Penrose’s 1965 singularity theorem: from geodesic incompleteness to cosmic censorship

Klaas Landsman¹

Received: 4 May 2022 / Accepted: 5 August 2022 / Published online: 29 September 2022
© The Author(s) 2022

Abstract
Supplementing earlier literature by e.g. Tipler et al. (1980), Israel (1987), Thorne (1994), Earman (1999), Senovilla and Garfinkle (2015), Curiel (2019) and Landsman (2021), I provide a historical and conceptual analysis of Penrose’s path-breaking 1965 singularity (or incompleteness) theorem. The emphasis is on the nature and historical origin of the assumptions and definitions used in—or otherwise relevant to—the theorem, as well as on the discrepancy between the (astro)physical goals of the theorem and its actual content: even if its assumptions are met, the theorem fails to prove the existence or formation of black holes. Penrose himself was well aware of this gap, which he subsequently tried to overcome with his visionary and influential cosmic censorship conjectures. Roughly speaking, to infer from (null) geodesic incompleteness that there is a “black” object one needs weak cosmic censorship, whereas in addition a “hole” exists (as opposed to a boundary of an extendible space-time causing the incompleteness of geodesics) if strong cosmic censorship holds.

Keywords  Roger Penrose · History of general relativity · Singularity theorems

This article belongs to the Topical Collection: Singularity theorems, causality, and all that (SCRI21).

This expository paper is based on my keynote talk Penrose in perspective at the conference Singularity theorems, causality, and all that: A tribute to Roger Penrose (SCRI21), June 14–18, 2021, organized by Piotr Chruściel, Greg Galloway, Michael Kunzinger, Ettore Minguzzi, and Roland Steinbauer. I am singularly grateful to these organizers for the invitation to speak. I am also indebted to Michel Janssen for showing me a copy of Einstein’s letter to Besso (March 1, 1931) quoted in footnote 10, as well as to José Senovilla for several important corrections on the first version of this paper. Finally, with gratitude I spoke to Penrose for two hours on July 2, 2022 in Oxford, followed by a walk, which led to further clarifications.

Klaas Landsman
landsman@math.ru.nl

¹ Department of Mathematics, Radboud University, Nijmegen, The Netherlands
1 Introduction

Penrose’s short paper ‘Gravitational collapse and space-time singularities’ from 1965 was surely one of the most important contributions to general relativity since Einstein had formulated the theory in 1915:

However, all bets were off concerning the possibility of a non-singular outcome to a collapse after the publication of a theorem by Roger Penrose (1965), which has claims to be considered the most influential development in general relativity in the 50 years after Einstein founded the theory. Penrose demonstrated that singularities were, after all, a generic feature of gravitational collapse and must appear soon after the formation of a trapped surface (a surface from which light cannot escape outwards). Although the theorem could offer no information about the nature of the singularity, there was a widespread belief that it would have an all-enveloping spacelike character, as in the Schwarzschild case, thus obstructing further development of the situation. Penrose’s paper was just as important for what it initiated as for what it accomplished. Powerful global techniques exploiting the causal structure of space-time had been introduced into the theory for the first time. (Israel [99, p. 253])

Penrose’s theorem said roughly this: Suppose that a star—any kind of star whatsoever—implodes so far that its gravity becomes strong enough to form an apparent horizon, that is, strong enough to pull outgoing light rays back inward. After this happens, nothing can prevent the gravity from growing so strong that it creates a singularity. Consequently (since black holes always have apparent horizons), every black hole must have a singularity inside itself. (Thorne [186, p. 227])

The singularity theorems constitute the first genuine post-Einsteinian content of classical GR, not foreseen in any way by Einstein (…) The global mathematical developments needed for the singularity theorems, and the ideas on incompleteness or trapping—and thus also their derived inferences—were not treated nor mentioned, either directly or indirectly, in any of Einstein’s writings. In 1965 GR left adolescence behind, emancipated from its creator, and became a mature physical theory full of vitality and surprises. (Senovilla and Garfinkle, [175, p. 2])
The comparably path-breaking series of papers—this time on the observational side—by the Event Horizon Telescope Collaboration (2019) on their famous image of the black hole in M87 opens as follows:

Black holes are a fundamental prediction of the theory of general relativity (GR; Einstein 1915). A defining feature of black holes is their event horizon, a one-way causal boundary in space-time from which not even light can escape (Schwarzschild 1916). The production of black holes is generic in GR (Penrose (965)).

(Event Horizon Telescope Collaboration [61], p. L1)

On the basis of this paper, to general acclaim Penrose won half of the 2020 Physics Nobel Prize:

for the discovery that black hole formation is a robust prediction of the general theory of relativity.

All records show that this is certainly how the (astro)physics community interpreted Penrose’s result.1 In particular, it put an end to claims even by leading relativists to the effect that curvature singularities of the kind found in the static Schwarzschild solution from 1916 (and later in the Kerr solution from 1963), as well as in the dynamical Oppenheimer–Snyder solution from 1939, were artificial consequences of unphysical idealizations (such as spherical or axial symmetry) and would be absent in realistic models.

But let us recall that Penrose [148] actually proved that in the presence of a suitable curvature condition, a space-time \((M, g)\) that contains a non-compact Cauchy surface as well as a trapped surface cannot be future null geodesically complete. Thus it is far from obvious how Penrose’s theorem relates to the above summaries by Thorne, the EHT, and the Nobel Prize committee (Israel is more accurate; the discontinuity between Einstein and Penrose claimed by Senovilla and Garfinkle will be addressed).

Putting it bluntly, the main differences between Penrose’s actual theorem and its reputation are that:

- The theorem says nothing about event horizons, which form the “black” ingredient of a black hole;
- It is inconclusive about “singularities”, which should form the “hole” part of a black hole.

Hence the link between the theorem and black holes is opaque. What is missing is the further inference, perhaps under additional assumptions, from the existence of incomplete null geodesics to the existence of an event horizon, as well as an argument relating incompleteness to the existence of a singularity in some physically meaningful sense (such as a curvature singularity), as opposed to, for example, the sudden end of a geodesic being caused by extendibility of the space-time under consideration. Penrose [148] was well aware of both problems, see Sect. 3 below. Addressing the first had to wait for a good definition of an event horizon, which he provided in [151].

---

1 Israel [99], Tipler, Clarke, & Ellis [187], Earman [47, 49], Senovilla [173], Senovilla & Garfinkle [175] and Curiel [32] survey singularities and singularity theorems in GR in a historical context. For textbook treatments see e.g. [3, 26, 93, 103, 107, 139, 173, 189], and [131]. For Penrose’s theorem see also [2] and [71].
and which led, in the same paper, to what is now called his *weak cosmic censorship* conjecture. Penrose subsequently addressed the second (extendibility) problem, albeit rather indirectly at first, by what we now call *strong cosmic censorship*.

Here is the plan of the remainder of this paper:

Section 2 summarizes the status of singularities and singularity theorems in GR before 1965.²

Section 3 goes through [148] in detail, paying special attention to what was and was not new.

Section 4 resumes the discussion of singularities, including Penrose’s [154, 155] “localization” thereof.

Section 5 discusses cosmic censorship mainly from the point of view of Penrose [151, 153–155].

Section 6 explains the transition from his cosmic censorship conjectures to the versions that are currently used.

Section 7 is an Epilogue, reflecting a little on scientific revolutions and historiography in the present context.

Let me stress, perhaps against first appearances, that this paper is intended to *increase* our admiration of Penrose’s work in GR, which—apart from equally deep and influential ideas on gravitational radiation and asymptotic flatness, spinor and twistor theory, black hole thermodynamics, and cosmology—may be seen as a sustained effort to accomplish what many people seem to believe his 1965 theorem already did.

## 2 Singularities and singularity theorems before 1965

The period from 1916 to 1965 is well covered by [47, 49, 51, 175, 181] but problems of interpretation remain. Like the history of GR itself [100, 101], I see a steady but bumpy road—this time from Einstein to Penrose—where ultimately the mathematician overruled the physicist and won.

Both Hilbert [96] and Einstein [55] saw singularities as points in *space-time* where some components $g_{\mu\nu}$ of the metric were ill-defined and/or $\det(g) = 0$, with some attention paid to the role of the coordinate system. Such singularities were seen in the Schwarzschild solution from 1916 at $r = 2m$, as well as in the de Sitter solution from 1917.³ Slightly rewritten by Einstein [55], de Sitter’s metric is

$$ds^2 = dr^2 + R^2 \sin^2 \left( \frac{r}{R} \right) d\Omega^2 - \cos^2 \left( \frac{r}{R} \right) dt^2,$$

(2.1)

where $d\Omega^2 = d\theta^2 + \sin^2 \theta d\varphi^2$, and $R$ is related to the cosmological constant $\lambda = 3/R^2$ (de Sitter [180]). Einstein’s understanding of singularities has (understandably) been questioned,⁴ but let me note that:

² This section adds only little to [187], [49], and [175].
³ The Friedman solution (rediscovered by Lemaître [118]) dates from 1922, so it was not part of the earliest discussions.
⁴ For example: ‘He tended to be more disturbed by (what today we would call) merely apparent singularities and less disturbed by (what we would call) real singularities. Einstein had strong a priori ideas about what
Not just Einstein but even Hilbert [96], who was also an expert in GR, failed to recognize the apparent “singularity” of the Schwarzschild metric at \( r = 2m \) as a mere coordinate singularity.\(^5\)

Not just Einstein but even Weyl [191] saw \( r = \pi R/2 \) as a genuine singularity in de Sitter’s metric.\(^6\) But Einstein was not confused by the coordinate singularity at \( r = 0 \) and did understand the need for different coordinate patches in order to define regularity (and hence singularity) of the metric:

Furthermore, the continuity condition for the \( g_{\mu\nu} \) and the \( g^{\mu\nu} \) [including their first derivatives, as he says earlier] must not be interpreted to mean that it must be possible to choose coordinates such that the conditions are satisfied in the entire space. Obviously, one must only demand that for the neighborhood of every point one can select coordinates such that the continuity conditions are met within this neighborhood. [55, p. 271]; [57, p. 37].

Counterfactually granting Einstein [55] that \( r = \pi R/2 \) is indeed a “singularity”—since he believed that ‘no choice of coordinates can remove this discontinuity’—the interesting point is his subsequent argument that it is ‘genuine’. Namely, he defines a point \( P \) to lie ‘in the finite domain when it can be connected by a curve with a fixed, chosen point \( P_0 \), so that the distance integral \( \int_{P_0}^{P} ds \) has a finite value.’ He then shows results a correct physical theory should deliver. (…) He tended to push aside technical problems and jump over essential difficulties. (…) His difficulties lay with cases that today beginning students of GTR take for granted.’ [49, p. 185].

\(^5\) And they were not the only ones: ‘until the early 1960s the general opinion was that it was a real singularity’ [49, p. 188]. Hilbert [96] required a coordinate transformation intended to remove an apparent singularity to be regular (i.e. continuous and continuously differentiable) and invertible even at the location of the (alleged) singularity. This defeats the purpose and is not even true for changing from Cartesian to polar coordinates. ‘How Hilbert, one of the great mathematical minds of the century, could have failed to appreciate this elementary point defies rational explanation.’ [49, p. 193]. The sphere at the Schwarzschild radius \( r = 2m \) was alternatively described as a “discontinuity” (Schwarzschild), “magic circle” (Eddington), “barrier” (Kottler), “limit circle” (Brillouin), and even “the death” (Nordmann); I learnt this in a seminar by Dennis Lehmkuhl on April 12th, 2021. After important but inconclusive work by Eddington (1924), Lemaître [119], Sect. 11, correctly concluded that ‘The singularity of the Schwarzschild field [i.e. at \( r = 2m \)] is thus a fictitious singularity, analogous to that which appears at the horizon of the centre in the original form of the de Sitter universe.’ However, his work was not well known and took decades to be digested. For example, Einstein [56], p. 922, apparently unaware of [119], still called the Schwarzschild metric “singular” at \( r = 2m \) since \( g_{00} \) vanishes and ‘both light rays and material particles take an infinitely long time (measured in “coordinate time”) in order to reach the point \( r = 2m \) when originating from a point \( r > 2m \)’ (scare quotes are Einstein’s but emphasis added to highlight the changes from his reply to de Sitter). See also [58, 82], Penrose [47], Sect. 1.2, Penrose [49], Sect. 2, and van Dongen and Lehmkuhl (in progress). It seems that Finkelstein [63] was the first author to introduce the modern understanding of \( r = 2m \) (now no longer seen as singular) as a one-way membrane. But he did not use the term “event horizon”, which had just been introduced by Rindler [161], who in turn did not mention the Schwarzschild solution! It was Penrose [150] who first put all this together.

Until the 1930s almost no attention was paid to what we now understand to be the real singularity of the Schwarzschild metric at \( r = 0 \), firstly because Newtonian gravity has a similar singularity seen as unproblematic, and secondly because this vacuum solution was not supposed to be valid in the interior of a star, i.e. as \( r \to 0 \); indeed no star was known or expected to have a radius \( r < 2m \). This was also Einstein’s (final) reply to Hadamard, who during a lecture in Paris on April 5, 1922, asked Einstein what happened if the radius of the Sun were less than the Schwarzschild radius [6].

\(^6\) See Janssen [100], pp. 202–207, for this episode. de Sitter space-time, independently introduced by Levi-Civita [121], has no singularities at all (it is also globally hyperbolic), although it is not geodesically convex [9].
that the (alleged) singularities at $r = \pi R/2$ lie ‘in the finite domain’, concluding that ‘the De Sitter solution has a genuine singularity’ (eine echte Singularität). This, then, he considers ‘a grave argument against the admissibility of [de Sitter’s] solution’. Thus Einstein’s logic seems crystal clear:

- He first defines precisely what he means by a singular space-time (or ‘solution’, as he calls it):
  1. There must be at least one point $P$ where in all coordinate systems some component $g_{\mu\nu}$ or $g^{\mu\nu}$ (or its first derivative) fails to be continuous (this is the case, for example, if $\det(g) = 0$);
  2. Such points $P$ can be connected to “regular” points $P_0$ by some curve of finite proper length.

- Admissible space-times must be non-singular, i.e., either the metric is regular at all points $P$, or, if there are any singular points, these must lie at infinite (proper) distance from all regular points.

Criterion 1 is at odds with the modern notion of a (smooth) Lorentzian manifold, which cannot contain such points. But this was far from obvious at the time and still needed to be spelled out as late as 1963: 7 the clue to clarity is to refuse ever to speak of a singularity but instead to phrase everything in terms of the properties of differentiable metric fields on manifolds. If one is given a manifold, and on it a metric which does not at all points satisfy the necessary differentiability requirements, one simply throws away all the points of singularity. The starting point for any further discussion is then the largest submanifold on which the metric is differentiable. This is done because there is not known any useful way of describing the singularities of a function except by describing its behavior at regular points near the singularity. The first problem then is to select a criteria which will identify in an intuitively acceptable way a “nonsingular space.” Evidently, differentiability is only a minimum prerequisite, since everything becomes differentiable when the singular points are discarded. The problem is rather to recognize the holes left in the space where singular (or even regular) points have been omitted. (Misner [132, p. 924])

I will discuss Misner’s paper below. Returning to Einstein [55], at least in a Whig interpretation of history his second criterion may be claimed to foreshadow geodesic incompleteness. It would have been more reasonable, though, if he had used causal geodesics instead of arbitrary curves; first deciding for causal curves on physical grounds and/or because this is enough for all interesting examples, 8 his criterion does not even make sense if arbitrary (as opposed to geodesic) causal curves are allowed. 9

7 Even for mathematicians, leaving out singular points from the space under consideration is not an obvious move, as for example algebraic geometry shows. Within differential geometry, stratified spaces provide another example [158].
8 In de Sitter space-time one can reach singular points ($r = \pi R/2$) from regular ones ($0 \leq r < \pi R/2$) with any kind of curve.
9 First, any two points that can be connected by a lightlike (i.e. null) curve can also be connected by either a lightlike geodesic or a timelike curve; see e.g. [131], Theorems 2.20 and 2.22, or [112], Proposition 5.13,
Even in the absence of satisfactory definitions of a singular space-time, towards which little progress was made until the 1960s, one may look for “singularity theorems” prior to Penrose’s. But because it is not clear what is meant by a “singularity”, there is considerable leeway in identifying such theorems, ranging from accepting Genesis 1:1 to rejecting anything before Penrose [148] to deserve the name.\(^\text{10}\) For example, both Tipler et al. [187] and Earman [49] attribute the first singularity theorem to Tolman and Ward [188], whereas Senovilla and Garfinkle emphatically start with Raychaudhuri [159].

Prior to Robertson [165] and Walker [190] and Tolman and Ward [188] analyzed closed homogenous and isotropic universes filled with matter in the form of a perfect fluid with \(\rho > 0\) and \(p \geq 0\) (and zero cosmological constant), to conclude that such models have a beginning in a state of zero volume and also recollapse after a state of maximal expansion. However, they did not conclude that there was a big bang, instead believing in some sort of cyclic process of repeated expansions, collapses, and bounces.\(^\text{11}\)

Earman [49] reviews various analogous cosmological results from the 1930s, due to Tolman, Robertson, de Sitter, Synge, and Lemaître, which with hindsight may be classified as “singularity theorems”, in which “singularities” are defined in diverse ways, including zero volume (or radius), infinite density, as well as problematic expressions for the metric. Lemaître [119], which stands out in its astronomy, mathematics, and physics, was the first paper to use an energy condition.\(^\text{12}\) But all these authors immediately give arguments why their results must be mathematical artifacts, typically by appealing to their unrealistic underlying idealizations or to unknown new physics. Even Lemaître [119] writes:

all going back to [152]. Second, as pointed out by Clarke [26, pp. 2–3] in precisely the present context of [55], any two points that can be connected by a timelike curve can also be connected by a timelike curve of finite proper length (namely by ‘wiggling the curve to make its speed close to the speed of light’), which makes Einstein’s criterion vacuous as stated. See also [49] and [138] for broader discussions of this episode.

\(^\text{10}\) Following the Friedman solution from 1922 and its rediscovery by Lemaître [118], who—unlike the mathematician Friedman—as an astronomer related the solution to the redshifts of extragalactic nebulae that had been discovered in 1917, it was well understood in the 1930s that theoretically speaking the expanding universe might have had a beginning, but this possibility was not taken very seriously until the 1960s [187]. See [137, 138], and [181] for Einstein’s views. Believing the universe was static, his initial reaction to Friedmann [66] and Lemaître [118] was that their solutions were ‘irrelevant’ and ‘abominable’, respectively, but after talking to Eddington in 1930 and Tolman in 1931 he came to accept an expanding universe. As a letter to Besso from March 1, 1931 shows, Einstein also clearly realized, like many others, that ‘the expansion of matter points to a beginning of time which lies \(10^{10}\), respectively \(10^{11}\) years in the past.’ (‘Der Hacken ist aber, dass die Expansion der Materie auf einen zeitlichen Anfang schliessen lässt, der \(10^{10}\) bzw. \(10^{11}\) Jahre zurückliegt’). The so-called Einstein–de Sitter universe (which is a flat FRLW model remaining popular until the end of the twentieth century) indeed starts from what we now call a big bang, but like other leading relativists including de Sitter (!) and Eddington, Einstein tried to avoid this conclusion, in his case by ‘pointing out that the inhomogeneity of the distribution of stellar material makes our approximate treatment illusory’; in other words, by appealing to the unphysical idealizations inherent in the symmetries of the Friedman–Lemaître–Robertson–Walker solution. As we will see, this became a common move against singularities in GR, which Penrose [148] and Hawking [89–91] made obsolete.

\(^\text{11}\) From a modern point of view such a process means that their space-time can be extended through \(t = 0\), which in the case at hand is impossible because of the curvature singularity as \(t \rightarrow 0\) (in this case it is the Ricci scalar that diverges).

\(^\text{12}\) This happens after eq. (2,10) on page 678 of the reprint in 1997, where he states that ‘in all reasonable applications \(T_1^4, T_2^4\) and \(T_3^4\) will be negative, and in all cases less than \(T_4^4 = \rho\) in absolute value’.
The matter has to find, though, a way of avoiding the vanishing of its volume. ([119], p. 678 of 1997 reprint)

Meanwhile, progress was also made on the astrophysical side. Light stars were found to retire as white dwarfs, in which nuclear burning has ended and inward gravitational pressure is stopped by a degenerate electron gas. In 1931, Chandrasekhar and Landau independently discovered that this only works for masses \( M \leq 1.46M_\odot \) (I take the current value); heavier stars collapse into neutron stars (typically after a supernova explosion). But also these have an upper mass, as first suggested by Oppenheimer and Volkoff [141]; the current upper bound is about 2.3\( M_\odot \). Heavier stars collapse into a black hole.

Yet this conclusion, so obvious with hindsight, was resisted as much as the idea of a big bang, despite a paper by Oppenheimer and Snyder [142] in which the collapse process was explicitly described within GR, albeit under the assumption of spherical symmetry and an idealized equation of state. It was:

a manuscript that has strong claims to be considered the most daring and uncannily prophetic paper in the field. There is nothing in this paper which needs revision today. (Israel [99, pp. 226–227])

But there is a big difference between our perspective and the contemporary one (see also Sect. 7):

At the end of the 1930s not only was there no agreement on how to define singularities, there was not even a consensus about the status of singularities in the key test case, the Schwarzschild solution. (Earman [49, p. 240])

Restricting this summary to events directly relevant to Penrose’s paper, all authors agree that the next key paper was Raychaudhuri [159]; as already mentioned, Senovilla and Garfinkle [175] even take this to contain the very first singularity theorem in GR. Raychaudhuri studies a universe filled with dust, i.e.

\[
T^{\mu\nu} = \rho u^\mu u^\nu,
\]

but allows \( \rho \) to be an arbitrary positive function \( \rho(t, x) \), as opposed to previous treatments where \( \rho \) depends on \( t \) alone. Indeed, the goal of Raychaudhuri’s paper is clear from its abstract, which states that:

a simple change over to anisotropy without the introduction of spin does not solve any of the outstanding difficulties of isotropic cosmological models. (Raychaudhuri [159, p. 1123])

---

13 See [99, 123, 186] for the relevant history, which is fascinating by itself.

14 As pointed out to me by Senovilla, the interior part of their solution is the contracting part of Friedmann’s solution. See also [43], whose work was all but ignored; the poor author ‘died soon after on the operation table’ [34].

15 Raychaudhuri’s reasoning breaks down if a pressure term \( p h_{\mu\nu} \) is added to \( T_{\mu\nu} \), since in that case the congruence he uses is no longer geodesic, which in turn adversely affects his analogue of the central equation (2.3).

16 By ‘spin’ Raychaudhuri means \( \omega_{ij} = \frac{1}{2}(\nabla_i u_j - \nabla_j u_i) \), noticing that \( \omega_{ij} = 0 \) iff the flow \( u^\mu \) is hypersurface orthogonal. He cites [83] as an example where ‘spin’ contributes to making this famously
Instead of the parameter $\theta = \nabla \mu u^\mu$ familiar from modern renditions of his work, Raychaudhuri uses the volume element $\sqrt{-\det(g)}$, or rather $G = (-\det(g))^{1/6}$, but these are related.\(^{17}\) For this reason, the famous equation for $\theta$ named after him, namely (assuming that $\nabla u u = 0$, which follows from (2.2))

$$\nabla_u \theta \equiv \dot{\theta} = -\frac{1}{3} \theta^2 - \sigma_{\mu\nu} \sigma^{\mu\nu} + \omega_{\mu\nu} \omega^{\mu\nu} - R_{\mu\nu} u^\mu u^\nu,$$

(2.3)
cannot be found in his paper;\(^{18}\) instead, one finds an equivalent (but less transparent) second-order differential equation for $G$. Either way, if $R_{\mu\nu} u^\mu u^\nu \geq 0$ (which Raychaudhuri infers from the energy condition $\rho \geq 0$ and the Einstein equations) and $\omega_{\mu\nu} = 0$ (i.e. if the flow is irrotational), then, since $\sigma_{\mu\nu} \sigma^{\mu\nu} \geq 0$ for any congruence, (2.3) implies that if $\theta > 0$ on some hypersurface orthogonal to the flow at some time, i.e. the universe is expanding, then $\theta \to \infty$ at some finite time in the past, which corresponds to $G \to 0$. Since $\rho \sqrt{-\det(g)} = \rho G^3$ is constant, this suggests $\rho \to \infty$, but $G$ depends on the coordinates and $G \to 0$ could merely mark a singularity of the congruence. On Raychaudhuri’s assumption that matter moves along the timelike geodesics tangent to $u$, however, he does obtain a density and curvature singularity, adding that the time from such a ‘singular state’ to ‘the present state’ is a maximum for isotropic models, so that adding anisotropy to isotropic cosmological models even brings the initial singularity forward!\(^{19}\)

Similar comments apply to Komar [105], whose Introduction is worth quoting as a sign of the times:

A principal success of the general theory of relativity in the realm of cosmology is given by the Friedmann solution of the field equations. This solution, which employs the assumptions that the universe is spatially isotropic and that the state of matter may be represented by incoherent dust, yields the result that the universe is not stationary, but is rather in a state either of expansion from a singular point in time (which would correspond to creation), or of contraction toward a singular point in time (which would correspond to annihilation). The question naturally arises whether such singular points are a consequence of the

\(^{17}\) If $u^\mu$ is hypersurface orthogonal and normalized to $g_{\mu\nu} u^\mu u^\nu = -1$, the relation is $\theta = \partial_t \ln(\sqrt{-\det(h)})$, where $h$ is the metric induced by $g$ on the hypersurface to which the flow is orthogonal. See e.g. [112], Sect. 6.1, Eq. (6.25).

\(^{18}\) Here $\sigma_{\mu\nu} := k_{(\mu\nu)} - \frac{1}{2} h_{\mu\nu}$, with $k_{(\mu\nu)} = h_{\mu\nu} h^\rho \sigma_{\rho\nu}$, where $h$ is the spatial metric $h_{\mu\nu} = g_{\mu\nu} + u_{\mu} u_{\nu}$, and $\omega_{\mu\nu} = k_{[\mu\nu]}$. Eq. (2.3) appears, for example, in [52], Eq. (36), which is the earliest source I know.

\(^{19}\) In general, $\theta \to \infty$ could indicate that the hypersurface orthogonal to the flow becomes lightlike (and then timelike), in which case neither the density nor the space-time needs to be singular [28, 49, 167, 187]. Instead of a singularity in the energy density this implies that the domain of influence of some initial spacelike hypersurface carrying the initial data for the Einstein equations comes to an end, so that a (necessarily lightlike) Cauchy horizon forms. This happens, for example, in the Taub–NUT space-time [132, 133, 176]. Cauchy horizons may still lead to singularities under a perturbation that destabilizes the Cauchy horizon, cf. Sect. 6. This cannot happen in Bianchi type I models [94] and type IX models [176], where Raychaudhuri’s argument does lead to a singularity in the energy density and even in space-time.
particular symmetry presupposed in Friedmann’s model, or whether perhaps for
more general distributions of matter one need not expect instants of creation or
annihilation of the universe. The purpose of this paper is to show that singularities
in the solution of the field equations of general relativity are to be expected under
very general hypotheses (enumerated specifically below), and in particular that
the singular instant of creation (or annihilation) necessarily would occur at a
finite time in the past (or future, respectively). (Komar [105, p. 104])

Replacing ‘cosmology’ by ‘gravitational collapse’ and ‘Friedmann’ by ‘Schwarzschild’,
this purpose is hardly different from Penrose’s: it echoes the general spirit that one
would like to go beyond the simplified singular models of the past and see if their
singularities persist (which Komar shows they do).

In modern fashion, Komar just relies on energy assumptions, due to which he need
not specify $T_{\mu\nu}$ explicitly—unlike Raychaudhuri [159], whom he does not cite.20
These are: $T_{00} \geq 0$ and $T_{ij} \geq 0$, and $T_{00} = 0$ implies $T_{ij} = 0$. But by fiat he forces
his geodesic congruence to be the same as Raychaudhuri’s, including its identification
with the matter flow lines. This then leads him to very similar conclusions.21

Both papers crossed mathematical work on relativistic fluid mechanics,22 which in
turn crossed or included the study of spatially homogeneous (Bianchi) cosmologies
initiated by Gödel [84], Taub [185] and the Hamburg group, summarized by Heckmann
and Schücking [94]. The general goal was
to find singularity free [!] fluid filled Bianchi cosmological models. (Ellis [59,
p. 2])

This research influenced all subsequent developments, where Raychaudhuri’s argu-
ment survived as a lemma for the mature singularity theorems, in whose proofs it is
used to prove focusing of geodesics.

To preface the next phase involving the young relativists Misner, Penrose, and
Hawking, I quote:

Wheeler and Sciama had a widespread impact on the development of the under-
standing of singularities, particularly through their students. Wheeler’s students
included Misner, Shepley, Thorne, and Geroch, while Sciama’s included Ellis,
Hawking, Carter, Rees, and Clarke. However, Sciama has claimed, on occasion,
that his most important contribution to relativity has been in influencing Roger
Penrose to work on the subject! (Ellis [187, p. 136])

The last pre-Penrose papers on singularities I mention are Misner [132] and Shepley
([176, 177]).23 Once again, but for a different reason, it is instructive to quote from
the Introduction of [176]:

---

20 [160] seems a little reprimand for this. [173] combines their results in his Theorem 5.1.

21 Similar results were found by Landau, whose Russian colleagues recognized their inconclusiveness [49,
§5]. This eventually led to the so-called BKL-approach to singularities in GR. See [4] or a recent review.

22 For example, prior to 1955 by Eisenhart, Synge, Lichnerowicz, and subsequently by the Hamburg group
in GR founded by Jordan, consisting notably of Behr, Ehlers, Heckmann, and Schücking. See [187] for
references.

23 Misner [132], footnote 4, thanks his PhD student Shepley for ‘preparing this review’ (i.e. of the definition
of a singularity in space-time in the Introduction), adding that ‘We have borrowed heavily from [lectures by]
Most cosmological models in general relativity have a point of singularity (...) which can be reached by a geodesic of finite total length from other points of the space-time manifold, where the metric is degenerate or otherwise irregular (for example, a point where a curvature scalar is infinite). All known dust-filled models, without cosmological constant, have singular points, and it may be conjectured that the presence of dustlike matter filling space always leads to a singularity. This paper shows that this conjecture is true for an important class of universes. (Shepley [176, p. 1403])

This shows that using geodesic incompleteness as a characterization of a singularity was familiar at the time, although e.g. Heckmann & Schucking [94] do not mention it. But it was seen as insufficient:

completeness is vulgarly used as the definition of non-singularity. (Ryan & Shepley [167, p. 80])

Thus Misner [132] refines this idea. After the call to arms already quoted in Sect. 2, followed by a comparison with the Riemannian case, geodesic incompleteness is proposed as a necessary condition for a space-time to be singular, in which case it is sufficient that ‘some scalar polynomial in the curvature tensor and its covariant derivatives be unbounded on an (open) geodesic segment of finite length,’ since this implies inextendibility of the geodesic in question into some potential extension of the space-time. 25

Shepley [176] claims that Bianchi type IX universes filled with dust are singular even if the congruence of matter flow lines is rotating. But although he finds points where \( \text{det}(g) = 0 \), he shows neither geodesic incompleteness, nor divergence of the density or of some curvature scalar, admitting that ‘The question of what happens as the singular point approaches is presently under investigation’ (p. 1409). 26

3 Gravitational collapse and space-time singularities: exegesis

The brevity (2.5 pages) and brilliance of Penrose [148] makes almost every comma relevant, but let me try some restraint. Following a brief discussion of gravitational collapse, Penrose explains his motivation:

L. Marcus’ (sic), i.e. Lawrence Markus (1922–2020). Shepley [177] is a Princeton PhD thesis supervised by Misner and Wheeler. Some relevant passages reappear in Ryan and Shepley [167], from which I will quote as if these were written in 1963–1964. Footnote 19 in Misner [132] shows that [133], which contains a similar analysis, must have been almost ready already in 1963.

24 Mentioning GR but not the search for a definition of a singular space-time, [95, 110], and [62] all study causal geodesic (in)completeness in Lorentzian manifolds in comparison with the Riemannian case.

25 For a detailed treatment of all of this see Beem, Ehrlich, & Easley [3], Chapter 6. See also Chruściel, [22], Sect. 4.4. If each finite segment of some incomplete geodesic is contained in a compact subset of space-time, then Misner does not seem to count this incompleteness as a singularity, since in that case space-time is typically extendible (as happens in Taub–NUT).

26 This may refer to [129], which is criticized in Collins and Ellis [28], or to Ryan and Shepley [167], Chapter 10, which at last proves geodesic incompleteness and inextendibility. See also [49].
The question has been raised as to whether [the Schwarzschild] singularity is, in fact, simply a property of the high symmetry assumed. The matter collapses radially inwards to the single point at the center, so that a resulting space-time catastrophe there is perhaps not surprising. Could not the presence of perturbations which destroy the spherical symmetry alter the situation drastically? (…)

It will be shown that, after a certain critical condition has been fulfilled, deviations from spherical symmetry cannot prevent space-time singularities from arising. (Penrose [148, p. 57, 58])

Penrose does not strictly define what he means by a singularity, and cites none of the papers discussed in the previous section except [142]. But the examples he gives (namely gravitational collapse, Schwarzschild, and Kerr) suggest that he has curvature singularities in mind. Indeed, the preamble to his theorem shows that he does not identify singularities with incomplete geodesics: 27

The existence of a singularity can never be inferred, however, without an assumption such as completeness [i.e. inextendibility] for the manifold under consideration. [[148] (p. 58)]

Similarly, in a list of four possibilities to avoid singularities despite his theorem, 28 Penrose includes:

(c) The space-time manifold is incomplete [i.e. extendible]. 29 (p. 58)

*Geodesic* completeness first appears in the statement of the theorem, in which it is a *reductio ad absurdum* assumption, see below, whose negation as a way out of the ensuing contradiction is left to the reader and is nowhere defined as a singularity. The logic of the above two quotes therefore seems to be that inextendibility of the manifold, as opposed to a ‘physical’ singularity, could be the cause of geodesic incompleteness and as such has to be excluded to make room for a curvature singularity as the physically relevant cause of geodesic incompleteness. This logic foreshadows strong cosmic censorship; cf. Sect. 5.

We now come to the theorem, which states that the following assumptions are together inconsistent:

27 I suppose ‘completeness’ here means *inextendibility* as opposed to *geodesic completeness*, since a few lines later Penrose states his theorem as the incompatibility between certain assumptions (see main text) and future null completeness, which is subsequently specified as ‘Every null geodesic in $M^4_+\!$ can be extended into the future to arbitrarily affine parameter values (null completeness)’ i.e. null geodesic completeness. Similarly, (c) above only makes sense if Penrose takes ‘incomplete’ to mean *extendible*. I thank Erik Curiel and José Senovilla for a discussion of this point (via email, May 2021), about which we agree.

28 The others are: (a) Negative local energy occurs; (b) Einstein’s equations are violated; (d) The concept of space-time loses its meaning at very high curvatures—possibly because of quantum phenomena.’ Penrose omits (e): There is no Cauchy surface. But, as Senovilla pointed out to me, on a modern understanding this may be taken to be a consequence of his (c).

29 A footnote at this point adds that ‘The “I’m all right, Jack” philosophy with regard to the singularities would be included under this heading!’. This reconfirms that Penrose considers extendibility a possible but cheap way to avoid singularities.
Theorem 1 For any space-time \((M, g)\), the following assumptions are contradictory:\(^{30}\)

1. A space-time is a four-dimensional time-orientable Lorentzian manifold;\(^{31}\)
2. \((M, g)\) is future null geodesically complete;
3. \(M\) contains a non-compact Cauchy surface \(C_3\);
4. One has \((-R_{\mu\nu} + \frac{1}{2}g_{\mu\nu}R)t^\mu t^\nu \geq 0\) for any timelike vector \(t\) (at any point);\(^{32}\)
5. There exists a trapped surface in \(M\).

Penrose [148] defines a Cauchy surface by the property that every inextendible timelike or null geodesic meets it.\(^{33}\) This specific concept of a Cauchy surface in \(\text{gr}\) seems to have originated with Penrose, who therefore, with Leray [120], Choquet-Bruhat [12] and Geroch [76] should be included on the list of architects of global hyperbolicity in \(\text{gr}\).\(^{34}\) Here is a picture from 1965 in the master’s hand:

---

\(^{30}\) In the preamble to this list Penrose supposes that \(M\) is ‘the future time development’ of \(C_3\). Short of anything like the results of Choquet-Bruhat & Geroch [16], this simply seems to refer to his definition of \(C_3\) as a Cauchy surface, see below.

\(^{31}\) Penrose always works in signature \((+−−−)\) and includes a cosmological constant \(\lambda\) in condition 4, which I put to zero.

\(^{32}\) For null vectors \(t\) this inequality implies \(R_{\mu\nu}t^\mu t^\nu \leq 0\) by continuity. This is what is actually used in the proof. Penrose [150] indeed states the curvature condition in the latter way. As we have seen, such conditions were hardly new in 1965, neither as curvature conditions nor, via the Einstein equations, as energy conditions (which here would simply be \(T_{\mu\nu}t^\mu t^\nu \geq 0\)).

\(^{33}\) On the tacit understanding that \(C_3\) is achronal (or semispacelike, in Penrose’s own terminology) this is equivalent to the property \(D(C_3) = M\) (where the domain of dependence is defined via timelike curves), and hence to the later definition to the effect that \(C_3\) is any subset of \(M\) that intersects every inextendible timelike curve once. See Penrose [150, 152]. Penrose [146] had already loosely defined a Cauchy surface \(C_3\) as a hypersurface that intersects every null geodesic exactly once.

\(^{34}\) Domains of dependence \(D(S)\) of hypersurfaces \(S \subset M\), and the closely related concept of a Cauchy surface (for which \(D(S) = M\) originate in hyperbolic PDE theory in flat space, for which [29] was the standard reference at the time. The equivalence between the existence of a Cauchy surface as Choquet-Bruhat [148] defines it and global hyperbolicity as used in the analysis of the Cauchy problem for \(\text{gr}\) by Leray [120] and Choquet-Bruhat [12], i.e. compactness (in a suitable topology) of the space of causal curves \(C(x, y)\) for any pair \((x, y)\) of causally connected points, is announced in Penrose ([150], p. 191), based on lectures in 1967 (a footnote on the same page says that ‘This will be shown elsewhere’). Penrose even mentions the factorization theorem \(M \cong \mathbb{R} \times S\) in the presence of a Cauchy surface \(S\). The proof of both results appeared in print only in Geroch [76], but it seems safe to assume that all of this must have been reasonably clear to Penrose already at the end of 1964. Indeed, the author of papers like [143] would have had no trouble adapting the description of domains of dependence and Cauchy surfaces in flat space in terms of characteristics in Courant and Hilbert [29] to curved space-times. For some history of the Cauchy problem in \(\text{gr}\) see [14, 15, 164, 184].
The concept of a trapped surface was also introduced by Penrose [148] himself, defined as a closed, spacelike, two-surface $T^2$ with the property that the two systems of null geodesics which meet $T^2$ orthogonally converge locally in future directions at $T^2$. (p. 58)

Qua idea and definition trapped surfaces belong to the most innovative and lasting contributions of [148]. In the presence of a radial coordinate $r$, as in the Schwarzschild, Reissner–Nordström, and Kerr solutions, Penrose’s convergence condition is equivalent to the gradient $\nabla r$ being timelike, which in the Schwarzschild solution happens for $r < 2m$, and which in the other two (subcritical) cases is the case at least for a while after crossing the event horizon. The convergence condition can be stated in terms of the null hypersurface $C$ generated by the future directed null congruence emanating from any spacelike surface $T^2$. Pick a tetrad $(e_1, e_2, L_+, L_-)$ with $(e_1, e_2)$ spacelike and tangent to $C$, $L_\pm$ null and tangent (and hence orthogonal) to $C$, such that $g(e_i, e_j) = \delta_{ij}$, $g(e_i, L_\pm) = 0$ for $i, j = 1, 2$, $g(L_+, L_-) = 0$, and $g(L_+, L_-) = -1$, all defined on $C$. With $\theta_\pm = \nabla_\mu L_\mu^\pm$, the surface $T^2$ is trapped iff $\theta_\pm < 0$ throughout $T^2$ for both signs (whereas in Minkowski space-time one of $\theta_\pm$ is $> 0$ whilst the other $\theta_\mp$ is $< 0$).

Thirty years later, Penrose recalled how he suddenly got the idea in the late autumn of 1964.\[36\]

---

\[35\] See Hawking and Ellis [93], chapter 9, for an early analysis of trapped surfaces in causal theory per se, and [174] for a recent review. The study of trapped surface formation from the PDE point of view began with [172] and was continued by e.g. Christodoulou and his students [5, 17, 18, 20, 37, 122]. Schoen and Yau [171] used the concept of a marginally outer trapped surface (MOTS), which is a Lorentzian analogue of a minimal surface in Riemannian geometry and is probably the most important variation on Penrose’s trapped surface (in the latter both expansion factors $\theta_\pm$ of the null congruence emanating from $T^2$ are strictly negative, whereas for a MOTS one equals zero). See [1] and [72] for reviews. MOTSS appear in later singularity theorems with weaker assumptions [54] as well as in the Penrose inequality, see Sect. 5.

\[36\] It is not obvious that this recollection is about trapped surfaces, but on July 2, 2022, Penrose confirmed that it was.
My conversation with Robinson stopped momentarily as we crossed a side road, and resumed again at the other side.\textsuperscript{37} Evidently, during those few moments an idea occurred to me, but then the ensuing conversation blotted it from my mind! Later in the day, after Robinson had left, I returned to my office. I remember having an odd feeling of elation that I could not account for. I began going through in my mind all the various things that had happened to me during the day, in an attempt to find what it was that had caused this elation. After eliminating numerous inadequate possibilities, I finally brought to mind the thought that I had had while crossing the street.\textsuperscript{38}

The proof of the singularity theorem displays some of the unprecedented causal and topological techniques Penrose had developed in the preceding years in connection with gravitational radiation \cite{136, 145, 147}. It is very similar to all later reformulations I am aware of:\textsuperscript{39}

- Assumptions 1, 2, 4, and 5 imply (directly) that $\partial I^+(T^2)$ (which Penrose calls $B^3$) is compact;
- Assumptions 1 and 3 imply (by contradiction) that $\partial I^+(T^2)$ is non-compact.

The argument for the first point is that the null Raychaudhuri equation for $\theta_{\pm}$ (which is due to Penrose, who writes $\rho = -\frac{1}{2}\theta_{\pm}$ for both cases) and the boundary condition $\theta_{\pm} < 0$ give a focal point on each null geodesic ruling $C$. It is left to the reader to realize that $\partial I^+(T^2) = J^+(T^2) \setminus I^+(T^2)$, which follows from global hyperbolicity (which makes $J^+(T^2)$ closed), and that beyond each focal point the null geodesic enters $I^+(T^2)$ and hence leaves $\partial I^+(T^2)$, in order to conclude with Penrose that $\partial I^+(T^2)$ is compact.\textsuperscript{40}

Penrose’s homotopical argument for the second point seems correct but has not survived;\textsuperscript{41} the most rigorous proof I know of (see e.g. \cite{71}) is obtained by combining the following results:\textsuperscript{42}

\textsuperscript{37} This refers to Ivor Robinson (1923–2016), as opposed to David Robinson. Both were relativists associated with Penrose via King’s College London, where Penrose was a postdoc from 1961–1963 (he wrote his 1965 paper during late 1964 at Birkbeck College). \cite{166} by David Robinson is a relativist’s memoir that also contains interesting information about Penrose.

\textsuperscript{38} It is hard to overlook the analogy with Poincaré’s famous recollection of his crucial flash of insight into the theory of Fuchsian functions, referring to an event in 1880: ‘At that moment I left Caen where I then lived, to take part in a geologic expedition organized by the École des Mines. The circumstances of the journey made me forget my mathematical work; arrived at Coutances we boarded an omnibus for I don’t know what journey. At the moment when I put my foot on the step the idea came to me, without anything in my previous thoughts having prepared me for it; that the transformations I had made use of to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify this; I did not have the time for it, since [as soon as] I sat down in the bus I resumed the conversation already begun, but I was entirely certain at once. On returning to Caen I verified the result at my leisure to salve my conscience.’ \cite{86, p. 217}.

\textsuperscript{39} See the second set of references in footnote 1.

\textsuperscript{40} See e.g. \cite{2}, paraphrased also in \cite{112}, Sects. 6.3–6.4, for a detailed outline of this reasoning.

\textsuperscript{41} In our discussion on July 2, 2022, Penrose even described his proof of the second point as ‘stupid’, and acknowledged Misner as the source of the simpler proof that later appeared in Hawking and Ellis \cite{93}, Sect. 8.2, Theorem 1.

\textsuperscript{42} The first is Lemma 3.17 in \cite{152}; see also Minguzzi \cite{131}, Theorem 2.87 (iii). The second is Theorem 1 in Budic et al. \cite{8}; see also Theorem 1 and Corollary 1 in Galloway \cite{70}.

\textsuperscript{186, p. 227}
1. Any achronal boundary such as $\partial I^+(T^2)$ is a (locally Lipschitz) topological hypersurface.

2. Any compact achronal topological hypersurface in a globally hyperbolic space-time is a Cauchy surface; hence by Geroch [76] Property 7, any Cauchy surface in this space-time is compact.

4 Intermezzo: singularities revisited and refined

In his PhD Thesis, Hawking writes without any further ado (or mention of inextendibility):

any model must have a singularity, that is, it cannot be a geodesically complete $C^1$, piecewise $C^2$ manifold. (Hawking [89, Sect. 4.1])

And similarly in the paper announcing his singularity theorem, citing Misner [132] at this point:

Space-time is said to be singularity free if it is time-complete (all timelike geodesics can be extended to arbitrary length) and if the metric is a $C^2$ tensor field. (Hawking [90, p. 444])

In his subsequent Adams Prize Essay, Hawking explicitly assumes inextendibility and proposes to:

take timelike and lightlike geodesic incompleteness as our definition of a singularity of space-time. (Hawking [91, Sect. 6.1])

All of this was directly inspired by Penrose [148], but it seems to have been Hawking, rather than Penrose, who first advocated causal geodesic incompleteness as the def-

---

43 See [59] and https://en.wikipedia.org/wiki/Adams_Prize for the history of this award. ‘The Adams Prize is awarded jointly each year by the Faculty of Mathematics and St John’s College to UK-based researchers, under the age of 40, doing first class international research in the Mathematical Sciences.’ Each year a topic is set, which in 1966 was Geometric problems of relativity. Although Penrose won the prize, Hawking was awarded an ‘auxiliary Adams Prize’ in the same year. The memoir [59] by George Ellis about Hawking is especially valuable since Ellis was around himself at the time: ‘After the publication of [Penrose’s] paper in January 1965, the members of Dennis Sciama’s general relativity group in the Department of Applied Mathematics and Theoretical Physics at Cambridge University (particularly Stephen Hawking, myself, and Brandon Carter) hurriedly tried to learn the new methods that Penrose had introduced. We were assisted in this by discussions with Felix Pirani and the group at King’s College, London; with John Wheeler and Charles Misner, who visited Cambridge from the USA for an extended period; and with Roger Penrose and Bob Geroch, who was visiting Penrose at Birkbeck College, London. In particular we had a one day seminar in Cambridge attended by the members of the King’s College group, where I and Brandon Carter summarized our understandings of the ingredients of Penrose’s theorem. (…) Stephen arrived at [his] results by discussions with the Cambridge group that under Dennis Sciama’s guidance met to discuss ideas at tea time each day, and with the London groups (as well as attending many seminars, we used to regularly catch the train to attend lectures on general relativity at King’s College, London).’ [59, pp. 3–4]. See also Tipler et al. [187], pp. 135–136.
inition of a space-time singularity. In a later defense of this definition, Hawking and Ellis make both a physical and a pragmatic point:44

Timelike geodesic incompleteness has an immediate physical significance in that it presents the possibility that there could be freely moving observers or particles whose histories did not exist after (or before) a finite interval of proper time. This would appear to be an even more objectionable feature than infinite curvature and so it seems appropriate to regard such a space as singular. (…) The advantage of taking timelike and/or null incompleteness as being indicative of the presence of a singularity is [also] that on this basis one can establish a number of theorems about their occurrence. (Hawking & Ellis [93, p. 258])

As we have seen, crucial nuance is lost in doing so. Although Geroch initially used practically the same definition as Hawking, also citing Misner [132], his subsequent analysis [74] is more critical:

(a) there is no widely accepted definition of a singularity in general relativity;
(b) each of the proposed definitions is subject to some inadequacy. (Geroch [74, p. 526])

This paper is still worth reading, as is its more technical successor [77]; see also [3, 26, 31, 32, 48, 173].

Almost 60 years later, none of the problems that arose from identifying space-time singularities with incomplete causal geodesics has been resolved: timelike, null (i.e. lightlike), causal, and general geodesic incompleteness are all inequivalent to each other; each is in turn inequivalent to incompleteness with respect to curves with either arbitrary or bounded acceleration; geodesically incomplete manifolds (of any kind) exist without curvature singularities; global hyperbolicity is compatible with geodesic incompleteness; geodesic completeness is compatible with a lack of global hyperbolicity; et cetera.

Penrose [154] gave a new definition of a singularity (i.e. Definition 2 below) ‘which is not quite the same as that suggested by the singularity theorems’ (p. 85). It has two novel and distinguishing features:

1. It is perspectival, in that not just the singularity matters but also its relationship to points which it can influence. This is crucial for understanding Penrose’s mature idea of cosmic censorship (Sect. 5).
2. Inextendible timelike curves replace incomplete causal geodesics as marks of singularities.

The change from ‘causal’ to ‘timelike’ is innocent,45 but the change from incomplete geodesics to inextendible curves is not (this change is required by the proof of Theorem 5 below).46 The key difference lies in the possibility that an inextendible

---

44 Also cf. Geroch [74], p. 534: ‘Finally, we remark that geodesic incompleteness (…) is commonly used as a definition of a singularity because with such a definition one can show that large classes of solutions of Einstein’s equations are singular.’

45 Cf. Theorem (2.3) in [80]; one could even use lightlike curves [65, §3.3]. Causal geodesics would lead to some weaker causality condition than global hyperbolicity, cf. Theorem 5 below.

46 It is the second (‘converse’) part of the proof of Theorem 5 below that does not work for causal geodesics instead of curves, since the curve c constructed there is not necessarily a geodesic. The fact that we just get
causal curve may go off to infinity and yet count as a singularity in the new sense, i.e. as a ‘locally naked’ one.\textsuperscript{47} For example, even $M = \mathbb{R}^4$ carrying plane gravitational waves, which is geodesically complete and strongly causal but not globally hyperbolic, will now count as singular,\textsuperscript{48} as does anti-de Sitter.\textsuperscript{49} Consequently, Definition 2 below accepts more singularities than the singularity theorems, and excluding these new singularities therefore certainly excludes all old-school singularities: it will be an \textit{unnecessarily strong} condition. However, in \textit{strongly causal space-times that are asymptotically flat at null infinity} the right intuition to think about the “new” singularities remains: causal curves that either crash into a curvature singularity or reach the edge of an extendible space-time.\textsuperscript{50}

Penrose states his new definition in the language of terminal indecomposable sets [80], but this language can be avoided. A simpler account is as follows:

1. Suppose that $(M, g)$ is \textit{past distinguishing}, i.e., $I^- (x) = I^- (y)$ implies $x = y$. This property allows us to exchange properties of points $x$ for properties of their timelike pasts $I^- (x)$.\textsuperscript{51}

2. By definition, if $z \in I^- (x)$ then $z$ can signal to $x$, or influence $x$; we may say that $z$ is \textit{naked} for $x$. For the next step, we note that $z \in I^- (x)$ is equivalent to the property $I^- (z) \subset I^- (x)$.

3. If $z$ is the endpoint of some future-directed timelike curve $c$, then $I^- (z) \subset I^- (x)$ if and only if

$$I^- (c) \subset I^- (x). \quad (4.1)$$

4. Crucially, this inclusion is also defined if $c$ has no endpoint (i.e. is inextendible). This suggests:

Footnote 46 continued

a causal curve at that point goes back to the definition of domains of dependence and Cauchy surfaces in terms of causal curves rather than geodesics.

\textsuperscript{47} Inextendible causal curves may also hover around in a compact set, but this is impossible in strongly causal space-times.

\textsuperscript{48} This example, which was kindly suggested to me in this context by Senovilla, is even due to Penrose [146]! A metric $g = dudv - dy^2 - dz^2 + H(u, y, z)du^2$ on $\mathbb{R}^4$, which describes gravitational or electromagnetic waves, is Ricci-flat iff $(\partial^2 y^2 + \partial^2 z^2)H = 0$, which is solved by $g = dudv - dy^2 - dz^2 + f(u)(y^2 - z^2) + g(u)yz$. See [3], Chapter 13, [64, 65], and [73] for details and generalizations. If $f \neq 0$ such space-times cannot be asymptotically flat at null infinity (cf. footnote 54), which would therefore be desirable. Also cf. Penrose [155], p. 623.

\textsuperscript{49} In this context, anti-de Sitter should be seen as a special case of the Malament–Hogarth space-times [97, 127], which are never globally hyperbolic and yet typically geodesically complete (as indeed is anti-de Sitter).

\textsuperscript{50} Penrose [154, 155] assumes that $(M, g)$ is strongly causal, which on the causal ladder is one step above being past distinguishing (Minguzzi [131], chapter 4, Definition 4.46), i.e. the condition that suffices to make sense of Definition 2. However, strong causality is necessary for Theorem 5 below. This is true for Penrose’s original proof because of his reliance on the theory of TIPs, and in my version of the proof this is because of the invocation of Theorem 2.53 in [131], where the case distinction I need relies on strong causality via Remark 2.54 and Theorem 2.80.

\textsuperscript{51} One may also use \textit{past-directed} inextendible causal curves, replacing $I^- (\cdot)$ by $I^+ (\cdot)$, etc. For strong cosmic censorship this definition is equivalent to the given one, as follows from Theorem 5 below, which follows either way [155].
Definition 2 A future-directed future-inextendible causal curve $c$ in $M$ is \textit{naked} for $x \in M$ if (4.1) holds.

For example, in Minkowski space-time, for fixed $x$ take $z \in I^{-}(x)$ and remove $z$. Then any fd future-inextendible timelike curve $c$ whose endpoint would have been $z$ satisfies (4.1).

5 Cosmic censorship: Penrose

Let us now return to Penrose [148]. In Sect. 3 I already noted that his emphasis on the inextendibility (or, as he put it, completeness) of the underlying space-time shows that Penrose was fully aware that (null) geodesic incompleteness alone was insufficient to point to a curvature (or “physical”) singularity, i.e., a “hole”. Similarly, Penrose realized that the “black” in our (later) term “black hole” was as yet lacking:52

Whether or not “visible” singularities inevitably arise under appropriate circumstances is an intriguing question not covered by the present discussion. (Penrose [148, footnote 9])

This was taken up in his magisterial and very influential paper ‘Gravitational collapse: The role of general relativity’ [151], in which Penrose introduced what we now call \textit{weak cosmic censorship}:53

We are thus presented with what is perhaps the most fundamental question of general-relativistic collapse theory, namely: does there exist a “cosmic censor” who forbids the appearance of naked singularities, clothing each one in an absolute event horizon? In one sense, a “cosmic censor” can be shown \textit{not} to exist. For it follows from a theorem of Hawking that the “big bang” singularity is, in principle, observable. But it is not known whether singularities observable from outside will ever arise in a generic \textit{collapse} which starts off from a perfectly reasonable nonsingular initial state. (Penrose [151, p. 1162])

For a mathematician like Penrose, this conjecture (here phrased as a question) relied on a precise definition of an “absolute event horizon”. Like his 1965 singularity theorem, also this concept used mathematical ideas that Penrose had initially developed for gravitational waves, especially his notion of null infinity $\mathscr{I}$, pronounced “scri” [144, 145, 150]. Perhaps because [151] is mainly addressed to physicists, he relegates this pretty important matter to a footnote, which reads:54

52 The rise of the term “black hole” is discussed in the appendix to [113], which was written by Erik Curiel. See also [99]and [186]. I will freely use this material in what follows. Briefly, although it seems that the term was coined by Dicke and/or Wheeler, [151] marked its first use in the scientific literature, where for the first few years it was usually surrounded by scare quotes. Other pioneers of the term were Bardeen, Carter, Christodoulou, and Hawking.

53 For reviews of cosmic censorship before the \textit{pde} people took over see Geroch & Horowitz [79], Sect. 5.4, Tipler, Clarke, & Ellis [187], Sect. 6, Earman [47], Chapter 3, and Joshi [102], Chapter 4. See Sect. 6 for the \textit{pde} approach.

54 Instead of Penrose’s weakly asymptotically simple space-times, I use the cleaner notion of space-times $(M, g)$ that are \textit{asymptotically flat at null infinity} [22, Sect. 3.1]. This means that $(M, g)$ has a conformal
In a general space-time with a well-defined external future infinity, the absolute event horizon would be defined as the boundary of the union of all timelike curves which escape to this external future infinity. In the terminology of [150], if \( M \) is a weakly asymptotically simple space-time, for example, then the absolute event horizon in \( M \) is \( \hat{I}_-(\mathcal{I}^+) \) \( [= \partial I^- (\mathcal{I}^+) \) \] \]. (Penrose [151, footnote 3, p. 1146 of the 2002 reprint)

Penrose leaves it to the reader to extract the mathematical definition of a black hole from this; and as in the case of the definition of a singularity, it was again Hawking who was subsequently more explicit:

A black hole on a spacelike surface \([S]\) is defined to be a connected component of the region of the surface bounded by the event horizon \([i.e. M \setminus I^- (\mathcal{I}^+) \cap S]\).

(Hawking [92, Abstract, p. 152])

Later (p. 156) he explains that this is a region ‘from which there is no escape to \( \mathcal{I}^+ \)’. This definition is repeated in Hawking & Ellis [93], Sect. 9.2, and is still standard, usually in the simple form \( M \setminus I^- (\mathcal{I}^+) \).

However, since there can be holes (in the sense of curvature singularities) that are not black in having no event horizons, and there may be event horizons that do not cover any kind of singularity, it might be better to call (a connected component of) \( M \setminus I^- (\mathcal{I}^+) \) a black object instead of a black hole. In any case, his definition of an event horizon led Penrose to a more precise definition of weak cosmic censorship:

I first review what is now the “standard” picture of gravitational collapse to a black hole—which I shall refer to as the establishment viewpoint. The picture is, in fact, a very good one, with a remarkable degree of internal consistency. But it is worth emphasizing that, among other things, this picture does depend on a very big assumption, that of cosmic censorship. In my own view, this

Footnote 54 continued

completion \((\hat{M}, \hat{g})\), i.e., a conformal embedding \( \iota : M \hookrightarrow \hat{M} \) such that \( \partial \hat{M} = \hat{M} \setminus (\mathcal{I}^+) \equiv \mathcal{I} = \mathcal{I}^+ \cup \mathcal{I}^- \), where \( \mathcal{I}^\pm := \partial \hat{M} \cap J^\pm (M) \). Writing \( \iota^* \hat{g} = \iota^* \Omega^2 g \) for some smooth function \( \Omega : \hat{M} \to [0, \infty) \), one asks \( \Omega > 0 \) on \( \mathcal{I}(M) \), \( \Omega = 0 \) on \( \partial M \), and \( \partial \mathcal{I} \neq 0 \) on \( \partial \hat{M} \). Adding fall-off conditions on the Ricci tensor guaranteeing that \( \mathcal{I}^\pm \) are null hypersurfaces in \( M \), one finally requires that \( \mathcal{I}^\pm \equiv \mathbb{R} \times S^2 \). To simplify the notation I always assume that null infinity \( \mathcal{I} \) is such that its timelike past \( I^- (\mathcal{I}^+) \) lies in \( M \), i.e. does not intersect the boundary \( \partial \hat{M} \). If this is not the case (as e.g. in anti-de Sitter space), just write \( I^- (\mathcal{I}^+) \cap M \) instead of \( I^- (\mathcal{I}^+) \) everywhere.

55 This does not mean it is uncontroversial [33]. The opposition between Rovelli and Wald is especially telling. Rovelli states that Penrose’s definition of an absolute event horizon is ‘very useless, because it assumes we can compute the future of real black holes, and we cannot’ [33, p. 30] to which Wald replies that ‘I have no idea why there should be any controversy of any kind about the definition of a black hole. There is a precise, clear definition in the context of asymptotically flat spacetimes (…) I don’t see this as any different than what occurs everywhere else in physics, where one can give precise definitions for idealized cases but these are not achievable/measurable in the real world.’ ([33], p. 32).

56 Simply take the Schwarzschild metric for \( m < 0 \), or the supercharged Reissner–Nordström metric (i.e. \( e^2 > m^2 > 0 \)), or the fast rotating Kerr solution (i.e. \( a^2 > m^2 > 0 \)); these are of course the simplest examples of Penrose’s naked singularities.

57 For an artificial example, in the usual conformal completion of Minkowski space-time \( \mathbb{M} \) [145], truncate \( \mathcal{I}^+ = \{(p,q) | p = \pi/2, q \in (-\pi/2, \pi/2)\} \) or \( \{(p,q) | p = \pi/2, q \in (-\pi/2, 0)\} \). This turns the region \( I^+ (0) \) into a fake black hole in flat space-time. What I call black objects also include regular (i.e. geodesically complete) black “holes”. See e.g. [10, 68, 179, 183], and references therein.
assumption might well be false in suitable circumstances. I believe that the consequences of possible violations of cosmic censorship are certainly worth considering seriously. In effect, this principle states that naked singularities do not develop out of an initially nonsingular state. A naked singularity may be said to exist in a space-time $M$ if there is a well-defined external infinity (that is, $\mathcal{I}^+$, if $M$ is weakly asymptotically simple) from which timelike curves may be drawn into the past that terminate on the singularity. I shall not make precise, here, the concept of “terminating on a singularity” but refer instead to Hawking’s concept of asymptotic predictability. This is a precise condition on a space-time $M$ that may be taken as a statement that $M$ accords with a form of cosmic censorship. Imposing such a condition, we deduce the existence of an absolute event horizon $E$, which encloses but does not intersect the singular region. This horizon $E$ may be defined as the boundary of the past of the external infinity.

(Penrose [153, p. 125])

The first part is tongue in cheek, since the “establishment” included—or even was Penrose himself! The ‘form of cosmic censorship’ probably refers to Hawking’s [92] theorem to the effect that any future trapped surface lies entirely within a black hole. Penrose [153] goes on to develop the Penrose inequality, which he saw as a test for cosmic censorship he expected to fail (which it has not, so far).

The challenge of making “the concept of “terminating on a singularity”” more precise was taken up in [154], see also [155], and led to Definition 2 above. A definition of cosmic censorship should then specify the nature of the points $x$ relative to which “singular” curves $c$ are naked. Hence it would have been natural if Penrose had proposed e.g. the following definition (cf. footnote 54):61

**Definition 3** Let $(M, g)$ be a strongly causal space-time that is asymptotically flat at null infinity.

1. An $N$-naked singularity in $M$ with respect to some region of exposure $N \subset \hat{M}$ is a future-inextendible causal curve $c$ in $M$ that is naked for some point $x \in N$, i.e., the pair $(c, x)$ satisfies $I^-(c) \subset I^-(x)$.
2. The $N$-weak cosmic censorship conjecture states that a “physically reasonable” space-time $(M, g)$ contains no $N$-naked singularities, i.e., no curve $c$ in $M$ and point $x \in N$ as specified in 1. exist.

Paraphrasing Geroch & Horowitz [79], p. 274, the set $N$ represents the region of space-time (or possibly of its conformal completion $\hat{M}$) in which observers could

---

58 A space-time $(M, g)$ is (future) asymptotically predictable from a partial Cauchy surface $S$ if $\mathcal{I}^+ \subset D^+(S)$, where the domain of dependence $D^+(S)$ is computed in the conformal completion $\hat{M}$ of $M$ [93, p. 310], cf. footnote 54. A violation of this condition would come from some null geodesic outside $D^+(S)$, presumably marking a singularity at least in the spirit of [148], reaching $\mathcal{I}^+$. This would makes such a singularity visible to asymptotic observers.

59 See also Proposition 9.2.1 in [93], corrected by [27].

60 See [7] and [128] for reviews of the Penrose inequality. See also [112], Sect. 10.11 for a brief summary. A special case, the Riemannian Penrose inequality, has been proved; see [117] for details.

61 Definition 3 is stated in words by Geroch & Horowitz [79], pp. 274–277. See also Earman [47], pp. 74–78.
detect that their universe is singular.’ This definition is quite subtle and excludes for example that the big bang or the big crunch are seen as naked singularities (and the ensuing cosmos as violating weak cosmic censorship): for the big bang, which we can ‘see’, no such curve \(c\) exists (any relevant \(c\) is past inextendible), whereas for the big crunch no \(x\) would exist. See Geroch & Horowitz [79], pp. 274–277 and Penrose [155], §12.3.2. 62

Various choices of the region \(N\) have (often implicitly) been proposed in the literature, for example:

- \(N = I^- (\mathcal{I}^+)\), i.e. the complement of the black hole region in \(M\) (my interpretation of [151]);
- \(N = I^- (\mathcal{I}^+) \cap I^+ (\mathcal{I}^-) \equiv \mathcal{D}(M, \mathcal{I})\), i.e. the domain of outer communication(s) in \(M\); 63
- \(N = \mathcal{D}(M, \mathcal{I}) \cup \mathcal{I}\) with \(\mathcal{I} = \mathcal{I}^+ \cup \mathcal{I}^- [157]\);
- \(N = J^- (\mathcal{I}^+) \cap J^+ (S)\), where \(S\) is some partial Cauchy surface in \(M\). 64
- \(N = J^- (\mathcal{I}^+) [187]\); 65 See also footnote 73.
- \(N = \mathcal{I}^+ [79]\); 66
- \(N = M [154, 155]\), see below (strong cosmic censorship); this is defined without any need for asymptotic flatness.

The first and last choices seem optimal in having a clean characterization via Theorems 4 and 5. First:

**Theorem 4** Let \((M, g)\) be a strongly causal space-time that is asymptotically flat at null infinity. Then \((M, g)\) contains no naked singularities with respect to \(I^- (\mathcal{I}^+)\) iff \(I^- (\mathcal{I}^+)\) is globally hyperbolic.

This is proved in the same way as Penrose’s Theorem 5 below, adding a case distinction \(x \in I^- (\mathcal{I}^+)\) and \(x \notin I^- (\mathcal{I}^+)\) that does not upset the result. If one accepts the choice \(N = I^- (\mathcal{I}^+)\), Theorem 4 might of course serve to define weak cosmic censorship as global hyperbolicity of \(I^- (\mathcal{I}^+)\). Except for \(N = M\) there seems to be no complete analogue of Theorem 4 for the other choices of \(N\); at best, there is a one-sided implication to global hyperbolicity of \(N\). In any case, I side with Earman (private communication) that one may regard weak cosmic censorship as a family of conjectures, one for each relevant region \(N\).

---

62 In Penrose [155], p. 618, a ‘physically reasonable’ space-time is taken to mean: ‘A system which evolves, according to classical general relativity with reasonable equation of state, from generic non-singular initial data on a suitable Cauchy surface.’ But this is stated in connection with weak cosmic censorship. In strong cosmic censorship, the above formulation is problematic since it suggests that the ‘evolved’ space-times in question are globally hyperbolic, which by Theorem 5 makes strong cosmic censorship trivially true. I return to this problem in Sect. 6 below. In [154] Penrose does not state genericity conditions.

63 See e.g. Chruściel [22], Sect. 3.1, for this notion. Modern black hole uniqueness theorems like Theorem 3.1 in [25] assume weak cosmic censorship defined as global hyperbolicity of \(I^- (\mathcal{I}^+) \cap I^+ (\mathcal{I}^-)\).

64 This comes tantalizingly close to Hawking’s future asymptotically predictability but is not equivalent to it, see [108], especially Lemma 2.10. See also Wald [189], Sect. 12.1 and Chruściel [22], Sect. 3.5.1.

65 In fact, Tipler, Clarke & Ellis [187], p. 176, define weak cosmic censorship as global hyperbolicity of \(J^- (\mathcal{I}^+)\).

66 Their definition of (weak) cosmic censorship is that the closure of \(N\) (defined in \(M\)) in \(\hat{M}\) does not intersect \(\mathcal{I}^+\) (p. 280).
But Definition 3 is ahistorical! Instead, Penrose [154, 155] immediately moved on, arguing that:

It seems to me to be comparatively unimportant whether the observer himself can escape to infinity. Classical general relativity is a scale-invariant theory, so if locally naked singularities occur on a very tiny scale, they should also, in principle, occur on a very large scale in which a ‘trapped’ observer could have days or even years to ponder upon the implications of the uncertainties introduced by the observations of such a singularity. (…) Indeed, for inhabitants of recollapsing closed universes (as possibly we ourselves are) there is no ‘infinity’, so the question of being locally ‘trapped’ is one of degree rather than principle. It would seem, therefore, that if cosmic censorship is a principle of Nature, it should be formulated in such a way as to preclude such \textit{locally} naked singularities (Penrose [155, p.619]).

Thus ‘locally naked singularities’ are defined via Definition 3 (assuming only strong causality) by choosing $N = M$. Then the key result (which \textit{a posteriori} justifies his use of curves instead of geodesics) is due to Penrose [154, 155]:

\textbf{Theorem 5} A strongly causal space-time has no \textit{locally} naked singularities iff it is globally hyperbolic.

\textit{Proof:} I first prove the implication “global hyperbolicity $\Rightarrow$ no locally naked singularities” by contradiction. Suppose that $(M, g)$ is globally hyperbolic and that (4.1) holds for some $c$ and $x$. Take $y$ on $c$ and then take a future-directed sequence $(y_n)$ of points on $c$, with $y_0 = y$. Because of (4.1) this sequence lies in $J^+(y) \cap J^-(x)$, which is compact by assumption. Hence $(y_n)$ has a limit point $z$ in $J^+(y) \cap J^-(x)$.

Now define curves $(\gamma_n)$ as the segments of $c$ from $y$ to $y_n$. By Theorem 2.53 in [131], case (i), which is a sharp version of the limit curve lemma of causal theory going back to [152], these curves have a uniform limit $\gamma$. Its arc length (as measured by an auxiliary complete Riemannian metric) is, on the one hand, infinite (since $c$ is endless and hence has infinite arc length, which is approached as the $y_n$ move up along $c$). But on the other hand it is finite, since $\gamma$ ends at $z$ (and fd continuous causal curves have finite arc length iff they have an endpoint). Hence (4.1) cannot be true for any $c$ and $x$.

The (contrapositive) proof of the converse implication relies on the following lemma:

\textbf{Lemma 6} Let $(M, g)$ be a space-time, let $S \subset M$ be closed and achronal, and let $x$, $y \in M$.

1. If $y \in \text{int}(D^-(S))$, then $J^+(y) \cap J^-(S)$ is compact. In particular, taking $S = \partial I^-(x)$ and assuming $y \in I^-(x)$, it follows that $J^+(y) \cap J^-(x)$ is compact.

\textit{Proof:} (i) do not endorse this: ‘One would not, however, wish to regard [Kerr space-time with $a^2 < m^2$] as a counterexample to cosmic censorship, for it is only the “nearby” observers who can detect the singular behavior, not the “distant” observers.’ (p. 277). Slowly rotating Kerr space-time is not globally hyperbolic but it does have an event horizon.

\textit{This proof is taken from [112], Sect. 10.4. Penrose [155] gives a different argument using TIPs.}

\textit{Here I do follow Penrose [155], p. 624. The lemma combines Propositions 5.20 and 5.5 (h) in [152].}
2. We have \( \text{int}(D^-(S)) = I^-(S) \cap J^+(D^-(S)) \).

To prove the converse direction of Theorem 5, assume that \((M, g)\) is not globally hyperbolic. Then, under the assumption of strong causality, there are \(x, y\) for which \(J^-(x) \cap J^+(y)\) is not compact.\(^{70}\) We may assume that \(y \in I^-(x)\). Part 1 of Lemma 6 gives \(y \not\in \text{int}(D^-(\partial I^-(x)))\). Part 2 gives some \(y' \in I^-(x)\) with \(y' \not\in D^-(\partial I^-(x))\), so that, by definition of \(D^-\), there exists some \(f_d\) future-inextendible curve \(c\) from \(y'\) that avoids \(\partial I^-(x)\). Since \(y' \in I^-(x)\), this curve does lie in \(I^-(x)\), and hence (4.1) holds.\(^{71}\)

**Definition 7** The *strong cosmic censorship conjecture* states that no “physically reasonable” space-time \((M, g)\) contains locally naked singularities, i.e. pairs \((c, x)\) as in Definition 2 such that \(I^-(c) \subset I^-(x)\). Equivalently, by Theorem 5 the conjecture states that any “physically reasonable” space-time is globally hyperbolic.

In this formulation (unlike many others), strong cosmic censorship (as defined by the last bullet above) implies weak cosmic censorship (as defined by the first bullet).\(^{72}\)

Under the assumptions of Penrose's singularity theorem one can prove an analogue of Hawking's \([92]\) theorem on the invisibility of trapped surfaces (cf. \([93, \text{Proposition 9.2.1}]\)).

**Theorem 8** Under the assumptions of Theorem 1 and the additional clause that \((M, g)\) be asymptotically flat at null infinity, any future trapped surface must lie entirely within the “black region” \(M \setminus I^-(\mathcal{I}^+)\).

I just sketch the proof, as the details are similar to Hawking's. If the trapped surface \(T^2\) were to (partly) lie in \(I^-(\mathcal{I}^+)\), then also part of \(\partial I^+(T^2)\) lies in \(I^-(\mathcal{I}^+)\). Hence some of the lightlike geodesics \(\gamma\) ruling \(\partial I^+(T^2)\) with past endpoint on \(T^2\) would reach \(\mathcal{I}^+\) and hence have infinite length. But the definition of a trapped surface excludes this, as in the proof of Theorem 1, since it forces each \(\gamma\) to be incomplete. \(\square\)

Since this forces \(M \setminus I^-(\mathcal{I}^+)\) to be non-empty, there must be an event horizon \(\partial I^-(\mathcal{I}^+)\), which by the same argument must cover not only any trapped surface but also all ensuing incomplete null geodesics.

### 6 Cosmic censorship: from Penrose to PDE

Theorem 8 bridges part of the gap between Penrose’s 1965 singularity theorem and its intended application to black holes, in that in the asymptotically flat setting an event horizon comes for free. However, according to this analysis (whose centerpiece is Penrose’s Theorem 5), strong cosmic censorship (à la Penrose, cf. Definition 7) is

---

\(^{70}\) Unless \(M\) is compact this is true even without the assumption of strong causality, see [98].

\(^{71}\) All this can be checked in the Minkowskian example following Definition 2, where, assuming \(z \in I^+(y)\), the removal of \(z\) ruins compactness of \(J^+(y) \cap J^-(x)\) and hence global hyperbolicity. The existence of \(c\) is trivial.

\(^{72}\) This follows from the definitions, but it also follows from Theorems 4 and 5 and the observation that if \((M, g)\) is globally hyperbolic, then so is \(I^-(\mathcal{I}^+)\): if \(x \in J^+(y)\) for \(x, y \in I^-(\mathcal{I}^+) \subset M\), then \(J^+(y) \cap J^-(x) \subset I^-(\mathcal{I}^+)\).
in fact one of the assumptions of the singularity theorem, so that its implication of weak cosmic censorship (once again in the sense of Penrose, i.e. defined by taking $N = I^-(\scri^+)$ in Definition 3) is hardly surprising. And we are also left with the issue of the (in)extendibility of space-time.

With hindsight, it would have been natural if, one the basis of [148], Penrose had defined:

- **Weak cosmic censorship** as the existence and stability of event horizons (in case of singularities);
- **Strong cosmic censorship** as the nonexistence or instability of Cauchy horizons (in all cases).

This would have directly addressed the above gap. In the first (weak) case this is indeed what he did, but he subsequently replaced this by the version of strong cosmic censorship that really interested him. However, via a detour explained below, the latter eventually did lead to the second bullet point.

The current versions of both cosmic censorship conjectures are based on the initial-value approach or PDE approach to GR, whose starting point may be taken to be the celebrated theorem of Choquet-Bruhat and Geroch [16]. This theorem states that for each initial data triple $(S, h, K)$ satisfying the vacuum constraints, where $(S, h)$ is a complete 3d Riemannian manifold carrying an additional symmetric 2-tensor $K$, there exists a maximal globally hyperbolic development or MGHD $(\mathcal{M}, g, i)$, where $(\mathcal{M}, g)$ is a space-time satisfying the (vacuum) Einstein equations and $i : S \hookrightarrow \mathcal{M}$ is an embedding such that $g$ returns the initial data $(h, K)$ on $i(S)$, in that $i^*g = h$ and $K$ is the extrinsic curvature of $i(S) \subset \mathcal{M}$. The space-time $(\mathcal{M}, g)$ contains $i(S)$ as a Cauchy surface and hence is globally hyperbolic (as the name MGHD suggests).

Moreover, the triple $(M, g, i)$ is unique up to (unique) time-orientation-preserving isometries fixing the Cauchy surface, i.e. for any other MGHD $(\mathcal{M}', g', i')$ there exists

---

73 Compare Tipler, Clarke, & Ellis [187], p. 167: ‘The strong cosmic censorship principle is said to hold if the *entire* space-time $(\mathcal{M}, g)$ is globally hyperbolic, i.e., if there are no nontrivial Cauchy horizons, and the weak cosmic censorship principle is said to hold if all breakdowns of global hyperbolicity occur inside a black hole, i.e. if the region $J^-(\scri^+)$ is globally hyperbolic.’

74 Penrose [156] still states that the inference from geodesic incompleteness to a black hole requires weak cosmic censorship in a version that looks like a form of Definition 3 with either $N = I^- (\scri^+)$ or $N = \scri^+$. But, as in [155], he then moves on to strong cosmic censorship (in his version, of course) as the version one really wants. Note that (his) strong cosmic censorship does not imply the $N = \scri^+$-weak version! In our discussion on July 2, 2022, Penrose said that for him the essence of strong cosmic censorship had always been the avoidance of timelike singularities, and that (in)extendibility was secondary.

75 The original source is [16], who merely sketched a proof (based on Zorn’s lemma). Even the 800-page textbook by Choquet-Bruhat [13] does not contain a proof of the theorem (which is Theorem XII.12.2); the treatment in Hawking & Ellis [93], Sect. 7.6, is slightly more detailed but far from complete, too. [162] is a book-length exposition, but ironically the proof of Theorem 16.6 is wrong; it is corrected in [https://people.kth.se/~hansr/mghd.pdf](https://people.kth.se/~hansr/mghd.pdf). A constructive proof was given by Sbierski [169]. For a summary see [112], Sect. 7.6, which also contains further information.

76 Though rarely if ever mentioned in statements of the theorem, the isometry $\psi$ is unique [116]. This can be shown by Proposition 3.62 in [139] or the equivalent argument in footnote 639 of [112], to the effect that an isometry $\psi$ is determined at least locally (i.e. in a convex nbhd of $x$) by its tangent map $\psi_\xi'$ at some fixed $x \in M'$. Take $x \in i'(S)$. Since $\psi$ is fixed all along $i'(S)$ by the condition $\psi \circ i' = i$ and since it also fixes the (future-directed) normal $N_\xi$ to $i'(S)$ by the condition $\psi^*g = h$, it is determined locally. Theorem 1 in [88] then applies, which is a rigidity theorem for isometries going back at least to [75], Appendix A.
an isometry $\psi : M' \to M$ that preserves time orientation and satisfies $\psi \circ i' = i$. This theorem has given rise to an ideology about GR in which:

- All valid assumptions about GR are assumptions about initial data $(S, h, K)$.
- All valid questions about GR are questions about “the” MGHD $(M, g, i)$ of these data.

This PDE-based program sometimes gives a different perspective from the Penrose–Hawking–Geroch mathematical approach to GR originating in the 1960s, in which typically larger (usually maximally extended) space-times are studied. In particular, in the PDE approach one would at first sight ask if some MGHD $(M, g, i)$ obtained from “physically reasonable” initial data satisfies strong cosmic censorship. By Theorem 5 this comes down to the question if $(M, g)$ is globally hyperbolic. But this is true by construction! Hence Penrose’s clear and distinct formulation of strong cosmic censorship seems problematic in a paradoxical way:

1. Global hyperbolicity of the MGHD $(M, g, i)$ of any initial data set $(S, h, K)$ is automatic.
2. Global hyperbolicity of maximally extended (black hole) space-times is often too strong.77

In what they call the ‘evolutionary approach’ to cosmic censorship, Geroch & Horowitz [79], Sect. 5.4, show a way out and en passant sketch the modern PDE versions of both the weak and strong case. As we shall see, their version of strong cosmic censorship can actually be seen as a special case of Penrose’s.

Their proposal for weak cosmic censorship is based on their earlier refinement of Penrose’s notion of asymptotic flatness at null infinity [78]. Using the notation of footnote 54, assuming that the scaling function $\Omega$ satisfies $\Delta \Omega = 0$, as can always be achieved by a redefinition of $\Omega$, the extra condition is that the null geodesics ruling the null hypersurface $\mathcal{I}^{\pm}$ are complete.78

Consider now the following statement: for any asymptotically flat initial-data set, topologically $\mathbb{R}^3$, its maximal evolution [i.e. MGHD] is an asymptotically flat spacetime [at null infinity] (i.e. it even satisfies completeness of null infinity). This statement, we claim, captures a sense of cosmic censorship. That the initial-data set be topologically $\mathbb{R}^3$ ensures that the evolution is not singular already on $S$; that $S$ be asymptotically flat ensures that one deals with isolated systems, and in particular that any singular behavior of the evolution must be due to the system itself and not external influences. Suppose, then, that this maximal evolution were singular, say to the future of $S$. It certainly cannot be nakedly singular for the future of $S$, for $S$ must be a Cauchy surface for its evolution. Furthermore, the statement asserts that this evolution must be sufficiently large that it includes the entire asymptotic regime, so in particular asymptotic observers can live out their entire lives within this maximal evolution. What this statement means,

77 For example, maximally extended Kerr space-time is not even globally hyperbolic in the physically relevant case $0 < a^2 < m^2$.
78 Without this, the Penrose–Hawking definition of a black hole as a connected component of $M \setminus (\mathcal{I}^+)$, with absolute event horizon $\partial I^- (\mathcal{I}^+)$, would be problematic. See e.g. footnote 57. At least, examples like this are avoided now.
then, is that asymptotic observers will forever be unaffected by any singular behavior of the spacetime. But this is a version of cosmic censorship. We have essentially just re-expressed our earlier statement (that the closure of $N$ have empty intersection with $\mathcal{I}^+$), but now in a way that avoids the counterexample obtained by removing a point from Minkowski spacetime. (Geroch & Horowitz [79, pp. 285–286])

In heuristic terms this means that, if we disregard exceptional initial conditions, no singularities are observed from infinity, even though observations from infinity are allowed to continue indefinitely. (Christodoulou [18, p. A26])

This statement of weak cosmic censorship is exemplary from a pde point of view: it starts with an assumption on initial data and formulates a question about the ensuing MGHD. However, like the definition of a black hole it depends on the concept of null infinity, which is “prescient” and may be seen as an undesirable idealization (cf. [33], and footnote 55). Thus [18] reformulated the above statement of weak cosmic censorship in such a way that the idealization $\mathcal{I}^+$ no longer occurs. 79

Either way, this turns the Penrosian version of weak cosmic censorship on its head! For whereas his version states that outgoing signals from a black hole singularity are blocked by a (future) event horizon $H_+(\mathcal{I}^0)$, the new version is about incoming (null) signals: the further these are away from $H_+(\mathcal{I}^0)$, the longer it takes them to enter $H_+(\mathcal{I}^0)$, and in the limit (i.e., in the original formulation, at null infinity) this takes infinitely long, making future null infinity $\mathcal{I}^+$ complete. Yet there is a connection with Penrose: lack of global hyperbolicity of $I^-(\mathcal{I}^+)$ gives a partial Cauchy surface in $M$ a Cauchy horizon which cuts off $\mathcal{I}^+$. This is clear from simple examples like Schwarzschild for $m < 0$, Reissner–Nordström at $e^2 > m^2$, or Kerr at $a^2 > m^2$; see e.g. [42] and [112], Sect. 10.6.

The status of the weak cosmic censorship conjecture(s) in any reasonable form is open.80 One especially problematic issue is the status of the genericity conditions that define “physically reasonable” space-times (or initial conditions) and hence the range of applicability of the conjecture(s). For example, the physical relevance of genericity conditions typically used in the mathematical literature, which are evidently based on pde techniques, has been questioned [87, Sect. 3.4].

The current (pde) version of strong cosmic censorship also goes back to [79]:

Consider, for motivation, an initial-data set whose maximal evolution is extendible to the future of $S$ (...) This extended spacetime cannot, by definition of the maximal evolution, have $S$ as a Cauchy surface. That is, from a point $p$ in the extension there must exist a maximally extended past-directed time-

---

79 Let $(S, h, K)$ be asymptotically flat initial data for the Einstein equations (satisfying the constraints), with $\text{MGHD}(M, g, i)$. Christodoulou then defines $(M, g)$ to have complete future null infinity if for any $s > 0$ there exists a region $B_0 \subset B \subset S$ such that $\partial D^+(B)$, which is ruled by null geodesics, has the property that each null geodesic starting in $\partial J^+(B_0) \cap \partial D^+(B)$ can be future extended beyond parameter value $s$. Here $D^+(B)$ is the future domain of dependence of $B$, and each null geodesic in question is supposed to have tangent vector $L = T - N$, where $T$ is the fd unit normal to $S$ in $M$ and $N$ is the outward unit normal to $\partial B$ in $S$.

80 For reviews see [5, 19, 20, 37, 103, 109, 140]. N.B. Black hole uniqueness theorems [25] rely on weak cosmic censorship!
like curve which cannot be assigned a past endpoint, and which fails to meet \( S \). In this rather mild sense the extended spacetime must be nakedly singular. One might therefore imagine formulating cosmic censorship as the assertion that every maximal evolution [i.e. \( \text{MGHD} \)] is inextendible, i.e. that, once the maximal evolution is completed, it is not possible to add any ‘extra regions’ as vantage points from which observers could detect that their spacetime is singular to the future of \( S \).

(Geroch & Horowitz [79, pp. 286–287])

Interestingly, they immediately refine this suggestion by giving the example of a small spacelike disk \( S \) in Minkowski spacetime \( \mathbb{M} \), whose MGHD is its domain of dependence \( D(S) \) (which is a double cone), which is clearly extendible to all of \( \mathbb{M} \). To avoid cases like that, they end up with the following:

**Conjecture 5.2.** For \( p \) any point in any extension of the maximal evolution of any non-compact initial data set \( S \), \( I^- (p) \cap S \) has non-compact closure. [79, pp. 288]

But this refinement seems not to have survived, perhaps because the models on which the conjecture is tested already exclude trivial (counter)examples like the one mentioned. Thus we now simple have:

The appropriate notion of cosmic censorship (…) is that the generic solution to Einstein’s equations is globally hyperbolic, i.e., that the maximal Cauchy development of a generic initial data set is inextendible. (Moncrief [134, p. 88])

The strong cosmic censorship conjecture says that ‘most’ spacetimes developed as solutions of Einstein’s equations from prescribed initial data cannot be extended outside of their maximal domains of dependence. (Chruściel, Isenberg, & Moncrief [24, p. 1671])

**Conjecture 17.1** (Strong Cosmic Censorship). For generic initial data for Einstein’s equations, the MGHD is inextendible. (Ringström [162, p. 188])

**Conjecture 3.5** (Strong cosmic censorship). For generic asymptotically flat vacuum data sets, the maximal Cauchy development \((M, g)\) is inextendible as a suitably regular Lorentzian manifold. (Dafermos [40, p. 11])

If the extension satisfies the Einstein equations this is a special case of Penrose’s formulation [155], applied to a judicious choice of space-time! To see this, recall the following concept [21]:

**Definition 9** A development of initial data \((S, h, K)\) satisfying the vacuum constraints is a triple \((M, g, i)\), where \((M, g)\) is a space-time solving the vacuum Einstein equations and \( i : S \to M \) is a spacelike embedding such that \( i^* g = h \) and \( i(S) \) has extrinsic curvature \( K \) in \( M \). Such a development is called maximal if it has no extension \((M', g')\) that also satisfies the vacuum Einstein equations.\(^{82}\)

\(^{81}\) In §3 they further specify ‘most’ in terms of open and dense subsets in the space of initial data

\(^{82}\) Thus the difference between a development and a Cauchy development is that in the former \( i(S) \) is no longer required to be Cauchy surface in \( M \), so that \((M, g)\) is not necessarily globally hyperbolic. Chruściel
Now apply Penrose’s strong cosmic censorship to such a maximal development, i.e. require it to be globally hyperbolic (cf. Theorem 5). The connection with inextendibility is then easily made:

**Proposition 10** A maximal development of initial data \((S, h, K)\) is globally hyperbolic (with Cauchy surface \(i(S)\)) if and only if the MGHD of \((S, h, K)\) is inextendible as a solution to the vacuum Einstein equations. In that case, “the” maximal development coincides with “the” MGHD (up to isomorphism).

**Proof** The set of isometry classes \([M, g, i]\) of Cauchy developments \((M, g, i)\) of given initial data \((S, h, K)\) is partially ordered, and the MGHD \([M_t, g_t, i_t]\) is its top element. Hence if some maximal development \((M_m, g_m, i_m)\) is globally hyperbolic, then \([M_m, g_m, i_m] \leq [M_t, g_t, i_t]\). On the other hand, since \((M_t, g_t, i_t)\) is a solution and \((M_m, g_m, i_m)\) is maximal also the converse holds, so \((M_m, g_m, i_m) \cong (M_t, g_t, i_t)\).  

Therefore, Penrose’s conjecture applied to the maximal development à la [21] states that for “generic” initial data the MGHD is inextendible as a solution to the vacuum Einstein equations. The tension that Penrose’s formulation may have given rise to reflected the ambiguity about the space-time his condition of global hyperbolicity is supposed to apply to: it seems that in the 1960s and 1970s people had maximally extended solutions in mind, see e.g. the Penrose diagrams in [93], whereas from the 1990s onward the MGHD in the PDE picture had become the space-time of choice.

However, the strong cosmic censorship conjecture as stated by all authors quoted above is stronger than the one suggested by Proposition 10, in that even extensions that do not satisfy the Einstein equations are excluded. Indeed, any space-time solves these equations for some energy-momentum tensor, however unfamiliar, and one has no idea what kind of (strange) matter is contained in an extension. But the Einstein equations should at least be defined: to this end one may either consider extensions in which the metric is \(C^2\) (i.e. the borderline case where Einstein equations make sense as classical PDEs), or allow \(C^0\) metrics for which the associated Christoffel symbols are locally \(L^2\) (which is the least regular case in which the metric can still be defined as a weak solution to Einstein’s equations).

[21] proves existence of maximal developments (but not uniqueness up to isometry, as in the globally hyperbolic case).

83 Note that although each extendible space-time has an inextendible extension [77], which may be taken as “maximal”, this inextendible extension is not unique (except perhaps if one imposes unphysical conditions like analyticity).

84 [24] and [23] consider smooth extensions.

85 See [81], [20], p. 9, and [124], footnote 1. Minkowski space-time turns out to be inextendible even in \(C^0\) [170], so in that case the validity of strong cosmic censorship is independent of the regularity of the extension. However, for two-ended asymptotically flat data for the spherically symmetric Einstein–Maxwell-scalar field system (to which the conjecture, so far discussed for the vacuum case, can be extended in the obvious way), the conjecture fails in \(C^0\), i.e. the MGHD is extendible with a \(C^0\) metric, but it holds in \(C^1\), in that the metric of the extension fails to be \(C^1\) [35, 36]. The situation for the Kerr metric is similar [41].
Sensitivity to the precise formulation of genericity conditions (which from a PDE point of view define which initial data are deemed “physically reasonable” and from a traditional GR perspective do so for space-times) is another issue; already the Kerr metric, in which strong cosmic censorship fails for all parameter values (as long as $a \neq 0$ and $m \neq 0$), shows that genericity conditions are necessary. Such counterexamples made it especially courageous of Penrose to conjecture strong cosmic censorship; but of course he had good arguments. His key observation, indeed one of his most prophetic insights, was first published in 1968 (i.e. before even weak cosmic censorship had been formulated by him):

There is a further difficulty confronting our observer who tries to cross [the Cauchy horizon] $H_+(\mathcal{H})$. As he looks out at the universe he is “leaving behind,” he sees, in one final flash, as he crosses $H_+(\mathcal{H})$, the entire later history of the rest of his “old universe.” If, for example, an unlimited amount of matter eventually falls into the star then presumably he will be confronted with an infinite density of matter along “$H_+(\mathcal{H})$”. Even if only a finite amount of matter falls in, it may not be possible in generic situations to avoid a curvature singularity in place of $H_+(\mathcal{H})$. This is at present an open question. But it may be, that the place to look for curvature singularities is in this region rather than (or as well as?) at the “center.”

(Penrose [150, p. 222])

Here one should realize that for Penrose $H_+(\mathcal{H})$ is a (future) Cauchy horizon (relative to some partial Cauchy surface $S$) in some “large” (typically maximally extended) space-time, whereas from a PDE point of view it is the boundary of the MGHD of the corresponding initial data on $S$, if it has one; that is, if it is extendible (in a suitable regularity class). Either way, the conjectured “blueshift instability” of $H_+(\mathcal{H})$ has been confirmed in a large number of studies and remains the key to strong cosmic censorship.

Failure of strong cosmic censorship is often taken to imply a failure of determinism in GR. The idea is that the (classical) world—including the (physical degrees of freedom of the) gravitational field itself—is governed by hyperbolic partial differential equations whose initial data should be given on a hypersurface $S$ and whose solutions should thereby be determined on its domain of dependence $D(S)$ [29]. From Penrose’s point of view, if a (typically maximally extended) space-time $(M', g')$ fails to be

---

86 Papers that state such conditions include [35], [162], Sect. 17.2.3 [163], Sect. 11 and [125].

87 Further to [178], see [130] down to [11] for heuristics. Mathematically rigorous work started with Defermos [35] see also [38, 39]; more recent papers may be traced back from [104]. The conclusion seems to be that at least in asymptotically flat cases Cauchy horizons turn into weak null singularities, which are null boundaries with $C^0$ metric but $\Gamma^\mu_{\nu\lambda}$ not locally $L^2$ [124]. At least for one-ended asymptotically flat initial data, behind such a weak null singularity there is also a strong curvature singularity at $r = 0$. See also [125] for the two-ended case, and [69] for extremal Reissner–Nordström black holes (i.e. $0 < m = |e|$). For cosmological constant $\lambda \neq 0$ the situation changes, e.g. in de Sitter space ($\lambda > 0$) the verdict on strong cosmic censorship depends critically on both the matter coupling and the regularity of the extension [44], whilst in anti-de Sitter ($\lambda < 0$) even weak cosmic censorship fails [30]. The so-called cosmological case, which is characterized by taking initial data on compact spacelike slices and indeed includes (toy) examples relevant to cosmology, was first studied by Misner and Taub [133], who noted the instability of the Cauchy horizon in Taub–NUT space (this work actually predated Misner & Taub, see footnote 23). See also [167] and [28].

88 See also [46, 47, 126, 162, 182].
globally hyperbolic, then, taking some partial Cauchy surface $S'$, neither the part $M' \setminus D(S') \neq \emptyset$ of space-time behind the Cauchy horizon of $S'$ itself, nor things happening within space-time behind this horizon, are determined by initial data on $S'$. From the PDE point of view, although any MGH (M, g) is globally hyperbolic with Cauchy surface $S \subset M$, if $(M, g)$ is extendible with extension $(M', g')$, then by definition $S$ fails to be a Cauchy surface for $M' \supset M$ and we are back to Penrose.

If the extension $(M', g')$ satisfies the Einstein equations and is globally hyperbolic, which is logically possible, then the lack of determinism is merely apparent: although $(M', g')$ is not determined by the initial data on $S$ originally expected to do so, it is determined by new initial data on some Cauchy surface $S' \subset M'$, which may be accessible to physicists in $M'$. But in interesting examples the MGH(M, g) typically has extensions $(M', g')$ that fail to be either globally hyperbolic or unique (up to isometry), or both. In those cases the ensuing violation of determinism seems real (though hard to interpret).

In any case, although extendibility of the MGH does seem to give rise to some form of indeterminism in GR, so that (contrapositively) determinism necessitates strong cosmic censorship in the PDE sense of inextendibility of the MGH, the latter is not sufficient for determinism. The indeterminism of someone falling into a black hole singularity is perfectly well compatible with global hyperbolicity, as the Schwarzschild solution shows. More generally, in classical (mathematical) physics indeterminism may come from either a lack of uniqueness of solutions or from a lack of existence thereof; the latter includes incompleteness of solutions, i.e. non-existence after (or before) some finite time. Strong cosmic censorship secures uniqueness of solutions of the Einstein equations (whose existence is given by PDE theory à la Choquet-Bruhat), but, as we have seen, it fails to imply existence (i.e. for all times) of solutions of the geodesic equation [46]. Hence at the classical level strong cosmic censorship does not imply determinism. But ironically, determinism might be rescued by quantum mechanics: because of the unitarity of their time-evolution quantum objects cannot disappear into a singularity, and hence they form standing waves. These only become sources of indeterminism if they undergo measurement.

### 7 Epilogue

Prior to the 1960s spacetime singularities were regarded as a minor embarrassment for GTR. They constituted an embarrassment because it was thought by Einstein and others that a true singularity, a singularity in the fabric of space-time itself, was an absurdity. But the embarrassment was a minor one that could be swept under the rug; for the then known models of GTR containing singularities all embodied very special and physically unrealistic features. Two developments

---

89 This happens e.g. in Taub–NUT and Gowdy cosmologies. See e.g. [23, 24, 45, 134, 162, 163, 187].
90 In non-relativistic mechanics bodies may disappear to infinity in finite time [168], and hence, by the same (time-reversed) token, may appear from nowhere in finite time and hence influence affairs in a way unforeseeable from any Cauchy surface. This analogy with GR, e.g. with white holes, is discussed by Earman [50], Sect. 3.6.
91 See [111, 114, 115] for the final proof of randomness of measurement outcomes in quantum mechanics.
forced a major shift in attitude. First, the observation of the cosmic low tempera-
ture blackbody radiation lent credence to the notion that our universe originated
in a big bang singularity.\textsuperscript{92} Second, and even more importantly, a series of the-
orems due principally to Stephen Hawking and Roger Penrose indicated that,
according to GTR, singularities cannot be relegated to the distant past because
under quite general conditions they can be expected to occur both in cosmology
and in the gravitational collapse of stars. Thus, singularities cannot be swept
under the rug; they are, so to speak, woven into the pattern of the rug. Of course,
these theorems might have been taken as turning what was initially only a minor
embarrassment into a major scandal. Instead, what occurred in some quarters
was a 180° reorientation in point of view: singularities were no longer relegated
to obscurity; rather they were to be recognized as a central feature of the GTR,
a feature which called attention to a new aspect of reality that was neglected in
all previous physical theories, Newtonian and special relativistic alike.

\textsuperscript{92} It might be appropriate to point out that it was rather the agreement between: (i) the \textit{interpretation} of this
observation as coming form a possible big bang [102, Part IV]; (ii) Friedman-type cosmological models
in GR [60]; (iii) the \textit{interpretation} of redshifts of galaxies as indicating an expanding universe [138]; (iv)
Gamow’s earlier calculations of the abundance of light elements [106, Chapter 4]; (v) Hawking’s singularity
theorem (which was only understood by relativists), which led to the consensus about the big bang from
1966 onwards.

\textsuperscript{123}
about, namely black holes) Penrose’s theorem would not have been able to launch a revolution. But what the relevant scientific community believed about the theorem, culminating in his 2020 Nobel Prize, carried the day.

Finally, in the context of the history of mathematics, Grattan-Guinness [85] calls attention to the historical actors’ knowledge or ignorance of their ignorance. Restricting ourselves to the two main actors:

Did Einstein [55] know that he did not know what a singular space-time was? The evidence is mixed (see Sect. 2): he combined egregious mistakes (in which he was not alone) with stunning foresight.

Was Penrose [148] aware of the gap between his theorem and its intended application to black holes? Here the evidence is clear: he was, and he subsequently did everything he could to overcome it.

Data availability The data for this article consist of the reasoning and the references. All references are easily found either on the internet or, in case of books, in standard academic libraries. Readers who have any trouble finding a reference are welcome to contact the author for further help. The only exception of data not freely available is the letter from Einstein to Besso quoted in footnote 10. Those interested in seeing the entire letter should contact the Einstein archives via Hanoch Gutfreund (hanoch.gutfreund@mail.huji.ac.il) and/or Diana Buchwald (Diana_buchwald@caltech.edu).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Andersson, L., Eichmair, M., Metzger, J.: Jang’s equation and its applications to marginally trapped surfaces. Contemp. Math. 554, 13–46 (2011)
2. Aretakis, S.: Lecture Notes on General Relativity. https://web.math.princeton.edu/~aretakis/columbiaGR.pdf (2013)
3. Beem, J.K., Ehrlich, P.E., Easley, K.: Global Lorentzian Geometry, 2nd edition. M. Dekker (1996)
4. Belinski, V., Henneaux, M.: The Cosmological Singularity. Cambridge University Press (2018)
5. Bieri, L.: Black hole formation and stability: a mathematical investigation. Bull. AMS (N.S.) 55, 1–30 (2018)
6. Biezunski, M.: Einstein’s reception in Paris in 1922. In: The Comparative Reception of Relativity, ed. Glick, T.F., pp. 169–188. Springer (1987)
7. Bray, H.L., Chruściel, P.T.: The Penrose inequality. In: The Einstein Equations and the Large Scale Behavior of Gravitational Fields, eds. Friedrich, H., Chruściel, P.T., pp. 39–70. Springer (2004)
8. Budic, R., Isenberg, J., Lindblom, L., Yasskin, P.B.: On the determination of Cauchy surfaces from intrinsic properties. Commun. Math. Phys. 61, 87–95 (1978)
9. Calabi, E., Markus, L.: Relativistic space forms. Ann. Math. 75, 63–76 (1962)
10. Carballo-Rubio, R., Di Filippo, F., Liberati, S., Visser, M.: Geodesically complete black holes. Phys. Rev. D 101, 084047 (2020)
11. Chesler, P.M., Narayan, R., Curiel, E.: Singularities in Reissner-Nordström black holes. Class. Quant. Gravity 37, 025009 (2020)
12. Choquet-Bruhat, Y.: Hyperbolic partial differential equations on a manifold. In: Batelle Rencontres: 1967 Lectures in Mathematics and Physics, eds. C. DeWitt and J.A. Wheeler, pp. 84–106. W.A. Benjamin (1968)

13. Choquet-Bruhat, Y.: General Relativity and the Einstein Equations. Oxford University Press (2009)

14. Choquet-Bruhat, Y.: Beginnings of the Cauchy problem (2014). arXiv:1410.3490

15. Choquet-Bruhat, Y.: A Lady Mathematician in this Strange Universe: Memoirs. World Scientific (2018)

16. Choquet-Bruhat, Y., Geroch, R.: Global aspects of the Cauchy problem in general relativity. Commun. Math. Phys. 14, 329–335 (1969)

17. Christodoulou, D.: The formation of black holes and singularities in spherically symmetric gravitational collapse. Commun. Pure Appl. Math. 44, 339–373 (1991)

18. Christodoulou, D.: On the global initial value problem and the issue of singularities. Class. Quant. Gravity 16, A23–A35 (1999)

19. Christodoulou, D.: The instability of naked singularities in the gravitational collapse of a scalar field. Ann. Math. 149, 183–217 (1999)

20. Christodoulou, D.: The Formation of Black Holes in General Relativity EMS (2009)

21. Chruściel, P.T.: On uniqueness in the large of solutions of Einstein’s equations (“Strong Cosmic Censorship”). Mathematical aspects of classical field theory. Contemp. Math. 132, 235–274 (1992)

22. Chruściel, P.T.: Geometry of Black Holes. Oxford University Press (2020)

23. Chruściel, P.T., Isenberg, J.: Nonisometric vacuum extensions of vacuum maximal globally hyperbolic space-times. Phys. Rev. D 48, 1616–1628 (1993)

24. Chruściel, P.T., Isenberg, J., Moncrief, V.: Strong cosmic censorship in polarised Gowdy space-times. Class. Quantum Gravity 7, 1671–168 (1990)

25. Chruściel, P.T., Lopes Costa, J., Heusler, M.: Stationary black holes: uniqueness and beyond. Living Rev. Relat. 15 (7) (2012)

26. Clarke, C.J.S.: The Analysis of Space-Time Singularities. Cambridge University Press (1993)

27. Claudel, C.-M.: Black holes and closed trapped surfaces: a revision of a classic theorem. arXiv:gr-qc/0005031 (2000)

28. Collins, C.B., Ellis, G.F.R.: Singularities in Bianchi cosmologies. Phys. Rep. 56, 65–105 (1979)

29. Courant, R., Hilbert, D.: Methods of Mathematical Physics. Volume II: Partial Differential Equations (Wiley Interscience) (1962)

30. Crisford, T., Santos, J.E.: Violating the weak cosmic censorship conjecture in four-dimensional Anti-de Sitter space. Phys. Rev. Lett. 118, 181101 (2017)

31. Curiel, E.: The analysis of singular space-times. Philosophy of Science 66, S119–S145. Revised and extended version at http://strangebeautiful.com/phil-phys.html (1999)

32. Curiel, E.: Singularities and Black Holes. The Stanford Encyclopedia of Philosophy (Spring 2019). https://plato.stanford.edu/archives/spr2019/entries/space-time-singularities/ (2019)

33. Curiel, E.: The many definitions of a black hole. Nat. Astron. 3, 27–34 (2019)

34. Dadhich, N.: Indian contribution to the physics of black holes: 2020 Nobel Prize. Curr. Sci. 119, 2030–2033 (2020)

35. Dafermos, M.: Stability and instability of the Cauchy horizon for the spherically symmetric Einstein-Maxwell-scalar field equations. Ann. Math. 158, 875–928 (2003)

36. Dafermos, M.: The interior of charged black holes and the problem of uniqueness in general relativity. Commun. Pure Appl. Math. 58, 445–504 (2005)

37. Dafermos, M.: The formation of black holes in General Relativity [after D. Christodoulou]. Séminaire Bourbaki 64, no. 1051. https://www.dpmms.cam.ac.uk/~md384/expose-chr.pdf (2012)

38. Dafermos, M.: The geometry and analysis of black hole space-times in general relativity (ETH Nachdiplom lectures). https://www.dpmms.cam.ac.uk/~md384/Ravello_Lectures_1.pdf (2013)

39. Dafermos, M.: Black holes without spacelike singularities. Commun. Math. Phys. 332, 729–757 (2014)

40. Dafermos, M.: The mathematical analysis of black holes in general relativity. Proceedings of the ICM, 2014, https://www.dpmms.cam.ac.uk/~md384/ICMArticleMihalis.pdf (2014)

41. Dafermos, M., Luk, J.: The interior of dynamical vacuum black holes I: The C^0-stability of the Kerr Cauchy horizon. arXiv:1710.01722 (2017)

42. Dafermos, M., Rodnianski, I.: Lectures on black holes and linear waves. arXiv:0811.0354 (2008)
43. Datt, B.: Über eine Klasse von Lösungen der Gravitationsgleichungen der Relativität. Z. Phys. 108, 314–321 (1938). English translation: On a class of solutions of the gravitation equations of relativity. Gen. Relativ. Gravit. 31, 1619–1627 (1999)
44. Dias, O.J.C., Reall, H.S., Santos, J.E.: Strong cosmic censorship: taking the rough with the smooth. J. High Energy Phys. 2018, 1 (2018)
45. Doboszewski, J.: Non-uniquely extendible maximal globally hyperbolic space-times in classical general relativity: a philosophical survey. Eur. Stud. Philos. Sci. 6, 193–212 (2017)
46. Doboszewski, J.: Relativistic space-times and definitions of determinism. Eur. J. Philos. Sci. 9, 24 (2019)
47. Earman, J.: Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes. Oxford University Press (1995)
48. Earman, J.: Tolerance for space-time singularities. Found. Phys. 26, 623–640 (1996)
49. Earman, J.: The Penrose–Hawking singularity theorems: History and implications. In: The Expanding Worlds of General Relativity. Einstein Studies, vol. 7, eds. Goenner, H., Renn, J., Ritter, T., Sauer, T., pp. 236–267. Birkhäuser (1999)
50. Earman, J.: Aspects of determinism in modern physics. In: Handbook of the Philosophy of Science. Vol. 2: Philosophy of Physics, Part B, eds. Butterfield, J., Earman, J., pp. 1369–1434. North-Holland/Elsevier (2007)
51. Earman, J., Eisenstaedt, J.: Einstein and Singularities. Stud. Hist. Philos. Mod. Phys. 30, 185–235 (1999)
52. Ehlers, J.: Beiträge zur relativistischen Mechanik kontinuierlicher Medien. Abh. Akad. Wiss. Lit. Mainz Math.-Naturwiss. Kl. 11, 792–837 (1961). English translation: Contributions to the relativistic mechanics of continuous media. Gen. Relativ. Gravit. 25, 1225–1266 (1993)
53. Ehlers, J.: A.K. Raychaudhuri and his equation. Int. J. Mod. Phys. D 15, 1573–1580 (2006)
54. Eichmair, M., Galloway, G.J., Pollack, D.: Topological censorship from the initial data point of view. J. Diff. Geomet. 95, 389–405 (2013)
55. Einstein, A.: Kritisches zu einer von Hrn. de Sitter gegebenen Lösung der Gravitationsgleichungen. Sitzungsber. Preuß. Akad. Wiss. Math. Kl. 270–272 (1918)
56. Einstein, A.: On a stationary system with spherical symmetry consisting of many gravitating masses. Ann. Math. 40, 922–936 (1939)
57. Einstein, A.: On a stationary system with spherical symmetry consisting of many gravitating masses. Ann. Math. 40, 922–936 (1939)
58. Einstein, A.: The Collected Papers of Albert Einstein, Vol. 7: The Berlin Years: Writings, 1918–1921. https://einsteinpapers.press.princeton.edu/vol7-trans/ (2002)
59. Eisenstaedt, J.: Lemaître and the Schwarzschild solution. In: The Attraction of Gravitation: New Studies in the History of General Relativity, eds. Earman, J., Janssen, M., Norton, J.D., pp. 353–389. Birkhäuser (1993)
60. Ellis, G.F.R.: Stephen Hawking’s 1966 Adams Prize Essay. Eur. J. Phys. H 39, 403–411 (2014)
61. Ellis, G.F.R., Maartens, R., MacCallum, M.A.H.: Relativistic Cosmology. Cambridge University Press (2012)
62. Event Horizon Telescope Collaboration: First M87 event horizon telescope results. I. The shadow of the supermassive black hole. Astrophys. J. Lett. 75:L1 (2019)
63. Fierz, M., Jost, R.: Affine Vollständigkeit und kompakte Lorentz’sche Mannigfaltigkeiten. Helv. Phys. Acta 38, 137–141 (1965)
64. Finkelstein, D.: Past-future asymmetry of the gravitational field of a point particle. Phys. Rev. 110, 956–967 (1958)
65. Flores, J.L., Sánchez, M.: Causality and conjugate points in general plane waves. Class. Quant. Gravity 20, 2275–2291 (2003)
66. Friedman, A.: Über die Krümmung des Raumes. Z. Phys. 10, 377–386 (1922). English translation: On the curvature of space. Gen. Relativ. Gravit. 31, 1991–2000 (1999)
67. Friedman, A.: Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. Z. Phys. 21, 326–332 (1924). English translation: On the possibility of a world with constant negative curvature of space. Gen. Relativ. Gravit. 31, 2001–2008 (1999)
68. Frolov, V.P.: Notes on nonsingular models of black holes. Phys. Rev. D 94, 104056 (2016)
69. Gajic, D., Luk, J.: The interior of dynamical extremal black holes in spherical symmetry. Pure Appl. Anal. 1, 263–326 (2019)
70. Galloway, G.J.: Some results on Cauchy surface criteria in Lorentzian geometry. Ill. J. Math. 29, 1–10 (1985)
71. Galloway, G.J.: Notes on Lorentzian causality. https://www.math.miami.edu/~galloway/vienna-course-notes.pdf (2014)
72. Galloway, G.J., Ming, P., Schoen, R.: Initial data and the Einstein constraint equations. In: General Relativity and Gravitation: A Centennial Perspective, eds. Ashtekar, A., Berger, B.K., Isenberg, J., MacCallum, M., pp. 412–448. Cambridge University Press (2015)
73. García-Parrado, A., Senovilla, J.M.M.: Causal structures and causal boundaries. Class. Quant. Gravity 22, R1–R84 (2005)
74. Geroch, R.: What is a singularity in General Relativity? Ann. Phys. 48, 526–540 (1968)
75. Geroch, R.: Limits of spacetimes. Commun. Math. Phys. 13, 180–193 (1969)
76. Geroch, R.: Domain of dependence. J. Math. Phys. 11, 437–449 (1970)
77. Geroch, R.: Singularities. In: Relativity, eds. Carmeli, M., Fickler, S., Witten, L., pp. 259–292. Plenum (1970)
78. Geroch, R., Horowitz, G.: Asymptotically simple does not imply asymptotically Minkowskian. Phys. Rev. Lett. 40, 203–206 (1978)
79. Geroch, R., Horowitz, G.: Global structure of spacetimes. In: General Relativity: An Einstein Centenary Survey, eds. Hawking, S.W., Israel, W., pp. 212–293. Cambridge University Press (1979)
80. Geroch, R., Kronheimer, E.H., Penrose, R.: Ideal points in space-time. Proc. Royal Soc. (London) A327, 545–567 (1972)
81. Geroch, R., Traschen, J.: Strings and other distributional sources in general relativity. Phys. Rev. D 36, 1017–1031 (1987)
82. Godart, O.: Contributions of Lemaître to General Relativity (1922–1934). In: Studies in the History of General Relativity, eds. Eisenstaedt, J., Kox, A.J., pp. 437–452. Birkhäuser (1992)
83. Gödel, K.: An example of a new type of cosmological solution of Einstein’s field equations of gravitation. Rev. Mod. Phys. 21, 447–450 (1949)
84. Gödel, K.: Rotating universes in general relativity theory. In: Proceedings of the International Congress of Mathematicians, Vol. 1, ed. Graves, L.M. et al., pp. 175–181 (1952). Reprinted in Gen. Relativ. Gravit. 32, 1419–1427 (2000)
85. Grattan-Guinness, I.: The mathematics of the past: distinguishing its history from our heritage. Hist. Math. 31, 163–185 (2004)
86. Grey, J.: Henri Poincaré: A Scientific Biography. Princeton University Press (2013)
87. Gundlach, C., Martin-Garcia, J.M.: Critical phenomena in gravitational collapse. Living Rev. Relativ. 10, 5 (2007)
88. Halvorson, H., Manchak, J.B.: Closing the Hole Argument. Br. J. Philos. Sci. in press. http://philsci-archive.pitt.edu/19790/ (2022)
89. Hawking, S.W.: Properties of Expanding Universes. PhD Thesis, University of Cambridge. https://www.repository.cam.ac.uk/handle/1810/251038 (1965)
90. Hawking, S.W.: Singularities in the universe. Phys. Rev. Lett. 17, 444–445 (1966)
91. Hawking, S.W.: Singularities and the geometry of space-time (Adams Prize Essay). Reprinted in Eur. J. Phys. H 39, 413–503 (2014) (1966)
92. Hawking, S.W.: Black holes in general relativity. Commun. Math. Phys. 25, 152–166 (1972)
93. Hawking, S.W., Ellis, G.F.R.: The Large Scale Structure of Space-Time. Cambridge University Press (1973)
94. Heckmann, O., Schücking, E.: Relativistic cosmology. In: Gravitation: An Introduction to Current Research, ed. Witten, L., pp. 438–469. Wiley (1962)
95. Hermann, R.: An incomplete compact homogeneous Lorentz metric. J. Math. Mech. 13, 497–501 (1964)
96. Hilbert, D.: Die Grundlagen der Physik (Zweite Mitteilung), Nachr. K. Ges. Wiss. Göttingen Math.-Phys. Kl. 53–76 (1917)
97. Hogarth, M.: Predictability, Computability, and Spacetime. PhD Thesis, University of Cambridge. http://www.hypercomputation.net/download/1996a_hogarth.pdf (1996)
98. Honnonkpe, R. A., Minguzzi, E.: Globally hyperbolic space-times can be defined without the ‘causal’ condition. Class. Quantum Grav. 36, 197001 (2019)
99. Israel, W.: Dark stars: The evolution of an idea. In: Three Hundred Years of Gravitation, eds. Hawking, S.W., Israel, W., pp. 199–276. Cambridge University Press (1987)
100. Janssen, M.: “No success like failure ...”: Einstein’s quest for general relativity, 1907–1920. In: The Cambridge Companion to Einstein, eds. Janssen, M., Lehner, C., pp. 167–227. Cambridge University Press (2014)
101. Janssen, M., Renn, J.: How Einstein Found His Field Equations: Sources & Interpretation. Springer (2022)
102. Jones, B.J.T.: Precision Cosmology: The First Half Million Years. Cambridge University Press (2017)
103. Joshi, P.S.: Gravitational Collapse and Spacetime Singularities. Cambridge University Press (2007)
104. Kehle, C., Van de Moortel, M.: Strong Cosmic Censorship in the presence of matter: the decisive effect of horizon oscillations on the black hole interior geometry. arXiv:2105.04604 (2021)
105. Komar, A.: Necessity of singularities in the solution of the field equations of general relativity. Phys. Rev. 104, 544–546 (1956)
106. Kragh, H.: Conceptions of the Cosmos. A History of Cosmology: From Myths to the Accelerating Universe. Oxford University Press (2007)
107. Kriele, M.: Spacetime: Foundations of General Relativity and Differential Geometry. Springer (1999)
108. Królak, A.: Towards the proof of the cosmic censorship hypothesis. Class. Quant. Gravity 3, 267–280 (1986)
109. Królak, A.: Cosmic censorship hypothesis. Contemp. Math. 359, 51–64 (2004)
110. Kundt, W.: Note on the completeness of spacetimes. Z. Phys. 172, 488–489 (1963)
111. Landsman, K.: Foundations of Quantum Theory: From Classical Concepts to Operator Algebras. Springer (2017)
112. Landsman, K.: Foundations of General Relativity: From Einstein to Black Holes. Radboud University Press (2021)
113. Landsman, K.: Singularities, black holes, and cosmic censorship: A tribute to Roger Penrose. Found. Phys. 51, 42 (2021)
114. Landsman, K.: Indeterminism and undecidability. In: Undecidability, Uncomputability, and Unpredictability, eds. Aguirre, A., Merali, Z., Sloan, D., pp. 17–46. Springer (2021). arXiv:2003.03554
115. Landsman, K.: Reopening the hole argument. http://philsci-archive.pitt.edu/20399/ (2022)
116. Landsman, K.: Indeterminism and undecidability. In: Undecidability, Uncomputability, and Unpredictability, eds. Aguirre, A., Merali, Z., Sloan, D., pp. 17–46. Springer (2021). arXiv:2003.03554
117. Lee, D.A.: Geometric Relativity. American Mathematical Society (2019)
118. Lemaître, G.: Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. Ann. Soc. Sci. Bruxelles A 47, 49–59 (1927)
119. Lemaître, G.: L’univers en expansion. Ann. Soc. Sci. Bruxelles A 53, 51–85 (1933). English translation: The expanding Universe, In: Gen. Relativ. Gravit. 29, 641–680 (1997), with editorial note and biography by Andrzej Krasiński, pp. 637–40
120. Leray, J.: Hyperbolic Differential Equations. Mimeographed Lecture Notes. The Institute for Advanced Study (1953)
121. Levi-Civita, T.: Rend. R. Acc. Lincei 26, 519–31 (1917)
122. Li, J., Yu, P.: Construction of Cauchy data of vacuum Einstein field equations evolving to black holes. Ann. Math. 181, 699–768 (2015)
123. Longair, M.S.: The Cosmic Century: A History of Astrophysics and Cosmology. Cambridge University Press (2006)
124. Luk, J.: Weak null singularities in general relativity. J. AMS 31, 1–63 (2017)
125. Luk, J., Oh, S.-J.: Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat Initial data I: the interior of the black hole region. Ann. Math. 190, 1–111. II: The exterior of the black hole region. Ann. PDE 5(6) (2019)
126. Manchak, J.B.: On the inextendibility of space-time. Philos. Sci. 84, 1215–1225 (2017)
127. Manchak, J.B.: Malament-Hogarth machines. Br. J. Philos. Sci. 71, 1143–1153 (2020)
128. Mars, M.: Present status of the Penrose inequality. Class. Quant. Gravity 26, 193001 (2009)
129. Matzner, R.A., Shepley, L.C., Warren, J.B.: Dynamics of SO(3, R)-homogeneous cosmologies. Ann. Phys. 57, 401–460 (1970)
130. McNamara, J.M.: Instability of black hole inner horizons. Proc. Royal Soc. A 358, 499–517 (1978)
131. Minguzzi, E.: Lorentzian causality theory. Living Rev. Relativ. 22, 3 (2019)
132. Misner, C.W.: The flatter regions of Newman, Unti, and Tamburino’s generalized Schwarzschild space. J. Math. Phys. 4, 924–937 (1963)
133. Misner, C.W., Taub, A.H.: A singularity-free empty universe. Soviet Phys. JETP 55, 233–255 (1969)
134. Moncrief, V.: Global properties of Gowdy space-times with $T^3 \times \mathbb{R}$ topology. Ann. Phys. 132, 87–107 (1981)
135. Netz, R.: The Shaping of Deduction in Greek Mathematics: A Study in Cognitive History. Cambridge University Press (1999)
136. Newman, E., Penrose, R.: An approach to gravitational radiation by a method of spin coefficients. J. Math. Phys. 3, 566–578 (1962)
137. Nussbaumer, H.: Einstein’s conversion from his static to an expanding universe. Eur. Phys. J. H. 39, 37–62 (2014)
138. Nussbaumer, H., Bieri, L.: Discovering the Expanding Universe. Cambridge University Press (2009)
139. O’Neill, B.: Semi-Riemannian Geometry. Academic Press (1983)
140. Ong, Y.C.: Space-time singularities and cosmic censorship conjecture: a review with some thoughts. Int. J. Mod. Phys. A 35, 2030007 (2020)
141. Oppenheimer, J.R., Volkoff, G.M.: On massive neutron cores. Phys. Rev. 55, 374–381 (1939)
142. Oppenheimer, J.R., Snyder, H.: On continued gravitational contraction. Phys. Rev. 56, 455–459 (1939)
143. Penrose, R.: Null hypersurface initial data for classical fields of arbitrary spin and for general relativity. Aerospace Research Laboratories 63–65 (1963). Reprinted in Gen. Relativ. Gravit. 12, 225–264 (1980)
144. Penrose, R.: Asymptotic properties of fields and space-times. Phys. Rev. Lett. 10, 66–68 (1963)
145. Penrose, R.: Conformal treatment of infinity. In: Relativity, Groups, and Topology, eds. DeWitt, B., DeWitt-Morette, C.M., pp. 565–584 Gordon & Breach (1964). Reprinted in Gen. Relativ. Gravit. 43, 901–922 (2011)
146. Penrose, R.: A remarkable property of plane waves in general relativity. Rev. Mod. Phys. 37, 215–220 (1965)
147. Penrose, R.: Zero rest-mass fields including gravitation: asymptotic behaviour. Proc. Royal Soc. London A 284, 159–203 (1965)
148. Penrose, R.: Gravitational collapse and space-time singularities. Phys. Rev. Lett. 14, 57–59 (1965)
149. Penrose, R.: An analysis of the structure of space-time. Adams Prize Essay (1966). In: Roger Penrose: Collected Works, Volume 1: 1953–1967, pp. 579–730. Oxford University Press (2011)
150. Penrose, R.: Structure of space-time. In: Battelle Rencontres: 1967 Lectures in Mathematics and Physics, eds. DeWitt, C., Wheeler, J.A., pp. 121–235. W.A. Benjamin (1968)
151. Penrose, R.: Gravitational collapse: The role of general relativity. Riv. Nuovo Cim. Numero Speciale I, 252 (1969). Reprinted in Gen. Relativ. Gravit. 34, 1141–1165 (2002)
152. Penrose, R.: Techniques of Differential Topology in Relativity (SIAM) (1972)
153. Penrose, R.: Naked singularities. Ann. N. Y. Acad. Sci. 224, 125–134 (1973)
154. Penrose, R.: Gravitational collapse. In: Gravitational Radiation and Gravitational Collapse, ed. DeWitt-Morette, C., pp. 82–91. International Astronomical Union (1974)
155. Penrose, R.: Singularities and time-asymmetry. In: General Relativity: An Einstein Centenary Survey, eds. Hawking, S.W., Israel, W., pp. 581–638. Cambridge University Press (1979)
156. Penrose, R.: The question of cosmic censorship. J. Astrophys. Astron. 20, 233–248 (1999)
157. Penrose, R., Sorkin, R.D., Woolgar, E.: A positive mass theorem based on the focusing and retardation of null geodesics. arXiv:gr-qc/9301015 (1993)
158. Pflaum, M.: Analytic and Geometric Study of Stratified Spaces. Springer (2001)
159. Raychaudhuri, A.K.: Relativistic cosmology. I. Phys. Rev. 98, 1123–1126 (1955)
160. Raychaudhuri, A.K.: Singular state in relativistic cosmology. Phys. Rev. 106, 172–173 (1957)
161. Rindler, W.: Visual horizons in world models. Mon. Not. R. Astron. Soc. 116, 662–677 (1956)
162. Ringström, H.: The Cauchy Problem in General Relativity. EMS (2009)
163. Ringström, H.: Cosmic censorship for Gowdy spacetimes. Living Rev. Relativ. 13, 2 (2010)
164. Ringström, H.: Origins and development of the Cauchy problem in general relativity. Class. Quant. Gravity 32, 124003 (2015)
165. Robertson, H.P.: Relativistic cosmology. Rev. Mod. Phys. 5, 62–90 (1933)
166. Robinson, D.C.: Gravitation and general relativity at King’s College London. Eur. Phys. J. H 44, 181–207 (2019)
167. Ryan, M.P., Shepley, L.C.: Homogeneous Relativistic Cosmologies. Princeton University Press (1975)
168. Saari, D.G., Xia, Z.: Off to infinity in finite time. Notices AMS 42, 538–546 (1995)
169. Sbierski, J.: On the existence of a maximal Cauchy development for the Einstein equations: a deornification. Ann. Henri Poincaré 17, 301–329 (2016)
170. Sbierski, J.: The C0-inextendibility of the Schwarzschild space-time and the spacelike diameter in Lorentzian geometry. J. Differ. Geom. 108, 319–378 (2018)
171. Schoen, R., Yau, S.T.: Proof of the positive mass theorem. II. Commun. Math. Phys. 79, 231–260 (1981)
172. Schoen, R., Yau, S.T.: The existence of a black hole due to condensation of matter. Commun. Math. Phys. 65, 575–579 (1983)
173. Senovilla, J.M.M.: Singularity theorems and their consequences. Gen. Relativ. Gravit. 30, 701–848 (1998). Corrected version: arXiv:1801.04912
174. Senovilla, J.M.M.: Trapped surfaces. Int. J. Mod. Phys.: Conf. Series 7, 1–30 (2012)
175. Senovilla, J.M.M., Garfinkle, D.: The 1965 Penrose singularity theorem. Class. Quant. Gravity 32, 124008 (2015)
176. Shepley, L.C.: Singularities in spatially homogeneous, dust-filled universes. Proc. Natl. Acad. Sci. 52, 1403–1409 (1964)
177. Shepley, L.C.: $SO(3, R)$-Homogeneous Cosmologies. PhD Thesis, Princeton University (1965)
178. Simpson, M., Penrose, R.: Internal instability in a Reissner-Nordström black hole. Int. J. Theor. Phys. 7, 183–197 (1973)
179. Simpson, A., Visser, M.: The eye of the storm: a regular Kerr black hole. J. Cosmol. Astropart. Phys. 03, 011 (2022)
180. Sitter, W. de.: On the relativity of inertia: Remarks concerning Einstein’s latest hypothesis. Proc. R. Netherlands Acad. Arts Sci. 19, 1217–1225 (1917). https://archive.org/details/proceedingsofsec192koni/page/1216/mode/2up
181. Smeenk, C.: Einstein’s role in the creation of relativistic cosmology. In: The Cambridge Companion to Einstein, eds. Janssen, M., Lehner, C., pp. 228–269. Cambridge University Press (2014)
182. Smeenk, C., Wüthrich, C.: Determinism and general relativity. Philos. Sci. 88, 638–664 (2021)
183. Spallucci, E., Smalagic, A.: Regular black holes from semi-classical down to Planckian size. Int. J. Mod. Phys. D 26, 1730013 (2017)
184. Stachel, J.: The Cauchy problem in general relativity–The early years. In: Studies in the History of General Relativity, eds. Eisenstaedt, J., Kox, A.J., pp. 407–418. Birkhäuser (1992)
185. Taub, A.H.: Empty space-times admitting a three parameter group of motions. Ann. Math. 53, 472–490 (1951). Reprinted in Gen. Relativ. Gravit. 36, 2699–2719 (2004)
186. Thorne, K.S.: Black Holes and Time Warps: Einstein’s Outrageous Legacy. W.W. Norton (1994)
187. Tipler, F.J., Clarke, C.J.S., Ellis, G.F.R.: Singularities and horizons–A review article. In: General Relativity and Gravitation: One Hundred Years After the Birth of Albert Einstein, Vol. 2, ed. Held, A., pp. 97–206. Plenum Press (1980)
188. Tolman, R.C., Ward, M.: On the behavior of non-static models of the universe when the cosmological term is omitted. Phys. Rev. 39, 835–843 (1932)
189. Wald, R.M.: General Relativity. University of Chicago Press (1984)
190. Walker, A.G.: On Milne’s theory of world-structure. Proc. London Math. Soc. Series 2(42), 90–127 (1937)
191. Weyl, H.: Raum - Zeit - Materie: Vorlesungen über Allgemeine Relativitätstheorie. Springer (1918)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.