Topological Censorship

John L. Friedman

Institute for Theoretical Physics, University of California, Santa Barbara, California 93106
and Department of Physics, University of Wisconsin, Milwaukee, Wisconsin 53201

Kristin Schleich and Donald M. Witt

Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

All three-manifolds are known to occur as Cauchy surfaces of asymptotically flat vacuum spacetimes and of spacetimes with positive-energy sources. We prove here the conjecture that general relativity does not allow an observer to probe the topology of spacetime: Any topological structure collapses too quickly to allow light to traverse it. More precisely, in a globally hyperbolic, asymptotically flat spacetime satisfying the null energy condition, every causal curve from $\mathcal{J}^-$ to $\mathcal{J}^+$ is homotopic to a topologically trivial curve from $\mathcal{J}^-$ to $\mathcal{J}^+$.

Because every three-manifold occurs as the spatial topology of a solution to the Einstein equations [1], one might ask why such topological structures are not part of our ordinary experience. A key part of the answer is a singularity theorem due to Gannon [2–4], showing that any asymptotically flat spacetime with a nonsimply connected Cauchy surface has singular time evolution if it satisfies the weak energy condition. Only topological structures comparable in size to the visible Universe or small enough that quantum effects play a crucial role in their dynamics can survive from the big bang to the present.

According to the cosmic censorship conjecture [5] singularities forming to the future of a regular initial data surface are hidden by an event horizon. If correct, the conjecture suggests that any topological structures will ultimately collapse within the horizon of a set of black holes. This collapse is too rapid to allow observers to traverse the wormhole throat for known exact analytically extended black hole solutions. Consequently one is led to a related *topological* censorship conjecture [6]—that no observer remaining outside a black hole has time to probe the topology of spacetime.

A precise formulation of topological censorship requires some standard definitions related to causal structure [7]. A spacetime M is *globally hyperbolic* if it has a Cauchy surface Σ; that is, M is the domain of dependence $D(\Sigma)$ of a spacelike hypersurface, Σ. A spacetime (M, g_{ab}) is *asymptotically flat* [8] if the following conditions hold: (i) There is a conformal completion $(\tilde{M}, \tilde{g}_{ab})$ where $\tilde{M}$ is compact with $\tilde{g}_{ab} = \Omega^2 g_{ab}$ where Ω is vanishing on $\mathcal{J}$ but has null gradient which is nonvanishing. (ii) The boundary, $\mathcal{J} = \tilde{M} - M$ is a disjoint union of past and future parts, $\mathcal{J}^+ \cup \mathcal{J}^-$, each having topology $S^2 \times \mathbb{R}$ with the $\mathbb{R}$'s complete null generators. (Topological censorship, however, does not require completeness in a future direction.) In this definition, $\mathcal{J}^+$ and $\mathcal{J}^-$ are future and past null infinity, respectively. A *causal curve* is any curve which is nonspacelike. The causal future, $J^+(S)$,

of a set S is the union of S with all points in M that lie on a future-directed causal curve originating in S. The causal past, $J^-(S)$, is defined as above with past substituted for future. The frontier or point set boundary of a set $A \subset X$, with respect to the set X is given by $\dot{A} = \bar{A} \cap \overline{X - A}$, where $\bar{A} = \text{int}(A) \cup \dot{A}$. Unless otherwise stated, the frontier will be defined with respect to $\tilde{M}$.

The null energy condition is the requirement that $T_{ab}k^a k^b \geq 0$, for all null vectors k^a. It is implied by each of the other common positive energy conditions: the weak energy condition, the strong energy condition, and the dominant energy condition [9]. The null energy condition implies that the convergence of a congruence of null geodesics cannot decrease and thus that initially converging null geodesics have conjugate points. A weaker form of energy conditions are averaged energy conditions; their use for proving the existence of conjugate points is due to Tipler [10]. In particular, Borde proved the *averaged* null energy condition (ANEC) also ensures this property along with even weaker averages [11].

A spacetime satisfies ANEC if the integral of $T_{ab}k^a k^b$ is non-negative along every inextendible null geodesic with affine parameter λ and corresponding tangent k^a: $\int d\lambda\, T_{ab}k^a k^b \geq 0$. Finally, we denote by γ_0 a timelike curve with past end point in $\mathcal{J}^-$ and future end point in $\mathcal{J}^+$ that lies in a simply connected neighborhood U of $\mathcal{J}$.

We can now state and prove the topological censorship theorem.

Theorem 1.—If an asymptotically flat, globally hyperbolic spacetime (M, g_{ab}) satisfies the averaged null energy condition, then every causal curve from $\mathcal{J}^-$ to $\mathcal{J}^+$ is deformable to γ_0 rel $\mathcal{J}$.

The proof is given below after Lemmas 1 and 2. It is similar to the argument used by Morris, Thorne, and Yurtsever (following a suggestion of Don Page) [12] to show that three-dimensional wormholes are not traversable and to the proof of Gannon's singularity theorem.

The proof of our theorem is by contradiction. We assume there is a nondeformable causal curve γ from $\mathcal{J}^-$ to

$\mathcal{J}^+$. This curve will unwrap in the universal covering space of M to yield a curve which connects two different asymptotic regions in the universal cover. One then shows that the unwrapped curve must pass through a trapped surface. Finally, one shows that if the spactime is globally hyperbolic, curves which pass through trapped surfaces cannot be observed. However, this is in contradiction with the assumption that γ was a causal curve. Thus the existence of such a curve is in contradiction and consequently the topology cannot be actively probed.

It may be helpful to illustrate this by considering the simple example in Fig. 1 of an RP^3 geon. Choose the $t = 0$ slice of Schwarzschild and, instead of extending it across $r = 2M$, identify antipodal points at the throat. The toplogy of this spatial slice is $RP^3 - pt$. The maximal evolution of this slice is a spacetime with spatial topology $RP^3 - pt$. Its universal covering space is the maximally extended Schwarzschild spacetime. Any nondeformable causal curve unwraps to a curve which connects the two disconnected asymptotic regions in the covering space and must pass through a trapped surface.

Since asymptotically flat spacetimes can have multiple disconnected asymptotic regions, $\mathcal{J}$ in general will have disconnected components. Let $\mathcal{J}_a$ be one such component, and let $\bar{M}_a = M \cup \mathcal{J}_a$ be a partial conformal completion, with $\Omega = 1$ outside an open neighborhood of $\mathcal{J}_a$ which intersects no other component of $\mathcal{J}$. The proof uses Prop. 9.2.8 of [7], slightly modified as Lemma 2, below, which in turn relies on Lemma 1.

Lemma 1. — Let $(\mathcal{M}, g_{ab})$ be any asymptotically flat spacetime with a simply connected Cauchy surface Σ. Let T be a smooth closed compact orientable two-surface in Σ. Then no inner directed null geodesic from T with respect to $\mathcal{J}_a$ is part of $\dot{J}^+(T)$.

Proof of Lemma 1. — Let t be a time function (see [7], p. 319) for which Σ is a surface of constant t, and let
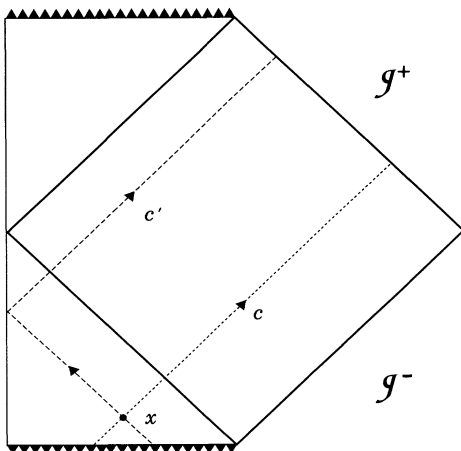

FIG. 1. The Penrose diagram for an RP^3 geon. Each point in the diagram is a two-sphere except for the left vertical boundary, whose points are RP^2's.

$T(t)$ be the orbit of T under diffeos generated by $\nabla^a t$. As Σ is simply connected, the timelike surface $T(t)$ separates the spacetime into disjoint parts, interior and exterior to the surface. Each inner-directed null geodesic γ from T that meets $\mathcal{J}^+$ must first intersect $T(t)$ for $t > 0$ at a point p. But p is in the timelike future of T: $p \in I^+(T)$. Thus p cannot lie on $\dot{J}^+(T)$, and γ cannot be a generator of $\dot{J}(T)$. Q.E.D.

A *congruence* of null geodesics in an open set of a spacetime is a family of null geodesics such that for each point in the open set there passes precisely one null geodesic in this family. Of particular relevance are congruences of null geodesics that are hypersurface orthogonal to a two-surface T. The expansion of such a null congruence is given by $\theta = s^{ab} \nabla_a k_b$ where s_{ab} is the metric of the two-surface and k^a is the tangent vector field to the congruence. A compact orientable two-surface T is *strongly outer trapped* if on the two-surface, $\theta < 0$ for an outer-directed hypersurface-orthogonal null congruence. The next lemma is valid even with vanishing expansion but the proof is simpler if we assume that the expansion is strictly positive.

Lemma 2. — Let $(\mathcal{M}, g_{ab})$ be an asymptotically flat spacetime that satisfies ANEC and has a simply connected Cauchy surface Σ. Then no surface T, strongly outer trapped with respect to $\mathcal{J}_a$, intersects $J^-(\mathcal{J}_a^+)$.

Proof of Lemma 2. — The proof uses essentially the same techniques as of Props. 9.2.1 and 9.2.8 in Hawking and Ellis [7]. If T intersects $J^-(\mathcal{J}_a^+)$, there is a causal curve connecting T to $\mathcal{J}_a^+$. Hence $J^+(T)$ intersects $\mathcal{J}_a^+$. By definition, $\mathcal{J}_a^+$ is closed, and since the spacetime is globally hyperbolic, $J^+(T)$ is closed. Hence $\mathcal{J}_a^+ \cap J^+(T)$ is closed. If $[\mathcal{J}_a^+ \cap J^+(T)]^\bullet$ is empty, then $\mathcal{J}_a^+ \cap J^+(T)$ is also open as a subset of $\mathcal{J}_a^+$, because $\mathcal{J}_a^+ \cap J^+(T) = \text{int}[\mathcal{J}_a^+ \cap J^+(T)] \cup [\mathcal{J}_a^+ \cap J^+(T)]^\bullet$. But if $\mathcal{J}_a^+ \cap J^+(T)$ is both closed and open, then $\mathcal{J}_a^+$ is disconnected. However, by definition $\mathcal{J}_a^+$ is connected. Therefore $[\mathcal{J}_a^+ \cap J^+(T)]^\bullet$ is not empty and it follows that $\mathcal{J}_a^+ \cap J^+(T)$ is also nonempty; $J^+(T)$ intersects $\mathcal{J}_a^+$. A past directed null generator of $J^+(T)$ with future end point at a point of $\mathcal{J}_a^+$ must have past end point at T and can contain no conjugate point. By Lemma 1, the geodesic must be outer directed from T. But ANEC and $\theta < 0$ imply that every outward null geodesic from the outer trapped surface T has a conjugate point within finite affine parameter length. This is a contradiction, because the generators of $J^+(T)$ have infinite affine parameter length. Q.E.D.

Proof of Theorem 1. — Consider the universal covering space [13] π: $\mathcal{M} \rightarrow M$ and the corresponding spacetime $(\mathcal{M}, g_{ab})$, with g_{ab} the pullback of g_{ab} to $\mathcal{M}$ by π. By construction $\mathcal{M}$ is simply connected, and any point in M has a simply connected neighborhood A whose inverse image $\pi^{-1}(A)$ is the disjoint union of open simply connected sets in $\mathcal{M}$. Each of these copies of A in $\mathcal{M}$ corresponds to a homotopically distinct way of reaching A from a fidu-

cial point of M, and we can choose the fiducial point to lie on $\mathcal{J}^+$. The projection π, restricted to any single copy of A, is an isometry.

Since the open neighborhood U of $\mathcal{J}(M)$ is chosen to be simply connected and M itself is not simply connected, U will be covered by multiple copies of itself in $\mathcal{M}$, which will therefore have multiple asymptotic regions. Let $\mathcal{U}_0 \subset \mathcal{M}$ be one of these copies, an open connected neighborhood of a single asymptotic region of $\mathcal{M}$. Construct a partial conformal completion $(\mathcal{M}_0, \tilde{g}_{ab})$ by adjoining a single copy of $\mathcal{J}(M)$ to $\mathcal{U}_0$. Then $(\mathcal{M}, g_{ab})$, with one asymptotic region singled out, satisfies the requirements of the Lemmas.

Suppose the theorem is false. Then there is a causal curve γ in M, from $\mathcal{J}^-(M)$ to $\mathcal{J}^+(M)$, which is not deformable to γ_0 relative to $\mathcal{J}(M)$. The curves γ_0 and γ can be lifted to curves Γ_0 and Γ in $\mathcal{M}$ that meet the same point of $\mathcal{J}_0^+$. Because the construction of $\mathcal{M}$ assigns distinct points to homotopically different ways of reaching the same point of M, the curves Γ_0 and Γ will join $\mathcal{J}_0^+(\mathcal{M})$ to different copies of the asymptotic region $\mathcal{J}^-(M)$. Because γ_0 lies in the simply connected neighborhood U of $\mathcal{J}(M)$, Γ_0 will lie in the neighborhood $\mathcal{U}_0$ of $\mathcal{J}_0(\mathcal{M})$, while Γ will join $\mathcal{J}_0^+(\mathcal{M})$ to another copy of U. In this second asymptotic region, it will intersect spheres of arbitrarily large radii.

These large spheres appear outer trapped as seen from the first asymptotic region, $\mathcal{U}_0$: Let $\bar{\Sigma}$ be the covering space of a Cauchy surface Σ of M and let $\mathcal{S}$ be a sphere in an asymptotic region of $\bar{\Sigma}$ different from the one containing $\mathcal{U}_0$. If we define outer-directed curves from any sphere $\mathcal{S}'$ to be those that reach $\mathcal{U}_0$ without intersecting $\mathcal{S}'$ a second time, then the outer directed curves from $\mathcal{S}'$ are curves from its concave surface—curves that would ordinarily be called inner directed by an observer in the asymptotic region near $\mathcal{S}'$. Since the spacetime is asymptotically flat, one can always pick $\mathcal{S}'$ so that the outer directed null congruence has $\theta < 0$. As Γ is causal, this implies that there are strongly outer trapped surfaces that intersect $J^-(\mathcal{J}_0^+(\mathcal{M}))$. But this contradicts Lemma 2. Hence any causal curve γ from $\mathcal{J}^-$ to $\mathcal{J}^+$ must be deformable to γ_0. Q.E.D.

An alternate proof to the more standard proof given above can be obtained using recent techniques developed by Penrose, Sorkin, and Woolgar [14] to prove a spacetime version of the positive energy theorem. They construct a partial ordering of all causal curves joining a given generator of $\mathcal{J}^-$ to a given generator of $\mathcal{J}^+$, calling a curve *faster* if it arrives earlier at $\mathcal{J}^+$ and leaves later from $\mathcal{J}^-$. A fastest curve is a null geodesic without conjugate points if the curve does not line on $\mathcal{J}$. The same argument, restricted to curves in a given homotopy class, shows the existence of a null geodesic without conjugate points lying in that homotopy class and joining $\mathcal{J}^-$ to $\mathcal{J}^+$ as needed to prove Theorem 1.

The consequences of Theorem 1 can be seen by consid-

ering a globally hyperbolic spacetime with non-Euclidean topology, assumed for simplicity to have one asymptotic region. Its universal cover will be a spacetime with multiple asymptotic regions. Suppose that an observer wishes to probe the topology of her spacetime and communicate the results of her measurements to a distant observer near $\mathcal{J}^+$. In order to detect a topological geon, her path or the path of her communication must traverse the geon and exit to $\mathcal{J}^+$; but this is forbidden by the theorem. Only observers and light rays that do not loop around a factor of the topology can communicate with $\mathcal{J}^+$, and such causal curves do not detect the existence of non-Euclidean topology. Thus general relativity prevents one from *actively* probing the topology of spacetime.

However, note that one can *passively* observe that topology by detecting light that originates at a past singularity. This is in keeping with the cosmic censorship conjecture which allows an observer to see a singularity in her past; similarly, the active topological censorship theorem proved above allows light rays to pass through a point x and then traverse homotopically distinct paths to a distant observer if the rays originate at a past singularity. The RP^3 geon of the identified Schwarzschild geometry provides an example. In Fig. 1, an observer O outside the black hole can passively detect the topology, receiving signals that traverse the homotopically distinct paths c and c' from a point x of the nonsimply connected Cauchy surface Σ. Followed back to the past, these null geodesics eventually hit the singularity at $r = 0$. In accordance with the theorem, to passively detect the topology one must see a signal that originates in a white hole rather than $\mathcal{J}^-$ (and hence at a singularity if cosmic censorship holds and if singularities are generically spacelike).

Even passive detection of spacetime topology is allowed only for a restricted set of topologies. A Cauchy surface of an asymptotically flat spacetime with a single asymptotic region is a connected sum of a set of prime factors. There are a countably infinite number of prime three-manifolds. By a natural way of counting, most are $K(\pi, 1)$'s, a class of manifolds that includes the hyperbolic spaces, the product of a circle and any closed two-manifold of nonzero genus, and the ten closed flat manifolds. A consequence, known to Schoen and Yau, of their proof of the positive-energy theorem [15] is that all $K(\pi, 1)$ factors are hidden by apparent horizons. Because apparent horizons are always within real horizons in asymptotically flat globally hyperbolic spacetimes, passive topological censorship is enforced: Topology in this class is invisible to observers outside black holes. Although the result not new, it is not explicitly stated in the literature.

Theorem 2.—Given any asymptotically flat initial data set (Σ, h_{ab}, p_{ab}) with sources which obey the dominant energy condition, all nontrivial topology due to a $K(\pi, 1)$ prime factor is surrounded by a two-sphere which is an apparent horizon.

Here h_{ab} and p_{ab} are the three-metric and extrinsic

curvature on Σ, and the definition of asymptotic flatness for initial data is that of [15]. A two-sphere $\mathcal{T} \subset \Sigma$ is an *apparent horizon* of a spacetime if the outgoing orthogonal null geodesics have zero convergence on $\mathcal{T}$. In terms of the initial data, the condition is $(k_{ab} + p_{ab})s^{ab} = 0$, where s_{ab} and k_{ab} are the two-metric and extrinsic curvature of $\mathcal{T}$ as a submanifold of Σ. Theorem 2 is a corollary of the following lemma, which summarizes part of the Schoen-Yau proof.

Lemma 3.—Let (Σ, h_{ab}, p_{ab}) be any asymptotically flat initial data set with sources which obey the dominant energy condition. Then Σ satisfies $\Sigma - \text{int}(K) \cong N - C$ where K is a compact submanifold of Σ, C is the interior of a disjoint union of three-balls, N is a three-manifold which admits an asymptotically flat metric with zero scalar curvature, and $\partial[\Sigma - \text{int}(K)]$ is a disjoint union of two-spheres which are apparent horizons.

To obtain Theorem 2 from Lemma 3, note that any three-manifold with an asymptotically flat metric of zero scalar curvature has a compactification admitting a metric with positive scalar curvature. Gromov and Lawson show that no such three-space can contain $K(\pi, 1)$'s as prime factors. Consequently, any $K(\pi, 1)$'s in Σ must be behind the apparent horizons.

Finally, note that the RP^3 geon is a counterexample to any hope that the Shoen-Yau result could be extended to all topologies. We believe that Theorem 2 is in fact sharp—that passive topological censorship hold for no other prime factors.

To prove active topological censorship, we used the averaged null energy condition, the weakest of the standard conditions. However, it is clear that any energy condition which implies that outward directed null congruences with $\theta < 0$ have conjugate points will suffice in our proof. Several such weaker conditions are known [11]. This suggests that the theorem may hold for semiclassical gravity, in which the source is the expectation value of a renormalized stress tensor. There are now several results which rely on conditions weaker than the averaged weak or averaged null energy conditions and which may be valid in a semiclassical theory: The chronology protection theorem of Hawking [16], the Penrose-Sorkin-Woolgar proof of a positive energy theorem [14], Gannon's theorem, and our active topological censorship theorem. Wald and Yurtsever [17] show that ANEC is violated by the renormalized stress tensor of free fields in generic spacetimes, and the question is whether there is a condition weak enough to be satisfied by the semiclassical stress tensor and strong enough to

enforce the theorems.

[1] D. M. Witt, Phys. Rev. Lett. **57**, 1386 (1986); (to be published).

[2] D. Gannon, J. Math. Phys. **16**, 2364 (1975). The averaged null energy conditions is enough to guarantee the null convergence condition and prove the relevant theorem (Corollary 1.2).

[3] C. W. Lee, Commun. Math. Phys. **51**, 157 (1976).

[4] G. J. Galloway, J. Phys. A **16**, 1435 (1983).

[5] R. Penrose, Riv. Nuovo Cimento **1**, 252 (1969).

[6] The conjecture, due to two of the authors (J.F. and D.W.) appears in J. L. Friedman, in *Conceptual Problems of Quantum Gravity*, edited by A. Ashtekar and J. Stachel (Bikhäuser, Boston, 1991).

[7] S. Hawking and G. Ellis, *The Large Scale Structure of Space-Time* (Cambridge Univ. Press, Cambridge, 1973).

[8] R. Geroch and G. T. Horowitz, in *General Relativity, an Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge Univ. Press, New York, 1979).

[9] The weak energy condition is $T_{ab} w^a w^b \geq 0$ for all timelike vectors w^a. The dominant energy condition is the weak condition with the additional requirement that $T^{ab} w_b$ be causal. The strong energy condition is $(T_{ab} - \frac{1}{2} T g_{ab}) \times w^a w^b \geq 0$ for timelike vectors w^a.

[10] F. J. Tipler, J. Diff. Geometry **30**, 165 (1978); Phys. Rev. D **17**, 2521 (1978).

[11] A. Borde, Classical Quantum Gravity **4**, 343 (1987).

[12] M. S. Morris, K. S. Thorne, and U. Yurtsever, Phys. Rev. Lett. **61**, 1446 (1988).

[13] The universal covering manifold $\mathcal{M}$ of a smooth path-connected manifold M can be constructed as follows. Pick a point $x_0 \in M$ and consider the set of smooth paths $P = \{c : [0,1] \to M \mid c(0) = x_0\}$. A projection map π: $P \to M$ is defined by $\pi(c(t)) = c(1)$. Let $\mathcal{M}$ be P modulo the equivalence relation, $c_1 \sim c_2$ if and only if $c_1(1) = c_2(1)$ and c_1 is homotopic to c_2 with end points fixed. The projection map π is then well defined and smooth as a map π: $\mathcal{M} \to M$.

[14] R. Penrose, R. D. Sorkin, and E. Woolgar, "A Positive Mass Theorem Based on the Focusing and Retardation of Null Geodesics," Syracuse report, 1993 (to be published).

[15] R. Schoen and S.-T. Yau, Commun. Math. Phys. **79**, 231 (1981).

[16] S. W. Hawking, Phys. Rev. D **46**, 603 (1992).

[17] R. M. Wald and U. Yurtsever, Phys. Rev. D **44**, 403 (1991).