Black holes: singularities, topology, and rigidity

Greg Galloway
University of Miami

Copenhagen Centre for Geometry and Topology,
University of Copenhagen, 16-20 June 2025

Topics

1. Elements of Lorentzian Geometry and Causality
2. The Geometry of Null Hypersurfaces
3. The Penrose Singularity Theorem and Related Results
4. The topology of black holes
5. The size of marginally outer trapped surfaces
6. Black hole topology and initial data rigidity

Elements of Lorentzian Geometry and Causality

Main references:

- B. O'Neill, Semi-Riemannian geometry, Pure and Applied Mathematics, vol. 103, Academic Press Inc., New York, 1983.
- G. J. Galloway, Notes on Lorentzian causality, ESI-EMS-IAMP Summer School on Mathematical Relativity (available at: `http://www.math.miami.edu/~galloway/`).

Further references:

- Beem, Ehrlich and Easley [9], Hawking and Ellis [38], Penrose [53], Wald [60]. See also Minguzzi's remarkably comprehensive treatment of causality theory [50].

A comment about prerequisites:

- Basic smooth manifold theory
- Basic Riemannian geometry

  - A Riemannian manifold is a smooth manifold M equipped with a Riemannian metric g.

  - g is a smooth assignment to each tangent space T_pM of positive definite inner product (symmetric, bilinear linear form) $g_p : T_pM \times T_pM \to \mathbb{R}$

A comment about prerequisites:

- Basic smooth manifold theory
- Basic Riemannian geometry

  - A Riemannian manifold is a smooth manifold M equipped with a Riemannian metric g.

  - g is a smooth assignment to each tangent space T_pM of positive definite inner product (symmetric, bilinear linear form) $g_p : T_pM \times T_pM \to \mathbb{R}$

Pseudo-Riemannian manifold:

- A pseudo-Riemannian manifold is a smooth manifold M equipped with a pseudo-Riemannian metric g.

  - g is a smooth assignment to each tangent space T_pM of a *nondegenerate* inner product (symmetric, bilinear linear form) $g_p : T_pM \times T_pM \to \mathbb{R}$

  - Each tangent space T_pM admits an orthonormal basis $\{e_1, e_2, \cdots, e_n\}$, $g(e_i, e_j) = 0$, $i \neq j$, $g(e_i, e_i) = \pm 1$ ([52].)

  - A pseudo-Riemannian manifold admits a Levi-Civita connection ∇, just as in the Riemannian case.

Lorentzian manifolds

In GR the space of events is represented by a Lorentzian manifold, i.e. smooth manifold M^{n+1} equipped with a metric g of Lorentzian signature. Thus, at each $p \in M$,

$$g : T_pM \times T_pM \to \mathbb{R}$$

is a scalar product of signature $(-,+,...,+)$. With respect to an orthonormal basis $\{e_0, e_1, ..., e_n\}$, as a matrix,

$$[g(e_i, e_j)] = \mathrm{diag}(-1, +1, ..., +1)\,.$$

Example: Minkowski space, the spacetime of Special Relativity. Minkowski space is $\mathbb{R}^{n+1}$, equipped with the Minkowski metric η: For vectors $X = X^i \frac{\partial}{\partial x^i}$, $Y = Y^i \frac{\partial}{\partial x^i}$ at p, (where x^i are standard Cartesian coordinates on $\mathbb{R}^{n+1}$),

$$\eta(X, Y) = \eta_{ij} X^i X^j = -X^0 Y^0 + \sum_{i=1}^{n} X^i Y^i\,.$$

Thus, each tangent space of a Lorentzian manifold is isometric to Minkowski space. This builds in the local accuracy of Special Relativity in General Relativity.

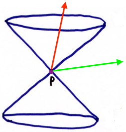
<u>*Causal character of vectors.*</u>

At each point, vectors fall into three classes, as follows:

$$X \text{ is } \begin{cases} \text{timelike} & \text{if } g(X,X) < 0 \\ \text{null} & \text{if } g(X,X) = 0 \\ \text{spacelike} & \text{if } g(X,X) > 0 \,. \end{cases}$$

A vector X is *causal* if it is either timelike or null.

The set of null vectors $X \in T_pM$ forms a double cone $\mathcal{V}_p$ in the tangent space T_pM:



called the null cone (or light cone) at p.
Timelike vectors point inside the null cone and spacelike vectors point outside.

<u>*Time orientability.*</u>

At each $p \in M$ we have a double cone; label one cone the *future* cone and the other a *past cone*.

If this assignment of a past and future cone can be made in a continuous manner over all of M then we say that M is *time-orientable*.

Time orientability (cont.).

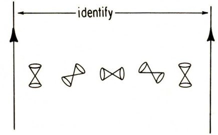
- There are various ways to make the phrase "continuous assignment" precise (see e.g., O'Neill p. 145), but they all result in the following:

  Fact: A Lorentzian manifold M^{n+1} is time-orientable iff it admits a smooth timelike vector field T.

- If M is time-orientable, the choice of a smooth timelike vector field T fixes a time orientation on M: A causal vector $X \in T_pM$ is *future pointing* if it points into the same half-cone as T, and *past pointing* otherwise.

  (Remark: If M is not time-orientable, it admits a double cover that is.)

- Example of a non-time-orientable spacetime:



By a spacetime we mean a connected time-oriented Lorentzian manifold (M^{n+1}, g).

Lorentzian inequalities. If X is causal, $g(X,X) \leq 0$, define its length as

$$|X| = \sqrt{-g(X,X)}\,.$$

Proposition 1.1

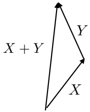
The following basic inequalities hold.

(1) *(Reverse Schwarz inequality) For all causal vectors $X, Y \in T_pM$,*

$$|g(X,Y)| \geq |X||Y|$$

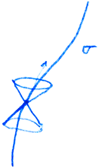
(2) *(Reverse triangle inequality) For all future directed causal vectors $X, Y \in T_pM$,*

$$|X+Y| \geq |X| + |Y|\,.$$



Proof. Exercise (see Gal-ESI, Prop 1.1).

The reverse triangle inequality (RTI) is the geometric origin of the so-called 'twin paradox'



Causal character of curves.

Let $\gamma : I \to M$, $t \to \gamma(t)$ be a smooth curve in M.

- γ is said to be timelike provided $\gamma'(t)$ is timelike for all $t \in I$.

  In GR, a timelike curve corresponds to the history (or *worldline*) of an observer.

- Null curves and spacelike curves are defined analogously.

  A causal curve is a curve whose tangent is either timelike or null ($\neq 0$) at each point.

- The length of a causal curve $\gamma : [a, b] \to M$, is defined by

$$L(\gamma) = \text{Length of } \gamma = \int_a^b |\gamma'(t)| dt = \int_a^b \sqrt{-\langle \gamma'(t), \gamma'(t) \rangle}\, dt\,.$$

  - Owing to the RTI, causal geodesics ($\nabla_{\gamma'}\gamma' = 0$) locally *maximize* length.

  - If γ is timelike one can introduce arc length parameter along γ. In general relativity, the arc length parameter along a timelike curve is called *proper time*, and corresponds to time kept by the observer.

Futures and Pasts

Let (M, g) be a spacetime. A timelike (resp. causal) curve $\gamma : I \to M$ is said to be *future directed* provided each tangent vector $\gamma'(t)$, $t \in I$, is future pointing. (*Past-directed* timelike and causal curves are defined in a time-dual manner.)

Causal theory is the study of the causal relations $\ll$ and $<$:

Definition 1.2

For $p, q \in M$,

1. *$p \ll q$ means there exists a future directed timelike curve in M from p to q (we say that q is in the timelike future of p),*
2. *$p < q$ means there exists a future directed causal curve in M from p to q (we say that q is in the causal future of p),*

We shall use the notation $p \leq q$ to mean $p = q$ or $p < q$.

The causal relations $\ll$ and $<$ are clearly transitive. Also, from variational considerations, it is heuristically clear that the following holds,

$$\text{if } p \ll q \text{ and } q < r \text{ then } p \ll r\,.$$

Proposition 1.3 (O'Neill, p. 294)

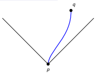
In a spacetime M, if q is in the causal future of p ($p < q$) but is not in the timelike future of p ($p \not\ll q$) then any future directed causal curve γ from p to q must be a null geodesic (when suitably parameterized).

Now introduce standard causal notation:

Definition 1.4

Given any point p in a spacetime M, the timelike future and causal future of p, denoted $I^+(p)$ and $J^+(p)$, respectively, are defined as,

$$I^+(p) = \{q \in M : p \ll q\} \quad \text{and} \quad J^+(p) = \{q \in M : p \leq q\}\,.$$

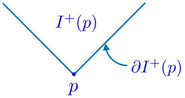


The timelike and causal *pasts* of p, $I^-(p)$ and $J^-(p)$, respectively, are defined in a time-dual manner in terms of past directed timelike and causal curves.

With respect to this notation, the above proposition becomes:

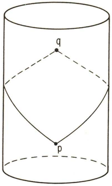
Proposition *If $q \in J^+(p) \setminus I^+(p)$ then any future directed causal curve from p to q is a null geodesic.*

Ex. Minkowski space. For p any point in Minkowski space, $I^+(p)$ consists of all points inside the future null cone, and $J^+(p)$ consists of all points on and inside the future null cone. $\partial I^+(p)$ is just the future null cone at p.



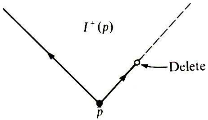
We note, however, that curvature and topology can drastically change the structure of 'null cones' in spacetime.

Ex. Consider the following example of a flat spacetime cylinder, closed in space.



For any point p, $\partial I^+(p)$, is compact and consists of the two future directed null geodesic segments emanating from p that meet to the future at a point q. By extending these geodesics beyond q we enter $I^+(p)$.

In general, sets of the form $I^+(p)$ are open (see e.g. Gal-ESI, p. 10). However, sets of the form $J^+(p)$ need not be closed, as can be seen by removing a point from Minkowski space.



For any subset $S \subset M$, we define the timelike and causal future of S, $I^+(S)$ and $J^+(S)$, respectively by

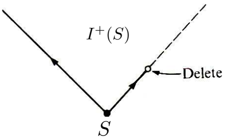
$$I^+(S) = \bigcup_{p \in S} I^+(p) = \{q \in M : p \ll q \text{ for some } p \in S\} \tag{1.1}$$

$$J^+(S) = \bigcup_{p \in S} J^+(p) = \{q \in M : p \leq q \text{ for some } p \in S\}\,. \tag{1.2}$$

Note:

- $S \subset J^+(S)$.
- $I^+(S)$ is open (union of open sets).

$I^-(S)$ and $J^-(S)$ are defined in a time-dual manner.



Although in general $J^+(S) \neq \overline{I^+(S)}$, the following relationships always hold between $I^+(S)$ and $J^+(S)$.

Proposition 1.5

For all subsets $S \subset M$,

(1) $\operatorname{int} J^+(S) = I^+(S)$,

(2) $J^+(S) \subset \overline{I^+(S)}$ *(and $= \overline{I^+(S)}$ if S is closed)*

Proof. Exercise.

As we'll discuss, there are certain conditions under which $J^+(S)$ is closed.

Achronal Boundaries

Achronal sets play an important role in causal theory.

Definition 1.6

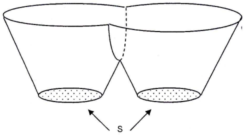
A subset $S \subset M$ is achronal provided no two of its points can be joined by a timelike curve.

Of particular importance are *achronal boundaries*.

Definition 1.7

An achronal boundary is a set of the form $\partial I^+(S)$ (or $\partial I^-(S)$), for some $S \subset M$.

The following figure illustrates some of the important structural properties of achronal boundaries.



Proposition 1.8

An achronal boundary $\partial I^+(S)$, if nonempty, is a closed achronal C^0 hypersurface in M.

We make some comments about the proof. The following fact is useful.

Fact: If $p \in \partial I^+(S)$ then $I^+(p) \subset I^+(S)$, and $I^-(p) \subset M \setminus \overline{I^+(S)}$

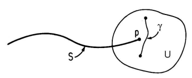
Proof: Exercise.

Claim A: An achronal boundary $\partial I^+(S)$ is achronal.

Proof. Suppose there exist $p, q \in \partial I^+(S)$, with $q \in I^+(p)$. By the above fact, $q \in I^+(S)$. But since $I^+(S)$ is open, $I^+(S) \cap \partial I^+(S) = \emptyset$, contradicting $q \in \partial I^+(S)$.

Definition 1.9 (edge points)

Let $S \subset M$ be achronal. Then $p \in \overline{S}$ is an edge point of S provided every neighborhood U of p contains a timelike curve γ from $I^-(p, U)$ to $I^+(p, U)$ that does not meet S.



We denote by $\mathrm{edge}\, S$ the set of edge points of S. Check that:

$$\overline{S} \setminus S \subset \mathrm{edge}\, S \subset \overline{S}$$

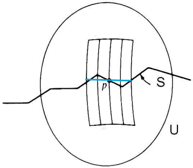
If $\mathrm{edge}\, S = \emptyset$ we say that S is *edgeless*.

Claim B: An achronal boundary is *edgeless*.

Proof of the claim. The above fact implies that for any $p \in \partial I^+(S)$, any timelike curve from $I^-(p)$ to $I^+(p)$ must meet $\partial I^+(S)$. It follows that $\partial I^+(S)$ is edgeless.

Claim C: An edgeless achronal set S, if nonempty, is a C^0 hypersurface in M.

Sketch of proof. Fix $p \in S$. Since p is not an edge point, there exists a neighborhood U of p such that every timelike curve from $I^-(p, U)$ to $I^+(p, U)$ meets S (exactly once by achronality). One can then use the integral curves of a timelike vector field to express S locally near p as a graph over a smooth hypersurface:

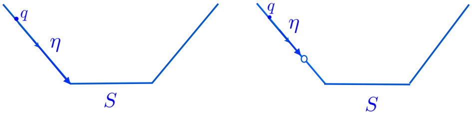


One can further use the achronality of S to show that the graphing function is continuous. See O'Neill [52, p. 413] for details.

We mention briefly the following result shows that, in general, large portions of achronal boundaries are ruled by null geodesics.

Proposition 1.10

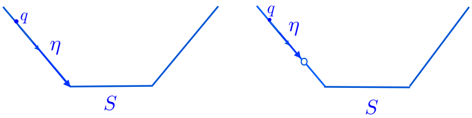
Let $S \subset M$ be closed. Then each $q \in \partial I^+(S) \setminus S$ lies on a null geodesic η contained in $\partial I^+(S)$, which either has a past end point on S, or else is past inextendible in M.



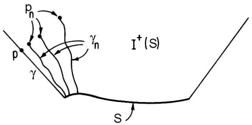
We mention briefly the following result shows that, in general, large portions of achronal boundaries are ruled by null geodesics.

Proposition 1.10

Let $S \subset M$ be closed. Then each $q \in \partial I^+(S) \setminus S$ lies on a null geodesic η contained in $\partial I^+(S)$, which either has a past end point on S, or else is past inextendible in M.



Proof. Read Gal-ESI, Prop. 3.4. The proof involves taking the limit of a certain sequence of timelike curves:



Taking the limit makes us of the *Limit Curve Lemma* (Lemma 3.5 in GG-ESI).

Causality conditions

A number of results in Lorentzian geometry and general relativity require some sort of causality condition.

Chronology condition: A spacetime M satisfies the *chronology condition* provided there are no closed timelike curves in M.

Compact spacetimes have limited interest in general relativity since they all violate the chronology condition.

Proposition 1.11

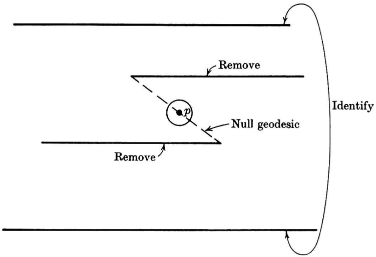
Every compact spacetime contains a closed timelike curve.

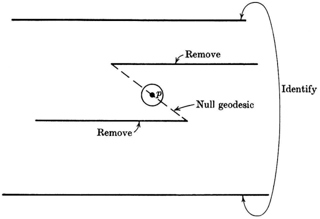
Proof: The sets $\{I^+(p); p \in M\}$ form an open cover of M from which we can abstract a finite subcover: $I^+(p_1)$, $I^+(p_2)$, ..., $I^+(p_k)$. We may assume that this is the minimal number of such sets covering M. Since these sets cover M, $p_1 \in I^+(p_i)$ for some i. It follows that $I^+(p_1) \subset I^+(p_i)$. Hence, if $i \neq 1$, we could reduce the number of sets in the cover. Thus, $p_1 \in I^+(p_1)$ which implies that there is a closed timelike curve through p_1. □

Causality condition: A spacetime M satisfies the causality condition provided there are no closed (nontrivial) causal curves in M.

Exercise: Construct a spacetime that satisfies the chronology condition but not the causality condition.

A spacetime that satisfies the causality condition can nonetheless be on the verge of failing it, in the sense that there exist causal curves that are "almost closed", as illustrated by the following figure.





Strong causality is a condition that rules out almost closed causal curves.

Definition 1.12

An open set U in spacetime M is said to be causally convex provided every causal curve segment with end points in U lies entirely within U.

Definition 1.13

Strong causality is said to hold at $p \in M$ provided p has arbitrarily small causally convex neighborhoods, i.e., for each neighborhood V of p there exists a causally convex neighborhood U of p such that $U \subset V$.

Note that strong causality fails at the point p in the figure above. It can be shown that the set of points at which strong causality holds is open.

Strong causality condition: A spacetime M is said to be strongly causal if strong causality holds at all of its points.

Although there are even stronger causality conditions, strong causality is sufficient for many applications.

Strong causality has the following useful consequence.

Lemma 1.14 (non-imprisonment)

Suppose strong causality holds at each point of a compact set K in a spacetime M. If $\gamma : [0, b) \to M$ is a future inextendible causal curve that starts in K then eventually it leaves K and does not return, i.e., there exists $t_0 \in [0, b)$ such that $\gamma(t) \notin K$ for all $t \in [t_0, b)$.

(γ is future inextendible if it cannot be continuously extended, i.e. if $\lim_{t \to b-} \gamma(t)$ does not exist.)

Proof. Exercise (or see O'Neill).

We say that a future inextendible causal curve cannot be "imprisoned" in a compact set on which strong causality holds.

Global hyperbolicity

We now come to a fundamental condition in spacetime geometry, that of global hyperbolicity.

Mathematically, global hyperbolicity is a basic 'niceness' condition that often plays a role analogous to geodesic completeness in Riemannian geometry. Physically, global hyperbolicity is closely connected to the issue of classical determinism and the strong cosmic censorship conjecture.

The following is the *classical* definition of global hyperbolicity.

Definition 1.15

A spacetime M is said to be globally hyperbolic provided

- *M is strongly causal.*
- *(Internal Compactness) The sets $J^+(p) \cap J^-(q)$ are compact for all $p, q \in M$.*

Remarks:

(1) Condition (2) says roughly that M has no holes or gaps.

(2) In fact, as shown by Bernal and Sanchez [12], internal compactness + causality imply strong causality. (Remarkably, under certain circumstances, causality assumptions can be removed altogether [42].)

- We consider a few basic consequences of global hyperbolicity.

Proposition 1.16

Let M be a globally hyperbolic spacetime. Then,

(1) *The sets $J^{\pm}(A)$ are closed, for all compact $A \subset M$.*

(2) *The sets $J^{+}(A) \cap J^{-}(B)$ are compact, for all compact $A, B \subset M$.*

Proof. First observe that $J^{\pm}(p)$ are *closed* for all $p \in M$: Suppose $q \in \overline{J^{+}(p)} \setminus J^{+}(p)$ for some $p \in M$. Choose $r \in I^{+}(q)$, and $\{q_n\} \subset J^{+}(p)$, with $q_n \to q$. Since $I^{-}(r)$ is an open neighborhood of q, $\{q_n\} \subset J^{-}(r)$ for n large. It follows that $q \in \overline{J^{+}(p) \cap J^{-}(r)} = J^{+}(p) \cap J^{-}(r)$, since $J^{+}(p) \cap J^{-}(r)$ is compact and hence closed. But this contradicts $q \notin J^{+}(p)$. Thus, $J^{+}(p)$ is closed, and similarly so is $J^{-}(p)$.

For the rest of the proof of the proposition, see e.g. [38, p. 207].

- Analogously to the case of Riemannian geometry, one can learn a lot about the structure of spacetime by studying its causal geodesics.

  Basic question: Given $q \in I^{+}(p)$ under what circumstances does there exist a maximal future directed timelike geodesic γ from p to q? *Maximal means:* $L(\gamma) \geq L(\sigma)$ for all future directed causal curves σ from p to q.

- Maximality can be expressed in terms of the Lorentzian *distance function*, $d : M \times M \to [0, \infty]$. For $p < q$, let $\Omega_{p,q}$ denote the collection of future directed causal curves from p to q. Then, for any $p, q \in M$, define

$$d(p,q) = \begin{cases} \sup\{L(\sigma) : \sigma \in \Omega_{p,q}\}, & \text{if } p < q \\ 0, & \text{if } p \not< q \end{cases}$$

  It can be shown that for globally hyperbolic spacetimes, d is continuous.

- Global hyperbolicity is the standard condition in Lorentzian geometry that ensures the existence of maximal timelike geodesic segments.

Theorem 1.17

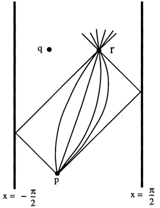
Let M be a globally hyperbolic spacetime. If $q \in I^+(p)$ then there is a maximal future directed causal geodesic from p to q (i.e., no causal curve from p to q can have greater length).

Brief comment on the proof. Consider a sequence of causal curves γ_n from p to $q \in J^+(p)$, such that $L(\gamma_n) \to d(p,q)$. Show that a subsequence converges to a causal curve γ with $L(\gamma) = d(p,q)$. (See Gal-ESI, Prop. 4.5, for further details.)

Remark: Contrary to the situation in Riemannian geometry, geodesic completeness does not guarantee the existence of maximal segments.

Ex. Two-dimensional anti-de Sitter space:

$$M = \{(t,x) : -\pi/2 < x < \pi/2\}, g = \sec^2 x(-dt^2 + dx^2)$$



All future directed timelike geodesics emanating from p refocus at r. The points p and q are timelike related, but there is no timelike geodesic segment from p to q.

Cauchy hypersurfaces

Global hyperbolicity is closely related to the existence of certain 'ideal initial value hypersurfaces', called Cauchy surfaces. There are slight variations in the literature in the definition of a Cauchy surface. Here we adopt the following definition.

Definition 1.18

A Cauchy surface for a spacetime M is an achronal subset S of M which is met by every inextendible causal curve in M.

Observations:

- If S is a Cauchy surface for M then $\partial I^+(S) = S$. Similarly, $\partial I^-(S) = S$ (Exercise.) It follows from Proposition 1.8 that a Cauchy surface S is a closed achronal C^0 hypersurface in M.
- If S is Cauchy then every inextendible timelike curve meets S exactly once.

Now, for a very classical result:

Theorem 1.19 (Geroch [37])

If a spacetime M is globally hyperbolic then it has a Cauchy surface S.

We make some comments about the proof. (As discussed later, the converse also holds.)

- Introduce a measure μ on M such that $\mu(M) = 1$, and consider the function $f : M \to (0, \infty)$ defined by
$$f(p) = \frac{\mu(J^-(p))}{\mu(J^+(p))} \,.$$
- Internal compactness is used to show that f is continuous, and strong causality is used to show that f is strictly increasing along future directed causal curves.
- One shows further that $f \to +\infty$ along every future inextendible causal curve and $f \to 0$ along every past inextendible causal curve.
- It follows that level sets of f, $\{f = t : t \in (0, \infty)\}$ are Cauchy surfaces for M. $\square$

Remark: The function f constructed in the proof is what is referred to as a *time function*, namely, a continuous function that is strictly increasing along future directed causal curves.

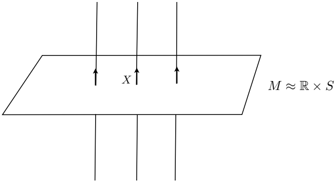
In fact it is possible to construct smooth time functions, i.e. smooth functions t with (past directed) timelike gradient ∇t. These are necessarily time functions; see Bernal and Sanchez [11] and Chruściel, Grant and Minguzzi [18].

Proposition 1.20

Let M be globally hyperbolic.

- *If S is a Cauchy surface for M then M is homeomorphic to $\mathbb{R} \times S$.*
- *Any two Cauchy surfaces in M are homeomorphic.*

Proof: To prove the first, one introduces a future directed timelike vector field X on M. Each integral curve of X meets S exactly once. These integral curves, suitably parameterized, provide the desired homeomorphism.



A similar technique may be used to prove the second. □

Remark: In view of Proposition 1.20, any nontrivial topology in a globally hyperbolic spacetime must reside in its Cauchy surfaces.

The following fact is often useful.

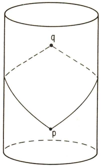
Proposition 1.21

If S is a compact achronal C^0 hypersurface in a globally hyperbolic spacetime M then S must be a Cauchy surface for M.

Comments on the proof:

- We have that $M = J^+(S) \cup J^-(S)$: $J^+(S) \cup J^-(S)$ is closed by Proposition 1.16, and is also easily shown to be open.
- Let γ be an inextendible causal curve. Suppose γ meets $J^+(S)$ at a point p. Then the portion of γ to the past of p must meet S, otherwise it is imprisoned in the compact set $J^-(p) \cap J^+(S)$, which would be a strong causality violation.

Ex. $S = \partial I^+(p)$ in the flat spacetime cylinder closed in space.



Domains of Dependence

Definition 1.22

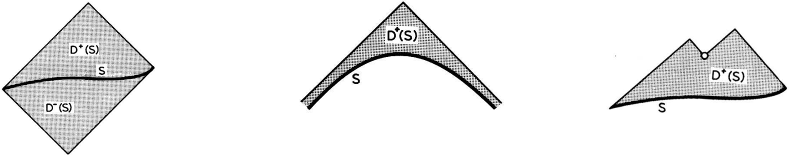
Let S be an achronal set in a spacetime M. The future domain of dependence of $D^+(S)$ of S is defined as follows,

$$D^+(S) = \{p \in M : \text{ every past inextendible causal curve from } p \text{ meets } S\}$$

In physical terms, since information travels along causal curves, a point in $D^+(S)$ only receives information from S. Thus if physical laws are suitably causal, initial data on S should determine the physics on $D^+(S)$.

The past domain of dependence of $D^-(S)$ is defined in a time-dual manner. The (total) domain of dependence of S is the union, $D(S) = D^+(S) \cup D^-(S)$.

Below we show a few examples of future and past domains of dependence.



Note: $S \subset D^+(S)$.

The following characterizes Cauchy surfaces in terms of domain of dependence.

Proposition 1.23

Let S be an achronal subset of a spacetime M. Then, S is a Cauchy surface for M if and only if $D(S) = M$.

Proof: Exercise.

The following basic result ties domains of dependence to global hyperbolicity.

Proposition 1.24

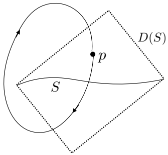
Let $S \subset M$ be achronal.

(1) *Strong causality holds on* $\operatorname{int} D(S)$.

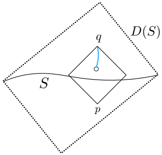
(2) *Internal compactness holds on* $\operatorname{int} D(S)$, *i.e., for all* $p, q \in \operatorname{int} D(S)$, $J^+(p) \cap J^-(q)$ *is compact.*

- See Gal-ESI, Prop. 5.5, for a discussion of the proof. A few heuristic remarks:

  (1) Strong causality: Suppose γ is a closed timelike curve through $p \in \mathrm{int}\, D(S)$.

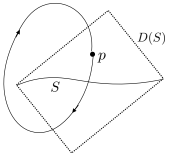


  (2): Internal compactness: A failure of internal compactness suggests the existence of a "hole" in $\mathrm{int}\, D(S)$:

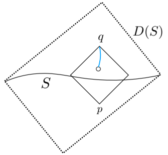


- See Gal-ESI, Prop. 5.5, for a discussion of the proof. A few heuristic remarks:

  (1) Strong causality: Suppose γ is a closed timelike curve through $p \in \operatorname{int} D(S)$.



  (2): Internal compactness: A failure of internal compactness suggests the existence of a "hole" in $\operatorname{int} D(S)$:



- From the fact that $\operatorname{int} D(S)$ is globally hyperbolic, we can now address the converse of Theorem 1.19.

Corollary 1.25

If S is a Cauchy surface for M then M is globally hyperbolic.

Proof: This follows immediately from Propositions 1.23 and 1.24: S Cauchy $\implies D(S) = M \implies \operatorname{int} D(S) = M \implies M$ is globally hyperbolic. $\square$

Thus we have that: M is globally hyperbolic if and only if M admits a Cauchy surface.

Cauchy horizons

We conclude this section with some comments about Cauchy horizons. If S is achronal, the *future Cauchy horizon* $H^+(S)$ of S is the future boundary of $D^+(S)$.

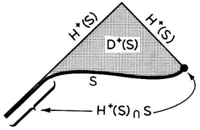
This is made precise in the following definition.

Definition 1.26

Let $S \subset M$ be achronal. The future Cauchy horizon $H^+(S)$ of S is defined as follows

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

The past Cauchy horizon $H^-(S)$ is defined time-dually. The (total) Cauchy horizon of S is defined as the union, $H(S) = H^+(S) \cup H^-(S)$.



We record some basic facts about domains of dependence and Cauchy horizons.

Proposition 1.27

Let S be an achronal subset of M. Then the following hold.

1. $H^+(S)$ *is achronal.*
2. $\partial D^+(S) = H^+(S) \cup S$.
3. $\partial D(S) = H(S)$.

Point 3 provides a useful mechanism for showing that an achronal set S is Cauchy: S is Cauchy iff $D(S) = M$ iff $\partial D(S) = \emptyset$ iff $H(S) = \emptyset$.

Cauchy horizons have structural properties similar to achronal boundaries, as indicated in the next two results.

Proposition 1.28

Let $S \subset M$ be achronal. Then $H^+(S) \setminus \mathrm{edge}\, H^+(S)$, if nonempty, is an achronal C^0 hypersurface in M.

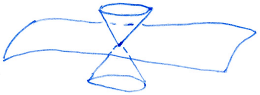
Proposition 1.29

Let S be an achronal subset of M. Then $H^+(S)$ is ruled by null geodesics, i.e., every point of $H^+(S) \setminus \mathrm{edge}\, S$ is the future endpoint of a null geodesic in $H^+(S)$ which is either past inextendible in M or else has a past end point on $\mathrm{edge}\, S$.

The Geometry of Null Hypersurfaces

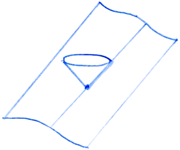
In addition to curves, one is interested in the geometry and causality of certain higher dimensional submanifolds.

- A *spacelike* hypersurface is a smooth hypersurface all of whose tangent vectors are spacelike (or, equivalently, whose normal vectors are timelike):



  In other words, a hypersurface is spacelike iff the induced metric is positive definite (i.e. Riemannian). In GR, a spacelike hypersurface represents space at a given instant of time.

- A null hypersurface is a smooth hypersurface such that the null cone is tangent to it at each of its



  Null hypersurfaces play an important role in GR as they represent horizons of various sorts. Null hypersurfaces have an interesting geometry which we discuss in this section.

Comments on Curvature and the Einstein Equations

- Let $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, $(X, Y) \to \nabla_X Y$, be the Levi-Civita connection with respect to the Lorentz metric g. ∇ is determined locally by the Christoffel symbols,

$$\nabla_{\partial_i}\partial_j = \sum_k \Gamma_{ij}^k \,\partial_k , \qquad \left(\partial_i = \frac{\partial}{\partial x^i}, etc.\right)$$

- *Geodesics* are curves $t \to \sigma(t)$ of zero covariant acceleration,

$$\nabla_{\sigma'(t)}\sigma'(t) = 0\,.$$

  Timelike geodesics correspond to *free falling* observers.
- The Riemann curvature tensor is defined by,

$$R(X, Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$$

  The components ${R^\ell}_{kij}$ are determined by,

$$R(\partial_i, \partial_j)\partial_k = \sum_\ell {R^\ell}_{kij}\partial_\ell$$

- The Ricci tensor Ric and scalar curvature R are obtained by taking traces,

$$R_{ij} = \sum_\ell {R^\ell}_{i\ell j} \quad \text{and} \quad R = \sum_{i,j} g^{ij} R_{ij}$$

- The Einstein equation, the field equations of GR, are given by:

$$\mathrm{Ric} - \frac{1}{2} R\, g = \kappa T\,,$$

  where T is the energy-momentum tensor.

- The *vacuum* Einstein equations are obtained by setting $T = 0$. This is equivalent to setting $\mathrm{Ric} = 0$. (I.e., vacuum iff Ricci flat).

  We will sometimes require that a spacetime satisfying the Einstein equations, obeys an *energy condition*.

- The *null energy condition* (NEC) is the requirement that

$$T(X,X) = \sum_{i,j} T_{ij} X^i X^j \geq 0 \quad \text{for all null vectors } X\,.$$

- The stronger *dominant energy condtion* (DEC) is the requirement,

$$T(X,Y) = \sum_{i,j} T_{ij} X^i Y^j \geq 0 \quad \text{for all future directed causal vectors } X, Y\,.$$

  (Note that these are actually curvature conditions.)

Null Hypersurfaces

Definition 2.1

A null hypersurface in a spacetime (M, g) is a smooth co-dimension one submanifold S of M, such that at each $p \in S$, $g : T_pS \times T_pS \to \mathbb{R}$ is degenerate.

This means that there exists a nonzero vector $K_p \in T_pS$ (the direction of degeneracy) such that

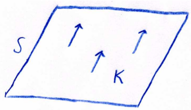
$$\langle K_p, X\rangle = 0 \quad \text{for all} \quad X \in T_pS \qquad\qquad (\langle\cdot,\cdot\rangle = g)$$

In particular,

- K_p is a null vector, $\langle K_p, K_p\rangle = 0$, which we can choose to be future pointing, and
- $[K_p]^\perp = T_pS$.
- Moreover, every vector $X \in T_pS$ that is not a multiple of K_p is spacelike.

Thus, every null hypersurface S gives rise to a smooth future directed null vector field K on S, unique up to a positive pointwise scale factor.

$$p \in S \xrightarrow{K} K_p \in T_pS,$$



Ex. $\mathbb{M}^{n+1}$ = Minkowski space.

- Null hyperplanes in $\mathbb{M}^{n+1}$: Each nonzero null vector $X \in T_p\mathbb{M}^{n+1}$ determines a null hyperplane $\Pi = \{q \in \mathbb{M}^{n+1} : \langle \overline{pq}, X \rangle = 0\}$.
- Null cones in $\mathbb{M}^{n+1}$: The past and future cones, $\partial I^-(p)$ and $\partial I^+(p)$, respectively, are smooth null hypersurfaces away from the vertex p.

The following fact is fundamental.

Proposition 2.2

Let S be a smooth null hypersurface and let K be a smooth future directed null vector field on S. Then the integral curves of K are null geodesics (when suitably parameterized),

Remark: The integral curves of K are called the *null generators* of S.

Proof: Suffices to show:

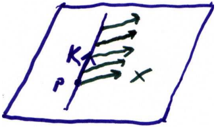
$$\nabla_K K = \lambda K$$

This follows by showing at each $p \in S$,

$$\nabla_K K \perp T_pS\,, \quad \text{i.e.,} \quad \langle \nabla_K K, X \rangle = 0 \quad \forall X \in T_pS$$

Extend $X \in T_pS$ by making it invariant under the flow generated by K,

$$[K, X] = \nabla_K X - \nabla_X K = 0$$



X remains tangent to S, so along the flow line through p,

$$\langle K, X\rangle = 0$$

Differentiating,

$$K\langle K, X\rangle = \langle \nabla_K K, X\rangle + \langle K, \nabla_K X\rangle = 0$$
$$\langle \nabla_K K, X\rangle = -\langle K, \nabla_X K\rangle = -\frac{1}{2} X\langle K, K\rangle = 0. \qquad \square$$

Remark: To study the 'shape' of the null hypersurface S we study how the null vector field K varies along S. Since K is actually orthogonal to S, this is somewhat analogous to how we study the shape of a hypersurface in a Riemannian manifold, or spacelike hypersurface in a Lorentzian manifold, by introducing the shape operator (or Weingarten map) and associated second fundamental form.

Null Weingarten Map/Null 2nd Fundamental Form

- We introduce the following equivalence relation on tangent vectors: For $X, Y \in T_pS$,
$$X = Y \bmod K \iff X - Y = \lambda K$$
Let $\overline{X}$ denote the equivalence class of $X \in T_pS$ and let,
$$T_pS/K = \{\overline{X} : X \in T_pS\}$$
Then,
$$TS/K = \cup_{p\in S} T_pS/K$$
is a rank $n-1$ vector bundle over S ($n = \dim S$). This vector bundle does not depend on the particular choice of null vector field K.
- There is a natural positive definite metric h on TS/K induced from $\langle\,,\rangle$: For each $p \in S$, define $h : T_pS/K \times T_pS/K \to \mathbb{R}$ by
$$h(\overline{X}, \overline{Y}) = \langle X, Y\rangle.$$
Well-defined: If $X' = X \bmod K$, $Y' = Y \bmod K$ then
$$\begin{aligned}\langle X', Y'\rangle &= \langle X + \alpha K, Y + \beta K\rangle \\ &= \langle X, Y\rangle + \beta\langle X, K\rangle + \alpha\langle K, Y\rangle + \alpha\beta\langle K, K\rangle \\ &= \langle X, Y\rangle\,.\end{aligned}$$

- The *null Weingarten map* $b = b_K$ of S with respect to K is, for each point $p \in S$, a linear map $b : T_pS/K \to T_pS/K$ defined by

$$b(\overline{X}) = \overline{\nabla_X K}\,.$$

b is well-defined: $X' = X \mod K \Rightarrow$

$$\begin{aligned}\nabla_{X'} K &= \nabla_{X+\alpha K} K \\ &= \nabla_X K + \alpha \nabla_K K = \nabla_X K + \alpha\lambda K \\ &= \nabla_X K \quad \mod K\end{aligned}$$

- b is self adjoint with respect to h, i.e., $h(b(\overline{X}), \overline{Y}) = h(\overline{X}, b(\overline{Y}))$, for all $\overline{X}, \overline{Y} \in T_pS/K$.

Proof: Extend $X, Y \in T_pS$ to vector fields tangent to S near p. Using $X\langle K, Y\rangle = 0$ and $Y\langle K, X\rangle = 0$, we obtain,

$$\begin{aligned}h(b(\overline{X}), \overline{Y}) &= h(\overline{\nabla_X K}, \overline{Y}) = \langle \nabla_X K, Y\rangle \\ &= -\langle K, \nabla_X Y\rangle = -\langle K, \nabla_Y X\rangle + \langle K, [X, Y]\rangle \\ &= \langle \nabla_Y K, X\rangle = h(\overline{X}, b(\overline{Y}))\,.\end{aligned}$$

- The *null second fundamental form* $B = B_K$ of S with respect to K is the bilinear form associated to b via h:

  For each $p \in S$, $B : T_pS/K \times T_pS/K \to \mathbb{R}$ is defined by,

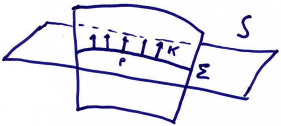
$$B(\overline{X}, \overline{Y}) := h(b(\overline{X}), \overline{Y}) = h(\overline{\nabla_X K}, \overline{Y}) = \langle \nabla_X K, Y \rangle \, .$$

  Since b is self-adjoint, B is symmetric, $B(\bar{X}, \bar{Y}) = B(\bar{Y}, \bar{X})$.

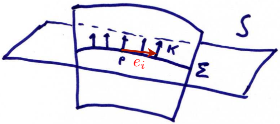
- The *null mean curvature* (or *null expansion scalar*) of S with respect to K is the smooth scalar field θ on S defined by,

$$\theta = \mathrm{tr}\, b$$

  θ has a natural geometric interpretation. Let Σ be the intersection of S with a hypersurface in M which is transverse to K near $p \in S$; Σ will be a co-dimension two spacelike submanifold of M, along which K is orthogonal.



The geometry of null hypersurfaces

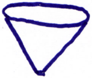


- Let $\{e_1, e_2, \cdots, e_{n-1}\}$ be an orthonormal basis for $T_p\Sigma$ in the induced metric. Then $\{\overline{e}_1, \overline{e}_2, \cdots, \overline{e}_{n-1}\}$ is an orthonormal basis for T_pS/K. Hence at p,

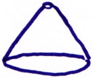
$$\begin{aligned}\theta = \operatorname{tr} b &= \sum_{i=1}^{n-1} h(b(\overline{e}_i), \overline{e}_i) = \sum_{i=1}^{n-1} B(\overline{e}_i, \overline{e}_i) = \sum_{i=1}^{n-1} \langle \nabla_{e_i} K, e_i \rangle. \\ &= \operatorname{div}_\Sigma K\,. \end{aligned} \tag{2.1}$$

  where $\operatorname{div}_\Sigma K$ is the divergence of K along Σ.

- Thus, θ measures the overall expansion of the null generators of S towards the future.



$\theta > 0$



$\theta < 0$

- *Effect of scaling:* If $\widetilde{K} = fK$, $f \in C^\infty(S)$, is any other future directed null vector field on S, then $b_{\tilde{K}} = fb_K$, and hence, $\widetilde{\theta} = f\theta$ (exercise!).

  It follows that the Weingarten map $b = b_K$ at a point p is uniquely determined by the value of K at p.

Comparison Theory

We now study how the null Weingarten map propagates along the null geodesic generators of S.

Let $\eta : I \to M$, $s \to \eta(s)$, be a future directed affinely parameterized null geodesic generator of S. For each $s \in I$, consider the Weingarten map $b = b(s)$ based at $\eta(s)$ with respect to a null vector field K which equals $\eta'(s)$ at $\eta(s)$,

$$b(s) = b_{\eta'(s)} : T_{\eta(s)}S/\eta'(s) \to T_{\eta(s)}S/\eta'(s)$$

Proposition 2.3

The one parameter family of Weingarten maps $s \to b(s)$, obeys the following Riccati equation,

$$b' + b^2 + R = 0\,, \qquad {}' = \nabla_{\eta'} \tag{2.2}$$

where $R : T_{\eta(s)}S/\eta'(s) \to T_{\eta(s)}S/\eta'(s)$ is given by $R(\overline{X}) = \overline{R(X, \eta'(s))\eta'(s)}$.

Remark on notation: In general, if $Y = Y(s)$ is a vector field along η tangent to S, we define, $(\overline{Y})' = \overline{Y'}$. Then, if $X = X(s)$ is a vector field along η tangent to S, b' is defined by,

$$b'(\overline{X}) := b(\overline{X})' - b(\overline{X'})\,. \tag{2.3}$$

Proof: Fix a point $p = \eta(s_0)$, $s_0 \in (a, b)$, on η. On a neighborhood U of p in S we can scale the null vector field K so that K is a geodesic vector field, $\nabla_K K = 0$, and so that $K_{\eta(s)} = \eta'(s)$ for each s near s_0.

Let $X \in T_pM$. Shrinking U if necessary, we can extend X to a smooth vector field on U so that $[X, K] = \nabla_X K - \nabla_K X = 0$. Then,

$$R(X, K)K = \nabla_X \nabla_K K - \nabla_K \nabla_X K - \nabla_{[X,K]} K = -\nabla_K \nabla_K X$$

Hence along η we have,

$$X'' = -R(X, \eta')\eta'$$

(which implies that X, restricted to η, is a Jacobi field along η).

Thus, from Equation (2.3), at the point p we have,

$$\begin{aligned} b'(\overline{X}) &= \overline{\nabla_X K}\,' - b(\overline{\nabla_K X}) = \overline{\nabla_K X}\,' - b(\overline{\nabla_X K}) \\ &= \overline{X''} - b(b(\overline{X})) = -\overline{R(X, \eta')\eta'} - b^2(\overline{X}) \\ &= -R(\overline{X}) - b^2(\overline{X}), \end{aligned}$$

which establishes Equation (2.2).

By taking the trace of (2.2) we obtain the following formula for the derivative of the null mean curvature $\theta = \theta(s)$ along η,

$$\theta' = -\mathrm{Ric}(\eta', \eta') - \sigma^2 - \frac{1}{n-1}\theta^2, \tag{2.4}$$

where $\sigma := (\mathrm{tr}\, \hat{b}^2)^{1/2}$ is the *shear scalar*, $\hat{b} := b - \frac{1}{n-1}\theta \cdot \mathrm{id}$ is the trace free part of the Weingarten map, and $\mathrm{Ric}(\eta', \eta')$ is the spacetime Ricci tensor evaluated on the tangent vector η'.

Equation 2.4 is known in relativity as the Raychaudhuri equation (for an irrotational null geodesic congruence). This equation shows how the Ricci curvature of spacetime influences the null mean curvature of a null hypersurface.

We consider a basic application of the Raychaudhuri equation.

Proposition 2.4

Let M be a spacetime which obeys the null enery condition (NEC), $\mathrm{Ric}\,(X, X) \geq 0$ for all null vectors X, and let S be a smooth null hypersurface in M. If the null generators of S are future geodesically complete then S has nonnegative null expansion, $\theta \geq 0$.

Proof: Suppose $\theta < 0$ at $p \in S$. Let $s \to \eta(s)$ be the null generator of S passing through $p = \eta(0)$, affinely parametrized. Let $b(s) = b_{\eta'(s)}$, and take $\theta = \mathrm{tr}\ b$. By the invariance of sign under scaling, one has $\theta(0) < 0$.

Raychaudhuri's equation and the NEC imply that $\theta = \theta(s)$ obeys the inequality,

$$\frac{d\theta}{ds} \le -\frac{1}{n-1}\theta^2 \,,$$

and hence $\theta < 0$ for all $s > 0$. Dividing through by θ^2 then gives,

$$\frac{d}{ds}\left(\frac{1}{\theta}\right) \ge \frac{1}{n-1} \,,$$

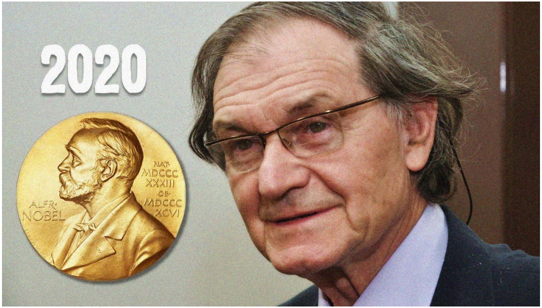
which implies $1/\theta \to 0$, i.e., $\theta \to -\infty$ in finite affine parameter time, contradicting the smoothness of θ. $\square$

Remark. Let Σ be a local cross section of the null hypersurface S (see earlier figure) with volume form ω. If Σ is moved under flow generated by K then $L_K\omega = \theta\,\omega$, where $L =$ Lie derivative.

Thus, Proposition 2.4 implies, under the given assumptions, that cross sections of S are nondecreasing in area as one moves towards the future.

Proposition 2.4 is the simplest form of Hawking's black hole area theorem [38]. For a study of the area theorem, with a focus on issues of regularity, see [15].

The Penrose Singularity Theorem and Related Results



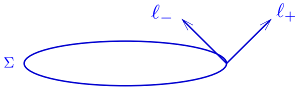
- In December, 2020, Roger Penrose was awarded (one half of) the 2020 Nobel prize for "the discovery that black hole formation is a robust prediction of the general theory of relativity", that is, for his 1965 gravitational collapse singularity theorem.
- Penrose's goal was to show that singularities appearing in exact solutions of the Einstein equation (such as the Schwarzschild solution) were not just artifacts of their exact symmetries (e.g. spherical symmetry).
- This led to his introduction of the notion of a trapped surface.

In this section we introduce Penrose's famous notion of a trapped surface and present the classical Penrose singularity theorem.

- Let (M^{n+1}, g) be an $(n+1)$-dimensional spacetime, $n \geq 3$.
  Let Σ^{n-1} be a closed (i.e., compact without boundary) co-dimension two spacelike submanifold of M.

- Each normal space of Σ, $[T_p\Sigma]^\perp$, $p \in \Sigma$, is timelike and 2-dimensional, and hence admits two future directed null directions orthogonal to Σ.

  Thus, under appropriate orientation assumptions, Σ admits two smooth nonvanishing future directed null normal vector fields ℓ_+ and ℓ_- (unique up to positive rescaling).

  By convention, we refer to ℓ_+ as outward pointing and ℓ_- as inward pointing.

- Associated to ℓ_+ and ℓ_-, are the two *null second fundamental forms*, χ_+ and χ_-, respectively, defined as

$$\chi_\pm : T_p\Sigma \times T_p\Sigma \to \mathbb{R}, \qquad \chi_\pm(X,Y) = g(\nabla_X \ell_\pm, Y)\,.$$

- *The null expansion scalars (or null mean curvatures)* $\theta_\pm$ of Σ are obtained by tracing $\chi_\pm$ with respect to the induced metric γ on Σ,

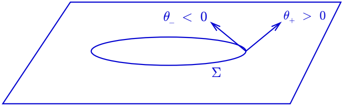
$$\theta_\pm = \mathrm{tr}_\Sigma \chi_\pm = \sum_{i=1}^{n-1} \chi_\pm(e_i, e_i) = \mathrm{div}\,_\Sigma \ell_\pm\,.$$

The sign of $\theta_\pm$ does not depend on the scaling of $\ell_\pm$ (exercise).

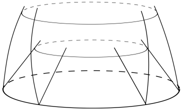
Physically, θ_+ (resp., θ_-) measures the divergence of the outgoing (resp., ingoing) light rays emanating orthogonally from Σ.

Remark: There is a natural connection between these null expansion scalars $\theta_\pm$ and the null expansion of null hypersurfaces: ℓ_+ locally generates a smooth null hypersurface S_+. Then θ_+ is the null expansion of S_+ restricted to Σ; θ_- may be described similarly.

- For round spheres in Euclidean slices in Minkowski space (and, more generally, large "radial" spheres in AF spacelike hypersurfaces),



- However, in regions of spacetime where the gravitational field is strong, one can have both



$$\theta_- < 0 \quad \text{and} \quad \theta_+ < 0\,,$$

in which case Σ is called a **trapped surface**.

Penrose observed that trapped surfaces occur inside the black hole region of the Schwarzchild solution.

- As we now discuss, assuming appropriate energy and causality conditions, if a trapped surface forms, then the development of "singularities" is inevitable.

Theorem 3.1 (Penrose singularity theorem)

Let M be a globally hyperbolic spacetime which satisfies the NEC, $\mathrm{Ric}(X, X) \geq 0$ for all null vectors X, and which has a noncompact Cauchy surface S. If M contains a trapped surface Σ then M is future null geodesically incomplete.

Proof: We first observe the following.

Claim: $\partial I^+(\Sigma)$ is noncompact.

Proof of Claim: $\partial I^+(\Sigma)$ is an achronal boundary, and hence, by Proposition 1.8, is an achronal C^0 hypersurface. If $\partial I^+(\Sigma)$ were compact then, by Proposition 1.21, $\partial I^+(\Sigma)$ would be a compact Cauchy surface. But this would contradict the assumption that S is noncompact (all Cauchy surfaces are homeomorphic). $\square$

We now construct a future inextendible null geodesic in $\partial I^+(\Sigma)$, which we show must be future incomplete.

- We have that

$$\partial I^+(\Sigma) = \overline{I^+(\Sigma)} \setminus \operatorname{int} I^+(\Sigma) = J^+(\Sigma) \setminus I^+(\Sigma)\,.$$

  It then follows from Proposition 1.3 that each $q \in \partial I^+(\Sigma)$ lies on a null geodesic in $\partial I^+(\Sigma)$ with past end point on Σ. Moreover this null geodesic meets Σ orthogonally (due to achronality, cf. O'Neill [52, Lemma 50, p. 298]).

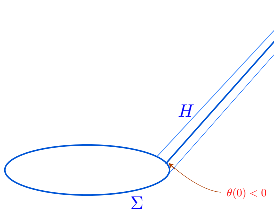
- Since $\partial I^+(\Sigma)$ is closed and noncompact, there exists a sequence of points $\{q_k\} \subset \partial I^+(\Sigma)$ that diverges to infinity. For each k, there is a null geodesic η_k from Σ to q_k, which is contained in $\partial I^+(\Sigma)$ and meets Σ orthogonally.

- By compactness of Σ, some subsequence η_{k_j} converges to a future inextendible null geodesic η contained in $\partial I^+(\Sigma)$, and meeting Σ orthogonally (at p, say).

  η must be future incomplete. Suppose not.

- By achronality of $\partial I^+(\Sigma)$,
  - No other future directed null normal geodesic starting on Σ can meet η.
  - There can be no *null focal point* to Σ along η (cf. O'Neill, Prop. 48, p. 296).

- It follows that η is contained in a smooth (perhaps very thin) null hypersurface $H \subset \partial I^+(\Sigma)$.

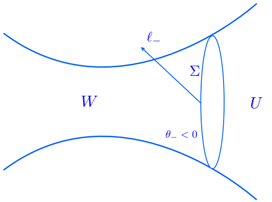


- Let θ be the null expansion of H along η. Since Σ is a trapped surface $\theta(p) < 0$. Arguing just as in the "area theorem" (Proposition 2.4), using Raychaudhuri + NEC, θ must go to $-\infty$ in finite affine parameter time $\rightarrow\leftarrow$. Hence η must be future incomplete. $\square$

For certain applications, the following variant of the Penrose singularity theorem is useful.

Theorem 3.2 ("One-sided Penrose")

Let M be a globally hyperbolic spacetime satisfying the null energy condition, with smooth spacelike Cauchy surface V. Let Σ be a smooth, closed, connected hypersurface in V which separates V into an "inside" W and an "outside" U, i.e., $V \setminus \Sigma = U \cup W$ where $U, W \subset V$ are connected disjoint sets. Suppose, further, that $\overline{W}$ is non-compact. If Σ is inner-trapped ($\theta_- < 0$) then M is future null geodesically incomplete.



Proof: Hints: Consider the achronal boundary $\partial I^+(\overline{U})$, and argue similarly to the proof of the Penrose singularity theorem that if M is future null geodesically complete then $\partial I^+(\overline{U})$ is compact. Show that this is not compatible with W being noncompact.

This version of the Penrose singularity theorem may be used to prove the following beautiful result of Gannon [36] and Lee [48].

Theorem 3.3 (Gannon-Lee)

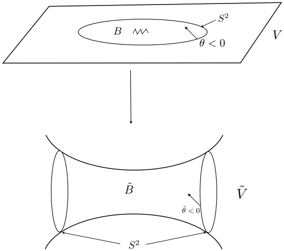
Let M be a globally hyperbolic spacetime which satisfies the null energy condition and which contains a smooth asymptotically flat spacelike Cauchy surface V. If V is not simply connected ($\pi_1(V) \neq 0$) then M is future null geodesically incomplete.

This version of the Penrose singularity theorem may be used to prove the following beautiful result of Gannon [36] and Lee [48].

Theorem 3.3 (Gannon-Lee)

Let M be a globally hyperbolic spacetime which satisfies the null energy condition and which contains a smooth asymptotically flat spacelike Cauchy surface V. If V is not simply connected ($\pi_1(V) \neq 0$) then M is future null geodesically incomplete.

<u>*Comment on the proof.*</u> Let $\tilde{V}$ be the universal cover of V. If $\pi_1(V) \neq 0$ then $\tilde{V}$ will have more than one AF end.



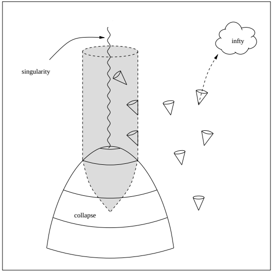
Associated to $\tilde{V}$ is a spacetime $\tilde{M} \approx \mathbb{R} \times \tilde{V}$ which covers $M \approx \mathbb{R} \times V$. Now apply Theorem 3.2 to $\tilde{M}$ with Cauchy surface $\tilde{V}$.

The topology of black holes

Introduction

Black holes are certainly one of the most remarkable predictions of General Relativity.

The following depicts the process of gravitational collapse and formation of a black hole.



A stellar object, after its fuel is spent, begins to collapse under its own weight. As the gravitational field intensifies the light cones bend "inward" (so to speak).

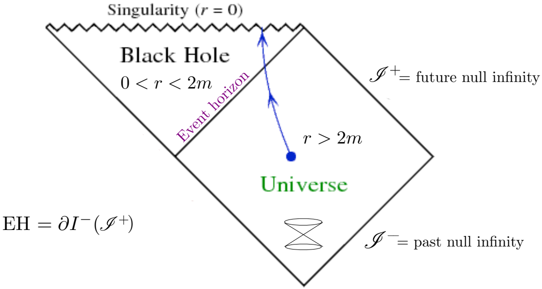
The shaded region is the black hole region. The boundary of this region is the black hole event horizon. It is the boundary between points that can send signals to infinity and points that can't.

<u>*Ex.*</u> *The Schwarzshild solution (1916).* Static (time-independent, nonrotating) spherically symmetric, vacuum solution to the Einstein equations.

$$g = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 d\Omega^2$$

This metric represents the region outside a (collapsing) spherically symmetric star.

The region $0 < r < 2m$ is the black hole region; $r = 2m$ corresponds to the event horizon.



Ex. The Kerr solution (1963). Stationary (time-independent, rotating), axisymmetric, vacuum solution.

$$
\begin{aligned}
\mathrm{d}s^2 \;=\; & -\left[1 - \frac{2mr}{r^2 + a^2\cos^2\theta}\right]\mathrm{d}t^2 - \frac{4mra\sin^2\theta}{r^2 + a^2\cos^2\theta}\,\mathrm{d}t\,\mathrm{d}\phi \\
& + \left[\frac{r^2 + a^2\cos^2\theta}{r^2 - 2mr + a^2}\right]\mathrm{d}r^2 + (r^2 + a^2\cos^2\theta)\,\mathrm{d}\theta^2 \\
& + \left[r^2 + a^2 + \frac{2mra^2\sin^2\theta}{r^2 + a^2\cos^2\theta}\right]\sin^2\theta\,\mathrm{d}\phi^2.
\end{aligned}
$$

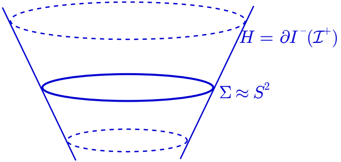
- The Kerr solution is determined by two parameters: mass parameter m and angular momentum parameter a. When $a = 0$, the Kerr solution reduces to the Schwarzschild solution. The Kerr solution contains an event horizon (provided $a < m$), and hence represents a steady state rotating black hole.

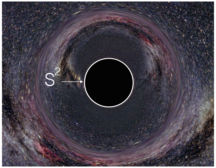
- It is a widely held belief that "true" astrophysical black holes "settle down" to a Kerr solution. This belief is based largely on results ("no hair theorems") that establish the uniqueness of Kerr among all asymptotically flat stationary, solutions to the vacuum Einstein equations; see e.g. [14]. (There has also been progress on establishing the nonlinear stability of the Kerr solution.)

A basic step in the proof of the uniqueness of the Kerr solution is Hawking's theorem on the topology of black holes in $3+1$ dimensions.

Theorem 4.1 (Hawking's black hole topology theorem)

Suppose (M, g) is a $(3+1)$-dimensional asymptotically flat stationary black hole spacetime obeying the dominant energy condition. Then cross sections Σ of the event horizon are topologically 2-spheres.





Comment on the proof: Hawking's proof is variational in nature. Using the dominant energy condition and the Gauss-Bonnet theorem, he shows that if Σ has genus ≥ 1 then Σ can be deformed outward to an *outer trapped surface*.

However, there can be no outer trapped surface outside the event horizon. Such a surface would be visible from 'null infinity', but there are arguments precluding that possibility [60, 17].

Higher Dimensional Black Holes

- String theory, and various related developments (e.g., the AdS/CFT correspondence, braneworld scenarios, entropy calculations) have generated a great deal of interest in gravity in higher dimensions, and in particular, in higher dimensional black holes.
- One of the first questions to arise was:

  Does black hole uniqueness hold in higher dimensions?

- With impetus coming from the development of string theory, in 1986, Myers and Perry [51] constructed natural higher dimensional generalizations of the Kerr solution. These models painted a picture consistent with the situation in $3+1$ dimensions. In particular, they have spherical horizon topology.

- But in 2002, Emparan and Reall [23] discovered a remarkable example of a $4+1$ dimensional AF stationary vacuum black hole spacetime with horizon topology $S^2 \times S^1$ (the *black ring*).

VOLUME 88, NUMBER 10 PHYSICAL REVIEW LETTERS 11 MARCH 2002

A Rotating Black Ring Solution in Five Dimensions

Roberto Emparan[1,*] and Harvey S. Reall[2]
[1]*Theory Division, CERN, CH-1211 Geneva 23, Switzerland*
[2]*Physics Department, Queen Mary College, Mile End Road, London E1 4NS, United Kingdom*
(Received 8 November 2001; published 21 February 2002)

Thus in higher dimensions, black hole uniqueness does not hold and horizon topology need not be spherical.

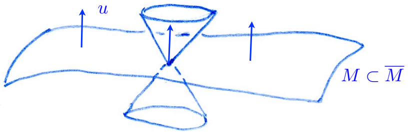
This caused a great surge of activity in the study of higher dimensional black holes.

- *Question:* What horizon topologies are allowed in higher dimensions? What restrictions are there?

Marginally Outer Trapped Surfaces

- We want to describe a generalization of Hawking's theorem to higher dimensions. This will be based on properties of marginally outer trapped surfaces or MOTS, for short.
- Under natural circumstances, MOTS model cross sections of black hole event horizons. We will be considering MOTS in initial data sets.
- Formally, an initial data set is a triple (M, g, K), where (M, g) is a smooth Riemannian manifold, and K is a symmetric $(0, 2)$-tensor field on M.
- Though not strictly necessary, we find it convenient to assume that our initial data sets (M, g, K) are embedded in a spacetime $(M, \overline{g})$, meaning that M is a spacelike hypersurface in $(\overline{M}, \overline{g})$, with induced metric g and second fundamental form K:

$$K(X, Y) = \overline{g}(\overline{\nabla}_X u, Y) \quad \text{for vectors } X, Y \in T_p M.$$



Initial data DEC

- Recall, a spacetime $(\overline{M}, \overline{g})$ is said to satisfy the dominant energy condition provided,

$$G(X, Y) \geq 0\,, \quad \text{for all future directed causal vectors } X, Y.$$

  where G is the Einstein tensor, $G := \mathrm{Ric}_{\overline{M}} - \frac{1}{2} R_{\overline{M}} \overline{g} \quad (= 8\pi T)$.

- The spacetime DEC implies the following holds along the spacelike hypersurface M,

$$\mu \geq |J|\,, \qquad\qquad (*)$$

  where μ = *local energy density* = $G(u, u)$, and J = *local momentum density* = 1-form $G(u, \cdot)$ on M.

- μ and J can be expressed solely in terms of initial data:

$$\mu = \frac{1}{2}\left(S + (\mathrm{tr}\, K)^2 - |K|^2\right)\,, \qquad \text{(Einstein constraints)}$$

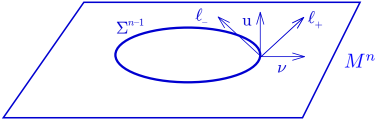
$$J = \mathrm{div} K - d(\mathrm{tr}\, K)\,.$$

  where S is the scalar curvature of M.

  In the *time-symmetric* case ($K = 0$), $(*)$ reduces to $S \geq 0$.

- Consider an initial data set (M^n, g, K) in a spacetime $(\overline{M}^{n+1}, \overline{g})$, $n \geq 3$.

  Let Σ^{n-1} be a closed 2-sided hypersurface in M^n. Σ admits a smooth unit normal field ν in M.



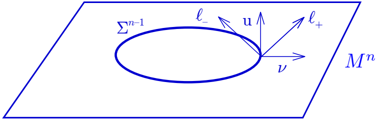
$$\ell_+ = u + \nu \qquad \text{f.d. outward null normal}$$
$$\ell_- = u - \nu \qquad \text{f.d. inward null normal}$$

- *Null second fundamental forms:* χ_+, χ_-

$$\chi_\pm(X, Y) = g(\nabla_X \ell_\pm, Y) \qquad X, Y \in T_p\Sigma$$

- *Null expansion scalars:* θ_+, θ_-

$$\theta_\pm = \mathrm{tr}_\Sigma \chi_\pm = \mathrm{div}_\Sigma \ell_\pm$$



$$\theta_\pm = \mathrm{tr}_\Sigma \chi_\pm = \mathrm{div}_\Sigma \ell_\pm$$

- Check that the *sign* of $\theta_\pm$ is invariant under positive rescaling of $\ell_\pm$.
  Physically, θ_+ measures the divergence of the outgoing light rays from Σ.

- In terms of initial data (M^n, g, K),

$$\chi_\pm = K|_\Sigma \pm A\,,$$

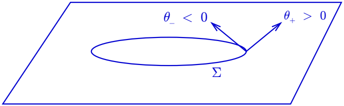
  where A is the second fundamental form of Σ within M, and hence,

$$\theta_\pm = \mathrm{tr}_\Sigma K \pm H\,,$$

  where $H =$ mean curvature of Σ within M.

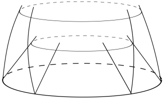
  Note: In the time-symmetric case, $K = 0$, $\theta_+ = H$.

- For round spheres in Euclidean slices in Minkowski space (and, more generally, large "radial" spheres in AF spacelike hypersurfaces),



- However, in a strong gravitational field one can have both,

$$\theta_- < 0 \qquad \text{and} \qquad \theta_+ < 0\,,$$



in which case Σ is trapped surface (Penrose).

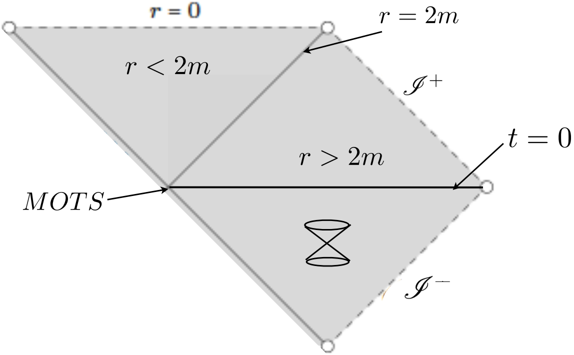
- Focusing attention on the outward null normal:
  - If $\theta_+ < 0$ - we say Σ is outer trapped
  - If $\theta_+ = 0$ - we say Σ is a marginally outer trapped surface (MOTS)

Note: In the time symmetric case a MOTS is simply a minimal surface.

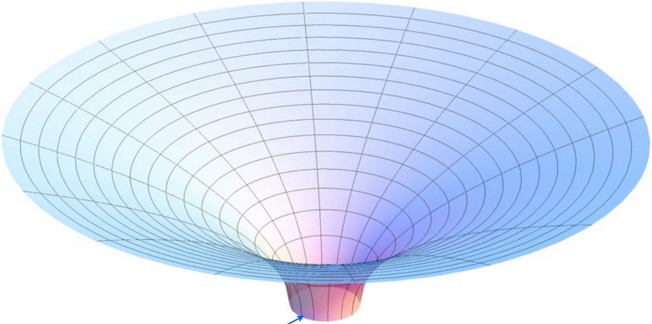
Let's consider some examples of MOTSs in the Schwarzschild spacetime.

Ex. The Schwarzschild spacetime.

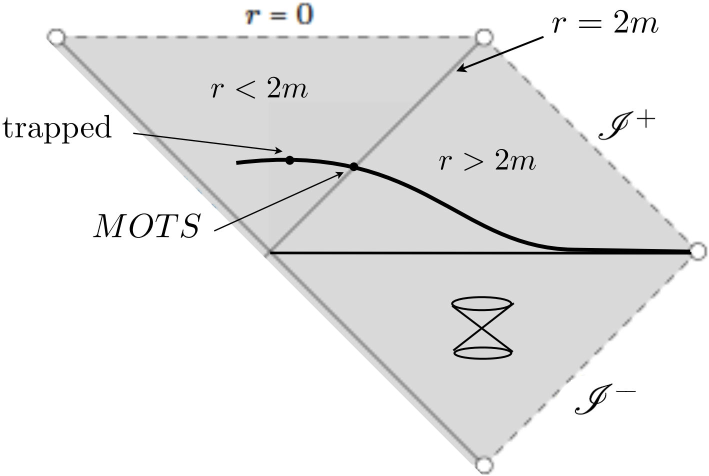
$$g = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 d\Omega^2$$



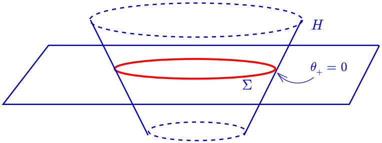
The $t = 0$ slice in Schwarzschild (the Flamm paraboloid):



$MOTS$ (minimal surface)

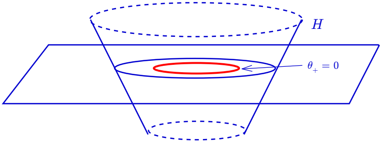

$r = 0$
$r = 2m$
$r < 2m$
trapped
$\mathscr{I}^+$
$r > 2m$
$MOTS$
$\mathscr{I}^-$

- In fact, in general, in *stationary* black hole spacetimes - cross sections of the event horizon are MOTS.

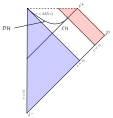


(In the stationary case, the null generators of the horizon have zero expansion.)

- In dynamical black hole spacetimes - MOTS typically occur inside the event horizon:



E.g. Vaidya spacetime



MOTSs admit a notion of stability based on variations of the null expansion (Andersson, Mars and Simon [4, 5]).

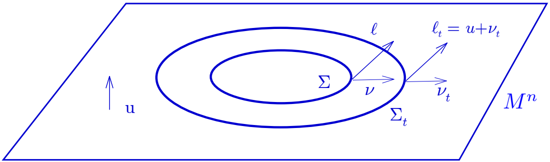
Let Σ be a MOTS in an initial data set (M, g, K) with outward normal ν. Consider normal variations of Σ in M, i.e., variations $t \to \Sigma_t$ of $\Sigma = \Sigma_0$ with variation vector field

$$\mathcal{V} = \frac{\partial}{\partial t}\Big|_{t=0} = \phi\nu, \quad \phi \in C^\infty(\Sigma)\,.$$

Let

$$\theta(t) = \text{ the null expansion of } \Sigma_t,$$

with respect to $\ell_t = u + \nu_t$, where ν_t is the unit normal field to Σ_t in M.



A computation shows ([5, 7]),

$$\left.\frac{\partial\theta}{\partial t}\right|_{t=0} = L(\phi)\,,$$

where $L : C^\infty(\Sigma) \to C^\infty(\Sigma)$ is given by,

$$L(\phi) = -\triangle\phi + 2\langle X, \nabla\phi\rangle + \left(Q + \mathrm{div}\, X - |X|^2\right)\phi\,,$$

$$Q = \frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi|^2\,,$$

- In the time-symmetric case ($K = 0$), $\theta = H$, $X = 0$, and L reduces to the classical stability operator of minimal surface theory.
- In analogy with minimal surface case, we refer to L as the MOTS stability operator. Note, however, that L is *not* in general self-adjoint.

Nevertheless, one has the following. (as a consequence of Krein-Rutman, see Andersson, Mars, Simon [5] and also [49]).

Lemma 4.2

Among eigenvalues with smallest real part, there is a real eigenvalue $\lambda_1(L)$, called the principal eigenvalue. The associated eigenfunction ϕ, $L(\phi) = \lambda_1 \phi$, is unique up to a multipicative constant, and can be chosen to be strictly positive.

In analogy with the minimal surface case:

$$\boxed{\text{We say that a MOTS } \Sigma \text{ is stable provided } \lambda_1(L) \geq 0}$$

Remarks:

- There is a physical characterization of stability: Σ is stable iff there is an outward variation with $\left.\frac{\partial \theta}{\partial t}\right|_{t=0} \geq 0$.
- In the minimal surface case this is equivalent to the second variation of area being nonnegative.

There is a basic criterion for a MOTS to be stable.

We say a MOTS Σ is *weakly outermost* provided there are no outer trapped ($\theta < 0$) surfaces outside of, and homologous, to Σ.

- **Fact.** A weakly outermost MOTS is stable.

  Proof: Suppose to the contrary, $\lambda_1 < 0$. Consider the variation $t \to \Sigma_t$ of Σ with variation vector field $\mathcal{V} = \phi\nu$, where ϕ is a positive eigenfunction associated to $\lambda_1 = \lambda_1(L)$. Then,

  $$\left.\frac{\partial\theta}{\partial t}\right|_{t=0} = L(\phi) = \lambda_1\phi < 0$$

  Since $\theta(0) = 0$, this implies $\theta(t) < 0$ for small $t > 0$., There is a basic criterion for a MOTS to be stable.

We say a MOTS Σ is *weakly outermost* provided there are no outer trapped ($\theta < 0$) surfaces outside of, and homologous, to Σ.

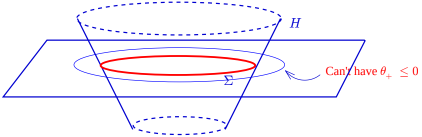
- **Fact.** A weakly outermost MOTS is stable.

  Proof: Suppose to the contrary, $\lambda_1 < 0$. Consider the variation $t \to \Sigma_t$ of Σ with variation vector field $\mathcal{V} = \phi\nu$, where ϕ is a positive eigenfunction associated to $\lambda_1 = \lambda_1(L)$. Then,

  $$\left.\frac{\partial \theta}{\partial t}\right|_{t=0} = L(\phi) = \lambda_1 \phi < 0$$

  Since $\theta(0) = 0$, this implies $\theta(t) < 0$ for small $t > 0$.,

- **Fact:** Cross sections of the event horizon in AF stationary black hole spacetimes obeying the DEC are (weakly) outermost MOTSs.



- More generally, weakly outermost MOTSs can arise as the boundary of the "trapped region" (Andersson and Metzger [6], Eichmair [19]).

A Generalization of Hawking's Black Hole Topology Theorem

Theorem 4.3 (G. and Schoen [35])

Let (M, g, K) be an n-dimensional initial data set, $n \geq 3$, satisfying the dominant energy condition (DEC), $\mu \geq |J|$. If Σ is a stable MOTS in M then (apart from certain exceptional circumstances) Σ must be of positive Yamabe type, i.e. must admit a metric of positive scalar curvature.

- The theorem may be viewed as a spacetime analogue of a classical result of Schoen and Yau [55] concerning stable minimal hypersurfaces in manifolds of positive scalar curvature.
- Exceptional circumstances: Various geometric quantities vanish: $\mathrm{Ric}_\Sigma = 0$ (i.e. Σ is Ricci flat), $\mu + J(\nu) = 0$, and $\chi_+ = 0$,

Thus, apart from these exceptional circumstances, Σ is of positive Yamabe type.

Σ being positive Yamabe implies many well-known restrictions on the topology.

Comments

- $\dim \Sigma = 2$ ($\dim \overline{M} = 3+1$): Then $\Sigma \approx S^2$ by Gauss-Bonnet, and one recovers Hawking's theorem.
- $\dim \Sigma = 3$ ($\dim \overline{M} = 4+1$): Then by the classical results of Schoen-Yau and Gromov-Lawson (and assuming orientability) Σ must be diffeomorphic to:
  - a spherical space, i.e. S^3 or a quotient of a S^3, or
  - $S^2 \times S^1$, or
  - a connected sum of the above two types.

  Thus, the basic horizon topologies in the case $\dim \Sigma = 3$ are (i) S^3 which is realized by the Myers-Perry black holes and (ii) $S^2 \times S^1$ realized by the Emparan-Reall "Black Ring".
- The question then became: Which other topologies can actually be *mathematically* realized by a black hole spacetime with appropriate properties (e.g. asymptotically flat, stationary, axisymmetric)?

- There has been a considerable amount of work on this question by a number of people.
- Under certain symmetry assumptions (stationary, biaxial symmetry) the list of possible topologies was reduced to S^3, $S^2 \times S^1$ and the lens spaces $L(p, q) \cong S^3/\mathbb{Z}_p$, $q < p$ (Hollands and Yazadjiev, [40]).

  Attention then focused on constructing black hole spacetimes, with appropriate properties, having lens space horizon topology $L(p, q)$.
- Recently Marcus Khuri and Jordan Rainone succeeded in constructing such black hole spacetimes having any lens space horizon topology ("Black Lenses in Kaluza-Klein Matter" [45]).

The topology of black holes

- There has been a considerable amount of work on this question by a number of people.
- Under certain symmetry assumptions (stationary, biaxial symmetry) the list of possible topologies was reduced to S^3, $S^2 \times S^1$ and the lens spaces $L(p, q) \cong S^3/\mathbb{Z}_p$, $q < p$ (Hollands and Yazadjiev, [40]).

  Attention then focused on constructing black hole spacetimes, with appropriate properties, having lens space horizon topology $L(p, q)$.
- Recently Marcus Khuri and Jordan Rainone succeeded in constructing such black hole spacetimes having any lens space horizon topology ("Black Lenses in Kaluza-Klein Matter" [45]).
- This work has been the focus of a recent Quanta Magazine article.

MATHEMATICAL PHYSICS

Mathematicians Find an Infinity of Possible Black Hole Shapes

By STEVE NADIS |

JANUARY 24, 2023 | 12 |

In three-dimensional space, the surface of a black hole must be a sphere. But a new result shows that in higher dimensions, an infinite number of configurations are possible.

Theorem 4.4 (G. and Schoen, [35])

Let (M, g, K), be an n dimensional, $n \geq 3$, initial data set, obeying the DEC, $\mu \geq |J|$. If Σ is a stable MOTS in M then (apart from certain exceptional circumstances) Σ must be of positive Yamabe type.

Comments on the proof: MOTS stability operator:

$$L(\phi) = -\triangle\phi + 2g(X, \nabla\phi) + \left(Q + \operatorname{div} X - |X|^2\right)\phi\,,$$

$$Q = \frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2\,,$$

Key fact: If Σ is stable ($\lambda_1(L) \geq 0$) then Σ satisfies the MOTS stability inequality:

$$\int_\Sigma |\nabla\psi|^2 + \left(\frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2\right)\psi^2 \geq 0\,, \quad \forall\psi \in C^\infty(\Sigma) \qquad (*)$$

- Consider the symmetrized operator, obtained by setting $X = 0$,

$$L_0(\phi) = -\triangle\phi + (\frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2)\,\phi \qquad (**)$$

- By the Raleigh formula for first eigenvalue,

$$
\begin{aligned}
\lambda_1(L_0) &= \inf_{\psi \not\equiv 0} \frac{\int_\Sigma \psi L_0(\psi)\, d\mu}{\int_\Sigma \psi^2\, d\mu} \\
&= \inf_{\psi \not\equiv 0} \frac{\int_\Sigma |\nabla \psi|^2 + \left(\frac{1}{2} S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2\right) \psi^2\, d\mu}{\int_\Sigma \psi^2\, d\mu} \geq 0\,.
\end{aligned}
$$

- At this stage fairly standard arguments can be employed. By making the conformal change: $\tilde{\gamma} = \phi^{\frac{2}{n-2}}\gamma$, where ϕ is a positive eigenfunction corresponding to $\lambda_1(L_0)$ $(L_0(\phi) = \lambda_1(L_0)\phi)$, a computation shows,

$$
\begin{aligned}
\tilde{S}_\Sigma &= \phi^{-\frac{n}{n-2}}\left(-2\triangle\phi + S_\Sigma \phi + \frac{n-1}{n-2}\frac{|\nabla\phi|^2}{\phi}\right) \\
&= \phi^{-\frac{2}{n-2}}\left(2\lambda_1(L_0) + 2(\mu + J(\nu)) + |\chi_+|^2 + \frac{n-1}{n-2}\frac{|\nabla\phi|^2}{\phi^2}\right) \geq 0
\end{aligned}
$$

$(\mu + J(\nu) \geq \mu - |J| \geq 0)$.

- By the Raleigh formula for first eigenvalue,

$$\lambda_1(L_0) = \inf_{\psi\not\equiv 0} \frac{\int_\Sigma \psi L_0(\psi)\, d\mu}{\int_\Sigma \psi^2\, d\mu}$$
$$= \inf_{\psi\not\equiv 0} \frac{\int_\Sigma |\nabla\psi|^2 + \left(\frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2\right)\psi^2\, d\mu}{\int_\Sigma \psi^2\, d\mu} \geq 0\,.$$

- At this stage fairly standard arguments can be employed. By making the conformal change: $\tilde\gamma = \phi^{\frac{2}{n-2}}\gamma$, where ϕ is a positive eigenfunction corresponding to $\lambda_1(L_0)$ ($L_0(\phi) = \lambda_1(L_0)\phi$), a computation shows,

$$\tilde S_\Sigma = \phi^{-\frac{n}{n-2}}(-2\triangle\phi + S_\Sigma\phi + \frac{n-1}{n-2}\frac{|\nabla\phi|^2}{\phi})$$
$$= \phi^{-\frac{2}{n-2}}(2\lambda_1(L_0) + 2(\mu + J(\nu)) + |\chi_+|^2 + \frac{n-1}{n-2}\frac{|\nabla\phi|^2}{\phi^2}) \geq 0$$

$(\mu + J(\nu) \geq \mu - |J| \geq 0)$.

- If $\tilde S_\Sigma > 0$ at some point, then by well known results [44] one can conformally change $\tilde h$ to a metric of strictly positive scalar curvature. If $\tilde S$ vanishes identically then, $\lambda_1(L_0) = 0$, $\mu + J(\nu) \equiv 0$, $\chi_+ \equiv 0$ and ϕ is constant, which then implies that $S \equiv 0$. By a result of Bourguinon (see [44]), it follows that Σ_Σ carries a metric of positive scalar curvature unless it is Ricci flat.

Key fact: If Σ is stable ($\lambda_1(L) \geq 0$) then Σ satisfies the MOTS stability inequality:

$$\int_\Sigma |\nabla\psi|^2 + Q\psi^2 \geq 0\,, \quad \forall \psi \in C^\infty(\Sigma) \qquad (*)$$

where $Q = \frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2$. Here is the proof:

- Let ϕ be a positive principle eigenvalue of L. Then by stability, $L(\phi) = \lambda_1(L)\phi \geq 0 \implies$

$$-\triangle\phi + 2\langle X, \nabla\phi\rangle + \left(Q + \mathrm{div}\, X - |X|^2\right)\phi \geq 0$$

- Completing the square on the LHS $\implies$

$$-\triangle\phi + (Q + \mathrm{div}\, X)\,\phi + \phi|\nabla \ln\phi|^2 - \phi|X - \nabla\ln\phi|^2 \geq 0$$

- Setting $u = \ln\phi \implies$

$$-\triangle u + Q + \mathrm{div}\, X - |X - \nabla u|^2 \geq 0$$

- Absorbing the Laplacian term $\triangle u = \mathrm{div}\,(\nabla u) \implies$

$$Q + \mathrm{div}\,(X - \nabla u) - |X - \nabla u|^2 \geq 0$$

Set $Y = X - \nabla u$.

- Setting $Y = X - \nabla u$, we arrive at,

$$-Q + |Y|^2 \leq \mathrm{div}\, Y\,.$$

- Multiplying through by ψ^2, where $\psi \in C^\infty(\Sigma)$, we derive,

$$\begin{aligned} -\psi^2 Q + \psi^2 |Y|^2 &\leq \psi^2 \mathrm{div}\, Y \\ &= \mathrm{div}\,(\psi^2 Y) - 2\psi\langle \nabla\psi, Y\rangle \\ &\leq \mathrm{div}\,(\psi^2 Y) + 2|\psi||\nabla\psi||Y| \\ &\leq \mathrm{div}\,(\psi^2 Y) + |\nabla\psi|^2 + \psi^2|Y|^2\,. \end{aligned}$$

- Integrating the above inequality yields,

$$\int_\Sigma |\nabla\psi|^2 + Q\psi^2 \geq 0 \quad \text{for all } \psi \in C^\infty(\Sigma)\,,$$

i.e., we obtain the MOTS stability inequality,

$$\int_\Sigma |\nabla\psi|^2 + \left(\frac{1}{2}S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2\right)\psi^2 \geq 0\,, \quad \forall \psi \in C^\infty(\Sigma) \quad \square$$

- *Exercise.* As observed, the stability inequality implies the following: $\lambda_1(L) \geq 0 \implies \lambda_1(L_0) \geq 0$. Show how to modify the *proof* of the stability inequality to obtain:

$$\boxed{\lambda_1(L_0) \geq \lambda_1(L)}$$

  Hint: in the first line of the proof replace "≥ 0" by "$= \lambda_1(L)$", and proceed.

- One drawback of our black hole topology theorem is that the 'exceptional circumstances' allow, for example, the possibility of a toroidal black hole in a vacuum spacetime.

  In Section 6 we describe how to remove such possibilities under natural circumstances.

The size of marginally outer trapped surfaces

- Our study of the topology of black holes was based primarily on properties of marginally outer trapped surfaces (MOTS).
- In fact, MOTS arose early in the development of the theory of black holes, in connection with gravitational collapse. See especially the discussion in Hawking and Ellis [38, Section 9.2]. As discussed there, under suitable circumstances, the occurrence of a MOTS signals the presence of a black hole.
- MOTSs arose in a more purely mathematical context in the work of Schoen and Yau [56] concerning the existence of solutions of Jang's equation, in connection with their proof of the positive mass theorem in the non-time symmetric case ($K \neq 0$).
- In this section we obtain some results concerning the *size* of MOTS.

Proposition 5.1 (G. and Mendes [31])

Let Σ be a (closed) stable MOTS in a 3-dimensional initial data (M, g, K). Suppose there exists $c > 0$, such that $\mu + J(\nu) \geq c$ on Σ, where ν is the outward unit normal to Σ. Then the area of Σ satisfies,

$$A(\Sigma) \leq \frac{4\pi}{c}\,.$$

Moreover, if equality holds,

(1) Σ is a round 2-sphere, with Gaussian curvature $\kappa_\Sigma = c$,

(2) the null second fundamental form of Σ vanishes $\chi_+ = 0$, and

(3) $\mu + J(\nu) = c$ on Σ.

Comments on the energy condition $\mu + J(\nu) \geq c$.

- Let (M, g, K) be an initial data set in a spacetime $(\bar{M}, \bar{g})$ which satisfies the Einstein equation,
$$G + \Lambda \bar{g} = \kappa T$$
where, $G = \mathrm{Ric}_{\bar{M}} - \frac{1}{2} R_{\bar{M}} \bar{g}$ is the Einstein tensor, and T is the energy-momentum tensor.
- Then, setting $\ell = u + \nu$, where ν is any unit vector tangent to M and u is the future directed unit normal to M, we have along Σ in M,
$$\mu + J(\nu) := G(u, u) + G(u, \nu) = G(u, \ell) = \kappa T(u, \ell) + \Lambda > 0$$
for ordinary matter fields, provided $\Lambda \geq 0$,
- Moreover, if T obeys DEC (including the vacuum case $T = 0$) and $\Lambda > 0$, then one has
$$\mu + J(\nu) \geq \Lambda > 0 \, .$$

Comments on the energy condition $\mu + J(\nu) \geq c$.

- Let (M, g, K) be an initial data set in a spacetime $(\bar{M}, \bar{g})$ which satisfies the Einstein equation,
$$G + \Lambda \bar{g} = \kappa T$$
where, $G = \mathrm{Ric}_{\bar{M}} - \frac{1}{2} R_{\bar{M}} \bar{g}$ is the Einstein tensor, and T is the energy-momentum tensor.
- Then, setting $\ell = u + \nu$, where ν is any unit vector tangent to M and u is the future directed unit normal to M, we have along Σ in M,
$$\mu + J(\nu) := G(u, u) + G(u, \nu) = G(u, \ell) = \kappa T(u, \ell) + \Lambda > 0$$
for ordinary matter fields, provided $\Lambda \geq 0$,
- Moreover, if T obeys DEC (including the vacuum case $T = 0$) and $\Lambda > 0$, then one has
$$\mu + J(\nu) \geq \Lambda > 0 \,.$$
- Finally note, $\mu + J(\nu) \geq \mu - |J|$ (exercise). Hence the energy condition is satisfied if the strict DEC, $\mu - |J| \geq c$ holds.

Proposition 5.2 (G. and Mendes [31])

Let Σ be a (closed) stable MOTS in a 3-dimensional initial data (M, g, K). Suppose there exists $c > 0$, such that $\mu + J(\nu) \geq c$ on Σ, where ν is the outward unit normal to Σ. Then the area of Σ satisfies,

$$A(\Sigma) \leq \frac{4\pi}{c} \, .$$

Moreover, if equality holds,

(1) Σ is a round 2-sphere, with Gaussian curvature $\kappa_\Sigma = c$,

(2) the null second fundamental form of Σ vanishes $\chi_+ = 0$, and

(3) $\mu + J(\nu) = c$ on Σ.

Proof.

- Apply the stability inequality: Σ stable $\Longrightarrow$

$$\int_\Sigma |\nabla\psi|^2 + \left(\frac{1}{2} S_\Sigma - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2\right) \psi^2 \geq 0 \, , \quad \forall \psi \in C^\infty(\Sigma)$$

Now set $\psi = 1$.

- Using $S_\Sigma = 2k_\Sigma$, we obtain

$$\int_\Sigma \left(\mu + J(\nu) + \frac{1}{2}|\chi|^2\right) dA \le \int_\Sigma k_\Sigma dA = 2\pi(2-2g) = 4\pi \quad (g=0).$$

- On the other hand, since $\mu + J(\nu) \ge c$,

$$\int_\Sigma \left(\mu + J(\nu) + \frac{1}{2}|\chi|^2\right) dA \ge \int_S c\, dA = cA(\Sigma)\,.$$

These two inequalities now imply $A(\Sigma) \le 4\pi/c$.

- Now assume $A(\Sigma) = 4\pi/c$. Then these inequalities combine to give,

$$\int_\Sigma \left(\mu + J(\nu) + \frac{1}{2}|\chi|^2\right) dA = 4\pi\,(= cA(\Sigma))\,,$$

or, equivalently,

$$\int_\Sigma \left((\mu + J(\nu) - c) + \frac{1}{2}|\chi|^2\right) dA = 0\,.$$

- Since $\mu + J(\nu) \ge c$ on Σ, this implies that $\mu + J(\nu) \equiv c$ and $\chi \equiv 0$.

- It remains to show that $k_\Sigma = c$. Since $\mu + J(\nu) = c$ and $\chi_+ = 0$, the symmetrized operator L_0 becomes,

$$L_0(\phi) = -\triangle\phi + (k_\Sigma - c)\phi$$

- Under our assumption that $A(\Sigma) = 4\pi/c$, check that the above inequalities imply $\int_\Sigma (k_\Sigma - c) dA = 0$.

  Consider then the Rayleigh formula for $\lambda_1(L_0)$,

$$\lambda_1(L_0) = \inf_{\psi \not\equiv 0} \frac{\int_\Sigma \left(|\nabla\psi|^2 + (k_\Sigma - c)\psi^2\, dA\right)}{\int_\Sigma \psi^2\, dA} \;\geq 0\,.$$

  We see that the minimum is achieved for $\psi = 1$. It follows that $\lambda_1(L_0) = 0$, and $\phi = 1$ is an associated eigenfunction of L_0. Setting $\phi = 1$ in

$$L_0(\phi) = -\triangle\phi + (k_\Sigma - c)\phi = 0$$

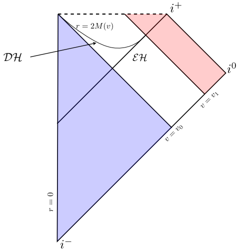
  we get that $k_\Sigma = c$.

The size of marginally outer trapped surfaces

Remarks.

- The Vaidya spacetime is an interesting example, which shows that there are stable MOTS that saturate the area inequality.

$$ds^2 = -\left(1 - \frac{2M(v)}{r}\right) dv^2 + 2dvdr + r^2 d\Omega^2\,, \qquad T = \frac{M'(v)}{4\pi r^2} dv \otimes dv$$



(∗) Vaidya may be viewed as a dynamical version of Schwarschild spacetime. It is a spherically symmetric spacetime containing an incoming null fluid.

(∗) Inside the black hole, within the fluid there is a spherically symmetric dynamical horizon; see [8]. A dynamical horizon is a spacelike hypersurface foliated by MOTS. The MOTS are necessarily stable (as can be seen from the maximum principle for MOTS).

(∗) A computation shows that each MOTS saturates the area inequality.

Remarks (cont.)

- Proposition 5.2 is used in an essential way to establish local and global splitting results for area minimizing MOTS in certain initial data sets; see [31, 32].
- These results may be viewed as extensions to the spacetime setting of result of H. Bray, S. Brendle, and A. Neves [10] concerning area minimizing 2-spheres in Riemannian 3-manifolds with positive scalar curvature.

A different measure of size.

- With regard to our area bound, a sphere can, of course, be very long, and still have very small area.

- In Kip Thorne's formulation of his famous Hoop Conjecture [59] he argues that, during gravitational collapse, 'horizons' (e.g. MOTS) can't form if the length of the collapsing object is very long relative to its mass.
- This suggests a possible MOTS diameter bound.

A different measure of size.

- With regard to our area bound, a sphere can, of course, be very long, and still have very small area.

- In Kip Thorne's formulation of his famous Hoop Conjecture [59] he argues that, during gravitational collapse, 'horizons' (e.g. MOTS) can't form if the length of the collapsing object is very long relative to its mass.
- This suggests a possible MOTS diameter bound.

Theorem 5.3 (G., [30])

Let Σ be a stable MOTS in a 3-dimensional initial data set (M, g, K). Suppose there exists $c > 0$, such that $\mu + J(\nu) \geq c$ on Σ, where ν is the outward unit normal to Σ. Then the diameter of Σ satisfies,

$$\mathrm{diam}(\Sigma) \leq \frac{2}{\sqrt{3}} \cdot \frac{\pi}{\sqrt{c}} \, .$$

Some comments on the proof.

Fix points $p, q \in \Sigma$ such that $\mathrm{diam}(\Sigma) = d_h(p,q)$. Since Σ is stable, by the corollary to the MOTS stability inequality, $\lambda_1(L_0) \geq 0$, where L_0 is the symmetrized operator.

- Let $\psi > 0$ be an associated eigenfunction:

$$L_0(\psi) = -\triangle\psi + (\kappa - (\mu + J(\nu)) - \frac{1}{2}|\chi_+|^2)\psi = \lambda_1(L_0)\psi$$

Hence,

$$\triangle\psi = (\kappa - P)\psi$$

where

$$P = (\mu + J(\nu)) + \frac{1}{2}|\chi_+|^2 + \lambda_1(L_0) \geq c\,.$$

- Now consider Σ in the metric $\tilde{h} = \psi^2 h$. The Gaussian curvature of $(\Sigma, \tilde{h})$ is given by,

$$\tilde{\kappa} = \psi^{-2}\kappa - \psi^{-3}\triangle\psi + \psi^{-4}|\nabla\psi|^2\,.$$

Substituting for $\triangle\psi$ gives

$$\tilde{\kappa} = \psi^{-2}(P + \psi^{-2}|\nabla\psi|^2)\,, \qquad (*)$$

- Let γ be a minimal geodesic from p to q in the metric $\tilde{h}$. Then by Synge's formula for the 2nd variation arc length,

$$\int_0^{\tilde{\ell}} \left(\frac{df}{d\tilde{s}}\right)^2 - \tilde{\kappa} f^2 \, d\tilde{s} \geq 0\,, \quad \text{for all } f \text{ vanishing at the endpoints}$$

where $\tilde{s}$ is $\tilde{h}$-arc length along γ and $\tilde{\ell}$ is the $\tilde{h}$-length of γ.

- Making the change of variable $s = s(\tilde{s})$, where s is h-arc length along γ ($d\tilde{s} = \psi ds$), and using the expression for $\tilde{\kappa}$, this becomes

$$\int_0^{\ell} \psi^{-1}(f')^2 - (P + \psi^{-2}|\nabla\psi|^2)\psi^{-1} f^2 \, ds \geq 0\,.$$

- Setting $k = \psi^{-1/2} f$, after a small computation and completing the square, we arrive at

$$\int_0^{\ell} \frac{4}{3}(k')^2 - P\,k^2 \, ds \geq 0\,.$$

Setting $k = \sin\frac{\pi s}{\ell}$ and using $P \geq c$, we obtain, $\ell \leq \frac{2}{\sqrt{3}} \cdot \frac{\pi}{\sqrt{c}}$.

<u>*Comments.*</u>

- Kip Thorne's Hoop Conjecture is actually an 'if and only if' statement. The statement of the conjecture is somewhat vague, but very roughly it says:

  Horizons form when, and only when, there is a sufficient concentration of mass in a region of fixed size, as measured by a "suitable notion of radius".

- There have a number of contributions to the Hoop conjecture, we mention in particular,

  - Schoen and Yau, *The existence of a black hole due to condensation of matter*, [57, CMP, 1983]

  and more recently,

  - Hirsch, Kazaras, Khuri, and Y. Zhang, *Spectral torical band inequalities and generalizations of the Schoen-Yau black hole existence*, [39, IMRN].

- Whereas the results of Schoen and Yau and of Hirsch et al. are important contributions to the "when" direction, our diameter estimate seems relevant to the "only when" direction.

Black hole topology and initial data rigidity

Let us recall the statement of our black hole topology theorem.

Theorem 6.1 (G. and Schoen [35])

Let (M, g, K) be an n-dimensional initial data set, $n \geq 3$, satisfying the dominant energy condition (DEC), $\mu \geq |J|$. If Σ is a stable MOTS in M then (apart from certain exceptional circumstances) Σ must be of positive Yamabe type, i.e. must admit a metric of positive scalar curvature.

- As we now discuss, the "apart from exceptional circumstances" can be removed if we assume that Σ is "weakly outermost".
- Recall, we say a MOTS Σ is *weakly outermost* provided there are no outer trapped ($\theta < 0$) surfaces outside of, and homologous, to Σ.

  As previously discussed, "weakly outermost" $\implies$ stable. Moreover, under natural circumstances, cross sections of the event horizon are weakly outermost. (Outer trapped surfaces cannot occur outside the event horizon.)

As a starting point we reinterpret our black hole topology theorem as an "infinitesimal" rigidity result:

Theorem 6.2 (Infinitesimal rigidity)

Let (M, g, K) be an n-dimensional, $n \geq 3$, initial data set satisfying the DEC, $\mu \geq |J|$. If Σ is a stable MOTS in M that does not admit a metric of positive scalar curvature then:

(a) $\chi_\Sigma^+ = 0$.

(b) Σ *is Ricci flat, and*

(c) $\mu + J(\nu) = 0$ *along* $\Sigma \quad (\implies \mu = |J|)$.

From this we are able to obtain a local splitting result.

The proof uses this infinitesimal rigidity in a number of ways. A key fact is that it implies $\lambda_1(L) = 0$. $\quad (0 = \lambda_1(L_0) \geq \lambda_1(L) \geq 0.)$

A local rigidity result.

Theorem 6.3 (G. [27, 2008], [28, 2018])

Let (M, g, K) be an n-dimensional, $n \geq 3$, initial data set satisfying the DEC, $\mu \geq |J|$. Suppose Σ is a weakly outermost MOTS in M that does not admit a metric of positive scalar curvature.

Then there exists an outer neighborhood $U \cong [0, \delta) \times \Sigma$ of Σ in M foliated by MOTS, i.e. each leaf $\Sigma_t = \{t\} \times \Sigma$ is a MOTS. In fact,

(a) $\chi^+_{\Sigma_t} = 0$.

(b) Each Σ_t is Ricci flat.

(c) $\mu + J(\nu_t) = 0$ *along each* Σ_t $(\implies \mu = |J|)$

Thus, if Σ is an *outermost MOTS* (i.e, if there are no outer trapped, or marginally outer trapped, surfaces outside of and homologous to Σ) then Σ must be of positive Yamabe type *without exception.*

Sketch of the proof.

- **Claim.** An outer neighborhood U of Σ is foliated by surfaces of constant null expansion, i.e. there exists $U \approx [0, \epsilon) \times \Sigma$ such that

$$\theta(t) = \text{null expansion of } \Sigma_t = \{t\} \times \Sigma,$$

is constant for each $t \in [0, \epsilon)$.

Idea: Consider the null mean curvature operator $f \to \Theta(f)$, where

$$\Theta(f) = \text{null mean curvature of } \Sigma_f = \text{graph } f \text{ over } \Sigma$$

  * Θ has linearization at $f = 0$,

$$\Theta'(0) = L\,,$$

where L is the MOTS stability operator.

  * Since $\lambda_1(L) = 0$, and is simple, $\Theta'(0)$ has a one dimensional kernel. Then an inverse function theorem argument leads to the desired neighborhood.

- Σ weakly outermost $\implies$ $\theta(t) \geq 0\ \forall t \in [0, \epsilon)$.
  WTS: $\theta(t) = 0\ \forall t \in [0, \epsilon)$.

- A computation shows,

$$\frac{d\theta}{dt} = L_t(\phi) + \theta\tau - \frac{1}{2}\theta^2$$

  where, for each $t \in [0, \varepsilon)$, L_t is the MOTS stability operator on Σ_t acting on the lapse function $\phi = \phi_t$.

- For simplicity let us agree to forget the '$\theta\tau$' term, e.g. take $\tau = 0$. Then:

$$\frac{d\theta}{dt} = L_t(\phi) - \frac{1}{2}\theta^2 \qquad (*)$$

- We have $\theta(0) = 0$ and $\theta(t) \geq 0$. Suppose $\frac{d\theta}{dt} > 0$ for some t. Then $L_t(\phi) > 0$, which implies $\lambda_1(L_t) > 0$ (see [5]). But, by previous arguments, this would imply that $\Sigma_t \approx \Sigma$ carries a metric positive scalar curvature. $\rightarrow\leftarrow$.

- Hence we have $\theta(0) = 0$, $\theta(t) \geq 0$ and $\frac{d\theta}{dt} \leq 0 \implies \theta(t) = 0$, i.e. each Σ_t is a MOTS. Moreover $(*)$ implies $L_t(\phi) = 0$, which in turn implies that $\lambda_1(L_t) \geq 0$ (see [5]).

- Hence each Σ_t is a stable MOTS that does not carry a metric of positive scalar curvature. Now apply our infinitesimal rigidity result.

  Finally you can make things work without dropping the '$\theta\tau$' term.

A global rigidity result.

We now present a *global* initial data rigidity result obtained with Michael Eichmair and Abraão Mendes [20]. For this, we were motivated in part by the spacetime positive mass theorem.

We have the following beautiful result of Eichmair, Huang, Lee and Schoen [22]

Theorem 6.4 (Spacetime PMT, EHLS)

Let (M, g, K) be an n-dimensional, $3 \leq n \leq 7$, asymptotically flat initial data set with ADM energy-momentum vector (E, P). If the DEC, $\mu \geq |J|$, is satisfied, then $E \geq |P|$.

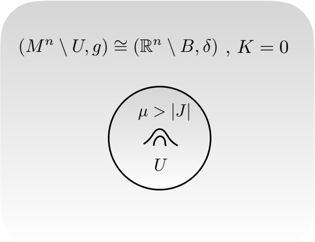
In very broad terms, it generalizes to the spacetime setting the proof of the Riemannian PMT of Schoen and Yau, where now MOTS play a role analogous to minimal surfaces in the Schoen-Yau proof. (Dimension restriction comes from existence and regularity results for MOTS.)

More recent works have considered the equality case, $E = |P|$, and the case with boundary, $\partial M \neq \emptyset$; see e.g. [43, Huang-Lee II] and references therein.

In the arXiv preprint: *arXiv:1612.07505*, Lohkamp presented an approach to proving the Spacetime PMT, somewhat analogous to his approach to proving the Riemannian PMT.

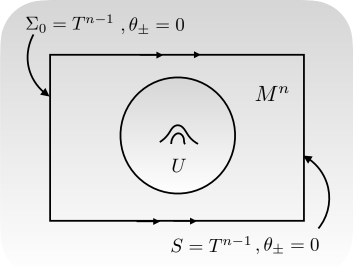
Proceeding by contradiction, his approach reduces to proving the following:

Nonexistence of $\mu - |J| > 0$ - islands: Let (M, g, K) be an initial data set that is isometric to Euclidean space, with $K = 0$, outside some bounded open set U. Then one cannot have $\mu > |J|$ on U.

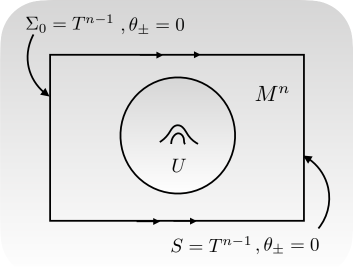


With Eichmair and Mendes, we established a "rigid" version of this no-island result in dimensions $3 \leq n \leq 7$.

- Under Lohkamp's assumptions we can compactify to obtain a compact manifold M with boundary $\partial M = \Sigma_0 \sqcup S$.



- Under Lohkamp's assumptions we can compactify to obtain a compact manifold M with boundary $\partial M = \Sigma_0 \sqcup S$.

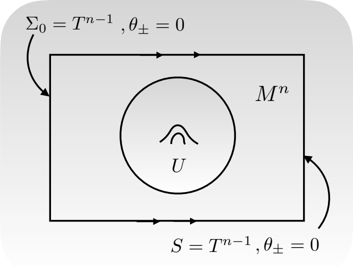


- In our result we assume that Σ_0 satisfies the cohomology condition, i.e. that there exist $\omega_1, \ldots, \omega_{n-1} \in H^1(\Sigma_0, \mathbb{Z})$ such that

$$\omega_1 \smile \cdots \smile \omega_{n-1} \neq 0$$

(Schoen-Yau [58], Dan Lee [49])

- Under Lohkamp's assumptions we can compactify to obtain a compact manifold M with boundary $\partial M = \Sigma_0 \sqcup S$.



- In our result we assume that Σ_0 satisfies the cohomology condition, i.e. that there exist $\omega_1, \ldots, \omega_{n-1} \in H^1(\Sigma_0, \mathbb{Z})$ such that

$$\omega_1 \smile \cdots \smile \omega_{n-1} \neq 0$$

(Schoen-Yau [58], Dan Lee [49])

- We also assume that M satisfies the homotopy condition with respect to Σ_0. (This is slightly more general than assuming there is a retract of M on to Σ_0.)

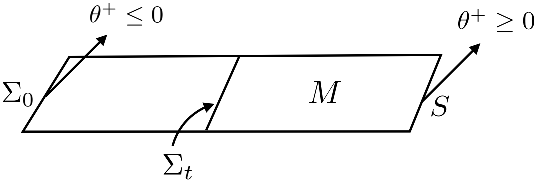
Theorem 6.5 (Eichmair, G., Mendes; CMP, 2021)

Let (M^n, g, K) be an n-dim'l, $3 \le n \le 7$, compact-with-boundary initial data set, which satisfies the DEC. Assume the boundary ∂M can be expressed as a disjoint union of hypersurfaces, $\partial M = \Sigma_0 \sqcup S$, such that the following conditions hold:

(i) *$\theta^+ \le 0$ along Σ_0 wrt the normal pointing into M, and $\theta^+ \ge 0$ along S wrt the normal pointing out of M,*

(ii) *Σ_0 satisfies the cohomology condition and M satisfies the homotopy condition with respect to Σ_0.*

Then $M \cong [0, \ell] \times \Sigma_0$, and (a) each leaf $\Sigma_t \cong \{t\} \times \Sigma_0$ is a flat torus, (b) $\chi^+_{\Sigma_t} = 0$ and (c) $\mu + J(\nu_t) = 0$; in particular $\mu = |J|$ on M.



Comments on the proof. The proof consists of three elements:

(1) Key lemma: Showing that Σ_0 is a weakly outermost MOTS in M.

- This is where the cohomology condition and homotopy condition come in.
- It also makes use of the basic existence theory for MOTS as developed by Schoen, Yau, Andersson, Metzger and Eichmair. (Existence based on 'blow-up' of Jang's equation.)

(2) Now use the local rigidity result for weakly outermost MOTS to obtain a neighborhood $U \cong [0, \delta) \times \Sigma_0$ on which the conclusions (a), (b), and (c) of the theorem hold. (By Bochner, see e.g. [54, Cor. 9.2.5], Σ_t Ricci flat + cohomology condition $\implies$ Σ_t is a flat torus.

(3) Extend to all of M. Use compactness results to show that the foliation $\{\Sigma_t\}$ extends to $t = \delta$. (Since $\chi_{\Sigma_t} = 0$, the 2nd FF forms of the Σ_t's within M are uniformly controlled by K.)

Now continue, and use the MOTS maximum principle when reaching S.

Weakly Outermost Lemma

Under the assumptions of the theorem, Σ_0 is a weakly outermost MOTS.

Comment on the proof:

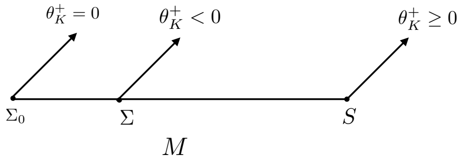
- Assume for the moment that Σ_0 is a MOTS. Suppose Σ_0 is not weakly outermost. Then there exists an outer trapped surface Σ homologous to Σ_0:

Weakly Outermost Lemma

Under the assumptions of the theorem, Σ_0 is a weakly outermost MOTS.

Comment on the proof:

- Assume for the moment that Σ_0 is a MOTS. Suppose Σ_0 is not weakly outermost. Then there exists an outer trapped surface Σ homologous to Σ_0:

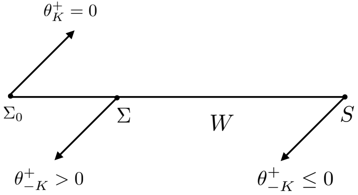


Weakly Outermost Lemma

Under the assumptions of the theorem, Σ_0 is a weakly outermost MOTS.

Comment on the proof:

- Assume for the moment that Σ_0 is a MOTS. Suppose Σ_0 is not weakly outermost. Then there exists an outer trapped surface Σ homologous to Σ_0:

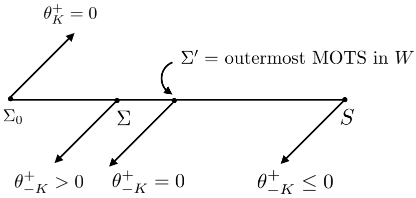


Weakly Outermost Lemma

Under the assumptions of the theorem, Σ_0 is a weakly outermost MOTS.

Comment on the proof:

- Assume for the moment that Σ_0 is a MOTS. Suppose Σ_0 is not weakly outermost. Then there exists an outer trapped surface Σ homologous to Σ_0:

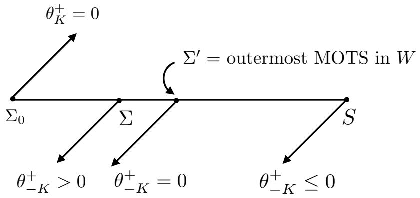


Weakly Outermost Lemma

Under the assumptions of the theorem, Σ_0 is a weakly outermost MOTS.

Comment on the proof:

- Assume for the moment that Σ_0 is a MOTS. Suppose Σ_0 is not weakly outermost. Then there exists an outer trapped surface Σ homologous to Σ_0:



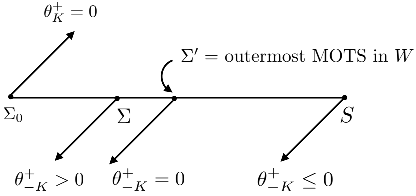
- The cohomology and homotopy conditions imply that Σ' does not carry a metric of postive scalar curvature. But then the local rigidity result implies that Σ' isn't outermost $\rightarrow\leftarrow$.

Weakly Outermost Lemma

Under the assumptions of the theorem, Σ_0 is a weakly outermost MOTS.

Comment on the proof:

- Assume for the moment that Σ_0 is a MOTS. Suppose Σ_0 is not weakly outermost. Then there exists an outer trapped surface Σ homologous to Σ_0:



- The cohomology and homotopy conditions imply that Σ' does not carry a metric of postive scalar curvature. But then the local rigidity result implies that Σ' isn't outermost $\rightarrow\leftarrow$.
- Finally, Σ_0 must be MOTS: If not, then we have $\theta^+ \leq 0$, and $\theta^+ < 0$ somewhere. In this case one can use null mean curvature flow to perturb Σ_0 to a strictly outer trapped surface (Andersson and Metzger, CMP (2009)).

MOTS-based results for general initial data sets have been shown to be useful in obtaining some purely Riemannian results. We mention a couple of examples.

- By imposing a certain convexity condition on K, Theorem 6.5 can be used to establish stronger rigidity; see [20, Theorem 1.3]. As a corollary, by setting $K = -\epsilon g$, $\epsilon = 0, 1$ we obtain the following scalar curvature rigidity result.

Theorem 6.6 ([20])

Let (M^n, g), be an n-dimensional, $3 \le n \le 7$, compact Riemannian manifold with boundary, with scalar curvature $R \ge -n(n-1)\epsilon$, where $\epsilon = 0$ or $\epsilon = 1$.

Assume the boundary ∂M can be expressed as a disjoint union of hypersurfaces, $\partial M = \Sigma_0 \sqcup S$, such that the following hold.

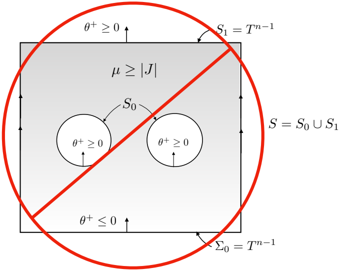
(i) *The mean curvature of Σ_0 satisfies $H \le (n-1)\epsilon$ wrt the inward pointing normal, and the mean curvature of S satisfies $H \ge (n-1)\epsilon$ wrt the outward pointing normal.*

(ii) *Σ_0 satisfies the cohomology condition and M satisfies the homotopy condition with respect to Σ_0.*

Then (M, g) is isometric to $([0, \ell] \times \Sigma_0, dt^2 + e^{2\epsilon t} g_0)$, where (Σ_0, g_0) is a flat torus.

- In the case $\epsilon = 1$, this theorem immediately yields the hyperbolic space rigidity result obtained by Andersson-Cai-G. [1, Theorem 1.1].

- Theorem 6.5, rules out certain compact initial data sets with multiple weakly outer untrapped boundary components.



Theorem 6.5 requires S to be connected.

- This was used with Piotr Chruściel to extend Chruściel and Delay's general version of the PMT for asymptotically hyperbolic manifolds (arXiv:1901.05263), to the case with boundary (with optimal boundary mean curvature); see [16].

For further related results on these topics, see, for example:

- "Initial data rigidity results", with Michael Eichmair and Abraão Mendes, [20].
- "Some Rigidity Results for compact initial data sets", with Abraão Mendes, [32].
- "Some rigidity results for charged initial data sets", with Abraão Mendes, [34].
- "Rigidity Aspects of Penrose's Singularity Theorem", with Eric Ling, [29].
- "Aspects of the geometry and topology of expanding horizons", with Abraão Mendes, [33].

References

[1] Lars Andersson, Ming-Liang Cai, and Gregory J. Galloway, *Rigidity and positivity of mass for asymptotically hyperbolic manifolds*, Ann. Henri Poincaré **9** (2008), no. 1, 1–33.

[2] L. Andersson, M. Dahl, G. J. Galloway, and D. Pollack, *On the geometry and topology of initial data sets with horizons*, 2015, arXiv:1508.01896.

[3] L. Andersson, M. Eichmair, and J. Metzger, *Jang's equation and its applications to marginally trapped surfaces*, Complex analysis and dynamical systems IV. Part 2, Contemp. Math., vol. 554, Amer. Math. Soc., Providence, RI, 2011, pp. 13–45.

[4] L. Andersson, M. Mars, and W. Simon, *Local existence of dynamical and trapping horizons*, Phys. Rev. Lett. **95** (2005), 111102.

[5] L. Andersson, M. Mars, and W. Simon, *Stability of marginally outer trapped surfaces and existence of marginally outer trapped tubes*, Adv. Theor. Math. Phys. **12** (2008), no. 4, 853–888.

[6] L. Andersson and J. Metzger, *The area of horizons and the trapped region*, Comm. Math. Phys. **290** (2009), no. 3, 941–972.

[7] L. Andersson and J. Metzger, *Curvature estimates for stable marginally trapped surfaces*, J. Diff. Geom. **84** (2010), no. 2, 231–265.

[8] A. Ashtekar and B. Krishnan, *Dynamical horizons and their properties*, Phys. Rev. D (3) **68** (2003), no. 10, 104030, 25.

[9] J. K. Beem, P. E. Ehrlich, and K. L. Easley, *Global Lorentzian geometry*, second ed., Monographs and Textbooks in Pure and Applied Mathematics, vol. 202, Marcel Dekker Inc., New York, 1996.

[10] H. Bray, S. Brendle, and A. Neves, *Rigidity of area-minimizing two-spheres in three-manifolds*, Comm. Anal. Geom. **18** (2010), no. 4, 821–830.

[11] A. N. Bernal and M. Sánchez, *Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes*, Comm. Math. Phys. **257** (2005), no. 1, 43–50.

[12] A. N. Bernal and M. Sánchez, *Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal'*, Classical Quantum Gravity **24** (2007), no. 3, 745–749.

[13] D. Christodoulou, *The formation of black holes in general relativity*, Geometry and analysis. No. 1, Adv. Lect. Math. (ALM), vol. 17, Int. Press, Somerville, MA, 2011, pp. 247–283.

[14] P. T. Chruściel, J. L. Costa, and M. Heusler, *Uniqueness and Beyond*, Living Rev. Relativ. **15** (2012) 7.

[15] P. T. Chruściel, E. Delay, G. J. Galloway, and R. Howard, *Regularity of horizons and the area theorem*, Ann. Henri Poincaré **2** (2001), no. 1, 109–178.

[16] P. T. Chruściel and G. J. Galloway, *Positive mass theorems for asymptotically hyperbolic riemannian manifolds with boundary*, Classical and Quantum Gravity **38** (2021), no. 23, 237001.

[17] P. T. Chruściel, G. J. Galloway, and D. Solis, *Topological censorship for Kaluza-Klein space-times*, Ann. Henri Poincaré **10** (2009), no. 5, 893–912.

[18] P. T. Chruściel, J. D. E. Grant, and E. Minguzzi, *On Differentiability of Volume Time Functions*, Annales Henri Poincare **17** (2016), no. 10, 2801–2824.

[19] M. Eichmair, *Existence, regularity, and properties of generalized apparent horizons*, Comm. Math. Phys. **294** (2010), no. 3, 745–760.

[20] Michael Eichmair, Gregory J. Galloway, and Abraão Mendes, *Initial data rigidity results*, Comm. Math. Phys. **386** (2021), no. 1, 253–268.

[21] M. Eichmair, G. J. Galloway, and D. Pollack, *Topological censorship from the initial data point of view*, J. Differential Geom. **95** (2013), no. 3, 389–405.

[22] Michael Eichmair, Lan-Hsuan Huang, Dan A. Lee, and Richard Schoen, *The spacetime positive mass theorem in dimensions less than eight*, J. Eur. Math. Soc. (JEMS) **18** (2016), no. 1, 83–121.

[23] R. Emparan and H. S. Reall, *A rotating black ring solution in five dimensions*, Phys. Rev. Lett. **88** (2002), no. 10, 101101, 4.

[24] J. L. Friedman, K. Schleich, and D. M. Witt, *Topological censorship*, Phys. Rev. Lett. **71** (1993), no. 10, 1486–1489.

[25] G. J. Galloway, *Notes on Lorentzian causality*, ESI-EMS-IAMP Summer School on Mathematical Relativity (available at: `http://www.math.miami.edu/~galloway/`).

[26] G. J. Galloway, *On the topology of the domain of outer communication*, Classical Quantum Gravity **12** (1995), no. 10, L99–L101.

[27] G. J. Galloway, *Rigidity of marginally trapped surfaces and the topology of black holes*, Comm. Anal. Geom. **16** (2008), no. 1, 217–229.

[28] G. J. Galloway, *Rigidity of outermost MOTS: the initial data version*, Gen. Relativity Gravitation **50** (2018), no. 3, Art. 32, 7.

[29] G. J. Galloway and Eric Ling, *Aspects of the geometry and topology of expanding horizons*, Proc. Am. Math. Soc. **153** (2025), 1239–1250.

[30] G. J. Galloway, *Remarks on the size of apparent horizons*, Lett. Math. Phys. **113** (2023), no. 6, Paper No. 118, 10.

[31] G. J. Galloway and Abraão Mendes, *Rigidity of marginally outer trapped 2-spheres*, Commun. Anal. Geom. **26** (2018), no. 1, 63–83 (English).

[32] G. J. Galloway and Abraão Mendes, *Some rigidity results for compact initial data sets*, Trans. Amer. Math. Soc. **377** (2024), no. 3, 1989–2007 (English).

[33] G. J. Galloway and A. Mendes, *Aspects of the geometry and topology of expanding horizons*, Proc. Am. Math. Soc. **153** (2025), no. 3, 1239–1250.

[34] G. J. Galloway and Abraão Mendes, *Some rigidity results for charged initial data sets*, Nonlinear Anal. **256** (2025), Paper No. 113780, 9.

[35] G. J. Galloway and R. Schoen, *A generalization of Hawking's black hole topology theorem to higher dimensions*, Comm. Math. Phys. **266** (2006), no. 2, 571–576.

[36] D. Gannon, *Singularities in nonsimply connected space-times*, J. Mathematical Phys. **16** (1975), no. 12, 2364–2367.

[37] R. Geroch, *Domain of dependence*, J. Mathematical Phys. **11** (1970), 437–449.

[38] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, Cambridge University Press, London, 1973, Cambridge Monographs on Mathematical Physics, No. 1.

[39] S. Hirsch, D. Kazaras, M. Khuri, and Y. Zhang, *Spectral torical band inequalities and generalizations of the Schoen-Yau black hole existence theorem*, Int. Math. Res. Not., to appear, arXiv:2301.08270.

[40] S. Hollands and S. Yazadjiev, *Uniqueness theorem for 5-dimensional black holes with two axial Killing fields*, Comm. Math. Phys. **283** (2008), no. 3, 749–768.

[41] G. Horowitz (ed.), *Black holes in higher dimensions*, Cambridge University Press, London, 2012.

[42] R. A. Hounnonkpe and E. Minguzzi, *Globally hyperbolic spacetimes can be defined without the 'causal' condition*, Classical Quantum Gravity **36** (2019), no. 19, 197001, 9.

[43] L.-H. Huang and D. A. Lee, *Equality in the spacetime positive mass theorem II*, Calc. Var. Part. Differ. Equ **64** (2025), no. 3, 92.

[44] J. Kazdan and F. Warner, *Prescribing curvatures*, Proc. Symp. in Pure Math. **27** (1975) 309–319.

[45] M. A. Khuri and J. F. Rainone, *Black lenses in kaluza-klein matter*, Phys. Rev. Lett. **131** (2023), 041402.

[46] A. Ionescu and S. Klainerman, *Rigidity results in general relativity: a review*, 2015, arXiv:1501.01587.

[47] S. Klainerman, J. Luk, and I. Rodnianski, *A fully anisotropic mechanism for formation of trapped surfaces in vacuum*, 2013, arXiv:1302.5951v1.

[48] C. W. Lee, *A restriction on the topology of Cauchy surfaces in general relativity*, Comm. Math. Phys. **51** (1976), no. 2, 157–162.

[49] D. A. Lee, *Geometric relativity*, Graduate Studies in Mathematics, vol. 201, American Mathematical Society, Providence, RI, 2019.

[50] E. Minguzzi, *Lorentzian causality theory*, Living Rev. Relativ. **22** (2019) 22.

[51] R. C. Myers and M. J. Perry, *Black holes in higher-dimensional space-times*, Ann. Physics **172** (1986), no. 2, 304–347.

[52] B. O'Neill, *Semi-Riemannian geometry*, Pure and Applied Mathematics, vol. 103, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983, With applications to relativity.

[53] R. Penrose, *Techniques of differential topology in relativity*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1972, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 7.

[54] Peter Petersen, *Riemannian geometry*, 3rd ed., Graduate Texts in Mathematics, Springer International Publishing, 2016.

[55] R. Schoen and S.-T. Yau, *On the structure of manifolds with positive scalar curvature*, Manuscripta Math. **28** (1979), no. 1-3, 159–183.

[56] R. Schoen and S.-T. Yau, *Proof of the positive mass theorem. II*, Comm. Math. Phys. **79** (1981), no. 2, 231–260.

[57] R. Schoen and S.-T. Yau, *The existence of a black hole due to condensation of matter*, Comm. Math. Phys. **90** (1983), no. 4, 575–579.

[58] R. Schoen and S.-T. Yau, *Positive scalar curvature and minimal hypersurface singularities*, Surveys in differential geometry 2019. Differential geometry, Calabi-Yau theory, and general relativity. Part 2, Surv. Differ. Geom., vol. 24, Int. Press, Boston, MA, [2022] ©2022, pp. 441–480.

[59] K. S. Thorne, *Nonspherical Gravitational Collapse - A Short Review*, in Magic Without Magic: John Archibald Wheeler , ed. J. Klauder (W.H. Freeman & Co., San Francisco, 1972), pp. 231-258.

[60] R. M. Wald, *General relativity*, University of Chicago Press, Chicago, IL, 1984.