, since $\tilde{D}^+(S)$ is closed, $\tilde{D}^+(S) \cap B$ is noncompact. For suppose it were compact. Then $H^+(S) \cap B$ would be a closed subset of $\tilde{D}^+(S) \cap B$, hence compact, in contradiction to the above result. Now assume $H^+(S) \cap A \neq \varnothing$. Since $A \cap U$ is spacelike and connected in some neighborhood U of $H^+(S) \cap A$, there is a neighborhood $W \subset U$ of $H^+(S) \cap A$ such that $I^-(H^+(S) \cap A) \cap W \subset \operatorname{int} B$. (That is, any timelike curve with future endpoint on $H^+(S) \cap A$ immediately enters $\operatorname{int} B$ as it moves into the past, and remains in $\operatorname{int} B$ for *some* proper time at least.)

Now let V be a neighborhood of $H^+(S) \cap A$ contained in U. Then $\overline{A - V \cap A}$ is compact, since it is a closed subset of the compact set A. We must have $I^-(H^+(S) \cap B) \cap A \subset \overline{A - V \cap A}$, since if this were *not* true, there would be a past-directed timelike curve from $\operatorname{int} B$ which intersects $A - V$ on *leaving* $\operatorname{int} B$. But $V \cap A$ is spacelike and past-directed timelike curves can intersect $V \cap A$ only upon entering $\operatorname{int} B$, since the space-time is time orientable, and A is orientable. Furthermore, we have

$$I^-(H^+(S) \cap B) \subset A \cap \operatorname{int} \tilde{D}^+(S) \cup S \tag{4.1}$$

since $I^-(H^+(S)) - I^-(S) \subset \operatorname{int} \tilde{D}^+(S) \cup S$ and in addition $A \cap I^-(S) = \varnothing$. Thus

$$I^-(H^+(S) \cap B) \cap A \subset \overline{A - V \cap A} \cap [\operatorname{int} \tilde{D}^+(S) \cup S]. \tag{4.2}$$

Define $\overline{A - V \cap A} \cap [\operatorname{int} \tilde{D}^+(S) \cup S] \equiv K$. Note that K is compact, since $\overline{A - V \cap A}$ consists of disjoint sets, some in $\operatorname{int} \tilde{D}^+(S) \cup S$, and the others not. (Recall that $\overline{A - V \cap A} \cap H^+(S) = \varnothing$.) Thus by Proposition 2, $\overline{J^-(K) \cap S}$ is compact.

Since $B \cap I^-(S) = \varnothing$, any timelike curve drom B which intersects S must intersect A. Thus

$$I^-(H^+(S) \cap B) \cap S \subset I^-([I^-(H^+(S) \cap B)] \cap A) \cap S. \tag{4.3}$$

By (4.2), the set on the RHS of (4.3) is contained in $I^-(K) \cap S$. Furthermore,

$$I^-(K) \cap S \subset \overline{J^-(K) \cap S} = L$$

so we have, finally

$$I^-(H^+(S) \cap B) \cap S \subset L.$$

By a similar argument $I^-(\text{int}\,\tilde{D}^+(S) \cap B) \cap S \subset L$. Combining these results, we can conclude that $I^-(\tilde{D}^+(S) \cap B) \cap S$ is contained within a compact set L. (If $H^+(S) \cap A = \varnothing$, we set $K \equiv A \cap [\text{int}\,\tilde{D}^+(S) \cup S]$ and $L \equiv \overline{J^-(K) \cap S}$.)

Now write $d(S, [\text{int}\,\tilde{D}^+(S)] \cap B) \equiv \alpha$. Suppose the space-time (M, g) is timelike geodesically complete. Let $\beta\colon L \times [0, \alpha] \to M$ be the differentiable map which takes a point $p \in L$ a distance $s \in [0, \alpha]$ up the future-directed geodesic through p orthogonal to L. Then $\beta(L \times [0, \alpha])$ is compact, and it contains the set $\text{int}\,\tilde{D}^+(S) \cap B$ since by Lemma 5, there is to each point $q \in \text{int}\,\tilde{D}^+(S) \cap B$ a future-directed geodesic γ orthogonal to S of maximal length from S to q, and by assumption, γ has length less than or equal to α. Furthermore, since $I^-(\tilde{D}^+(S) \cap B) \cap S \subset L$, γ intersects S *in* L. Thus

$$\tilde{D}^+(S) \cap B = \overline{\tilde{D}^+(S) \cap B} \subset \beta(L \times [0, \alpha])$$

and hence $\tilde{D}^+(S) \cap B$ is compact. But we have already shown that $\tilde{D}^+(S) \cap B$ is noncompact. We have a contradiction, so (M, g) cannot be timelike geodesically complete; there must be an incomplete timelike geodesic in B. ∎

5. Singularities, CTL, and Topology Change

Besides its intrinsic interest, Theorem 3 has a number of important applications, most notably to the question of topology change. By Lemma 6, topology change can occur only if a Cauchy horizon forms. However, this Cauchy horizon could be due either to the formation of singularities or to causality violation (or both). Many workers have discussed the causal case [11, 22, 23], but whether or not causality violation could give rise to nonsingular topology change has up to now been an open question. We can use Theorem 3 to show singularities *must* accompany topology change in a physically realistic space-time, whether or not causality violation occurs.

To show this for closed universes we will require a lemma due to Geroch [11, 22].

Lemma 12. *Let B be a compact subset of a space-time such that the boundary of B is the disjoint union of two compact spacelike 3-manifolds, S and S'. Suppose that the causality condition holds on B. Then S and S' are diffeomorphic, and further B is topologically $S \times [0, 1]$.*

We first use Theorem 3 to show that if B is compact, then, in a physically realistic space-time, topology change cannot occur *at all.*

Theorem 4. *Let B be a compact subset of a space-time such that the boundary of B is the disjoint union of two compact spacelike 3-manifolds, S and S'. If*

(1) *the weak energy condition and the Einstein equations hold on B;*

(2) $K^aK^bK_{[c}R_{d]ab[e}K_{f]} \neq 0$ *at at least one point $p \in B$ on every null geodesic which is totally past-imprisoned in B.*

Then S and S' are diffeomorphic, and further B is topologically $S \times [0, 1]$.

Comment. Since we are using the notation of HE, both S and S' are connected. The theorem is still true even if one (or both) is not connected, but the proof is a bit more complex than the one given here. (Condition (2) could be replaced by: (3) the generic condition holds on B.)

Proof. Since S and S' are spacelike and orientable, and the space-time is time orientable, any causal curve from S into B can intersect S again only after first leaving B through S'. Similarly for S'. Thus we can attach B to a space-time (M, g') in which S and S' are acausal and $B \cap I^-(S) = \varnothing$. If S and S' are not diffeomorphic, then by Lemma 12, causality violation occurs in B. The space-time (M', g') is c.b. vicious, with B in the definition of c.b. vicious space-time chosen to equal the B of the present theorem, and $A = S \cup S'$. Since $S \subset \tilde{D}^+(S)$, condition (ii) holds, S is a partial Cauchy surface since it is a 3-D acausal spacelike manifold without boundary. Since S' is connected, if $H^+(S) \cap A = H^+(S) \cap S'$ is nonempty, then $S' \cap U$ is obviously spacelike and connected in some neighborhood U of $H^+(S) \cap A$. Conditions (1) and (2) are the same as those in the Scholium to Theorem 3, so the conclusions to Theorem 3 hold here. (Note that any generator of $H^+(S)$ which intersects B is totally past-imprisoned in B; in fact, such a generator is totally *future* imprisoned in B.) Thus B must be noncompact, contrary to assumption. Hence. S and S' are diffeomorphic, and further B is topologically $S \times [0, 1]$. ∎

We can state Theorem 4 in a somewhat different form. If topology change *does* occur, then B is not compact and contains a singularity if B is finite.

THEOREM 5. *Let B be a four-dimensional region of a space-time such that the boundary of B is the disjoint union of two compact partial Cauchy surfaces, S and S'. Suppose that*

(1) *the weak energy condition and the Einstein equations hold on the space-time:*
(2) *the generic condition holds on the space-time:*
(3) *$d(S, [\mathrm{int}\, \tilde{D}^+(S)] \cap B)$ is finite;*
(4) *S and S' are not diffeomorphic.*

Then the space-time is timelike geodesically incomplete and B is not compact.

The proof is similar to that of Theorem 4, and so is omitted. ∎

We can prove a theorem analogous to Theorem 4 for open universes, provided we assume the possible topology change to be localized in a compact region. Following Geroch [22, pp. 51–53] we define a 3-manifold S to be *externally Euclidean* if there exists a connected compact set C of S such that $S - C$ is diffeomorphically $S^2 \times R$. That is, outside of the region C, S is identical with ordinary Euclidean space with a 3-ball removed. Let M be a four-dimensional subset of space-time whose boundary is the disjoint union of two externally Euclidean spacelike 3-manifolds S and S'. Suppose that there exists a connected compact set K of M such that $M - K$ is diffeomorphic to $S^2 \times R \times [0, 1]$, where for each fixed number $\alpha \in [0, 1]$ the submanifold

$S^2 \times R$ of M is spacelike and for each fixed point p of $S^2 \times R$ the line $[0, 1]$ of M is timelike. Then M will be called *externally Lorentzian*. K may be described as a timelike world tube between S and S' in which any "topology change" must take place. Analogous to Lemma 12, we have

LEMMA 13 (Geroch [22; p. 55]). *Let M be an externally Lorentzian portion of space-time, the boundary of M being the disjoint union of two spacelike externally Euclidean 3-manifolds, S and S'. Suppose M has no closed timelike curves. Then S and S' are diffeomorphic, and M is diffeomorphically $S \times [0, 1]$.*

We can strengthen this result.

THEOREM 6. *Let M be an externally Lorentzian portion of space-time, the boundary of M being the disjoint union of two spacelike externally Euclidean 3-manifolds, S and S'. If S is a partial Cauchy surface for the entire space-time, and a Cauchy surface for $M - K$, and in addition we assume*

(1) *the weak energy condition and the Einstein equations hold on M;*

(2) *the generic condition holds on the space-time.*

Then S and S' are diffeomorphic, and S is a Cauchy surface for M.

(The Cauchy surface condition makes certain that the breakdown in prediction is restricted entirely to K; $M - K \subset \operatorname{int} \tilde{D}^+(S)$.)

Proof. If the conclusion does not hold, then causality is violated in K. We now proceed as in Theorem 4, showing that M can be regarded as part of a c.b. vicious space-time. Choose $B = K$ and $A = \partial K$. (A will have to be smoothed out at $\partial K \cap S$ and $\partial K \cap S'$, to make sure that A is a C^∞ manifold, but this is a minor technicality.) The remainder of the proof is very similar to that of Theorem 4, and will omitted. ∎

6. CTL AND COMPACT SPACE-TIMES

Another connection between the concepts of compactness and causality violation is provided by compact space-times.

LEMMA 14. (Geroch [11]; HE [p. 189]). *If (M, g) is compact, then the chronology condition is violated in the space-time.* ∎

The proof is easy. We simply note that M can be covered by open sets of the form $I^+(q)$ with $q \in M$ and recall that the chronology condition holds at q only if q is not in $I^+(q)$. Thus M can be covered by a *finite* number of sets $I^+(q)$ only if the chronology condition is violated in M.

The important thing to notice about this proof is that it is entirely topological in nature; there is no reference to the dynamics of space-time—there is no reference to Einstein's equations. If one uses Einstein's equations, we can prove a stronger theorem than Lemma 14; we can prove that in a generic, compact space-time, the chronology condition is violated at *every* point.

DEFINITION. A space-time (M, g) will be called *totally vicious* if $I^+(q) \cap I^-(q) = M$ for some $q \in M$.

Notice that if $I^+(q) \cap I^-(q) = M$ is true for *one* $q \in M$, it will be true for *all* $q \in M$ since the choronoly violating set, denoted by V, is the disjoint union of open sets of the form $I^+(q) \cap I^-(q)$. Thus if (M, g) is totally vicious, then every point in M can be connected to every other point by both a future-directed and a past-directed timelike curve. The causality violation is *total.*

THEOREM 7. *If (M, g) is compact, then (M, g) is totally vicious provided the following conditions hold*

(1) *the Einstein equations and the weak energy condition,*
(2) *the generic condition.*

Proof. By Lemma 14, (M, g) violates the chronology condition, so there is a point $q \in M$ for which $I^+(q) \cap I^-(q) \neq \varnothing$. If (M, g) were not totally vicious, the set $\dot{I}^+(q) \cup \dot{I}^-(q)$ would be nonempty. Suppose $\dot{I}^+(q) \neq \varnothing$. Now $\dot{I}^+(q)$ is generated by null geodesic segments which have no past endpoints. For a generator of $\dot{I}^+(q)$ can have an endpoint only at q, and q cannot be an endpoint without violating the achronolity of $\dot{I}^+(q)$ (since $q \ll q$). We can now argue as in Proposition 3 to show that all generators of $\dot{I}^+(q)$ are geodesically complete in the past direction. But this contradicts conditions (1) and (2) as an argument similar to the one found in the proof of Theorem 3 will show. If $\dot{I}^+(q) \neq \varnothing$, we can obtain a contradiction in the same way. Thus $\dot{I}^+(q) \cup \dot{I}^-(q)$ must be empty, which means that (M, g) is totally vicious. ∎

7. CAUSALITY VIOLATION AND THE HAWKING–PENROSE–GEROCH SINGULARITY THEOREMS

Besides the Geroch Topology Change Theorems, there is another important class of theorems in General Relativity which use the causality condition: the Hawking–Penrose–Geroch Singularity Theorems. In an earlier section of this paper, I showed the causality condition to the nonessential in the first class of theorems; in this section, I shall argue that the causality condition is an unnecessary assumption in the singularity theorems. More precisely, I shall generalize the two most important singularity theorems, showing that singularities can be prevented by causality violation *only* if the causality violation begins at "infinity," an unlikely possibility.

As the first step in the generalization, recall the two most important singularity theorems.

HAWKING–PENROSE THEOREM [HE, p. 266]. *Space-time (M, g) is not timelike and null geodesically complete if*

(1) $R_{ab}K^aK^b \geqslant 0$ *for every non-spacelike vector* K^a *(this can be inferred from the strong energy condition and the Einstein equations)*;

(2) *the generic condition is satisfied*;

(3) *the chronology condition holds on* M;

(4) *there exists at least one of the following*:

(i) *a compact achronal set without edge,*

(ii) *a closed trapped surface,*

(iii) *a point* p *such that on every past (or every future) null geodesic from* p *the divergence* θ *of the null geodesics from* p *becomes negative (i.e., the null geodesics from* p *are focussed by the matter or curvature and start to reconverge).*

PENROSE'S THEOREM [HE, p. 263]. *Space-time cannot be null geodesically complete if*

(1) $R_{ab}K^aK^b \geqslant 0$ *for all null vectors* K^a *(this can be inferred from the weak energy condition and the Einstein equations).*

(2) *there is a noncompact Cauchy surface in* M;

(3) *there is a closed trapped surface in* M.

(For cosmological applications, condition (3) in Penrose's theorem can be replaced by condition 4(iii) of the Hawking–Penrose theorem; in *open* universes, we need only the *weak* energy condition to infer the existence of singularities.)

The generalization of the above theorems will be based on the following two definitions.

DEFINITION. A space-time (M, g) is said to be *asymptotically deterministic* if

(i) (M, g) contains a partial Cauchy surface S such that

(ii) either $H(S) \equiv H^+(S) \cup H^-(S)$ is empty, or, if not, then

$$\lim_{s \to a} [\inf T_{ab}K^aK^b] > 0$$

on at least one of the null geodesic generators $\gamma(s)$ of $H(S)$, where a is the past limit of the affine parameter along γ if $\gamma \in H^+(S)$, and the future limit if $\gamma \in H^-(S)$. (K^a is the tangent vector to γ.)

DEFINITION. The matter tensor will be said to be *past stochastic* along a causal geodesic segment γ if there exist numbers $a > 0$, $b > 0$, and an integral number c of disjoint affine parameter intervals $(s_1, s_2), (s_3, s_4), \ldots, (s_i, s_{i+1}), \ldots$ along γ, each interval satisfying $| s_j - s_{j+1} | \geqslant b$ (and $s_j > s_{j+1}$ for all j) with $T_{ab}K^aK^b \geqslant a$ at every point in every interval. (K^a is the tangent vector to γ.) Furthermore, c is finite if γ has a past endpoint or is past incomplete, and infinite if γ is past complete. *Future stochastic* matter tensors are defined similarly.

These definitions are based on the following physical reasoning. If causality violation occurs in a space-time with a partial Cauchy surface S, then $H(S)$ must be nonempty. For the sake of argument, suppose the causality violation occurs to the future of S, giving $H^+(S) \neq \emptyset$. If the formation of a Cauchy horizon is due to causality violation, then we would expect that the region where $H^-(S)$, and hence the causality violation, begins would contain matter. It requires nonzero gravitational fields to "tip over the light cones" sufficiently far to give causality violation, and it seems unlikely that these fields would occur in empty space, for nonzero gravitational fields will give rise to matter via pair creation. (There are exceptions to this; for example, the Taub–NUT universe and the modified Minkowski space pictured in Fig. 5, but in these cases there are singularities on the Cauchy horizon, and hence $H^+(S)$ is not due entirely to causality violation. Furthermore, in the actual universe, particle creation would be expected to occur around such singularities.)

The region where $H^+(S)$ "begins" is an open set of the space-time which contains the past inextendible portions of the generators of $H^+(S)$. In asymptotically deterministic space-times, this region is nonempty in the sense that a certain component of the matter tensor as measured in a pseudo-orthonormal frame parallel propagated into the past along at least one of the generators of $H^-(S)$ does not have a vanishing lower bound. It is possible, of course, that the particular component we have selected could vanish as the affine parameter approached its past limit without *all* components vanishing. However, physical considerations show this to be unlikely. Suppose, for example, that the matter can be represented as a perfect fluid, with matter tensor [HE, p. 70]

$$T_{ab} = (\mu + p)\, V_a V_b - p g_{ab}\,.$$

Thus

$$T_{ab}K^aK^b = (\mu + p)(V_aK^a)^2$$

which can vanish $s \to a$ if

$$\begin{aligned}
&\text{case 1:} \quad \mu \to 0 \qquad \text{and} \qquad p \to 0 \qquad (\text{thus } T_{ab} \to 0),\\
&\text{case 2:} \quad \mu \to -p, \qquad \text{with} \quad p \not\to 0,\ \mu \not\to 0,\\
&\text{case 3:} \quad (V_aK^a) \to 0,
\end{aligned}$$

(or both 2 and 3). Case 2 cannot occur if we impose the strong and weak energy condition on the space-time. As for case 3, pick a point b on $\gamma(s)$ with $V_aK^a \neq 0$. Then the frequency shift which a photon traveling along $\gamma(s)$ undergoes is given by [24]

$$\frac{\nu_b}{\nu_a} = \frac{{}_bV_cK^c}{{}_sV_cK^c}$$

where the frequency is measured by observers moving with the fluid. Thus if $V_cK^c \to 0$ as $s \to a$ a photon would suffer an infinite blue shift as it moved from a to b; a space-time which allowed this type of behavior would be unstable, because a photon from a would arrive at b with infinite energy. (See Ellis and King [25, p. 154] for a detailed discussion of this type of instability.) Note that in asymptotically deterministic space-times, the limit

$$\liminf_{s \to a} T_{ab}K^aK^b \tag{7.1}$$

is not required to exist; it is merely required that (7.1) be greater than zero.

However, it is possible to have $T_{ab}K^aK^b \not\to 0$ as $s \to a$ while (7.1) equals zero, for it is conceivable that $T_{ab}K^aK^b$ could "fluctuate" as $s \to a$, with $T_{ab}K^aK^b = 0$ at an infinite number of points if $a = -\infty$. Even if this occurred, we would still expect that the matter tensor would be *at least* stochastic along at least *one* generator $\gamma(s)$ of $H^+(S)$ unless $T_{ab}K^aK^b$ approaches zero in some average sense as $s \to a$; for example, if $\gamma(s)$ intersected a proton every so often, then the matter tensor would be past stochastic along $\gamma(s)$. Note that were $\gamma(s)$ past complete, it would be possible to have

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} | s_i - s_{i+1} |}{| s_1 - s_{N+1} |} = 0$$

so that the nonzero regions are of zero measure in the entire history of $\gamma(s)$, and *still* have the matter tensor past stochastic along $\gamma(s)$.

In summary, then, it seems very reasonable to assume that a space-time in which the Cauchy horizon is due to causality violation is asymptotically deterministic, or at least it has a generator of $H(S)$ along which the matter tensor is past (future) stochastic as the generator of $H^+(S)$ ($H^-(S)$) approaches its past (future) affine parameter limit. With this assumption, we can eliminate the causality condition from Penrose's theorem and the Hawking–Penrose theorem.

THEOREM 8. (Generalized Penrose's Theorem). *Space-time (M, g) cannot be null geodesically complete if*

(1) $R_{ab}K^aK^b \geqslant 0$ *for all null vectors* K^a;

(2) *there is a closed trapped surface in* M;

(3) *the space-time is asymptotically deterministic, and the Einstein equations hold*;

(4) *the partial Cauchy surface defined by* (3) *is noncompact.*

Conditions (3) and (4) can be replaced by: There exists a noncompact partial Cauchy surface S with the property

(3′) if $H(S)$ is nonempty, then there is at least one generator γ of $H^+(S)$ (or $H^-(S)$) along which the matter tensor is past (future) stochastic as γ approaches its past (future) affine parameter limit. Further, the Einstein equations hold.

Proof. If $H(S)$ is empty, then S is a noncompact Cauchy surface for (M, g); by Penrose's theorem, the space-time is null geodesically incomplete. If $H(S)$ is nonempty, for all partial Cauchy surfaces S, there must exist a partial Cauchy surface S' and a generator $\gamma(s)$ of $H^-(S')$ (say) along which

$$\liminf_{s\to a} T_{ab}K^aK^b > 0.$$

If $\gamma(s)$ were future complete, then by the Einstein equations, we have along $\gamma(s)$

$$\lim_{s\to a=+\infty} \int^a H(s)\, ds = +\infty \tag{7.2}$$

where $H(s) = \frac{1}{2}(R_{ab}K^aK^b + 2\sigma^2)$. By Lemma 9, condition (1), and Eq. (2.3), $\gamma(s)$ must have an infinite number of conjugate points, since zeros of (2.3) correspond to conjugate points of γ. But this is impossible because of Lemma 11 and the achonality of $H^-(S^-)$. Thus $\gamma(s)$ must be future incomplete. (Were $\gamma(s)$ in $H^+(S')$ rather than in $H^-(S')$ we could proceed in the same way to deduce the *past* incompleteness of γ.)

If condition (3′) rather than (3) holds, then Eq. (7.2) still holds, and the argument proceeds as before. In all cases, the space-time is null incomplete, since $\gamma(s)$ is a portion of a null geodesic. ▮

THEOREM 9 (Generalized Hawking–Penrose Theorem). *Space-time (M, g) is not timelike and null geodesically complete if*

(1) $R_{ab}K^aK^b \geqslant 0$ *for every non-spacelike vector* K^a;

(2) *the generic condition is satisfied*;

(3) *there exists at least one of the following*:

(i) *a compact achronal set without edge*,

(ii) *a closed trapped surface*,

(iii) *a point p such that on every past (or every future) null geodesic from p the divergence θ of the null geodesics from p becomes negative*,

(4) *the space-time is asymptotically deterministic, and the Einstein equations hold.*

Condition (4) can be replaced by condition (3′) of Theorem 8. The proof of Theorem 9 is similar to the proof of Theorem 8, and so will be omitted. ▮

There is additional evidence that $H(S)$ which arises from causality violation can be nonsingular only if it begins in empty space.

DEFINITION. The matter tensor will be said to be *future vanishing* along a future complete causal geodesic segment $\gamma(s)$ if there are *no* numbers s_1, s_2 $(s_2 > s_1)$ in any affine parameter interval of the form $[s_0, +\infty)$ such that $s_1 > s_0$ and

$$\frac{1}{s_1 - s_0} < \int_{s_1}^{s_2} T_{ab}K^aK^b\, ds \tag{7.3}$$

where K^a is the tangent to $\gamma(s)$. *Past vanishing* matter tensors are defined analogously.

PROPOSITION 4. *Let S be a partial Cauchy surface. Then the matter tensor must be future vanishing along* every *future complete generator of $H^-(S)$, provided the weak energy condition and the Einstein equations hold.*

Proof. Suppose that $\gamma(s)$ is a future complete generator of $H^-(S)$ along which T_{ab} is *not* future vanishing. Then there are affine parameter values $s_0 < s_1 < s_2$ such that (7.3) holds. By the Einstein equations, the weak energy condition, Eq. (2.3), and Proposition 1, $\gamma(s)$ must have a pair of conjugate points, since zeros of (2.3) correspond to conjugate points of $\gamma(s)$. But this is impossible because of Lemma 11 and the achronality of $H^-(S)$. Thus T_{ab} must be future vanishing along $\gamma(s)$. (If "future" is replaced by "past" and "$H^-(S)$" is replaced by "$H^+(S)$" in Proposition 4, the resulting statement is also true.) ∎

Thus there can be no "patches" of matter near the upper bound of the affine parameter of a future complete generator $\gamma(s)$ of $H^-(S)$, for (7.3) would be expected to hold in this case. Note that requiring the matter tensor to be *not* future vanishing along a future complete null geodesic segment $\gamma(s)$ is a much weaker restriction on the tensor than requiring it to be future stochastic along the segment; if we have $T_{ab}K^aK^b \geqslant a$ on *one* affine parameter interval $|s_2 - s_1| \geqslant b$, the matter tensor will *not* be future vanishing, provided the interval is sufficiently near $s = +\infty$. If $\gamma(s)$ intersects *one* proton near infinity, T_{ab} will not be future vanishing. However, either requirement imposed on $\gamma(s)$ would make it impossible for $\gamma(s)$ to be a generator on $H^-(S)$, since in either case $\gamma(s)$ would have a pair of conjugate points.

Not only must the matter vanish as $s \to +\infty$ along a future complete generator of $H^-(S)$; we can also show that the tidal force components of the Weyl tensor cannot approach a nonzero limit as $s \to +\infty$.

PROPOSITION 5. *Let γ be a future complete null geodesic segment with tangent vector K^a and affine parameter s. Suppose that there exists a parallel propagated pseudo-orthonormal frame with $\mathbf{E}_4 = \mathbf{K}$ in which some components of the tensor $K^cK^dK_{[a}C_{b]cd[e}K_{f]}$ satisfy*

$$\liminf_{s\to+\infty} |K^cK^dK_{[a}C_{b]cd[e}K_{f]}| \neq 0. \tag{7.4}$$

Then if the weak energy condition and the Einstein equations, $\gamma(s)$ cannot be a generator of $H^-(S)$, where S is a partial Cauchy surface. (There is a similar proposition for $H^+(S)$.)

Proof. Suppose on the contrary that $\gamma(s)$ is a generator of $H^-(S)$. If the limit (7.4) is to be nonzero, then there must be a value s_0 of the affine parameter for which some component of the tensor $K^cK^dK_{[a}C_{b]cd[e}K_{f]}$ is nonzero for all $s \in (s_0, +\infty)$. Now in a pseudo-orthonormal frame, this implies that some component of $C_{mcdn}K^cK^d$ is nonzero in the same interval [HE, p. 101].

Let h_{mn} be the component of C_{m44n} which is nonzero along $\gamma(s)$ (and satisfies $\liminf_{s\to+\infty} |h_{mn}| \neq 0$; that is, $|h_{mn}| \geqslant c > 0$ for all $s \in (s_0, +\infty)$). Now Eq. (2.2)

$$(d/ds)\,\sigma_{mn} = -C_{m4n4} - \theta\sigma_{mn} \equiv h_{mn} - \theta\sigma_{mn} \tag{2.2}$$

tells us h_{mn} gives rise to a shear σ_{mn} of a geodesic congruence about $\gamma(s)$. This equation is a linear, first-order ordinary differential equation which is easily integrated, giving

$$\sigma_{mn} = e^{-\int \theta ds} \int e^{\int \theta ds} h_{mn} \, ds + A_{mn} e^{-\int \theta ds}$$

where A_{mn} is a constant (the solution can be found in any elementary book on ordinary differential equations; for example, Hildebrand [26, p. 7]).

We want to prove that under the above assumptions $\gamma(s)$ has a pair of conjugate points between s_0 and $s = +\infty$; we will assume that there are no conjugate points in $(s_0, +\infty)$ and derive a contradiction. If there are no conjugate points in $(s_0, +\infty)$, then $\theta \geqslant 0$ in some interval $(s_1, +\infty)$, $s_1 \geqslant s_0$, by the weak energy condition the Einstein equations, and Eq. (2.1). Thus the function

$$f(s) = \int_{s_1}^{s} \theta \, ds$$

is monotone increasing. Furthermore, $f(s)$ must diverge as $s \to +\infty$, since if it converged (a monotone function must either diverge or converge; it cannot merely not have a limit), we would have

$$\begin{aligned}
|\sigma_{mn}| &= \left| e^{-\int_{s_1}^{s} \theta ds} \int_{s_1}^{s} e^{\int_{s_1}^{s} \theta ds} h_{mn} \, ds + A'_{mn} e^{-\int_{s_1}^{s} \theta ds} \right| \\
&\geqslant \left| e^{-\int_{s_1}^{s} \theta ds} \int_{s_1}^{s} e^{\int_{s_1}^{s} \theta ds} h_{mn} \, ds \right| - \left| A'_{mn} e^{-\int_{s_1}^{s} \theta ds} \right| \\
&\geqslant \left| e^{-\int_{s_1}^{\infty} \theta ds} \int_{s_1}^{s} h_{mn} \, ds \right| - | A'_{mn} | \\
&= e^{-\int_{s_1}^{\infty} \theta ds} \int_{s_1}^{s} | h_{mn} | \, ds - | A'_{mn} |.
\end{aligned} \tag{7.5}$$

But

$$\lim_{s \to +\infty} \int_{s_1}^{s} | h_{mn} | \, ds = +\infty$$

which implies that there exists an affine parameter $s_2 > s_1$ for which $|\sigma_{mn}| > 0$ for all $s \in (s_2, +\infty)$ and further $\lim_{s \to +\infty} \inf |\sigma_{mn}| > 0$ (in fact, $\lim_{s \to +\infty} \inf |\sigma_{mn}| = +\infty$). Using this together with the weak energy condition, we find that $\lim_{s \to +\infty} \inf H(s) > 0$, where $H(s) \equiv \frac{1}{2}(R_{ab}K^aK^b + 2\sigma^2) \equiv \frac{1}{2}(R_{ab}K^aK^b + \sigma_{mn}\sigma^{mn})$. Thus $\lim_{s \to +\infty} s^2 H(s) = +\infty$, which by Lemma 8 implies that all solutions to (2.3) have an infinite number of zeros, hence conjugate points, in $(s_1, +\infty)$, contrary to assumption.

Therefore, we must have $\lim_{s\to+\infty} f(s) = +\infty$. Write

$$K(s) \equiv e^{-\int_{s_1}^{s}\theta ds'} \int_{s_1}^{s} e^{\int_{s_1}^{s'}\theta ds''}\, ds' \equiv e^{-f(s)} \int_{s_1}^{s} e^{f(s')}\, ds'.$$

From Eq. (7.5) we can obtain

$$\begin{aligned} |\,\sigma_{mn}\,| &\geqslant c\left| e^{-\int_{s_1}^{s}\theta ds'} \int_{s_1}^{s} e^{\int_{s_1}^{s'}\theta ds''}\, ds' \right| - |\,A'_{mn}\,|\,|\,e^{-\int_{s_1}^{s} ds'}\,| \\ &= c\,|\,K\,| - |\,A'_{mn}\,|\,e^{-f(s)}. \end{aligned}$$

Using l'Hôpital's rule, we get

$$\begin{aligned} \lim_{s\to\infty} K &= \lim_{s\to\infty} \frac{\int_{s_1}^{s} f(s')\, ds'}{e^{f(s)}} = \lim_{s\to\infty} \frac{(d/ds)\int_{s_1}^{s} e^{f(s')}\, ds}{(d/ds)\, e^{f(s)}} \\ &= \lim_{s\to\infty} \frac{e^{f(s)}}{e^{f(s)} f'(s)} = \lim_{s\to\infty} \frac{1}{f'(s)} = \lim_{s\to\infty} \frac{1}{\theta}. \end{aligned}$$

By assumption $\theta \geqslant 0$. By Eq. (2.1), and the weak energy condition, θ is monotone decreasing (i.e., $d\theta/ds \leqslant 0$). Thus $\lim_{s\to\infty} 1/\theta$ must either exist or diverge to $+\infty$. Therefore,

$$\lim_{s\to\infty} |\,\sigma_{mn}\,| \geqslant c \lim_{s\to\infty} K - |\,A'_{mn}\,| \lim_{s\to\infty} e^{-f(s)}$$

since $f(s)$ diverges, the second term vanishes. The first term is bounded below by some positive number since $\theta < \infty$. Hence $\lim_{s\to\infty} |\,\sigma_{mn}\,| > 0$, and this implies $\int^{\infty} H(s)\, ds \geqslant \int^{\infty} \sigma^2\, ds = +\infty$. By Lemma 9 and Eq. (2.3), this means an infinite number of conjugate points in $(s_0 + \infty)$. Thus whether or not $f(s)$ diverges, $\gamma(s)$ must have an infinite number of conjugate points; this contradicts the assumption that it has no conjugate points. Thus there must be a pair of conjugate points on $\gamma(s)$, which would be impossible if $\gamma(s)$ were a generator of $H^{-}(S)$, by Lemma 11, and the achronality of $H^{-}(S)$. We conclude (finally!) that the $\gamma(s)$ cannot be a generator of $H^{-}(S)$. ∎

Proposition 5 does not claim that the entire Weyl tensor must vanish as $s \to +\infty$; first of all, it refers only to the tidal force components, and second, it does not eliminate the possibility that C_{m4n4} could fluctuate about zero, C_{m4n4} could be alternately positive and negative as $s \to +\infty$; for there is no Weyl tensor analog of the weak energy condition to prevent this behavior. However, if Weyl tensor did fluctuate in this manner, we would expect to see some matter present, because a varying gravitational field should give rise to particle creation, not much, to be sure, but by Proposition 4, we do not need much to prevent a nonsingular $H(S)$. Furthermore, it seems unlikely that the tidal force components of the Weyl tensor would vanish in *all* pseudo-orthonormal frames with $\mathbf{E}_4 = \mathbf{K}$ without the entire tensor vanishing also. Since the curvature tensor is determined by the Ricci tensor (or matter tensor) and the Weyl tensor, the theorems and propositions proved in this section collectively strongly suggest that $H(S)$ can be generated by portions of complete geodesics only if the curvature vanishes in the region where $H(S)$ begins. As argued earlier, this seems very unlikely if $H(S)$ arises from a violation of the causality condition, since in general CTL would occur only if curvature were present to "tip over the light cones."

However, violation of the causality condition is not the only pathology which could give rise to a Cauchy horizon; for example, $H(S)$ could be due to singularities which arise at "infinity," as in the Reissner–Nordström solution (Fig. 8) and the

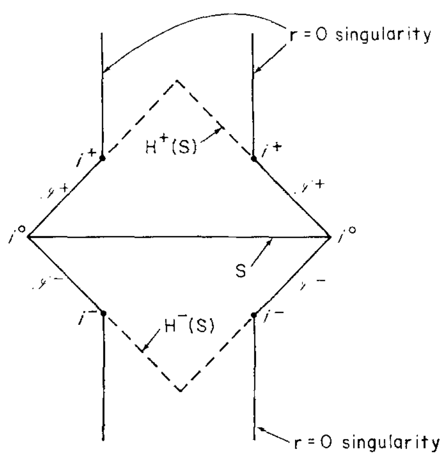

FIG. 8. The Reissner–Nordström solution, a space-time in which $H(S)$ begins at infinity.

Kerr solution. But the Cauchy horizon in the Reissner–Nordström solution is thought to be unstable [HE, p. 161; Simpson and Penrose [27]). In fact, the generic condition is not satisfied on any null geodesic $\gamma(s)$ such that a segment of $\gamma(s)$ is a generator of $H(S)$. Thus what the theorems and propositions of this section probably show is that $H(S)$ is generated at least in part by incomplete null geodesics, whether or not causality violation occurs. (Unless S is a "bad" partial Cauchy surface in empty space, as discussed in Ref. [10].)

There is another singularity theorem, due to Hawking, which does not need a causality assumption.

HAWKING'S THEOREM [HE, p. 272]. *Spacetime is not timelike geodesically complete if*

(1) $R_{ab}K^aK^b \geqslant 0$ *for every non-spacelike vector* K^a;

(2) *there exists a compact spacelike three-surface* S (*without edge*);

(3) *the unit normals to* S *are everywhere converging* (*or everywhere diverging*) *on* S.

The problem with this theorem is that condition (3) is much too strong; the space-time must be contracting (or expanding) *everywhere*. It is possible to generalize this theorem to arbitrary initial data in a closed universe in the following sense. Suppose a closed universe contracts (not necessarily everywhere), causality violation occurs in the regions of high density, and then the universe reexpands. We can show that, provided the causality violation begins in a finite region, the "bounce" must be accompanied by singularities.

THEOREM 10. *Let B be a four-dimensional region of a space-time such that the boundary of B is the disjoint union of two compact partial Cauchy surfaces, S and S'. Suppose that*

(1) $R_{ab}K^aK^b \geqslant 0$ *for every null vector* K^a;

(2) *the generic condition holds*;

(3) $d(S, [\operatorname{int} \tilde{D}^+(S)] \cap B)$ *is finite*;

(4) *the causality condition is violated in* B.

Then the space-time is timelike geodesically incomplete.

(The proof is similar to that of Theorem 5, and so will be omitted.) ∎

It was tacitly assumed above that if causality violation arises from regular initial data, then it is possible to find a partial Cauchy surface S whose Cauchy horizon lies on the boundary of a chronology-violating set V, at least in the region where the causality violation begins. That is, it was assumed that $\dot{I}^+(V) \cap H^+(S)$ was not empty and contained a geodesic segment of nonzero affine parameter length. However, finding such a partial Cauchy surface might be difficult. Therefore, I shall restate here some of the theorems, propositions, and definitions given above in a form which does not depend on the existence of $H(S)$, but only on the structure of $\dot{I}^+(V)$. The physical justifications for the theorems propositions, etc., are the same as those for their analogs given above. For example, we would expect matter to be present in the region where causality violation begins. This suggests the following analog to "asymptotically deterministic"

DEFINITION. A space-time (M, g) is said to be *asymptotically causal* if there is no point p for which $I^+(p) = M$ and each point q such that $I^+(q) \cap I^-(q) \neq \varnothing$, we have

$$\lim_{s \to a} [\inf T_{ab}K^aK^b] > 0 \tag{7.6}$$

on at least one of the null geodesic generators $\gamma(s)$ of $\dot{I}^+(q)$, where a is the past limit of the affine parameter along γ. We then have

THEOREM 11. *Suppose that a space-time (M, g) has a point q such that $I^+(q) \cap I^-(q) \neq \varnothing$, but $I^+(q) \neq M$. Then (M, g) is not null geodesically complete, provided*

(1) *the Einstein equations hold*;

(2) *the weak energy condition holds*;

(3) *at least one of the following holds*:

(a) *(M, g) is asymptotically causal*,

(b) *the matter tensor is past stochastic along at least one of the generators of $\dot{I}^+(q)$*.

Proof. Since $I^+(q) \neq M$, $\dot{I}^+(q)$ is nonempty. Further, $\dot{I}^+(q)$ is achronal and is generated by null geodesic segments which have no past endpoints since $q \in I^+(q) \cap I^-(q) = \text{int}[I^+(q) \cap I^-(q)]$. If condition (3a) holds, then along at least one of these segments, $\gamma(s)$, the matter tensor must satisfy (7.6). If $\gamma(s)$ were past complete, then by (1) and (2) we would have along $\gamma(s)$

$$\left| \lim_{s \to a = -\infty} \int^a H(s)\, ds \right| = +\infty \tag{7.7}$$

where $H(s) = \frac{1}{2}(R_{ab}K^aK^b + 2\sigma^2)$. By Lemma 9, conditions (1), (2), and Eq. (2.3), $\gamma(s)$ must have an infinite number of conjugate points since zeros of (2.3) correspond to conjugate points of $\gamma(s)$. But this is impossible because of Lemma 11 and the achronality of $\dot{I}^+(q)$. Thus $\gamma(s)$ must be past incomplete.

If condition (3b) holds, then Eq. (7.7) still holds, and the argument proceeds as before. ∎

Theorem 9, the Generalized Hawking–Penrose Theorem, has an obvious analog. Conditions (4) and (4′) of this theorem are respectively replaced by conditions (5) and (5′).

(5) If (M, g) does not satisfy the chronology condition, then (M, g) is asymptotically causal and the Einstein equations hold.

(5′) If (M, g) has a point q for which $I^+(q) \cap I^-(q)$ is nonempty, then $I^+(q) \neq M$, and there is at least one generator γ of $\dot{I}^+(q)$ along which the matter tensor is past stochastic as γ approaches its past affine parameter limit. Further, the Einstein equations hold.

The proof of this analog to Theorem 9 is omitted since it is essentially the same as the proof of Theorem 9. ∎

PROPOSITION 6. (analogous to Proposition 4). *Suppose there is a $q \in M$ such that $I^+(q) \cap I^-(q)$ is nonempty and $I^-(q) \neq M$. Then the matter tensor must be future vanishing along every furture complete generator of $\dot{I}^-(q)$, provided the weak energy condition and the Einstein equations hold. (Note that no generator of $\dot{I}^-(q)$ has a future endpoint.)* ∎

The analog to Proposition 5 is obtained by replacing "a partial Cauchy surface S" with "a point $q \in M$ such that $I^+(q) \cap I^-(q) \neq \varnothing$ and $I^-(q) \neq M$," and "$H^-(s)$" with "$\dot{I}^-(q)$."

8. Conclusion

The main purpose of this paper is to answer the question: "Is it possible to construct a time machine?" If by this we mean: "Is it possible to evolve CTL from regular initial data everywhere using known materials?," then the answer is almost certainly

No!

The import of the singularity theorems proved herein is this: CTL cannot in general arise in finite regions from regular initial data without *some* matter first passing through such extreme conditions that we cannot trust our knowledge of material behavior. The singularity theorems can be believed unless the weak energy condition is violated or the manifold picture of space-time breaks down. Most proposals [28] for the former would first require the matter to reach a density of 10^{54} g/cm^3, and the latter is not expected to occur until the matter density becomes 10^{94} g/cm^3. These numbers are respectively 40 and 80 orders of magnitude *above* the most extreme matter densities with which we are familiar (nuclear densities, and it's debatable how familiar we are with nuclear matter). Clearly, matter at these densities must be considered "unknown material."

Nevertheless, this result does not preclude the existence of CTL which arise from regular initial data; it is quite possible that the singularity on $H^+(S)$ is restricted to a very small region of space-time, with most of the matter forming the time machine avoiding the singularity. Indeed, a singularity is the future end of the event horizon when a black hole evaporates [5,29], and most of the matter in the universe avoids this singularity.

Furthermore, Hawking has argued [29] that what comes out of the singularity is completely random; he contends that, assuming global causality, the singularity emits with equal probability every configuration of particles compatible with such external constrants as energy and angular momentum conservation. However, it seems to me that if what comes out of a singularity is to be truly completely random, we should allow for the possibility that CTL could "come out of" a singularity. That is, we should extend Hawking's "Randomicity Principle" beyond the limited domain of particle emission and say that vitually any metric compatible with various external constraints such as those listed above could arise from a singularity. Among these metrics will be metrics containing CTL. Thus it is possible that CTL occur in the regions of space-time containing black hole explosions. But conversely, CTL, if they *do* occur, almost certainly must be associated with singularities, those "points" at which our knowledge of physics breaks down.

"The demonstration that no possible combination of known substances, known forms of machinery, and known forms of force can be united in a practicable machine by which men shall [travel back in time], seems to the writer as complete as it is possible for the demonstration of any physical fact to be."[1]

[1] The sentence in quotes is a slight modification of the conclusion of Newcomb's classic paper [30] proving the impossibility of heavier-than-air flying machines.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis advisor, Professor Dieter R. Brill, for his extremely helpful comments and criticisms. I would also like to thank Professor Robert H. Gowdy, and J. Isenberg, L. Lindblom, J. Nester, and P. Yasskin for helpful discussions.

REFERENCES

1. B. CARTER, *Phys. Rev.* **174** (1968), 1559.
2. G. W. GIBBONS AND R. A. RUSSELL-CLARK, *Phys. Rev. Lett.* **30** (1973), 398.
3. F. J. TIPLER, *Phys. Rev. D* **9** (1974), 2203.
4. C. J. S. CLARKE, *Comm. Math. Phys.* **41** (1975), 65. A cautionary remark is in order. Clark shows only that some component of the curvature tensor diverges in a frame parallel-propagated along any incomplete *timelike* curve contained in int $\tilde{D}(S)$. His result is not strictly applicable to most of the theorems of the present paper, which generally show that some of the *null* generators of $H^+(S)$ are incomplete. However, I would expect p.p. curvature singularities to develop in this case also.
5. B. S. DEWITT, *Phys. Rep. C* **19** (1975), 295.
6. R. GEROCH, Spacetime structure from a global viewpoint, *in* "General Relativity and Cosmology, Proc. Int. School in Phys. 'Enrico Fermi'" (R. K. Sachs, Ed.), pp. 71–103, Academic Press, New York, 1971.
7. R. PENROSE, "Techniques of Differential Topology in Relativity," Vol. 7 of the Regional Conference Series in Applied Math., Society of International Applied Mathematics, Philadelphia, 1972.
8. R. PENROSE, The structure of spacetime, *in* "Battelle Recontres" (C. M. DeWitt and J. A. Wheeler, Eds.), Benjamin, New York, 1968.
9. S. W. HAWKING AND G. F. R. ELLIS, "The Large Scale Structure of Space-time," Cambridge Univ. Press, Cambridge, 1973.
10. F. J. TIPLER, *Phys. Rev. Lett.* **37** (1976), 879; Ph.D. thesis, University of Maryland, 1976 (available from University Microfilms).
11. R. GEROCH, *J. Math. Phys.* **8** (1967), 782.
12. R. GEROCH, Singularities, *in* "Relativity" (S. Fickler, M. Carmeli, and L. Witten, Eds.), p. 265, Plenum, New York, 1970.
13. I. D. NOVIKOV AND YA. B. ZEL'DOVICH, *Ann. Rev. Astron. Astrophys.* **11** (1973), 387.
14. C. W. MISNER, K. S. THORNE, AND J. A. WHEELER, "Gravitation," Freeman, San Francisco, 1973.
15. P. HARTMAN, "Ordinary Differential Equations," Wiley, New York, 1964.
16. A. WINTER, *Quart. J. Appl. Math.* **7** (1949), 115.
17. H. EPSTEIN, V. GLASER, AND J. JAFFE, *Nuovo Cimento* **36** (1965), 1016.
18. J. J. SAKURAI, "Advanced Quantum Mechanics," Addison–Wesley, Reading, Mass., 1967.
19. S. W. HAWKING, *Phys. Rev. Lett.* **17** (1966), 444.
20. An asymptotically flat space-time is said to be *partially asymptotically predictable* from a partial Cauchy surface S if $\overline{\tilde{D}^+(S)} \cap \lambda \neq \varnothing$ and $\overline{\tilde{D}^-(s)} \cap \gamma \neq \varnothing$ for all generators λ of $\mathscr{I}^+$ and all generators γ of $\mathscr{I}^-$. A more detailed discussion of the meaning of this definition can be found in Ref. [10].
21. B. CARTER, *Phys. Rev.* **174** (1968), 1559; *J. Gen. Rev. Grav.* **1** (1971), 349.
22. R. GEROCH, Ph.D. Thesis, Princeton University, Princeton, 1967.
23. W. KUNDT, *Comm. Math. Phys.* **4** (1967), 143; D. R. BRILL, Thoughts on topology change, *in* "Magic Without Magic: John Archibald Wheeler" (J. R. Klauder, Ed.), Freeman, San Francisco, 1972; R. H. GOWDY, *Ann. Physics* **83** (1974), 203.
24. D. R. BRILL, A simple derivation of the general redshift formula, *in* "Methods of Local and Global Differential Geometry in General Relativity" (D. Farnsworth, J. Fink, J. Porter, and A. Thompson, Eds.), Springer–Verlag, Berlin, 1972.

25. G. F. R. Ellis and A. R. King, *Comm. Math. Phys.* **38** (1974), 119.
26. F. B. Hildebrand, "Advanced Calculus for Applications," Prentice-Hall, Englewood Clifts, N. J., 1965.
27. M. Simpson and R. Penrose, *Int. J. Theoret. Phys.* **7** (1973), 183.
28. F. W. Hehl, P. von der Heyde, and G. D. Kerlick, *Phys. Rev. D* **10** (1974), 1066.
29. S. W. Hawking, *Phys. Rev. D* **14** (1976), 2460.
30. S. Newcomb, "Side-lights on Astronomy: Essays and Addresses by Simon Newcomb," p. 345, Harper, New York, 1906.