Radial gravitational collapse causes timelike incompleteness

Leonardo García-Heveling[1,2,*]

[1]*Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany*
[2]*Department of Mathematics, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands*

We show that a globally hyperbolic spacetime containing a trapped surface and satisfying the strong energy condition and a condition on certain radial tidal forces must be timelike geodesically incomplete. This constitutes a "timelike" version of Penrose's celebrated singularity theorem. Recall that the latter concludes that certain spacetimes are null incomplete, providing the first theoretical evidence that black holes actually exist in our Universe. By concluding timelike instead of null incompleteness, we obtain, at the expense of stronger assumptions, a clearer physical interpretation and the existence of an event horizon.

I. INTRODUCTION

In his seminal 1964 article [1], Roger Penrose showed that gravitational collapse in general relativity leads to the formation of spacetime singularities. At the time, some singular solutions to Einstein's equations were already known, such as the Schwarzschild solution. It was believed, however, that such singularities were mathematical pathologies arising due to the high degree of symmetry and the lack of matter, and that physically realistic solutions would be singularity free. Penrose forever changed this point of view through his famous singularity theorem.

Penrose Singularity Theorem. Let (M, g) be a spacetime containing a noncompact Cauchy surface. Assume that the null convergence condition is satisfied (meaning $\mathrm{Ric}(X, X) \geq 0$ for every null vector X) and that (M, g) contains a trapped surface $\mathcal{T}$. Then (M, g) is null geodesically incomplete.

The key concept of *trapped surface* was introduced for the first time through this theorem. The original definition is that $\mathcal{T}$ is a codimension 2 spacelike submanifold with negative ingoing and outgoing null expansion. Physically, it corresponds to the boundary of a spatial region from which even light cannot escape (hence the name). The incomplete null geodesic predicted by the theorem corresponds to the trajectory of a light ray which ends abruptly, the interpretation being that it must have met a spacetime singularity. Since the remaining assumptions are rather mild, it provided the first compelling evidence that the formation of singularities is in fact a crucial feature of general relativity [2].

The nature of the singularity predicted by Penrose's theorem, however, remains mysterious to this day. While it is widely believed to be at the center of a black hole, the theorem itself gives no information about features usually associated to such objects: Is it shielded by an event horizon? Does the curvature blow up? Penrose himself was aware of the problem and conjectured some of these properties, but these so-called cosmic censorship conjectures remain open to this day [2].

Another problem is that Penrose's theorem only predicts null incompleteness. It thus remained open whether massive objects or even human observers, which follow timelike curves, can fall into the singularity. Moreover, the affine parameter of a null geodesic (i.e. the quantity that characterizes incomplete null geodesics) does not correspond to a physical observable. The affine parameter of a *timelike* geodesic, on the other hand, is the proper time measured by an observer.

The issue of timelike vs null incompleteness had already been identified early on by Geroch [3] and Beem [4], who gave some negative results, showing that null incompleteness does not always imply timelike incompleteness. The counterexamples given, however, are physically rather artificial. In this paper, we reopen the case by proving a positive result.

Theorem 1. Let (M, g) be an $n + 1$-dimensional spacetime. Assume that

  (i) (M, g) is globally hyperbolic.
  (ii) $\mathrm{Ric}(X, X) \geq 0$ for every timelike vector X.
  (iii) (M, g) contains a trapped surface $\mathcal{T}$ with (necessarily past-pointing timelike) mean curvature vector denoted by H and $H_0 := \min \sqrt{|g(H, H)|}$.
  (iv) There exists a constant $0 < k < H_0$ such that every unit speed timelike geodesic σ normal to $\mathcal{T}$ of length greater than $\frac{1}{k}$ satisfies

$$\int_0^{\frac{1}{k}} (1 - k^2 u^2)\mathrm{Riem}(\dot{\sigma}, U, \dot{\sigma}, U)\,du \leq 0, \quad (1)$$

*leonardo.garcia@uni-hamburg.de

where U is the parallel transport along σ of the unique (up to sign) spacelike unit vector orthogonal to $\dot{\sigma}(0)$ and $\mathcal{T}$.

Then every causal curve starting on $\mathcal{T}$ has length at most $\ell := \frac{1}{k} + \frac{n}{H_0 - k}$. In particular, (M, g) is timelike geodesically incomplete.

Note that our theorem predicts that *every* observer starting on $\mathcal{T}$ will fall into the singularity after a proper time of at most ℓ. From this, we can deduce the existence of an event horizon, because it means that an observer reaching a trapped surface can no longer escape to infinity. We return to this important point in the conclusions section. For now, let us comment on each of the assumptions:

(i) Global hyperbolicity is equivalent to the existence of a Cauchy surface [5,6], which can be compact or noncompact (unlike in Penrose's theorem).

(ii) This is the *timelike convergence condition*, which is stronger than the null convergence condition. Through the Einstein equations it can be translated into a bound on the energy-momentum tensor known as the *strong energy condition*, which is satisfied by many forms of matter [[7], pp. 95–96].

(iii) A trapped surface $\mathcal{T}$ is a compact, orientable, space-like codimension 2 submanifold with everywhere past-pointing timelike mean curvature vector H. This definition is equivalent to the usual one in terms of negative inward and outward null expansions [[8], p. 435]. Note that by compactness $H_0 > 0$.

(iv) The sectional curvatures $\text{Riem}(\dot{\sigma}, U, \dot{\sigma}, U)$ have the physical interpretation of tidal forces. In our case, U should be interpreted as the radial direction pointing inward (or outward) of the trapped surface. We are thus requiring that the radial tidal forces are repulsive, at least for some proper time $\frac{1}{k}$ in a weighted average sense (where the weight decreases toward the future).

Assumption (iv) is justified by the physical intuition that the gravitational attraction grows as one approaches the source, causing a test object freely falling toward a star to be stretched [[9], § 32.6]. In Schwarzschild spacetime, (iv) is satisfied, and in fact the radial tidal forces even blow up to $-\infty$. In Kerr spacetime, however, (iv) can be violated for geodesics staying close to the axis of rotation of the black hole [10]. Thus further investigation to allow the case of rotating black holes is needed (see also the discussion section at the end). From a conceptual point of view, assumption (iv) also has the disadvantage that it is an assumption on the geometry of our particular spacetime. In comparison, the energy condition (ii) depends only on the matter model considered, but not on the specific solution. By framing (iv) as an integral assumption (rather than point-wise), we do, luckily, avoid the need for extreme fine tuning.

We prove the theorem in the next section, and conclude the paper with a discussion about event horizons and possibilities to weaken our assumptions.

II. PROOF OF MAIN THEOREM

We use techniques from the book of O'Neill [8], which are used there to prove the Hawking singularity theorem. The latter is similar to the Penrose singularity theorem, but its assumptions and conclusions are usually interpreted from a cosmological point of view, instead of in the context of gravitational collapse.

We start with the global causality theoretic argument of our proof. To that end, recall that the Lorentzian length $L_g(\gamma)$ of a causal curve $\gamma : [0, b] \to M$ is given by

$$L_g(\gamma) := \int_0^b g(\dot{\gamma}, \dot{\gamma}) du, \qquad (2)$$

and equals b when γ is parametrized with unit speed.

Lemma 2. For every point q in the timelike future of $\mathcal{T}$, there exists a maximizing timelike geodesic σ from $\mathcal{T}$ to q. Moreover, σ is normal to $\mathcal{T}$.

Proof. Because (M, g) is globally hyperbolic, the Lorentzian distance from q (to the past),

$$\tau(x, q) := \sup \{L_g(\sigma) | \sigma \text{ causal from } x \text{ to } q\}, \qquad (3)$$

is continuous [[8], Lem. 14.21]. Thus $\tau(\cdot, q)$ attains a maximum on the compact set $\mathcal{T}$. For every point $p \in \mathcal{T}$ where said maximum is attained, again by global hyperbolicity there exists a geodesic σ from p to q with $L_g(\sigma) = \tau(p, q)$ [[8], Lem. 14.19]. That σ must be normal to $\mathcal{T}$ is standard [[8], Cor. 10.26]. ∎

Our goal is to show that the length of all maximizing geodesics provided by Lemma 2 is smaller than ℓ, from which the statement of Theorem 1 follows immediately, since all other curves must be shorter than the maximizing one (here $\ell := \frac{1}{k} + \frac{n}{H_0 - k}$ as in the theorem). It suffices to show that if there was a normal geodesic $\sigma : [0, \ell] \to M$ of length ℓ, it would have a focal point $\sigma(r)$ for $r < \ell$ (defined below), and hence would not be maximizing.

Let $\sigma : [0, \ell] \to M$ be a unit-speed geodesic normal to $\mathcal{T}$. We need to consider infinitesimal variations of σ that remain normal to $\mathcal{T}$. To each such variation corresponds a *variation vector field* $V : [0, \ell] \to TM$ along σ such that $V(0)$ is tangent to $\mathcal{T}$ and $V(\ell) = 0$, and vice versa any such variation vector field gives rise to an infinitesimal variation of σ. The intuition is that we are pushing σ along V, while keeping the starting direction normal to $\mathcal{T}$ and the endpoint fixed. Because σ is a normal geodesic, the first derivative of the length for these variations vanishes. The second variation L_V'' of the length is then given by Synge's formula

$$L_V'' = -\int_0^\ell (g(\dot{V}, \dot{V}) - \text{Riem}(\dot{\sigma}, V, \dot{\sigma}, V)) du$$
$$+ g(\dot{\sigma}(0), II(V(0), V(0))), \qquad (4)$$

where dots mean differentiation along σ and II is the second fundamental form of $\mathcal{T}$ (see [[8], Chap. 10]).

A focal point of $\mathcal{T}$ along σ is a point $\sigma(r)$ such that there exists a variation vector field V with $V(r) = 0$ and $L_V'' \geq 0$. When there is a focal point $\sigma(r)$, then σ no longer maximizes the distance to $\mathcal{T}$ after $\sigma(r)$.

Lemma 3. Every unit-speed timelike geodesic $\sigma : [0, \ell] \to M$ normal to $\mathcal{T}$ encounters a focal point $\sigma(r)$ with $0 < r \leq \ell$, provided σ is defined on $[0, \ell]$.

Proof. Fix any point $p \in \mathcal{T}$. Let $e_1 = U$ and let $e_2, \ldots, e_n$ be an orthonormal basis for $T_p\mathcal{T}$. Parallel transport $e_1, \ldots, e_n$ along σ, obtaining vector fields $E_1, \ldots, E_n$. In particular, $E_1 = U$ [cf. assumption (iii)]. For $u \in [0, \ell]$ the affine parameter of σ, and k as in assumption (iv), define

$$\varphi(u) := \begin{cases} ku & \text{for } u < 1/k, \\ 1 - \left(u - \frac{1}{k}\right)\frac{H_0 - k}{n} & \text{for } u \geq 1/k, \end{cases} \quad (5)$$

$$\psi(u) := \begin{cases} 1 & \text{for } u < 1/k, \\ 1 - \left(u - \frac{1}{k}\right)\frac{H_0 - k}{n} & \text{for } u \geq 1/k. \end{cases} \quad (6)$$

Note that $\psi(0) = 1$ while $\varphi(0) = \varphi(\ell) = \psi(\ell) = 0$. We use these as coefficients to define n variation vector fields in the following way:

$$V_1 := \varphi E_1, \qquad V_j := \psi E_j \quad \text{for } j = 2, \ldots, n. \quad (7)$$

Note here that φ, ψ are not smooth, but we can smoothly approximate them (more to that below). By (4) we have

$$L_{V_1}'' = -\int_0^\ell \left((\varphi')^2 - \varphi^2 \text{Riem}(E_1, \dot{\sigma}, E_1, \dot{\sigma})\right)du \quad (8)$$

$$= -k - \frac{H_0 - k}{n} + \int_0^\ell \psi^2 \text{Riem}(E_1, \dot{\sigma}, E_1, \dot{\sigma})du$$

$$- \int_0^\ell (\psi^2 - \varphi^2)\text{Riem}(E_1, \dot{\sigma}, E_1, \dot{\sigma})du \quad (9)$$

$$= -k - \frac{H_0 - k}{n} + \int_0^\ell \psi^2 \text{Riem}(E_1, \dot{\sigma}, E_1, \dot{\sigma})du$$

$$- \int_0^{\frac{1}{k}} (1 - k^2 u^2)\text{Riem}(E_1, \dot{\sigma}, E_1, \dot{\sigma})du \quad (10)$$

and for $j = 2, \ldots, n$,

$$L_{V_j}'' = -\int_0^\ell \left((\psi')^2 - \psi^2 \text{Riem}(E_j, \dot{\sigma}, E_j, \dot{\sigma})\right)du$$

$$+ g(\dot{\sigma}(0), II(e_j, e_j)) \quad (11)$$

$$= -\frac{H_0 - k}{n} + \int_0^\ell \psi^2 \text{Riem}(E_j, \dot{\sigma}, E_j, \dot{\sigma})du$$

$$+ g(\dot{\sigma}(0), II(e_j, e_j)) \quad (12)$$

Summing these two expressions, and using that

$$\text{Ric}(\dot{\sigma}, \dot{\sigma}) = \sum_{i=1}^n \text{Riem}(E_i, \dot{\sigma}, E_i, \dot{\sigma}), \quad (13)$$

$$H = \sum_{j=2}^n II(e_j, e_j), \quad (14)$$

and assumption (ii)–(iv), we obtain

$$\sum_{i=1}^n L_{V_i}'' = -(H_0 - k) + \int_0^\ell \psi^2 \text{Ric}(\dot{\sigma}, \dot{\sigma})du$$

$$+ g(\dot{\sigma}(0), H) - k$$

$$- \int_0^{\frac{1}{k}} (1 - k^2 u^2)\text{Riem}(E_1, \dot{\sigma}, E_1, \dot{\sigma})du \quad (15)$$

$$> -H_0 + g(\dot{\sigma}(0), H) \geq 0, \quad (16)$$

where the last inequality follows from the reverse Cauchy-Schwartz inequality [[8], Prop. 5.30] together with the fact that $H, \sigma(0)$ are past- and future-directed timelike respectively. We conclude that at least one of the L_{V_i}'' must be positive, and hence σ has a focal point, as desired. Since the inequality is strict, the argument also works for smooth approximations of φ, ψ close enough in the C^1-topology on the space of continuous functions $[0, \ell]$ (this is just a technicality needed because the theory developed in [8] assumes smoothness). ∎

III. DISCUSSION

At the expense of strengthening the assumptions, we have been able to improve on the conclusions of the Penrose singularity theorem in two important aspects:

(i) We conclude timelike incompleteness, which (unlike null incompleteness) has a clear physical interpretation in terms of the proper time experienced by observers.

(ii) We can further conclude the existence of an event horizon, since once an observer reaches a trapped surface, they cannot escape anymore and are doomed to fall into the singularity after a proper time of at most ℓ. This is true even for accelerated observers.

Here by horizon we simply mean the boundary that separates the regions $\mathcal{B}$ and $\mathcal{E}$ defined as follows. We call *black hole interior* $\mathcal{B}$ the region such that every future-directed timelike curve starting in $\mathcal{B}$ is incomplete, while by *black hole exterior* $\mathcal{E} := M \backslash \mathcal{B}$ we mean its complement, implying that from every point in $\mathcal{E}$, one can find at least one future-directed timelike curve of infinite length (i.e. an observer that lives forever). Under the assumptions of our theorem, $\mathcal{B} \neq \emptyset$, and $\mathcal{E}$ should also be assumed nonempty, unless we are in a big crunch scenario where the whole Universe is swallowed by the singularity. Incidentally, Senovilla [11] has argued that in certain situations, gravitational collapse *can* lead to a big crunch, but we shall

ignore this in the present discussion. The general treatment of event horizons as described here was pioneered by Müller [12], and has the advantages over the traditional approach using null infinity that we do not need to assume asymptotic flatness, or to worry about making a choice of spacetime compactification.

A word of caution about the horizon predicted by our theorem: It is in principle possible that there exist some timelike geodesics (or, more generally, curves with bounded acceleration) which are incomplete despite not emanating from a trapped surface. Our theorem gives us no control over those. In particular, they might never cross the horizon as we have defined it (meaning they could stay in $\mathcal{E}$ and never enter $\mathcal{B}$). Thus while the singularities *which our theorem predicts* are hidden behind a horizon, there could be other singularities in our spacetime (or the predicted singularity could be "larger"). Recall that Penrose's weak cosmic censorship conjecture states states that *every* singularity in a physically reasonable spacetime should be hidden behind an event horizon.

To conclude the paper, let us mention that many variations of the singularity theorems of Hawking and Penrose are known. These have relaxed assumptions, particularly the causality assumptions [13,14], the energy conditions [15–17], the regularity of the spacetime metric [18,19], and the dimensionality of the trapped submanifold [20] (this list of references is in no way exhaustive). The goal is often to accommodate for violations in the classical assumptions due to quantum phenomena. Not only should many of these modifications be possible for our theorem, but in fact it paves the way for further progress. On the one hand, our theorem is in the spirit of Penrose in that it is about gravitational collapse. On the other hand, our proof is more similar to that of the Hawking singularity theorem, which has generally proven to be easier to modify. Thus the chance arises to adapt improved variants of Hawking's theorem to the context of gravitational collapse.

We have, however, also created the need to relax our new assumption (iv) (which has no analogue in the original singularity theorems of Hawking and Penrose). From our proof it is already possible to derive some sufficient conditions that are technically weaker than (iv), but they make the statement of the theorem more convoluted without improving on its physical interpretation (for example, we can allow k to depend on the geodesic, as long as it remains uniformly bounded away from H_0). More interesting (at least from a mathematical point of view) is to substitute (ii) and (iv) together by an intermediate curvature bound. By this we mean a bound on the trace of the Riemann tensor over a $n-1$ dimensional subspace [instead of the full n dimensions that yield the Ricci tensor, cf. (13)]. Then we can replicate our proof but completely ignoring the "radial" direction E_1. But while the Ricci bound $\mathrm{Ric}(X,X) \geq 0$ is equivalent to the strong energy condition via the Einstein Equations, this is not true for intermediate curvature bounds, leaving their physical meaning in the air. Note that such an approach was taken by Galloway and Senovilla [20], but there the directions left out are to be interpreted as compactified string theory dimensions, which should not contribute to macroscopic physics anyhow, while in our case leaving out the radial direction seems artificial. Satisfactorily weakening assumption (iv) thus remains as an open point for further investigation.

ACKNOWLEDGMENTS

[1] R. Penrose, Gravitational collapse and spacetime singularities, Phys. Rev. Lett. **14,** 57 (1965).

[2] K. Landsman, Penrose's 1965 singularity theorem: From geodesic incompleteness to cosmic censorship, Gen. Relativ. Gravit. **54,** 115 (2022).

[3] R. Geroch, What is a singularity in general relativity?, Ann. Phys. (N.Y.) **48,** 526 (1968).

[4] J. K. Beem, Some examples of incomplete space-times, Gen. Relativ. Gravit. **7,** 501 (1976).

[5] A. N. Bernal and M. Sánchez, On smooth Cauchy hypersurfaces and Geroch's splitting theorem, Commun. Math. Phys. **243,** 461 (2003).

[6] R. Geroch, Domain of dependence, J. Math. Phys **11,** 437 (1970).

[7] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2023).

[8] B. O'Neill, *Semi-Riemannian Geometry. With Applications to Relativity* (Academic Press, San Diego, 1983).

[9] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).

[10] B. O'Neill, *The Geometry of Kerr Black Holes* (A. K. Peters, Wellesley, MA, 1995).

[11] J. M. M. Senovilla, Ultra-massive spacetimes, Portugaliae mathematica **80,** 133 (2023).

[12] O. Müller, Horizons, Adv. Theor. Math. Phys. **19,** 747 (2015).

[13] M. Lesourd, A new singularity theorem for black holes which allows chronology violation in the interior, Classical Quantum Gravity **35,** 245003 (2018).

[14] E. Minguzzi, A gravitational collapse singularity theorem consistent with black hole evaporation, Lett. Math. Phys. **110,** 2383 (2020).

[15] C. J. Fewster and E.-A. Kontou, A semiclassical singularity theorem, Classical Quantum Gravity **39,** 075028 (2022).

[16] M. Graf, E.-A. Kontou, A. Ohanyan, and B. Schinnerl, Hawking-type singularity theorems for worldvolume energy inequalities, arXiv:2209.04347.

[17] S.-H. Paeng, Hawking's singularity theorem under a bounded integral norm of Ricci curvature, J. Geom. Phys. **183,** 104708 (2023).

[18] M. Graf, Singularity theorems for C^1- Lorentzian metrics, Commun. Math. Phys. **378,** 1417 (2020).

[19] M. Kunzinger, A. Ohanyan, B. Schinnerl, and R. Steinbauer, The Hawking–Penrose singularity theorem for C^1-Lorentzian metrics, Commun. Math. Phys. **391,** 1143 (2022).

[20] G. J. Galloway and J. M. M. Senovilla, Singularity theorems based on trapped submanifolds of arbitrary co-dimension, Classical Quantum Gravity **27,** 152002 (2010).