

# On the Geometry and Topology of Initial Data Sets in General Relativity

IHP Summer School on Mathematical General Relativity

September 14-18, 2015

Gregory J. Galloway  
Department of Mathematics  
University of Miami

September 19, 2015

# Contents

<b>1</b>	<b>Lorentzian Causality</b>	<b>4</b>
1.1	Futures and Pasts . . . . .	9
1.2	Achronal Boundaries . . . . .	14
1.3	Causality conditions and Global Hyperbolicity . . . . .	18
1.4	Domains of Dependence . . . . .	26
<b>2</b>	<b>The Geometry of Null Hypersurfaces</b>	<b>30</b>
<b>3</b>	<b>The Penrose Singularity Theorem and Related Results</b>	<b>45</b>
<b>4</b>	<b>The Topology of Black Holes</b>	<b>53</b>
4.1	Marginally Outer Trapped Surfaces . . . . .	59
4.2	A Generalization of Hawking's Black Hole Topology Theorem . . . . .	65
<b>5</b>	<b>The Geometry and Topology of the Exterior Region</b>	<b>70</b>

## Resources for Causal Theory

### References

- [1] G. Galloway, *Notes on Lorentzian causality*, ESI-EMS-IAMP Summer School on Mathematical Relativity (available at: <http://www.math.miami.edu/~galloway/>).
- [2] B. O'Neill, *Semi-Riemannian geometry*, Pure and Applied Mathematics, vol. 103, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.
- \*\*\*\*\*
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Robert M. Wald, *General relativity*, University of Chicago Press, Chicago, IL, 1984.

# 1 Lorentzian Causality

## Lorentzian manifolds

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

$$g : T_p M \times T_p M \rightarrow \mathbb{R} \quad (1.1)$$

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

$$[g(e_i, e_j)] = \text{diag}(-1, +1, \dots, +1). \quad (1.2)$$

Example: Minkowski space, the spacetime of Special Relativity. Minkowski space is  $\mathbb{R}^{n+1}$ , equipped with the Minkowski metric  $\eta$ : 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  $\mathbb{R}^{n+1}$ ),

$$\eta(X, Y) = \eta_{ij} X^i X^j = -X^0 Y^0 + \sum_{i=1}^n X^i Y^i. \quad (1.3)$$

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

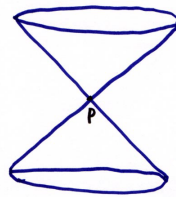
## Causal character of vectors.

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  $X \in T_p M$  forms a double cone  $\mathcal{V}_p$  in the tangent space  $T_p M$ :



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

*Time orientability.* 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*.

There are various ways to make the phrase “continuous assignment” precise (see e.g., O’Neill [22, 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 time-like 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.*

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

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

*Lorentzian inequalities.* If  $X \in T_p M$  is causal ( $g(X, X) \leq 0$ ), we 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  $X, Y \in T_p M$ ,*

$$|g(X, Y)| \geq |X||Y| \tag{1.4}$$

(2) *(Reverse triangle inequality) For all causal vectors  $X, Y$  that point into the same half cone of the null cone at  $p$ ,*

$$|X + Y| \geq |X| + |Y|. \tag{1.5}$$

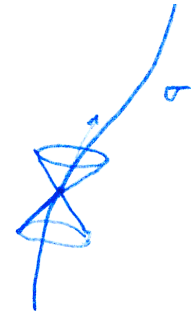
*Comment on the proof:* Note (1.4) trivially holds if  $X$  is null. For  $X$  timelike, decompose  $Y$  as  $Y = \lambda X + Y^\perp$ , where  $Y^\perp$  (the component of  $Y$  perpendicular to  $X$ ) is necessarily spacelike. Inequality (1.5) follows easily from (1.4).

The reverse triangle inequality is the source of the twin paradox.

## Causal character of curves.

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

$\gamma$  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 at each point. Heuristically, in accordance with relativity, information flows along causal curves, and so such curves are the focus of our attention. The notion of a causal curve extends in a natural way to piecewise smooth curves.

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

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

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



## 1.1 Futures and Pasts

Let  $(M, g)$  be a spacetime. A timelike (resp. causal) curve  $\gamma : I \rightarrow M$  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  $<$ :

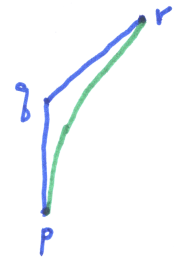
**Definition 1.1.** For  $p, q \in M$ ,

- (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 (nontrivial) 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,

if  $p \ll q$  and  $q < r$  then  $p \ll r$ .



This and related statements are a consequence of the following fundamental causality result; see O'Neill [22, p. 294] for a careful proof.

**Proposition 1.2.** *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  $\gamma$  from  $p$  to  $q$  must be a null geodesic (when suitably parameterized).*

Now introduce standard causal notation:

**Definition 1.2.** *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,*

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

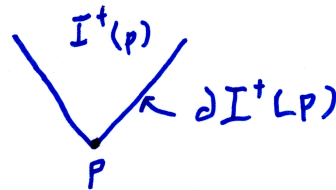
Hence,  $I^+(p)$  consists of all points in  $M$  that can be reached from  $p$  by a future directed timelike curve, and  $J^+(p)$  consists of the point  $p$  and all points in  $M$  that can be reached from  $p$  by a future directed causal curve.

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, Proposition 1.2 becomes:

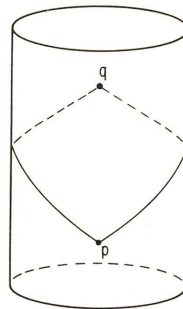
**Proposition 1.2** *If  $q \in J^+(p) \setminus I^+(p)$  ( $q \neq p$ ) then there exists a future directed null geodesic from  $p$  to  $q$ .*

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  $p$ .



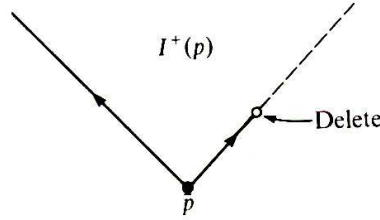
We note, however, that curvature and topology can drastically change the structure of ‘null cones’ in spacetime.

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 [11]). 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

$$I^+(S) = \bigcup_{p \in S} I^+(p) = \{q \in M : p \ll q \text{ for some } p \in S\} \quad (1.6)$$

$$J^+(S) = \bigcup_{p \in S} J^+(p) = \{q \in M : p \leq q \text{ for some } p \in S\}. \quad (1.7)$$

Note:

- (1)  $S \subset J^+(S)$ .
- (2)  $I^+(S)$  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 1.3.** *For all subsets  $S \subset M$ ,*

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

(2)  $J^+(S) \subset \overline{I^+(S)}$ .

## 1.2 Achronal Boundaries

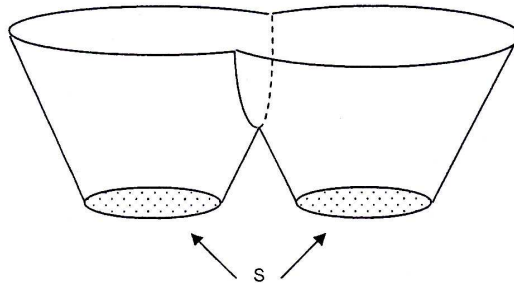
Achronal sets play an important role in causal theory.

**Definition 1.3.** 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 1.4.** 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 1.4.** An achronal boundary  $\partial I^+(S)$ , if nonempty, is a closed achronal  $C^0$  hypersurface in  $M$ .

We make a few remarks about the proof.

**Lemma 1.5.** *If  $p \in \partial I^+(S)$  then  $I^+(p) \subset I^+(S)$ , and  $I^-(p) \subset M \setminus \overline{I^+(S)}$ .*

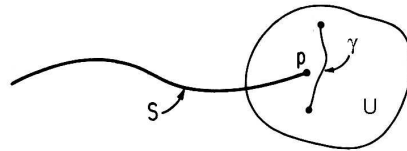
*Proof.* To prove the first part of the lemma, note that if  $q \in I^+(p)$  then  $p \in I^-(q)$ , and hence  $I^-(q)$  is a neighborhood of  $p$ . Since  $p$  is on the boundary of  $I^+(S)$ , it follows that  $I^-(q) \cap I^+(S) \neq \emptyset$ , and hence  $q \in I^+(S)$ . The second part of the lemma can be proved similarly.  $\square$

**Claim 1:** An achronal boundary  $\partial I^+(S)$  is achronal.

*Proof of the claim:* Suppose there exist  $p, q \in \partial I^+(S)$ , with  $q \in I^+(p)$ . By Lemma 1.5,  $q \in I^+(S)$ . But since  $I^+(S)$  is open,  $I^+(S) \cap \partial I^+(S) = \emptyset$ , contradicting  $q \in \partial I^+(S)$ .  $\square$

Lemma 1.5 also implies that achronal boundaries are *edgeless*. We need to introduce the edge concept.

**Definition 1.5.** *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  $\gamma$  from  $I^-(p, U)$  to  $I^+(p, U)$  that does not meet  $S$  (see the figure).*



We denote by *edge  $S$*  the set of edge points of  $S$ . We note that

$$\overline{S} \setminus S \subset \text{edge } S \subset \overline{S}$$

If *edge  $S = \emptyset$*  we say that  $S$  is *edgeless*.

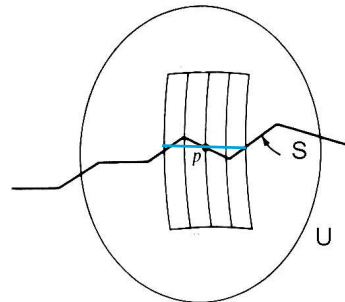
**Claim 2:** An achronal boundary  $\partial I^+(S)$  is edgeless.

*Proof of the claim:* Lemma 1.5 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 edgless.  $\square$

Proposition 1.4 now follows from the following basic result.

**Proposition 1.6.** *An edgeless achronal set  $S$ , if nonempty, is a  $C^0$  hypersurface in  $M$ .*

*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:

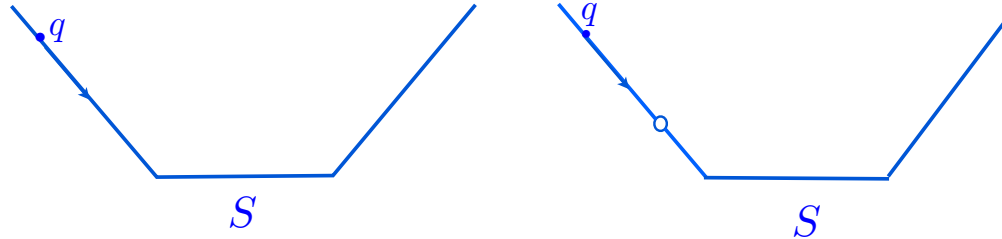


One can further use the achronality of  $S$  to show that the graphing function is continuous. See O'Neill [22, p. 413] for details.



The next result shows that, in general, large portions of achronal boundaries are ruled by null geodesics. A future (resp., past) directed causal curve  $\gamma : (a, b) \rightarrow M$  is said to be **future (resp., past) inextendible** in  $M$  if  $\lim_{t \rightarrow b^-} \gamma(t)$  does not exist.

**Proposition 1.7.** *Let  $S \subset M$  be closed. Then each  $q \in \partial I^+(S) \setminus S$  lies on a null geodesic contained in  $\partial I^+(S)$ , which either has a past end point on  $S$ , or else is past inextendible in  $M$ .*



*Proof.* See Gal-ESI [11] for the proof. Here we consider a special case: Suppose that  $J^+(S)$  is closed. Then

$$\partial I^+(S) = \overline{I^+(S)} \setminus \text{int } I^+(S) = J^+(S) \setminus I^+(S).$$

Hence for any  $q \in \partial I^+(S)$  there exists a point  $p \in S$  such that  $q \in J^+(p) \setminus I^+(p)$ . Then by Proposition 1.2, there exist a future directed null geodesic  $\eta$  from  $p$  to  $q$ ;  $\eta$  can't enter the timelike future of  $S$ , and hence must be contained in  $\partial I^+(S)$ .  $\square$

### 1.3 Causality conditions and Global Hyperbolicity

A number of results in Lorentzian geometry and general relativity require some sort of causality condition. It is perhaps natural on physical grounds to rule out the occurrence of closed timelike curves. Physically, the existence of such a curve signifies the existence of an observer who is able to travel into his/her own past, which leads to variety of paradoxical situations.

**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.

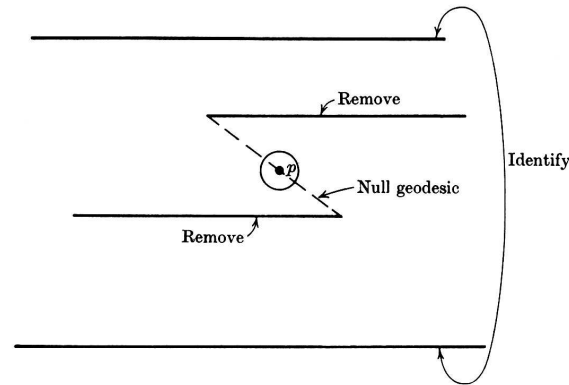
**Proposition 1.8.** *Every compact spacetime contains a closed timelike curve.*

*Proof.* The sets  $\{I^+(p); p \in M\}$  form an open cover of  $M$  from which we can abstract a finite subcover:  $I^+(p_1), I^+(p_2), \dots, I^+(p_k)$ . 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  $p_1$ .  $\square$

A somewhat stronger condition than the chronology condition is the

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

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 1.6.** *An open set  $U$  in spacetime  $M$  is said to be causally convex provided no causal curve in  $M$  meets  $U$  in a disconnected set.*

**Definition 1.7.** *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. In fact strong causality fails at all points along the dashed null geodesic. 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.

This is the “standard” causality condition in spacetime geometry, and, although there are even stronger causality conditions, it is sufficient for most applications.

The following lemma is often useful.

**Lemma 1.9.** *Suppose strong causality holds at each point of a compact set  $K$  in a spacetime  $M$ . If  $\gamma : [0, b) \rightarrow M$  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  $t \in [t_0, b)$ .*

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 (that, generically, spacetime solutions to the Einstein equations do not admit *naked singularities*).

**Definition 1.8.** *A spacetime  $M$  is said to be globally hyperbolic provided*

(1)  *$M$  is strongly causal.*

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

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

Ex. Minkowski space is globally hyperbolic but the spacetime obtained by removing one point from it is not.

We consider a few basic consequences of global hyperbolicity.

**Proposition 1.10.** *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  $A, B \subset M$ .*

Analogously to the case of Riemannian geometry, one can learn much about the global structure of spacetime by studying its causal geodesics. Global hyperbolicity is the standard condition in Lorentzian geometry that ensures the existence of maximal causal geodesic segments.

**Theorem 1.11.** *Let  $M$  be a globally hyperbolic spacetime. If  $q \in J^+(p)$  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).*

See Gal-ESI [11] for discussions of the proofs.

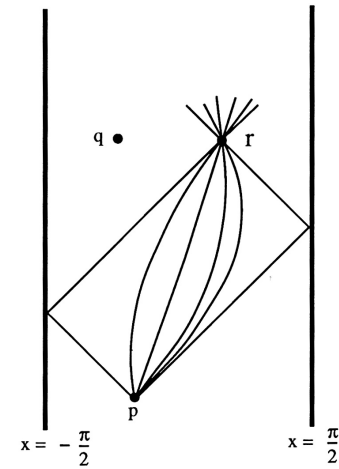
Remarks:

- (1) There are simple examples showing that if either of the conditions (1) or (2) fail to hold in the definition of global hyperbolicity then maximal segments may fail to exist.
- (2) Contrary to the situation in Riemannian geometry, geodesic completeness does not guarantee the existence of maximal segments. as is well illustrated by anti-de Sitter space.

Two-dimensional anti-de Sitter space:

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

It can be shown that 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 1.9.** *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:*

- (1) If  $S$  is a Cauchy surface for  $M$  then  $\partial I^+(S) = S$ . (Also  $\partial I^-(S) = S$ .) It follows from Proposition 1.4 that a Cauchy surface  $S$  is a closed achronal  $C^0$  hypersurface in  $M$ .
- (2) If  $S$  is Cauchy then every inextendible timelike curve meets  $S$  exactly once.

**Theorem 1.12** (Geroch [15]). *If a spacetime  $M$  is globally hyperbolic then it has a Cauchy surface  $S$ .*

We make some comments about the proof.

- Introduce a measure  $\mu$  on  $M$  such that  $\mu(M) = 1$ , and consider the function  $f : M \rightarrow \mathbb{R}$  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.
- Moreover, if  $\gamma : (a, b) \rightarrow M$  is a future directed inextendible causal curve in  $M$ , one shows  $f(\gamma(t)) \rightarrow 0$  as  $t \rightarrow a^+$ , and  $f(\gamma(t)) \rightarrow \infty$  as  $t \rightarrow b^-$ .
- It follows that ‘slices’ of  $f$ , e.g.,  $S = \{p \in M : f(p) = 1\}$ , are Cauchy surfaces for  $M$ . □

Remarks:

- (1) 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. See e.g, [1, 5] for recent developments concerning the construction of *smooth time functions*, i.e., smooth functions with (past directed) timelike gradient (which hence are necessarily time functions) and smooth spacelike Cauchy surfaces.
- (2) As we shall discuss shortly, the converse of Theorem 1.12 also holds.



**Proposition 1.13.** *Let  $M$  be globally hyperbolic.*

(1) *If  $S$  is a Cauchy surface for  $M$  then  $M$  is homeomorphic to  $\mathbb{R} \times S$ .*

(2) *Any two Cauchy surfaces in  $M$  are homeomorphic.*

*Proof.* To prove (1), one introduces a future directed timelike vector field  $X$  on  $M$ .  $X$  can be scaled so that the time parameter  $t$  of each integral curve of  $X$  extends from  $-\infty$  to  $\infty$ , with  $t = 0$  at points of  $S$ . Each  $p \in M$  is on an integral curve of  $X$  that meets  $S$  in exactly one point  $q$ . This sets up a correspondence  $p \leftrightarrow (t, q)$ , which gives the desired homeomorphism. A similar technique may be used to prove (2)  $\square$

In view of Proposition 1.13, any nontrivial topology in a globally hyperbolic spacetime must reside in its Cauchy surfaces. (Note also that a Cauchy surface must be connected.)

The following fact is often useful.

**Proposition 1.14.** *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 1.10, and is also easily shown to be open.
- Let  $\gamma$  be an inextendible causal curve. Suppose  $\gamma$  meets  $J^+(S)$  at a point  $p$ . Then the portion of  $\gamma$  to the past of  $p$  must meet  $S$ , otherwise it is imprisoned in the compact set  $J^+(S) \cap J^-(p)$ .

## 1.4 Domains of Dependence

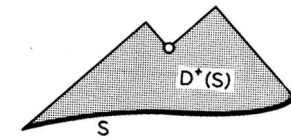
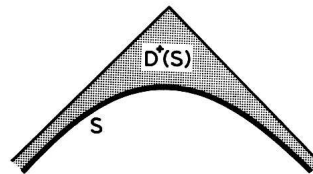
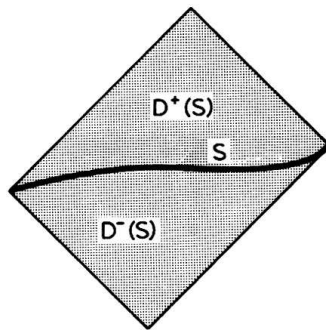
**Definition 1.10.** Let  $S$  be an achronal set in a spacetime  $M$ . The future domain of dependence of  $D^+(S)$  of  $S$  is defined as follows,

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

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:  $D^+(S) \supset S$ .

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

**Proposition 1.15.** *Let  $S$  be an achronal subset of a spacetime  $M$ . Then,  $S$  is a Cauchy surface for  $M$  if and only if  $D(S) = M$ .*

The proof follows straight-forwardly from the definitions.

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

**Proposition 1.16.** *Let  $S \subset M$  be achronal.*

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

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

*Proof.* See Gal-ESI [11] for a discussion of the proof. A few heuristic remarks: With regard to (1), suppose  $\gamma$  is a closed timelike curve through  $p \in \text{int } D(S)$ . By repeating loops we obtain an inextendible timelike which hence must meet  $S$ , infinitely often, in fact. But this would violate the achronality of  $S$ . Thus the chronology condition holds on  $\text{int } D(S)$ . More refined arguments show that strong causality holds on  $\text{int } D(S)$ .

With regard to (2), a failure of internal compactness suggests the existence of a “hole” in  $\text{int } D(S)$ . This, one expects, would lead to the existence of a past inextendible (resp. future inextendible) causal curve starting at  $p \in \text{int } D^+(S)$  (resp.  $p \in \text{int } D^-(S)$ ) that does not meet  $S$ , which would contradict the definition of the future (resp. past) domain of dependence.

We can now address the converse of Theorem 1.12.

**Corollary 1.17.** *If  $S$  is a Cauchy surface for  $M$  then  $M$  is globally hyperbolic.*

*Proof.* This follows immediately from Propositions 1.15 and 1.16:  $S$  Cauchy  $\implies D(S) = M \implies \text{int } D(S) = M \implies M$  is globally hyperbolic.  $\square$

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

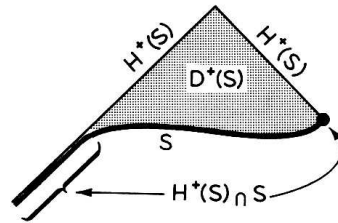
### Cauchy horizons

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 1.11.** *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 1.18.** *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)  $\partial D(S) = H(S)$ .

Proposition 1.18(c) 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. From Proposition 1.6, we obtain the following.

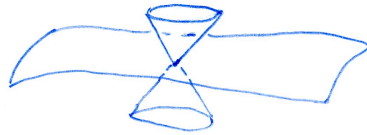
**Proposition 1.19.** *Let  $S \subset M$  be achronal. Then  $H^+(S) \setminus \text{edge } H^+(S)$ , if nonempty, is an achronal  $C^0$  hypersurface in  $M$ .*

In a similar vein to Proposition 1.7, we have the following.

**Proposition 1.20.** *Let  $S$  be an achronal subset of  $M$ . Then  $H^+(S)$  is ruled by null geodesics, i.e., every point of  $H^+(S) \setminus \text{edge } S$  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 edge  $S$ .*

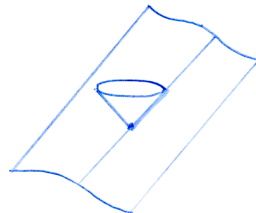
## 2 The Geometry of Null Hypersurfaces

In addition to curves, one can discuss the causality of certain higher dimensional submanifolds. For example, a *spacelike* hypersurface is a hypersurface all of whose tangent vectors are spacelike, or, equivalently, whose normal vectors are timelike:



More precisely, a hypersurface is spacelike if 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 hypersurface such that the null cone is tangent to at each of its points:



Null hypersurfaces play an important role in GR as they represent horizons of various sorts. For example the event horizons in the Schwarzschild and Kerr spacetimes are null hypersurfaces. Null hypersurfaces have an interesting geometry which we would like to discuss in this section.

## Comments on Curvature and the Einstein Equations

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

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

- *Geodesics* are curves  $t \rightarrow \sigma(t)$  of zero 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  $\text{Ric}$  and scalar curvature  $R$  are obtained by taking traces,

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

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

$$R_{ij} - \frac{1}{2} Rg_{ij} = 8\pi T_{ij},$$

where  $T_{ij}$  is the energy-momentum tensor.

- From the PDE point of view, the Einstein equations form a system of second order quasi-linear equations for the  $g_{ij}$ 's. This system may be viewed as a (highly complicated!) generalization of Poisson's equation in Newtonian gravity.
- The *vacuum* Einstein equations are obtained by setting  $T_{ij} = 0$ . It is easily seen that this equivalent to setting  $R_{ij} = 0$ .

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 condition* (DEC) is the requirement,

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



## 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  $p \in S$ ,  $g : T_p S \times T_p S \rightarrow \mathbb{R}$  is degenerate.

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

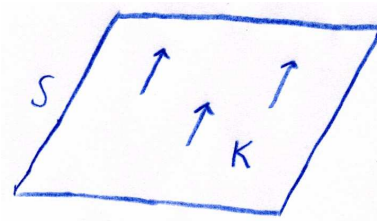
$$\langle K_p, X \rangle = 0 \quad \text{for all } X \in T_p S$$

where we have introduced the shorthand metric notation:  $\langle U, V \rangle = g(U, V)$ . In particular,

- (1)  $K_p$  is a null vector,  $\langle K_p, K_p \rangle = 0$ , which we can choose to be future pointing, and
- (2)  $[K_p]^\perp = T_p S$ .
- (3) Moreover, every (nonzero) vector  $X \in T_p S$  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.

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



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

- (1) Null hyperplanes in  $\mathbb{M}^{n+1}$ : Each nonzero null vector  $X \in T_p\mathbb{M}^{n+1}$  determines a null hyperplane  $\Pi = \{q \in \mathbb{M}^{n+1} : \langle \overline{pq}, X \rangle = 0\}$ .
- (2) 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 2.1.** *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  $S$ . Apart from parameterizations, the null generators are intrinsic to the null hypersurface.

*Proof.* Suffices to show:

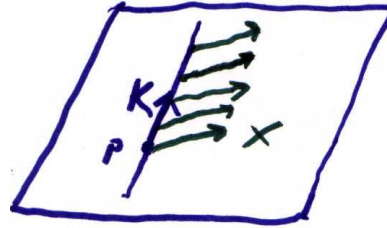
$$\nabla_K K = \lambda K$$

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

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

Extend  $X \in T_p S$  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  $p$ ,

$$\langle K, X \rangle = 0 \tag{2.1}$$

Differentiating,

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

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

□

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. We proceed to introduce null analogues of these objects. For technical reasons one works “mod  $K$ ”, as described below.

## Null Weingarten Map/Null 2nd Fundamental Form

- We introduce the following equivalence relation on tangent vectors: For  $X, Y \in T_p S$ ,

$$X = Y \text{ mod } K \iff X - Y = \lambda K \quad (2.2)$$

Let  $\bar{X}$  denote the equivalence class of  $X \in T_p S$  and let,

$$T_p S/K = \{\bar{X} : X \in T_p S\} \quad (2.3)$$

Then,

$$TS/K = \cup_{p \in S} T_p S/K \quad (2.4)$$

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  $\langle \cdot, \cdot \rangle$ : For each  $p \in S$ , define  $h : T_p S/K \times T_p S/K \rightarrow \mathbb{R}$  by

$$h(\bar{X}, \bar{Y}) = \langle X, Y \rangle.$$

Well-defined: If  $X' = X \text{ mod } K$ ,  $Y' = Y \text{ mod } K$  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  $b : T_p S / K \rightarrow T_p S / K$  defined by

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

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

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

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

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

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

- 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  $p \in S$ ,  $B : T_p S/K \times T_p S/K \rightarrow \mathbb{R}$  is defined by,

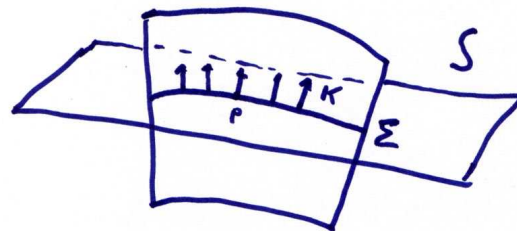
$$B(\bar{X}, \bar{Y}) := h(b(\bar{X}), \bar{Y}) = h(\overline{\nabla_X K}, \bar{Y}) = \langle \nabla_X K, Y \rangle.$$

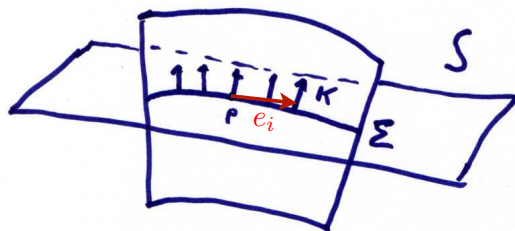
Since  $b$  is self-adjoint,  $B$  is symmetric.

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

$$\theta = \text{tr } b$$

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



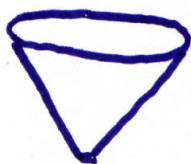


- Let  $\{e_1, e_2, \dots, e_{n-1}\}$  be an orthonormal basis for  $T_p\Sigma$  in the induced metric. Then  $\{\bar{e}_1, \bar{e}_2, \dots, \bar{e}_{n-1}\}$  is an orthonormal basis for  $T_pS/K$ . Hence at  $p$ ,

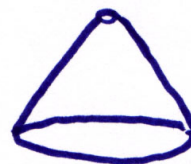
$$\begin{aligned} \theta &= \text{tr } b = \sum_{i=1}^{n-1} h(b(\bar{e}_i), \bar{e}_i) = \sum_{i=1}^{n-1} \langle \nabla_{e_i} K, e_i \rangle. \\ &= \text{div}_\Sigma K. \end{aligned} \tag{2.5}$$

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

Thus,  $\theta$  measures the overall expansion of the null generators of  $S$  towards the future.



$$\theta > 0$$

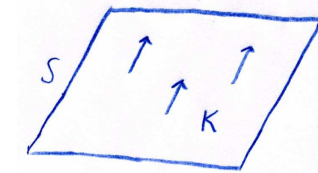


$$\theta < 0$$

*Effect of scaling:* If  $\tilde{K} = fK$ ,  $f \in C^\infty(S)$ , is any other future directed null vector field on  $S$ , then  $b_{\tilde{K}} = fb_K$ , and hence,  $\tilde{\theta} = f\theta$ . It follows that the Weingarten map  $b = b_K$  at a point  $p$  is uniquely determined by the value of  $K$  at  $p$ .

## Re: The Geometry of Null Hypersurfaces

- $S$  = null hypersurface in spacetime  $M$   
 $K$  = associated future directed null vector field



- Equivalence relation:

$$X = Y \text{ mod } K \iff X - Y = \lambda K$$

- mod  $K$  tangent space:  $T_p S/K = \{\bar{X} : X \in T_p S\}$   
 mod  $K$  tangent bundle:  $TS/K = \cup_{p \in S} T_p S/K$
- Metric  $h$  on  $TS/K$ :  $h(\bar{X}, \bar{Y}) = \langle X, Y \rangle$  (positive definite).
- Null Weingarten map  $b : TS/K \rightarrow TS/K$ , defined by

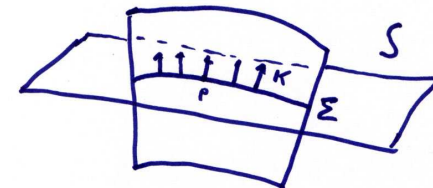
$$b(\bar{X}) = \overline{\nabla_X K}$$

- Associated null second fundamental form  $B : TS/K \times TS/K \rightarrow \mathbb{R}$ ,

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

- Null expansion scalar  $\theta$ ,

$$\theta := \text{tr } b = \text{tr } B = \text{div}_\Sigma K$$





## Comparison Theory

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

Let  $\eta : I \rightarrow M$ ,  $s \rightarrow \eta(s)$ , 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 the null vector field  $K$  which equals  $\eta'(s)$  at  $\eta(s)$ ,

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



**Proposition 2.2.** *The one parameter family of Weingarten maps  $s \rightarrow b(s)$ , obeys the following Riccati equation,*

$$b' + b^2 + R = 0, \quad ' = \nabla_{\eta'} \quad (2.6)$$

where  $R : T_{\eta(s)}S/\eta'(s) \rightarrow T_{\eta(s)}S/\eta'(s)$  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  $\eta$  tangent to  $S$ , we define,  $(\overline{Y})' = \overline{Y'}$ . Then, if  $X = X(s)$  is a vector field along  $\eta$  tangent to  $S$ ,  $b'$  is defined by,

$$b'(\overline{X}) := b(\overline{X})' - b(\overline{X'}). \quad (2.7)$$

*Proof.* Fix a point  $p = \eta(s_0)$ ,  $s_0 \in (a, b)$ , on  $\eta$ . 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  $K$ , restricted to  $\eta$ , is the velocity vector field to  $\eta$ , i.e., for each  $s$  near  $s_0$ ,  $K_{\eta(s)} = \eta'(s)$ . Let  $X \in T_p M$ . 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  $\eta$  we have,

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

(which implies that  $X$ , restricted to  $\eta$ , is a Jacobi field along  $\eta$ ).

Thus, from Equation (2.7), at the point  $p$  we have,

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

which establishes Equation 2.6.

By taking the trace of (2.6) we obtain the following formula for the derivative of the null mean curvature  $\theta = \theta(s)$  along  $\eta$ ,

$$\theta' = -\text{Ric}(\eta', \eta') - \sigma^2 - \frac{1}{n-1}\theta^2, \quad (2.8)$$

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

Equation 2.8 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 2.3.** *Let  $M$  be a spacetime which obeys the null energy condition (NEC),  $\text{Ric}(X, X) \geq 0$  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  $s \rightarrow \eta(s)$  be the null generator of  $S$  passing through  $p = \eta(0)$ , affinely parametrized. Let  $b(s) = b_{\eta'(s)}$ , and take  $\theta = \text{tr } b$ . 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,

$$\frac{d\theta}{ds} \leq -\frac{1}{n-1}\theta^2, \quad (2.9)$$

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

$$\frac{d}{ds} \left( \frac{1}{\theta} \right) \geq \frac{1}{n-1}, \quad (2.10)$$

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

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

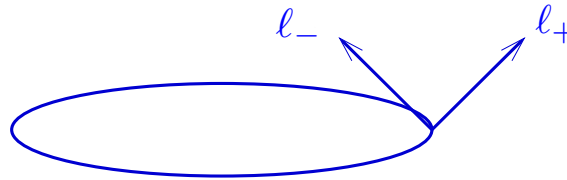
Thus, Proposition 2.3 implies, under the given assumptions, that cross sections of  $S$  are nondecreasing in area as one moves towards the future. Proposition 2.3 is the simplest form of Hawking's black hole area theorem [16]. For a study of the area theorem, with a focus on issues of regularity, see [3].

### 3 The Penrose Singularity Theorem and Related Results

In this section we introduce the important 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  $\Sigma^{n-1}$  be a closed (i.e., compact without boundary) co-dimension two spacelike submanifold of  $M$ .
- Each normal space of  $\Sigma$ ,  $[T_p\Sigma]^\perp$ ,  $p \in \Sigma$ , is timelike and 2-dimensional, and hence admits two future directed null directions orthogonal to  $\Sigma$ .

Thus, under suitable orientation assumptions,  $\Sigma$  admits two smooth nonvanishing future directed null normal vector fields  $\ell_+$  and  $\ell_-$  (unique up to positive rescaling).



By convention, we refer to  $\ell_+$  as outward pointing and  $\ell_-$  as inward pointing.

- Associated to  $\ell_+$  and  $\ell_-$ , are the two *null second fundamental forms*,  $\chi_+$  and  $\chi_-$ , respectively, defined as

$$\chi_{\pm} : T_p\Sigma \times T_p\Sigma \rightarrow \mathbb{R}, \quad \chi_{\pm}(X, Y) = g(\nabla_X \ell_{\pm}, Y). \quad (3.1)$$

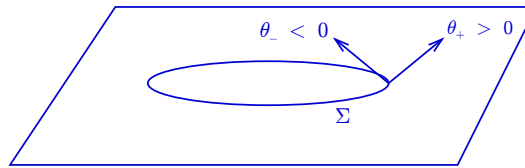
- The *null expansion scalars (or null mean curvatures)*  $\theta_{\pm}$  of  $\Sigma$  are obtained by tracing  $\chi_{\pm}$  with respect to the induced metric  $\gamma$  on  $\Sigma$ ,

$$\theta_{\pm} = \text{tr}_{\gamma} \chi_{\pm} = \gamma^{AB} \chi_{\pm AB} = \text{div}_{\Sigma} \ell_{\pm}. \quad (3.2)$$

The sign of  $\theta_{\pm}$  does not depend on the scaling of  $\ell_{\pm}$ . Physically,  $\theta_+$  (resp.,  $\theta_-$ ) measures the divergence of the outgoing (resp., ingoing) light rays emanating orthogonally from  $\Sigma$ .

*Remark:* There is a natural connection between these null expansion scalars  $\theta_{\pm}$  and the null expansion of null hypersurfaces:  $\ell_+$  locally generates a smooth null hypersurface  $S_+$ . Then  $\theta_+$  is the null expansion of  $S_+$  restricted to  $\Sigma$ ;  $\theta_-$  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  $\Sigma$  is called a **trapped surface**. (See e.g. [2, 19] for results concerning the dynamical formation of trapped surfaces.)

As we now discuss, under appropriate energy and causality conditions, the occurrence of a trapped surface signals the onset of gravitational collapse.

### *The Penrose Singularity Theorem.*

- The Penrose singularity theorem [23] is the first of the famous singularity theorems of general relativity. The singularity theorems establish, under generic circumstances, the existence in spacetime of incomplete timelike or null geodesics.
- Such incompleteness indicates that spacetime comes to an end either in the past or future. Future incompleteness is often associated with gravitational collapse and the formation of a black hole (such as in Schwarzschild and Kerr).
- All the classical singularity theorems require energy conditions. The Penrose singularity theorem assumes the NEC,  $\text{Ric}(X, X) \geq 0$  for all null vectors  $X$ , holds.
- It also assumes the existence of a *noncompact* Cauchy surface. This is consistent with modeling an isolated gravitating system, where one takes spacetime (or space) to be asymptotically flat.
- What Penrose proves is that if the gravitational field becomes sufficiently strong that trapped surfaces appear (as they do in the Schwarzschild solution) then the development of singularities is inevitable.



**Theorem 3.1** (Penrose singularity theorem). *Let  $M$  be a globally hyperbolic spacetime which satisfies the null energy condition and which has a noncompact Cauchy surface  $S$ . If  $M$  contains a trapped surface  $\Sigma$  then  $M$  is future null geodesically incomplete.*

*Proof:* We first observe the following.

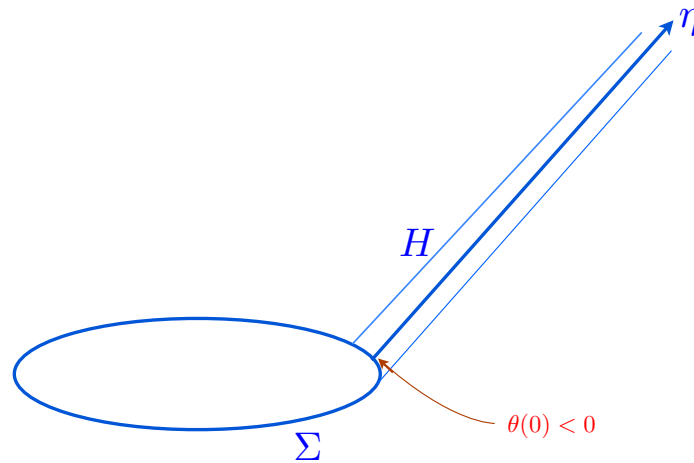
**Claim:**  $\partial I^+(\Sigma)$  is noncompact.

*Proof of Claim:*  $\partial I^+(\Sigma)$  is an achronal boundary, and hence, by Proposition 1.4, is an achronal  $C^0$  hypersurface. If  $\partial I^+(\Sigma)$  were compact then, by Proposition 1.14,  $\partial I^+(\Sigma)$  would be a compact Cauchy surface. But this would contradict the assumption that  $S$  is noncompact (all Cauchy surfaces are homeomorphic).

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

- Each  $q \in \partial I^+(\Sigma)$  lies on a null geodesic in  $\partial I^+(\Sigma)$  with past end point on  $\Sigma$ ; cf. Propositions 1.7 and 1.10. Moreover this null geodesic meets  $\Sigma$  orthogonally (due to achronality, cf. O’Neill [22, Lemma 50, p. 298]).
- Since  $\partial I^+(\Sigma)$  is closed and noncompact, there exists a sequence of points  $\{q_n\} \subset \partial I^+(\Sigma)$  that diverges to infinity. For each  $n$ , there is a null geodesic  $\eta_n$  from  $\Sigma$  to  $q_n$ , which is contained in  $\partial I^+(\Sigma)$  and meets  $\Sigma$  orthogonally.
- By compactness of  $\Sigma$ , some subsequence converges to a future inextendible null geodesic  $\eta$  contained in  $\partial I^+(\Sigma)$ , and meeting  $\Sigma$  orthogonally (at  $p$ , say). Suppose  $\eta$  is future complete.

- By achronality of  $\partial I^+(\Sigma)$ ,
  - No other null geodesic in  $\partial I^+(\Sigma)$  starting on  $\Sigma$  can meet  $\eta$ .
  - There can be no *null focal point* to  $\Sigma$  along  $\eta$  (cf. [22, Prop. 48, p. 296]).
- It follows that  $\eta$  is contained in a smooth (perhaps very thin) null hypersurface  $H \subset \partial I^+(\Sigma)$ .

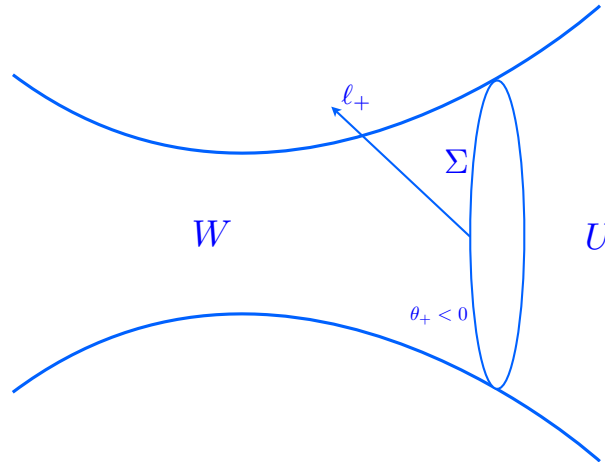


- Let  $\theta$  be the null expansion of  $H$  along  $\eta$ . Since  $\Sigma$  is a trapped surface  $\theta(p) < 0$ . Arguing just as in the “area theorem” (Proposition 2.3), using Raychaudhuri + NEC,  $\theta$  must go to  $-\infty$  in finite affine parameter time  $\rightarrow\leftarrow$ . Hence  $\eta$  must be future incomplete.  $\square$

### A variant of the Penrose singularity theorem

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

**Theorem 3.2.** *Let  $M$  be a globally hyperbolic spacetime satisfying the null energy condition, with smooth spacelike Cauchy surface  $V$ . Let  $\Sigma$  be a smooth closed (compact without boundary) hypersurface in  $V$  which separates  $V$  into an “inside”  $U$  and an “outside”  $W$ , 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  $\Sigma$  is **outer-trapped** ( $\theta_+ < 0$ ) then  $M$  is future null geodesically incomplete.*



*Proof.* 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. But, by considering the integral curves of a timelike vector field  $X$  on  $M$ , one can establish a homeomorphism between  $\partial I^+(\overline{U})$  and  $\overline{W} \rightarrow \leftarrow$ .

This version of the Penrose singularity theorem may be used to prove the following beautiful result of Gannon [14] and Lee [20].

**Theorem 3.3.** *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.*

*Comment on the proof.* Pass to the universal covering spacetime  $\overline{M}$ . It will contain a Cauchy surface  $\overline{V}$ , which is the universal cover of  $V$ . If  $\pi_1(V) \neq 0$  then  $\overline{V}$  will have more than one AF end. Now apply Theorem 3.2 to  $\overline{M}$  with Cauchy surface  $\overline{V}$ .

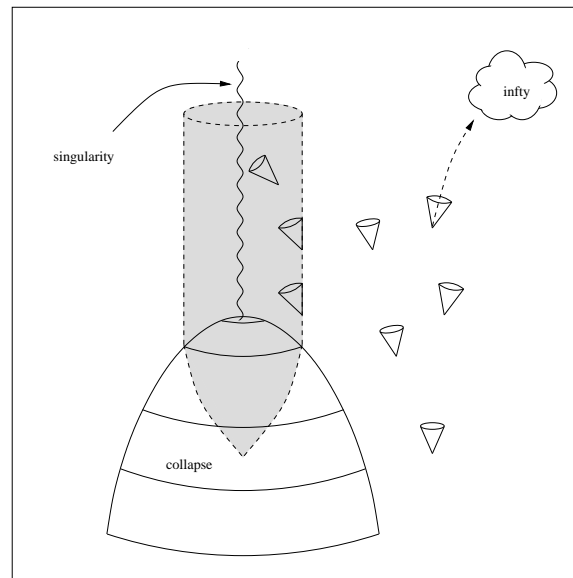
As suggested by Theorem 3.3, nontrivial topology tends to induce gravitational collapse. This in turn leads to the notion of *topological censorship*. We return to these issues in Section 5.

## 4 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.



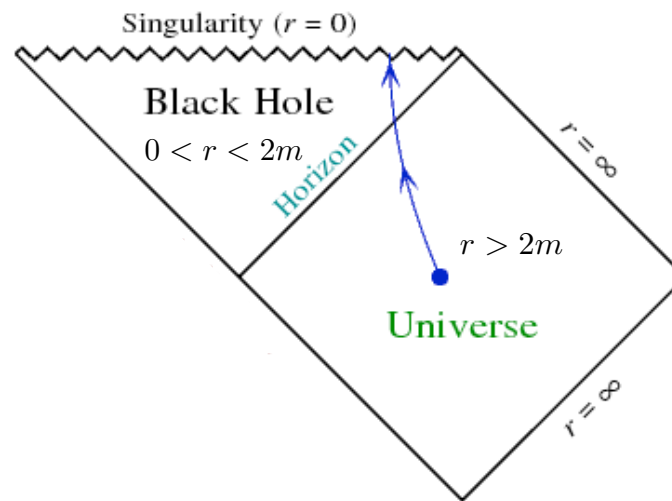
A stellar object, after its fuel is spent, begins to collapse under its own weight. As the gravitational field intensifies the light cones bend “inward” (so to speak).

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.

Ex. *The Schwarzschild solution (1916)*. Static (time-independent, nonrotating) spherically symmetric, vacuum solution to the Einstein equations.

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

The region  $0 < r < 2m$  is the black hole region;  $r = 2m$  corresponds to the event horizon. This metric represents the region outside a (collapsing) spherically symmetric star.



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

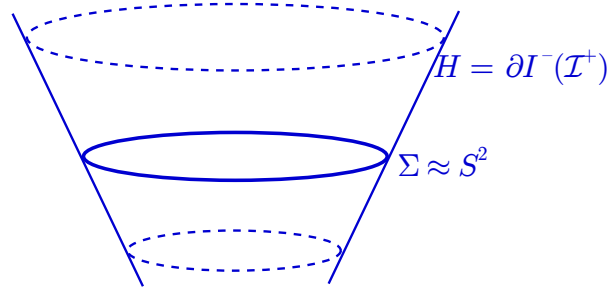
$$\begin{aligned}
 ds^2 = & - \left[ 1 - \frac{2mr}{r^2 + a^2 \cos^2 \theta} \right] dt^2 - \frac{4mra \sin^2 \theta}{r^2 + a^2 \cos^2 \theta} dt d\phi \\
 & + \left[ \frac{r^2 + a^2 \cos^2 \theta}{r^2 - 2mr + a^2} \right] dr^2 + (r^2 + a^2 \cos^2 \theta) d\theta^2 \\
 & + \left[ r^2 + a^2 + \frac{2mra^2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

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. (The proof assumes analyticity, but there has been recent progress in removing this assumption; see the excellent recent review article of Ionescu and Klainerman [18].)

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  $\Sigma$  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  $\Sigma$  has genus  $\geq 1$  then  $\Sigma$  can be deformed outward to an *outer trapped surface* (to be defined shortly). 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 [24, 4].



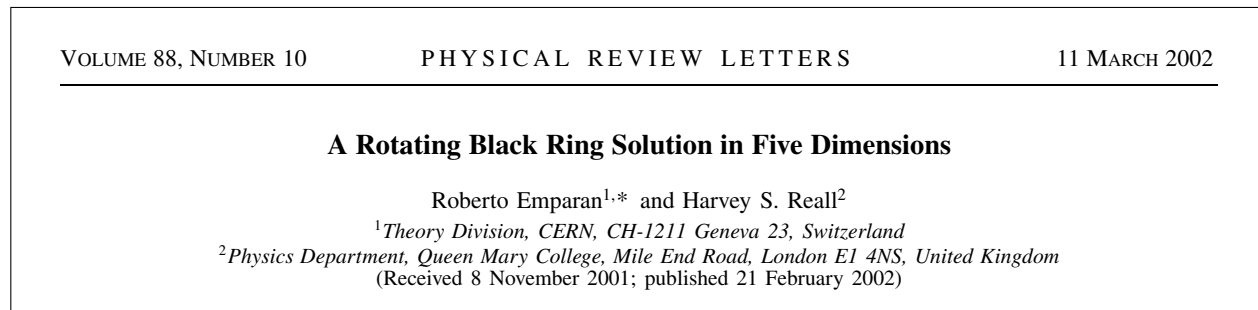
## Higher Dimensional Black Holes

- 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 [21] 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 [7] 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*).



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?

## 4.1 Marginally Outer Trapped Surfaces

Want to describe a generalization of Hawking's theorem to higher dimensions. This will be based on properties of *marginally outer trapped surfaces*.

By an *initial data set* in a spacetime  $M^{n+1}$  we mean a triple  $(V^n, h, K)$ , where  $V^n$  is a spacelike hypersurface,  $h$  is the induced metric and  $K$  is the second fundamental form: For vectors  $X, Y \in T_p V$ ,

$$K(X, Y) = h(\nabla_X u, Y)$$


Given an initial data set  $(V, h, K)$ , the spacetime dominant energy condition implies,

$$\mu \geq |J|$$

along  $V$ , where  $\mu = \text{local energy density} = G(u, u)$ , and  $J = \text{local momentum density} = 1\text{-form } G(u, \cdot)$  on  $V$ , where  $G = Ric - \frac{1}{2}Rg$ .

Using the Gauss-Codazzi equations,  $\mu$  and  $J$  can be expressed solely in terms of the initial data

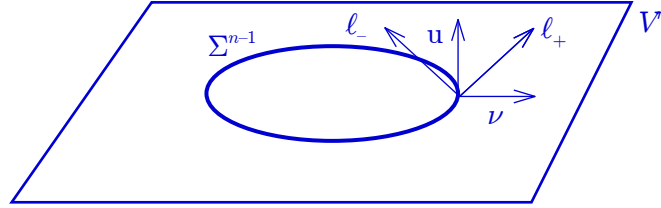
$$\mu = \frac{1}{2} (S + (\text{tr } K)^2 - |K|^2) ,$$

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

These are the *Einstein constraint* equations. In the *time-symmetric* case ( $K = 0$ ), the DEC reduces to  $S \geq 0$ .

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

Let  $\Sigma^{n-1}$  be a closed 2-sided hypersurface in  $V^n$ .  $\Sigma$  admits a smooth unit normal field  $\nu$  in  $V$ .



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

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

Null second fundamental forms:  $\chi_+, \chi_-$

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

Null expansion scalars:  $\theta_+, \theta_-$

$$\theta_{\pm} = \text{tr}_{\gamma} \chi_{\pm} = \gamma^{AB} (\chi_{\pm})_{AB} = \text{div}_{\Sigma} \ell_{\pm}$$

In terms of initial data  $(V^n, h, K)$ ,

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

where  $H$  = mean curvature of  $\Sigma$  within  $V$ . (*Note:* In the ‘time-symmetric’ case ( $K = 0$ ),  $\theta_+ = H$ .)

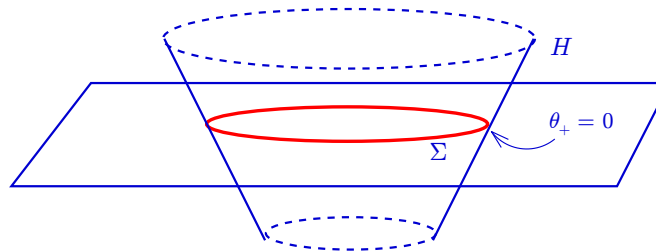
Recall,  $\Sigma$  is a trapped surface if both  $\theta_+$  and  $\theta_-$  are negative. Focusing attention on the outward null normal only:

- (1) If  $\theta_+ < 0$  - we say  $\Sigma$  is outer trapped.
- (2) If  $\theta_+ = 0$  - we say  $\Sigma$  is a marginally outer trapped surface (MOTS),

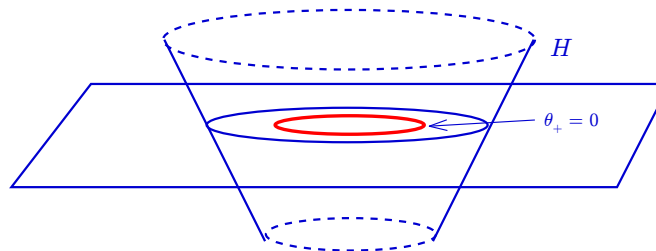
(*Note:* In the time symmetric case a MOTS is simply a minimal surface.)

MOTSs arise naturally in several situations.

- In stationary black hole spacetimes - cross sections of the event horizon are MOTS.



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



## Stability of MOTS

MOTSs admit a notion of stability based on variations of the null expansion (Andersson, Mars and Simon (2005, 2008)).

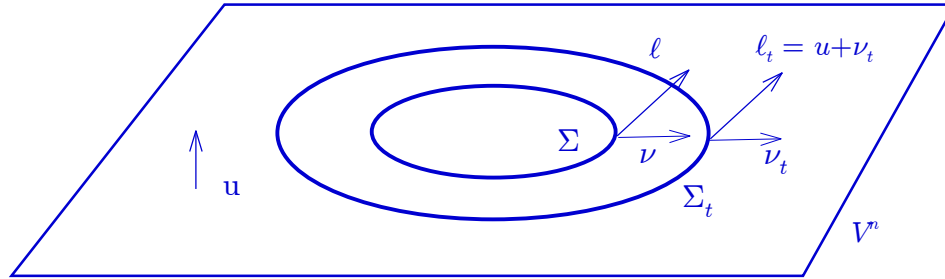
Let  $\Sigma$  be a MOTS in an initial data set  $(V, h, K)$  with outward normal  $\nu$ . Consider normal variations of  $\Sigma$  in  $V$ , i.e., variations  $t \rightarrow \Sigma_t$  of  $\Sigma = \Sigma_0$  with variation vector field

$$\mathcal{V} = \left. \frac{\partial}{\partial t} \right|_{t=0} = \phi \nu, \quad \phi \in C^\infty(\Sigma). \quad (4.1)$$

Let

$$\theta(t) = \text{the null expansion of } \Sigma_t, \quad (4.2)$$

with respect to  $\ell_t = u + \nu_t$ , where  $\nu_t$  is the unit normal field to  $\Sigma_t$  in  $V$ .



A computation shows,

$$\left. \frac{\partial \theta}{\partial t} \right|_{t=0} = L(\phi), \quad (4.3)$$

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

$$L(\phi) = -\Delta \phi + 2\langle X, \nabla \phi \rangle + (Q + \operatorname{div} X - |X|^2) \phi, \quad (4.4)$$

$$Q = \frac{1}{2} S_\Sigma - (\mu + J(\nu)) - \frac{1}{2} |\chi|^2, \quad (4.5)$$

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 observed by Andersson, Mars and Simon.

**Lemma 4.2.** *Among eigenvalues with smallest real part, there is a real eigenvalue  $\lambda_1(L)$ , called the *principal eigenvalue*. The associated eigenfunction  $\phi$ ,  $L(\phi) = \lambda_1 \phi$ , is unique up to a multiplicative constant, and can be chosen to be strictly positive.*

In analogy with the minimal surface case, *we say that a MOTS  $\Sigma$  is **stable** provided  $\lambda_1(L) \geq 0$ .*

Heuristically, a stable MOTS is *infinitesimally* outermost.

There is a basic criterion for a MOTS to be stable. We say a MOTS  $\Sigma$  is *weakly outermost* provided there are no outer trapped ( $\theta < 0$ ) surfaces outside of, and homologous, to  $\Sigma$ .

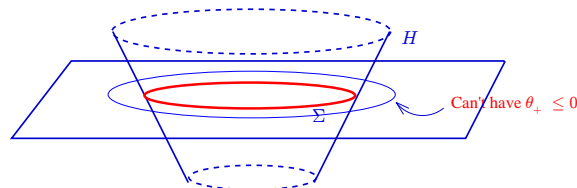
- **Fact.** A weakly outermost MOTS is stable.

*Proof:* Suppose to the contrary,  $\lambda_1 < 0$ . Consider the variation  $t \rightarrow \Sigma_t$  of  $\Sigma$  with variation vector field  $\mathcal{V} = \phi\nu$ , where  $\phi$  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  $t > 0$ , and hence there are outer trapped surfaces outside of, and homologous to,  $\Sigma$ .

- **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, Eichmair).



## 4.2 A Generalization of Hawking's Black Hole Topology Theorem

**Theorem 4.3** (G. and Schoen [13]). *Let  $(V, h, K)$  be an  $n$ -dimensional initial data set,  $n \geq 3$ , satisfying the dominant energy condition (DEC),  $\mu \geq |J|$ . If  $\Sigma$  is a stable MOTS in  $V$  then (apart from certain exceptional circumstances)  $\Sigma$  must be of *positive Yamabe type*, i.e. must admit a metric of positive scalar curvature.*

Remarks.

- The theorem may be viewed as a spacetime analogue of a fundamental result of Schoen and Yau concerning stable minimal hypersurfaces in manifolds of positive scalar curvature.
- The 'exceptional circumstances' are ruled out if, for example, the DEC holds strictly at some point of  $\Sigma$  or  $\Sigma$  is not Ricci flat.
- $\Sigma$  being of positive Yamabe type implies many well-known restrictions on the topology (see e.g.; see e.g. [17, Chapter 7]).

We consider here two basic examples, and for simplicity we assume  $\Sigma$  is orientable.

Case 1  $\dim \Sigma = 2$  ( $\dim M = 3 + 1$ ): In this case,  $\Sigma$  being of positive Yamabe type means that  $\Sigma$  admits a metric of positive Gaussian curvature. Hence, by the Gauss-Bonnet theorem,  $\Sigma$  is topologically a 2-sphere, and we recover Hawking's theorem.

Case 2.  $\dim \Sigma = 3$  ( $\dim M = 4 + 1$ ).

**Theorem** (Gromov-Lawson, Schoen-Yau) If  $\Sigma$  is a closed orientable 3-manifold of positive Yamabe type then  $\Sigma$  must be diffeomorphic to:

- a spherical space, or
- $S^2 \times S^1$ , or
- a connected sum of the above two types.

Remark: By the prime decomposition theorem,  $\Sigma$  must be a connected sum of (1) spherical spaces, (2)  $S^2 \times S^1$ 's, and (3)  $K(\pi, 1)$  manifolds. But since  $\Sigma$  is positive Yamabe, it cannot contain any  $K(\pi, 1)$ 's in its prime decomposition.

Thus, the basic horizon topologies in  $\dim \Sigma = 3$  case are  $S^3$  and  $S^2 \times S^1$ .

(Kundari and Lucietti have recently constructed an AF black hole spacetime satisfying the DEC with  $\mathbb{RP}^3$  horizon topology, arXiv 2014.)

*Proof of Theorem 4.3:* Need to produce a metric of positive scalar curvature on  $S_\Sigma$ .

- Formally setting  $X = 0$  in the MOTS stability operator  $L$ , we obtain the "symmetrized" operator

$$L_0(\phi) = -\Delta\phi + Q\phi. \quad (4.6)$$

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

- **Key Fact:**  $\lambda_1(L_0) \geq \lambda_1(L)$ .
- Since, by assumption,  $\lambda_1(L) \geq 0$ , we have  $\lambda_1(L_0) \geq 0$ .

This, in particular, implies the *MOTS stability inequality*,

$$\int_\Sigma |\nabla\psi|^2 + Q\psi^2 dA \geq 0, \quad \forall \psi \in C^\infty(\Sigma) \quad (4.7)$$

(In essence, the situation has been reduced to the Riemannian case.)

- Now make the conformal change:  $\tilde{\gamma} = f^{\frac{2}{n-2}}\gamma$ , where  $f$  is a positive eigenfunction corresponding to  $\lambda_1(L_0)$ :  $-\Delta f + Qf = \lambda_1(L_0)f$ . A standard computation shows,

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

*Proof of the Key Fact.* Let  $\phi$  be a positive eigenfunction associated to  $\lambda_1(L)$ ,  $L(\phi) = \lambda_1(L)\phi$ . Completing the square on the right hand side of (4.4)) gives,

$$-\Delta\phi + (Q + \operatorname{div} X)\phi + \phi|\nabla \ln \phi|^2 - \phi|X - \nabla \ln \phi|^2 = \lambda_1(L)\phi \quad (4.8)$$

where  $Q := \frac{1}{2}S - (\mu + J(\nu)) - \frac{1}{2}|\chi|^2$ . Letting  $u = \ln \phi$ , we obtain,

$$-\Delta u + Q + \operatorname{div} X - |X - \nabla u|^2 = \lambda_1(L). \quad (4.9)$$

Absorbing the Laplacian term  $\Delta u = \operatorname{div}(\nabla u)$  into the divergence term gives,

$$Q + \operatorname{div}(X - \nabla u) - |X - \nabla u|^2 = \lambda_1(L). \quad (4.10)$$

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

$$-Q + |Y|^2 + \lambda_1(L) = \operatorname{div} Y. \quad (4.11)$$

Given any  $\psi \in C^\infty(\Sigma)$ , we multiply through by  $\psi^2$  and derive,

$$\begin{aligned} -\psi^2 Q + \psi^2 |Y|^2 + \psi^2 \lambda_1(L) &= \psi^2 \operatorname{div} Y \\ &= \operatorname{div}(\psi^2 Y) - 2\psi \langle \nabla \psi, Y \rangle \\ &\leq \operatorname{div}(\psi^2 Y) + 2|\psi| |\nabla \psi| |Y| \\ &\leq \operatorname{div}(\psi^2 Y) + |\nabla \psi|^2 + \psi^2 |Y|^2. \end{aligned}$$

Integrating the above inequality yields,

$$\lambda_1(L) \leq \frac{\int_\Sigma |\nabla \psi|^2 + Q\psi^2}{\int_\Sigma \psi^2} \quad \text{for all } \psi \in C^\infty(\Sigma), \psi \not\equiv 0. \quad (4.12)$$

$$\implies \lambda_1(L) \leq \lambda_1(L_0).$$

One drawback of Theorem 4.3 is that it allows certain possibilities that one would like to rule out. E.g., the theorem does not rule out the possibility of a vacuum black hole spacetime with toroidal horizon topology. In a subsequent paper such possibilities were ruled out.

**Theorem 4.4** (G. [10]). *Let  $(V, h, K)$  be an  $n$ -dimensional,  $n \geq 3$ , initial data set in a spacetime obeying the DEC. Suppose  $\Sigma$  is a **weakly outermost MOTS** in  $V$ . If  $\Sigma$  is not of positive Yamabe type then an outer neighborhood  $U \approx [0, \epsilon) \times \Sigma$  of  $\Sigma$  is foliated by MOTS.*

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

This rigidity result belongs to a *family* of related rigidity results; see e.g. [12] and references therein.

## 5 The Geometry and Topology of the Exterior Region

In principle, the topology of space could be quite complicated. In fact GR does not put any restrictions on the topology of space.

**Theorem 5.1** (Isenberg-Mazzeo-Pollack (2003)). *Let  $N$  be any compact  $n$ -dimensional manifold, and  $p \in N$  arbitrary. Then  $N \setminus \{p\}$  admits an AF initial data set satisfying the vacuum constraint equations.*

However, according to the principle of *topological censorship*, the topology of the region outside of all black holes (and white holes) should be simple. The rationale is roughly as follows:

- The Theorem of Gannon [14] and Lee [20] discussed in Section 3 suggests that non-trivial topology induces gravitational collapse.
- *Weak cosmic censorship conjecture*: In the standard collapse scenario, the process of gravitational collapse leads to the formation of an event horizon which shields the singularities from view.
- *Topological censorship*: The nontrivial topology that induced collapse should end up behind hidden the event horizon, and the region outside the black hole should have simple topology.

This circle of ideas was formalized by the Topological Censorship Theorem of Friedman, Schleich and Witt [8], which says roughly that observers who remain outside the event horizon are unable to probe nontrivial topology.

In [9] the following strengthened version was obtained:

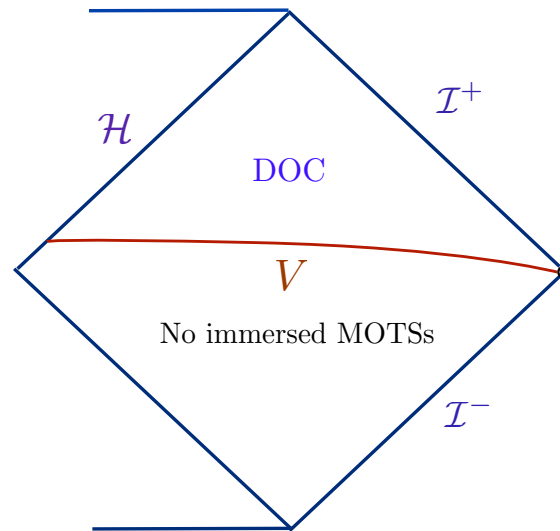
**Theorem 5.2.** *Let  $M$  be an asymptotically flat spacetime, and suppose that the domain of outer communications (the region outside of all black holes and white holes) is globally hyperbolic and satisfies the NEC. Then the DOC is simply connected.*

Remark: Here, by an asymptotically flat spacetime we mean one that has an asymptotic structure similar to that of Minkowski space in the sense that it admits a regular past and future null infinity  $\mathcal{I}^-$  and  $\mathcal{I}^+$ , respectively; cf. Wald [24]. (The DOC is then the region  $I^-(\mathcal{I}_+) \cap I^+(\mathcal{I}_+)$ ).

These and other results supporting the notion of topological censorship are *spacetime* results; they involve assumptions that are essentially global in time. It is a natural but difficult question to determine whether a given initial data set will give rise to a spacetime that satisfies these conditions.

The aim of recent work with Michael Eichmair and Dan Pollack was to obtain a result supportive of the principle of topological censorship *at the pure initial data level*, thereby circumventing these difficult issues of global evolution.

Consider the following schematic Penrose-type diagram of a black hole spacetime.



- Should think of the initial data manifold  $V$  as representing an asymptotically flat spacelike slice in the domain of outer communications (DOC) whose boundary  $\partial V$  corresponds to a cross section of the event horizon.
- At the initial data level, we represent this cross section by a MOTS.
- We assume there are no *immersed* MOTSs in the ‘exterior region’  $V \setminus \partial V$ .



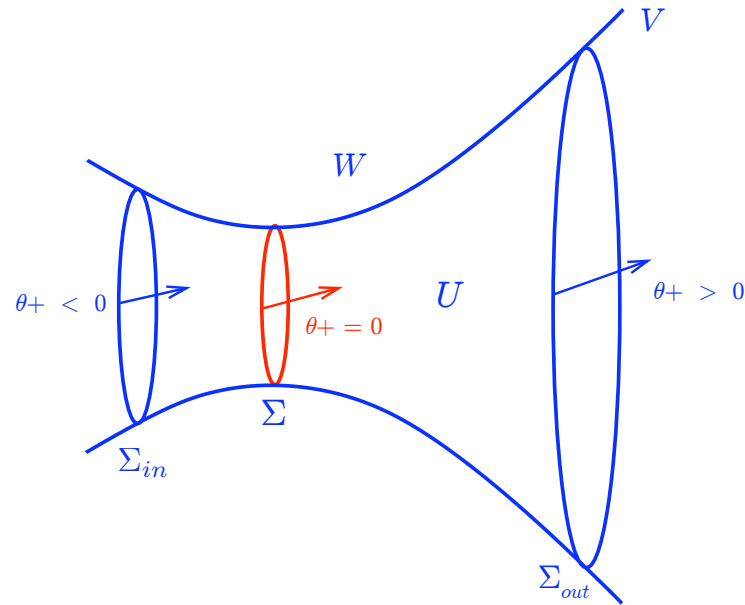
**Theorem 5.3** (Eichmair, G., Pollack [6]). *Let  $(V, h, K)$  be a 3-dimensional asymptotically flat (AF) initial data set such that  $V$  is a manifold-with-boundary, whose boundary  $\partial V$  is a compact MOTS. If there are no (immersed) MOTS in  $V \setminus \partial V$ , then  $V$  is diffeomorphic to  $\mathbb{R}^3$  minus a finite number of open balls.*

AF:  $V$  outside a compact set is diffeomorphic to  $\mathbb{R}^3 \setminus B_1(\mathbf{0})$ , such that  $h \rightarrow \delta$  and  $K \rightarrow 0$  at suitable rates as  $r \rightarrow \infty$ .

### Remarks

- The proof makes use of powerful existence results for MOTSs (Schoen, Andersson and Metzger, Eichmair).
- The proof also makes use of an important consequence of *geometrization*, namely that the fundamental group of every compact 3-manifold is *residually finite*.
- Dominant energy condition not required!
- MOTSs detect nontrivial topology (reminiscent of how minimal surfaces have been used in Riemannian geometry to detect nontrivial topology, cf., Meeks-Simon-Yau).
- In the ‘no horizon’ case ( $\partial V = \emptyset$ ) the conclusion is that  $V$  is diffeomorphic to  $\mathbb{R}^3$ .

**Theorem 5.4** (Schoen, Andersson and Metzger, Eichmair). *Let  $W^n$  be a connected compact manifold-with-boundary in an initial data set  $(V^n, h, K)$ ,  $3 \leq n \leq 7$ . Suppose,  $\partial W = \Sigma_{in} \cup \Sigma_{out}$ , such that  $\Sigma_{in}$  is outer trapped ( $\theta_+ < 0$ ) and  $\Sigma_{out}$  is outer untrapped ( $\theta_+ > 0$ ). Then there exists a smooth compact MOTS in  $W$  that separates  $\Sigma_{in}$  from  $\Sigma_{out}$ .*



Remarks on the proof of existence:

- Given the initial data set  $(V^n, h, K)$ , Jang's equation is the equation,

$$\hat{h}^{ij} \left( \frac{D_i D_j u}{\sqrt{1 + |Df|^2}} - K_{ij} \right) = 0 \quad (5.1)$$

where  $u : V \rightarrow \mathbb{R}$ ,  $D$  is the Levi-Civita connection of  $h$ , and  $\hat{h}^{ij} = h^{ij} - \frac{u^i u^j}{1 + |Du|^2}$ .

- In their proof of the positive mass theorem in the general (non-time symmetric case), Schoen and Yau studied in detail the existence and regularity properties of Jang's equation, which they interpreted as a prescribed mean curvature equation of sorts,

$$H_\Sigma + \text{tr}_\Sigma \bar{K} = 0, \quad (5.2)$$

where  $H_\Sigma =$  the mean curvature of  $\Sigma = \text{graph } u$  in  $V \times \mathbb{R}$ .

- They showed that the only obstruction to global existence is the presence of MOTSs in the initial data.
- As Schoen described in his 2005 Miami Waves lecture, one can turn this 'drawback' into a 'feature': One can establish existence by inducing blow-up of Jang's equation.
- Complete proofs were given by Andersson and Metzger (dim = 3), Eichmair (dim  $\geq 3$ ); cf. the excellent survey article by Andersson, Eichmair, Metzger (2011).

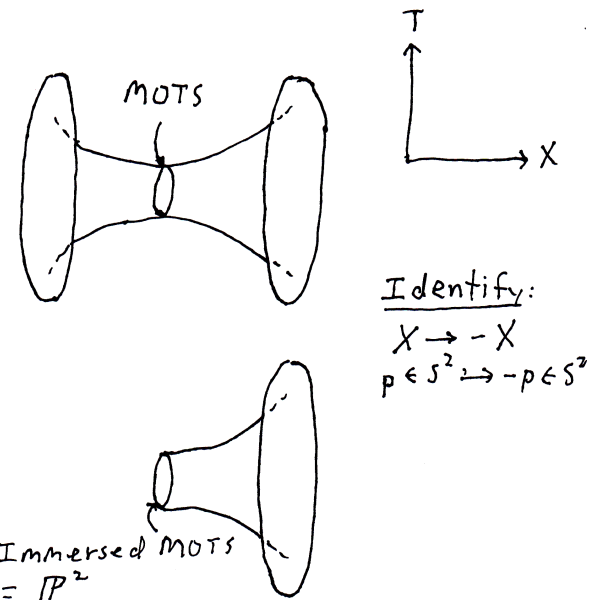
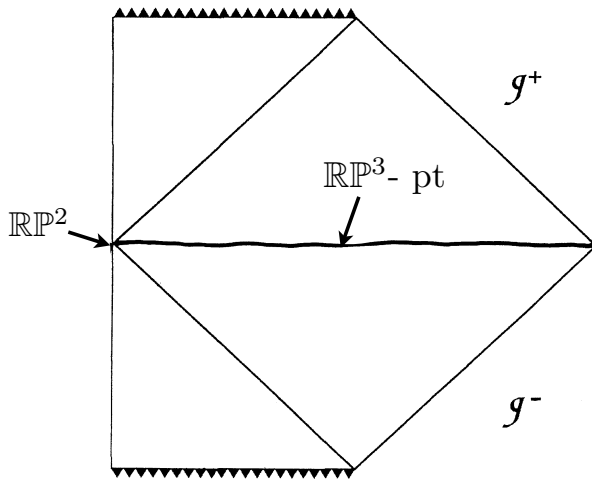
## Immersed MOTS

**Definition 5.1.** A subset  $\Sigma \subset V$  is an immersed MOTS in an initial data set  $(V, h, K)$  if there exists a finite cover  $\tilde{V}$  of  $V$  with covering map  $p : \tilde{V} \rightarrow V$  and a MOTS  $\tilde{\Sigma}$  in  $(\tilde{V}, p^*h, p^*K)$  such that  $p(\tilde{\Sigma}) = \Sigma$ .

A good example is given by the so called  $\mathbb{RP}^3$  geon. This is obtained from the extended Schwarzschild solution:

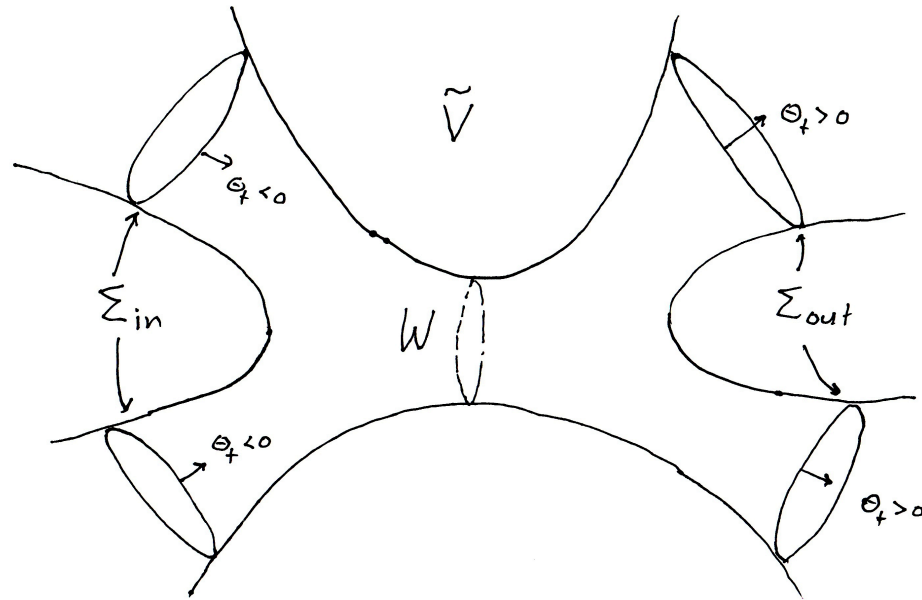
$$M_{Sch} = \mathbb{R}^2 \times S^2 \quad \text{with} \quad ds^2 = (-32m^3/r)e^{r/2m}(-dT^2 + dX^2 + r^2d\Omega^2)$$

by making the identifications,  $X \rightarrow -X$ ,  $p \in S^2 \rightarrow -p \in S^2$



**Lemma 5.5.** *Let  $(V^n, h, K)$  be an  $n$ -dimensional,  $3 \leq n \leq 7$ , AF initial data set. If  $V$  admits a finite nontrivial cover then  $V$  contains an immersed MOTS.*

*Proof.* Let  $\tilde{V}$  be a finite nontrivial cover of  $V$ . Then  $\tilde{V}$  (with data pulled back via the covering map) will be AF with more than one, but finitely many ends, as illustrated below.



Hence  $\tilde{V}$  contains a MOTS. Then the projection via the covering map is an immersed MOTS in  $V$ . □

**Theorem 5.3 (no horizon case).** Let  $(V^3, h, K)$  be a 3-dimensional AF initial data set. If there are no immersed MOTS in  $V$  then  $V$  is diffeomorphic to  $\mathbb{R}^3$ .

*Proof.* By AFness, may write  $V$  as  $V = \mathbb{R}^3 \# N$  where  $N$  is closed (compact without boundary) 3-manifold.

By the **residual finiteness** of  $\pi_1(N)$ , if  $\pi_1(N) \neq 0$ , then  $N$  admits a **finite** nontrivial cover. It follows that  $V$  admits a finite nontrivial cover. But then, by the Lemma,  $V$  contains an immersed MOTS  $\rightarrow\leftarrow$ . Hence,  $\pi_1(N) = 0$  and so  $N \approx S^3$  by Poincaré.  $\therefore V \approx \mathbb{R}^3 \# S^3 \approx \mathbb{R}^3$ . □

Consult [6] to see how the argument is modified to deal with the horizon case.

### Higher dimensions.

By similar techniques the following is shown in [6]

**Theorem 5.6** (Eichmair-G-Pollack). *Let  $(V, h, K)$  be an  $n$ -dimensional,  $3 \leq n \leq 7$ , AF initial data set. If  $V$  does not contain any immersed MOTS. then  $V$  has vanishing first Betti number,  $b_1(V) = 0$ .*

*Proof:*  $b_1(V) \neq 0 \implies V$  contains a nonseparating hypersurface  $\implies V$  admits a finite nontrivial cover  $\implies V$  contains an immersed MOTS by the Lemma.

In very recent work with Andersson, Dahl and Pollack (arXiv, 2015), we have taken an entirely different approach to the study of the topology of the exterior region. We have shown roughly the following.

**Theorem 5.7** (Andersson, Dahl, G., Pollack). *Let  $(V^n, h, K)$  be an  $n$ -dimensional initial data set,  $3 \leq n \leq 7$ , satisfying the DEC, with MOTS boundary  $\Sigma = \partial V$ , and assume there are no MOTS in the ‘exterior region’  $V \setminus \partial V$ .*

*Then the compactification  $V'$  admits a metric  $h'$  of positive scalar curvature, such that the induced metric on  $\Sigma$  is conformal to that induced by  $h$ , and such that  $h'$  is a product metric in a neighborhood of  $\Sigma$ .*

*Remarks:*

- In this set-up, one can then apply certain index theory obstructions, and minimal surface theory obstructions, to obtain restrictions on the topology of  $V$ .
- This result may be viewed as an extension of Theorem 4.3 on the topology of higher dimensional black holes, whereby the metric of positive scalar curvature metric on the boundary can be extended to the one-point compactification of the region exterior to the boundary.
- Much machinery is needed: Existence results for Jang’s equation, Eichmair’s regularity and compactness theory for MOTSs, Eichmair-Huang-Lee-Schoen density result, ...

## References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] P. T. Chruściel, D. E. Grant, and E. Minguzzi, *On differentiability of volume time functions*, 2013, arXiv:1301.2909v1.
- [6] 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.
- [7] R. Emparan and H. S. Reall, *A rotating black ring solution in five dimensions*, *Phys. Rev. Lett.* **88** (2002), no. 10, 101101, 4.
- [8] J. L. Friedman, K. Schleich, and D. M. Witt, *Topological censorship*, *Phys. Rev. Lett.* **71** (1993), no. 10, 1486–1489.
- [9] G. J. Galloway, *On the topology of the domain of outer communication*, *Classical Quantum Gravity* **12** (1995), no. 10, L99–L101.
- [10] ———, *Rigidity of marginally trapped surfaces and the topology of black holes*, *Comm. Anal. Geom.* **16** (2008), no. 1, 217–229.
- [11] ———, *Notes on lorentzian causality*, 2015, ESI-EMS-IAMP Summer School on Mathematical Relativity; notes available at <http://www.math.miami.edu/~galloway/>.
- [12] G. J. Galloway and A. Mendes, *Rigidity of marginally outer trapped 2-spheres*, 2015, arXiv:1506.00611.
- [13] 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.
- [14] D. Gannon, *Singularities in nonsimply connected space-times*, *J. Mathematical Phys.* **16** (1975), no. 12, 2364–2367.
- [15] R. Geroch, *Domain of dependence*, *J. Mathematical Phys.* **11** (1970), 437–449.



- [16] 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.
- [17] G. Horowitz (ed.), *Black holes in higher dimensions*, Cambridge University Press, London, 2012.
- [18] A. Ionescu and S. Klainerman, *Rigidity results in general relativity: a review*, 2015, arXiv:1501.01587.
- [19] S. Klainerman, J. Luk, and I. Rodnianski, *A fully anisotropic mechanism for formation of trapped surfaces in vacuum*, 2013, arXiv:1302.5951v1.
- [20] C. W. Lee, *A restriction on the topology of Cauchy surfaces in general relativity*, Comm. Math. Phys. **51** (1976), no. 2, 157–162.
- [21] R. C. Myers and M. J. Perry, *Black holes in higher-dimensional space-times*, Ann. Physics **172** (1986), no. 2, 304–347.
- [22] Barrett 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.
- [23] R. Penrose, *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14** (1965), 57–59. MR MR0172678 (30 #2897)
- [24] Robert M. Wald, *General relativity*, University of Chicago Press, Chicago, IL, 1984.