Null Geometry and the Einstein Equations

Gregory J. Galloway*
Department of Mathematics
University of Miami

1 Introduction

This paper is based on a series of lectures given by the author at the Cargèse Summer School on Mathematical General Relativity and Global Properties of Solutions of Einstein's Equations, held in Corsica, July 29 - August 10, 2002. The general aim of those lectures was to illustrate with some current examples how the methods of global Lorentzian geometry and causal theory may be used to obtain results about the global behavior of solutions to the Einstein equations. This, of course, is a long standing program, dating back to the singularity theorems of Hawking and Penrose [24]. Here we consider some properties of asymptotically de Sitter solutions to the Einstein equations with (by our sign conventions) positive cosmological constant, $\Lambda > 0$. We obtain, for example, some rather strong topological obstructions to the existence of such solutions, and, in another direction, present a uniqueness result for de Sitter space, associated with the occurence of *eternal* observer horizons. As described later, these results have rather strong connections with Friedrich's results [11, 13] on the nonlinear stability of asymptotically simple solutions to the Einstein equations with $\Lambda > 0$; see also Friedrich's article elewhere in this volume. The main theoretical tool from global Lorentzian geometry used to prove these results is the so-called null splitting theorem [16]. This theorem is discussed here, along with relevant background material.

The paper is divided into sections as follows. In Section 2 we present the basic elements of causal theory, emphasizing those parts of the subject needed for our work. In Section 3 we give a self-contained treatment of the geometry of smooth null hypersurfaces, and present a maximum principle for such hypersurfaces. In Section 4 we extend this maximum principle to the essential C^0 setting, and discuss the null splitting theorem. In Section 5 we consider the aforementioned applications.

*Supported in part by NSF grant # DMS-0104042

2 Elements of causal theory

Much of our work makes use of results from the causal theory of Lorentzian manifolds. In this section we recall some of the basic notions and notations of this subject, with an emphasis on what shall be needed later. There are many excellent treatments of causal theory, all varying somewhat in perspective and degree of rigor; see for example, [31, 24, 29, 33, 3].

Let M^{n+1} be a Lorentzian manifold, i.e., a smooth Hausdorff manifold equipped with a smooth metric $g = \langle\ ,\ \rangle$ of Lorentz signature $(- + \cdots +)$. A vector $X \in T_pM$ is timelike (resp., null, causal, spacelike) provided $\langle X, X\rangle < 0$ (resp., $\langle X, X\rangle = 0$, ≤ 0, > 0). This terminology extends to curves: A smooth curve $t \to \sigma(t)$ is timelike (resp., null, causal, spacelike) provided each of its velocity vectors $\sigma'(t)$ is timelike (resp., null, causal, spacelike). The causal character of curves extends in a natural way to piecewise smooth curves. For each $p \in M$, the set of null vectors at p forms a double cone in T_pM. M is said to be time orientable provided the assignment of a future cone and past cone at each point of M can be made in a continuous manner over M. By a *spacetime*, we mean a connected, time oriented Lorentzian manifold M. Hencforth we restrict attention to spacetimes.

Let ∇ denote the Levi-Civita connection of M. Hence, for vector fields $X = X^a$ and $Y = Y^b$, $\nabla_X Y = X^a \nabla_a Y^b$ denotes the covariant derivative of Y with respect to X. For the most part we use index free notation. By definition, geodesics are curves $t \to \sigma(t)$ of zero covariant acceleration, $\nabla_{\sigma'}\sigma' = 0$.

The Riemann curvature tensor $(X, Y, Z) \to R(X, Y)Z$ is defined by,

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z\,. \tag{2.1}$$

The components of the curvature tensor are determined by the equation, $R(\partial_i, \partial_j)\partial_k = R^\ell{}_{kij}\partial_\ell$. The Ricci tensor and scalar curvature are obtained by tracing, $R_{ij} = R^\ell{}_{i\ell j}$ and $R = g^{ij}R_{ij}$.

We now introduce the notations for futures and pasts. $I^+(p)$ (resp., $J^+(p)$), the timelike (resp., causal) future of $p \in M$, is the set of points $q \in M$ for which there exists a future directed timelike (resp., causal) curve from p to q. Since small deformations of timelike curves remain timelike, the sets $I^+(p)$ are always open. However, the sets $J^+(p)$ need not in general be closed. To emphasize the particular spacetime involved, one sometimes writes $I^+(p, M)$, etc. More generally, for $A \subset M$, $I^+(A)$ (resp., $J^+(A)$), the timelike (resp., causal) future of A, is the set of points $q \in M$ for which there exists a future directed timelike (resp., causal) curve from a point $p \in A$ to q. Note, $I^+(A) = \cup_{p\in A} I^+(p)$, and hence is open.

By variational techniques one can establish the following fundamental causality result; cf. [29, p. 294].

Proposition 2.1 *If $q \in J^+(p) \setminus I^+(p)$, then any future directed causal curve from p to q must be a null geodesic (when suitably parametrized).*

The timelike and causal pasts $I^-(p)$, $J^-(p)$, $I^-(A)$, $J^-(A)$ are defined in a time dual manner.

Many causally defined sets of interest, for example, horizons of various sorts, arise essentially as *achronal boundaries*. By definition, an achronal boundary is a set of the form $\partial I^+(A)$ (or $\partial I^-(A)$), for some $A \subset M$. We wish to describe several important structural properties of achronal boundaries.

Proposition 2.2 *An achronal boundary $\partial I^+(A)$, if nonempty, is a closed achronal C^0 hypersurface in M.*

Recall, an achronal set is a subset of spacetime for which no two points can be joined by a timelike curve. We discuss briefly the proof of the proposition, beginning with the following simple lemma.

Lemma 2.3 *If $p \in \partial I^+(A)$ then $I^+(p) \subset I^+(A)$, and $I^-(p) \subset M \setminus \overline{I^+(A)}$.*

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 $I^+(A)$, it follows that $I^-(q) \cap I^+(A) \neq \emptyset$, and hence $q \in I^+(A)$. The second part of the lemma is proved similarly.

Since $I^+(A)$ is open, it does not meet its boundary. The first part of Lemma 2.3 then implies that $\partial I^+(A)$ is achronal. Lemma 2.3 also implies that $\partial I^+(A)$ is *edgeless*. The edge of an achronal set $S \subset M$ is the set of points $p \in \overline{S}$ such that every neighborhood U of p, contains a timelike curve from $I^-(p, U)$ to $I^+(p, U)$ that does *not* meet S. But Lemma 2.3 shows that for any $p \in \partial I^+(A)$, any timelike curve from $I^-(p)$ to $I^+(p)$ must meet $\partial I^+(A)$. The remainder of Proposition 2.2 then follows from the fact (cf. [29, p. 413]) that an achronal set without edge points is a C^0 hypersurface of M.

The next result shows that, in general, large portions of achronal boundaries are ruled by null geodesics.

Proposition 2.4 *Let $A \subset M$ be closed. Then each $p \in \partial I^+(A) \setminus A$ lies on a null geodesic contained in $\partial I^+(A)$, which either has a past end point on A, or else is past inextendible in M.*

We give a sketch of the proof. Choose a sequence of points $\{p_n\} \subset I^+(A)$ such that $p_n \to p$, and let γ_n be a past directed timelike curve from p_n to a point of A. By passing to a subsequence if necessary, $\{\gamma_n\}$ converges, in a suitable sense, to a past directed causal curve γ from p, which must be contained in $\partial I^+(A)$. We overlook here the technical difficulty that the limit curve γ need not be smooth; this, can be dealt with, however; see, [3, Sections 3.3, 14.1]. Since each segment of γ is both causal and achronal, it follows from Proposition 2.1 that γ is a null geodesic. Now, each γ_n is past inextendible in $M \setminus A$, and hence so is γ. Thus γ either has a past end point on A or is past inextendible in M.

Finally we make some remarks and recall some facts about global hyperbolicity. A spacetime M is *strongly causal* at $p \in M$ provided there are arbitrarily small neighborhoods U of p such that any causal curve γ which starts in, and leaves, U never returns to U. M is *strongly causal* if it is strongly causal at each of its points. Thus, heuristically speaking, M is strongly provided there are no closed or "almost closed" causal curves in M.

A spacetime M is said to be *globally hyperbolic* provided (i) M is strongly causal and (ii) the sets $J^+(p) \cap J^-(q)$ are compact for all $p, q \in M$. The latter condition rules out the occurence of naked singularities, and hence global hyperbolicity is closely related to the notion of cosmic censorship. Global hyperbolicity is also related to the existence of an ideal initial value hypersurface in spacetime. There are slight variations in the literature on the definition of a *Cauchy surface* for a spacetime M. Here we adopt the following definition: A Cauchy surface for M is an achronal C^0 hypersurface S of M which is met by every inextendible causal curve in M. We now recall several fundamental results.

Proposition 2.5 *M is globally hyperbolic if and only if M admits a Cauchy surface. If S is a Cauchy surface for M then M is homeomorphic to $\mathbb{R} \times S$.*

Along similar lines, one has that any two Cauchy surfaces in a given globally hyperbolic spacetime are homeomorphic. Hence, according to Proposition 2.5, any nontrivial topology in a globally hyperbolic spacetime must reside in its Cauchy surfaces. The following fact is often useful.

Proposition 2.6 *If S is a compact achronal hypersurface in a globally hyperbolic spacetime M then S must be a Cauchy surface for M.*

Cauchy surfaces can be characterized in terms of the domain of dependence. Let S be an achronal subset of M. The *future domain of dependence* of S is the set $D^+(S)$ consisting of all points $p \in M$ such that every past inexendendible causal curve from p meets S. Physically, $D^+(S)$ is the part of spacetime to the future of S that is predictable from S. The *future Cauchy horizon* of S, $H^+(S)$, is the future boundary of $D^+(S)$; formally, $H^+(S) = \overline{D^+(S)} \setminus I^-(D^+(S))$. Physically, $H^+(S)$ is the future limit of the part of spacetime predictable from S. The past domain of dependence $D^-(S)$ and past Cauchy horizon $H^-(S)$ are defined in a time-dual manner. Set $D(S) = D^+(S) \cup D^-(S)$ and $H(S) = H^+(S) \cup H^-(S)$; one has $\partial D(S) = H(S)$. Then, it is a basic fact that an achronal set S is a Cauchy surface for M iff $D(S) = M$ iff $H(S) = \emptyset$. Cauchy horizons have structural properties similar to achronal boundaries, as indicated in the following.

Proposition 2.7 *Let S be an achronal subset of a spacetime M. Then $H^+(S) \setminus \mathrm{edge}\, S$, if nonempty, is an achronal C^0 hypersurface of M ruled by null geodesics, each of which either is past inextendible M or has past end point on* $\mathrm{edge}\, S$.

The proof of Proposition 2.7 is roughly similar to the proofs of Propositions 2.2 and 2.4. Proposition 2.6 can now be easily proved by showing, with the aid of Proposition 2.7, that $H(S) = \emptyset$.

We conclude this brief presentation with the following basic facts about global hyperbolicity.

Proposition 2.8 *Let M be a globally hyperbolic spacetime. Then,*

(i) M is causally simple, i.e., the sets $J^{\pm}(A)$ are closed, for all compact $A \subset M$.

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

3 The geometry of smooth null hypersurfaces.

Here we review some aspects of the geometry of null hypersurfaces, along the lines developed in [16, 26], and present a maximum principle for such hypersurfaces.

Let (M^{n+1}, g) be a spacetime, with $n \geq 2$. A *(smooth) null hypersurface* in M is a smooth co-dimension one embedded submanifold S of M such that the pullback of the metric g to S is degenerate. Because of the Lorentz signature of g, the null space of the pullback is one dimensional at each point of S. Hence, every null hypersurface S admits a smooth nonvanishing future directed null vector field $K \in \Gamma TS$ such that the normal space of K at $p \in S$ coincides with the tangent space of S at p, i.e., $K_p^{\perp} = T_pS$ for all $p \in S$. It follows, in particular, that tangent vectors to S not parallel to K are spacelike. Note also that the vector field K is unique up to a positive (pointwise) scale factor. The following fact is fundamental.

Proposition 3.1 *The integral curves of K when suitably parameterized, are null geodesics.*

Proof: It suffices to show that $\nabla_K K = \lambda K$. This will follow by showing that at each $p \in S$, $\nabla_K K \perp T_pS$, i.e., $\langle \nabla_K K, X\rangle = 0$ for all $X \in T_pS$. 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 we obtain,

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

and hence,

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

□

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

Since K is orthogonal to S we can introduce the null Weingarten map and null second fundamental form of S with respect K in a manner roughly analogous to what is done for spacelike hypersurfaces or hypersurfaces in a Riemannian manifold. For technical reasons, one works "mod K", as described below.

We introduce the following equivalence relation on tangent vectors: For $X, X' \in T_pS$, $X' = X \bmod K$ if and only if $X' - X = \lambda K$ for some $\lambda \in \mathbb{R}$. Let $\overline{X}$ denote the equivalence class of X. Let $T_pS/K = \{\overline{X} : X \in T_pS\}$, and $TS/K = \cup_{p\in S} T_pS/K$. TS/K, the mod K tangent bundle of S, is a smooth rank $n-1$ vector bundle over 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 $h(\overline{X}, \overline{Y}) = \langle X, Y\rangle$. A simple computation shows that h is well-defined.

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 $b(\overline{X}) = \overline{\nabla_X K}$. It is easily checked that b is well-defined. Note if $\widetilde{K} = fK$, $f \in C^\infty(S)$, is any other future directed null vector field tangent to S, then $\nabla_X \widetilde{K} = f\nabla_X K \bmod K$. It follows that the Weingarten map b of S is unique up to positive scale factor and that b at a given point $p \in S$ depends only on the value of K at p.

A standard computation shows, $h(b(\overline{X}), \overline{Y}) = \langle \nabla_X K, Y\rangle = \langle X, \nabla_Y K\rangle = h(\overline{X}, b(\overline{Y}))$. Hence b is self-adjoint with respect to h. 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 $B(\overline{X}, \overline{Y}) = h(b(\overline{X}), \overline{Y}) = \langle \nabla_X K, Y\rangle$. Since b is self-adjoint, B is symmetric. We say that S is *totally geodesic* iff $B \equiv 0$. This has the usual geometric meaning: If S is totally geodesic then any geodesic in M starting tangent to S stays in S. This follows from the fact that, when S is totally geodesic, the restriction to S of the Levi-Civita connection of M defines an affine connection on S. Null hyperplanes in Minkowski space are totally geodesic, as is the event horizon in Shwarzschild spacetime.

The *null mean curvature* of S with respect to K is the smooth scalar field θ on S defined by, $\theta = \operatorname{tr} b$. Let Σ be the intersection of S with a hypersurface in M which is transverse to K near $p \in S$; Σ will be an $n-1$ dimensional spacelike submanifold of M. 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,

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

Thus, the null mean curvature gives a measure of the divergence towards the future of the null generators of S. Note that if $\widetilde{K} = fK$ then $\widetilde{\theta} = f\theta$. Thus the null mean curvature inequalities $\theta \geq 0$, $\theta \leq 0$, etc., are invariant under positive rescaling of K. In Minkowski space, a future null cone $S = \partial I^+(p) \setminus \{p\}$ (resp., past null cone $S = \partial I^-(p) \setminus \{p\}$) has positive null mean curvature, $\theta > 0$ (resp., negative null mean curvature, $\theta < 0$).

The null second fundamental form of a null hypersurface obeys a well-defined comparison theory roughly similar to the comparison theory satisfied by the second fundamental forms of a family of parallel spacelike hypersurfaces (cf., Eschenburg [9], which we follow in spirit).

Let $\eta : (a, b) \to M$, $s \to \eta(s)$, be a future directed affinely parameterized null geodesic generator of S. For each $s \in (a, b)$, let

$$b(s) = b_{\eta'(s)} : T_{\eta(s)}S/\eta'(s) \to T_{\eta(s)}S/\eta'(s) \tag{3.2}$$

be the Weingarten map based at $\eta(s)$ with respect to the null vector $K = \eta'(s)$. The one parameter family of Weingarten maps $s \to b(s)$, obeys the following Ricatti equation,

$$b' + b^2 + R = 0. \tag{3.3}$$

Here $'$ denotes covariant differentiation in the direction $\eta'(s)$: 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,

$$b'(\overline{X}) = b(\overline{X})' - b(\overline{X'})\,. \tag{3.4}$$

$R : T_{\eta(s)}S/\eta'(s) \to T_{\eta(s)}S/\eta'(s)$ is the curvature endomorphism defined by $R(\overline{X}) = \overline{R(X, \eta'(s))\eta'(s)}$.

We indicate the proof of Equation 3.3. 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, restricted to η, is the velocity vector field to η, i.e., for each s near s_0, $K_{\eta(s)} = \eta'(s)$. 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 3.4, 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 3.3.

By taking the trace of 3.3 we obtain the following formula for the derivative of the null mean curvature $\theta = \theta(s)$ along η,

$$\theta' = -\mathrm{Ric}(\eta', \eta') - \sigma^2 - \frac{1}{n-1}\theta^2, \tag{3.5}$$

where σ, the shear scalar, is the trace of the square of the trace free part of b. Equation 3.5 is the well-known Raychaudhuri equation (for an irrotational null geodesic congruence) of relativity. This equation shows how the Ricci curvature of spacetime influences the null mean curvature of a null hypersurface.

The following proposition is a standard application of the Raychaudhuri equation, a C^0 version of which will be needed later.

Proposition 3.2 *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 mean curvature, $\theta \geq 0$.*

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,

$$\frac{d\theta}{ds} \leq -\frac{1}{n-1}\theta^2\,, \tag{3.6}$$

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

$$\frac{d}{ds}\left(\frac{1}{\theta}\right) \geq \frac{1}{n-1}\,, \tag{3.7}$$

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

Proposition 3.2 implies, under the given assumptions, that cross sections of S do not decrease in area as one moves towards the future. Proposition 3.2 is the most rudimentary form of Hawking's black hole area theorem. For a recent study of the area theorem, with a focus on issues of regularity, see [7].

3.1 Maximum principle for smooth null hypersurfaces

We present a maximum principle for smooth null hypersurfaces analogous to that for hypersurfaces in Riemannian manifolds and spacelike hypersurfaces in Lorentzian manifolds. Because of its natural invariance we restrict attention to the zero mean curvature case.

Theorem 3.3 *Let S_1 and S_2 be smooth null hypersurfaces in a spacetime M. Suppose,*

(1) S_1 and S_2 meet at $p \in M$ and S_2 lies to the future side of S_1 near p, and

(2) the null mean curvature scalars θ_1 of S_1, and θ_2 of S_2, satisfy, $\theta_2 \leq 0 \leq \theta_1$.

Then S_1 and S_2 coincide near p and this common null hypersurface has null mean curvature $\theta = 0$.

The heuristic here is that since the generators of S_1 are nonconverging, and the generators of S_2, which lie to the future of S_1 are nondiverging, the two sets of generators are forced to agree and form a nonexpanding congruence.

Proof: We give a sketch of the proof; for details, see [16]. (N.B. There is a bad typo in the statement of Theorem II.1 in [16], in which the mean curvature inequalities appear reversed.)

S_1 and S_2 have a common null direction at p. Let Q be a timelike hypersurface in M passing through p and transverse to this direction. By taking Q small enough, the intersections $\Sigma_1 = S_1 \cap Q$ and $\Sigma_2 = S_2 \cap Q$ will be smooth spacelike hypersurfaces in Q, with Σ_2 to the future side of Σ_1 near p.

Σ_1 and Σ_2 may be expressed as graphs over a fixed spacelike hypersurface V in Q (with respect to Gaussian normal coordinates), $\Sigma_1 = \text{ graph}\,(u_1)$, $\Sigma_2 = \text{ graph}\,(u_2)$. Let,

$$\theta(u_i) = \theta_i|_{\Sigma_i = \text{ graph}\,(u_i)}\,, \quad i = 1, 2\,.$$

By a computation,

$$\theta(u_i) = H(u_i) + \text{ lower order terms}\,,$$

where H is the mean curvature operator on spacelike graphs over V in Q. (The lower order terms involve the second fundamental form of Q.) Thus θ is a second order quasi-linear elliptic operator. In the present situation we have:

(i) $u_1 \le u_2$, and $u_1(p) = u_2(p)$.

(ii) $\theta(u_2) \le 0 \le \theta(u_1)$.

A suitable version of the strong maximum principle then implies, $u_1 = u_2$. Thus, Σ_1 and Σ_2 agree near p. The normal geodesics to Σ_1 and Σ_2 in M will then also agree. This implies that S_1 and S_2 agree near p. □

4 C^0 null hypersurfaces and the null splitting theorem

The usefulness of the maximum principle for smooth null hypersurfaces presented in the previous section is limited by the fact that the most interesting null hypersurfaces arising in general relativity, e.g., horizons of various sorts, are C^0, but in general not C^1. Such hypersurfaces often arise as (the null portions of) achronal boundaries $\partial I^\pm(A)$. For example, (i) in black hole spacetimes, the (future) event horizon is defined as, $H = \partial I^-(\mathcal{J}^+)$, where $\mathcal{J}^+$ is *future null infinity* (roughly, the ideal boundary of future end points of null geodesics that escape to infinity), and (ii) the observer horizon of an observer (future inextendible timelike curve) γ is defined as $\partial I^-(\gamma)$. The aim of this section is to present a maximum principle for C^0 null hypersurfaces, similar in spirit to the maximum principle for C^0 spacelike hypersurfaces obtained in [2], and to describe how this is used to prove the null splitting theorem.

From the properties of achronal boundaries discussed in Section 2, a set of the form $S = \partial I^-(A) \setminus A$, with A closed, is an achronal C^0 hypersurface ruled by null geodesics which are future inextendible in S (in fact, which are future inextendible in the sub-spacetime $M \setminus A$). Though future inextendible in S, the null geodesics ruling S (i.e., the *null generators* of S) may have past end points on S. Consider, for example, the set $S = \partial I^-(A) \setminus\ A$, where A consists of two disjoint closed disks in the $t = 0$ slice of Minkowski 3-space. This surface, which represents the merger of two truncated cones, has a "crease", i.e., a curve of nondifferentiable points (corresponding to the intersection of the two cones) but which otherwise is a smooth null hypersurface. The null generators of S that reach the crease, leave S when extended to the past.

Sets of the form $\partial I^{\pm}(A) \setminus A$, A closed, are models for our notion of C^0 null hypersurfaces.

Definition 4.1 *A C^0* **future** *(resp.,* **past***) null hypersurface is a locally achronal C^0 hypersurface ruled by null geodesics which are future (resp., past) inextendible in S.*

C^0 null hypersurfaces do not in general have null mean curvature in the classical sense, but may obey null mean curvature inequalities in the *support sense*, as described below.

Let S be a C^0 future null hypersurface, and let $p \in S$. A smooth null hypersurface W is said to be a past support hypersurface for S at p provided W passes through p and lies to the past of S near p. We note that any C^0 future null hypersurface S is supported from below at each point p by a smooth null hypersurface. Indeed, one may take $W = \partial I^-(q, U) \setminus \{q\}$, where U is a convex normal neighborhood of p, and $q \in U$ is on a null generator of S from p, slightly to the future of p. If S is actually smooth then, by an elementary comparison, $\theta_S(p) \geq \theta_W(p)$, provided the future directed null vector fields K_S and K_W used to define the null second fundamental forms on S and W, respectively, are scaled to agree at p. (Time-dual statement holds for C^0 past null hypersurfaces.) These considerations lead to the following definition.

Definition 4.2 *Let S be a C^0 future null hypersurface in M. We say that S has null mean curvature $\theta \geq 0$* **in the support sense** *provided for each $p \in S$ and for each $\epsilon > 0$ there exists a smooth (at least C^2) null hypersurface $W_{p,\epsilon}$ such that,*

(1) $W_{p,\epsilon}$ is a past support hypersurface for S at p, and

(2) the null mean curvature of $W_{p,\epsilon}$ at p satisfies $\theta_{p,\epsilon} \geq -\epsilon$.

For this definition, it is assumed that the null vectors have been uniformly scaled, e.g., have unit length with respect to a fixed background Riemannian metric, otherwise the inequality in (2) would be meaningless. Note that if S is smooth and satisfies

Definition 4.2 then $\theta_S \geq 0$ in the usual sense. If S is a C^0 *past* null hypersurface, one defines $\theta \leq 0$ *in the support sense* in an analogous manner in terms of *future* support hypersurfaces.

As a simple illustration of Definition 4.2, consider the future null cone $S = \partial I^+(p)$, where p is a point in Minkowski space. S is a C^0 future null hypersurface having mean curvature $\theta \geq 0$ in the support sense: One may use null hyperplanes, even at the vertex, as support hypersurfaces. A less trivial illustration is provided by the next proposition, which is a C^0 version of Proposition 3.2.

Proposition 4.1 *Let M be a spacetime which satisfies the NEC. Suppose S is a C^0 future null hypersurface in M whose null generators are future geodesically complete. Then S has null mean curvature $\theta \geq 0$ in the support sense.*

Proof: By restricting attention to a sufficiently small neighborhood of S, we may assume without loss of generality that S is globally achronal. Recall that if a null geodesic η contains a pair null conjugate points, then points of η are timelike related [29, p. 296]. Thus it follows that the null generators of S are free of null conjugate points.

Given $p \in S$, let $\eta : [0, \infty) \to S \subset M$, $s \to \eta(s)$, be a future directed affinely parametrized null geodesic generator of S from $p = \eta(0)$. For any $r > 0$, consider a small pencil of past directed null geodesics from $\eta(r)$ about η. This pencil, taken sufficiently small, will form a smooth (caustic free) null hypersurface $W_{p,r}$ containing $\eta([0, r))$. Moreover, since $W_{p,r} \subset J^-(S)$, it will be a lower support hypersurface for S at p.

Let $\theta = \theta(s)$, $0 \leq s < r$, be the null mean curvature of $W_{p,r}$ along $\eta|_{[0,r)}$, where, as in the notaton of Equation (3.2), $b(s) = b_{\eta'(s)}$. The differential inequality (3.6), which holds in the present situation, together with the initial condition $\theta(r) = -\infty$, implies,

$$\theta(0) \geq -\frac{n-1}{r}\,.$$

Since r can be taken arbitrarily large, the proposition follows $\Box$

Proposition 4.1 applies, in particular, to future event horizons in black hole spacetimes, in which the null generators are future complete. This fact provided the initial impetus for the development of a proof of the black hole area theorem which does not require the imposition of smoothness assumptions on the horizon; cf. [7].

We now present the maximum principle for C^0 null hypersurfaces.

Theorem 4.2 (Maximum Principle for C^0 null hypersurfaces.) *Let S_1 be a C^0 future null hypersurface and S_2 be C^0 past null hypersurface in a spacetime M. Suppose,*

(1) S_1 and S_2 meet at $p \in M$, with S_2 lying to the future side of S_1 near p, and

(2) S_1 and S_2 have null mean curvatures satisfying, $\theta_2 \leq 0 \leq \theta_1$ in the support sense.

Then S_1 and S_2 coincide near p and form a smooth null hypersurface with null mean curvature $\theta = 0$.

Comments on the proof: We first mention that, for simplicity, we have omitted a technical assumption from the statement of the theorem; in the usual geometric applications, this technical condition is satisfied automatically; cf., [16] for details. Although there are some significant technical issues, the proof proceeds more or less along the lines of the proof of Theorem 3.3.

One first observes that the point p is an interior point of a null generator common to both S_1 and S_2 near p. As before, one intersects S_1 and S_2 with a timelike hypersurface Q through p which is transverse to this generator. By taking Q small enough, the intersections $\Sigma_1 = S_1 \cap Q$ and $\Sigma_2 = S_2 \cap Q$ will be acausal C^0 hypersurfaces in Q passing through p, with Σ_2 to the future of Σ_1. One can again express Σ_1 and Σ_2 as graphs over a fixed smooth spacelike hypersurface $V \subset Q$ (with respect to Gaussian normal coordinates about V), $\Sigma_i = \mathrm{graph}\, u_i$, $i = 1, 2$. The functions u_1 and u_2 are Lipschitz functions on V satisfying, $u_1 \leq u_2$ and $u_1(p) = u_2(p)$. As in the proof of Theorem 3.3, let θ denote the null mean curvature operator. The null mean curvature assumption then implies that u_1 and u_2 satisfy the differential inequalities, $\theta(u_2) \leq 0 \leq \theta(u_1)$, in the support function sense. By the weak version of the strong maximum principle obtained in [2], which is a nonlinear generalization of Calabi's [5] weak version of the Hopf maximum principle, one concludes that u_1 and u_2 are smooth and agree near p. Thus, Σ_1 and Σ_2 are smooth spacelike hypersurfaces in Q which agree near p. One can then show that S_1 and S_2 are obtained locally by exponentiating normally out along a common smooth null orthogonal vector field along $\Sigma_1 = \Sigma_2$. The conclusion of Theorem 4.2 now follows. $\square$

4.1 The null splitting theorem

The main motivation for establishing a maximum principle for C^0 null hypersurfaces was the realization that such a result could be used to settle a problem that arose in the 80's concerning the occurence of lines in spacetime. Recall, in a Riemannian manifold, a *line* is an inextendible geodesic, each segment of which has minimal length, while in a spacetime, a *timelike line* is an inextendible timelike geodesic, each segment of which has maximal length among causal curves joining its end points. The classical Cheeger-Gromoll splitting theorem [6] describes the rigidity of Riemannian manifolds of nonnegative Ricci curvature which contain a line. (Note that a complete Riemannian manifold with strictly positive Ricci curvature cannot contain any lines.) The standard Lorentzian splitting theorem [10, 15, 28], which is an exact Lorentzian analogue of the Cheeger-Gromoll splitting theorem, describes the rigidity of spacetimes obeying the strong energy condition, $\mathrm{Ric}\,(X, X) \geq 0$ for all timelike vectors X, which contain a timelike line. Yau [34] posed the problem of establishing a Lorentzian analogue of the Cheeger-Gromoll splitting theorem as an approach to establishing the rigidity of the Hawking-Penrose singularity theorems; see [3] for a more detailed discussion of these matters, as well as a nice presentation of the proof of the Lorentzian splitting theorem.

But here we are interested in null geometry. Motivated by the more standard cases discussed above, a *null line* in spacetime is defined to be an inextendible null geodesic which is globally achronal, *i.e.*, for which no two points can be joined by a timelike curve. (Hence, each segment of a null line is maximal with respect to the Lorentzian arc length functional.) We emphasize that the condition of being a null line is a global one. Although each sufficiently small segment of a null geodesic is achronal, this achronality need not hold in the large: Consider, for example a null geodesic winding around a flat spacetime cylinder (closed in space); eventually points on the null geodesic are timelike related. Null lines arise naturally in causal arguments; recall, for example, that sets of the form $\partial I^{\pm}(A) \setminus A$, A closed, are ruled by null geodesics which are necessarily achronal. Null lines have arisen, by various constructions, in the proof of numerous results in general relativity; see, for example, [25, 18, 32, 22, 20]. All of the null geodesics in Minkowski space, de Sitter space and anti-de Sitter space are null lines. The null generators of the event horizon in extended Schwarzchild spacetime are null lines.

In analogy with the Lorentzian splitting theorem, one expects spacetimes which obey the NEC and contain a null line to exhibit some sort of *rigidity*, as suggested by the following considerations: The NEC tends to focus congruences of null geodesics, which can lead to the occurence of null conjugate points. But a null geodesic containing a pair of null conjugate points cannot be achronal. Thus we expect that a spacetime which obeys the NEC and contains a complete null line should be special in some way. The question, which arose in the 80's, after the proof of the Lorentzian splitting theorem, as to what this rigidity should be, is addressed in the following theorem.

Theorem 4.3 *Let M be a null geodesically complete spacetime which obeys the NEC. If M admits a null line η then η is contained in a smooth properly embedded achronal* **totally geodesic** *null hypersurface S.*

The simplest illustration of Theorem 4.3 is Minkowski space: Each null line ℓ in Minkowski space is contained in a unique null hyperplane Π.

Proof: The proof is an application of the maximum principle for C^0 null hypersurfaces. For simplicity we shall assume M is strongly causal; this however is not required; see [16] for details.

By way of motivation, note that the null plane Π in Minkowski space considered above can be realized as the limit of the future null cone $\partial I^+(x)$ as x goes to past null infinity along the null line ℓ. Π can also be realized as the limit of the past null cone $\partial I^-(x)$ as x goes to future null infinity along the null line ℓ. In fact, one sees that $\Pi = \partial I^+(\ell) = \partial I^-(\ell)$.

Thus, in the setting of Theorem 4.3, consider the achronal boundaries $S_+ = \partial I^+(\eta)$ and $S_- = \partial I^-(\eta)$. By results discussed in Section 2, S_+ and S_- are closed achronal C^0 hypersurfaces in M. Since η is achronal, it follows that S_+ and S_- both contain η. For simplicity, assume S_+ and S_- are connected (otherwise restrict attention to the component of each containing η). The proof then consists of showing that S_+ and S_- agree and form a smooth totally geodesic null hypersurface. (In a vague sense, this corresponds to showing, in the proof of the Lorentzian splitting theorem, that the level sets of the Busemann functions $b^{\pm} = 0$ associated to the timelike line coincide, which partially motivates thinking of Theorem 4.3 as a splitting theorem.)

We claim that S_- is a C^0 *future* null hypersurface whose generators are future complete. The assumption of strong causality implies that η is a closed subset of spacetime. Then, by (the time-dual of) Proposition 2.4, each point $p \in S_- \setminus \eta$ is on a null geodesic $\sigma \subset S_-$ which either is future inextendible in M or else has a future endpoint on η. In the latter case, σ meets η at an angle, and Proposition 2.1 then implies that there is a timelike curve from a point on σ to a point on η, violating the achronality of S_-. Thus, S_- is ruled by null geodesics which are future inextendible in M, and, hence by the completeness assumption, future complete, which establishes the claim. In a time dual manner, S_+ is a C^0 *past* null hypersurface whose generators are past complete.

Thus, by Proposition 4.1 and its time-dual, S_- and S_+ have null mean curvatures satisfying, $\theta_+ \le 0 \le \theta_-$, in the support sense. Let q be a point of intersection of S_+ and S_-. S_+ necessarily lies to the future side of S_- near q. We may now apply Theorem 4.2 to conclude that S_+ and S_- agree near q, to form a smooth null hypersurface having null mean curvature $\theta = 0$. A fairly straight forward continuation argument shows that $S_+ = S_- = S$ is a smooth null hypersurface with $\theta = 0$. By setting $\theta = 0$ in the Raychaudhuri equation (3.5), and using the NEC, we see that the shear σ must vanish, and hence S is totally geodesic. □

In the next section we consider some applications of Theorem 4.3.

5 Some global properties of asymptotically de Sitter spacetimes

In this section we present some global results for spacetimes M^{n+1} obeying the Einstein equations,

$$R_{ij} - \frac{1}{2}Rg_{ij} + \Lambda g_{ij} = 8\pi T_{ij}\,. \tag{5.8}$$

We will be concerned primarily with spacetimes that satisfy the null energy condition (NEC). In view of the Einstein equations, the NEC may be expressed in terms of the energy-momentum tensor $\mathcal{T} = T_{ij}$, as the condition, $\mathcal{T}(X,X) \ge 0$ for all null vectors X. (Note, in particular, that the NEC is insensitive to the sign of the cosmological constant.) In some situations we will specialize to the vacuum case, $T_{ij} = 0$, in which case the Einstein equations become,

$$R_{ij} = \lambda g_{ij} \tag{5.9}$$

where $\lambda = 2\Lambda/(n-1)$.

We mainly restrict attention to solutions of the Einstein equations with *positive* comsological constant, $\Lambda > 0$. By our sign conventions, de Sitter space, which may be expressed in local coordinates as,

$$M = \mathbb{R} \times S^n\,, \qquad ds^2 = -dt^2 + \cosh^2 t\, d\Omega^2 \tag{5.10}$$

is a vacuum solution of the Einstein equations with $\Lambda > 0$. (We have actually taken $\Lambda = n(n-1)/2$ in (5.10)). Thus, we will be typically dealing with spacetimes which behave asymptotically like de Sitter space. There has been increased interest in such spacetimes in recent years due, firstly, to observations concerning the rate of expansion of the universe, suggesting the presence of a positive cosmological constant in our universe, and, secondly, due to recent efforts to understand quantum gravity on de Sitter space via, for example, some de Sitter space version of the AdS/CFT correspondence (see [4] and references cited therein).

5.1 Asymptotically simple and de Sitter spacetimes

We use Penrose's notion of conformal infinity [30] to make precise what it means for spacetime to be asymptotically de Sitter. Recall, this notion is based on the way in which the standard Lorentzian space forms, Minkowski space, de Sitter space and anti-de Sitter space, conformally imbed into the Einstein static universe $(\mathbb{R} \times S^n, -du^2 + d\Omega^2)$; see the article of Friedrich in this volume for further discussion. Under the transformation $u = \tan^{-1}(e^t) - \pi/4$, the metric (5.10) becomes

$$ds^2 = \frac{1}{\cos^2(2u)}(-du^2 + d\Omega^2)\,. \tag{5.11}$$

Thus, de Sitter space conformally imbeds onto the region $\pi/4 < u < \pi/4$ in the Einstein static universe. Future conformal infinity $\mathcal{J}^+$ (resp., past conformal infinity $\mathcal{J}^-$) is represented by the *spacelike* slice $u = \pi/4$ (resp., $u = -\pi/4$). This serves to motivate the following definition.

Definition 5.1 *A spacetime* (M, g) *is* **asymptotically de Sitter** *provided there exists a spacetime-with-boundary* $(\tilde{M}, \tilde{g})$ *and a smooth function* Ω *on* $\tilde{M}$ *such that*

- *(a)* M *is the interior of* $\tilde{M}$*; hence* $\tilde{M} = M \cup \mathcal{J}$, $\mathcal{J} = \partial \tilde{M}$.
- *(b)* $\tilde{g} = \Omega^2 g$*, where (i)* $\Omega > 0$ *on* M*, and (ii)* $\Omega = 0$, $d\Omega \neq 0$ *along* $\mathcal{J}$.
- *(c)* $\mathcal{J}$ *is spacelike.*

In general, $\mathcal{J}$ decomposes into two disjoint sets, $\mathcal{J} = \mathcal{J}^+ \cup \mathcal{J}^-$ where $\mathcal{J}^+ \subset I^+(M, \tilde{M})$ and $\mathcal{J}^- \subset I^-(M, \tilde{M})$. $\mathcal{J}^+$ is future conformal infinity and $\mathcal{J}^-$ is past conformal infinity. It is to be understood in the above definition that both $\mathcal{J}^+$ and $\mathcal{J}^-$ are nonempty. If a spacetime M obeys Definition 5.1 with $\mathcal{J}^- = \emptyset$ (resp., $\mathcal{J}^+ = \emptyset$), we will say that M is *future* (resp., *past*) asymptotically de Sitter. Expanding dust filled FRW models, which are solutions to the Einstein equations with $\Lambda > 0$, typically begin with a big bang singularity, cf. [8, Chpt. 23]. These cosmological models are future asymptotically de Sitter, but not past asymptotically de Sitter. We remark that no a priori assumption is made about the topology of $\mathcal{J}^+$ and/or $\mathcal{J}^-$.

Definition 5.2 *An asymptotically de Sitter spacetime is* **asymptotically simple** *provided each inextendible null geodesic in M has a future end point on $\mathcal{J}^+$ and a past end point on $\mathcal{J}^-$.*

Thus, spacetime is asymptotically simple provided each null geodesic extends to infinity both to the future and the past. Schwarzschild de Sitter spacetime (see e.g., [23, 4]), which represents a Schwarzschild black hole in a de Sitter background, is an interesting example of a spacetime which is asymptotically de Sitter, but not asymptotically simple: Null geodesics entering the black hole cannot escape to infinity. In an obvious modification of the definition, we may also refer to spacetimes which are *future* (resp., *past*) asymptotically simple.

There are connections between asymptotic simplicity and the causal structure of spacetime, as illustrated in the next proposition.

Proposition 5.1 *Let M be a future asymptotically de Sitter spacetime, with future conformal infinity $\mathcal{J}^+$.*

(i) If M is future asymptotically simple then M is globally hyperbolic.

(ii) If M is globally hyperbolic and $\mathcal{J}^+$ is compact then M is future asymptotically simple.

In either case, the Cauchy surfaces of M are homeomorphic to $\mathcal{J}^+$.

Proof: By extending $M \cup \mathcal{J}^+$ a little beyond $\mathcal{J}^+$, one may obtain a spacetime without boundary M' such that $\mathcal{J}^+$ is achronal and has no future Cauchy horizon in M', $H^+(\mathcal{J}^+, M') = \emptyset$.

Suppose M is future asymptotically simple. We claim that $\mathcal{J}^+$ is a Cauchy surface for M'. Since, by construction, $H^+(\mathcal{J}^+, M') = \emptyset$, we need only show that $H^-(\mathcal{J}^+, M') = \emptyset$. If $H^-(\mathcal{J}^+, M') \neq \emptyset$, then, by Proposition 2.7 and the fact that $\mathcal{J}^+$ is edgeless, there exists a null geodesic η contained in $H^-(\mathcal{J}^+, M')$ which is future inextendible in M'. By asymptotic simplicity, η must meet $\mathcal{J}^+$, and hence enter $I^+(\mathcal{J}^+, M')$. But this violates the achronality of $\mathcal{J}^+$. Thus, M' is globally hyperbolic, from which it easily follows that M is, as well. This proves part (i).

Now assume M is globally hyperbolic and $\mathcal{J}^+$ is compact. Since any Cauchy surface for M is clearly a Cauchy surface for M', M' is globally hyperbolic. Then, by Proposition 2.6, $\mathcal{J}^+$ is a Cauchy surface for M'. Hence, any future inextendible null geodesic in M' starting in M meets $\mathcal{J}^+$. It follows that M is future asymptotically simple, which proves part (ii). We leave the proof of the final statement to the reader. □

5.2 A uniqueness theorem for de Sitter space.

We present here a uniqueness theorem for de Sitter space associated with the occurence of null lines. Every inextendible null geodesic in de Sitter space is a null line. This fact may be understood in terms of the causal structure of de Sitter space. The *observer horizon* of an observer (future inextendible timelike curve) γ is, by definition, the achronal boundary $\partial I^-(\gamma)$. The observer horizon describes the limit of the region of spacetime ultimately observable by γ. In a future asymptotically de Sitter spacetime, every observer has a nontrivial observer horizon, as follows from the fact that $\mathcal{J}^+$ is spacelike. In de Sitter space, the observer horizon of every observer γ is *eternal*, i.e., extends from $\mathcal{J}^+$ all the way back to $\mathcal{J}^-$. If q is the future end point of γ on $\mathcal{J}^+$, then the observer horizon $\partial I^-(\gamma)$ may be viewed as the past null cone from q, which, in de Sitter space, reconverges right on $\mathcal{J}^-$ at a point q' "antipodal" to q. By properties of achronal boundaries, $\partial I^-(\gamma)$ is ruled by achronal null geodesics which, in de Sitter space, extend all the way from $\mathcal{J}^-$ to $\mathcal{J}^+$. Thus, to summarize, the observer horizon of every observer in de Sitter space is eternal and, as a consequence, is ruled by null lines.

We now consider the following rigidity result for asymptotically de Sitter spactimes, cf. [17].

Theorem 5.2 *Suppose M^4 is an asymptotically simple and de Sitter spacetime satisfying the vacuum Einstein equation (5.9), with $\lambda > 0$. If M contains a null line (i.e., if there is at least one eternal observer horizon) then M is isometric to de Sitter space.*

Theorem 5.2 may be interpreted in terms of the initial value problem for the vacuum Einstein equations, with $\lambda > 0$. According to the fundamental work of Friedrich [11], the set of asymptotically simple and de Sitter solutions to (5.9), with $\lambda > 0$, is open in the set of all maximal globally hyperbolic solutions with compact spatial sections. Thus, by Theorem 5.2, in conjunction with the work of Friedrich, a sufficiently small perturbation of the Cauchy data on a fixed Cauchy hypersurface in de Sitter space will in general destroy all the null lines of de Sitter space, i.e., the resulting spacetime that develops from the perturbed Cauchy data will not contain any null lines (or, equivalently, will not contain any eternal observer horizons). While one would expect many of the null lines to be destroyed, it is somewhat surprising that *none* of the null lines persist. The absence of null lines (or eternal observer horizons) implies, in particular, that the "past null cones" $\partial J^+(p)$ will be compact for all $p \in M$ sufficiently close to $\mathcal{J}^+$. As all such sets in de Sitter space are noncompact, this further serves to illustrate the special nature of the causal structure of de Sitter space (see also [22, Corollary 1]). Finally, we remark, that a similar uniqueness result has also been obtained for Minkowski space, see [16, 17].

Proof: We present some comments on the proof; see [17] for further details. The main step is to show that M has constant curvature. Since M is Einstein, it is sufficient to show that M is conformally flat.

Let η be the assumed null line in M. By Theorem 4.3, η is contained in a smooth totally geodesic null hypersurface S in M. By asymptotic simplicity, η acquires a past end point p on $\mathcal{J}^-$ and a future end point q on $\mathcal{J}^+$. Let us focus attention on situation near p. By the proof of Theorem 4.3, and the fact that p is the past end point of η, we have that,

$$S = \partial I^+(\eta) = \partial I^+(p, \tilde{M}) \cap M\,.$$

It follows that $N_p := S \cup \{p\}$ is a smooth null cone in $\tilde{M}$, generated by the future directed null geodesics emanating from p.

Since S is totally geodesic and the shear σ is a conformal invariant, the null generators of N_p have vanishing shear in the unphysical metric. The trace free part of the Riccati equation (3.3) then implies (see [24, p. 88]) that the components of the conformal tensor suitably contracted in the direction of the null generators vanishes,

$$\tilde{C}_{abcd}K^bK^d = 0 \quad \text{on} \quad S, \tag{5.12}$$

where K is a smooth tangent field to the null generators of S. An argument of Friedrich [12], in which N_p plays the role of an initial characteristic hypersurface, now implies that the conformal tensor of spacetime vanishes on the future domain of dependence of N_p,

$$C^i_{jkl} = 0 \quad \text{on} \quad D^+(N_p, \tilde{M}) \cap M\,. \tag{5.13}$$

Friedrich's argument makes use of the conformal field equations, specifically the divergencelessness of the rescaled conformal tensor,

$$\tilde{\nabla}_i d^i_{jkl} = 0, \qquad d^i_{jkl} = \Omega^{-1} C^i_{jkl}\,.$$

In a time-dual manner one obtains that C^i_{jkl} vanishes on $D^-(N_q, \tilde{M}) \cap M$. Since it can be shown that M is contained in $D^+(N_p, \tilde{M}) \cup D^-(N_q, \tilde{M})$, we conclude that M is conformally flat. Together with equation (5.9), this implies that M has constant (positive) curvature. Moreover, further global arguments show that M is geodesically complete and simply connected. It then follows from uniqueness results for Lorentzian space forms that M is isometric to de Sitter space. □

As illustrated by Schwarzschild-de Sitter spacetime, the assumption of asymptotic simplicity cannot be dropped from Theorem 5.2. However, it appears that this assumption can be substantially weakened; the essential point is to assume the existence of a null line which extends from $\mathcal{I}^-$ to $\mathcal{I}^+$; the null lines in Schwarzschild-de Sitter spacetime do not have end points on $\mathcal{I}$. It is also possible to weaken the vacuum assumption, for example, to allow for the possible presence of matter fields. These extensions of Theorem 5.2 are being considered in [19].

5.3 On the topology of asymptotically de Sitter spacetimes

The results of this subsection were obtained in joint work with Lars Andersson [1].

The result of Friedrich on the nonlinear stability of asymptotic simplicity mentioned in the previous subsection establishes the existence of an open set of solutions to the vacuum Einstein equations with compact Cauchy surfaces, which are asymptotically simple and de Sitter. One is naturally interested in the general features or properties of this class of solutions. Here we address the question of which Cauchy surface topologies are allowable within this class. Obviously, since de Sitter space is in this class, the spherical topology S^n is allowable. Moreover, since isometries of S^n extend in an obvious way to isometries of de Sitter space, any spherical space form S^n/Γ can be achieved. The next theorem shows that, at least in $3+1$ dimensions, these are all the topologies one can expect to get.

Theorem 5.3 *Let M^{n+1}, $n \geq 2$, be an asymptotically de Sitter spacetime (to both the past and future) satisfying the NEC. If M is asymptotically simple either to the past or future, then M is globally hyperbolic, and the Cauchy surfaces of M are compact with finite fundamental group.*

Remarks:

1. Theorem 5.3 implies that the universal cover S^* of S *finitely* covers S. Hence, S^* is compact and simply connected. In three spatial dimensions, this means that S^* is a homotopy 3-sphere (and in fact diffeomorphic to the 3-sphere if the Poincaré conjecture is valid). Thus, in $3+1$ dimensions, the Cauchy surfaces are homotopy 3-spheres, perhaps with identifications.

2. Theorem 5.3 may be reformulated as follows: If M is an asymptotically de Sitter spacetime obeying the NEC, having compact Cauchy surfaces with infinite fundamental group, then M cannot be asymptotically simple, either to the future or the past. This is well illustrated by Schwarzschild de Sitter spacetime, which has Cauchy surface topology $S^{n-1} \times S^1$. Formulated this way, Theorem 5.3 implies that, in the conformal framework of Friedrich [12], if one evolves, via the Einstein equations, suitable initial data on a compact $\mathcal{I}^-$, with infinite fundamental group, something catastophic must develop to the future, as the resulting maximal development cannot be asymptotically de Sitter to the future, i.e., cannot admit a regular $\mathcal{I}^+$. Presumably the resulting physical spacetime M is *globally* singular; it cannot simply develop

a localized black hole, similar to that of Schwarzschild de Sitter spacetime. In any case, the time dual of Theorem 5.4 presented after the proof of Theorem 5.3 shows that some singularity (in the usual sense of causal incompleteness) must occur to the future.

Proof of Theorem 5.3: For the sake of definiteness, assume M is asymptotically simple to the future. That M is globally hyperbolic follows from Proposition 5.1. We show that the Cauchy surfaces of M are compact. One can extend M a little beyond $\mathcal{I}^{\pm}$ to obtain a spacetime without boundary M' which contains $\tilde{M}$, such that any Cauchy surface for M is also a Cauchy surface for M'. Thus, it suffices to show the Cauchy surfaces of M' are compact.

Fix $p \in \mathcal{I}^-$, and consider $\partial I^+(p, M')$. If $\partial I^+(p, M')$ is compact, then by Proposition 2.6, $\partial I^+(p, M')$ is a compact Cauchy surface, and we are done. If $\partial I^+(p, M')$ is noncompact then, by considering a sequence of points going to infinity in $\partial I^+(p, M')$, we can construct a null geodesic generator $\gamma \subset \partial I^+(p, M')$ which is future inextendible in M'. Since M is future asymptotically simple, γ will meet $\mathcal{I}^+$ at q, say. Then, γ_0, the portion of γ between p and q is a null line in M. By Theorem 4.3, γ_0 is contained in a smooth totally geodesic null hypersurface S in M. By arguments like those of the preceeding subsection, the set $N = S \cup \{p, q\}$ forms a compact achronal hypersurface in M'; it represents a future null cone in M' emanating from the point p, and reconverging to a past null cone at q. (We do *not* use the fact here that S is totally geodesic.) By Proposition 2.6, N is a compact Cauchy surface for M'.

Thus, we are lead to the conclusion that the Cauchy surfaces of M', and hence the Cauchy surfaces of M, are compact. Now pass to the universal covering spacetime M^* of M. Since all of the hypotheses of Theorem 5.3 lift to M^*, the Cauchy surfaces of M^* are compact, as well. But since the Cauchy surfaces of M^* cover the Cauchy surfaces of M, and are simply connected, it follows that the universal covering of any Cauchy surface S for M is finite. This implies that the fundamental group of S is finite. □

To conclude this subsection we present a singularity theorem for future asymptotically simple and de Sitter spacetimes.

Theorem 5.4 *Let M^{n+1}, $2 \le n \le 7$, be a future asymptotically simple and de Sitter spacetime with compact orientable Cauchy surfaces, which obeys the NEC. If the Cauchy surfaces of M have positive first Betti number, $b_1 > 0$, then M is past null geodesically incomplete.*

Note that if a Cauchy S contains a wormhole, i.e., has topology of the form $N\#(S^1 \times S^{n-1})$, then $b_1(S) > 0$. The theorem is somewhat remiscent of previous results of Gannon [21], which show, in the asymptotically flat setting, how nontrivial spatial topology leads to the occurence of singularities.

The proof of Theorem 5.4 is an application of the Penrose singularity theorem, stated below in a form convenient for our purposes.

Theorem 5.5 *Let M be a globally hyperbolic spacetime with noncompact Cauchy surfaces, satisfying the NEC. If M contains a past trapped surface, M is past null geodesically incomplete.*

Recall [24, 3, 29] that a past trapped surface is a compact co-dimension two spacelike submanifold W of M with the property that the two congruences of null normal geodesics issuing to the past from W have negative divergence along W.

Comments on the proof of Theorem 5.4: Since M is asymptotically de Sitter and $\mathcal{J}^+$ is compact (see Proposition 5.1), one can find in the far future a smooth compact spacelike Cauchy surface Σ for M, with second fundamental form which is *positive definite* with respect to the *future* pointing normal. This means that Σ is *contracting* in all directions towards the *past*.

By Poincaré duality, and the fact that there is never any co-dimension one torsion, $b_1(\Sigma) > 0$ if and only if $H_{n-1}(\Sigma, \mathbb{Z}) \neq 0$. By well known results of geometric measure theory (see [27, p. 51] for discussion; this is where the dimension assumption is used), every nontrivial class in $H_{n-1}(\Sigma, \mathbb{Z})$ has a least area representative which can be expressed as a sum of smooth, orientable, connected, compact, embedded minimal (mean curvature zero) hypersurfaces in Σ. Let W be such a hypersurface; note W is spacelike and has co-dimension two in M. As described in [14], since W is minimal in Σ, and Σ is contracting in all directions towards the past in M, W must be a past trapped surface in M.

Since W and Σ are orientable, W is two-sided in Σ. Moreover, since W represents a nontrivial element of $H_{n-1}(\Sigma, \mathbb{Z})$, W does not separate Σ, for otherwise it would bound in Σ. This implies that there is a loop in Σ with nontrivial intersection number with respect to W. There exists a covering space Σ^* of Σ in which this loop gets unraveled. Σ^* has a simple description in terms of cut-and-paste operations: By making a cut along Σ, we obtain a compact manifold Σ' with two boundary components, each isometric to W. Taking $\mathbb{Z}$ copies of Σ', and gluing these copies end-to-end we obtain the covering space Σ^* of Σ. In this covering, W is covered by $\mathbb{Z}$ copies of itself, each one separating Σ^*; let W_0 be one such copy. We know by global hyperbolicity that M is homeomorphic to $\mathbb{R} \times \Sigma$, and hence the fundamental groups of Σ and M are isomorphic. This implies that the covering spaces of M are in one-to-one correspondence with the covering spaces of Σ. In fact, there will exist a covering spacetime M^* of M in which Σ^* is a Cauchy surface for M^*. Thus, M^* is a spacetime obeying the NEC, which contains a noncompact Cauchy surface (namely Σ^*) and a passed trapped surface (namely W_0). By the Penrose singularity theorem, M^* is past null geodesically incomplete, and hence so is M. □

Acknowledgement

We wish to thank Piotr Chruściel and Helmut Friedrich for the invitation to speak at the aforementioned Cargése Summer School on Mathematical General Relativity. The School's recognition of the 50th anniversary of Professor Choquet-Bruhat's seminal work on the Einstein equations, made it, indeed, a special pleasure to participate.

References

[1] L. Andersson and G. J. Galloway, *DS/CFT and spacetime topology*, Adv. Theor. Math. Phys. **6** (2002) 307–327.

[2] L. Andersson, G. J. Galloway, and R. Howard, *A strong maximum principle for weak solutions of quasi-linear elliptic equations with applications to Lorentzian and Riemannian geometry*, Comm. Pure Appl. Math. **51** (1998) 581–624.

[3] J. K. Beem, P. E. Ehrlich, and K. L.Easley, *Global Lorentzian geometry*, 2 ed., Pure and Applied Mathematics, vol. 202, Marcel Dekker, New York, 1996.

[4] Raphael Bousso, *Adventures in de Sitter space*, preprint, hep-th/0205177.

[5] E. Calabi, *An extension of E. Hopf's maximum principle with an application to Riemannain geometry*, Duke Math. Jour. **25** (1958), 45–56.

[6] J. Cheeger and D. Gromoll, *The splitting theorem for manifolds of nonnegative Ricci curvature*, J. Diff. Geom. **6** (1971) 119–128

[7] P. T. Chruściel, E. Delay, G. J. Galloway, and R. Howard. *Regularity of horizons and the area theorem.* Annales H. Poincaré, **2** (2001) 109–178.

[8] Ray D'Inverno, *Introducing Einstein's relativity*, Oxford University Press, Oxford, 1992.

[9] J.-H. Eschenburg, *Comparison theorems and hypersurfaces*, Manuscripta Math. **59** (1987), no. 3, 295–323.

[10] ———, *The splitting theorem for space-times with strong energy condition*, J. Diff. Geom. **27** (1988), 477–491.

[11] H. Friedrich, *On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure.*, Comm. Math. Phys. **107** (1986) 587–609.

[12] ———, *Existence and structure of past asymptotically simple solutions of Einstein's field equations with positive cosmological constant*, J. Geom. Phys. **3** (1986) 101–117.

[13] ———, *On the global existence and the asymptotic behaviour of solutions to the Einstein-Maxwell-Yang-Mills equations*, J. Diff. Geom., **34** (1991) 275–345.

[14] G. J. Galloway, *Minimal surfaces, spatial topology and singularities in space-time*, J. Phys. A **16** (1983), no. 7, 1435–1439.

[15] ———, *The Lorentzian splitting theorem without the completeness assumption*, J. Diff. Geom. **29** (1989), 272–389.

[16] ———*Maximum Principles for null hypersurfaces and null splitting theorems*, Ann. Henri Poincaré **1** (2000) 543–567.

[17] ———*Some global results for asymptotically simple spacetimes*, in: The conformal structure of space-times: Geometry, analysis, numerics, ed. by J. Frauendiener and H. Friedrich, Lecture Notes in Physics, vol. 604, pp 51–60, 2002, Springer Verlag.

[18] G. Galloway, K. Schleich, D. Witt, and E. Woolgar, *Topological Censorship and Higher Genus Black Holes*, Phys. Rev. D **60** (1999) 104039.

[19] G. J. Galloway and D. Solis, in preparation.

[20] G. J. Galloway, S. Surya and E. Woolgar, *A uniqueness theorem for the adS soliton*, Phys. Rev. Lett. **88** (2002) 101102.

[21] D. Gannon, *Singularities in nonsimply connected space-times*, J. Math. Phys., **16** (1975) 2364–2367.

[22] S. Gao and R. M. Wald, *Theorems on gravitational time delay and related issues*, Class. Quant. Grav. **17** (2000) 4999-5008.

[23] G. Gibbons and S.W. Hawking, *Cosmological event horizons, thermodynamics, and particle creation.*, Phys. Rev. D **15** (1977) 2738.

[24] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press, Cambridge, 1973.

[25] S.W. Hawking and R. Penrose, *The singularities of gravitational collapse and cosmology*, Proc. Roy. Soc. Lond. **A314** (1970) 529–548.

[26] D.N. Kupeli, *On null submanifolds in spacetimes*, Geom. Dedicata **23** (1987), 33–51.

[27] H. Blaine Lawson, Jr., *Minimal varieties in real and complex geometry*, Les Presses de l'Université de Montréal, Montreal, Que., 1974, Séminaire de Mathématiques Supérieures, No. 57 (Été 1973).

[28] R.P.A.C. Newman, *A proof of the splitting conjecture of S.-T. Yau*, J. Diff. Geom. **31** (1990) 163–184.

[29] B. O'Neill, *Semi–Riemannian geometry*, Academic Press, New York, 1983.

[30] R. Penrose, *Zero rest-mass fields including gravitation: asymptotic behavior*, Proc. Roy. Soc. Lond. A **284** 159–203.

[31] ———, *Techniques of differential topology in relativity*, SIAM, Philadelphia, 1972, (Regional Conf. Series in Appl. Math., vol. 7).

[32] R. Penrose, R. D. Sorkin and E. Woolgar, *A positive mass theorem based on the focusing and retardation of null geodesics*, preprint, gr-qc/9301015.

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

[34] S.-T. Yau *Problem section*, in Ann. Math. Studies, vol. 102, ed. S.-T. Yau, Princeton, 1982, pp 669–706.